

Semiautomated tremor detection using a combined cross-correlation and neural network approach

T. Horstmann,¹ R. M. Harrington,² and E. S. Cochran³

Received 30 April 2013; revised 6 August 2013; accepted 13 August 2013.

[1] Despite observations of tectonic tremor in many locations around the globe, the emergent phase arrivals, low-amplitude waveforms, and variable event durations make automatic detection a nontrivial task. In this study, we employ a new method to identify tremor in large data sets using a semiautomated technique. The method first reduces the data volume with an envelope cross-correlation technique, followed by a Self-Organizing Map (SOM) algorithm to identify and classify event types. The method detects tremor in an automated fashion after calibrating for a specific data set, hence we refer to it as being “semiautomated”. We apply the semiautomated detection algorithm to a newly acquired data set of waveforms from a temporary deployment of 13 seismometers near Cholame, California, from May 2010 to July 2011. We manually identify tremor events in a 3 week long test data set and compare to the SOM output and find a detection accuracy of 79.5%. Detection accuracy improves with increasing signal-to-noise ratios and number of available stations. We find detection completeness of 96% for tremor events with signal-to-noise ratios above 3 and optimal results when data from at least 10 stations are available. We compare the SOM algorithm to the envelope correlation method of Wech and Creager and find the SOM performs significantly better, at least for the data set examined here. Using the SOM algorithm, we detect 2606 tremor events with a cumulative signal duration of nearly 55 h during the 13 month deployment. Overall, the SOM algorithm is shown to be a flexible new method that utilizes characteristics of the waveforms to identify tremor from noise or other seismic signals.

Citation: Horstmann, T., R. M. Harrington, and E. S. Cochran (2013), Semiautomated tremor detection using a combined cross-correlation and neural network approach, *J. Geophys. Res. Solid Earth*, 118, doi:10.1002/jgrb.50345.

1. Introduction

[2] Since *Obara* [2002] first observed tremor along the southwest Japan subduction zone, tectonic tremor has been detected in many other subduction zones, such as Cascadia [*Rogers and Dragert*, 2003; *McCausland et al.*, 2005], Costa Rica [*Brown et al.*, 2009; *Walter et al.*, 2011], Alaska [*Peterson and Christensen*, 2009], and Mexico [*Payero et al.*, 2008]. Many studies have also shown that tectonic tremor occurs outside of subduction zones; for example, *Nadeau and Dolenc* [2005] detected tectonic tremor along the San Andreas fault, a transform plate boundary, and *Peng and Chao* [2008] observed tectonic tremor in Central Range in Taiwan, an arc-continental type collision environment. Tectonic tremor that occurs spontaneously is denoted as

“ambient” tremor, in contrast to “triggered” tremor, which is excited by surface waves of large distant earthquakes or other static or dynamic stress changes [*Ghosh et al.*, 2009; *Gomberg et al.*, 2008; *Peng et al.*, 2008; *Peng et al.*, 2009; *Wang et al.*, 2013]. The distinctive waveform characteristics of tectonic tremor include emergent phase arrivals and a low-amplitude signal, with the main portion of the energy often concentrated in the 2–8 Hz band [*Obara*, 2002; *Schwartz and Rokosky*, 2007; *Beroza and Ide*, 2011]. The tremor signal duration ranges from short bursts of a few seconds up to several tens of minutes or hours [*Ryberg et al.*, 2010]. In the following, we refer to tectonic tremor simply as tremor.

[3] While some progress has been made on understanding the mechanics of tremor, detection of tremor is not trivial; the emergent phase arrivals and low amplitudes make automated detection a difficult task. In attempts to overcome difficulties associated with detection, a variety of detection methods have been developed, many of which are based on one of two techniques: (1) a cross correlation of envelope waveforms [*Obara*, 2002] or (2) searching for low-frequency earthquakes (LFEs) [*Katsumata and Kamaya*, 2003]. Additional techniques which do not rely on cross correlation of envelopes of templates include those used by *Brudzinski and Allen* [2007], *Ghosh et al.* [2009], and *Walter et al.* [2011].

¹Geophysical Institute, Karlsruhe Institute of Technology, Karlsruhe, Germany.

²Earth and Planetary Sciences, McGill University, Montreal, Canada.

³U.S. Geological Survey, Pasadena, California, USA.

Corresponding author: T. Horstmann, Geophysical Institute, Karlsruhe Institute of Technology, Hertzstr. 16, DE-76185 Karlsruhe, Germany. (Tobias.Horstmann@kit.edu)

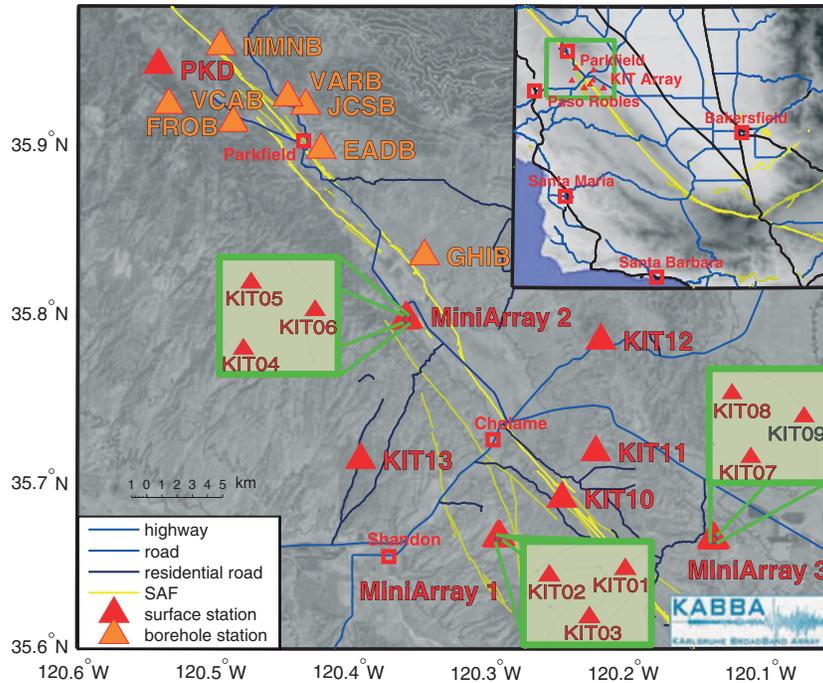


Figure 1. Station distribution used in our analysis, including the temporary array, the permanent HRSN stations, and station PKD. Three sites contain a mini-array of three stations each, spaced approximately 150 m apart.

[4] Two of the most common methods include a method by *Wech and Creager* [2008] analyzing network coherence with envelope cross correlation and a method by *Shelly et al.* [2007] which uses a template-matching, cross-correlation technique to identify LFEs. *Wech and Creager* [2008] introduced a method analyzing network coherence through epicentral reliability and spatial repeatability. They first calculate envelope functions for 5 min time windows that are low-pass filtered between 1 and 8 Hz. Next, they use a grid search over all possible source locations based on S wave lag times calculated from the correlation values of the envelope functions. Source locations with estimated epicentral error estimates exceeding 5 km are rejected. Tremor activity is then detected when at least two locations occur within a $0.1 \times 0.1^\circ$ area per day. The advantage of the method is its ability to detect and locate tremor at the same time and that it provides robust results even when an individual station has a poor signal-to-noise ratio. A limitation of the method is that it is restricted to detection of extended tremor episodes.

[5] *Shelly et al.* [2007] introduced a different detection method using LFEs in combination with a template-matching, cross-correlation technique. The set of LFEs from a number of different event families function as master templates in the detection algorithm which then identifies tremor events in continuous waveform data. The advantages of the method are that it is very precise in event time determination and it can be applied using individual stations. The shortcoming of the method is the required a priori knowledge of the master event templates which restricts detection to LFEs which correlate highly with the defined templates. *Brown et al.* [2008] modified the LFE template technique to overcome the disadvantage of requiring a priori knowledge of the master templates by eliminating the requirement that

master templates be predefined. They remove the need for using predefined master templates by applying a running autocorrelation technique to identify the event families and average the repeating events to create master templates. However, the method is computationally expensive and thus difficult to apply to large data sets.

[6] Here we introduce a new method using a neural network algorithm based on frequency content and motion products of tremor waveforms. The fundamental advance of the method is that it does not rely on master templates and is not based on any assumptions about a minimum signal length.

[7] Many tremor detection methods employ a minimum time window length, limiting event detection to those tremor episodes with duration exceeding the minimum window length [e.g., *Wech and Creager*, 2008]. Removing such restrictions on duration permits the detection of a wider range of event types than present methods, thereby increasing the potential for discovering tremor at different depths within the fault zone. Detecting tremor at shallow depths, particularly in subduction zones, could have potential implications for seismic and tsunami hazard; therefore, methods capable of detecting a wider variety of events have the potential to advance our understanding the role of tremor in fault slip. By detecting a larger variety of events, we increase the observations available for analysis and interpretation and advance our understanding of tremor sources. Here we describe a new method for tremor detection, while the detailed analysis of the tremor will be the subject of forthcoming work. The method detects tremor in an automated fashion, however, calibration of the method is required for use with a specific data set, so we refer to the method as “semiautomated.”

Table 1. Complete List of Stations, Their Positions, and Data Availability

Station	Latitude (°N)	Longitude (°W)	Elevation	Start Date	End Date
KIT A1	35.665945	-120.293227	510 m	24 May 2010	02 May 2011
KIT A2	35.795039	-120.360526	375 m	24 May 2010	31 Jun 2011
KIT A3	35.664648	-120.137173	387 m	24 May 2010	31 Jun 2011
KIT10	35.688881	-120.247177	599 m	24 May 2010	05 Jan 2011
KIT11	35.716370	-120.222643	619 m	24 May 2010	11 May 2011
KIT12	35.782349	-120.218689	581 m	25 May 2010	31 Jun 2011
KIT13	35.711880	-120.393730	448 m	24 May 2010	31 Jun 2011
GHIB	35.832249	-120.347282	330 m	24 May 2010	31 Jun 2011
EADB	35.895222	-120.422623	224 m	24 May 2010	31 Jun 2011
FROB	35.910950	-120.486877	231 m	24 May 2010	31 Jun 2011
MMNB	35.956501	-120.496002	480 m	29 May 2010	31 Jun 2011
JCSB	35.921169	-120.433998	299 m	24 May 2010	31 Jun 2011
VCAB	35.921619	-120.533920	555 m	24 May 2010	31 Jun 2011
VARB	35.926079	-120.447052	177 m	29 May 2010	31 Jun 2011
PKD	35.945171	-120.541603	583 m	24 May 2010	31 Jun 2011

[8] The paper is structured as follows: We first present the data set and methods description in sections 2 and 3, followed by a method performance evaluation in section 4 using a 3 week test data set. Finally, we apply the method to over 1 year of continuously recorded waveform data to identify all tremor events within that time span.

2. Data Set

[9] The data set consists of continuous broadband recordings of 13 STS-2 seismometers from the Karlsruhe Broadband Array (KABBA) at a sampling rate of 200 sps. The surface stations were deployed along the Cholame segment of the San Andreas fault at seven sites within a 20 km by 25 km area centered on the town of Cholame (Figure 1). The highest amplitude tremor identified in previous studies occurs directly beneath Cholame [Nadeau and Guilhem, 2009; Shelly and Hardebeck, 2010; Ryberg et al., 2010]. In contrast to the area around Parkfield, the area near Cholame lacks dense station coverage. The array was designed to supplement the seismic network south of the High Resolution Seismic Network (HRSN) to better record the vigorous tremor occurring here. The station installation at three of the seven sites consist of mini-arrays of three stations each spaced approximately 150 m apart. The stations recorded continuously from May 2010 to July 2011. In addition to the array broadband surface stations, we use seven borehole stations from the HRSN and one additional broadband surface station from the Berkeley Digital Seismic Network (BDSN). The HRSN stations GHIB, EADB, FROB, VCAB, VARB, MMNB, and JCSB are located north of the array near Parkfield and record at a sampling rate of 250 sps. Station PKD of the BDSN network is also located north of the array and records continuous data at a sampling rate of 100 sps. Additional data exist from the Northern California Seismic Network surface stations, however, we do not use them, as selected stations provide us with the widest range of azimuthal coverage using the least number of stations possible in order to optimize the computational efficiency. A complete list of stations is provided in Table 1.

3. Methods

[10] In this section, we outline the individual steps of the semiautomated detection method. We first detail the data

reduction step which reduces the volume of continuously recorded data (section 3.1). Second, we describe the data classification steps required to prepare the data input for the neural network, or Self-Organized Map (SOM) clustering algorithm (section 3.2). Next, we present the SOM clustering, including the determination of the signal classes (section 3.3). Finally, we describe the postprocessing steps that reduce the number of earthquakes and false picks in the signal classes, which may occur if the adjustable sensitivity of the algorithm is set too high (section 3.4).

