Detection of low-frequency earthquakes by matched filter technique using the product of mutual information and correlation coefficient

Ryo Kurihara (rkuri@eri.u-tokyo.ac.jp)
The University of Tokyo: Tokyo Daigaku  https://orcid.org/0000-0003-2954-8486

Aitaro Kato
The University of Tokyo: Tokyo Daigaku

Sumito Kurata
The University of Tokyo: Tokyo Daigaku

Hiromichi Nagao
The University of Tokyo: Tokyo Daigaku

Full paper

Keywords: Matched filter technique, Low-frequency earthquakes, Mutual information, Kirishima volcano

DOI: https://doi.org/10.21203/rs.3.rs-769134/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Title: Detection of low-frequency earthquakes by matched filter technique using the product of mutual information and correlation coefficient

Author #1: Ryo Kurihara, Earthquake Research Institute, the University of Tokyo, rkuri@eri.u-tokyo.ac.jp
Author #2: Aitaro Kato, Earthquake Research Institute, the University of Tokyo, akato@eri.u-tokyo.ac.jp
Author #3: Sumito Kurata, Graduate School of Information Science and Technology, the University of Tokyo, kurata@mist.i.u-tokyo.ac.jp
Author #4: Hiromichi Nagao, Earthquake Research Institute, the University of Tokyo, nagaoh@eri.u-tokyo.ac.jp

Indicate the corresponding author: Ryo Kurihara
Abstract

Matched filter technique is often used to detect microearthquakes such as deep low-frequency (DLF) earthquakes. It compares correlation coefficients (CC) between waveforms of template earthquakes and the observed data. Conventionally, the sum of CC at multiple seismic stations is used as an index to detect the DLF earthquakes. A major disadvantage of conventional method is drastically reduced detection accuracy when there are too few seismic stations. A new matched filter method proposed in this study can accurately detect microearthquakes using only a single station. It adopts mutual information (MI) in addition to CC to measure the similarity between the template and target waveforms. The method uses the product of MI and CC (MICC) as an index to detect DLF earthquakes. This index shows a distinct peak corresponding to an earthquake signal in a synthetic data set consisting of artificial noise and the waveform of a DLF earthquake.

Application of this single-station method to field observations of Kirishima volcano, one of the most active volcanoes in Japan, detected a total of 354 DLF earthquakes from the data in December 2010, whereas the catalog of the Japan Meteorological Agency shows only two. The catalog of DLF earthquakes constructed here shows similar temporal behavior to that found by conventional matched filter method using the sum of the CC of the six stations near the volcano. The proposed method successfully identified approximately 80% of the earthquakes in the conventionally constructed catalogs. These results suggest that the proposed method can greatly contribute to the accurate cataloging of DLF earthquakes using only a single seismic station.

Keywords
Matched filter technique, Low-frequency earthquakes, Mutual information, Kirishima volcano
**Introduction**

One of established methods for the automatic detection of microearthquakes, is the matched filter technique employing Pearson's correlation coefficients (CC) between template waveforms—which are previously seismic waveforms—and target waveforms. It can detect microearthquakes such as afterquakes (Peng and Zhao 2009), low-frequency earthquakes and low-frequency tremor in the plate subduction zone and volcanic regions (Gibbons and Ringdal 2006; Shelly et al. 2007; Shapiro et al. 2017; Yukutake et al. 2019; Kurihara et al. 2019; Kato and Nakagawa 2020). The matched filter method is especially useful in situations of continued intense seismic activity such as aftershocks and seismic swarms because the waveforms of low-magnitude earthquakes are masked by the large-amplitude waves from multiple events. The method has been used to detect many low-frequency earthquakes; the resulting catalogs of earthquakes provide a precise view of spatial and temporal evolution. Detailed analysis of the low-frequency earthquake activity has improved understanding of wider geophysical activity such as the occurrence of slow slip events and the migration of volcanic fluids.

