


FULL PAPER

Open Access



Segmentation characteristics of deep, low-frequency tremors in Shikoku, Japan using machine learning approaches

Kate Huihsuan Chen^{1*} , Hao-Yu Chiu², Kazushige Obara³ and Yi-Hung Liu⁴

Abstract

Shikoku island, southwestern Japan lies in the western Nankai Trough and showcases along-strike segmentation of slow earthquake behavior. Whether the spatial variation of tremor behavior reflects the regional differences in structure/source properties and how much such differences can be recognized by the seismic signals themselves are two questions addressed in this paper. Taking advantage of advanced methods in recognizing and classifying signals using machine learning approaches, we attempt to answer them by conducting signal classification experiments in Shikoku. Based on the tremor catalog from 1 June 2014 to 31 March 2015, the tremors recorded in four different areas were treated as different classes and segmented into 60-s-long signals. The number of tremors in four different areas (A to D, from west to east) reached 15,000, 31,000, 10,000, and 16,000, respectively. To efficiently distinguish between tremors from different areas, we applied a k -nearest neighbor (k -NN) classifier with Fisher's class separability criteria to select the optimal feature subset. The resulting classification performance reached more than 90% at all 12 stations. We further designed a triangle test to select the features that can better represent the differences in source properties between areas. We found that the most efficient features were associated with (1) the number of peaks in the temporal evolution of discrete Fourier transforms and (2) the energy distribution in the autocorrelation function (ACF). To match the difference in behavior revealed by the ACF, the size of the tremor zone, which mainly controls how long the seismic energy lasts in a tremor episode, was determined to be largest in Area B and smallest in Area C. The heterogeneity of the asperities in a tremor zone, which may control how spiky the tremor signals developed over time, was determined to be strong in Areas B and C. Together with previously documented variations in slow earthquake behavior in the same area, we finally propose a conceptual model that provides a better understanding of the regional differences in the tremor source properties in Shikoku, Japan.

Keywords Machine learning, Signal classification, Deep low-frequency tremor, k -nearest neighbor, Hi-net, Shikoku

*Correspondence:

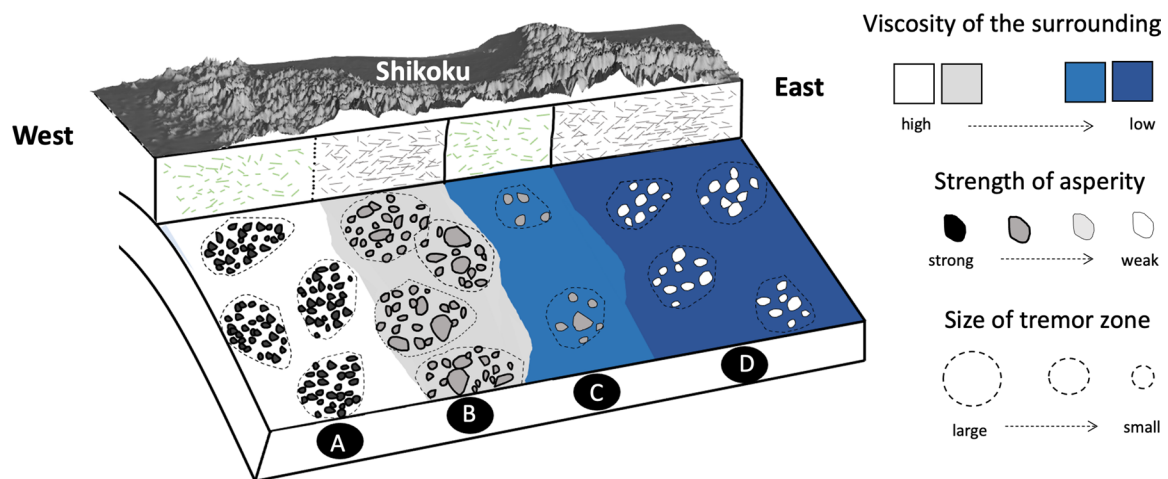
Kate Huihsuan Chen
katepill@gmail.com

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

Graphical Abstract

Segmentation characteristics of deep, low-frequency tremors in Shikoku, Japan using machine learning approaches

Kate Huihsuan Chen, Hao-Yu Chiu, Kazushige Obara, Yi-Hung Liu



Introduction

Deep low-frequency tremor characterized by dominant frequencies of ~ 2 – 10 Hz is a member of slow earthquake family. With the lack of emergent P and S seismic wave arrivals, long-lasting but small-amplitude waveform characters, they have been recognized as representing slow slip processes below the seismogenic zone. The deep low-frequency tremors have been discovered in a variety of tectonic settings, including subduction zones, transform faults, and mountain belts (Obara 2002; Nadeau and Dolenc 2005; Ide et al. 2007; Kao et al. 2007; Shelly et al. 2006; Wech and Creager 2008; Nadeau and Guilhem 2009; Ide 2012; Wech et al. 2012; Chuang et al. 2014; Chen et al. 2018). Due to the strong spatiotemporal correlation between tremors and geodetically inferred slow slip events (e.g., Rogers and Dragert 2003; Obara and Hirose 2006; Brudzinski and Allen 2007; Obara and Sekine 2009), tremors are regarded as a seismic signature of aseismic slip. The observation of tremors, therefore, provides crucial information about the strain accumulation at greater depths, especially in mountainous and offshore areas where ground-based geodetic data are not available or do not have sufficiently high resolution.

The tremors in Shikoku, southwestern Japan, exhibit remarkable diversity in their slow slip phenomena in terms of spatial characteristics, radiated energy, duration, and tremor migration patterns (Ide 2008, 2010;

Obara 2010; Obara et al. 2010, 2011; Idehara et al. 2014; Annoura et al. 2016; Kano et al. 2018; Hirose and Kimura 2020; Supino et al. 2021). Such observations provide a great opportunity to establish the along-strike variation in physical properties on the subducting plate interface. Obara (2010) separated the tremors in Shikoku into western, central, and eastern segments (black dashed line in Fig. 1a) based on their different spatial distributions, slip rates, and migration patterns. Unlike the strong association with tidal stress found in eastern Shikoku, the tremors in western Shikoku were not correlated with tides (Nakata et al. 2008). The total seismic energy of the tremors was found to be larger in western Shikoku compared with eastern Shikoku (Annoura et al. 2016). The spatiotemporal variation in episodic tremor and slip (ETS) also divides Shikoku into three segments (Hirose and Obara 2005; Sekine et al. 2010; Hirose and Kimura 2020), suggesting a strong regional difference in mechanical properties. Based on precisely located ETS of tremors, the combination of central and western Shikoku first demarcated by Obara (2010) was further divided into another three segments by Kano et al. (2018) (yellow dashed line in Fig. 1a). In total, four segments in Shikoku are considered in this study, as shown by the red line in Fig. 1a.

Two key questions were of interest in this study: Does the spatial variation of tremor behavior reflect regional

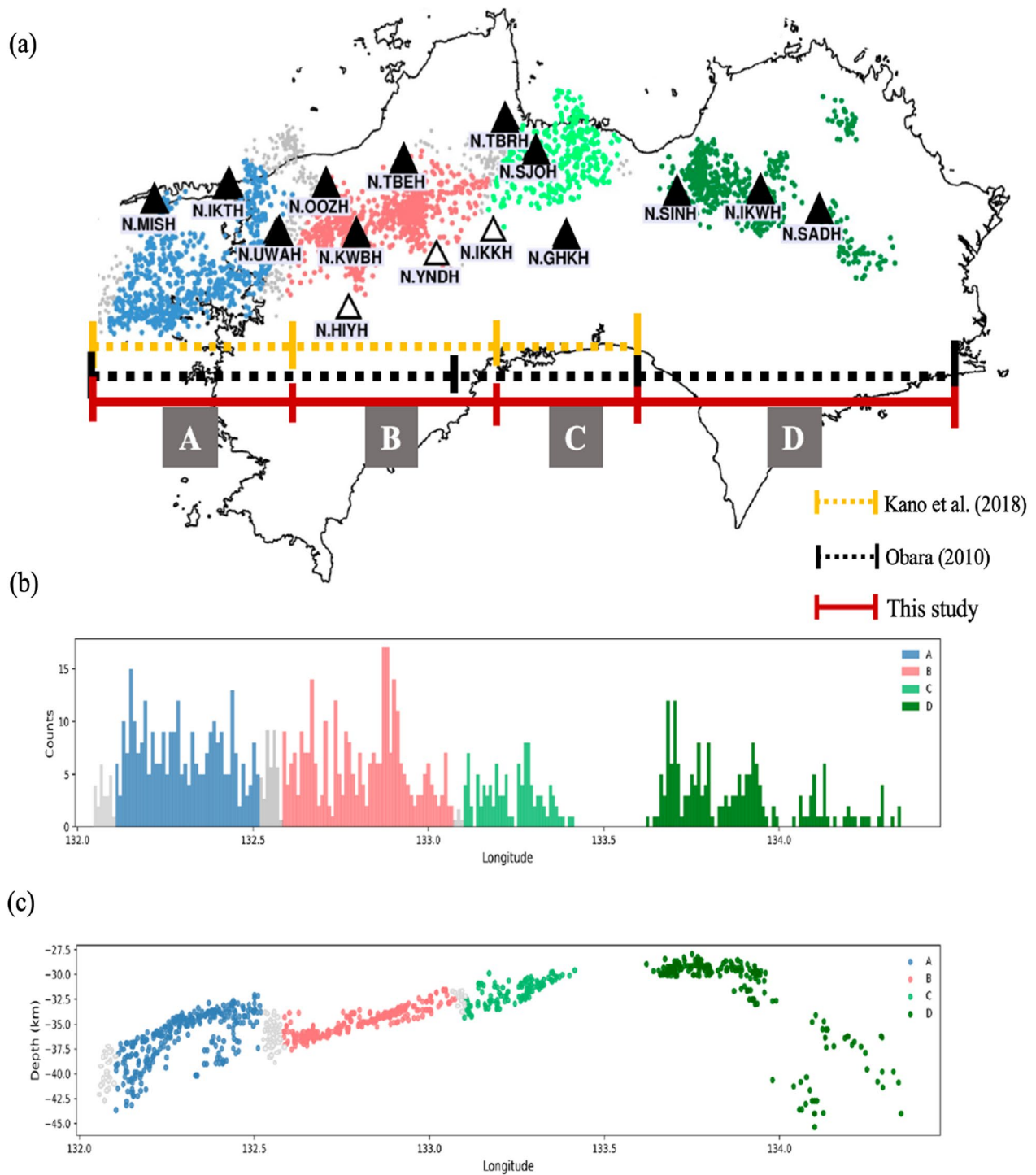


Fig. 1 **a** Distribution of stations and tremor events catalogued in this study. Blue, light coral, light green, and dark green dots represent the labeled tremor events in the three sub-areas of western Shikoku (western, central, and eastern sub-areas) and eastern Shikoku, respectively (Areas A to D). The colored lines indicate different divisions: Obara (2010) in black, Kano et al. (2018) in yellow, and this study in red. **b** The number of tremor event along the longitude. **c** The depth distribution of tremor events along the longitude. Events belong to four different areas are denoted by different colors

differences in structure and source property? To what extent can such differences be recognized by the seismic signals themselves? Through advanced methods of recognizing and classifying signals using machine learning approaches, we attempted to answer these questions by conducting signal classification experiments in Shikoku. We first demonstrated the relevant regional characteristics using a tremor catalog published by Annoura et al. (2016). By evaluating their classification performance using a supervised classifier, k -nearest neighbor (k -NN), we explored the seismic features that allow us to discriminate one subarea from another. We also design a triangle test for features extraction for a better presentation of the differences in source properties between areas. The most efficient features discovered here are likely associated with the heterogeneity of the asperities in a tremor zone and the size of tremor zone. Together with previously documented variations in slow earthquake behavior in the same area, we finally proposed a conceptual model that provides a better understanding of the regional differences in the tremor sources of Shikoku.

Data labeling and classifier

Using the tremor catalog published by Annoura et al. (2016), we selected 878 tremor events that featured continuous seismic waveforms from 15 Hi-net stations (black and open triangles in Fig. 1a) (e.g., Okada et al. 2004; Obara et al. 2005). Following Kano et al. (2018), the tremors in Shikoku were classified into four areas from A to D, as shown by the differently colored dots in Fig. 1a. During the study period from 1 June 2014 to 31 March 2015, the number of tremor events in A to D were 289, 299, 106, and 184, corresponding to total durations of 2220 s, 2145 s, 1646 s, and 1755 s, respectively. The number of events and depth ranges are shown in Fig. 1b, c. The events were distributed from depths of 27.5 km to 45.0 km, and a tremor gap appeared between Areas C and D. By visual inspection, 12 stations showing clear tremor signals were selected, as listed in Table 1 denoted by black triangles in Fig. 1a).