3.1. Data Reduction

[11] Our data set contains roughly 14 months of continuous recordings on a maximum of 21 stations, including both the temporary array and permanent stations. The goal of the data reduction step is to keep time periods with potential seismic events while reducing the data volume as much as possible. In the following, the term “seismic event” refers to both tremor and earthquakes, which are assumed to be of unknown type until they are classified by the method. The data reduction step uses a cross-correlation technique similar to that applied by previous studies; however, here the cross correlation is not designed to detect or classify seismic events exclusively. Instead, portions of a continuous time series are classified using the SOM clustering.

[12] We start by searching for time windows containing coherent signals across the station array. We first filter the waveforms between 2 and 8 Hz, as the signal to noise ratio for tectonic tremor for our data set is highest in the 2–8 Hz band [Obara, 2002; Schwartz and Rokosky, 2007; Beroza and Ide, 2011]. Second, the envelope of each trace is calculated and decimated to 0.2 sps, following which, the envelopes of individual components are stacked for each station. Third, we perform a cross correlation of waveform envelopes between each station and a designated master station. The cross-correlation step is repeated using each station once as a master station. We do not remove the instrument response, as it is flat within the frequency band of interest and removing it would increase the computational time. Moreover, all cross correlations are normalized and based on smoothed envelopes, which removes the influence of different amplitudes and reducing the effect of phase shifts. We tested different time window lengths and the time step to determine the most effective time window length. A trial-and-error comparison of handpicked time windows and time

windows picked by the automated method indicates that a 520 s window and 5 s time step optimally removes noise while retaining the highest number of the seismic events. The interstation distance divided by a speed of 3 km/s is the maximum permitted time lag between peak correlation values, reducing the influence of infrasound events and noise while still detecting S waves from shallow sources close to the stations. After averaging the cross-correlation values for each time window and each master station over all station pairs, we then select the maximum mean cross-correlation value, producing a mean coefficient function over time. We retain event windows for which the coefficient function exceeds the mean correlation value by a threshold of 0.15 for at least 30 s. The threshold was determined by testing values on a 1 week test data set from 24 May to 14 June 2010. We determine by visual inspection which values retain the majority of events, while best reducing the data volume.

[13] We merge windows separated by less than 300 s into a single event window. Doing so may preserve extended tremor episodes which may have been otherwise fragmented into separate time windows. Visual examination of the tremor episodes recorded in our data set suggests that tremor episodes do not typically last longer than approximately 100 s; hence, we use a generous value of 300 s to bridge time windows. One could use a longer interwindow time, however, doing so increases the cost of computational time during the SOM processing.

3.2. Data Preparation

[14] The data preparation steps outlined here prepare the data for input into the Self-Organized Map (SOM) algorithm [Kohonen, 2001], which is available as a Matlab toolbox [Vesanto, 2000]. The data preparation is not essential to the clustering algorithm but is included to enhance the detection algorithm's performance, including adjusting the sensitivity of the algorithm. The goal of the SOM is to group time series data with distinguishing features. The SOM algorithm may use any similar feature, such as frequency content in a particular band or polarity, to classify the data. The larger the feature differences between various types of signals, the more effectively they distinguish those signals. Below, we describe the preconditioning steps and the feature selection and calculation. The main data preconditioning steps entail reducing the effects of noise and calculating and normalizing the feature values used by the SOM algorithm. We first describe the noise reduction in section 3.2.1, followed by trace alignment in section 3.2.2. We then explain the feature calculation in section 3.2.3 and, finally, we describe the feature vector normalization in section 3.2.4.

3.2.1. Noise Reduction

[15] To remove the influence of noise and to enhance signal amplitudes for the feature calculation, we implement a noise reduction technique introduced by Martin [2001], which is based on minimum statistics and employs spectral subtraction methods. The noise reduction step improves the detection result, thereby increasing the sensitivity of the detection algorithm. The technique assumes that the power spectral density of a given signal quickly decays to the background noise level over time. It tracks the minimum spectral amplitude value in narrow frequency bands, using the tracked values as an estimated minimum noise level. A mean noise level estimation is then calculated by multiply-

ing the minimum noise estimation with a bias compensating factor, which is based on the variance of the spectral amplitude in that given frequency band [Martin, 1994]. Assuming that noise and signal are statistically independent and additive, one may remove noise by subtracting the mean noise estimation from the original spectra. The phase spectra are not modified, thereby permitting a transformation back to the time domain, with the noise removed.

[16] The algorithm applies a short time Fourier transform with a moving time window length of 0.6 s and 0.3 s overlap in which the power spectra are recursively smoothed. For example, let $M_{k,l}$ denote the power spectrum for a given frequency bin l within some time window k . The smoothed power spectrum is given by $P_{k,l} = \alpha \cdot P_{k-1,l} + (1 - \alpha) \cdot M_{k,l}$, where α represents an updating factor that controls the influence of previous time windows. The updating factor α is typically set between 0.9 and 0.95 [Martin, 1994]. Here we use a value of 0.9.

[17] A requirement of the noise reduction method is that the time window used to track the minimum noise level must be longer than the expected observed signal [Martin, 1994]. Tremor episodes consist of successive energy bursts which may last up to several minutes. The time window for the noise reduction must be larger than the longest observed tremor signal in order to bridge to the next minimum containing only noise. However, the shorter the time window, the more accurate the estimation of the noise level. We reviewed longer tremor episodes in our test data set and tested various window lengths and found a maximum of 420 s an optimum duration. The disadvantage of using such a long time window becomes apparent when considering cases where the noise level is monotonically increasing. For example, the noise level is calculated for a given sample from the time window spanning 420 s prior to the sample. If the noise level constantly increases, the current estimation of the minimum noise level in a given window occurs at the beginning of the window. However, the noise estimation is subtracted from the sample at the end of the time window. Thus, the noise level is underestimated. Consequently, the noise level estimation will be underestimated at a time lag approximately equal to the time window length. However, even in the extreme case of monotonically increasing noise, the noise reduction will reduce the noise amplitude leading to a better signal-to-noise ratio and therefore an enhanced tremor signal.

[18] Figure 2 illustrates the functionality of the technique. Figure 2b shows the original time series recorded at station KIT10 with a 2–8 Hz band-pass filter applied exhibiting multiple tremor bursts. The same tremor bursts highlighted in Figure 2b are enhanced in Figure 2c after applying the noise reduction. The enhanced signal will be more easily recognized by the SOM algorithm. Note that the PSD shown in Figure 2 reflects a narrow frequency band centered at 7 Hz. Therefore, some tremor bursts which are not obvious in the PSD can be seen in the waveforms in Figures 2b and 2c.

3.2.2. Trace Alignment

[19] The goal of the SOM algorithm is to cluster specific data features independent of the signal duration. Specifically, we want to be able to detect short tremor bursts, as well as longer episodes. Therefore, in order to cluster features of the same signal recorded at individual stations, all traces must be corrected for moveout. We calculate an individual

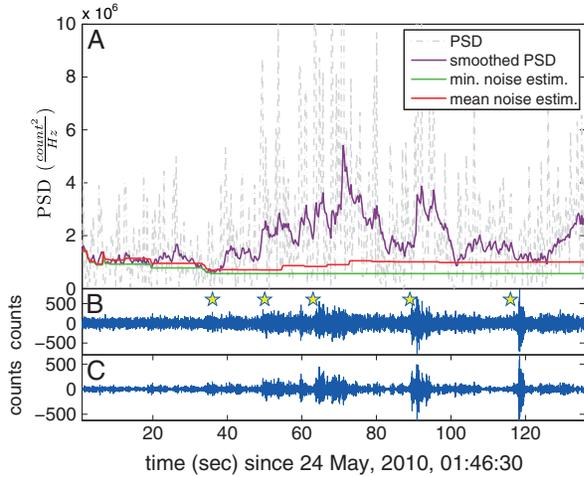


Figure 2. (a) Power spectral density (PSD) for one narrow frequency bandpass centered around 7 Hz of the waveform data shown in Figure 2b (dashed line). Smoothed PSD (purple), tracked minimum value (green), and a bias factor computed mean noise estimation (red). (b) Z component of station KIT 10 filtered between 2 and 8 Hz during a tremor episode including individual tremor bursts (yellow stars). (c) Noise-reduced trace after subtracting the mean noise estimation. The trace in Figure 2c has lower noise amplitudes than the trace in Figure 2b, enabling the SOM algorithm to perform better.

moveout correction for each of the time windows retained by the data reduction step discussed in section 3.1. We then align traces according to the moveout before inputting them into the SOM. We determine the moveout between station pairs by cross-correlating envelopes that have been smoothed over 15 samples with a master station.

[20] In order to allow greater flexibility and account for the current noise conditions at the individual stations, any of the stations may serve as the master station. We determine the moveout used to align the traces from the offset associated with the maximum of the cross-correlation function. Preliminary inspection of the data showed a number of infrasound events (see example in Appendix C and Figure C2, as well as a discussion of the feature calculation of infrasound events). Setting a maximum permissible time difference allows us to remove infrasound events with much lower propagation velocities, while retaining most other seismic events. For each time window, we choose the master station for which the envelope correlation coefficients are on average higher than the other stations.

3.2.3. Feature Calculation

[21] Once we prepare the data for input into the SOM by reducing the effects of noise and aligning the traces, we calculate the features on which the event detection is based. Any characteristic of a time series which varies with signal source is well suited for use in SOM clustering. Such characteristics may include polarity, frequency-wave number, and complex seismic trace properties. We term such particular characteristics as “features” after Köhler *et al.* [2010] and calculate multiple feature values continuously throughout the time series. We combine all feature values into a fea-

ture vector for each window where the feature values are calculated. For an extended discussion on features, including various types and their interdependence, we refer the reader to Köhler *et al.* [2010].

[22] Determining what features work best for a given data set requires testing the features for uniqueness using a sample data set. We perform a preliminary test of the SOM algorithm on a 3 week long sample data set that contains both tremor and earthquakes. The sample data set was taken from the beginning of our measurement period, spanning the period from 24 May to 14 June 2010. We use waveforms from our temporary deployment, as well as from permanent stations installed locally near the Parkfield segment of the San Andreas fault (Table 1). The sample data set contains noise, tremor events, and local and regional earthquakes and is therefore suited for testing the method. In the subsequent text, we refer to the 3 week long test data set discussed here as simply the test data set. Example tremor events are shown in Figures 7 and 8, and an example earthquake recording is shown in Appendix C in Figure C1. Systematic tests using various features of tremor episodes indicate that two features are capable and sufficient to distinguish tremor signals from regional earthquakes and noise: (1) spectral amplitudes in five narrow frequency bands calculated via a Stockwell transform [Stockwell *et al.*, 1996] and (2) combined horizontal to vertical component motion products [Jepsen and Kennett, 1990].

[23] The five frequency bands used for the first feature include the main energy band of tremor, namely, the 2–8 Hz band, as well as the 15–30 Hz and 0.5–1.5 Hz bands. The 2–8 Hz band is subdivided into three equally sized bands. The 15–30 Hz band is useful for distinguishing tremor from noise, and the 0.5–1.5 Hz band is useful for discriminating tremor from regional and teleseismic earthquakes. The high-resolution spectral amplitudes are calculated using a Stockwell transform and then averaged over the given frequency bins [Stockwell *et al.*, 1996]. The Stockwell transform produces the time-frequency distribution of a signal using a moving Gaussian window that is scalable for different frequencies. Compared to a short time Fourier transform, the Stockwell transform retains better time resolution, similar to wavelet transformations.

[24] The second feature we calculate consists of combined motion products. Combined motion product values differ widely for different seismic phases; thus, they are ideal for use within the SOM algorithm. White [1964] defined two motion product detectors, HV and HiV having a 90° phase shift:

$$P_N = u_N \cdot u_D \quad P_E = u_E \cdot u_D \quad (\text{HV}) \quad (1)$$

$$Q_N = H(u_N) \cdot u_D \quad Q_E = H(u_E) \cdot u_D \quad (\text{HiV}) \quad (2)$$

where u_N , u_E , and u_D are the displacements in the north, east, and vertical directions. $H(u_N)$ denotes the Hilbert transform of u_N . Jepsen and Kennett [1990] used both HV and HiV detectors to define combined motion products:

$$P_{NE} = \sqrt{P_N^2 + P_E^2} \quad \text{and} \quad Q_{NE} = \sqrt{Q_N^2 + Q_E^2}. \quad (3)$$

They define a wave classification parameter PQ_{abs} using the combined motion products, where $PQ_{\text{abs}} = P_{NE} \times \text{abs}(Q_{NE})$.

