



RESEARCH ARTICLE

10.1029/2019JB017866

Investigation of Single-Station Classification for Short Tectonic Tremor in Taiwan

Key Points:

- Recent observation from a mountain array allows us to document the tremor events recorded at the closest distance
- We evaluate the possibility of detecting minutes-long tremor using a single station
- We obtain a collection of efficient seismic features for successful classification

Correspondence to:

K. H. Chen,
katepili@gmail.com

Citation:

Liu, Y.-H., Yeh, T.-C., Chen, K. H., Chen, Y., Yen, Y.-Y., & Yen, H.-Y. (2019). Investigation of single-station classification for short tectonic tremor in Taiwan. *Journal of Geophysical Research: Solid Earth*, 124, 8803–8822. <https://doi.org/10.1029/2019JB017866>

Received 16 APR 2019

Accepted 25 JUL 2019

Accepted article online 30 JUL 2019

Published online 22 AUG 2019

Yi-Hung Liu¹ , Ting-Chen Yeh², Kate Huihsuan Chen² , Yaochieh Chen², Yuan-Yi Yen², and Horng-Yuan Yen³

¹Graduate Institute of Mechatronics Engineering, National Taipei University of Technology, Taipei, Taiwan, ²Department of Earth Sciences, National Taiwan Normal University, Taipei, Taiwan, ³Department of Earth Sciences, National Central University, Jhongli, Taiwan

Abstract Tectonic tremor in Taiwan is known for its short duration and weak amplitude. Recent observations from the mountain area (Central Range) allows us to document such type of tectonic tremor recorded at a close distance and to investigate a single station technique for a minute-long tremor. To effectively distinguish tremor from earthquakes and noise, we applied the Fisher's class separability criterion to automatically select the optimal feature subset from a set of commonly adopted feature candidates. During the study period of 1 January to 16 September 2016 when a local seismic array was deployed, we successfully differentiated tremor from local earthquakes and noise with high accuracy of 86.6% to 98.9% from three stations using a k -nearest neighbors classifier. Other than the maximum amplitude, number of peaks, and energy of the 2- to 8-Hz passband, the spikiness of discrete Fourier transforms median in time is also found to be important to separate tremor from noise.

1. Introduction

Tectonic tremor is long-lasting (several minutes to days), noise-like seismic signal with 1- to 10-Hz frequency band. The lack of impulsive P and S arrivals in the signal represent a slow slip process at the fault interface (e.g., Dragert et al., 2004). They have been observed in various tectonic regions, typically adjacent to the locked portion of the fault (e.g., Miyazawa & Mori, 2005; Nadeau & Dolenc, 2005; Obara, 2002; Peng & Chao, 2008; Rubinstein et al., 2009; Wech & Creager, 2008), where the triggering potential for large earthquakes is found to be likely if the adjacent locked zone is critically stressed (Ito et al., 2013; Kato et al., 2012; Kato & Nakagawa, 2014; Obara & Kato, 2016; Socquet et al., 2017; Uchida et al., 2016). In places where geodetic measurements do not have enough resolution or slow slip events are small or short in duration, tectonic tremor is often regarded as a powerful tool for monitoring slow slip events. The detection of tectonic tremor relies heavily on the similarity of and the time lapse between the arrival of tremor bursts from multiple stations (Obara, 2002). The tectonic tremor is generally identified by a high envelope waveform similarity with nearly the same arrival at different stations, a higher energy than the incoherent background noise, and a much longer duration than ordinary earthquakes. Station coverage and density, therefore, are the key factors for the detection capability. In places where tremor signals are weak in amplitude and short in duration, the signals were manually checked to exclude swarms of small earthquakes or loud noise (e.g., Chen et al., 2018), resulting in a time-consuming and subjective determination of tremor catalogs.

Single station detection of tectonic tremor has been proposed using a variety of methods on Cascadia data. Brudzinski and Allen (2016) used the amplitude ratio of 2- to 5-Hz passband to a long-term average to scan for tremor activity. Kao et al. (2007) used (1) the mean of the absolute amplitudes of 1.5- to 5.0-Hz band-pass-filtered signal and (2) the square root of the normalized variance of the signal intensity to recognize tremor. Sit et al. (2012) used the ratio of 2- to 5-Hz passband to 10- to 15-Hz and 0.02- to 0.1-Hz passbands to discard the influence of seismic activity and successfully identify tremor. Such frequency scanning approach is found to reduce erroneous signals and detect smaller tremor episodes unrecognized previously. Comparing with network-wide processing, the frequency scanning at a single station shows common identification of large episodic tremor and slip events in Cascadia with the detection capability up to 50 km from the station (Sit et al., 2012). However, the above single station detection methods are applied to hour-long tremor episodes. In our study area, the tectonic tremor lasts only several minutes, ranging from 1 to 37 min with the median of 4 min (Chen et al., 2018). Such short duration makes it difficult to avoid

influences of local earthquakes and cultural noise during the single station processing mentioned previously.

Advances in machine learning techniques allow for an automatic search for patterns in large data sets. In volcanic areas, the recognition of seismic signatures has long been studied for different classes of volcanic seismic events (Curilem et al., 2009; Masotti et al., 2006, 2008). Automatic classification methods have also been applied to differentiate blasts and earthquakes (Fäh & Koch, 2002; Laasri et al., 2015; Reynen & Audet, 2017; Vallejos & McKinnon, 2013); shallow and deep earthquakes (Mousavi et al., 2016); local, regional, and teleseismic earthquakes (Giudicepietro et al., 2017); rockfall, earthquakes, and noise (Hibert et al., 2017; Provost et al., 2016); and earthquakes and explosions (Kortström et al., 2016). Among the above-mentioned studies, a variety of supervised classifiers were used, including Random Forest (RF; Svetnik et al., 2003), Artificial Neural Network (Haykin, 1999), hidden Markov Models (Eddy, 1996), and support vector machines (Vapnik, 1998). The previous methods of single station detection in Cascadia focus only on the amplitude for tremor passband (1–5 Hz; Brudzinski & Allen, 2016; Kao et al., 2007; Sit et al., 2012), where the detection criteria heavily depends on station quality and signal-to-noise ratio. Using Convolutional Neural Network, Nakano et al. (2019) successfully discriminated tremors from local earthquakes and noise in spectral images. Which seismic features exactly are needed for tremor recognition remain unexplored? Taking advantage of stations deployed at a mountain area on the top of the tremor zone, we attempt to document a more complete tremor catalog and to investigate the single station technique for a minute-long tremor. In this paper, we demonstrate the performance of a supervised classifier k -nearest neighbor (k -NN) on a data set comprising of $M \geq 2$ local earthquakes, ambient tremor, and natural noise in Taiwan, to evaluate the possibility of detecting minutes-long tremor using a single station. Our main goal is to discriminate ambient tremor from local earthquakes and natural noise by investigating how this k -NN classifier performs with seismic data at different stations. We also aim to obtain a collection of efficient seismic features for successful classification in future monitoring systems.

2. Tectonic Tremor in Taiwan

Tectonic tremor in Taiwan have been previously identified for different purposes, while the detection schemes vary. Chuang et al. (2014) attempted to obtain the significant tremor activities in Taiwan based on the following detection criteria: (1) higher than waveform cross-correlation coefficient (ccc) > 0.95 between stations, (2) averaged signal-to-noise ratio greater than 1.15, and (3) durations longer than 300 s. In Idehara et al. (2014) and Ide et al. (2015), an envelope correlation hypocenter method was adopted that requires ccc higher than 0.5 between stations. After locating tremor using differential arrival time measurements that are derived from waveform cross correlation, a spatiotemporal clustering technique is applied to remove isolated events. Similar to the above method, Chao et al. (2017) used the waveform envelope correlation and clustering method by Wech (2010) to obtain a preliminary event catalog, following by a spatiotemporal clustering criterion to exclude ordinary earthquakes. Using temporary arrays, Sun et al. (2015) used both the broadband frequency-wave number beamforming and moving-window grid search methods to detect weak tremor signals. To detect tremor with stable azimuth and deep origin, they ask for (1) the detected duration longer than 3 min while the array envelop function is coherent among nearby stations, (2) the back azimuth from two methods that is consistent ($< 30^\circ$ difference), (3) wave number that should be smaller than 1.5 cycles per kilometer, and (4) waveform cross-correlation values that should be higher than 0.02. Chen et al. (2018) used slightly different criteria from Ide et al. (2015) and Chao et al. (2017) to improve the completeness of tremor catalog. In this study, we adopted the method in Chen et al. (2018) to include short-duration and isolated events and exclude events that are too short to be discriminated from earthquakes. Given that a more complete catalog leads to the involvement of short-duration tremor in Taiwan, it is challenging to prevent the misidentification of tremor, earthquake swarms, and loud noise.

Using three-component broadband seismograms recorded by the Broadband Array in Taiwan for Seismology, Central Weather Bureau Seismic Network (CWBSN), and a temporary array deployed on the southern Central Range from 1 January to 16 September 2016 (colored triangles in Figure 1), identification of tectonic tremors was carried out by two stages. At the first stage we used 28 stations from Broadband Array in Taiwan for Seismology, 12 stations from CWBSN, and three stations from the mountain array. The three-

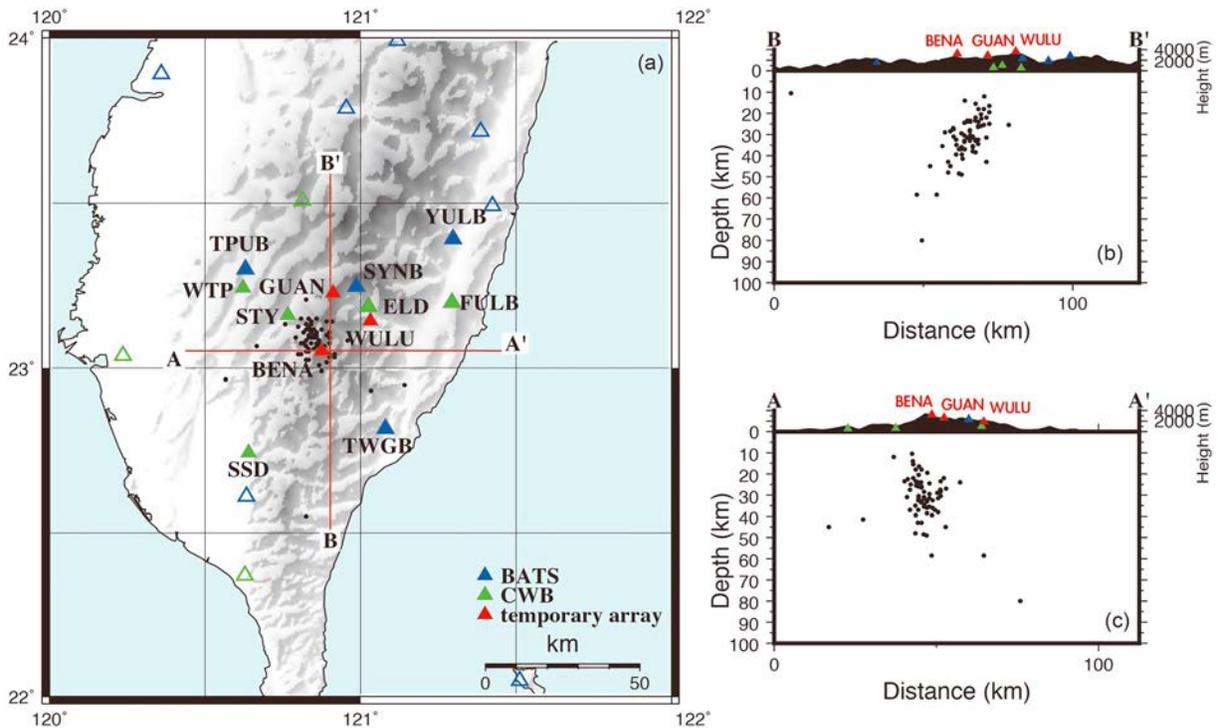


Figure 1. (a) Map of stations distribution and tremor catalogued in this study. (b) cross-sections A-A' and B-B' showing a southward dipping ellipsoidal structure. The stations deployed on the southern Central Range from 1 January to 16 September 2016 are denoted by red triangles (BN = station BENA; KA = station GUAN; WL = station WULU). Stations in the Broadband Array in Taiwan for Seismology (BATS) and Central Weather Bureau Seismic Network (CWBSN) are denoted by blue and green triangles, respectively. Open triangles indicate the stations used in envelope cross-correlation approach for early detection, while the filled triangles are the stations used for a visual inspection.

component seismograms were band-pass filtered from 2 to 8 Hz, enveloped, and low-pass filtered to 2 Hz. Using a 300-s moving window with a 150-s overlap, we then applied the detection criteria below: (1) Waveform ccc should be higher than 0.6 for more than 10 stations, and (2) a signal-to-noise ratio should be greater than 1.2. With these criteria, the events were obtained for hypocenter determination using envelope cross-correlation approach by Ide (2010). To avoid the occasional contamination of tremor signals by swarm of microearthquakes, a visual inspection was later applied using only the 12 stations near the tremor zone (colored triangles in Figure 1a) at the second stage.