Estimates of kernel density for event duration and root mean square (RMS) amplitude are shown in Fig. 2. We

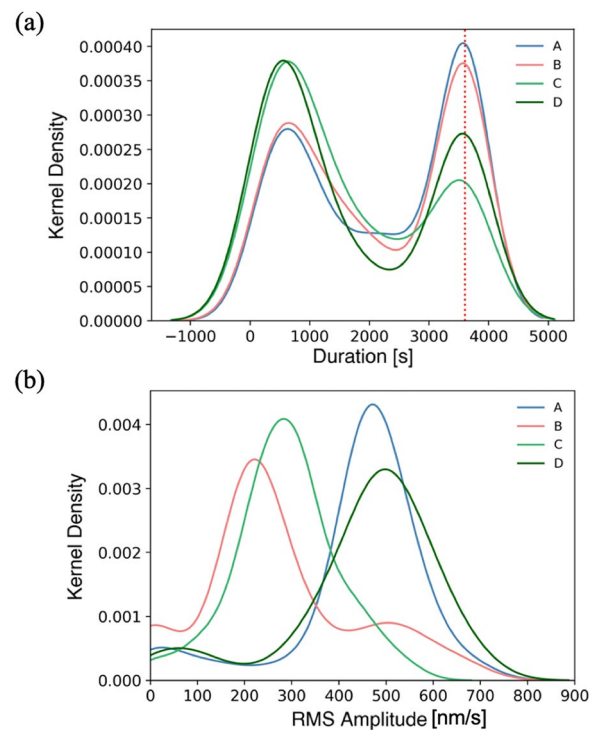


Fig. 2 The kernel density of **a** event duration and **b** RMS amplitude. The red dashed line indicates the longest duration of 3600 s

found that Areas C and D tended to have shorter duration tremors, whereas Areas A and B had longer duration tremors (Fig. 2a). Regarding RMS amplitude, however, Areas A and D appeared to show larger amplitudes of tremors compared with Areas B and C. In other words, Area A tended to have the largest number of long-lasting tremors with relatively large amplitude; Area B is also characterized by more long lasting tremors, but with the smallest amplitudes.

To build the training and test data sets in the following analyses, the seismic data were segregated into 60-s segments and examined visually. The total number of 60-s-long tremors in each area is given in Table 1. Tremors in the four areas were treated as different classes. Each tremor event class was split into training (70%) and testing (30%) datasets. This hold-out method allowed us to find an efficient model parameter and hyper-parameter for better generalization performance. We used a simple classifier, k -NN (k -nearest neighbor), that reflects the classification capability of the seismic feature and is often used as a benchmark for further complex classifiers. The nearest neighbors of a given test datum are defined in terms of the standard Euclidean distance (Mitchell 1997). The k -NN classifier involves a simple and non-parametric approach, which is easy to implement for multi-class problems and no assumptions are needed. It

Table 1 Four subareas in the Shikoku region (A to D)

| Area | Stations used in this study | Total number of labeled data |
|------|-----------------------------|------------------------------|
| A | N.MISH, N.IKTH, N.UWAH | 15,000 |
| B | N.OOZH, N.TBEH, N.KWBH | 31,000 |
| C | N.TBRH, N.SJOH, N.GHKH | 10,000 |
| D | N.IKWH, N.SADH, N.SINH | 16,000 |

70% of the labeled data were for training; 30% were for testing

is a memory-based approach that can adapt new training data and is thus useful for real-time applications. The k -NN classification can be implemented by (1) preparing the N -dimensional data point in the training set for the M different classes (2) calculating the standard Euclidean distance between a test data point to the k nearest training data point that were previously labeled into M classes (3) deciding which class this test data point belongs to based on the majority of the neighbors. The only free parameter is k , which is an odd number. The sensitivity test on six binary models shows a general trend of decreasing model accuracy with increasing k , while the maximum accuracy occurs at $k=5$. Therefore the k is set as 5 in the present study.

The performance of the classifier was evaluated using leave-one-out cross validation (LOO) (Efron and Tibshirani 1993). In the training stage, LOO uses a single observation from a given data set as the test data, and the remaining observations are used as the training data. The classification rate (CR) is obtained when each observation in the data set has been treated as the test data.

Seismic features

The signal features were chosen based on the seismic signatures commonly exploited for event classification. To train this model, 29 seismic features (Table 2) were computed based on the characteristics of the temporal waveforms (first family), spectral content (second family), and energy concentration of the frequency and time (third family). Twenty-seven of the features were the same as those Liu et al. (2019) selected for discriminating tremors from local earthquake and noise, which mainly referred to Provost et al. (2016) and Hibert et al. (2017) with regard to the features used for the signal classification of rockfalls, earthquakes, and noise.

By analyzing each 60-s labeled datapoint, we obtained an $N \times 29$ feature matrix (N representing the event number). The 29 features belonged to three separate families, their expressions and descriptions are shown in Table 2 (more details can be seen in the supplementary material (Additional file 1: Figs. S1–4). In the 1st family we considered the amplitude, skewness, kurtosis, and autocorrelation function of the waveforms. In the 2nd family we focused on spectral characteristics including maximum amplitude, number of peaks, and the total energy, and kurtosis of the signals. In the 3rd family we computed the energy concentration in frequency and time.

To evaluate the classification performance of the seismic features, two approaches were used for feature extraction. Sequential forward selection (SFS), which aims at selecting a subset of features that achieves the highest classification accuracy, was conducted first. To reduce feature dimensions and select the most

representative features for each area, evaluating the classification performance for an individual feature was also needed. Fisher's class separability method (Fang et al. 2015; Liu et al. 2017), which computes the ratio of inter-class scatter and intraclass scatter, was adopted for this purpose. This ratio is referred as Fisher score and abbreviated to F -score hereafter. In this study, an F -score higher than 0.2 was designated as representing strong separability between classes. This determined the correspondence between the p -value (a commonly used statistical value that describes how likely your data have occurred by random chance) and F -score for all tests in this study. As a p value > 0.05 indicates strong evidence for the null hypothesis, an F -score > 0.2 was found to correspond with a p -value < 0.05 .

Experimental design for feature extraction

For the purpose of signal discrimination, models based on the k -NN classifier were trained to distinguish the tremors from different areas of Shikoku (Areas A to D). To identify the efficient features that better represented the properties of the tremors in each area, we adopted binary classification and developed a set of experiments. The four classes of tremors belonging to Areas A to D are represented by T_A , T_B , T_C , and T_D as the source effect, whereas the three stations located in each subarea are abbreviated as S_A , S_B , S_C , and S_D as the site effect. The expression of T_X-S_Y on the other hand, shows the consideration of path effect (X and $Y = A$ or B or C or D).

As shown in Fig. 3a, Experiment-1 considers the signals from the source to the stations in the same area, as denoted by T_A-S_A , T_B-S_B , T_C-S_C , and T_D-S_D . There exists six combination of two-class pairs (tests): T_A-S_A vs T_B-S_B , T_A-S_A vs T_C-S_C , T_A-S_A vs T_D-S_D , T_B-S_B vs T_C-S_C , T_B-S_B vs T_D-S_D , and T_C-S_C vs T_D-S_D (Tests 1–6 in Table 3). Experiment-2 considered the signals from the sources in T_A to various stations S_A , S_B , S_C , and S_D , aiming to identify the differences between T_A-S_A , T_A-S_B , T_A-S_C , and T_A-S_D (blue dashed lines in Fig. 3b). Thus there were three tests in Experiment-2, T_A-S_A vs T_A-S_B , T_A-S_A vs T_A-S_C , and T_A-S_A vs T_A-S_D , which are listed as Tests 7–9 in Table 3. Experiment-3 to Experiment-5 followed the same protocol as Experiment-2 but with the source areas set as T_B , T_C , and T_D , respectively, with the corresponding descriptions listed in Table 3 (Tests 10–18).

Note that comparing with NS and vertical components, the waveforms in EW component reveal slightly better classification performance in our test when 100 tremor events were randomly selected and the 29 features (Table 2) were computed for Test 1 in Experiments 1 (83.7%, 80.2%, and 77.8% for EW, NS, and Z components, respectively). Therefore, we measured the classification performance using EW component of the data

Table 2 Computed seismic features in this study

| ID | Expression | Description |
|---|--|---|
| 1st family | | |
| 1. <i>EnvMean/Max</i> | $f_1 = \frac{\text{mean}(\text{Env})}{\text{max}(\text{Env})}$ | Ratio of the mean over the max of the envelope signal |
| 2. <i>EnvMed/Max</i> | $f_2 = \frac{\text{median}(\text{Env})}{\text{max}(\text{Env})}$ | Ratio of the median over the max of the envelope signal |
| 3. <i>EnvMax</i> | $f_3 = \text{max}(\text{Env})$ | Maximum envelope amplitude |
| 4. <i>EnvMaxTime</i> | $f_4 = \arg \text{max}(\text{Env})$ | Time of the maximum envelope amplitude |
| 5. <i>RawKurt</i> | $f_5 = \left(\frac{\mu_{\text{Raw}}}{\sigma_{\text{Raw}}}\right)^4$ | Kurtosis of the raw signal |
| 6. <i>EnvKurt</i> | $f_6 = \left(\frac{\mu_{\text{Env}}}{\sigma_{\text{Env}}}\right)^4$ | Kurtosis of the envelope |
| 7. <i>RawSkew</i> | $f_7 = \left(\frac{\mu_{\text{Raw}}}{\sigma_{\text{Raw}}}\right)^3$ | Skewness of the raw signal |
| 8. <i>EnvSkew</i> | $f_8 = \left(\frac{\mu_{\text{Env}}}{\sigma_{\text{Env}}}\right)^3$ | Skewness of the envelope |
| 9. <i>ACF_{1/3}</i> | $f_9 = \int_1^{2/3} C(\tau) d\tau$ | Energy in the first third of the autocorrelation function |
| 10. <i>ACF_{2/3}</i> | $f_{10} = \int_{2/3}^1 C(\tau) d\tau$ | Energy in the remaining part of the autocorrelation function |
| 11. <i>ACF_{1/3}/ACF_{2/3}</i> | $f_{11} = \frac{f_9}{f_{10}}$ | Ratio of the above two |
| 2nd family | | |
| 12. <i>Max BP_{2–8Hz}</i> | $f_{12} = \text{max}(BP_{2-8\text{Hz}})$ | Maximum amplitude of the 2–8 Hz filtered signal |
| 13. <i>NPks BP_{2–8Hz}</i> | $f_{13} = \text{length}(\text{findpeaks}(\text{Env}_{BP}))$ | Number of peaks in the envelope filtered by 2–8 Hz |
| 14. <i>BP_{0.1–1Hz}</i> | $f_{14} = \int_0^T BP_{0.1-1\text{Hz}}(t) dt$ | Energy of the signal filtered by 0.1–1 Hz |
| 15. <i>BP_{2–8Hz}</i> | $f_{15} = \int_0^T BP_{2-8\text{Hz}}(t) dt$ | Energy of the signal filtered by 2–8 Hz |
| 16. <i>BP_{5–20Hz}</i> | $f_{16} = \int_0^T BP_{5-20\text{Hz}}(t) dt$ | Energy of the signal filtered by 5–20 Hz |
| 17. <i>Kurt BP_{0.1–1Hz}</i> | $f_{17} = \left(\frac{\mu_{BP1}}{\sigma_{BP1}}\right)^4$ | Kurtosis of the signal filtered by 0.1–1 Hz |
| 18. <i>Kurt BP_{2–8Hz}</i> | $f_{18} = \left(\frac{\mu_{BP2}}{\sigma_{BP2}}\right)^4$ | Kurtosis of the signal filtered by 2–8 Hz |
| 19. <i>Kurt BP_{5–20Hz}</i> | $f_{19} = \left(\frac{\mu_{BP3}}{\sigma_{BP3}}\right)^4$ | Kurtosis of the signal filtered by 5–20 Hz |
| 3rd family | | |
| 20. <i>KurtDFT</i> | $f_{20} = \text{Kurtosis}(\max_{t=0 \dots T} [\text{SPEC}(t, f)])$ | Kurtosis of the maximum of all DFT as a function of time |
| 21. <i>DFTmax/mean</i> | $f_{21} = \frac{\max_{t=0 \dots T} [\text{SPEC}(t, f)]}{\text{mean}_{t=0 \dots T} [\text{SPEC}(t, f)]}$ | Mean ratio between the maximum and the mean of all DFT |
| 22. <i>DFTmax/med</i> | $f_{22} = \frac{\max_{t=0 \dots T} [\text{SPEC}(t, f)]}{\text{median}_{t=0 \dots T} [\text{SPEC}(t, f)]}$ | Mean ratio between the maximum and the median of all DFT |
| 23. <i>NPks_DFTmax</i> | $f_{23} = \text{length}(\text{findpeaks}(\max_{t=0 \dots T} [\text{SPEC}(t, f)]))$ | Number of peaks in the curve showing the temporal evolution of the DFT max |
| 24. <i>NPks_DFTmean</i> | $f_{24} = \text{length}(\text{findpeaks}(\text{mean}_{t=0 \dots T} [\text{SPEC}(t, f)]))$ | Number of peaks in the curve showing the temporal evolution of the DFT mean |
| 25. <i>NPks_DFTmed</i> | $f_{25} = \text{length}(\text{findpeaks}(\text{median}_{t=0 \dots T} [\text{SPEC}(t, f)]))$ | Number of peaks in the curve showing the temporal evolution of the DFT median |
| 26. <i>NPks_DFTmax/mean</i> | $f_{26} = \frac{f_{23}}{f_{24}}$ | Ratio between #23 and #24 |
| 27. <i>NPk_DFTmax/med</i> | $f_{27} = \frac{f_{23}}{f_{25}}$ | Ratio between #23 and #25 |
| 28. <i>Sum BP_{2–8Hz}/BP_{5–20Hz}</i> | $f_{28} = \frac{\sum_{2 \leq f \leq 8} (\text{STFT}(t, f))}{\sum_{5 \leq f \leq 20} (\text{STFT}(t, f))}$ | Ratio between the sum of energy in 2–8 Hz and the sum of energy in 5–20 Hz |
| 29. <i>Sum BP_{2–8Hz}/BP_{5–50Hz}</i> | $f_{29} = \frac{\sum_{5 \leq f \leq 20} (\text{STFT}(t, f))}{\sum_{5 \leq f \leq 50} (\text{STFT}(t, f))}$ | Ratio between the sum of energy in 5–20 Hz and higher than 5 Hz |