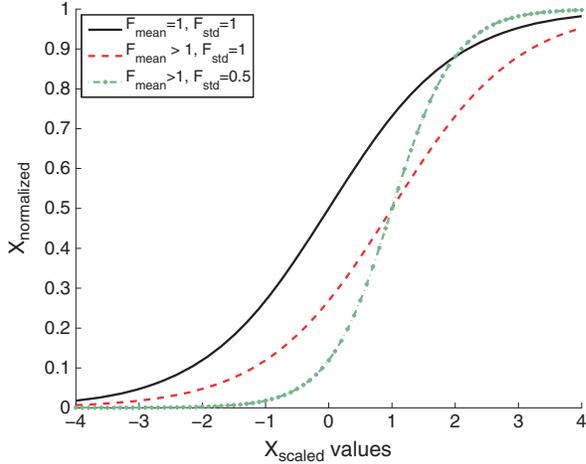


Figure 3. Graphic description of extended softmax normalization. Transformation from X_{scale} to X_{norm} : (a) original softmax normalization (solid line); (b) $F_{\text{mean}} \times \text{mean}(X_{\text{org}}) = 1$ and $F_{\text{std}} = 1$ (dashed line); (c) $F_{\text{mean}} \times \text{mean}(X_{\text{org}}) = 1$ and $F_{\text{std}} = 0.5$ (dash-dotted line).

For a rectilinear polarized wavefield, PQ_{abs} is used to discriminate between vertically and horizontally polarized S waves, with expected PQ_{abs} values of > 0.25 and < 0.25 , respectively. Waves with two-dimensional motion, such as Love or Rayleigh waves, are expected to have PQ_{abs} values greater than 0.25. The PQ_{abs} parameter exhibits large increases during earthquake and tremor episodes, while maintaining low values for noise. PQ_{abs} is therefore a useful parameter for distinguishing seismic events from noise. We also use it to distinguish the clusters identified by the SOM algorithm containing seismic signals, which we discuss in detail in section 3.3.1.

[25] As discussed above, we use five frequency bands for the spectral density calculation and one combined motion product, for a total of six values per time interval, per station. We first decimate the time series to 50 sps after applying a low-pass filter with a 20 Hz corner frequency to avoid aliasing effects. We then calculate the six feature values at the sample rate of the decimated time series, with the exception of the upper frequency. The upper frequency band feature is calculated from the time series decimated to 100 sps, which then has a low-pass filter applied with a corner at 40 Hz. The decimated and filtered time series of the upper frequency band is also resampled to 50 sps. We then create feature vectors for an individual station by taking the average value of each feature over 0.5 s time intervals, in order to optimize the balance between computational cost and accuracy. The feature vectors from each individual station are combined into a single feature vector for the entire array. For example, a feature vector would consist of 60 feature values per time interval for a 10 station array.

3.2.4. Feature Vector Normalization

[26] The input for the SOM algorithm is the feature vectors. The SOM clustering algorithm is based on Euclidean distances of the feature values comprising the feature vectors in a high-dimensional parameter space [Vesanto and Alhoniemi, 2000]. In order to weight all features equally, it is critical to normalize all feature values to the same range, i.e.,

between 0 and 1. We use an extended softmax normalization to normalize feature values.

[27] The softmax normalization is defined as

$$X_{\text{norm}} = \frac{1}{1 - e^{-X_{\text{scaled}}}}, \quad \text{with} \quad X_{\text{scaled}} = \frac{X_{\text{org}} - \text{mean}(X_{\text{org}})}{\text{std}(X_{\text{org}})}, \quad (4)$$

where X_{org} are the original feature values [Pyle, 1999, pp. 271–274, 355–359]. The main benefits of the softmax normalization over a standard normalization are the ability to reduce the influence of extreme values without neglecting them and regulate the detection sensitivity of the method. By inserting two scaling factors within the softmax normalization, we can define the extended softmax normalization as follows:

$$X_{\text{norm}} = \frac{1}{1 - e^{-X_{\text{scaled}}}} \quad \text{with} \quad X_{\text{scaled}} = \frac{X_{\text{org}} - F_{\text{mean}} \cdot \text{mean}(X_{\text{org}})}{F_{\text{std}} \cdot \text{std}(X_{\text{org}})}, \quad (5)$$

where F_{mean} and F_{std} are the scaling factors that determine the range of the scaled feature values. Adjusting the F_{mean} and F_{std} parameters are what allows one to adjust the sensitivity of the detection algorithm for a given a data set. The scaling factors F_{mean} and F_{std} influence the transformation behavior of the data set to a normalized data space and can be used to carve out small but important differences which would be lost using the standard softmax normalization. At the same time, bigger yet less important differences in the data values can be downweighted.

[28] Our data set contains over 1 year of data, making calculating the mean and standard deviation of X_{org} for all feature vectors for the entire time period computationally expensive. Nevertheless, applying the same normalization to the entire data set is critical in order to cluster events that may occur widely spaced in time during the study period. As one does not know the range of feature values a priori, the F_{mean} and F_{std} scaling factors must be determined empirically using a sample data set. We therefore calibrate our algorithm using the same 3 week representative test data set used for the feature selection in section 3.2.3. We calculate the mean ($\text{mean}(X_{\text{org}})$) and standard deviation ($\text{std}(X_{\text{org}})$) of the feature values for the 3 week time period and use those values to normalize the entire data set. We manually pick tremor events during the 3 week period and determine the best F_{mean} value based on a comparison of manually picked events and those picked by the SOM. We start by normalizing our data set with a range of F_{mean} and F_{std} values and run the SOM for each of the F -value pairs. We evaluate which of the F -value pairs correspond to the highest number of accurately picked events compared to the handpicked events. The F -values deemed best from the comparison are then used for the entire data set, assuming similar characteristics amongst earthquakes and tremor over the entire time period.

[29] Figure 3 illustrates the normalization. If the factors F_{mean} and F_{std} are set to 1 for the regular softmax normalization (black line), the normalization scales nearly linearly for values close to the mean value and tapers off asymptotically to 0 and 1 for values far from the mean. The portion of the curve that scales linearly is controlled by the standard deviation scaling factor, F_{std} . When $F_{\text{std}} \neq 1$, as in the case of the extended softmax transformation, then the curvature of the nonlinear region scales asymptotically to both 0 and 1. The factor F_{mean} is used to shift the linearly scaled region

Table 2. F_{mean} and F_{std} Values Determined by a Comparison of Manually and SOM Detected Events for the Test Data Set^a

Feature	F_{mean}	F_{std}
2–4 Hz frequency band	0.5	0.5
4–6 Hz frequency band	0.5	0.5
6–8 Hz frequency band	0.5	0.5
15–30 Hz frequency band	8.0	1.5
0.5–1.5 Hz frequency band	2.5	1
PQ_{abs}	1.8	0.6

^aThe values shown are the normalization values used for the remaining data set in the extended softmax normalization (equation (5)).

of the normalized data to the position in the feature data set where it best discriminates between noise and seismic signals. Increasing the F_{mean} value has the effect of offsetting the scaled data to a value above 0; in the case of the red and green curves shown in Figure 3, the inflection point of the curves is shifted to a value of 1. The most important benefit of using the extended softmax normalization is the capability to regulate the sensitivity of the detection method via the scaling factors F_{mean} and F_{std} . Higher or lower values for F_{mean} and F_{std} translate to higher or lower detection sensitivity. One must keep in mind, however, that with increased sensitivity comes an increased number of false detections. Given their importance, Appendix A contains an extended discussion on the determination of F -values. Table 2 indicates the optimal values for our data set to achieve the desired level of sensitivity.

3.3. SOM Clustering Algorithm

[30] Following feature normalization, the feature vectors are ready for input into the SOM clustering program [Vesanto, 2000]. The SOM is an unsupervised learning method that clusters data into groups with similar feature values. Following feature input, the algorithm first determines the number of so-called prototype vectors in an N -dimensional parameter space, where N is the number of feature values for each time window (e.g., the number of components comprising the feature vectors for an individual station). The prototype vectors form a two-dimensional grid with a hexagonal structure, where each prototype vector has the same dimension as each of the feature vectors and is connected to the six nearest neighbor vectors. After the grid is randomly initialized in the parameter space, the algorithm updates the prototype vectors and moves them toward the feature vectors during an iterative training period. At the end of the training period, the grid is spread to the data cloud. The data cloud is composed of the feature vectors, and it approximates the probability density function of the input data. This process is also known as vector quantization [Köhler et al., 2010]. Once the grid position in the parameter space has been calculated, each data point is linked to the nearest prototype vector. We illustrate an example of a data set with three features in Figure 4. Following the grid formation, a hierarchical clustering algorithm clusters the SOM prototype vectors [Vesanto and Alhoniemi, 2000]. The algorithm iteratively determines appropriate clusters based on the average distances between prototype vectors within existing clusters; however, the user may set an acceptable minimum and maximum number of clusters. The optimal

number of clusters is then chosen by evaluating the Davies-Bouldin index (DB index) [Davies and Bouldin, 1979]. The DB index is a metric for evaluating the appropriate number of clusters based on the distance between cluster members and between cluster centers. It has a minimum value when the clusters are most compact and widely spaced. All of the above mentioned aspects of the clustering algorithm, including the DB index, are contained within the SOM toolbox as described by Köhler et al. [2009, 2010]. Due to the computational limitation of the working memory, we split the data set into day-long sections, permitting the SOM to determine the optimum number of clusters for each 24 h period of data.

[31] Theoretically, the DB index may indicate any number of clusters for each day’s worth of data. In practice, we restrict the range for which the DB index is evaluated in order to avoid using an inappropriate number of groups. Based on the number of different signals we expect to be present in the continuous data (noise, tremor, earthquakes, infrasound), we empirically determine the appropriate range of numbers of clusters necessary. We thereby attempt to prevent the algorithm from mixing different signals in one cluster that we want to distinguish. We use the test data set to evaluate the appropriate maximum and minimum number of clusters. Figure 5 illustrates the influence of the number of clusters on the detection accuracy, as well as the number of accurate detections found by the SOM in the 3 week data set. The detection accuracy varies only marginally ($\pm 2\%$) when the minimum allowable number of clusters is greater than four, and the maximum allowable number of clusters is greater than eight, suggesting that the number of clusters has little influence on the detection accuracy (Figure 5). The method detects the largest number of events when the minimum number of clusters is eight or nine. The detection accuracy increases by approximately 1% if the minimum number is set to 7 or 10. However, the number of detected signals decreases significantly when the minimum number of clusters is less than eight. One might argue based on Figure 5 that the best minimum value ranges from 19 to 20.

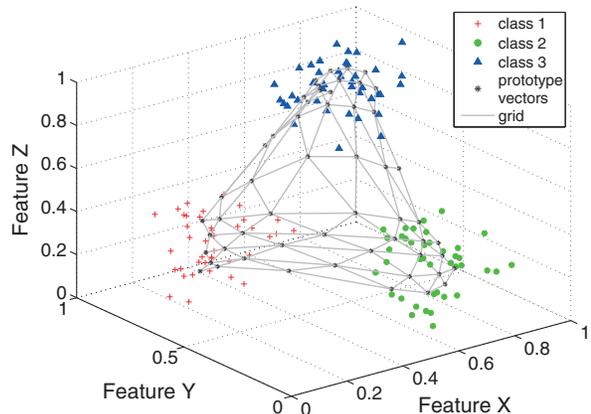


Figure 4. SOM after training: the SOM can be seen as a regular two-dimensional grid of prototype vectors (black stars connected by a mesh) initialized in a 3-D data space in this example. Red, blue, and green points represents the data set. Each component of an individual data vector (shown here as X , Y , and Z) represents a feature. The color code of the data points indicates the clustering result.

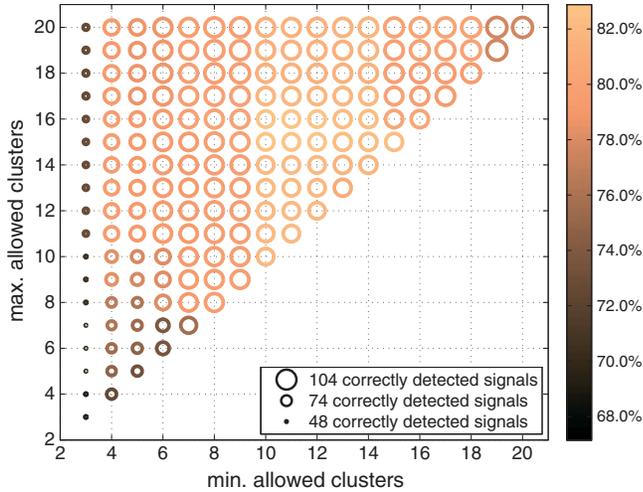


Figure 5. Minimum and maximum number of permitted clusters: circle size corresponds to number of correctly detected events, and color code indicates the detection accuracy.

However, analyses of various normalization factor values show a general trend of decreased detection accuracy when using such a high minimum number of clusters. We therefore use eight as the minimum number of clusters because the SOM detects the largest number of events with a minimum loss in detection and accuracy for the remaining data set. A similar argument holds for the maximum number of clusters. We choose a maximum value of 20 to allow greater flexibility in the number of clusters. Therefore, the DB index falls within the range of 8 and 20 clusters.