As a result, 97 tremor events with duration ranging from 221–771 s were identified in the study period. Using the waveform envelope correlation and clustering location scheme from Wech and Creager (2008) and a local 1-D shear wave velocity model (Huang et al., 2014), the 97 tremor events were relocated (black dots in Figure 1). The distribution of tremor events reveals a NS striking and SE dipping ellipsoidal concentrating at the depth from 20–40 km. Such fault orientation is consistent with the SE dipping thrust Tulungwan fault that acts as a boundary between a slate belt of moderate metamorphic grade and a relatively unmetamorphosed fold-and-thrust belt (Chen et al., 2018; Huang & Byrne, 2014). At the stations deployed on the mountain (BENA, GUAN, and WULU), the tremor signals appear to be higher in amplitude comparing with other stations, as shown by the waveform examples in Figure 2. As Taiwan tremor is known for its short duration (1–37 min) and weak amplitude (Chen et al., 2018; Chuang et al., 2014), we attempt to investigate the single station technique that is well suited to detect the short and weak tremor. To do so, we select three stations for different classes of seismic signals to demonstrate the performance of a supervised classifier.

3. Labeled Data

Using the tremor catalog built in this study and earthquake catalog from CWBSN, three classes of events including local earthquake, tectonic tremor, and ambient noise were labeled. Figures 3a–3c show example

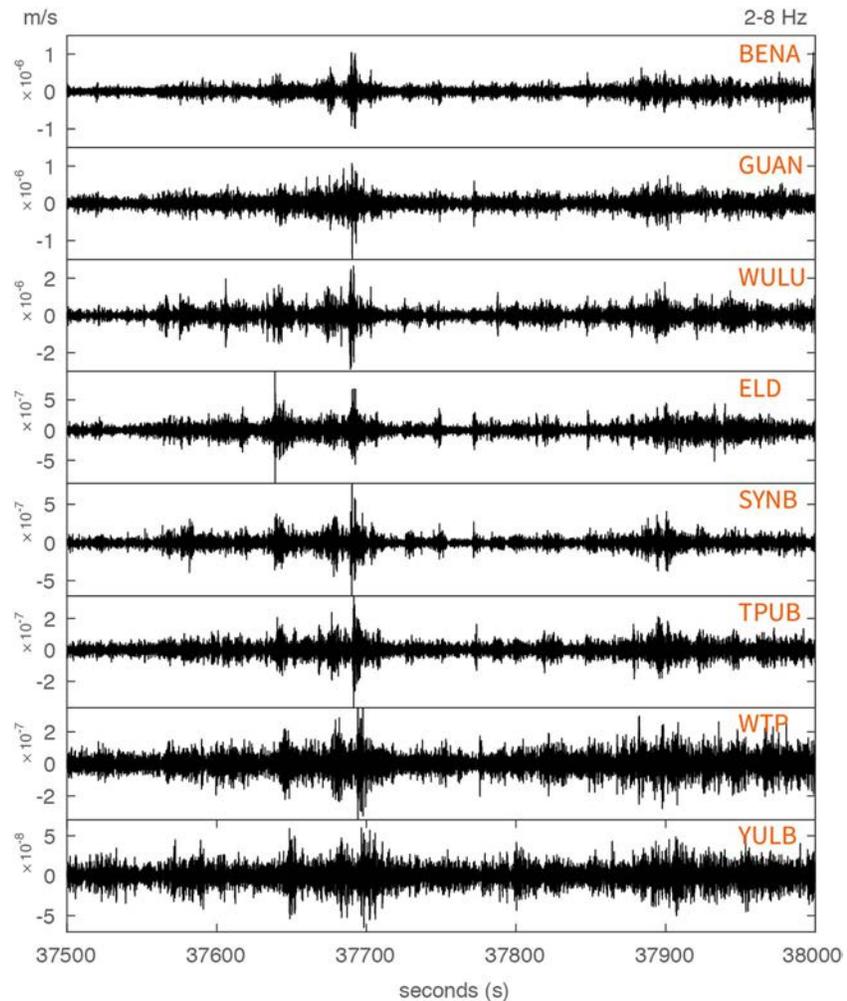


Figure 2. Example of the 2- to 8-Hz-filtered, E-W component velocity seismograms for a ~500-s-long tremor recorded at eight stations. This event occurred on 24 February 2016. For location of the seismic stations, please see filled triangles in Figure 1.

of their waveforms at three selected stations of BENA, GUAN, and WULU (blue triangles in Figure 3d). The stations BENA and GUAN are selected due to their short distance to Tremor, while YULB located further northeast is selected for comparison.

The first class of events, local earthquakes (hereafter called “Earthquake”), is composed of 291 $M \geq 2$ events with three components that were randomly chosen from the CWBSN earthquake catalog. To minimize the difference in path effect from different propagation distances, we only considered earthquakes with an epicentral distance less than 30 km (gray circles in Figure 3d) for three examined stations. We manually discarded earthquakes that were not distinguishable from noise to ensure the quality of the training data. The resulting Earthquake population shows that 78%, 70%, and 87% of events are smaller than $M3$ for BENA, GUAN, and YULB, respectively, while 100%, 100%, and 44% events are shallower than 20 km for BENA, GUAN, and YULB, respectively.

The second class, ambient tremor (hereafter called “Tremor”), was identified during the deployment period from 1 January to 16 September 2016 (as described in the previous section). We use the time window of 60 s that is comparable to the shortest duration in the published 2007–2012 tremor catalog. Note that, unlike Earthquakes, minute-long Tremor is difficult to be identified by a single station, given that their amplitude is similar to Noise in the 2- to 8-Hz frequency range (Figures 3f and 3g).

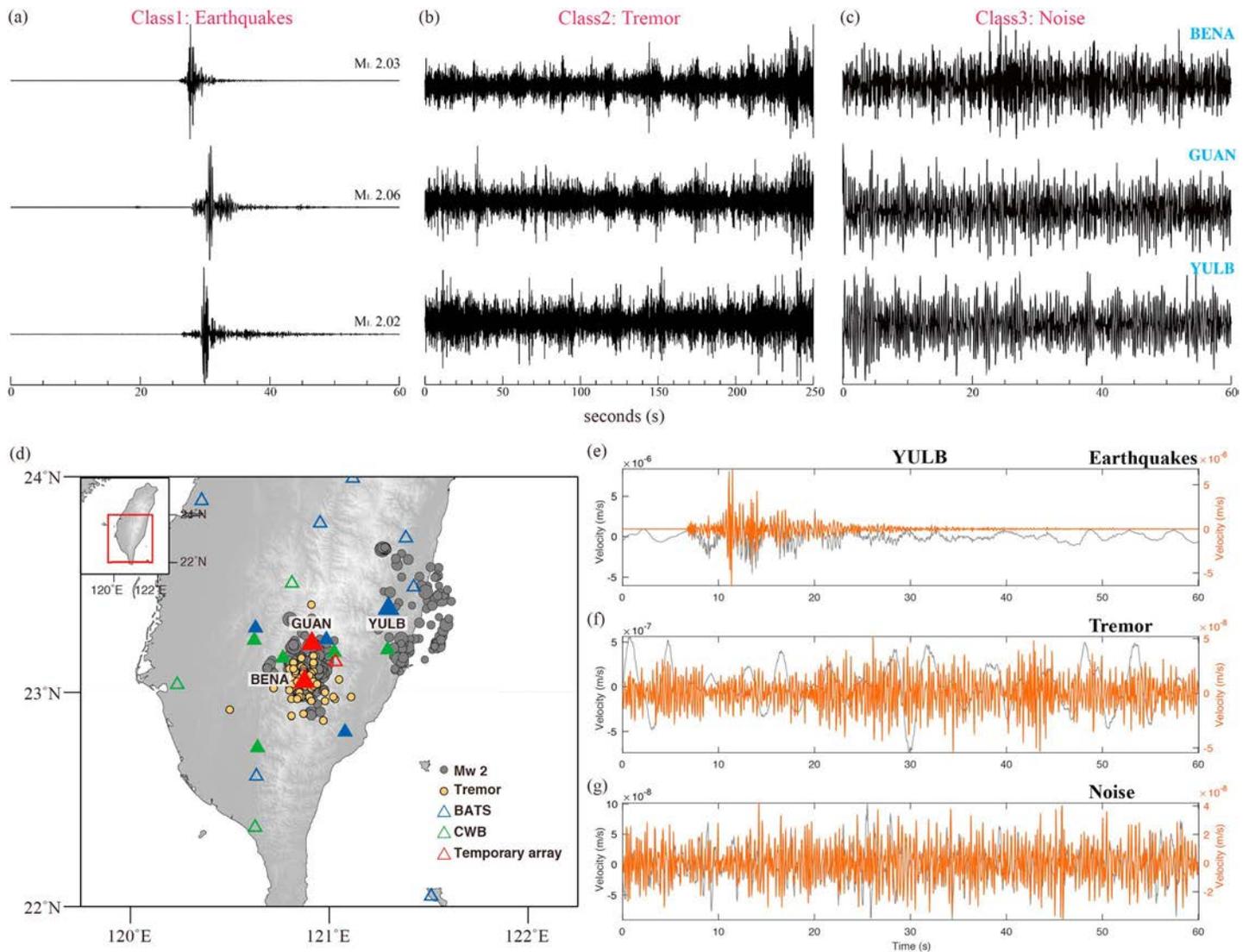


Figure 3. Examples of 2- to 8-Hz-filtered, N-S component waveforms for (a) Local earthquakes, (b) Ambient Tremors, and (c) Natural Noise. (d) Map of the study area (denoted by a red rectangle in the upper left diagram). Gray and yellow circles represent the labeled local earthquakes and ambient tremors, respectively. Open triangles indicate the stations used in envelope cross-correlation approach for early detection, while the filled triangles are the stations used for a visual inspection. Stations used in the classification are GUAN, BENA, and YULB, as denoted by the text. (e–g) Sixty-second waveform examples at the YULB station for three classes of data. Gray and red traces show raw data and 2- to 8-Hz-filtered waveforms, respectively.

The third class, ambient noise (hereafter called “Noise”), was selected 60 s before the arrival of the 291 events in class Earthquake and manually checked to exclude small earthquakes. Based on the timing of the local earthquakes, most of the Noise occurred in daytime.

Once the three classes of data were identified and labeled, three components of the waveforms were prepared (291×3 events for each class) and cut into 60-s window for the computation of the seismic features (detailed in section 5). In each class and each component, the 204 of the total 291 events were selected for training data, while 87 events were saved for test data. The additional set of 87×3 events was used to evaluate the single-station *k*-NN classifier.