from each station in the following analysis. The classification performance was first determined using SFS, which adds features from a dataset sequentially, to minimize the number of overall features in a subset (Whitney 1971). This was done by predefining the number of features N ($N < 29$) at each timestep and evaluating the classification

performance in each test. To meet the optimal performance requirements, the maximum classification rate for each test is obtained and shown in Table 3. The SFS for the 18 tests produced a classification rate (CR) of 89.3% to 99.5% (Fig. 4 and Table 3), showing the effectiveness of the proposed architecture and feature choices.

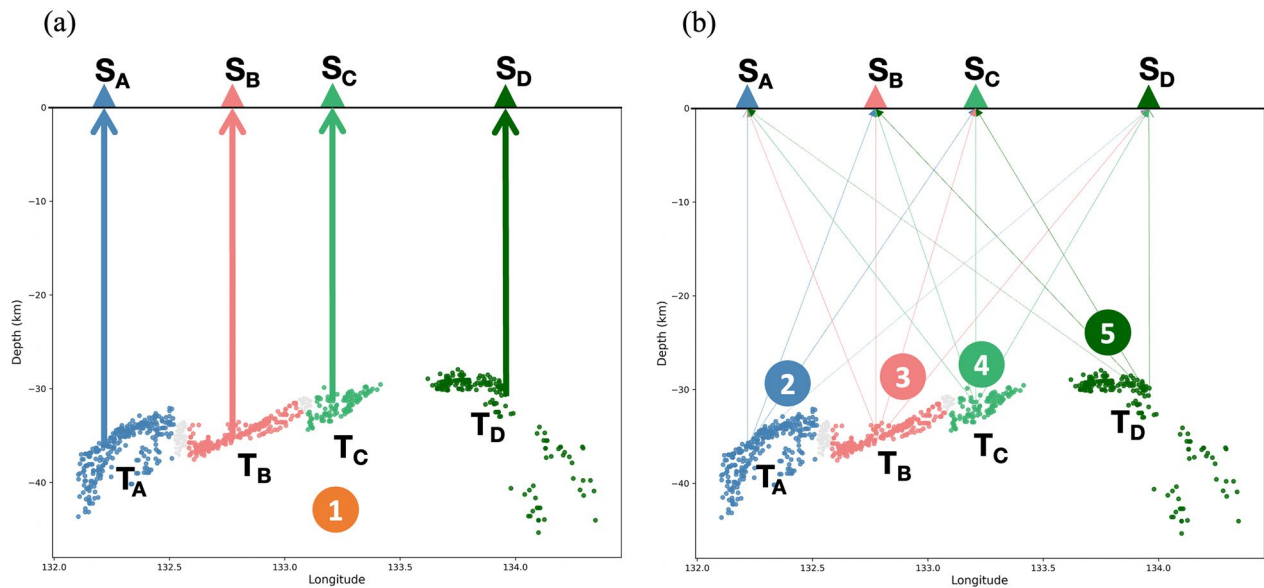


Fig. 3 Experiments of two-class classification in this study corresponding to Table 3. **a** Experiment-1: The experiment involved six tests that compared the signals from the source to the stations at the same area. **b** Experiments-2–5: These experiments considered sources from one area to stations in other areas. In total, there were twelve tests considered in Experiments-2–5. T_A , T_B , T_C , and T_D represents the four classes of tremors in Areas A to D, while S_A , S_B , S_C , and S_D represents the stations in each area. The names of stations are listed in Table 1

Table 3 Description of different two-class tests in the experiments

| Experiment ID | Test ID | Description | CR on SFS (%) | CR on all features (%) |
|---------------|---------|------------------------|---------------|------------------------|
| 1 | 1 | T_A-S_A vs T_B-S_B | 91.3 | 84.3 |
| | 2 | T_A-S_A vs T_C-S_C | 97.6 | 95.2 |
| | 3 | T_A-S_A vs T_D-S_D | 94.2 | 88.4 |
| | 4 | T_B-S_B vs T_C-S_C | 94.6 | 88.6 |
| | 5 | T_B-S_B vs T_D-S_D | 89.3 | 84.1 |
| | 6 | T_C-S_C vs T_D-S_D | 95.1 | 89.1 |
| 2 | 7 | T_A-S_A vs T_A-S_B | 92.7 | 87.5 |
| | 8 | T_A-S_A vs T_A-S_C | 97.8 | 95.5 |
| | 9 | T_A-S_A vs T_A-S_D | 92.3 | 85.6 |
| 3 | 10 | T_B-S_B vs T_B-S_A | 94.4 | 89.7 |
| | 11 | T_B-S_B vs T_B-S_C | 90.0 | 80.4 |
| | 12 | T_B-S_B vs T_B-S_D | 90.4 | 82.7 |
| 4 | 13 | T_C-S_C vs T_C-S_A | 99.5 | 98.3 |
| | 14 | T_C-S_C vs T_C-S_B | 96.6 | 93.4 |
| | 15 | T_C-S_C vs T_C-S_D | 95.7 | 89.8 |
| 5 | 16 | T_D-S_D vs T_D-S_A | 96.9 | 91.5 |
| | 17 | T_D-S_D vs T_D-S_B | 89.9 | 81.8 |
| | 18 | T_D-S_D vs T_D-S_C | 93.3 | 84.5 |

The corresponding number of features optimized for the performance of the model can be also seen in Fig. 4 (number above each circle). Out of the 29 features, 14

were found to be necessary to improve the CR. They are *EnvMax*, $ACF_{1/3}$, $ACF_{2/3}$, and $ACF_{1/3}/ACF_{2/3}$ in the 1st family, $Max\ BP_{2-8Hz}$, $NPksBP_{2-8Hz}$, $BP_{0.1-1Hz}$, BP_{2-8Hz} , BP_{5-20Hz} , $KurtBP_{2-8Hz}$, and $KurtBP_{5-20Hz}$ in the 2nd family, and $KurtDFT$, $NPks_DFTmean$, and $NPks_DFTmed$ in the 3rd family. Note that BP_{5-20Hz} was found to be useful in all 18 tests, whereas $BP_{0.1-1Hz}$ and $NPks_DFTmean$ were useful in 17 tests. At the first glance, the useful features separating tremors from different areas are associated with the autocorrelation function, band-pass filtered energy, and temporal evolution of DFT.

As shown in Table 3, the CR using all features are generally lower than those using SFS. Experiment-1 reveals that the CR was highest for T_A-S_A vs T_C-S_C (97.6% on SFS and 95.2% on all features) and lowest for T_B-S_B vs T_D-S_D (89.3% on SFS and 84.1% on all features). In Experiments-2–5, the classification rate was highest for T_C-S_C vs T_C-S_A (99.5% on SFS and 98.3% on all features) and lowest for T_D-S_D vs T_D-S_B (89.9% on SFS and 81.8% on all features). This suggests that the tremor signals between Areas A and C are more distinguishable compared with those from Areas B and D. To extract the features that are most representative of the tremor source in each area, evaluating the performance of individual features is necessary. The SFS-selected feature subsets were thus further used in second-step feature selection. By ranking individual features using the Fisher score, the highest F-score and the corresponding classification rate is shown as vertical bars in Fig. 5. Compared with the classification

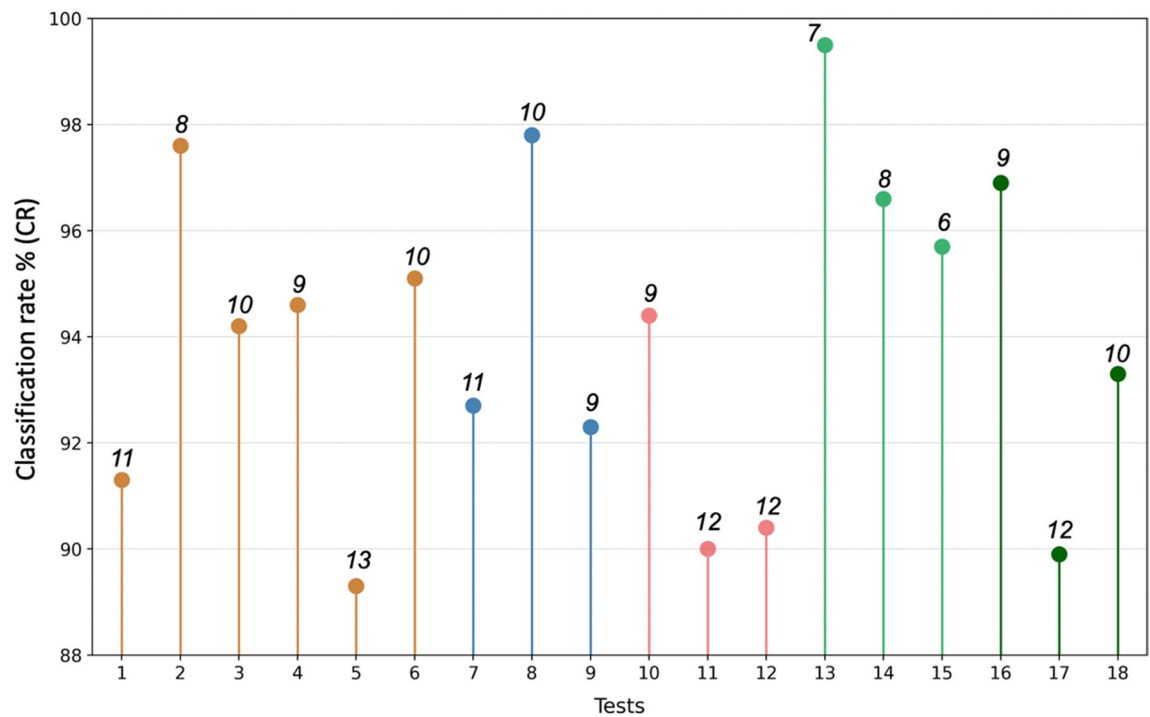


Fig. 4 The highest classification rate using SFS for various tests. The number above each circle denotes the total number of features in each subset for each test that met the optimum classification rate