3.3.1. Signal Class Determination

[32] The SOM groups the data points into clusters, but these clusters do not have an a priori association with any particular signal type (e.g., tremor, earthquake, noise). Therefore, the user is free to determine the signal cluster of interest. Here the algorithm identifies the content of the clusters, and we manually classify the groups of clusters into “classes.” We designate three main classes into which the various clusters determined by the SOM are sorted. Two of the classes contain seismic signals, and the third class contains noise. We label these classes as S1, S2, and N (seismic signal classes 1 and 2, and noise). Clusters containing seismic signals are identified by isolating classes with high PQ_{abs} values. The PQ_{abs} feature is able to discriminate between noise and seismic signals for all feature values, making it a robust tool for identifying seismic events. Examination of the test data set indicates that normalized $PQ_{\text{abs}} \gg 0.5$ for seismic signals and $PQ_{\text{abs}} < 0.5$ for noise signals. We therefore designate a cluster as belonging to one of the two seismic classes if the following two criteria are met: (1) mean $PQ_{\text{abs}} \geq 0.5$ at a minimum of three stations and (2) average $PQ_{\text{abs}} \geq 0.5$ on all borehole stations. We impose the additional requirement for borehole stations in order to avoid classifying infrasound events as tremor.

[33] We use the frequency features to further group the seismic signals into two classes (S1 and S2). Class S1 contains tremor and small local earthquakes, and class S2 contains regional and larger local earthquakes. Similar

to tremor signals, earthquake signals also have values of $PQ_{\text{abs}} > 0.5$. Although the PQ_{abs} feature is similar between the two types of signals, the spectral characteristics of teleseismic and regional earthquakes differ from tremor in the low-frequency band (0.5–1.5 Hz). We identify clusters containing regional and larger local earthquakes using the mean spectral amplitude values in the lower frequency band; events with (normalized) mean spectral amplitude exceeding 0.6 at a minimum of three stations comprise the class S2. The remaining clusters of seismic events comprise class S1 (tremor and local earthquake) class. Unfortunately, the spectral characteristics of small local earthquakes are similar to those of tremor, and there is no definitive cutoff value which can discriminate between the two. Additionally, visual inspection shows similar spectral energy of tremor and local earthquakes in the 10 to 15 Hz band. We initially tried using a feature in the 15–50 Hz band in an attempt to differentiate local earthquakes from tremor. The spectral amplitudes in the 15–50 Hz band also did not differ significantly enough to cluster local earthquakes into separate groups. We therefore use an additional postprocessing step to distinguish the small local earthquakes from tremor event in the S1 class (section 3.4.1). We describe an estimation of what fraction of the signal grouped into the S1 class is composed of local earthquake signal in sections 4.2 and 4.2.5.

[34] Finally, we discard all detected time windows less than 4 s in length and merge time windows separated by less than 30 s in the S1 class. We compile the beginning and end of the resulting time windows in our tremor catalog. Table 3 summarizes the criteria used to sort the SOM-detected time windows into the three user-determined classes.

3.4. Postprocessing

[35] Following the SOM clustering and sorting into the S1, S2, and N classes, the S1 class still contains a number of false detections (noise and small local earthquakes). The goal of the postprocessing step is to identify false detections and nontremor seismic signals and move them into the N and S2 classes respectively. The postprocessing step consists of two parts: (1) The first part sorts the earthquakes remaining in the S1 class into the S2 class, and (2) the second part sorts noise (false detections) into the N class. We label the two parts earthquake postprocessing and noise postprocessing, respectively. The noise postprocessing step is optional and may be implemented depending on the sensitivity set by the user (adjustable via the F_{mean} and F_{std} parameters discussed in section 3.2.4).

3.4.1. Earthquake Postprocessing

[36] Because the amplitude/frequency characteristics of small local earthquakes are not dissimilar enough to that of tremor to be discriminating, some local earthquakes are grouped into class S1. The earthquake postprocessing step

Table 3. Criteria Used for Grouping SOM Clusters Into the Respective Classes

Class	Class Description	0.5–1.5 Frequency	
		Mean PQ_{abs}	Band
S1	Tremor, small local earthquakes	> 0.5	< 0.6
S2	Earthquakes (large local, regional)	> 0.5	> 0.6
N	Noise	< 0.5	-

Table 4. Summary of Parameter Settings of the Method Discussed in Section 3

Parameter	Setting
<i>Data Reduction (Section 3.1)</i>	
Filter bandpass	2–8 Hz
Data decimation	0.2 sps
Cross-correlation time window length	520 s
Cross-correlation time window step	5 s
Cross-correlation max. lag	interstation distance/3 km/s
Correlation coefficient threshold	0.15
Signal merging period	300 s
<i>Noise Reduction (Section 3.2.1)</i>	
Fourier transform time window length	0.6 s
Fourier transform time window step	0.3 s
Smoothing factor α	0.9
Minimum tracking time window length	420 s
<i>Trace Alignment (Section 3.2.2)</i>	
Envelope smoothing	15 samples
Cross-correlation max. lag	interstation distance/3 km/s
<i>Feature Calculation (Section 3.2.3)</i>	
Data decimation	50 sps
Time window length for feature averaging	0.5 s
<i>SOM Clustering Algorithm (Section 3.3)</i>	
Minimum number of clusters	8
Maximum number of clusters	20
<i>Signal Class Determination (Section 3.3.1)</i>	
PQ_{abs} threshold	0.5
Amplitude threshold for the 0.5–1.5 Hz band	0.6
Minimum station number	3
<i>EQ Postprocessing (Section 3.4.1)</i>	
STA-window length	0.5 s
LTA-window length	30 s
C2	6
C5	5.5
Minimum station number with detections	3
Time window length for connecting detections at different stations	6 s
Time window extension	30 s
PGV threshold	1400 nm/s
<i>Noise Postprocessing (Section 3.4.2)</i>	
Cross-correlation value averaging	3 best stations
Mean cross-correlation value threshold	0.8

identifies earthquakes in class S1 and moves them into class S2. We first apply a STA/LTA (short-term average / long-term average) trigger to S1 class time windows that are less than 30 s in length to identify impulsive earthquake arrivals. We use an STA/LTA trigger based on the algorithm described by *Allen* [1982] with a STA-window length of 0.5 seconds and a LTA-window length of 30 s. We set the weighting factor C2 from *Allen* [1982] between the two terms of the characteristic function to 6 and the threshold constant C5 to 5.5. The two terms of the characteristic function are sensitive to changes in amplitude and frequency, respectively, while the C5 constant sets the threshold for event declaration. We apply the algorithm to the *Z* component and declare an earthquake when the STA/LTA threshold is reached on at least three stations within a 6 s time window.

[37] The second part of the earthquake postprocessing compares the remaining time windows in the S1 class to the Advance National Seismic System (ANSS) catalog [*Advanced National Seismic System*, 2012]. The detection algorithm presented here does not determine origin times for the detected events; therefore, we compare the results by calculating traveltimes from the earthquake epicenter to the center of the KIT array. We estimate the time window in which a cataloged event should occur by calculating upper bound and lower bound traveltimes using a range of seismic

velocities between 4.4 and 6.6 km/s. We account for both *P* and *S* wave velocities for cases where only the *S* wave is detected. Additionally, we extend the time window of the expected earthquake by 30 s for two purposes: to compensate for uncertainties in the travel path and to account for the range of arrival times for epicenters which are in close proximity to the array. The range of 4.4 to 6.6 km/s was chosen quasi-empirically (based on realistic values) to maximize earthquake detection by increasing the tolerance for error in the traveltimes estimation without making the time windows so large that we might have erroneously included coincidental tremor events. We count an earthquake as detected if it occurs within the calculated time window. We limit our earthquake catalog to the area bounded within the latitude and longitude ranges of 32°N and 38°N, and 123.3°W and 113.3°W. Additionally, we consider only earthquakes with a calculated peak ground velocity (PGV) > 1400 nm/s at the station array. We calculate the PGV following *van der Elst and Brodsky* [2010], using $\log_{10} \text{PGV} = -2.29 + 0.85M - 1.29 \log_{10} r$ where *M* represents the magnitude and *r* represents the hypocentral distance in km.

[38] We assume the ANSS catalog is complete for larger earthquakes. We also apply the STA/LTA detection algorithm to time windows of less than 30 s in length to identify small events missing from the catalog. If no catalog information is available, one could apply the STA/LTA trigger to

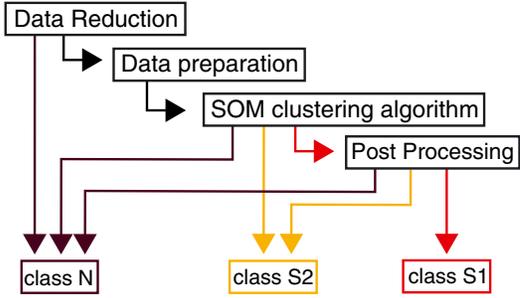


Figure 6. Flow chart of the method: neglected data within the reduction step in section 3.1 are sorted into the noise class N, and potential signals are kept and prepared in section 3.2. Data preparation entails noise reduction, trace alignment, and feature calculation and normalization. Prepared data then gets classified into the three classes within the SOM clustering algorithm in section 3.3. Afterward, the optional postprocessing step in section 3.4 sorts misclassified tremor to the earthquake and noise classes using an STA/LTA trigger and envelope cross-correlation.

all time windows. The STA/LTA detection alone decreases the correctly classified tremor by approximately 10%, while also increasing the number of earthquakes in the S1 class by approximately 10%. We provide a quantitative assessment of the detection algorithm performance using the earthquake postprocessing step in section 4.2.5.

3.4.2. Noise Postprocessing

[39] The noise postprocessing step moves noise events falsely classified into classes S1 or S2 into class N. The false classification of noise events results from the sensitivity of the algorithm, which can be adjusted via the F_{mean} and F_{std} parameters discussed in section 3.2.4. There is an

inherent trade-off between accuracy and sensitivity in the detection algorithm; however, higher sensitivity enables the detection of more lower amplitude tremor events. For cases where a high sensitivity is desired, the optional noise postprocessing step detects and removes the majority of false picks, increasing the detection accuracy. The basis for the noise postprocessing step is the assumption that events have a similar shape at different stations, while uncorrelated noise bursts do not. Thus, the noise postprocessing consists of the following steps: (1) calculating a smoothed envelope for all traces in the time windows selected by the SOM, (2) cross-correlating the envelopes and calculating a mean coefficient of the three best cross-correlation coefficients, and (3) discarding time windows with an average coefficient below 0.8. We implement additional steps to determine the time window for which the envelope correlation is calculated to account for possible misalignment of traces. Given that the noise postprocessing step is optional, the details are included in Appendix B.

[40] All parameter values described in section 3 are summarized in Table 4, and Figure 6 provides a flow chart of the method.

4. Results

[41] In this section, we evaluate the performance of the tremor detection algorithm using test data set referenced in section 3 to determine the best parameter settings. We first discuss two examples of detected waveforms in order to illustrate the capabilities and limitations of the method (section 4.1) and present tremor and earthquake detection statistics, including correct, false, and missed detections (section 4.2). Additionally, to evaluate the method’s effectiveness, we compare the performance of our method

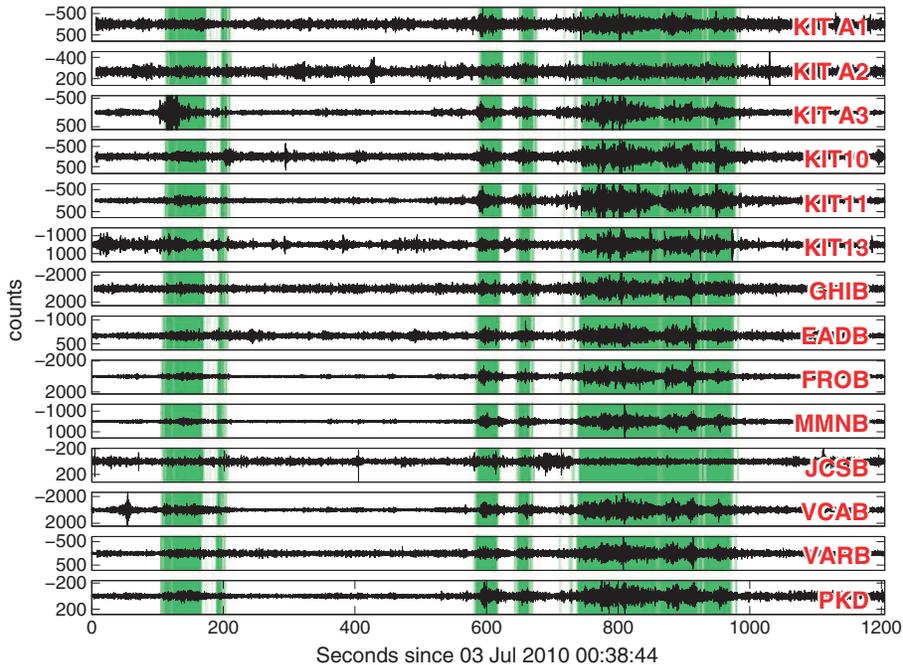


Figure 7. Example of a tremor episode exhibiting a good signal-to-noise ratio. Detected tremor is shown in green, and noise is shown in white.