4. *k*-NN Classifier

From a pattern-recognition point of view, a simple classifier is preferred for feature evaluation, since sophisticated classifiers (e.g., support vector machine, hidden Markov model, and Neural Net) may compensate for

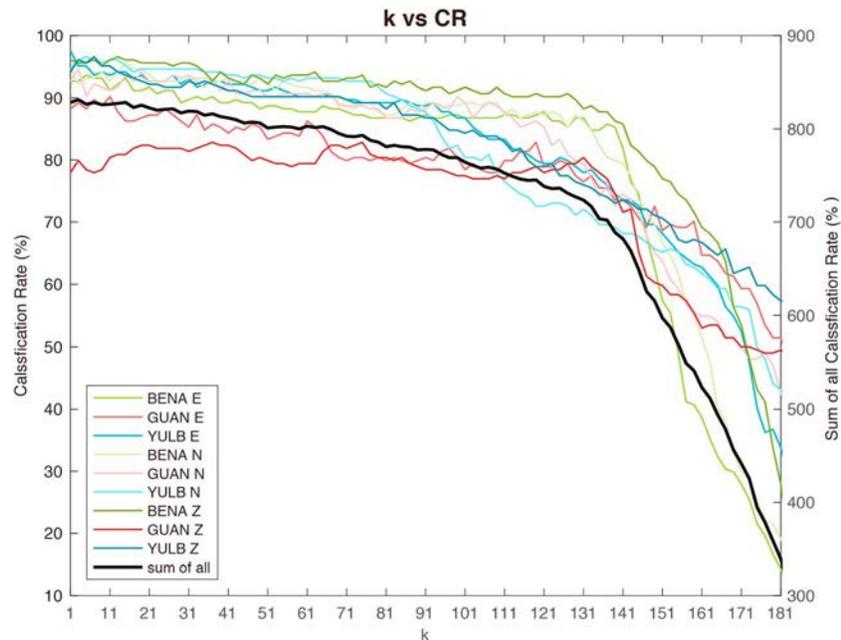


Figure 4. Classification rate as a function of k . Different color represents data from different station and component, while black curve represents the sum of data points from different stations and components.

the weakness of a feature and be difficult to reflect the true classification capability of the feature. Therefore, the commonly used k -NN method is chosen as the classifier herein.

The k -NN classification algorithm is implemented in two steps (Mitchell, 1997): (1) Define a training set containing set of N -dimensional data points; (2) given a test data point \mathbf{x} to be classified, let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$ denote the k data points from the training set that are nearest to \mathbf{x} ; then, return the class that the majority of the k -NN are from. The nearest neighbors of \mathbf{x} are defined in terms of the standard Euclidean distance. Moreover, to avoid possible tie situations, the free parameter k should be set as an odd number. If k is set as 1, the k -NN classifier becomes the nearest-neighbor classifier, which might suffer from the overfitting problem due to outliers. The performance of classifier can be evaluated by Leave-One-Out Cross Validation strategy (LOO-CV). LOO-CV uses a single observation from a given data set as the test data, and the remaining observations as the training data. This is repeated such that each observation in the data set is used once at a time as the test data (Efron & Tibshirani, 1993) to infer the classification rate (CR). In Figure 4 we performed a sensitivity test on the selection of k . There appears to show a general trend of decreasing CR with increasing k except for the Z component at GUAN station where the signal-to-noise ratio is lowest as <0.1 . When the CR from different components and stations are summed, the highest cumulative CR occurs when $k = 3$. Accordingly, we set k as 3 in the present study, as the same setting in some studies related to pattern recognition (Liao et al., 2017).

In this study, we extended the binary k -NN classifier to solve the three-class classification problem of Tremor versus Earthquake versus Noise by using one-against-one method and a majority voting strategy (Liu & Chen, 2007). For an L -class classification, $L(L - 1)/2$ binary classifiers need to be performed. In our experiments, test data are sent into three k -NN classifiers, one for Tremor versus Earthquake, one for Earthquake versus Noise, and the remaining one for Tremor versus Noise. The final classification output for these test data is then determined based on the majority of the three class labels generated from the three k -NN classifiers.

5. Seismic Signal Features

Signal features are chosen based on seismic signatures that are commonly exploited for event classification. We follow the features proposed in Provost et al. (2016) and Hibert et al. (2017) where the signals of rockfall, earthquakes, and noise were classified using more than 60 seismic features. The selected features are

associated with waveform, spectral content, spectrogram content, polarization, and network geometry. In this study, the lack of station coverage near the source area may produce large uncertainty in describing source and station property. The selected features here are simplified only to waveform, spectral, and spectrogram content with the time window and frequency range that is suitable for the signals of interest. To train the model, 27 seismic features are 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).

The first family of features is focused on seismic waveforms. As shown in Figure 3a, while Earthquake is characterized by emergent onset, short duration, and clear peak amplitude with exponential decay of energy, Tremor, and Noise usually exhibit no sharp arrival, no clear peak amplitude with long duration (Figures 3b and 3c). The parametrized features include the ratio of the mean and median over the maximum for the envelope signal, the maximum envelope amplitude, the time of the maximum envelope amplitude, the kurtosis and skewness of the raw signal and envelope, and the energy in the first third and the remaining parts of the autocorrelation function, as identified by features 1–8 in Table 1.

The second family of features focuses on the spectral content of the signals (features 9–19). Tremor is generally deficient at a high frequency band (>1 Hz). Compared with the background noise of the local night time, Tremor are enriched over 1–10 Hz with a higher amplitude; the background noise of the local daytime, however, reveals very similar patterns that makes the identification of Tremor in Taiwan somewhat challenging. This is the reason why more Tremor were detected at night due to the lower noise level (Chao et al., 2011; Chuang et al., 2014). Figure 5a shows a comparison of the spectra using 60-s window signals at station GUAN, the curve represents the medium value from 97 events for each class and each component. Earthquake appears to be rich in higher frequency range (>0.5 Hz), which is distinguishable from the other two classes. In a range of 2–8 Hz, Tremor can be further separated from Noise by showing relatively higher amplitude. In contrast, Noise shows higher amplitude than Tremor in lower frequency band below 0.5 Hz. This large-amplitude microseism peak at 0.2–0.5 Hz is commonly seen at land stations and the ocean-bottom seismometers in the offshore area of eastern Taiwan (Lin et al., 2010). The difference between classes, however, is only visible using medium value. Plotting all 97 events in the same diagram reveals strong variation (Figure 5b), suggesting the difficulty of finding a criterion in spectrum for classification. We computed the maximum amplitude of 2–8 Hz and number of peaks in the 2- to 8-Hz-filtered envelope (features 12 and 13) that are commonly considered by human operators for tremor detection. We also consider the energy and kurtosis of the signal over three frequency bands (0.1–1, 2–8, and 0.05–0.1 Hz). After computing the discrete Fourier transform (DFT) of the signals, we also calculated the kurtosis of the maximums for all DFTs as a function of time. The mean ratio between the maximum and mean and the maximum and median for all DFTs was calculated as well.

The third family of features focuses on spectrogram that have three dimensions in time, frequency, and energy (features 20–27). Earthquake spectrogram reveals a sharp increase of high frequency energy (>10 Hz) at arrival of signal, followed by an exponential decay of energy. Three different components reveal similar characteristics in spectrum. Tremor and Noise are deficient in high frequency, while Tremor is relative strong. Using DFT, we computed the energy concentration in the frequency domain by counting the number of peaks in the curve to show the temporal evolution of the DFT maximums, means, and medians.

It is noticed that the selected features have different units and presumably each feature takes a different range of values. In the present study, the feature values were normalized to have zero mean and unit variance using the z -score normalization method before the features are fed into the k -NN classifier. Assuming f is a feature value of a specific feature, the normalization is given by $(f - m)/sd$, where m and sd are the mean and standard deviation of all the values of that feature in the training data, respectively. Note that data scaling plays a critical role in data preprocessing (Cao et al., 2007). Models trained on scaled data usually have higher classification performance compared to the models trained on unscaled data,

6. Fisher's Class Separability Criterion

To evaluate the classification performance of the feature candidates and select an optimal feature set by using Fisher's class separability criterion, we aim at finding the top-ranked features that achieve the

Table 1
Computed Features for Training Data

ID	Description	
Temporal waveforms (1st family)		
#1	Ratio of the mean over the max of the envelope	—
#2	Ratio of the median over the max of the envelope	—
#3	Maximum envelope amplitude	—
#4	Time of the maximum envelope amplitude	—
#5	Kurtosis of the raw signal	$\frac{E(x-\mu)^4}{\sigma^4}$, where μ is the mean of x , σ is the standard deviation of x , and $E(t)$ represents the expected value of the quantity t . (using MATLAB R2017b inbuilt function kurtosis)
#6	Kurtosis of the envelope	See #5
#7	Skewness of the raw signal	$\frac{E(x-\mu)^3}{\sigma^3}$, where μ is the mean of x , σ is the standard deviation of x , and $E(t)$ represents the expected value of the quantity t . (using MATLAB R2017b inbuilt function skewness)
#8	Skewness of the envelope	See #7
Spectral content (2nd family)		
#9	Energy in the first third of the autocorrelation function	$\int_0^{\frac{T}{3}} C(\tau) d\tau$, with T : signal duration, C : autocorrelation function. (using autocorrelation function code by Calvin Price from MathWorks)
#10	Energy in the remaining parts of the autocorrelation function	See #9
#11	Ratio of the above two features	—
#12	Maximum amplitude of the 2- to 8-Hz waveform	—
#13	Number of peaks of the 2- to 8-Hz-filtered envelope	(using MATLAB R2017b inbuilt function findpeaks)
#14–16	Energy of the signal filtered at 0.1–1, 2–8, and 0.05–0.1 Hz	$\int_0^T y_f(t) dt$, with y_f : filtered signal in the frequency range ($f1$ – $f2$)
#17–19	Kurtosis of the signal at 0.1–1, 2–8, and 0.05–0.1 Hz	See #5
Energy concentration in frequency and time (3rd family)		
#20	Kurtosis of the maximums of all discrete Fourier transforms (DFTs) as a function of time	DFT: discrete Fourier transform Kurtosis ($\max_{t=0:T} [SPEC(t, f)]$) with $SPEC(t, f)$: the spectrogram
#21	Mean ratio between the maximum and the mean of all DFTs	$\text{mean}\left(\frac{\max(SPEC)}{\text{mean}(SPEC)}\right)$
#22	Mean ratio between the maximum and the median of all DFTs	See #21
#23	Number of peaks in the curve showing the temporal evolution of the DFTs maximum	(using MATLAB R2017b inbuilt function findpeaks)
#24	Number of peaks in the curve showing the temporal evolution of the DFTs mean	(using MATLAB R2017b inbuilt function findpeaks)
#25	Number of peaks in the curve showing the temporal evolution of the DFTs median	(using MATLAB R2017b inbuilt function findpeaks)
#26	Ratio between features #23 and #24	—
#27	Ratio between features #23 and #25	—

Note. The bold text is simply the function name in the matlab program.

highest classification accuracy. Fisher class separability criterion is a filter method for feature selection in the machine learning community (Fang et al., 2015). The Fisher method computes the Fisher score (F score) for an individual feature that is defined by the ratio of interclass scatter and intraclass scatter; the higher the value of the F score, the higher the between-class separability of the feature. Accordingly, the features can be ranked based on their F scores. For more details of the formula of the two scatter matrices, please refer to the article by Liu et al. (2017).