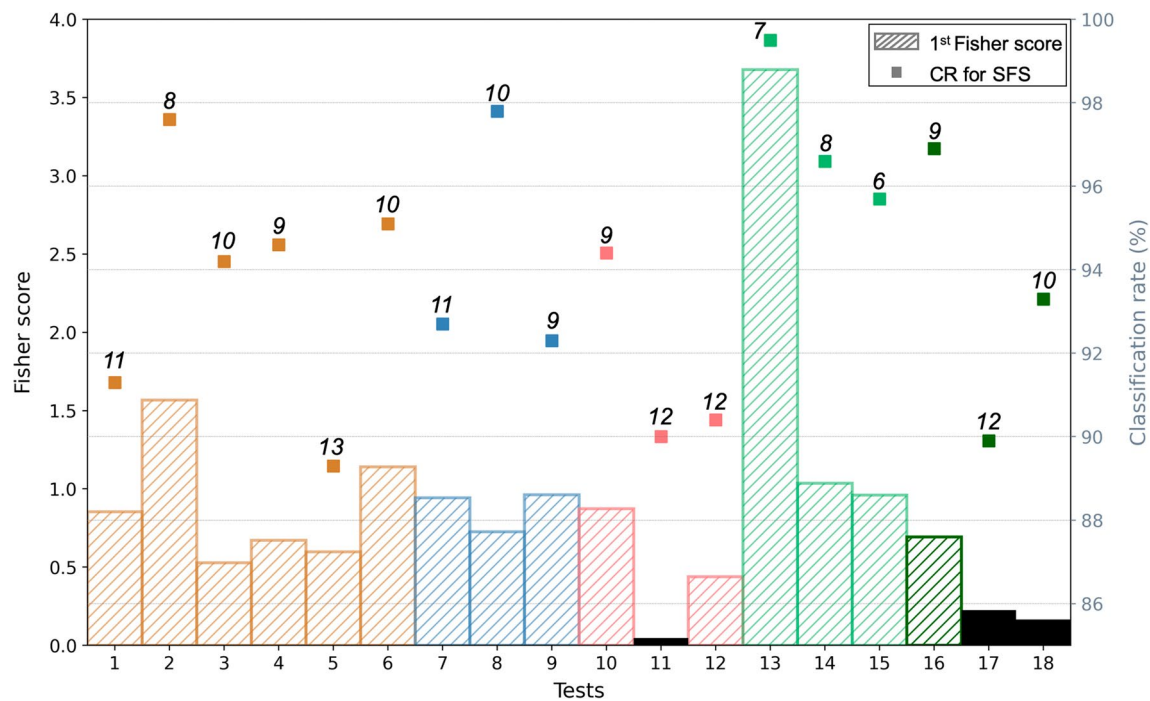


Fig. 5 Single feature selected by 1st Fisher score and the corresponding classification rate on different tests (vertical bar) and SFS-based classification rate (square). The bars in brown are the test under Fisher criterion on experiment 1; blue on experiment 2; light coral on experiment 3; light green on experiment 4; dark green on experiment 5. The sequences of tests on each experiment are in Table 1. The bars in black are the classifications with low p-values and Fisher scores. The number above each circle denotes the total number of features in each subset for each test that met the optimum classification rate

rate obtained using SFS (squares in Fig. 5), we found that the highest F -score in each test was highly correlated with the CR based on a top-ranking feature. As shown in Fig. 5, when the SFS-based CR is high, the top-ranking F -score is also high. A low F -score for a single feature, however, did not necessarily lead to a low CR. This can be illustrated by Test 18, where a top-ranking feature in isolation yielded a low F -score so as CR (black bar), but the CR obtained using multiple features reached 93.3% on SFS (square). Note that the purpose of this study is to design a classifier model for discriminating the tremors from different areas, and to explore the meaning behind the selected features. Therefore, the individual performance of the features carried out by F -score is adopted in the following analysis.

Triangle test to isolate the source effect

We designed a triangle test to determine the features that could better represent the differences in the source property of the tremors. Note that for each test we are able to select the efficient features using the criterion

of an F -score greater than 0.2. As the example shown in Fig. 6a, the efficient features in Test 1 represent the difference between S_A and S_B (source effect), between T_A and T_B (site effect), and between the path from T_A to S_A and that from T_B to S_B (path effect). To isolate the source effect, we next considered Test 7 (Fig. 6c), which only involved the difference in the site effect (S_A vs S_B) and path effect (T_A to S_A vs T_A to S_B). As denoted by the tabular form of Fig. 6b, the features that allow us to distinguish T_A - S_A from T_B - S_B are #24, #25, and #13 (Test 1 in Fig. 6a), whereas the feature distinguishing T_A - S_A and T_A - S_B is #24 (Test 7 in Fig. 6c). When Test 10 is also considered, the features distinguishing T_B - S_A and T_B - S_B are found to be #25, #24, and #16 (Test 10 in Fig. 6e and f). Note that the common feature for Test 1 and Test 7 is #24, which represents a differential site effect. Subtracting the features from Test 1 and Test 7 cancels out the site effect and minimizes the path effect of T_A - S_A vs T_B - S_B and T_A - S_A vs T_A - S_B . As listed in tabular form of between Fig. 6b, the candidate features that

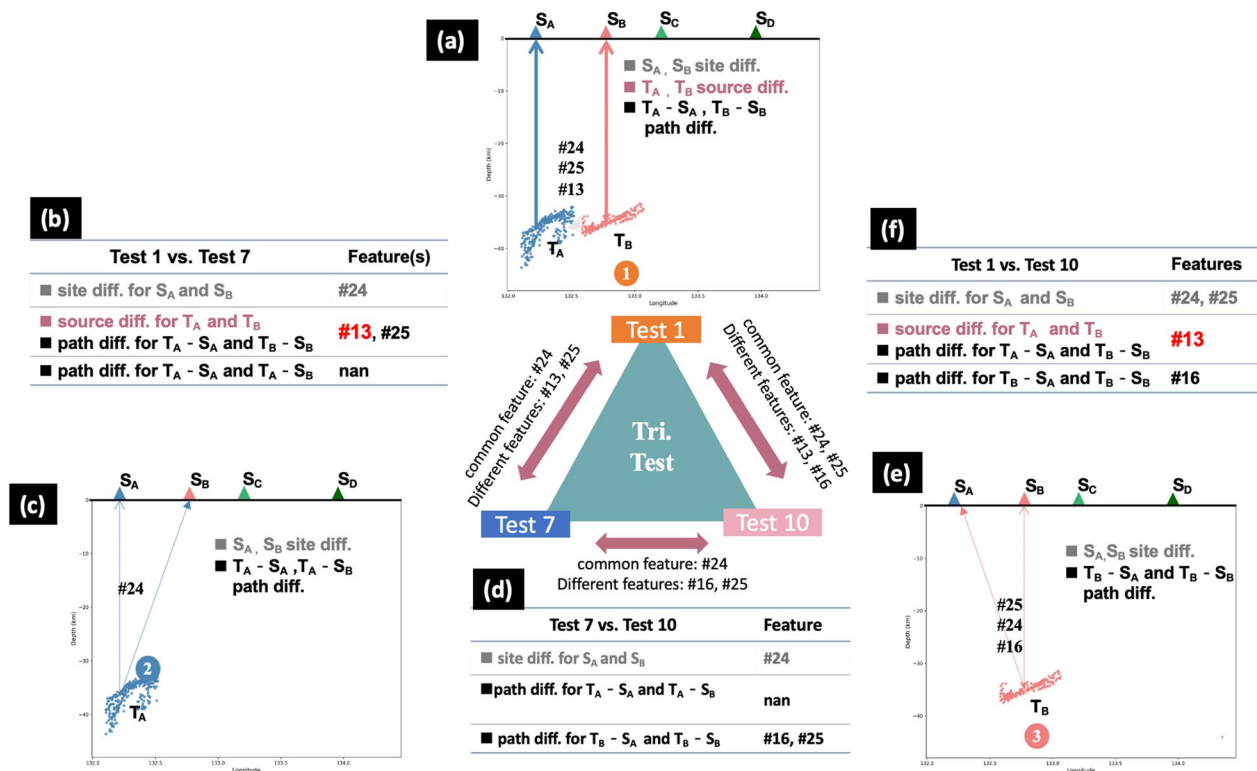


Fig. 6 Explanation of the triangle test using Tests 1, 7, and 10. **a** Test 1 in Experiment-1: the features selected using the F -score were #24, #25, and #13, which represent differences of source, path, and site effects for the two classes T_A - S_A vs T_B - S_B . **b** The list of common and different features for Test 1 and Test 7. **c** Test 7 in Experiment-2: the feature selected using the F -score was #24, which represents differences of path and site effects for the two classes T_A - S_A vs T_A - S_B . **d** The list of common and different features for Test 7 and Test 10. **e** Test 10 in Experiment-2: the features selected using the F -score were #25, #24, and #26, which represent differences of source, path, and site effects for the two classes T_B - S_A vs T_B - S_B . **f** The list of common and different features for Test 1 and Test 10

represent the different source effects of S_A and S_B are #13 and #25.

Similarly, subtracting the features from Test 1 and Test 10 (tabular form in Fig. 6f) leads to the same source effect discrimination. As Test 10 (Fig. 6e) contains the difference in site effects (S_A vs S_B) and the difference in path effects (T_B to S_A vs T_B to S_B), subtracting the features from Test 1 and Test 10 isolates the difference in source effect, #13. The common feature for Test 1 vs Test 7 and Test 1 vs Test 10 is feature #13 (highlighted in red in Figures b, d, and f), which is therefore considered responsible for the differences between T_A to T_B .

Such a triangle test that contains the three two-class tests in Table 3 generates a list of features that may contribute to the source differences. The binary classifications of source comparison of T_A vs T_B , T_A vs T_C , T_A vs T_D , T_B vs T_C , T_B vs T_D , and T_C vs T_D are illustrated in Fig. 7a–f, respectively. The resulting seven features that represent the source differences are listed in Table 4. How much these features show the regional difference in tremor properties will be discussed in the following section.

Selected features for tremor discrimination between areas

Using the triangle tests, we were able to determine the most representative features that allow us to discriminate between tremors from different areas. For example, as listed in Table 4, to distinguish the tremors in Area A (T_A) from those in Area B (T_B), the key feature is #13 ($NPks_{BP_{2-8Hz}}$, the number of peaks in the 2–8 Hz filtered envelope). To distinguish T_A from T_C , however, the key features are #11 ($ACF_{1/3}/ACF_{2/3}$, ratio of energy in the first third of the ACF to that in the remaining part) and #24 ($NPks_{DFTmean}$, number of peaks in the temporal evolution of the DFT mean). In total, there exist seven features that are useful for separating tremors of one area from another: $ACF_{1/3}$, $ACF_{2/3}$, $ACF_{1/3}/ACF_{2/3}$, $NPks_{BP_{2-8Hz}}$, BP_{5-20Hz} , $NPks_{DFTmean}$, and $NPks_{DFTmed}$.

The feature histograms in Fig. 8 can be used to visualize the frequency distribution of the normalized feature values for different areas. Note that the features $ACF_{1/3}$, $ACF_{2/3}$, and $ACF_{1/3}/ACF_{2/3}$ were computed in the time domain. As shown in Table 2, $ACF_{1/3}$ represents the energy in the first third of the ACF, $ACF_{2/3}$ represents the energy in the remaining part of the ACF, whereas $ACF_{1/3}/ACF_{2/3}$ is the ratio of the above two. As shown in

Fig. 8a–c, Tremors in Areas C and D are characterized by smaller $ACF_{1/3}$ and $ACF_{2/3}$ but greater $ACF_{1/3}/ACF_{2/3}$. The autocorrelation computes the similarity between a time series and a delayed version of itself. Here, the ACF was processed using a 60-s-long signal with a 0.6-s shift, which leads to 100 measures of time lapse. An example of the ACF can be seen in Additional file 1: Fig. S5a where we randomly selected 1000 events and then applied the ACF. Their averaged value is shown in Additional file 1: Fig. S5b, where tremors in Areas A and B have greater values not only in the first third of the ACF but also in the remaining two thirds, which is consistent with the comparison in Fig. 8a, b.