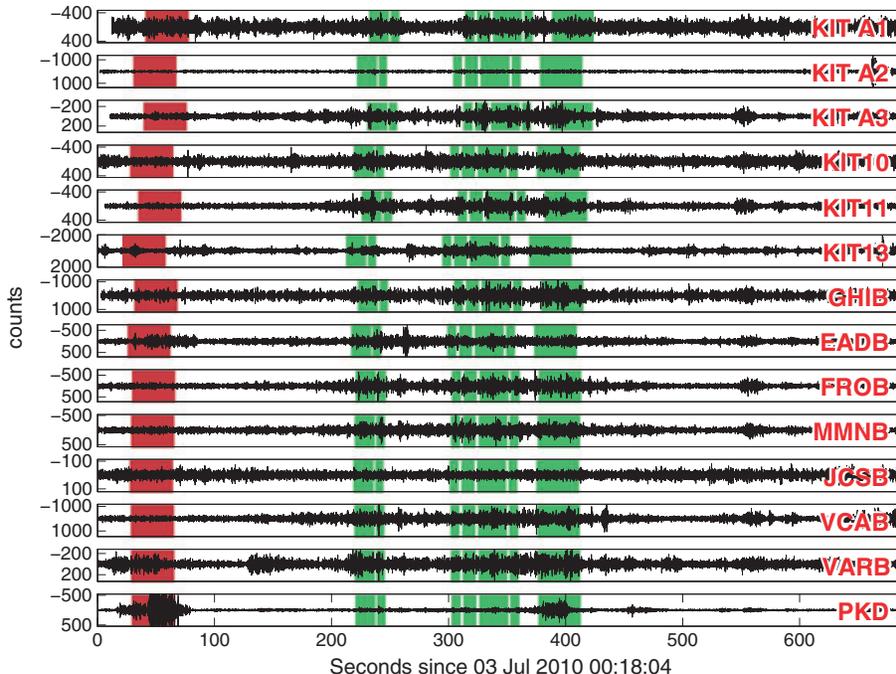


Figure 8. Example tremor waveforms with low signal-to-noise ratio. The figure illustrates the performance of the SOM clustering algorithm at the SNR detection threshold. Due to local noise conditions, tremor is often only visible on a subset of stations. Data is filtered between 2 and 8 Hz. Color code indicates the clustering result following the postprocessing step: detected tremor (green), noise (white), and tremor detected by the SOM and rejected by the noise postprocessing step (red). The color-coded clustering result is shifted in time at different stations illustrating the trace alignment determined by the algorithm. The unusual moveout for KIT A1, A2, and A3 shows the failing of the trace alignment step, although the method still detects the tremor. A probable event at 570 s is not detected, likely due to erroneous moveout correction within the trace alignment step.

an implementation of *Wech and Creager* [2008] method (section 4.2.2).

4.1. Tremor Waveforms

[42] Although tremor waveforms share characteristics, such as a long-duration (seconds to tens of minutes) and a lack of distinct phase arrivals, individual tremor episodes vary greatly in their appearance (e.g., duration, number of high amplitude pulses or tremor bursts, maximum amplitude, etc.). In this section, we show two examples of detected tremor waveforms to illustrate the capabilities and limitations of the method.

[43] Figure 7 shows an example waveform from 3 July 2010, filtered between 2 and 8 Hz. The tremor episode between 750 and 980 s has a median signal-to-noise ratio (SNR) of ~ 3 based on root-mean-square (RMS) value in contrast to the much lower amplitude tremor around 180 s (SNR of ~ 1.5). The SOM method does not provide estimates of source location; however, the detection method of *Wech and Creager* [2008] provides an estimate of tremor locations. For the episode shown in Figure 7, the average epicentral location is $35.64^\circ\text{N} \pm 0.30^\circ\text{N}$, $120.59^\circ\text{W} \pm 1.07^\circ\text{W}$, with a depth of 19.16 ± 13.54 km. The epicentral location is outside the boundary of Figure 1 and is therefore not shown.

[44] Given our emphasis on greater detection sensitivity (i.e., a higher number of detections) in our current algorithm settings, we also obtain a number of false detections.

The tremor signal shown in Figure 8 has a low SNR of ~ 1.8 and is difficult to recognize even by visual inspection. Figure 8 illustrates the performance of the method near the detection threshold. Both Figures 7 and 8 show waveforms where the detected tremor bursts are not obvious on all stations, often due to local noise; we show all traces to illustrate the performance of the method even if noisy data are present. The postprocessing step (correctly) rejects the first time window originally identified as a tremor event, while retaining the remaining tremor windows. Furthermore, the algorithm misses a probable tremor event around 570 s, likely due to misaligned traces. Figures 7 and 8 show time-shifted traces to illustrate the trace alignment determined by the algorithm. Misaligned traces could occur at a station with a low SNR or where multiple events occur in short succession. Either case would lead to different moveouts within the same time window.

4.2. Performance Evaluation

4.2.1. Comparison to Manually Detected Events

[45] Using the feature and normalization parameters outlined in the methods section, we compare the detected event time windows to manually pick events in a 3 week test data set from the beginning of the deployment, 24 May to 14 June. The manual picking, thus the manual detection and classification of earthquakes and tremor, was performed by careful inspection of the raw and filtered waveforms. A correct detection is defined when a time window picked

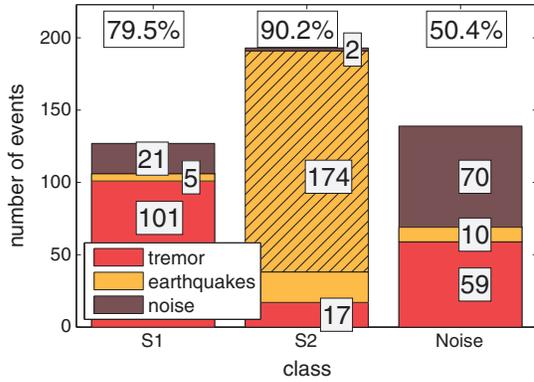


Figure 9. Detection performance for the test data set. Tremor events shown in red, earthquakes in yellow, and noise in brown. Cluster composition of S1 class, S2 class, and rejected events within the noise class. Class S1 contains 101 tremor events, 5 earthquakes, and 21 noise windows resulting in a detection accuracy of 79.5%. Class S2 contains 17 tremor events, 174 earthquakes, and 2 noise events. Class N contains 59 tremor events, 10 earthquakes, and 70 noise time windows. Of the 174 earthquakes (class S2), 48 have been identified and moved from class S1 using the ANSS catalog and 105 with the STA/LTA algorithm (hatched area).

by the algorithm overlaps with a manually picked event. In some cases, the semiautomated method defines an event time window that is visibly shorter than the duration of a tremor episode. Detected time windows separated by more than 30 s are considered as individual events. While events separated by less than 30 s are grouped into a single event.

[46] Figure 9 shows the composition of classes S1, S2, and N. The events shown in class N consist only of those events which were moved from class S1 following the noise postprocessing step; the events classified originally as noise by the SOM are not included. In total, the SOM and postprocessing steps correctly detect 101 tremor events and obtains 26 false detections (5 earthquakes and 21 noise events) in class S1, equivalent to a detection accuracy of 79.5%. Class S2 consists of 174 earthquakes, 17 tremor, and 2 noise events, corresponding to a 90.2% detection accuracy. During the earthquake postprocessing, the STA/LTA algorithm moves 105 earthquakes and 8 tremor events from class S1 into class S2. The comparison with the ANSS catalog shifts 48 earthquakes from class S1 to class S2, as well as 9 tremor events.

[47] Figure 10 plots the detected event date versus event length for the 3 week long test data set. Roughly two thirds of the detected events in classes S1 and S2 are shorter than 1 min. The longest tremor signal has a duration of 13 min, but we note that the method often divides tremor episodes into separate events. Seventy time windows rejected by the postprocessing step are removed correctly while 69 manually picked events (i.e., correctly detected events) are removed. Of the 139 rejected time windows, 101 windows are shorter than 30 s (Figure 10b). The noise postprocessing step rejects events mostly due to alignment failure (section 3.2.2), high noise conditions, and/or detection of a small fraction of a signal.

4.2.2. Comparison to Wech and Creager [2008] Method

[48] We compare the SOM method results for the 3 week long test data set with another commonly implemented detection algorithm developed by *Wech and Creager* [2008], referred to here as WECC. We implement the WECC method as it has been successfully applied for tremor detection in Cascadia and is most comparable to our method as it detects more extended coherent tremor episodes rather than single LFEs. The WECC method calculates the cross correlation of the envelopes of all station pairs within a 5 min time window. Next, the method performs a grid search over all potential source locations, searching for the *S* wave lag times in order to optimize the cross correlation. There are a number of adjustable parameters in the WECC method, namely, cross-correlation window length, cross-correlation value threshold, and minimum number of stations for location. We tested different parameter settings finding that a time-window length of 300 s with a 150 s time step worked best for our data set. We only use observations with a maximum cross-correlation coefficient exceeding 0.5 and require a detection on a minimum of five stations.

[49] One shortcoming of the WECC method is that it detects earthquakes and other coherent signals without any means of distinguishing such signals from tremor. Therefore, it is necessary to implement a postprocessing step as well. Similar strategies to those used with the SOM may be adopted for distinguishing earthquakes and tremor identified by the WECC algorithm: one option is to use a STA/LTA trigger to exploit the distinct phase arrivals of an earthquake. A second strategy is to use a remote station to distinguish earthquakes from lower amplitude tremor. A third possibility is the use of an earthquake catalog. Here we use the same strategy as the postprocessing step described in section 3.4.1 namely, and STA/LTA trigger paired with the comparison of

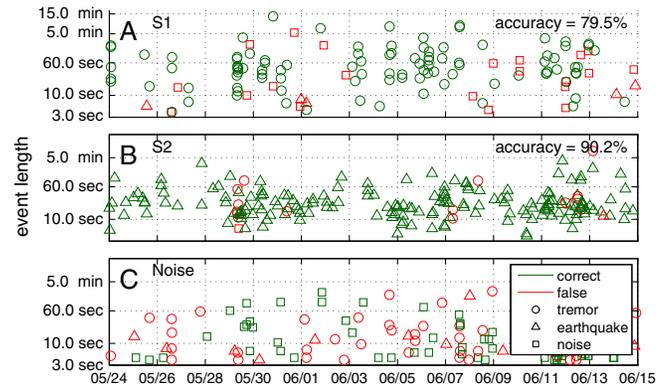


Figure 10. Method performance on the test data set. (a) Detected event time windows of class S1 plotted according to date and detected event length. Marker type denotes event type: tremor (circles), earthquakes (triangles), and noise (squares). Of a total of 127 automatically detected events, 101 are correctly detected tremor (green), and 26 are falsely detected (red) based on a comparison with manually picked events. Detection accuracy is 79.5%, and the sum of detected time windows is 141 min. (b) Events in class S2 and (c) events moved from class S1 by the noise postprocessing step.

Table 5. Results of the WECC Method Applied to the Test Data Set^a

Method	Number of Detections	Manual Classification			Detection Accuracy
		Tremor	EQ	Noise	
WECC	245	54	160	31	22%
WECC with STA/LTA	81	34	38	9	42%
SOM S1 class	127	101	5	21	79.5%

^aThe results of the SOM method from section 4.2 are displayed in the last line for comparison.

an earthquake catalog. However, we now set the C5 constant from *Allen* [1982] to a value of 1.5.

[50] We applied the WECC method to the test data set with and without the optional STA/LTA trigger. The method without the STA/LTA detects 245 events and only 54 are tremor according to our analyst picks. Of the 245 events, 160 are local and regional earthquakes, and 31 events are noise (see Table 5). The STA/LTA reduces the detected events to 81, 34 of which are tremor events, 38 earthquakes, and 9 noise events, leaving a detection accuracy of 44%. In comparison, the SOM method detects 101 tremor events in the same time period. Changing the parameter settings for the WECC method to adjust for higher detection sensitivity leads not only to an increase in detected tremor events (approximately 80) but also to an exponential increase in false detections (over 500).

[51] A direct comparison of the algorithms is difficult given that the SOM may detect multiple tremor events (i.e., multiple time windows) within the 5 min time window used by the WECC method. Therefore, we compare each detected tremor event by the WECC method directly with the detection result of the SOM method. Of the 34 tremor events detected by WECC, 25 were detected by the SOM method and correctly classified in class S1, 3 tremor events were misclassified as earthquakes by the SOM, 4 were thrown out by the SOM postevaluation step, and 2 were not detected (see Table 6). However, the SOM identifies an additional 76 tremor events that were not identified by WECC. Thus, 101 tremor events were correctly identified by SOM compared to just 34 identified by WECC.

[52] A reason for the discrepancy may be the limitation that the WECC method requires spatial and temporal clustering. The advantage of the WECC method is its ability to locate and detect tremor at the same time and its computational efficiency. However, the SOM method detects more tremor events with a higher accuracy since it uses characteristics of the waveforms to classify different event types.

4.2.3. Influence of Noise

[53] Furthermore, we determine the influence of signal to noise ratio on the detection accuracy. Using the test data set with manually detected events, we define a detection

completeness by dividing the number of automatically and correctly detected events by the total number of manually detected events, regardless of the difference in time window length. We calculate the detection completeness for bins of events with a similar SNR. The SNR is calculated by dividing the RMS amplitude of an event time window by the RMS amplitude of a noise window. The SNR is computed at each station on the vertical (*Z*) component. The individual noise window is manually selected. As the detection requires a minimum of three stations, we assign the third highest SNR of all stations to an individual event.