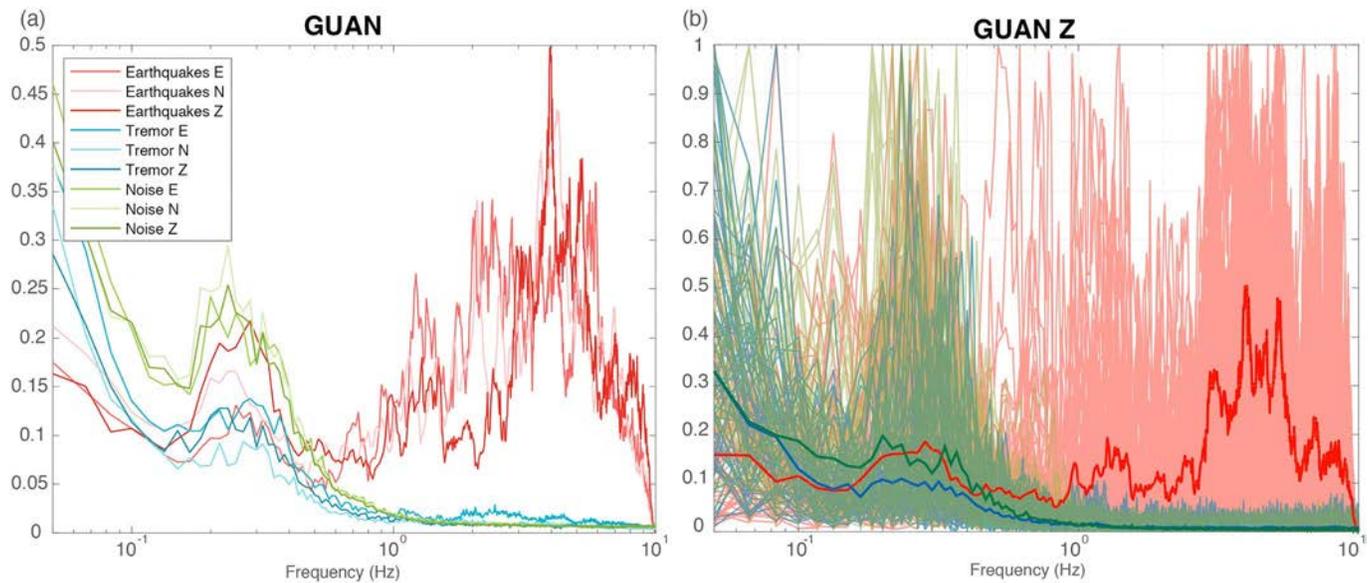


Figure 5. (left) Averaged Fourier spectrum of the broadband waveforms at GUAN station for Earthquake (red), Tremor (blue), and green (Noise). Three different components are denoted by slightly different colors. For each class and each component, 97 events are normalized with the maximum as 1. (right) Plot of Z-component spectrum for each class (thin lines) and median value (thick line) for each frequency bin.

7. Results

7.1. Experiment I: All 27 Features on the Training Data

Using the 27 features listed in Table 1, we used three components of data from stations BENA, GUAN, and YULB and evaluated the classification performance of the features using the k -NN classifier set at $k = 3$. We designed and conducted three different experiments. In Experiment I, we applied LOO-CV on the training set consisting of 204 labeled events for each class and each station. From these events, one is chosen as test data, while the remaining events are treated as training data for the calculation of classification accuracy (CR). This LOO-CV procedure is repeated until each event in the data set is used once as the test data (Efron & Tibshirani, 1993). Every misclassified event gives a classification error of 0.16% (1/612). In this experiment, all 27 features are extracted from every event; each event is represented by a 27-dimension feature vector. In other words, the experiment was conducted to test the CR on the training set without feature selection.

The confusion matrix in Table 2 shows that at station YULB, 75 Tremor are misclassified as Noise, while 97 Noise are misclassified as Tremor. At station GUAN, 38 Tremor are misclassified as Noise, while 44 Noise are misclassified as Tremor. At station BENA, 13 Tremor are misclassified as Noise, while 19 Noise are misclassified as Tremor. This leads to the highest CR at station BENA (93.7%) and lowest at stations YULB (68.6%). To explain the strong variation of CR among stations, we further examine the variability in waveforms and spectra by computing ccc for event pairs (Figure 6) and coefficient of variation (cv) in each frequency bin of 0.01 Hz (Figure 7), respectively. Here the cv is determined by the standard deviation of measurements by the mean value. We hypothesize that, the higher variation of data in the same class would result in distinct value for a given feature (in time or frequency domains), leading to higher misidentification rate. In Figure 6, we found that 1.5%, 5.5%, and 0.2% of Noise signals have cc higher than 0.5 for BENA, GUAN, and YULB station, respectively, while 37.3%, 33.6%, and 53.3% of Noise signals show low cc (<0.1). The variation in frequency domain appears to be significantly larger at station YULB, as reflected by $cv > 2$ at lower frequency band (<0.1 Hz). At higher frequency band (>0.2 Hz) the cv remain smaller than 1.5 at three stations. Summarizing above, small waveform similarity and large amplitude difference in each frequency bin can be the diagnosis for higher variability of Noise signals at station YULB. This may produce higher interclass scatter in seismic features selected in this study to produce lower CR. The same hypothesis, if applied to Tremor signals, the differences in waveform similarity and spectral behavior are minor comparing with Noise.

Table 2
Comparison of Classification Rate for Different Experiments Using 2016 Tremor Catalog

		BENA				GUAN				YULB			
		Eq.	Tremor	Noise		Eq.	Tremor	Noise		Eq.	Tremor	Noise	
I	Earthquake	204	0	0	93.7	203	0	1	86.4	187	11	6	68.8
	Tremor	3	188	13		0	166	38		2	127	75	
	Noise	3	19	182		0	44	160		0	97	107	
II ^a	Earthquake	204	0	0	98.8	203	1	0	89.4	195	7	2	86.6
	Tremor	0	200	4		0	168	36		1	164	40	
	Noise	0	7	197		0	34	170		0	45	159	
III ^b	Earthquake	86	0	1	95.7	86	0	1	70.8	81	5	0	76.6
	Tremor	1	78	8		0	29	58		0	46	41	
	Noise	0	1	86		0	17	70		0	14	73	

^aExperiment I: using Leave-One-Out Cross Validation on training data and all 27 features. ^bExperiment II: using Leave-One-Out Cross Validation on training data and features selected using Fisher score. ^cExperiment III: using test data and features selected by Fisher score. The bold text is simply the function name in the matlab program.

7.2. Experiment II: Selected Features on the Training Data

7.2.1. Feature Ranking

In Experiment II, the seismic features were selected based on the Fisher's class separability criterion. As shown in Figure 8a, at BENA station, the features with highest F scores that discriminate Earthquake from Tremor are #1, #2, #8, and #9 corresponding to the ratio of the mean and median over the maximum of the envelope, skewness of the envelope, and energy of the first third of the autocorrelation function. The fifth-ranked feature #6 has much lower F score (less than half of the maximum in this plot) and therefore is regarded as less significant and neglected in the comparison below. In Figure 8b, to discriminate Earthquake from Noise, the features with top three F scores are #1, #2, and #8 that are the same with the top three in Figure 8a. For Tremor versus Noise classification (Figure 8c), the top F scores that are higher than half of the maximum correspond to the features #25, #15, #12, and #13. They are number of peaks in the curve showing the temporal evolution of the DFTs median (#25), energy of the signal filtered at 2–8 Hz (#15), maximum amplitude of the 2- to 8-Hz waveform (#12), and number of peaks of the 2–8 Hz filtered (#13). None of these are common to the high-ranking features for Earthquake versus Noise and Earthquake versus Tremor. For Tremor versus Noise classification, the highest F score (2.8) appears to be much lower than those for Earthquake versus Tremor (7.9) and Earthquake versus Noise (10.5), indicating a low discriminability comparing with other two class pairs.

Two other stations in Figures 8d–8i reveal similar high- F score features. That is, the high-ranking features at station BENA (#1, #2, #8, and #9) are commonly found at stations GUAN and YULB for Earthquake versus

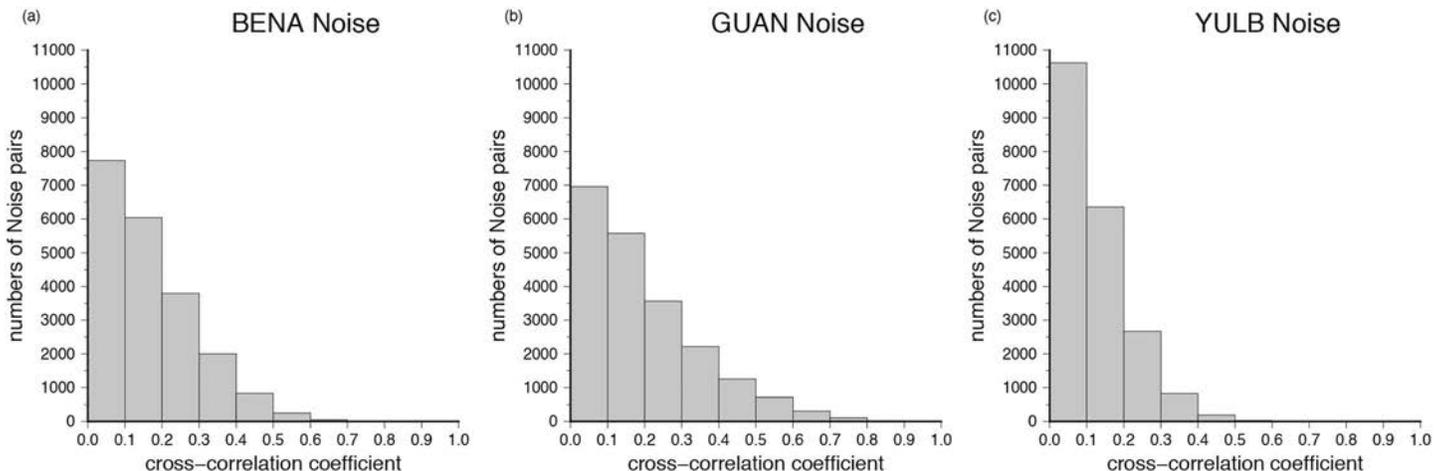


Figure 6. Histogram of cross-correlation coefficient between Noise event pairs for each station.

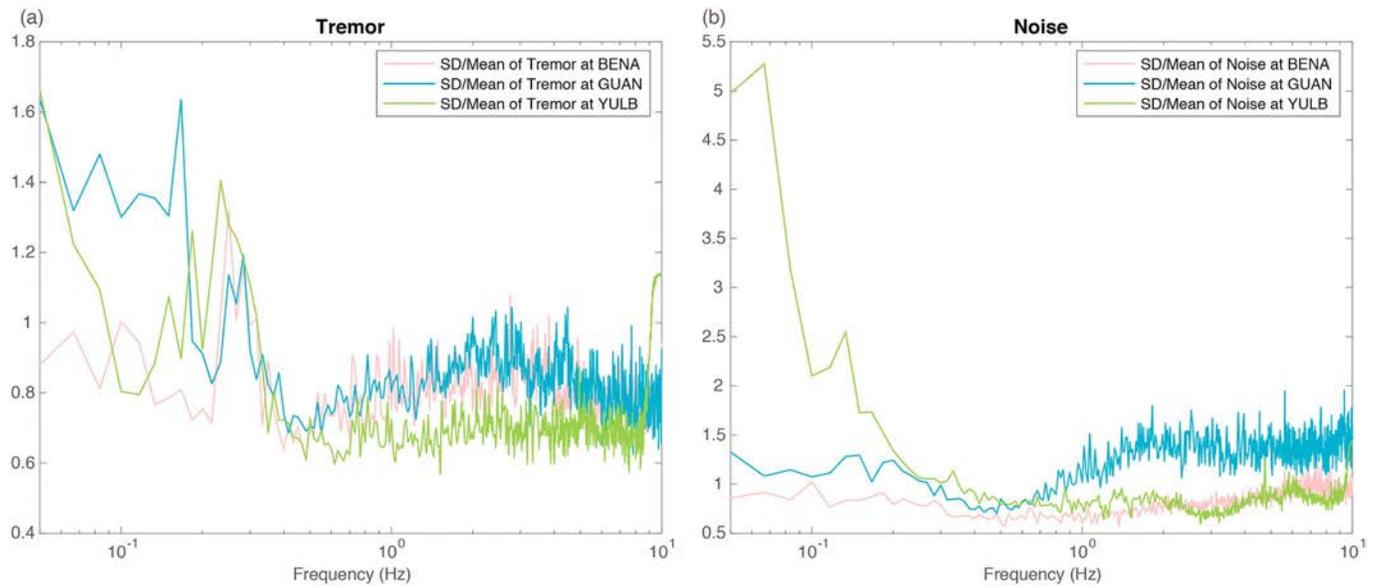


Figure 7. Coefficient of variation in amplitude for each frequency bin (0.01 Hz) for Tremor (a) and Noise (b). The coefficient of variation is determined by the standard deviation (SD) divided by the mean.