Features $NPks_{BP_{2-8Hz}}$ and BP_{5-20Hz} were computed in the frequency domain, as $NPks_{BP_{2-8Hz}}$ counted the number of peaks in the 2–8 Hz filtered envelope, and BP_{5-20Hz} computed the energy of the 5–20 Hz filtered waveforms. As shown in Fig. 8d, e, their difference between areas was less significant when comparing with three features associated with autocorrelation function as discussed earlier. An example of the $NPks_{BP_{2-8Hz}}$ feature description can be seen in Additional file 1: Figs. S6a, b, where the number of peaks in the envelope was counted when a given sample at $t=i$ had a higher amplitude than the neighboring samples at $t=i-1$ and $t=i+1$. Note that the filtered envelopes reveal high similarity between different areas, whereas the number of peaks indicate very minor differences. The spectra in Additional file 1: Fig. S6c, d, however, show that the total energy of the 5–20 Hz signals was largest in Area C. Together with the larger amplitudes for the 2–8 Hz envelope seen in Additional file 1: Fig. S5b, Area C appears to have been enriched by high frequency energy (> 2 Hz).

Features $NPks_{DFTmean}$ and $NPks_{DFTmed}$ were computed in the time–frequency domain. $NPks_{DFTmean}$ represents the number of peaks in the temporal evolution of the DFT mean values and $NPks_{DFTmed}$ for the DFT median values. As the DFT transforms a time series into frequencies and corresponding amplitudes, the DFT was sorted by amplitude into a finite time interval to obtain the mean and median amplitude. Additional file 1: Fig. S7 establishes the temporal pattern of the DFT in each area and the corresponding peak count. Note that Area C revealed the largest number of peaks in both $NPks_{DFTmean}$ and $NPks_{DFTmed}$, whereas for $NPks_{DFTmed}$, Area C also showed the largest amplitude. If the energy concentration is confined in a narrow frequency range, the median/mean amplitude retrieved from the time–frequency

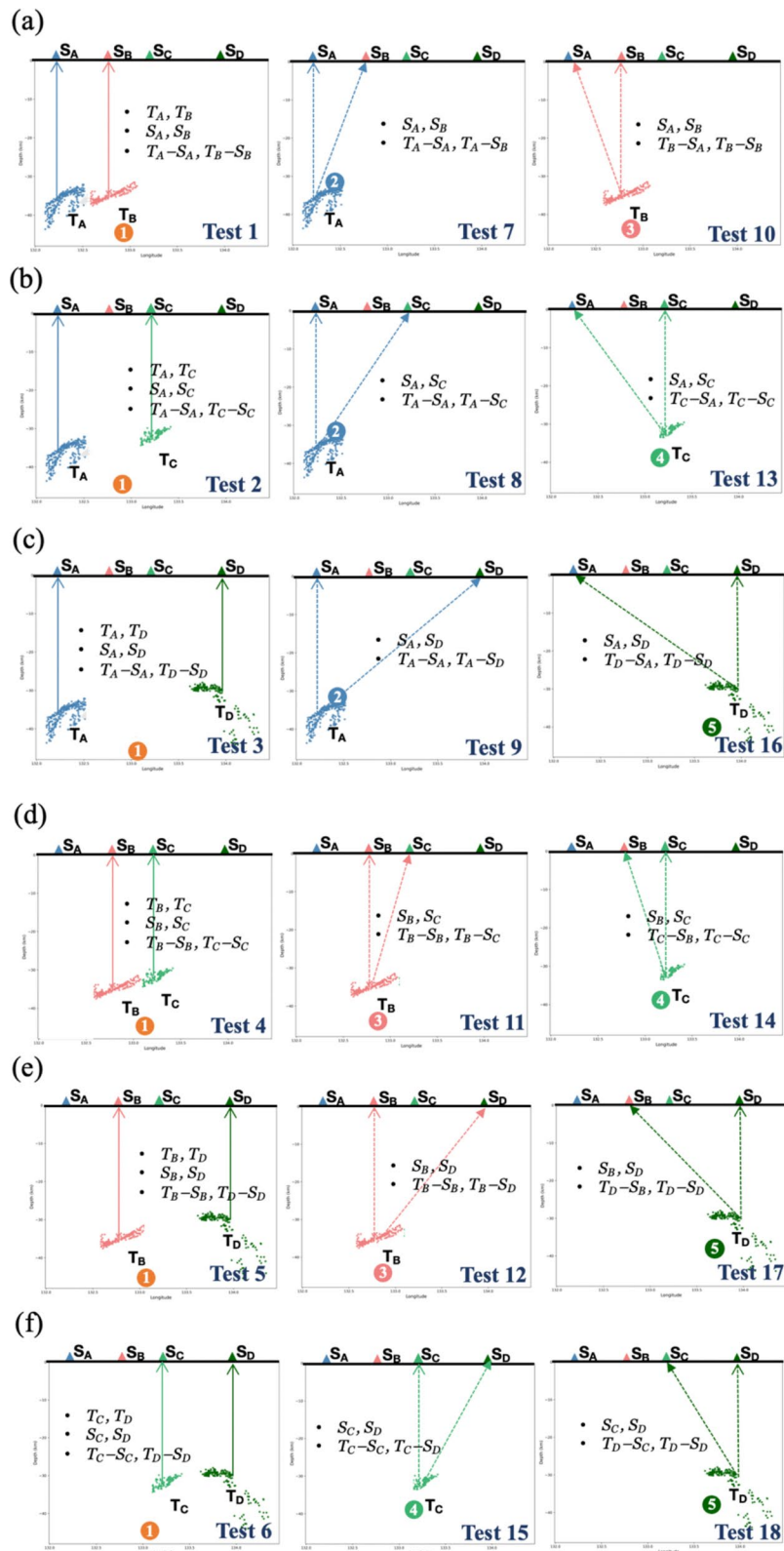
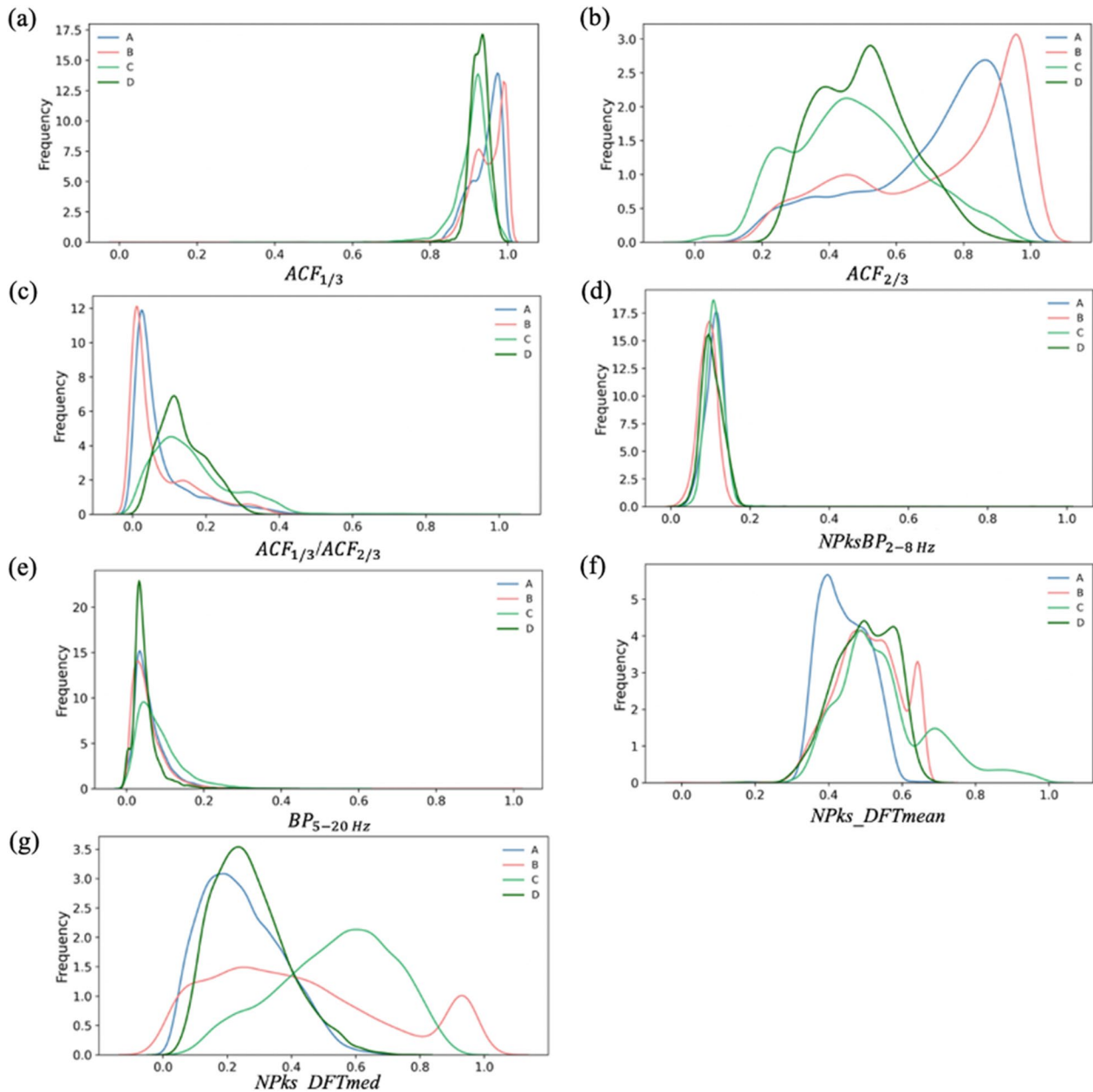


Fig. 7 Schematic diagram of various triangle tests that discriminate between the source effect of different areas: **a** T_A and T_B ; **b** T_A and T_C ; **c** T_A and T_D ; **d** T_B and T_C ; **e** T_B and T_D ; and **f** T_C and T_D . Note that the number in the filled circle corresponds to the Experiment ID

Table 4 Features responsible for the source differences using various triangle tests (feature number corresponding to ID number in Table 2)

| | T_A | T_B | T_C | T_D |
|-------|----------|--------------|--------------|----------|
| T_A | | #13 | #11, #24 | #10, #11 |
| T_B | #13 | | #9, #10, #11 | #10, #11 |
| T_C | #11, #24 | #9, #10, #11 | | #16, #25 |
| T_D | #10, #11 | #10, #11 | #16, #25 | |

plot would be similar over time. On the other hand, if the frequency concentration changes rapidly in time, the median/mean amplitude would be highly variable, leading to larger number of peaks in temporal evolution of DFTs. This implies that higher temporal fluctuation in main frequency content is needed, to explain the larger values of $NPks_DFTmean$ and $NPks_DFTmed$ in the tremor signals from T_C . T_A however, is

**Fig. 8** Distribution of feature space showing the discrimination between areas using the selected features listed in Table 4: **a** $ACF_{1/3}$, **b** $ACF_{2/3}$, **c** $ACF_{1/3}/ACF_{2/3}$, **d** $NPks_{BP2-8Hz}$, **e** BP_{5-20Hz} , **f** $NPks_DFTmean$, **g** $NPks_DFTmed$. Each normalized feature was obtained by dividing the feature value by the difference between the maximum and minimum value

characterized by smallest number of peaks in the both features, suggesting very minor variation in frequency content. To conclude, when the features from different areas are normalized and compared (Fig. 8), $ACF_{2/3}$, $ACF_{1/3}/ACF_{2/3}$, $NPks_DFTmean$, and $NPks_DFTmed$ are found to be more significant and will be used to obtain the new segmentation model in section "Source properties from seismic features".

Discussion

Slow earthquake phenomena

Shikoku island in the western Nankai Trough showcases a diversity of slow earthquake behavior. The long-term SSEs with durations of approximately 1 year are located on the updip side of Area A (the Bungo Channel), which is characterized by large magnitude earthquakes (M_w 7.0–7.1) and longer recurrence intervals of approximately 6 years (Hirose and Obara 2010; Ozawa et al. 2013). Long-term SSEs of similar duration are also found at the updip of Area B and Area C. They have smaller amounts of slip with the magnitude equivalence of M_w 6.0–6.3 (Takagi et al. 2016). A comparison of long-term SSE behavior between areas is listed in the "ID-1" row of Table 5. Short-term SSEs with a duration of approximately one week, on the other hand, are commonly observed in both Areas A and B with a similar recurrence time of 6 months. In Areas C and D, where short-term SSEs are less active, the recurrence interval appears to

be shorter (approximately three months) (Obara 2010; Hirose et al. 2010), as shown in row ID-2 in Table 5.