[54] Figure 11a shows the detection completeness versus the SNR. We differentiate between results with and without the noise postprocessing step. Note that the detection accuracy is lower if the noise postprocessing step is not implemented. However, evaluating the detection completeness of the method with and without the noise postprocessing step is instructive, as it allows us to understand how the noise postprocessing step influences detection accuracy.

[55] Figure 11c indicates that for tremor, the method provides a detection completeness of 96% for events with a SNR above 3. The detection completeness is approximately 80% for signals with a SNR value of 2 or higher. However, it is important to note that the majority of tremors have a signal to noise ratio less than 2. We also find that the noise postprocessing step may reduce the detection completeness in some cases. For example, events missed by the detection algorithm with a SNR ~ 3 are explained by failure of the noise postprocessing step. The noise level similarly influences the detection completeness for earthquakes. The influence of noise is minimal with a SNR > 3 , although a small number of undetected earthquakes with SNR > 3 result from poor trace alignment. The majority of earthquakes have a SNR above 4.5 and are concentrated in the last bin.

4.2.4. Influence of the Number of Stations

[56] We also determine the influence of the number of available stations used. In order to assess how the number of stations affects the detection accuracy as well as the total number and duration of detections, we employ a jack-knife test. We run the detection algorithm using data from a set number of 3 to 15 stations. For a given number of stations, we use a random subset of the stations and repeat the

Table 6. Tremor Detected by the WECC Method Compared to SOM Detection Method^a

	SOM Classification			
	Class S1	Class S2	Postevaluation	Detected Events Missed by the SOM
WECC	34	4	10	6
WECC with STA/LTA	25	3	4	2

^aNumber of tremor events correctly detected (class S1), detected events incorrectly classified as earthquakes (class S2), events discarded by the postevaluation step (postevaluation), and correctly detected events missed by the SOM (missed by the SOM).

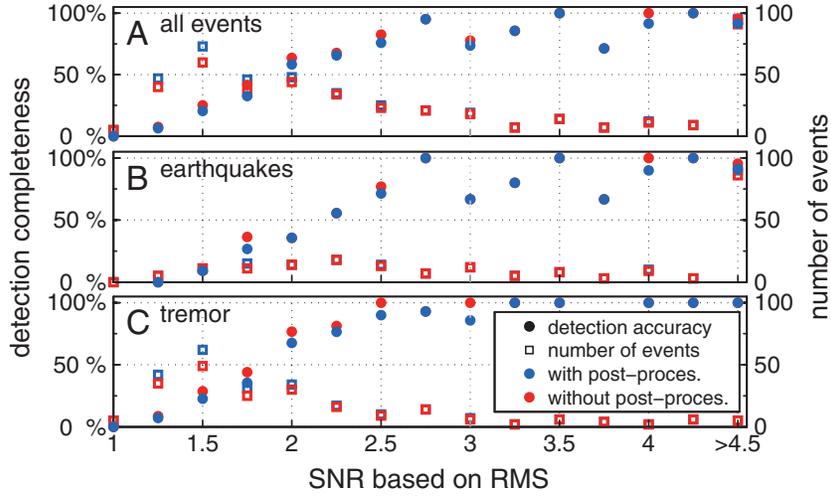


Figure 11. Mean detection accuracy of events for a given signal-to-noise ratio (SNR) (dots). Squares indicate number of events within each SNR bin. Blue corresponds to results with the noise postprocessing step included, and red corresponds to results before the noise postprocessing step is applied. (a) Detection accuracy versus SNR for all events, (b) detection accuracy versus SNR for earthquakes, and (c) detection accuracy versus SNR for tremor. Event classification is based on visual inspection. The y axis indicates the detection accuracy in percent and the total number of events determined by manual detection. Note that the results without the noise postprocessing step include more false picks, resulting in a lower detection accuracy.

analysis up to 250 times. We then calculate the mean and standard deviation values for each subset of stations (Figure 12). Figure 12a shows the detection accuracy, Figure 12b the number of correctly detected events, and Figure 12c the total detected signal length versus number of stations used. We perform the jackknife test for both classes S1 and S2. The figure shows consistently higher accuracy for class S2 (~ 90%) while the accuracy for the tremor class S1 increases from 50% for three stations to 80% for 15

stations. Using the minimum number of stations (3) results in a detection of roughly one third of all events with a high level of accuracy (90%) for class S2, but with a poor level of accuracy for class S1 (50%). As the number of stations is increased, the effects of noise at individual stations are reduced, particularly when borehole stations are considered.

[57] We also test the influence of individual stations on the detection accuracy. We perform a jackknife test for each

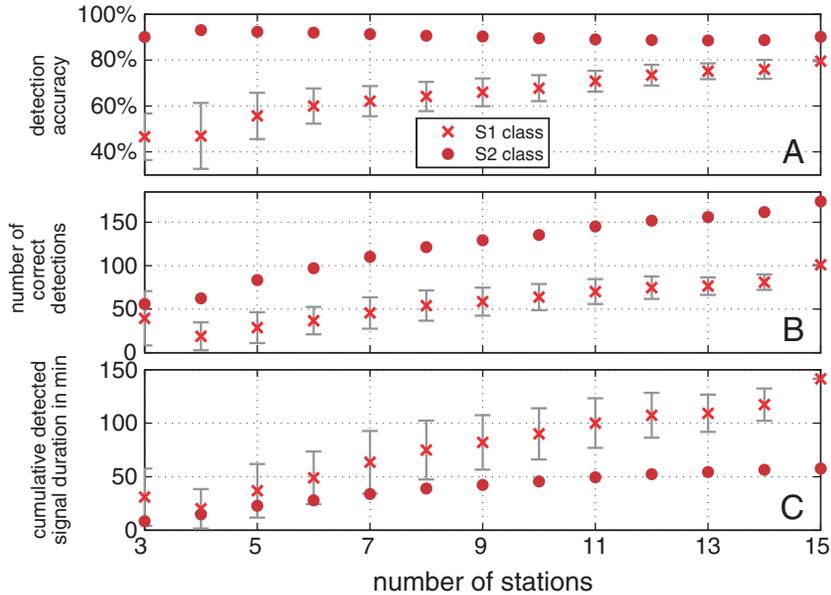


Figure 12. (a) Detection accuracy, (b) number of correctly detected events, and (c) detected signal length versus the number of stations used. Crosses display the result for the S1 (tremor) class, and dots indicate the results for the S2 (earthquake) class. Bars indicate the standard deviation within the averaged value.

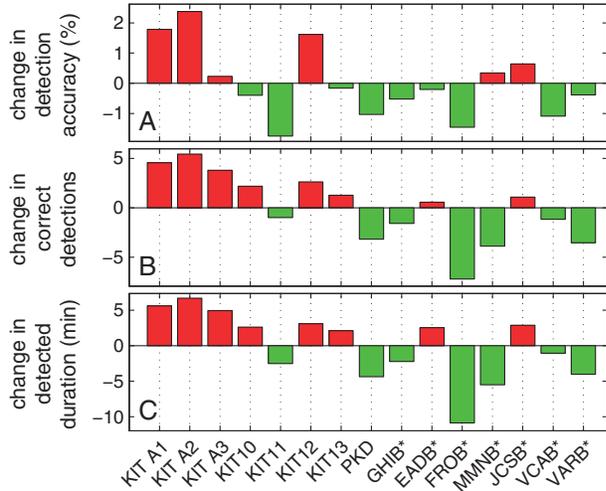


Figure 13. Comparison of mean performance values from the jackknife test. (a) deviation from the mean detection accuracy value for each (neglected) station, (b) deviation in the number of correctly detected events, and (c) the deviation in detected signal length. Asterisks designate borehole stations. Negative values (green) indicate improved accuracy when a station is included.

omitted station, with each test consisting of 250 random samples of (remaining) stations. We use the minimum number of occurrences for which any given station is included (N) in the random sampling to calculate the average values for each station over N samples (instead of 250). The jackknife test is performed in separate stages, in which a successively larger number of stations are omitted. The effect of omitting a specific station is shown in Figure 13. Figure 13a displays the deviation from the mean detection accuracy. Negative values indicate a decrease in the detection accuracy when a given station is omitted, indicating the importance of that particular station for the overall detection accuracy. Positive values suggest that a given station is less important for performance accuracy. An asterisk is used to denote the borehole stations.

[58] Figure 13 suggests that the stacked mini-array stations KIT A2 and KIT A1 have a negative influence on the detection accuracy while KIT 10 improves the accuracy slightly. Including all three stations in each of the mini-arrays decreases the number of correctly detected events and the detected signal duration, likely due to noisy conditions at the sites. Moreover, there is a clear difference between surface and borehole stations. Not surprisingly, the borehole stations perform better than the surface stations. The test indicates that the most valuable surface stations during the test data set period are KIT11 and PKD and the most valuable borehole stations are FROB and VARB.

4.2.5. Comparison With the ANSS Earthquake Catalog

[59] In order to test the sensitivity of the detection algorithm to seismic signals in general (i.e., classes S1 and S2), we compare the detection results with the ANSS earthquake catalog as described in section 3.4.1.

[60] Crosschecking the events found by manual inspection with the events detected by the SOM indicates that the SOM finds a total of 179 earthquakes within classes S1

and S2 (Figure 9). A comparison with the ANSS catalog indicates that 135 of those events were cataloged and that 9 tremor events were falsely classified as earthquakes. The remaining 44 events are not cataloged. Figure 14a shows the proportion of earthquakes from the ANSS catalog detected by the SOM. Figure 14b shows the ability of the method to detect almost every earthquake above a certain magnitude-distance threshold. Most of the earthquakes that occur within 150 km are classified correctly as earthquakes by the STA/LTA algorithm within the earthquake postprocessing step. Most of the earthquakes incorrectly identified as class S1 that occur at distances between 150 and 350 km are moved to class S2 by the earthquake catalog comparison described in section 3.4.1. Note that some small earthquakes occurring at large distances are falsely marked as detected because they occur within the window of a detected tremor event. We therefore apply the earthquake catalog comparison in the earthquake postprocessing only to earthquakes with a PGV > 1400 nm/s (indicated by the solid red line in Figure 14b). Figure 14b suggests that small earthquakes not in the ANSS catalog that occur close to the stations are often discarded during the noise postprocessing step.

4.3. Analysis of the 14 Month Data Set

[61] After quantitatively assessing the detection algorithm performance using the 3 week long test data set, we apply the method to the entire 14 month data set collected between 24

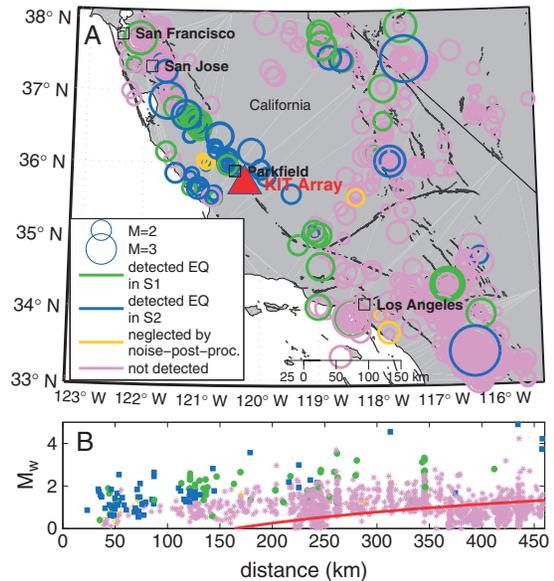


Figure 14. Earthquake detection results compared to the ANSS earthquake catalog. (a) All earthquakes occurring within our test data period as circles. Circle size corresponds to catalog magnitude. Earthquakes detected and classified in the S1 class are shown in green, detected earthquakes classified in S2 are shown in blue, detected events rejected by postprocessing are shown in yellow, and undetected events are shown in pink. (b) Earthquake magnitude versus distance; the color code corresponds to Figure 14a. Distance is from the center of the KIT array. Red solid line displays the PGV threshold of 1400 nm/s.