Tremor and Earthquake versus Noise. For Tremor versus Noise, the top F scores go to the features #25, #13, #12, and #15 at station GUAN and #12 and #15 at station YULB, which shares common features with station BENA (#25, #15, #12, and #13). As the F score represents discriminability between classes, the extremely low F score at stations GUAN (<0.9) and YULB (<0.4) could be associated with the high variation in Noise (as revealed in Figure 7) that increases the interclass scatter.

7.2.1.1. Feature Selection

To select features that improve the classification accuracy, we perform LOO-CV to calculate the classification accuracy using the top- N - F score-ranked features. More precisely, we compute 27 times of CR for each station and each component, starting with top-ranked feature. In each run we add a lower-ranked feature; once the CR is higher than the previous calculation, this feature is kept in the best feature subset. This is called add-one-feature-in strategy for reaching the best N features (Lin et al., 2010; Wu et al., 2018). The optimal feature subset refers to the top- N - F score-ranked features that give the highest LOO-CV classification accuracy for $1 \leq N \leq 27$. The first feature that gives a lower LOO-CV is determined to be the cutoff. For more details, please refer to the studies by Lin et al. (2010) and Wu et al. (2018).

For different stations and different binary classification models (Earthquake vs. Tremor, Earthquake vs. Noise, and Tremor vs. Noise), the optimal feature subsets could be different; namely, the numbers of best features N could be different. The N features required for improving CR are colored squares in Figure 9, while the number in the box indicates the ranking. As can be observed, the top-ranked features are mostly common over stations for Earthquake versus Tremor. If the top five ranked features appear at three stations, we regard them as efficient features. The efficient features for discriminating Earthquake from Tremor are #1, #2, #8, and #9, the same with those for discriminating Earthquake from Noise. The efficient features for discriminating Tremor from Noise are #12, #13, #15, and #25. Note that #25 is also considered efficient because it performs as top one feature at two stations, even not recognized as important at station YULB. These four features are distinct from the features needed to classify Earthquakes. Summarizing the above, for a binary classification models and for a specific station, the steps in Experiment II are (1) using the training set to compute the F score of each feature, (2) ranking the features according to their F scores, and (3) using the add-one-feature-in strategy to find the best feature subset (i.e., the top N features).

The resulting confusion matrix is shown in Table 2. At station YULB, 40 Tremor and 1 Earthquake are misclassified as Noise, while 45 Noise are misclassified as Tremor. At station GUAN, 36 Tremor are misclassified as Noise, while 34 Noise are misclassified as Tremor. At station BENA, four Tremor are misclassified as

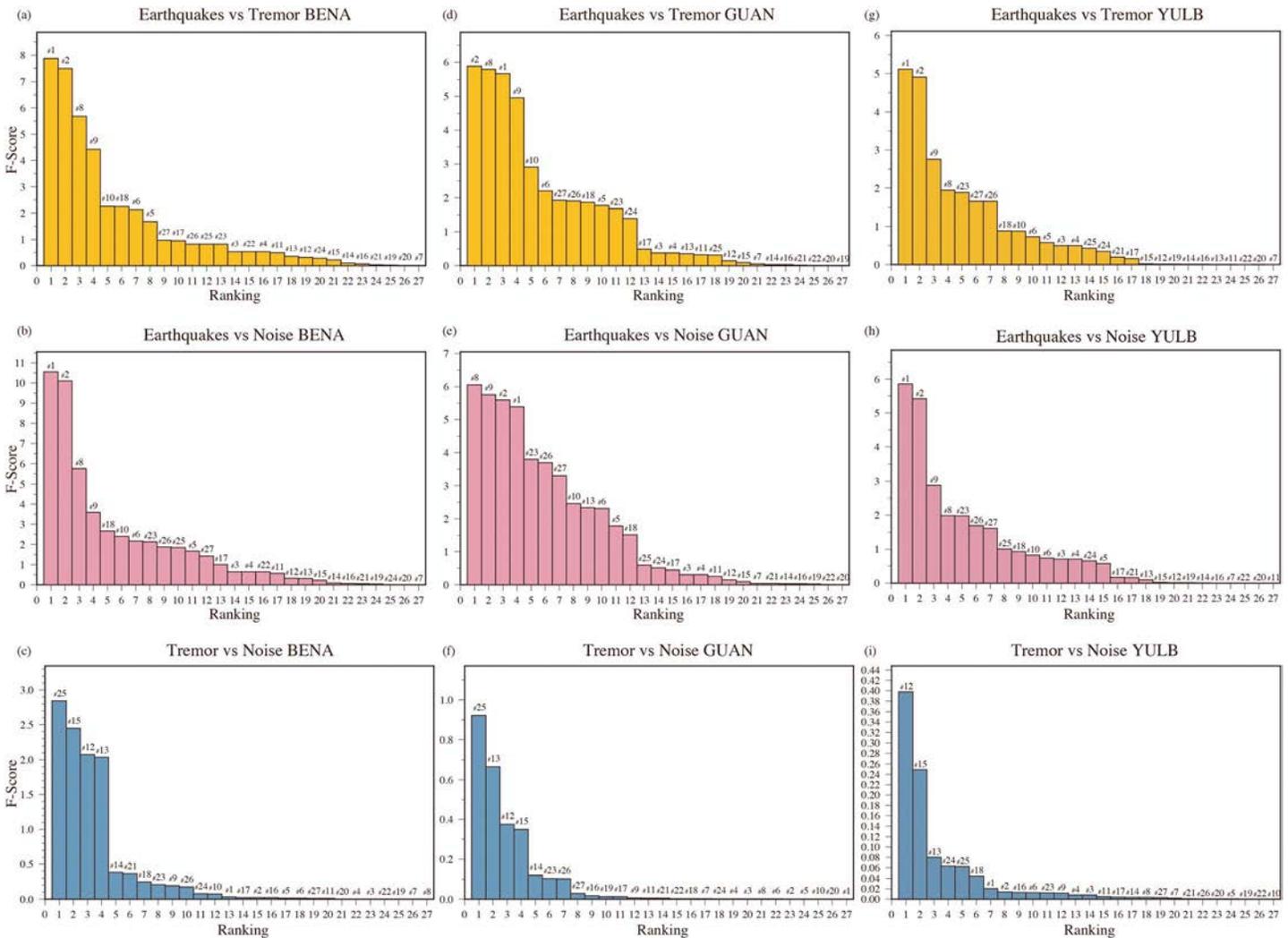


Figure 8. Fisher score as a function of feature ranking for the training data at stations BENA (a–c), GUAN (d–f), and YULB (g–i). (upper panel) Earthquakes versus Tremor. (middle) Earthquake versus Noise. (lower) Tremor versus Noise.

Noise, while seven Noise are misclassified as Tremor. However, the CR is progressively improved for Experiment II (86.6–98.8% with average of 91.6%) using selected features ranked by the F score.

7.3. Experiment III: Selected Features on the Test Data

In Experiment III, we fed all the test data consisting of 261 events for each station into the k -NN classifier, where the training data are the ones used in Experiment II. Note that the test data do not participate in the calculation of the F scores for another run of feature selection. For each binary classification model and each station, each test event is represented by an N -dimensional feature vector in which the elements are the optimal feature subset determined in Experiment II.

With the test data, Experiment III reveals an averaged CR of 81.0% due to the low CR at station GUAN (70.8%; Table 2). At station YULB, 41 Tremor are misclassified as Noise, while 14 Noise are misclassified as Tremor. At station GUAN, 58 Tremor are misclassified as Noise, while 17 Noise are misclassified as Tremor. At station BENA, eight Tremor are misclassified as Noise, while one Noise are misclassified as Tremor. The high misclassification rate for Tremor at station GUAN might be most responsible for the low CR of 70.8%.

three components																												
CI vs C2																												
Station \ Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
BENA	1	2			8	7		3	4	5							10	6								11	9	
GUAN	3	1	14	15	10	6	21	2	4	5	17	19	16	22	20	23	13	9			24	25	11	12	18	8	7	
YULB	1	2	12	13	11	10		4	3	9		3						8					5		14	7	6	
CI vs C3																												
Station \ Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
BENA	1	2						3	4									5										
GUAN	4	3	16	17	11	10	21	1	2	8	18	19	9	23	20	24	15	12	25				5	14	13	6	7	
YULB	1	2	12	13		11		4	3	10								9					5		8	6	7	
C2 vs C3																												
Station \ Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	
BENA									9	12		3	4	5	2			7			6		8	11	1	10		
GUAN			20	19			17		12		13	3	2	5	4	9	11	16	10			14	15	6	18	1	7	8
YULB												1	3		2									4				

Figure 9. Feature ranking result based on Fisher score for different binary classifications: Earthquakes (C1) versus Tremors (C2), Earthquakes(C1) versus Noise (C3), and Tremors (C2) versus Noise (C3). For each classification all the 27 features were extracted from the signals. The number in each square indicates the ranking based on Fisher score. The higher the F score of a feature is, the lower the number is (and higher ranking). The colored squares indicate the best subset of features determined by Leave-One-Out Cross Validation strategy. For example, for Earthquake versus Tremor classification the best three-feature subset for BENA station is {#1, #2, #8}.

Using the same selected features, the CR of the training model in Experiment II is higher than that in the test data (Experiment III). The F score performance for test data (Figure 10), although not used for reselection of features, can be used to compare with the training data (Figure 8) for the similarity and difference. The low CR in Experiment III can be reflected by low F score for Tremor versus Noise as 0.75 at station GUAN and 0.30 at station YULB in Figure 10. The high-ranking features for Earthquake versus Tremor and Earthquake versus Noise in Figures 8 and 10 are highly similar with those for Experiment II (#1, #2, #8, and #9). However, for discriminating Tremor and Noise, the high-ranking features in Experiment III are not the same with those discovered in Experiment II. At station BENA, the top four ranked features in Experiment III coincide with the efficient features selected in Experiment II (#25, #13, #15, and #12 in Figures 10c, 10f, and 10i) but not with those identified at stations GUAN and YULB. The less consistent high-ranking features for Tremor versus Noise may explain the lower CR at stations GUAN and YULB in Experiment III (Table 2).

Despite the lower F score for Tremor versus Noise in Experiment III (and therefore lower CR), the common high-ranking feature #25 is observed at three different stations for Tremor versus Noise. This feature, number of peaks in the curve showing the temporal evolution of the DFTs median, is also found to be top one ranking at stations BENA and GUAN in Experiment II (Figures 8c and 8f). This suggests that, different from #12, #13, and #15 identified in Experiment II that are all associated with characteristics of the 2- to 8-Hz-filtered signals, the spikiness of energy in time-frequency domain could be important for discriminating Tremor from Noise.