The spatial distribution of tremor events reveals clear segmentation bounded by abrupt changes in tremor activity. ETS is frequently observed across Areas A and B but not clearly in Areas C and D (Obara 2010). The tremors in Areas A and B are continuously distributed in the map in Fig. 1, and they were found to be largely modulated by the occurrence of SSEs (Takagi et al. 2016). The depth distribution of tremors approximated by the along-dip width in Annoura et al. (2016), however, reveals different characteristics, being more diffuse in Area A but narrower in Area B (Fig. 1c). Continuing to Area C, the tremors tend to be fewer and shorter in spatial extent along the strike, suggesting a clear separation from Area B. Further east, a discontinuity of the tremor zone occurs between Areas C and D. Note that the slip rate inferred from the tremor activity in Area C (~ 3.3 cm/yr) shows a strong contrast with Areas A and B (~ 4.2 cm/yr) and Area D (~ 4.9 cm/yr) (Hirose et al. 2010), as listed in row ID-3 of Table 5.

The tremor characteristics are also summarized in Table 5 (rows ID-4 to ID-7). The tremors in Areas A and B are greater in number and exhibit higher radiated energy, longer duration, and lower sensitivity to tidal stress. On the other hand, Areas C and D show fewer active tremors, lower radiated energy, and shorter duration but higher sensitivity to tidal stress (Obara 2002,

Table 5 The characteristics of slow earthquake phenomena and regional structure in different areas of Shikoku

| ID | Description of characters | A | B | C | D |
|----|---|-------------------------------------|---------------|---------------------------|---------------------|
| 1 | Equivalent magnitude of long-term SSEs (Hirose et al. 2010; Takagi et al. 2016) | M_w 7.0–7.1 | M_w 6.0–6.3 | M_w 6.0–6.3 | x |
| 2 | Recurrence interval of short-term SSEs (Hirose et al. 2010) | 5.6 ± 1.7 months | | 2.7 ± 1.4 month | 2.8 ± 1.6 month |
| 3 | Slip rate inferred from tremors (Hirose et al. 2010) | 4.2 ± 0.6 cm/yr | | 3.3 ± 0.4 cm/yr | 4.9 ± 0.6 cm/yr |
| 4 | Number of tremors (Obara 2010; Kano et al. 2018) | Large | Largest | Small | Smallest |
| 5 | Radiated energy per tremor (Obara 2010; Kano et al. 2018) | High | Moderate | Small | Smallest |
| 6 | Duration (Fig. 2 of this study) | More long-lasting events | | More short-lasting events | |
| 7 | Sensitivity to tidal stress (Ide 2010, 2012; Miyazawa et al. 2008) | Lowest | Low | High | High |
| 8 | Migration speed of tremor (Obara 2010; Kano et al. 2018) | 70% eastward (fast in A, slow in B) | | 64% westward | 67% eastward |
| 9 | Strength of patches (Kano et al. 2018) | Strong | Moderate | Weak | Weak |
| 10 | Density of brittle patches (Ohta and Ide 2017) | Dense | Moderate | Sparse | Sparse |
| 11 | Viscosity of background region (Ohta and Ide 2017) | High | Moderate | Low | Low |
| 12 | Q_p structure above the slab (Kita and Matsubara 2015) | High | Low | High | Low |
| 13 | Heterogeneity of asperities inferred from $NPks_DFTmean$ and $NPks_DFTmed$ in this study | Small | Large | Large | Small |
| 14 | Size of tremor zone inferred from duration, $ACF_{2/3}$, and $ACF_{1/3}/ACF_{2/3}$ in this study | Large | Largest | Smallest | Small |

2010; Kano et al. 2018; Ide 2010, 2012). The migration speed of the tremors further illustrates the distinction between areas. In Areas A, B, and D, the majority of tremors migrated from west to east, whereas in Area C the migration pattern was from east to west. As further established by Kano et al. (2018), Area A is characterized by relatively fast migration speed, greater energy release, and a smaller number of tremors compared with Area B (row ID-8 in Table 5). The structural heterogeneity (e.g., the density and strength of the brittle patches) that has led to differences in the condition of stress concentrations has been proposed to explain the above-mentioned characteristics (Ando et al. 2012; Ohta and Ide 2017; Kano et al. 2018). They inferred that the low strength of the tremor areas in Areas C and D is the reason why small tidal stress perturbations can easily promote their rupture. In addition, the tremor patches in Areas C and D are sparsely distributed with low viscosity in ductile surroundings, leading to pulse-like, short-lasting moment releases (Ohta and Ide 2017). The inferred properties of the host environment are listed in rows ID-9 to ID-11 of Table 5.

Conditions within the overriding plate

Using an attenuation structure, Kita and Matsubara (2016) established the relationship between the characteristics of Q_p in the lower crust of the overlying plate and the long-term SSE segmentation in Shikoku. High- Q_p zones are located in the crust of the island arc directly above the long-term SSE area, mostly in Area A. As addressed in Table 5, the lower crust above the slab in Area C is characterized by high Q_p and high V_p , whereas the adjacent Areas B and D show a moderate-to-low Q_p , low V_p , and low V_s . In Area A to the west, however, moderate-to-high Q_p and high V_p are observed in the lower crust. Inhomogeneity of the crust in the overlying plate is likely a result of orogeny, where crust thickening coincides with low Q_p , low V_p , and low V_s . Such heterogeneity corresponds to a change in physical properties of the rocks in the vicinity of the subduction interface, which controls the rheology and hence tremor and SSE activity (Kita and Matsubara 2016; Nakajima and Hasegawa 2016; Poiata et al. 2021).

The slip deficit rate in the transition zone may also reflect the conditions of the ductile background region. Using the moment release rate from tremors, the western (Areas A and B), central (Area C), to eastern Shikoku (Area D) exhibit slip deficit rates of 2.6 cm/yr, 3.4 cm/yr, and 2.0 cm/yr, respectively (Hirose et al. 2010). The area with the smallest slip deficit rate, Area D, corresponds to high attenuation, whereas the areas with the fastest slip deficit rate (Area C) and moderate slip rates (Areas A

and B) correspond to low attenuation. Note that ID-12 in Table 5 describes the averaged Q_p in each area from Kita and Matsubara (2016). Most of the Areas A and B, in fact, are characterized by high Q_p with a low- Q_p anomaly at the boundary with Area C (Fig. 14 in Kita and Matsubara (2016)).

Conceptual model for tremor segmentation

For a given station, the tremor signals reveal a combined effect from the source characteristics and the attenuation structure above which the tremors occur. Together with the collective information from sections "Slow earthquake phenomena" and "Conditions within the overriding plate", the efficient features identified in this study provide a better understanding of the tremor segmentation in Shikoku. To illustrate the source properties in each area, we followed the physical model proposed by Ando et al. (2012) that considers tremors as a result of cascade ruptures of small velocity-weakening patches (asperities) interacting with the velocity strengthening regions in the surroundings. As proposed by Nakata et al. (2011) and Yabe et al. (2015), the greater size of the tremor zone may result in longer duration of tremor episodes, and the higher density of asperities inside the zone may lead to greater amplitudes and reduced sensitivity to tidal stress. In addition, the higher effective strength of individual tremor asperities is likely responsible for the greater radiated energy (Kano et al. 2018). We thus used the conceptual model in Fig. 9 to illustrate the variety of tremor behavior in the along-strike direction.

Here each tremor zone is composed of multiple asperities, represented by the grey polygons enclosed by the dashed circle. The strength of the asperities is represented by different colors in the polygons. As summarized in Table 5, Area A is characterized by a large number of tremors with the highest radiated energy and longer lasting events, leading to strong asperities in relatively wide tremor zones (black polygons enclosed by large, dashed circles in Fig. 9). Similarly, Area B has the largest number of tremors with moderate radiated energy and longer lasting tremor episodes, leading to asperities of moderate strength in relatively wide tremor zones (dark grey polygons enclosed by large, dashed circles in Fig. 9). Tremors in both Areas A and B were found to last longer and be less sensitive to tidal stress and thus can be represented by larger tremor zones (dashed circles) and more densely distributed asperities in the tremor zones (filled polygons), respectively. By contrast, Areas C and D are characterized by a smaller number of tremors and shorter lasting events with strong sensitivity to tidal stress, leading to weaker asperities in a smaller number of relatively small tremor zones (dark white polygons

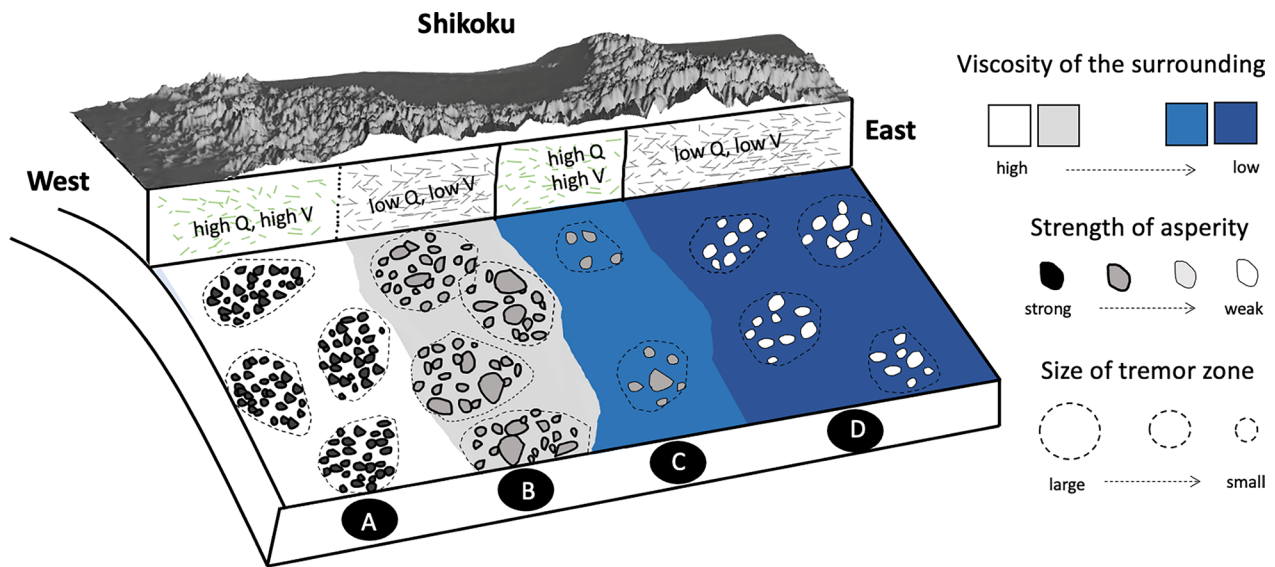


Fig. 9 Conceptual model of the tremor properties below Shikoku. The viscosity of the background region was proposed by Ohta and Ide (2017) based on the very low-frequency characteristics, where Areas A–B are characterized by high viscosity and densely distributed asperities. The strength of the asperities was proposed by Kano et al. (2018), and the size of the tremor zone was derived in this study

enclosed in smaller dashed circles in Fig. 9). As the SSEs may reflect the cascaded ruptures of a tremor zone that may be facilitated by shorter distance between tremor zones, Areas A and B (where the SSEs tend to be larger with longer recurrence intervals) are illustrated by the closely spaced dashed circles.

Source properties from seismic features

Using machine learning approaches and triangle tests designed in this study, the most representative features that best distinguish between tremors from different areas are found to be $ACF_{2/3}$, $ACF_{1/3}/ACF_{2/3}$, $NPks_DFTmean$, and $NPks_DFTmed$, as shown by the feature performance in Fig. 8.

As previously discussed in section "Selected features for tremor discrimination between areas", the large value of $NPks_DFTmean$ and $NPks_DFTmed$ may indicate the higher temporal fluctuation in main frequency content of the tremor signals. Therefore, with the largest number of peaks in both $NPks_DFTmean$ and $NPks_DFTmed$, Area C is likely characterized by higher temporal variation in the frequency content. In comparison, Area A is characterized by the smallest number of peaks in both $NPks_DFTmean$ and $NPks_DFTmed$, suggesting a small variation in frequency over time. We argue that these two features may indicate a difference in tremor source properties. As shown in Fig. 8g, Areas B and C are similar in terms of their $NPks_DFTmed$ performance, whereas Area A is similar to Area D. It is likely that the tremors

in Areas B and C have greater heterogeneity of asperities, which could be due to the diverse strength and/or size of their asperities. In contrast, tremors in Areas A and D are composed of more evenly distributed asperities of similar strength and/or size. We assumed the strength of asperities in each area remained similar, but altered asperity size to express the stronger heterogeneity of asperities in Areas B and C. This is represented in Fig. 9, where there is a large variation in the size of polygons in Areas B and C, but similar-sized grey polygons occur in Areas A and D. In ID-13 of Table 5 we summarize the difference in asperity heterogeneity based on the performance of features $NPks_DFTmean$ and $NPks_DFTmed$.