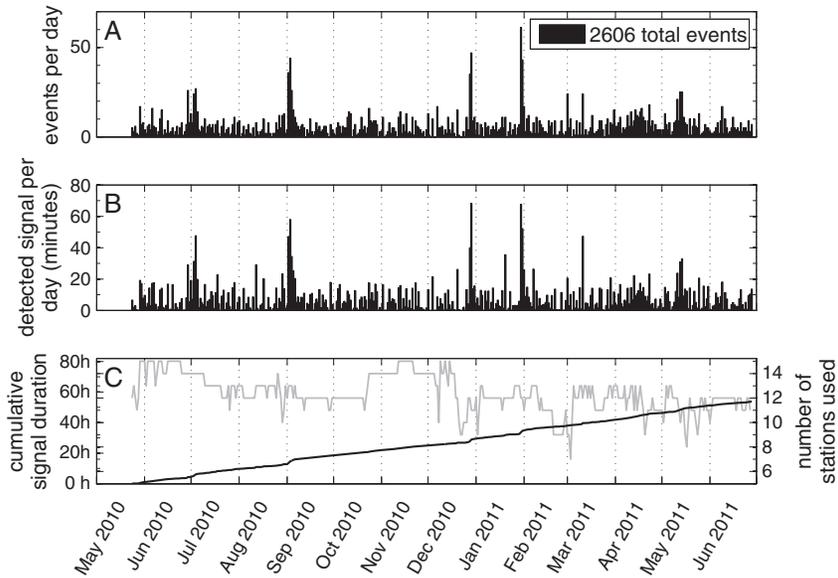


Figure 15. Detection results for the complete data set: (a) detected tremor events per day, (b) minutes of detected signal per day, and (c) cumulative signal duration (black curve). Figure 15c also shows the number of station used each day (gray curve).

May 2010 and 30 June 2011, using parameters established during calibration of the method. The parameter values are given in section 3 and Appendix A.

[62] The results for the complete data set are plotted in Figure 15. During the 13 month long study period, the SOM detects some tremor almost every day. We find over 2606 tremor detections occurring in windows totaling over 55 h, with an estimated detection accuracy of 80%. As mentioned in section 4.1, the method tends to split low amplitude tremor episodes into several individual bursts; thus, multiple detected time windows could be affiliated with a single extended event. The estimated number of events per day, ranging between 0 and 61, reflect high amplitude tremor arrivals or “bursts” that may occur within longer-duration, low-amplitude episodes. The gap in late December coincides with a drop in the number of stations used, which is due to station outages. However, generally we do not see a correlation between the number of stations used and the number of detections. The nearly linear increase in cumulative signal duration shown in Figure 15c suggests a constant rate of seismic activity as well as a stable detection sensitivity over the study period. There is an average of 6.5 events with a cumulative signal duration of 8.2 min per day. There are days when the number of detections are much higher than average, for example, in the beginning of September 2010 and at the end of January 2011. To identify teleseismic or regional wave arrivals at the array, we estimate the peak ground velocity (PGV). We find no evidence suggesting increase in detections is correlated with the occurrence of large earthquakes (i.e., with high PGV). Moreover, visual inspection does not suggest an increase in false detections for days with a higher than average number of detections. Thus, the step increases in the cumulative signal duration likely reflect the true behavior of ambient tremor. Similar observations of episodic tremor episodes in the area are previously described by *Nadeau*

and *Dolenc* [2005], *Nadeau and Guilhem* [2009], and *Shelly* [2010].

5. Conclusions

[63] We present a new method for tremor detection based on a neural network approach. The method identifies and distinguishes tremor, earthquakes, and noise based on frequency content and horizontal to vertical component products. We initially use a waveform envelope cross correlation to reduce the data volume. The method does not rely on a priori information such as event templates and is capable of identifying tremor bursts of variable duration, assuming some variation in tremor amplitude within the length of the noise window (420 s). Moreover, the sensitivity of the method can be adjusted by a set of normalization factors, F_{mean} and F_{std} .

[64] The method detects 2606 events within the continuously recorded data set from May 2010 to July 2011. The tremor detection accuracy is nearly 80% for the SOM and postprocessing steps, estimated by comparison to manual picks in a 3 week long test data set. We expect similar detection accuracy for the entire study period. A comparison of the earthquake detections with the ANSS catalog shows a detection accuracy of approximately 90%. The event detection is based on a minimum of three stations. However, optimum detection requires approximately 10 stations for our particular network configuration. And the overall best performance is achieved when four key stations (KIT11, PKD, FROB, and VARB) are included in the analysis. One limitation of the method is that it generally detects only a portion of the tremor episode, typically identifying high-amplitude portions of a longer, low-amplitude tremor episode.

[65] The method is in general very flexible and could be adjusted by supplementing further features and/or by changing the detection sensitivity. Although we tested a variety of

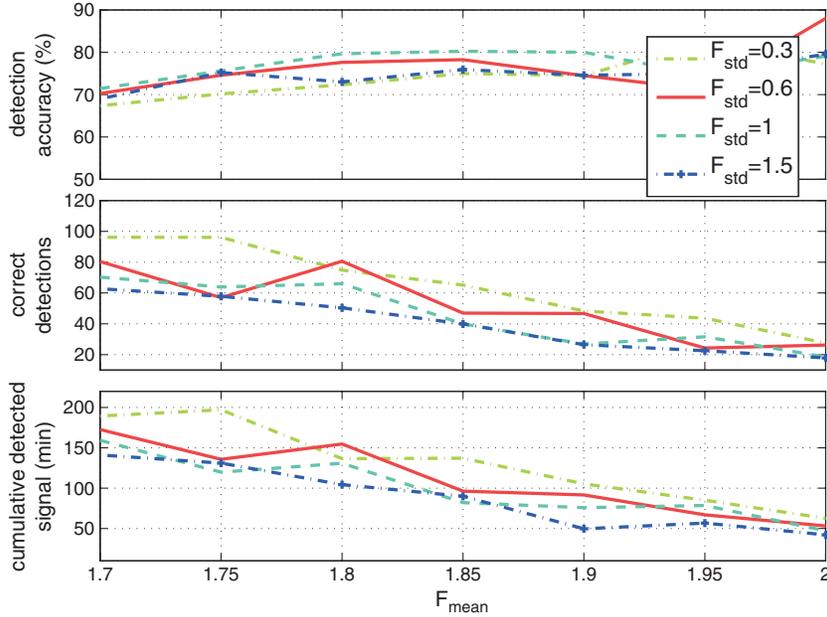


Figure A1. Normalization factor values for PQ_{abs} feature. The color coding corresponds to the F_{std} values shown in the legend.

features to discriminate tremor from noise in our data set, the features determined to be most discriminating may differ for other data sets. The fundamental advance of the method is that it does not rely on master templates and is not based on any assumptions about a minimum signal length. Removing such restrictions permits the detection of a wider range of event types than present methods, thereby increasing the potential for discovering tremor at different depths within the fault zone.

[66] The method does not provide locations for the detected tremor; however, the method could be useful to restrict the time period for a LFE search with a template-matching method. Other methods to locate tremor such as envelope triangulation could also be applied to the detected tremor periods. A follow-up paper is in preparation that describes tremor location with a time reversal approach.

Appendix A: Normalization Factor Determination

[67] In section 3.2.4 we introduce the extended softmax normalization, where we briefly discuss the possibility of shifting the range of linear normalization to the optimal position in the feature data set for discriminating between noise and different seismic signals. One can change the range of linear normalization by changing the F_{mean} and F_{std} factors. Here we discuss determining the F_{mean} and F_{std} values in greater detail. The optimal F -values are not known a priori, because the range of feature values is not known a priori, requiring manual determination.

[68] As mentioned in section 3.2.4, we use a test data set containing 3 weeks of seismograms with multiple tremor episodes, in order to determine the optimal F -values. We start by comparing manually picked time windows with SOM-picked time windows to determine a measure of the detection accuracy. We repeat the process and detection accuracy estimation, each time changing the F -value pairs, while performing a grid search of the best F -values for an

individual feature. The values corresponding to the highest accuracy empirically determine the values for F_{mean} and F_{std} .

[69] We face a trade-off between detection accuracy and sensitivity. Figure A1 shows the SOM performance for various F_{mean} and F_{std} values for the PQ_{abs} feature value normalization. By choosing a lower F_{mean} value, the number of correctly detected signals increases, as does the number of incorrectly detected signals. As a result, the detection accuracy is lower than for higher F_{mean} values. The grid search depicted in Figure A1 shows that values of $F_{\text{mean}} = 1.8$ and $F_{\text{mean}} = 0.6$ provide an optimal balance between sensitivity and accuracy. Table 2 summarizes the values for other features. Table A1 shows the calculated mean and standard deviation values used in the extended softmax normalization in equation (5) for each station and feature.

Table A1. F_{mean} and F_{std} Values Used in the Extended Softmax Normalization for Each Feature and Station (in Equation (5))^a

Feature	2–8 Hz		15–30 Hz		0.5–1.5 Hz		PQ_{abs}	
	F_{mean}	F_{std}	F_{mean}	F_{std}	F_{mean}	F_{std}	F_{mean}	F_{std}
KIT A1	3.17	17.08	4.69	27.08	5.44	7.39	9.85	7.27
KIT A2	5.68	21.88	2.47	20.03	11.08	29.03	10.93	7.80
KIT A3	2.98	9.57	6.66	18.05	2.59	7.66	9.42	7.56
KIT10	1.90	6.70	1.62	3.98	6.70	6.06	9.29	6.10
KIT11	1.47	4.83	0.53	1.09	2.72	2.69	7.14	5.73
KIT12	0.78	1.53	10.46	24.99	1.31	1.39	6.76	6.12
KIT13	2.84	12.51	4.12	12.06	6.81	7.01	9.54	6.83
GHIB	7.08	19.49	1.13	2.65	4.47	4.70	12.22	7.99
EADB	3.21	8.75	1.14	2.41	3.62	3.15	10.95	7.53
FROB	3.54	15.55	1.57	7.57	3.87	3.90	9.97	7.52
MMNB	2.26	10.35	0.97	2.77	5.18	4.94	8.94	6.87
JCSB	1.82	8.66	1.05	3.55	0.53	0.50	8.24	5.26
VCAB	8.33	31.35	5.74	25.71	5.46	5.17	13.45	9.42
VARB	1.30	6.23	1.07	2.05	0.53	0.88	6.88	5.42
PKD	0.73	3.29	0.26	0.69	3.72	3.28	4.02	4.73

^aKIT A1, A2, and A3 represent the mini-arrays for stations KIT01–KIT09.

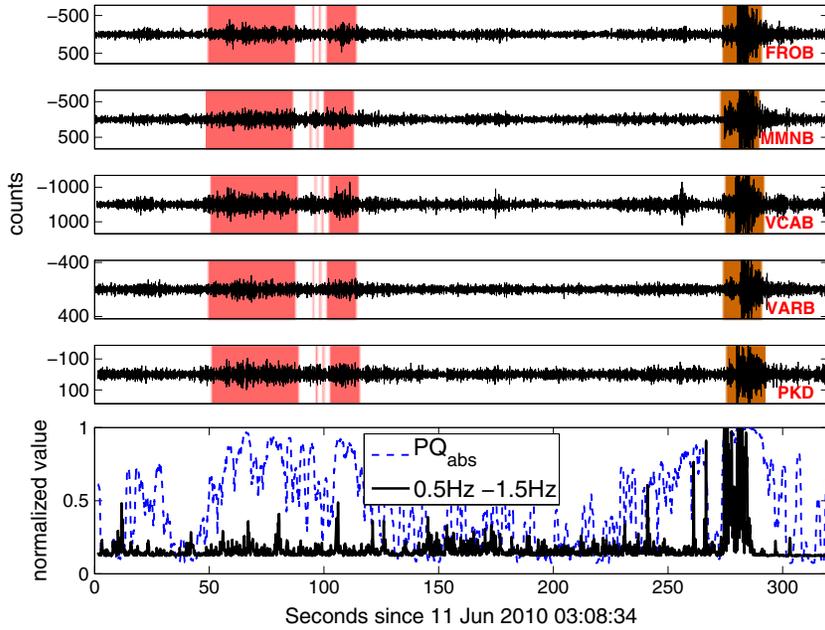


Figure C1. Example tremor waveforms with low signal-to-noise ratio followed by a small earthquake. Data are filtered between 2 and 8 Hz. Color code indicates the clustering result following the postprocessing step: detected tremor (red), earthquake (brown), and noise (white). The sixth panel shows the normalized feature values for the PQ_{abs} feature and the lower frequency band for station PKD.