8. Discussion

8.1. Seismic Features Useful for Tremor Detection

Very recently an improved Convolutional Neural Network method was applied on running spectral images to identify tectonic tremor, local earthquake, and noise by Nakano et al. (2019). They successfully obtained a high accuracy up to 99.5% using all components and stations available in the tremor zone. However, they pointed out that the tremor signals strongly contaminated with noise, if included in the training data set, would produce much lower recall down to 67.9%. This challenge was met here as the Tremor in Taiwan being highly similar with Noise in amplitude (Chuang et al., 2014) and spectral behavior (Figure 5).

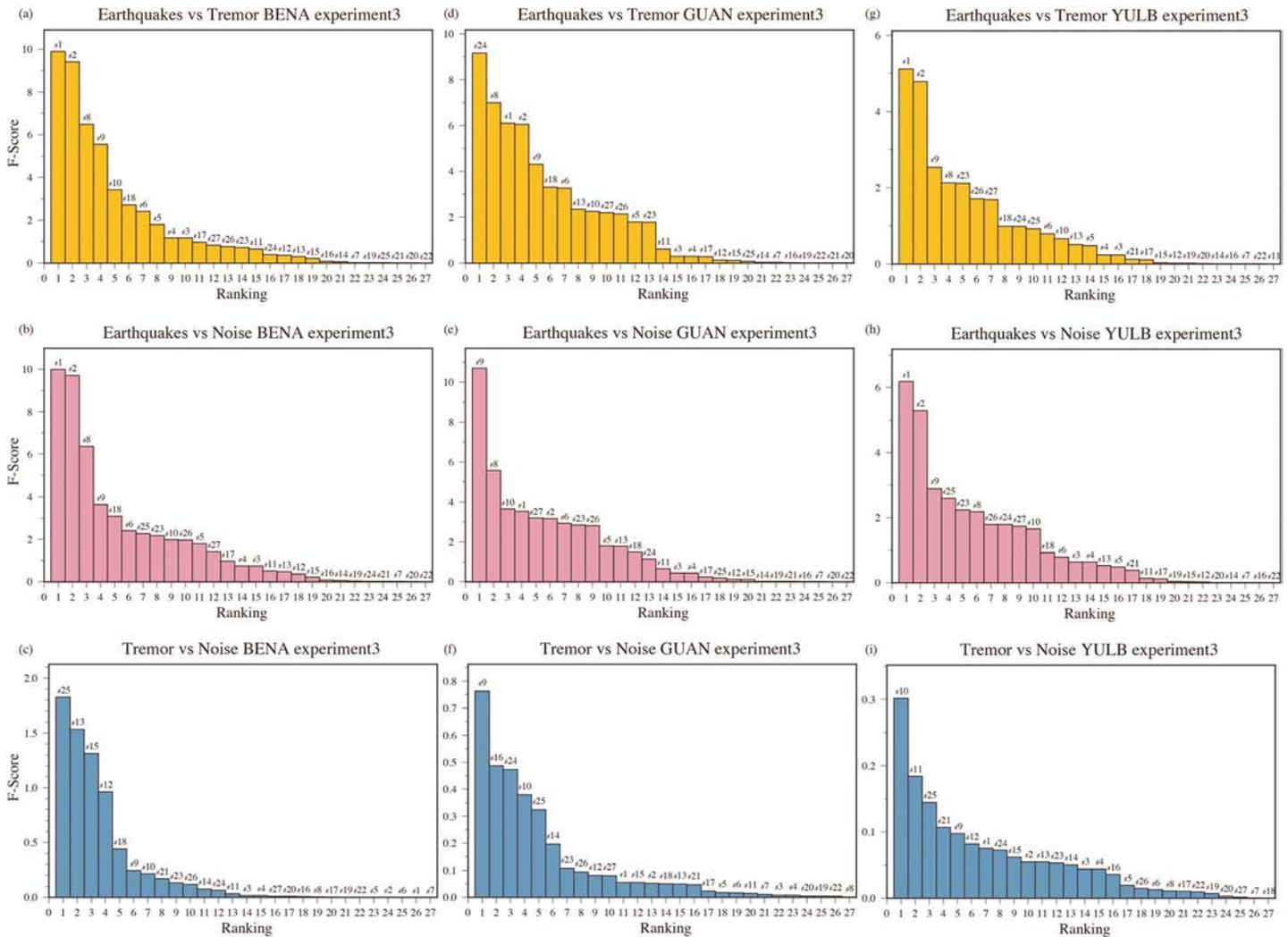


Figure 10. Fisher score as a function of feature ranking for the test data at stations BENA (a–c), GUAN (d–f), and YULB (g–i). (upper panel) Earthquakes versus Tremor. (middle) Earthquake versus Noise. (lower) Tremor versus Noise.

Different from Nakano et al. (2019) where the features selection was not conducted, this study attempts to select the best features using the Fisher's criterion.

Tremor in Taiwan have been previously identified using different study periods with different purposes (Chao et al., 2017; Chen et al., 2018; Chuang et al., 2014; Ide et al., 2015; Idehara et al., 2014; Sun et al., 2015). The common practice of tremor identification is based on waveform similarity, time lapse, and event duration using three components for 2- to 8-Hz-filtered envelope data at multiple stations. Using machine learning techniques, we found that four features that allow us to successfully separate Tremor from Noise, including the maximum amplitude, number of peaks, and energy of the 2- to 8-Hz-filtered signals (#12, #13, and #15) and number of peaks in the curve showing the temporal evolution of the DFTs median (#25). Among the above features, #25 is the one that has never been previously considered in tremor identification schemes. Figure 11a shows the clustered Tremor and Noise data over the feature space for #13 and #25. The limited overlap suggests that the selected features are useful to separate Tremor from Noise, while the optimal boundary seems to be nonlinear. Figures 11b–11d show the example of maximum, mean, and median of DFTs for three classes of events at station BENA. Comparing with Noise, Tremor has highly similar shape in time evolution of the DFTs but smaller magnitude. In this example, the number of peaks for three measurements of DFTs shows distinct value. Among features #23 (DFTs maximum), #24 (DFTs mean), and #25 (DFTs median), only #25 show significant difference between Tremor and Noise.

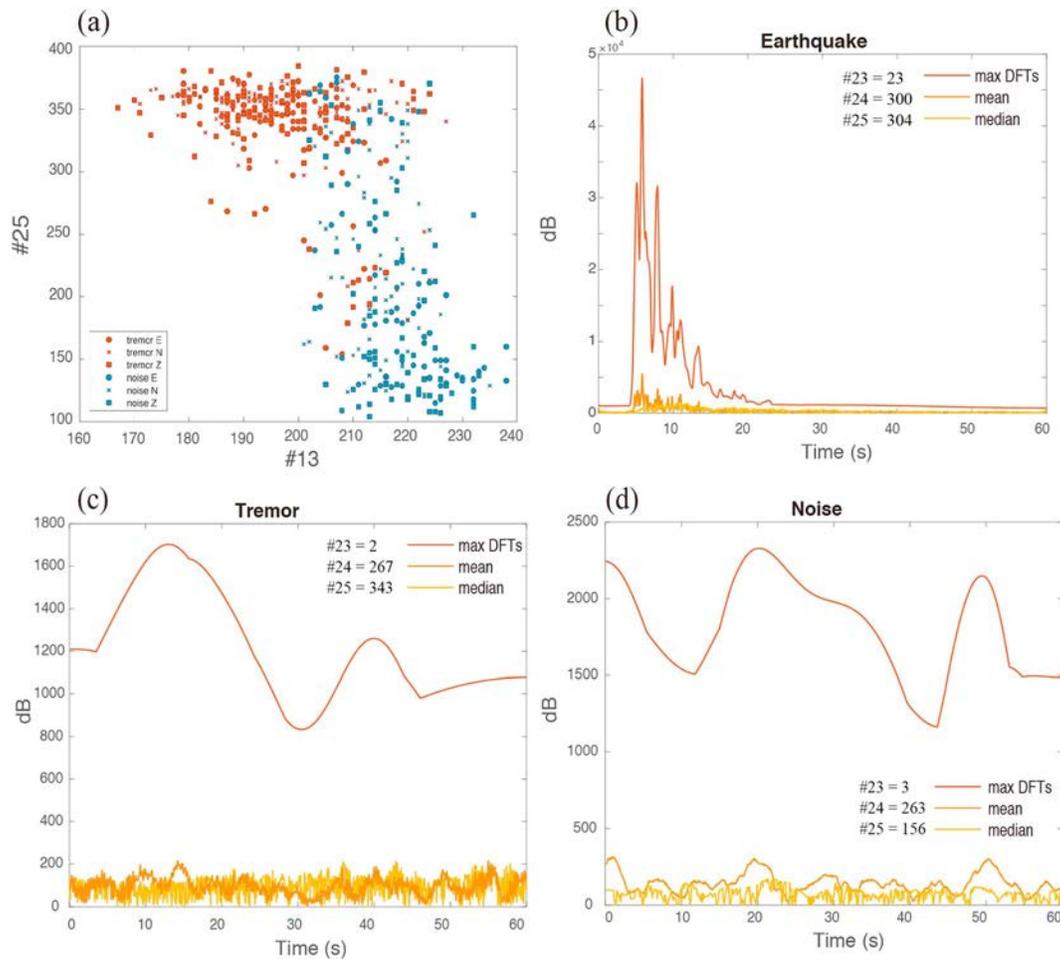


Figure 11. (a) A reduced feature space showing the discrimination between Tremor (red) and Noise (blue) using features #13 and #25. Different symbols indicate the different components. (b–d) Example for the maximum (red line), mean (orange line), and median (yellow line) values of discrete Fourier transforms (DFTs) for three classes of events recorded at station BENA. (b) M2.5 Earthquake occurred on 15 February 2016 at 20:24:32. (c) Tremor event occurred on 15 February 2016 at 00:53:29. (d) Noise event occurred on 1 January 2016 at 10:12:00. Number of peaks for maximum DFTs (#23), mean DFTs (#24), and median DFTs (#25) is indicated by the text next to the label.

We also test the robustness of #25, using the classifier trained on the catalog from different year in 2013 composed of 272 Earthquake, 548 Tremor, and 610 Noise events. From *F* score performance, the feature #25 is found to be characterized by high *F* score (3.8) with top two ranking at the station YULB. This suggests that #25 is a robust feature that is less influenced by time variability of the Noise data.