Additionally, $ACF_{1/3}$ represents the energy in the first third of the autocorrelation function (ACF), $ACF_{2/3}$ describes the energy in the remaining part of the ACF, and $ACF_{1/3}/ACF_{2/3}$ is the ratio of the above two. Note that the ACF can be used to map the decay rate of coda waves as an indicator of the heterogeneity of a local structure. Wegler and Sens-Schönfelder (2007) showed that the decay rate of the ACF is in good agreement with attenuation structure. Low attenuation corresponds to fast decay of the ACF and thus smaller values for $ACF_{2/3}$ and larger values for $ACF_{1/3}/ACF_{2/3}$, corresponding to the behavior of Areas C and D. Therefore, we suggest that in terms of attenuation structure, Areas A and B are similar to each other (low attenuation), whereas Areas C and D are more alike (high attenuation). This interpretation, however, is not consistent with the attenuation

structure inferred by Kita and Matsubara (2016) (their Fig. 10). They found that long-term SSEs (Area A) were located below the high- Q_p zone, and the rapid change of Q_p corresponds to the segment boundaries of tremor and short-term slow slip events. The cross-section of the Q_p -structure 10 km above the plate interface in Kita and Matsubara (2016) reveals that Areas A and C coincide with low attenuation, whereas Areas B and D coincide with high attenuation (ID-12 in Table 5). To reconcile the two models, we argue that the heterogeneity of the source itself may contribute to the performance of features $ACF_{2/3}$ and $ACF_{1/3}/ACF_{2/3}$. Assuming that stronger asperities radiate higher seismic energy and a larger tremor zone leads to a longer lasting tremor (Obara 2010; Yabe et al. 2015; Kano et al. 2018), the large-amplitude and/or long-lasting signals traveling through a strongly attenuated structure may have a similar ACF decay rate compared with small-amplitude and/or short-lasting signals traveling through a weakly attenuated body.

Compared with Area B, the tremors in Area A are characterized by similarly long duration but higher radiated energy, as listed in ID-5 and ID-6 of Table 5. To observe similar ACF performance for the tremors in Areas A (higher attenuation) and B (low attenuation), the tremors with higher radiated energy in Area A (i.e., stronger asperities) need to have shorter duration (i.e., smaller size of the tremor zone), so that the signals traveling through high- Q structure can produce a similar coda decay rate ($ACF_{2/3}$). Similarly, the size of the tremor zones should be greater in Area D and smaller in Area C to produce similar coda decay rates. The slightly different size of the tremor zones is illustrated by the symbols explanation in Fig. 9, where the tremor zones in Area B are larger than in Area A and the tremor zones in Area D are larger than in Area C. In addition, as the size of a tremor zone is proposed to control the duration (Yabe et al. 2015), longer lasting tremors in Areas A and B may be dominated by larger tremor zones compared with Areas C and D. Therefore, using $ACF_{2/3}$ and $ACF_{1/3}/ACF_{2/3}$ we propose that the tremor zone is largest in Area B and smallest in Area C. In ID-14 of Table 5, we summarize the difference in asperity heterogeneity based on the performance of $ACF_{2/3}$ and $ACF_{1/3}/ACF_{2/3}$.

Two-class vs. four-class classification models

A four-class classification configuration is also proposed here, to compare with the current binary models. As the four-class model classifies an input feature vector into one of the four classes, it usually requires low computational complexity. For example, a multi-layer neural network (with three output neurons in the output layer) can accomplish such task. However, the “feature set” may not be the optimum for the four classes, given that the best

feature subset for classifying two classes (e.g., A and B) may not be the optimum for another binary classification of classes A and C. Accordingly, we chose the pair-wise classification strategy commonly adopted in the machine learning community in our four-class classification model.

The pairwise classification classifies two classes with a set of features that are the best for the two classes’ classification. Such multi-class classification based on a set of binary classifications has been commonly used such as a support vector machine (SVM) with One-against-one (OAO) method. This is done by extending SVM to multi-class classification: to classify N classes, you need to train $N(N-1)/2$ SVM classifiers, and then use a majority voting strategy to make a final decision (Liu and Chen 2007; Liu et al. 2007). Since the OAO method trains a set of binary classifiers and for each binary classification the feature subset is different (the optimum for that binary classification), the overall classification result, i.e., the multi-class classification accuracy, would be better than that of a single multi-class model (e.g., a neural network). One comparison example can be found in Liu et al. (2007), where OAO-SVM outperforms a neural network trained by error back-propagation algorithm. Since the OAO strategy may achieve better accuracy, we decided to use such approach to accomplish the multi-class classification in the current study. The main drawback however, is that we need to train “many” binary classifiers in such pair-wise classification configuration for multi-class classification, which still results in a high computational complexity. Fortunately, the k -NN classifier used in the current study is an instance-based learning algorithm, which the training effort and computational complexity can be largely reduced.

Here, we extended the binary k -NN classifier to solve the four-class classification problem using OAO method and a majority voting strategy (Liu and Chen 2007). This is done by sending the test data into six k -NN classifiers (i.e., for N -class classification, $N(N-1)/2$ binary classifiers can be formed). The majority of the four class labels generated from the six k -NN classifiers will determine the final classification output. The illustration of four-class models are shown in Additional file 1: Fig. S8, corresponding to the combination of tests in Table 3. In experiment 1, the data from T_A-S_A , T_B-S_B , T_C-S_C , and T_D-S_D are regarded as four different classes and each class is composed of 3000 test data. Out of 3000 events recorded at each station group, 2450 of them are predicted correctly as T_A-S_A , while 1815, 2344, and 2221 are predicted correctly as T_B-S_B , T_C-S_C , T_D-S_D , respectively, leading to the CR of 73.6%. The T_A , T_B , T_C , and T_D tremors recorded at different station-groups can also generate different sets

of four-class model, as listed by experiments 2 to 5 in Additional file 1: Fig. S8. Their CR however, are generally lower than that in Experiment 1. This suggests that the differences of path effect are likely less obvious than that of source effect, as the experiments 2–5 consider different classes from different path effects.

The greater travel distance usually leads to rapid attenuation of higher frequency signals, which may produce a trend of precision reduction with distance. In the four-class model, however, such trend is not clear, the lowest precision does not always occur at the longest travel distance (lowest for T_A-S_D but not for T_D-S_A). Instead, the lowest precision occurs at T_A-S_D (Additional file 1: Fig. S8b), T_B-S_B (Additional file 1: Fig. S8c), T_C-S_D (Additional file 1: Fig. S8d), and T_D-S_D (Additional file 1: Fig. S8e), which is consistent with the two lowest precisions in Additional file 1: Fig. S8a. This suggests that the paths to S_B and S_D are less distinguishable comparing with other paths. On the other hand, the highest precisions are mostly observed at the paths to S_A (T_A-S_A in experiment 1, T_A-S_A in experiment 2, T_B-S_A in experiment 3, T_D-S_A in experiment 5), suggesting that they are strongly distinguishable from others in both source and path effects. T_C-S_C in experiment 4 also produces highest precision but not in experiment 1, indicating the source effect of T_C may play more important role in the high precision in experiment 4. The four-class model reveals consistent conclusion with that from binary models—tremors signals from Areas A and C are more distinguishable compared with those from Areas B and D.

There exists various machine learning approaches that have been used to solve classification problems, including linear classification (e.g., logistic regression), distance estimation (k -nearest neighbors), support vector machines (SVM), decision tree (e.g., Random Forest), and neural networks. Among all, the k -NN is relatively easy and straightforward to interpret. The triangle tests designed here allows us to obtain the region-specific features importance, leading to the interpretation of tremor source properties in four areas of Shikoku. However, the selected 7 features (Table 4) that are responsible for source difference in Areas A to D may be also responsible for path difference. To fully understand the impact of feature importance methods on the interpretation, different feature importance methods should be examined on the same dataset and classifier. For example, Random Forest (RF) could provide feature importance ranks during the classification process, which has demonstrated to be highly efficient for multi-classes problem. It is worthy to explore if the various feature importance methods are interchangeable and can lead to the consistent conclusion, in the future.

Conclusions

In this study, we first demonstrated the relevant regional characteristics using the tremor catalog published by Annoura et al. (2016). During the study period from 1 June 2014 to 31 March 2015, the number of tremors in four different areas in Shikoko (A to D, from east to west) was 15,000, 31,000, 10,000, and 16,000, respectively. The tremors of these four areas were treated as different classes and the tremors in the catalog were segmented into 60-s-long signals as the labeled data. Each class of tremor event was split into training (70%) and testing (30%) datasets. The supervised k -nearest neighbor (k -NN) classifier was trained to distinguish tremors from the four areas. To identify efficient features that better represent the properties of the tremors in each area, we adopted binary classification with a set of experiments. When the k -NN classifier was applied to the original 29 seismic features, the classification performance reached a classification rate greater than 90% at all 12 stations.

We further designed a triangle test to determine the features that could better represent the difference in the source properties of the tremors. As a result, two groups of features that were most useful at distinguishing tremors of one area from another were identified. The first group considered the number of peaks in the temporal evolution of discrete Fourier transform (DFT) mean values ($NPks_DFTmean$) and DFT median values ($NPks_DFTmed$), which imply a certain degree of temporal fluctuation in main frequency content of the tremor signals. The second group considered the energy in the first third of the autocorrelation function (ACF) ($ACF_{1/3}$), the energy in the remaining part of the ACF ($ACF_{2/3}$), and the ratio of above two ($ACF_{1/3}/ACF_{2/3}$). Given that the decay rate of the ACF shows good agreement with attenuation structure, the fast decay of the ACF is likely a result of a low attenuation structure along the path of the tremors.

The efficient features obtained in this study provide insight into the tremor source properties. To explain $NPks_DFTmed$ performance, the tremors in Areas B and C may need to have greater heterogeneity of asperities resulting from the diverse strength and size of the asperities in tremor zones. In contrast, tremors in Areas A and D are composed of more evenly distributed asperities of similar strength/size, leading to weak heterogeneity of asperities. Using features $ACF_{1/3}$ and $ACF_{2/3}$, we also inferred that the size of tremor zone may decide the behavior of coda decay, to reconcile the attenuation structure 10 km above the plate interface by Kita and Matsubara (2016). We argue that the tremor zone needs to be large in Areas A and B and small in Area C and D, to explain the significant difference in $ACF_{1/3}$ performance.

Together with previously documented variations in slow earthquake behavior between different areas, we finally proposed a conceptual model that provides a better understanding of the regional differences in the tremor sources of Shikoku, southwestern Japan.

Abbreviations

| | |
|--------------|--------------------------------|
| ETS | Episodic tremor and slip |
| <i>k</i> -NN | <i>k</i> -Nearest neighbor |
| RMS | Root mean square |
| LOO | Leave-one-out cross validation |
| CR | Classification rate |
| DFT | Discrete Fourier Transform |
| ACF | Autocorrelation function |
| SFS | Sequential forward selection |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40623-023-01776-w>.

Additional file 1: Text S1. Feature extraction dataset. **Figure S1.** The example of original tremor waveforms. **Figure S2.** The example of the autocorrelation function. **Figure S3.** The example of the band-pass filtered waveforms. **Figure S4.** The example of the evolution of the Discrete Fourier Transform. **Figure S5.** The examples of features $ACF_{1/3}$ and $ACF_{2/3}$. **Figure S6.** The example of feature $NPKs_{BP_{2-8\text{ Hz}}}$. **Figure S7.** The example of $NPKs_DFTmean$ and $NPKs_DFTme$. **Figure S8.** The result of four-class classification model.

Acknowledgements

We wish to thank ERI summer intern program that provided the opportunity for the collaboration. We also appreciate Jyr-Ching Hu's helpful suggestions on this work. Finally, we would like to thank Uni-edit (www.uni-edit.net) for editing and proofreading this manuscript. TEC Contribution Number is 00181.

Author contributions

KHC wrote this paper and supervised the work. HYC collected and investigated the data. KO helped on the conceptualization and design of this work. YHL provided the guidance of machine learning algorithms. All authors read and approved the final manuscript.