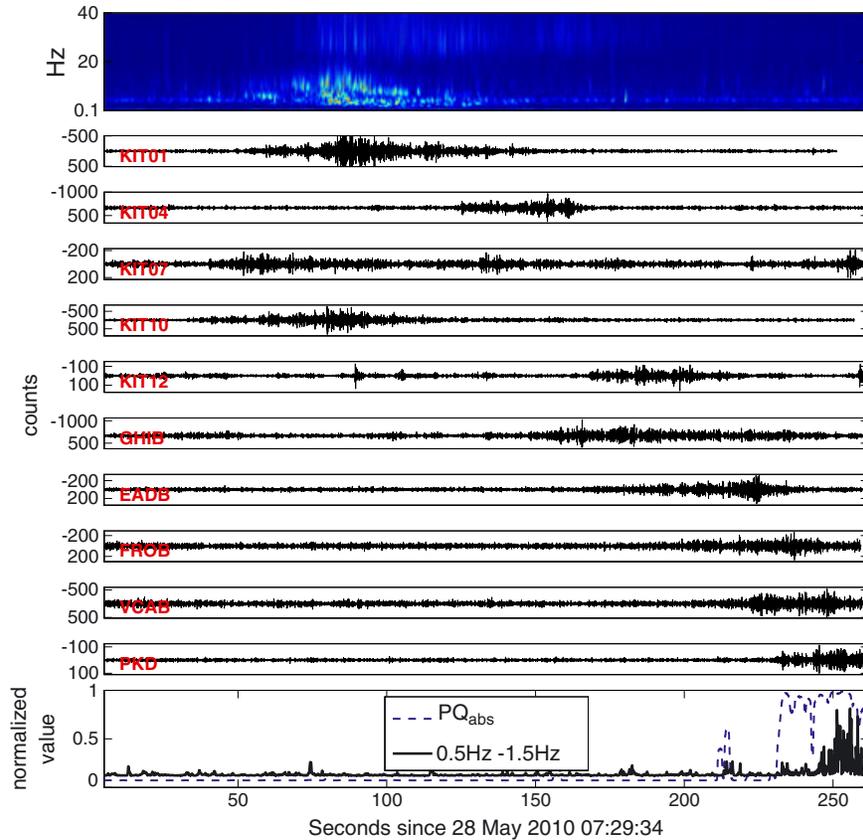


Figure C2. Example infrasound event waveforms, filtered between 2 and 8 Hz. White background color indicates the classification as noise as the clustering result. The first panel displays a spectrogram of the raw waveforms recorded at station KIT 01, showing most infrasound energy concentrated between 1 and 10 Hz. The twelfth panel shows the normalized feature values for the PQ_{abs} feature and the lower frequency band for station PKD. Note the large moveout between the stations.

Appendix B: Noise Postprocessing

[70] In section 3.4 we introduce a noise postprocessing step applied to time windows detected by the SOM. The basis of the noise postprocessing step is the assumption that event waveforms exhibit coherency across stations, whereas noise bursts do not. Thus, we test every detected time window for plausibility with a cross correlation of waveform envelopes between stations.

[71] First, we extend the detected time window by two percent of the time window length at the beginning and end. To account for short time windows, we also add 3 s at the beginning and end of each signal. The extension is necessary as sometimes the detected time windows are only fraction of an extended tremor signal.

[72] Second, we calculate the envelopes within the extended time windows for each Z component trace and smooth the envelopes with a window of 0.6% of the time window length.

[73] Third, we calculate the cross-correlation coefficient between all combinations of envelopes, using each station once as a master station. We permit a 4 s time lag between the envelopes in the cross correlation. The 4 s time lag is meant to account for cases where alignment between traces is incorrect. An example of where signal misalignment could occur is at a noisy station or where multiple events in short succession are recorded at various source-station geometries. The alignment described in section 3.2.2 is then sensitive to the bigger event with the higher amplitudes leading to a larger moveout for other events contained within the same window.

[74] After calculating the cross-correlation values, we average the three highest cross-correlation coefficients for each master station and select the highest average cross-correlation coefficient for each detected time window. Whether or not the noise postprocessing step accepts or rejects an event is based only on three station values, i.e., the minimum number of stations required for event detection. Finally, we neglect time windows with an average cross-correlation value below 0.8.

Appendix C: Additional Events Detected by the SOM

[75] In addition to the two tremor waveform examples shown in section 4.1, here we show two additional examples of detected tremor, as well as a detected earthquake (Figure C1), and an infrasound event (Figure C2). The first to fifth panels in Figure C1 show the waveform data, filtered between 2 and 8 Hz for selected stations. The sixth panel documents the normalized values for the PQ_{abs} feature and the lower frequency value between 0.5 and 1.5 Hz. During the seismic events, the PQ_{abs} feature values are increased, while the lower frequency band values are only increased during the earthquake. The difference in the normalized amplitude in the lower frequency band is used by the SOM to distinguish between tremor and earthquakes, although, in this particular example the earthquake was classified by the STA/LTA trigger in the postprocessing step. Figure C2 shows an example of an infrasound event. The twelfth panel documents the increase of the PQ_{abs} feature values during the infrasound wave train, which leads the SOM to misclassify infrasound events as tremor.

[76] **Acknowledgments.** The work was supported by funding from the “Concept of the Future” of the Karlsruhe Institute of Technology (KIT) within the framework of the national “Initiative for Excellence”. Additionally, this research was supported by the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-0529922 and USGS Cooperative Agreement 07HQAG0008. The SCEC contribution number for this paper is 1639. The authors would like to thank David R. Shelly, Arthur F. McGarr, and two anonymous reviewers for their constructive comments, which helped to improve the quality of the manuscript. The KABBA network provided seismic data from the temporary station deployment, and the permanent seismic data are distributed by the Northern California Earthquake Data Center (NCEDC). The HRSN data as well as the BDSN data are contributed by the Berkeley Seismological Laboratory, University of California, Berkeley. The ANSS Worldwide Earthquake Catalog is provided by Advanced National Seismic System (ANSS). The authors would like to thank the landowners for permission to install seismic stations on their property. In addition, we would like to thank Werner Scherer, Kayla Kroll, Tien-Hui Wang, and Peter Duffner for their help during station services and deployment. All calculations of the neural network algorithm in section 3.3 were performed using the SOM Toolbox of *Vesanto* [2000].

References

- Advanced National Seismic System (2012). available at: <http://quake.geo.berkeley.edu/anss> [17 Jan 2012].
- Allen, R. (1982), Automatic phase pickers: Their present use and future prospects, *Bull. Seismol. Soc. Am.*, 72(6B), S225–S242.
- Beroza, G., and S. Ide (2011), Slow earthquakes and nonvolcanic tremor, *Annu. Rev. Earth Planet. Sci.*, 39, 271–296.
- Brown, J., G. Beroza, and D. Shelly (2008), An autocorrelation method to detect low frequency earthquakes within tremor, *Geophys. Res. Lett.*, 35, L16305, doi:10.1029/2008GL034560.
- Brown, J., G. Beroza, S. Ide, K. Ohta, D. Shelly, S. Schwartz, W. Rabbal, M. Thorwart, and H. Kao (2009), Deep low-frequency earthquakes in tremor localize to the plate interface in multiple subduction zones, *Geophys. Res. Lett.*, 36, L19306, doi:10.1029/2009GL040027.
- Brudzinski, M., and R. Allen (2007), Segmentation in episodic tremor and slip all along Cascadia, *Geology*, 35(10), 907–910.
- Davies, D., and D. Bouldin (1979), A cluster separation measure, *IEEE T. Pattern. Anal.*, PAMI-1(2), 224–227.
- Ghosh, A., J. Vidale, Z. Peng, K. Creager, and H. Houston (2009), Complex nonvolcanic tremor near Parkfield, California, triggered by the great 2004 Sumatra earthquake, *J. Geophys. Res.*, 114, B00A15, doi:10.1029/2008JB006062.
- Gomberg, J., J. L. Rubinstein, Z. Peng, K. C. Creager, J. E. Vidale, and P. Bodin (2008), Widespread triggering of nonvolcanic tremor in California, *Science*, 319(5860), 173, doi:10.1126/science.1149164.
- Jepson, D., and B. Kennett (1990), Three-component analysis of regional seismograms, *Bull. Seismol. Soc. Am.*, 80(6B), 2032–2052.
- Katsumata, A., and N. Kamaya (2003), Low-frequency continuous tremor around the Moho discontinuity away from volcanoes in the southwest Japan, *Geophys. Res. Lett.*, 30(1), 1020, doi:10.1029/2002GL0159812.
- Köhler, A., M. Ohrnberger, and F. Scherbaum (2009), Unsupervised feature selection and general pattern discovery using self-organizing maps for gaining insights into the nature of seismic wavefields, *Comput. Geosci.*, 35(9), 1757–1767.
- Köhler, A., M. Ohrnberger, and F. Scherbaum (2010), Unsupervised pattern recognition in continuous seismic wavefield records using self-organizing maps, *Geophys. J. Int.*, 182(3), 1619–1630.
- Kohonen, T. (2001), *Self-Organising Maps.*, vol. 30, Springer, Verlag New York, Inc. Secaucus, NJ, USA.
- Martin, R. (1994), Spectral subtraction based on minimum statistics, *Proc. Eur. Signal Processing Conf.*, 1182–1185.
- Martin, R. (2001), Noise power spectral density estimation based on optimal smoothing and minimum statistics, *IEEE T. Speech. Audi. P.*, 9(5), 504–512.
- McCausland, W., S. Malone, and D. Johnson (2005), Temporal and spatial occurrence of deep non-volcanic tremor: From Washington to Northern California, *Geophys. Res. Lett.*, 32, L24311, doi:10.1029/2005GL024349.
- Nadeau, R., and D. Dolenc (2005), Nonvolcanic tremors deep beneath the San Andreas fault, *Science*, 307(5708), 389.
- Nadeau, R. M., and A. Guilhem (2009), Nonvolcanic tremor evolution and the San Simeon and Parkfield, California, earthquakes, *Science*, 325(5937), 191–193, doi:10.1126/science.1174155.
- Obara, K. (2002), Nonvolcanic deep tremor associated with subduction in southwest Japan, *Science*, 296(5573), 1679–1681.

- Payero, J., V. Kostoglodov, N. Shapiro, T. Mikumo, A. Iglesias, X. Pérez-Campos, and R. Clayton (2008), Nonvolcanic tremor observed in the Mexican subduction zone, *Geophys. Res. Lett.*, *35*, L07305, doi:10.1029/2007GL032877.
- Peng, Z., and K. Chao (2008), Non-volcanic tremor beneath the central range in Taiwan triggered by the 2001 M_w 7.8 Kunlun earthquake, *Geophys. J. Int.*, *175*(2), 825–829.
- Peng, Z., J. Vidale, K. Creager, J. Rubinstein, J. Gombert, and P. Bodin (2008), Strong tremor near Parkfield, CA, excited by the 2002 Denali Fault earthquake, *Geophys. Res. Lett.*, *35*, L23305, doi:10.1029/2008GL036080.
- Peng, Z., J. E. Vidale, A. G. Wech, R. M. Nadeau, and K. C. Creager (2009), Remote triggering of tremor along the San Andreas Fault in central California, *J. Geophys. Res.*, *114*, B00A06, doi:10.1029/2008JB006049.
- Peterson, C., and D. Christensen (2009), Possible relationship between non-volcanic tremor and the 1998–2001 slow slip event, south central Alaska, *J. Geophys. Res.*, *114*, B06302, doi:10.1029/2008JB006096.
- Pyle, D. (1999), *Data Preparation for Data Mining*, vol. 1, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Rogers, G., and H. Dragert (2003), Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip, *Science*, *300*(5627), 1942–3.
- Ryberg, T., C. Haberland, G. Fuis, W. Ellsworth, and D. Shelly (2010), Locating non-volcanic tremor along the San Andreas fault using a multiple array source imaging technique, *Geophys. J. Int.*, *183*(3), 1485–1500.
- Schwartz, S., and J. Rokosky (2007), Slow slip events and seismic tremor at circum-Pacific subduction zones, *Rev. Geophys.*, *45*(3), RG3004, doi:10.1029/2006RG000208.
- Shelly, D. (2010), Migrating tremors illuminate complex deformation beneath the seismogenic San Andreas fault, *Nature*, *463*(7281), 648–652.
- Shelly, D., and J. Hardebeck (2010), Precise tremor source locations and amplitude variations along the lower-crustal central San Andreas Fault, *Geophys. Res. Lett.*, *37*, L14301, doi:10.1029/2010GL043672.
- Shelly, D., G. Beroza, and S. Ide (2007), Non-volcanic tremor and low-frequency earthquake swarms, *Nature*, *446*(7133), 305–307.
- Stockwell, R., L. Mansinha, and R. Lowe (1996), Localization of the complex spectrum: The S transform, *IEEE T. Signal. Proces.*, *44*(4), 998–1001.
- van der Elst, N., and E. Brodsky (2010), Connecting near-field and far-field earthquake triggering to dynamic strain, *J. Geophys. Res.*, *115*, B07311, doi:10.1029/2009JB006681.
- Vesanto, J. (2000), Neural network tool for data mining: SOM toolbox, in *Proceedings of Symposium on Tool Environments and Development Methods for Intelligent Systems (TOOLMET2000)*, pp. 184–196.
- Vesanto, J., and E. Alhoniemi (2000), Clustering of the self-organizing map, *IEEE T. Neural. Networ.*, *11*(3), 586–600.
- Walter, J., S. Schwartz, J. Protti, and V. Gonzalez (2011), Persistent tremor within the northern Costa Rica seismogenic zone, *Geophys. Res. Lett.*, *38*(1), L01307, doi:10.1029/2010GL045586.
- Wang, T., E. Cochran, D. Agnew, and D. Oglesby (2013), Infrequent triggering of tremor along the San Jacinto Fault near Anza, California, *Bull. Seismol. Soc. Am.*, *103*(4), 2482–2497, doi:10.1785/0120120284.
- Wech, A., and K. Creager (2008), Automated detection and location of Cascadia tremor, *Geophys. Res. Lett.*, *35*(20), L20302, doi:10.1029/2008GL035458.
- White, J. (1964), Motion product seismograms, *Geophysics*, *29*, 288–298.