8.2. Selection of Noise Events in the Training Data

To understand whether the choice of Noise influences the key features for discriminating Earthquake, Tremor, and Noise, we also demonstrate the different spectral behavior between daytime and nighttime Noise. Here the 204 daytime Noise events are randomly selected from 6 a.m. to 6 p.m. during the study period, while the 204 nighttime events are from 7 p.m. to 5 a.m. The daytime Noise is generally noisier than those in nighttime, while their patterns of spectrum remain similar. The variability in spectra in Figure 12 reveals strong variation in daytime data (i.e., *cv* higher than 5.5) other than nighttime data. The *cv* behavior in daytime Noise (Figure 12a) is similar to the original data in Figure 7b—a largest *cv* at low frequency (<0.1 Hz) for station YULB. The nighttime events, however, reveals very similar *cv* patterns between stations. The *cv* behavior likely controls the consistency of feature ranking result. If we replace Noise in the original training data (i.e., mostly daytime events) by the 204 nighttime events, the top-ranking features are now shown in Figure 13 (colored squares with numbers). The efficient features for discriminating Earthquake from Tremor

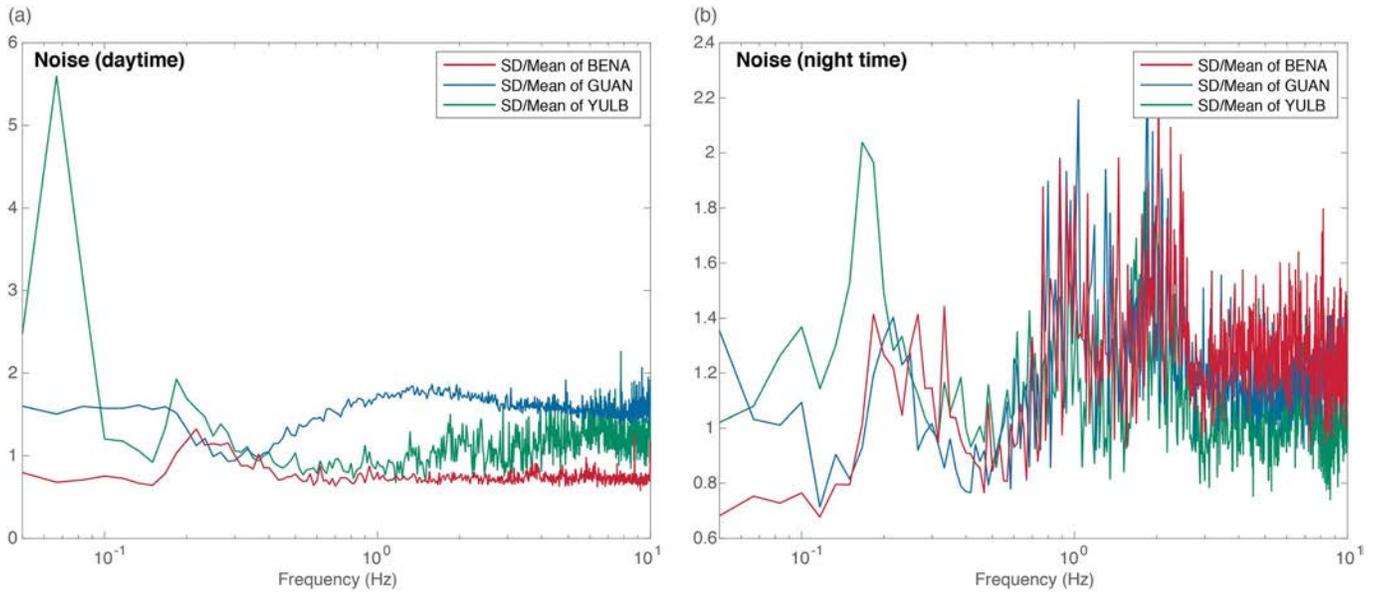


Figure 12. Coefficient of variation in amplitude for each frequency bin of 0.01 Hz for (a) daytime Noise from 6 a.m. to 6 p.m. and (b) nighttime Noise from 7 p.m. to 5 a.m. Here the standard deviation (SD) and mean of amplitude from 204 daytime and 204 nighttime events for each station are computed.

are #1, #2, #8, and #9, the same with those for discriminating Earthquake from Noise except for #8 (top one features at BENA and GUAN but top six at YULB). The efficient features for discriminating Tremor from Noise are #12, #13, #15, and #25. The selected features for three binary models are the same with what obtained from our original data set in Figure 9, suggesting that the choice of Noise may not influence the key features for discriminating Earthquake, Tremor, and Noise.

8.3. Possible Bias in the Classification Model

Our current results were based on a data set limited to the availability of local array on the top of tremor. The size of the training data is confined to 9-month data period of temporary array. We test the time stability of

night noise																											
CI vs C2																											
Station \ Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
BENA	1	2		15	8	7		3	4	5	13						9	6					12	14		11	10
GUAN	2	1	16	15	11	7		3	4	5	13		17				14	6					10	12	18	9	8
YULB	1	2	12	11		10		5	3	9								8					4	13	14	6	7
CI vs C3																											
Station \ Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
BENA	2	4	15	16	9	7		1	3	5	14						12	6					8		13	10	11
GUAN	3	2	18	17	10	9	22	1	4	8	13	19	12		20		16	11			21		5	14	15	7	6
YULB	1	2	11	12	14	13		6	3	10							16	9			15		4		8	5	7
C2 vs C3																											
Station \ Feature	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
BENA										9		4	2		3						7		5		1	6	8
GUAN					8				12	9		3	4		2	11				13			5	10	1	6	7
YULB												1												2	3		

Figure 13. Feature ranking result based on Fisher score for different binary classifications: Earthquakes versus Tremors, Earthquakes versus Noise, and Tremors versus Noise. Here Noise events are selected only from nighttime. For each classification all the 27 features were extracted from the signals. The number in each square indicates the ranking based on Fisher score. The higher the *F* score of a feature is, the lower the number is (and higher ranking). The colored squares indicate the best subset of features determined by Leave-One-Out Cross Validation strategy.

the classifier by using classifier trained on the 2016 catalog to identify events of 2013 catalog. Given that the BENA and GUAN were not available in 2013, for this test data set we use the Z-component seismograms from a single station YULB. This test data set is composed of 272 Earthquake, 548 Tremor, and 610 Noise. The resulting CR is reduced to 52.3%, mainly due to the misclassification of Noise (23% of Noise are incorrectly classified as Earthquake and 43% are incorrectly classified as Tremor). Out of 548 Tremor, 10% and 24% are incorrectly classified as Earthquake and Noise, respectively. The time instability of Noise characteristics results in poor classification performance, as the standard deviation in Noise's spectrum for the 2016 data is significantly smaller than that in 2013 data. If classifier is now trained by the 2013 data without feature selection, the CR for YULB reaches 95.6%, much higher than the 68.8% using the 2016 data (Experiment I). When Fisher's class separability criterion is applied, we found that the F score for Tremor versus Noise increases to 21, leading to very high discriminability between the two classes. The different F score performance from two different time period suggests that the nature of Noise and Tremor have strong variation in time.

The reported results out of the 2016 data have shown that our method is capable of dealing with the tremor classification for three single stations. However, we also found that the variation of Noise is considerably large. The complex spatiotemporal behavior of Noise cannot be represented by a single class; hence, the current data sets (the 2013 and 2016 catalogs) may not represent the real distribution of Noise. Thus, whether our current method can generalize well to large-scale and/or different time period data sets needs further investigation. If attempting to use the trained classifier for detecting tremor in continuous data, there exists some limitation that requires careful treatment for further improvement.

8.4. Further Improvement Using One-Class Classifier

The objective of the current work is not to design a robust classifier that can achieve better tremor detection performance than the state-of-the-art classifiers. Instead, we attempt to investigate the possibilities of using one single station's data to predict tremor and identify the optimal feature subset for the detection. Given that a sophisticated classifier may compensate the weakness of features, we used a very simple classification method, the k -NN, in the present study. However, we found that the multiclass classification approach applied in this study is not robust in practical in terms of real-time tremor monitoring using continuous data. The variety of earthquakes and noise-like signals (e.g., far-field earthquakes, landslides, and typhoons) that occurred frequently but are not considered in the current training data may produce significant classification errors in the current three-class classification framework. The trained decision boundary could be also biased due to the undersampled data in a multiclass classification.

To improve the generalization performance, one possible approach is to have tremors trained in a one-class machine learning model. One-class classification trains a data description to model a target class. Samples that are accepted and rejected by the description are considered as targets and nontargets, respectively. The advantages of one-class classification are as follows: (1) There is no need to collect other classes' samples. (2) It is computationally cheaper because only the target samples (i.e., tremors) are required in the training stage. One-class classification can be performed by either a density or a boundary method. Examples for the former and the later are Gaussian mixture model and support vector domain description, respectively. In the future, the classification performance using one-class classifiers can be further explored, if the aim is at tremor detection.

8.5. Further Improvement Using RFs

In the current study, the tasks of (1) feature selection and (2) three-class classification are performed separately. Fisher's class separability criterion is applied for feature selection, while one-against-one strategy is applied to solve the three-class classification problem. Although the process chain works well, the classification accuracy could be improved by better feature selection with higher robustness and versatility. There is one algorithm which, by construction, performs feature selection and classification in one single step and has demonstrated to be highly relevant and efficient for multiclass problem: RF (e.g., Hibert et al., 2017). The success of RF is attributed to its unique capabilities of inherently performing selection of the best features in multiclass classification problem. More importantly, RF also provides a direct measure of how important the feature is during the classification process. RF, by not removing features, ensures a higher

robustness and versatility, as some features which are relevant for a case will not be the same in a different one (different station and different time period). It is also worthy to explore the usage of this algorithm in future work.

9. Conclusion

In places where tremor signals are relatively weak and short, the powerful tools to segregate them from Noise are demanded. The challenge in identifying Tremor in Taiwan serves as a good example of further improvement using machine learning approach. Recent observation from the seismic stations deployed on the top of the tremor zone allows us to document the tremor events recorded at the closest distance. To identify Tremor, manual checks of multistations waveforms were necessary to exclude loud noise and swarms of small or regional earthquakes. This led to time-consuming and subjective determinations of tremor catalogs in Taiwan. The advances of machine learning approaches enable an automatic search for patterns in large data sets. In this study, we obtained a 9-month tremor catalog using a temporary array on the top of mountain and evaluated the performance of the simple supervised classifier k -NN. The k -NN classifies objects by a majority of votes to a training class that is defined by the distance to the training examples.

Using three-component broadband seismograms from 1 January to 16 September 2016, three classes of events were labeled: $M \geq 2$ Earthquakes, Tremor, and Noise. Based on the computation of 27 seismic features from three different stations and components, we used 612 training events in each class to evaluate the performance of the k -NN classifier. We found that the CR increased from 68.8–93.7% to 86.6–98.8% when the features were selected by F score. Using a new data set composed of 261 events in each class, the classification accuracy remained high from 76.6% to 90%. The efficient features allowing for a better classification between tremor and noise are suggested to be (1) maximum amplitude, number of peaks, and energy of the 2- to 8-Hz-filtered signals and (2) number of peaks in the curve showing the temporal evolution of the DFTs median. Our work validates the performance of a single station k -NN classifier that the classification accuracy can be improved to 98.8% at the station closest to Tremor, with features selected by F score based on the preselected Tremor, Earthquake, and Noise events. However, once we apply the single station approach to the catalog from different time period, the CR dramatically decreases to 53.2%. This suggests that to make this single-station classifier applicable for future tremor monitoring systems, further improvement in increasing the number of classes in the training data and considering time instability of Noise is needed.

Acknowledgments

Support for this work was provided by MOST Grants 103-2116-M-003-001-MY5, 107-2911-I-003-507, and 107-3011-F-027-002. Seismic data are archived at the Central Weather Bureau Seismic Network (<http://gdms.cwb.gov.tw/index.php>) and the Broadband Array in Taiwan for Seismology operated by IES (<http://bats.earth.sinica.edu.tw/Data/index.html>). We greatly appreciate helpful comments from Clément Hibert and the other anonymous reviewer. We also like to thank editor Ben-Zion Yehuda for his effort on improving this manuscript. We would like to thank Uni-edit (www.uni-edit.net) for editing and proofreading this manuscript.