Funding

This study is supported by Taiwan National Science and Technology Council (NSTC) Grant 111-2116-M-003-008-MY3 and JSPS KAKENHI Grant# JP16H06473 in Grant-in-Aid for Scientific Research on Innovative Areas "Science of Slow Earthquakes".

Availability of data and materials

We used Hi-net data provided by the National Research Institute for Earth Science and Disaster Resilience (<http://www.hinet.bosai.go.jp/?LANG=en>). Tremor catalog used here is from Annoura et al. (2016).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

All the authors declare no competing interests.

Author details

¹Department of Earth Sciences, National Taiwan Normal University, New Taipei, Taiwan. ²Department of Geosciences, National Taiwan University, New Taipei, Taiwan. ³Earthquake Research Institute, the University of Tokyo, Tokyo, Japan. ⁴Department of Mechanical Engineering, National Taiwan University of Science and Technology, New Taipei, Taiwan.

Received: 2 September 2022 Accepted: 22 January 2023

Published online: 08 March 2023

References

- Ando R, Takeda N, Yamashita T (2012) Propagation dynamics of seismic and aseismic slip governed by fault heterogeneity and Newtonian rheology. *J Geophys Res* 117:B11308. <https://doi.org/10.1029/2012JB009532>
- Annoura S, Obara K, Maeda T (2016) Total energy of deep low-frequency tremor in the Nankai subduction zone, southwest Japan. *Geophys Res Lett* 43:2562–2567. <https://doi.org/10.1002/2016GL067780>
- Brudzinski MR, Allen RM (2007) Segmentation in episodic tremor and slip all along Cascadia. *Geology* 35:907–910. <https://doi.org/10.1130/G23740A.1>
- Chen KH, Tai H-J, Ide S, Byrne TB, Johnson CW (2018) Tidal modulation and tectonic implications of tremors in Taiwan. *J Geophys Res* 123:5945–5964. <https://doi.org/10.1029/2018JB015663>
- Chuang LY, Chen KH, Wech A, Peng W (2014) Ambient tremors in a collisional orogenic belt. *Geophys Res Lett* 41:1485–1491. <https://doi.org/10.1002/2014GL059476>
- Efron B, Tibshirani R (1993) An introduction to the bootstrap. Chapman & Hall, London. <https://doi.org/10.1007/978-1-4899-4541-9>
- Fang L, Zhao H, Wang P, Yu M, Yan J, Cheng W, Chen P (2015) Feature selection method based on mutual information and class separability for dimension reduction in multidimensional time series for clinical data. *Biomed Signal Process Control* 21:82–89. <https://doi.org/10.1016/j.bspc.2015.05.011>
- Hibert C, Malet J-P, Bourrier F, Provost F, Berger F, Bornemann P et al (2017) Single-block rockfall dynamics inferred from seismic signal analysis. *Earth Surf Dyn Discuss* 5:1–15. <https://doi.org/10.5194/esurf-2016-64>
- Hirose H, Obara K (2005) Repeating short- and long-term slow slip events with deep tremor activity around the Bungo channel region, southwest Japan. *Earth Planets Space* 57:961–972
- Hirose H, Asano Y, Obara K, Kimura T, Matsuzawa T, Tanaka S, Maeda T (2010) Slow earthquakes linked along dip in the Nankai subduction zone. *Science* 330:1502. <https://doi.org/10.1126/science.119710>
- Hirose H, Kimura T (2020) Slip distributions of short-term slow slip events in Shikoku, southwest Japan, from 2001 to 2019 based on tilt change measurements. *J Geophys Res* 125:e2020JB019601. <https://doi.org/10.1029/2020JB019601>
- Hirose H, Obara K (2010) Recurrence behavior of short-term slow slip and correlated nonvolcanic tremor episodes in western Shikoku, southwest Japan. *J Geophys Res* 115:B00A21. <https://doi.org/10.1029/2008JB006050>
- Ide S (2008) A Brownian walk model for slow earthquakes. *Geophys Res Lett* 35:L17301. <https://doi.org/10.1029/2008GL034821>
- Ide S (2010) Striations, duration, migration and tidal response in deep tremor. *Nature* 466(7304):356–359. <https://doi.org/10.1038/nature09251>
- Ide S (2012) Variety and spatial heterogeneity of tectonic tremor worldwide. *J Geophys Res* 117:B03302. <https://doi.org/10.1029/2011JB008840>
- Ide S, Shelly DR, Beroza GC (2007) Mechanism of deep low frequency earthquakes: further evidence that deep non-volcanic tremor is generated by shear slip on the plate interface. *Geophys Res Lett* 34:L03308. <https://doi.org/10.1029/2006GL028890>
- Idehara K, Yabe S, Ide S (2014) Regional and global variations in the temporal clustering of tectonic tremor activity. *Earth Planet Space* 66:66. <https://doi.org/10.1186/1880-5981-66-66>
- Kano M, Kato A, Ando R, Obara K (2018) Strength of tremor patches along deep transition zone of a megathrust. *Sci Rep* 8:3655. <https://doi.org/10.1038/s41598-018-22048-8>

- Kao H, Shan S-J, Rogers G, Dragert H (2007) Migration characteristics of seismic tremors in the northern Cascadia margin. *Geophys Res Lett* 34:L03304. <https://doi.org/10.1029/2006GL028430>
- Kita S, Matsubara M (2016) Seismic attenuation structure associated with episodic tremor and slip zone beneath Shikoku and the Kii peninsula, southwestern Japan, in the Nankai subduction zone. *J Geophys Res* 121:1962–1982. <https://doi.org/10.1002/2015JB012493>
- Liu YH, Chen YT (2007) Face recognition using total margin-based adaptive fuzzy support vector machines. *IEEE Trans Neural Netw* 18(1):178–192. <https://doi.org/10.1109/TNN.2006.883013>
- Liu YH, Huang HP, Weng CH (2007) Recognition of electromyographic signals using cascaded kernel learning machine. *IEEE/ASME Trans Mechatron* 12:253–264
- Liu YH, Huang S, Huang Y-D (2017) Motor imagery EEG classification for patients with amyotrophic lateral sclerosis using fractal dimension and Fisher's criterion-based channel selection. *Sensors* 17(7):1557. <https://doi.org/10.3390/s17071557>
- Liu Y-H, Yeh T-C, Chen KH, Chen Y, Yen Y-Y, Yen H-Y (2019) Investigation of single-station classification for short tectonic tremor in Taiwan. *J Geophys Res*. <https://doi.org/10.1029/2019JB017866>
- Mitchell TM (1997) *Machine learning*. McGraw-Hill, New York
- Miyazawa M, Brodsky EE, Mori J (2008) Learning from dynamic triggering of low-frequency tremor in subduction zones. *Earth Planets Space* 60:e17–e20
- Nadeau RM, Dolenc D (2005) Nonvolcanic tremors deep beneath the San Andreas Fault. *Science* 307(5708):389. <https://doi.org/10.1126/science.1107142>
- Nadeau RM, Guilhem A (2009) Nonvolcanic tremor and the 2003 San Simeon and 2004 Parkfield, California earthquakes. *Science* 325:191–193. <https://doi.org/10.1126/science.1174155>
- Nakajima J, Hasegawa A (2016) Tremor activity inhibited by well-drained conditions above a megathrust. *Nat Commun* 7:13863. <https://doi.org/10.1038/ncomms13863>
- Nakata R, Suda N, Tsuruoka H (2008) Non-volcanic tremor resulting from the combined effect of Earth tides and slow slip events. *Nat Geosci* 1:676–678
- Nakata R, Ando R, Hori T, Ide S (2011) Generation mechanism of slow earthquakes: numerical analysis based on a dynamic model with brittle-ductile mixed fault heterogeneity. *J Geophys Res*. <https://doi.org/10.1029/2010JB008188>
- Obara K (2002) Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* 296:1679–1681
- Obara K (2010) Phenomenology of deep slow earthquake family in southwest Japan: spatiotemporal characteristics and segmentation. *J Geophys Res* 115:B00A25. <https://doi.org/10.1029/2008JB006048>
- Obara K, Hirose H (2006) Non-volcanic deep low-frequency tremors accompanying slow slips in the southwest Japan subduction zone. *Tectonophysics* 417:33–51. <https://doi.org/10.1016/j.tecto.2005.04.013>
- Obara K, Sekine S (2009) Characteristic activity and migration of episodic tremor and slow-slip events in central Japan. *Earth Planets Space* 61:853–862
- Obara K, Kasahara K, Hori S, Okada Y (2005) A densely distributed high sensitivity seismograph network in Japan: Hi-net by National Research Institute for Earth Science and Disaster Prevention. *Rev Sci Instrum* 76:021301–021301-12. <https://doi.org/10.1063/1.1854197>
- Obara K, Tanaka S, Maeda T, Matsuzawa T (2010) Depth-dependent activity of non-volcanic tremor in southwest Japan. *Geophys Res Lett*. <https://doi.org/10.1029/2010GL043679>
- Obara K, Matsuzawa T, Tanaka S, Kimura T, Maeda T (2011) Migration properties of non-volcanic tremor in Shikoku, southwest Japan. *Geophys Res Lett*. <https://doi.org/10.1029/2011GL047110>
- Ohta K, Ide S (2017) Resolving the detailed spatiotemporal slip evolution of deep tremor in Western Japan. *J Geophys Res* 122(12):10009–10036. <https://doi.org/10.1002/2017jb014494>
- Okada Y, Kasahara K, Hori S, Obara K, Sekiguchi S, Fujiwara H, Yamamoto A (2004) Recent progress of seismic observation networks in Japan—Hinet, F-net, K-NET and KIK-net. *Earth Planets Space* 56:15–28. <https://doi.org/10.1186/BF03353076>
- Ozawa S, Yairi H, Imakiire T, Tobita M (2013) Spatial and temporal evolution of the long-term slow slip in the Bungo Channel, Japan. *Earth Planets Space* 65:67–73. <https://doi.org/10.5047/eps.2012.06.009>
- Poiata N, Vilotte J-P, Shapiro NM, Supino M, Obara K (2021) Complexity of deep low-frequency earthquake activity in Shikoku (Japan) imaged from the analysis of continuous seismic data. *J Geophys Res* 126(11):e2021B022138
- Provost F, Hilbert C, Malet J-P (2016) Automatic classification of endogenous landslide seismicity using the Random Forest super-vised classifier. *Geophys Res Lett* 44:113–120. <https://doi.org/10.1002/2016GL070709>
- Rogers G, Dragert H (2003) Episodic tremor and slip on the Cascadia subduction zone: the chatter of silent slip. *Science* 300:1942–1943. <https://doi.org/10.1126/science.1084783>
- Sekine S, Hirose H, Obara K (2010) Along-strike variations in short-term slow slip events in the southwest Japan subduction zone. *J Geophys Res* 115:B00A27. <https://doi.org/10.1029/2008jb006059>
- Shelly DR, Beroza GC, Ide S, Nakamura S (2006) Low frequency earthquakes in Shikoku, Japan, and their relationship to episodic tremor and slip. *Nature* 442:188–191. <https://doi.org/10.1038/nature04931>
- Supino M, Shapiro N, Vilotte JP, Poiata N, Obara K. (2021) Tectonic low-frequency earthquakes in Shikoku, Japan: source scaling, size distribution and observational limits. <https://doi.org/10.1002/essoar.10506594.1>
- Takagi R, Obara K, Maeda T (2016) Slow slip event within a gap between tremor and locked zones in the Nankai subduction zone. *Geophys Res Lett* 43:1066–1074. <https://doi.org/10.1002/2015GL066987>
- Wech AG, Creager KC (2008) Automated detection and location of Cascadia tremor. *Geophys Res Lett* 35:L20302. <https://doi.org/10.1029/2008GL035458>
- Wech AG, Boese CM, Stern TA, Townend J (2012) Tectonic tremor and deep slow slip on the Alpine Fault. *Geophys Res Lett* 39:L10303. <https://doi.org/10.1029/2012GL051751>
- Wegler U, Sens-Schönfelder C (2007) Fault zone monitoring with passive image interferometry. *Geophys J Int* 168:1029–1033. <https://doi.org/10.1111/j.1365-246X.2006.03284.x>
- Whitney AW (1971) A direct method of nonparametric measurement selection. *IEEE Trans Actions Comput* 100(9):11001103
- Yabe S, Tanaka Y, Houston H, Ide S (2015) Tidal sensitivity of tectonic tremors in Nankai and Cascadia subduction zones. *J Geophys Res* 120:7587–7605. <https://doi.org/10.1002/2015JB012250>

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)