References

- Brudzinski, M. R., & Allen, R. M. (2007). Segmentation in episodic tremor and slip all along Cascadia. *Geology*, 35(10), 907–910. <https://doi.org/10.1130/G23740A.1>
- Cao, X. H., Stojkovic, I., & Obradovic, Z. (2016). A robust data scaling algorithm to improve classification accuracies in biomedical data. *BMC Bioinformatics*, 17(1), 359. <https://doi.org/10.1186/s12859-016-1236-x>
- Chao, K., Peng, Z., Hsu, Y.-J., Obara, K., Wu, C., Ching, K.-E., et al. (2017). Temporal variation of tectonic tremor activity in southern Taiwan around the 2010 M_L 6.4 Jiashian earthquake. *Journal of Geophysical Research: Solid Earth*, 122, 5417–5434. <https://doi.org/10.1002/2016JB013925>
- Chao, K., Peng, Z., Wech, A., Tang, C.-C., Lin, C.-H., & Chen, C.-H. (2011). Deep tremor activities beneath the Central Range in Taiwan and their relationship to local, regional, and teleseismic earthquakes. *Seismological Research Letters*, 82, 326.
- Chen, K. H., Tai, H.-J., Ide, S., Byrne, T. B., & Johnson, C. W. (2018). Tidal modulation and tectonic implications of tremors in Taiwan. *Journal of Geophysical Research: Solid Earth*, 123(7), 5945–5964. <https://doi.org/10.1029/2018JB015663>
- Chuang, L. Y., Chen, K. H., Wech, A., Byrne, T., & Peng, W. (2014). Ambient tremors in a collisional orogenic belt. *Geophysical Research Letters*, 41, 1485–1491. <https://doi.org/10.1002/2014GL059476>
- Curilem, G., Vergara, J., Fuentealba, G., Acuña, G., & Chacón, M. (2009). Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms. *Journal of Volcanology and Geothermal Research*, 180(1), 1–8. <https://doi.org/10.1016/j.jvolgeores.2008.12.002>
- Dragert, H., Wang, K., & Rogers, G. (2004). Geodetic and seismic signatures of episodic tremor and slip in the northern Cascadia subduction zone. *Earth, Planets and Space*, 56(12), 1143–1150. <https://doi.org/10.1186/BF03353333>
- Eddy, S. R. (1996). Hidden Markov models. *Current Opinion in Structural Biology*, 6(3), 361–365. [https://doi.org/10.1016/S0959-440X\(96\)80056-X](https://doi.org/10.1016/S0959-440X(96)80056-X)
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. London, United Kingdom: Chapman & Hall. <https://doi.org/10.1007/978-1-4899-4541-9>
- Fäh, D., & Koch, K. (2002). Discrimination between earthquakes and chemical explosions by multivariate statistical analysis: A case study for Switzerland. *Bulletin of the Seismological Society of America*, 92(5), 1795–1805. <https://doi.org/10.1785/0120010166>
- 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. *Biomedical Signal Processing and Control*, 21, 82–89. <https://doi.org/10.1016/j.bspc.2015.05.011>

- Giudicepietro, F., Esposito, A. M., & Ricciolino, P. (2017). Fast discrimination of local earthquakes using a neural approach. *Seismological Research Letters*, *88*(4), 1089–1096. <https://doi.org/10.1785/0220160222>
- Haykin, S. (1999). *Neural networks: A comprehensive foundation*, (Second ed.). New Jersey, USA: Prentice-Hall.
- 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 Surface Dynamics Discussions*, *5*, 1–15. <https://doi.org/10.5194/esurf-2016-64>
- Huang, C., & Byrne, T. B. (2014). Tectonic evolution of an active tectonostratigraphic boundary in accretionary wedge: An example from the Tulungwan-Chaochou Fault system, southern Taiwan. *Journal of Structural Geology*, *69*(B), 320–333. <https://doi.org/10.1016/j.jsg.2014.06.007>
- Huang, H.-H., Wu, Y.-M., Song, X., Chang, C.-H., Lee, S.-J., Chang, T.-M., & Hsieh, H.-H. (2014). Joint Vp and Vs tomography of Taiwan: Implications for subduction-collision orogeny. *Earth and Planetary Science Letters*, *392*, 177–191. <https://doi.org/10.1016/j.epsl.2014.02.02>
- 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., Yabe, S., Tai, H.-J., & Chen, K. H. (2015). Thrust-type focal mechanisms of tectonic tremors in Taiwan: Evidence of subduction. *Geophysical Research Letters*, *42*, 3248–3256. <https://doi.org/10.1002/2015GL063794>
- Idehara, K., Yabe, S., & Ide, S. (2014). Regional and global variations in the temporal clustering of tectonic tremor activity. *Earth, Planets and Space*, *66*(1). <https://doi.org/10.1186/1880-5981-66-66>
- Ito, Y., Hino, R., Kido, M., Fujimoto, H., Osada, Y., Inazu, D., et al. (2013). Episodic slow slip events in the Japan subduction zone before the 2011 Tohoku-Oki earthquake. *Tectonophysics*, *600*, 14–26. <https://doi.org/10.1016/j.tecto.2012.08.022>
- Kao, H., Thomson, P. J., Rogers, G., Dragert, H., & Spence, G. (2007). Automatic detection and characterization of seismic tremors in northern Cascadia. *Geophysical Research Letters*, *34*, L16313. <https://doi.org/10.1029/2007GL030822>
- Kato, A., & Nakagawa, S. (2014). Multiple slow-slip events during a foreshock sequence of the 2014 Iquique, Chile Mw 8.1 earthquake. *Geophysical Research Letters*, *41*, 5420–5427. <https://doi.org/10.1002/2014GL061138>
- Kato, A., Obara, K., Igarashi, T., Tsuruoka, H., Nakagawa, S., & Hirata, N. (2012). Propagation of slow slip leading up to the 2011 Mw 9.0 Tohoku-Oki Earthquake. *Science*, *335*(6069), 705–708. <https://doi.org/10.1126/science.1215141>
- Kortström, J., Uski, M., & Tiira, T. (2016). Automatic classification of seismic events within a regional seismograph network. *Computers and Geosciences*, *87*, 22–30. <https://doi.org/10.1016/j.cageo.2015.11.006>
- Laasri, E. H. A., Akhouayri, E.-S., Agliz, D., Zonta, D., & Atmani, A. (2015). A fuzzy expert system for automatic seismic signal classification. *Expert Systems with Applications*, *42*(3), 1013–1027. <https://doi.org/10.1016/j.eswa.2014.08.023>
- Liao, S. C., Wu, C.-T., Huang, H.-C., Cheng, W.-T., & Liu, Y.-H. (2017). Major depression detection from EEG Signals using kernel eigenfilter-bank common spatial patterns. *Sensors*, *17*(6), 1385. <https://doi.org/10.3390/s17061385>
- Lin, Y.-P., Wang, C.-H., Jung, T.-P., Wu, T.-L., Jeng, S.-K., Duann, J.-R., & Chen, J.-H. (2010). EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, *57*(7), 1798–1806. <https://doi.org/10.1109/TBME.2010.2048568>
- Liu, Y.-H., & Chen, Y.-T. (2007). Face recognition using total margin-based adaptive fuzzy support vector machines. *IEEE Transactions on Neural Networks*, *18*(1), 178–192. <https://doi.org/10.1109/TNN.2006.883013>
- Liu, Y.-H., 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>
- Masotti, M., Campanini, R., Mazzacurati, L., Falsaperla, S., Langer, H., & Spampinato, S. (2008). TREMOREC: A software utility for automatic classification of volcanic tremor. *Geochemistry, Geophysics, Geosystems*, *9*, Q04007. <https://doi.org/10.1029/2007GC001860>
- Masotti, M., Falsaperla, S., Langer, H., Spampinato, S., & Campanini, R. (2006). Application of support vector machine to the classification of volcanic tremor at Etna, Italy. *Geophysical Research Letters*, *33*, L20304. <https://doi.org/10.1029/2006GL027441>
- Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.
- Miyazawa, M., & Mori, J. (2005). Detection of triggered deep low-frequency events from the 2003 Tokachi-oki earthquake. *Geophysical Research Letters*, *32*, L10307. <https://doi.org/10.1029/2005GL022539>
- Mousavi, S. M., Horton, S. P., Langston, S. A., & Samei, B. (2016). Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression. *Geophysical Journal International*, *207*(1), 29–46. <https://doi.org/10.1093/gji/ggw258>
- Nadeau, R. M., & Dolenc, D. (2005). Nonvolcanic tremors deep beneath the San Andreas Fault. *Science*, *307*(5708), 389–389. <https://doi.org/10.1126/science.1107142>
- Nakano, M., Sugiyama, D., Hori, T., Kuwatani, T., & Tsuboi, S. (2019). Discrimination of seismic signals from earthquakes and tectonic tremor by applying a convolutional neural network to running spectral images. *Seismological Research Letters*, *90*(2A), 530–538. <https://doi.org/10.1785/0220180279>
- Obara, K. (2002). Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science*, *296*(5573), 1679–1681. <https://doi.org/10.1126/science.1070378>
- Obara, K., & Kato, A. (2016). Connecting slow earthquakes to huge earthquakes. *Science*, *353*(6296), 253–257. <https://doi.org/10.1126/science.aaf1512>
- Peng, Z., & Chao, K. (2008). Non-volcanic tremor beneath the Central Range in Taiwan triggered by the 2001 Mw 7.8 Kunlun earthquake. *Geophysical Journal International*, *175*(2), 825–829. <https://doi.org/10.1111/j.1365-246X.2008.03886.x>
- Provost, F., Hibert, C., & Malet, J.-P. (2016). Automatic classification of endogenous landslide seismicity using the Random Forest supervised classifier. *Geophysical Research Letters*, *44*, 113–120. <https://doi.org/10.1002/2016GL070709>
- Reynen, A., & Audet, P. (2017). Supervised machine learning on a network scale: Application to seismic event classification and detection. *Geophysical Journal International*, *210*(3), 1394–1409. <https://doi.org/10.1093/gji/ggx238>
- Rubinstein, J. L., Gombert, J., Vidale, J. E., Wech, A. G., Kao, H., Creager, K. C., & Rogers, G. (2009). Seismic wave triggering of nonvolcanic tremor, episodic tremor and slip, and earthquakes on Vancouver Island. *Journal of Geophysical Research*, *114*, B00A01. <https://doi.org/10.1029/2008JB005875>
- Sit, S., Brudzinski, M., & Kao, H. (2012). Detecting tectonic tremor through frequency scanning at a single station: Application to the Cascadia margin. *Earth and Planetary Science Letters*, *353–354*, 134–144. <https://doi.org/10.1016/j.epsl.2012.08.002>
- Socquet, A., Valdes, J. P., Jara, J., Cotton, F., Walpersdorf, A., Cotte, N., et al. (2017). An 8 month slow slip event triggers progressive nucleation of the 2014 Chile megathrust. *Geophysical Research Letters*, *44*, 4046–4053. <https://doi.org/10.1002/2017GL073023>
- Sun, W.-F., Peng, Z., Lin, C.-H., & Chao, K. (2015). Detecting deep tectonic tremor in Taiwan with a dense array. *Bulletin of the Seismological Society of America*, *105*(3), 1349–1358. <https://doi.org/10.1785/0120140258>

- Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random Forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, *43*(6), 1947–1958. <https://doi.org/10.1021/ci034160g>
- Uchida, N., Iinuma, T., Nadeau, R. M., Bürgmann, R., & Hino, R. (2016). Periodic slow slip triggers megathrust zone earthquakes in northeastern Japan. *Science*, *351*(6272), 488–492. <https://doi.org/10.1126/science.aad3108>
- Vallejos, J. A., & McKinnon, S. D. (2013). Logistic regression and neural network classification of seismic records. *International Journal of Rock Mechanics and Mining Sciences*, *62*, 86–95. <https://doi.org/10.1016/j.ijrmms.2013.04.005>
- Vapnik, V. N. (1998). *Statistical learning theory*. New York: John Wiley & Sons Inc.
- Wech, A. G. (2010). Interactive tremor monitoring. *Seismological Research Letters*, *81*(4), 664–669. <https://doi.org/10.1785/gssrl.81.4.664>
- Wech, A. G., & Creager, K. C. (2008). Automated detection and location of Cascadia tremor. *Geophysical Research Letters*, *35*, L20302. <https://doi.org/10.1029/2008GL035458>
- Wu, C.-T., Dillon, D. G., Hsu, H.-C., Huang, S., Barrick, E., & Liu, Y.-H. (2018). Depression detection using relative EEG power induced by emotionally positive images and a conformal kernel support vector machine. *Applied Sciences*, *8*(8), 1244. <https://doi.org/10.3390/app8081244>