Although matched filter method can be applied to data from only one station (for example, when the seismic network is small) (e.g. Vuan et al., 2018; Wech et al., 2020), the sum of CC between template waveforms and in the three components data measured at multiple observation stations was often used in most cases (e.g. Gibbons and Ringdal, 2006; Shelly et al., 2007). The analysis of low-frequency earthquakes usually employs a time window of 4–5 s for filtered waveforms at 1–4 Hz or 2–8 Hz. The similarities of waveforms measured at multiple stations are evaluated by CC stacked over a seismic network. Earthquakes are identified by the summed CC being larger than a threshold value. Using stacked CC from multiple stations is advantageous over using of a single station to detect small-magnitude earthquakes including low-frequency earthquakes, because the CC from a single station always become high, as noise shares the same frequency band as the signal, resulting in substantial false detections. In other words, conventional matched filter technique is applicable only in regions sufficiently covered by observation stations and with well-determined seismic catalogs. Single-station matched filter method could potentially be effective in various regions in which microearthquakes occur, thus allowing detection of much smaller events that are only recorded by a single station (e.g. Vuan et al., 2018). However, much improvements of
technique is required to maintain the quality of the catalog.

As CC can be calculated quickly, they can be used easily to evaluate waveform similarity. However, they do not necessarily evaluate the overall similarity because CC are generally sensitive to a portion of the waveform depending on their calculation formula. To reduce the contribution of large-amplitude parts, Gao and Kao (2020) proposed a method of dividing the time window, which could effectively distinguish seismic waves generated from different epicenters. The present work, unlike the previous study, concerns the detection of low-frequency earthquakes, so we cannot expect to improve detection accuracy by dividing the time window, due to the sharing of the same frequency band between signal and noises. Therefore, we tried to effect improvements by introducing another index in addition to CC.

Statistical studies have proposed using other indices in addition to CC such as the mutual information (MI), the maximum information coefficient (Reshef et al., 2011), and the total information coefficient (Reshef et al., 2016). These indices show the similarity of two data sets, including non-linear relationships not evaluated by CC. MI has been used to evaluate electron correlation in the fields of chemical physics (Sagar and Guevara 2005) and medical imaging (Pluim et al. 2003). Various studies have used other indices, but they are generally more computationally costly than MI. Therefore, we introduce MI, which can evaluate with low computational cost the degree of waveform similarity in small-amplitude parts. In order to take advantage of both MI and CC, we propose a new method for detecting deep low-frequency (DLF) earthquakes using their product (called MICC) as an index.

**Data and Method**

This study considers the waveforms of DLF earthquakes in Kirishima volcano, one of the most active volcanoes in Japan. Waveform data are from the high-sensitivity seismograph network (Hi-net) of NIED (Okada et al. 2004; National Research Institute for Earth Science and Disaster Resilience 2019). We apply a band-pass filter of 1–8 Hz, and decimate the waveform from 100 to 25 Hz sampling before calculations.

Conventional matched filter usually evaluates detection using stacked CC between observed and template waveforms of three components measured at multiple seismic stations. The template is selected from previously observed earthquakes. The template events used here are DLF earthquakes in the unified catalog of the Japan
The CC of component \( j \) at seismic station \( i \) are calculated as

\[
CC(i, j, t_{tg} + \Delta t_i) = \frac{\sum_{\tau} (v_{tp}(i, j, t_{tp} + \Delta t_i + \tau)v_{tg}(i, j, t_{tg} + \Delta t_i + \tau))}{\sqrt{\sum_{\tau} (v_{tp}^2(i, j, t_{tp} + \Delta t_i + \tau))} \sqrt{\sum_{\tau} (v_{tg}^2(t_{tg} + \Delta t_i + \tau))}}
\] (1).

Here, \( v_{tp}(i, j, t) \) and \( v_{tg}(i, j, t) \) are the velocities of the template earthquake and target data at time \( t \) of component \( j \) in station \( i \), respectively; \( t_{tp} \) and \( t_{tg} \) are the times of occurrence of the template earthquake and the target event, respectively; \( \Delta t_i \) is the time of S-wave propagation from the origin of the template earthquake to its arrival at station \( i \); \( \tau \) corresponds to each time step in the window length. This analysis sets the length of the time window to be 8 s, and allocates its center to the arrival time of the template S-wave earthquake observed by the JMA. When the JMA did not determine the arrival time at a station, we estimate it using the occurrence time in the JMA catalog and the JMA’s 1D velocity structure model (Ueno et al. 2002). When using multiple stations, CC are summed as follows:

\[
\text{Summed } CC(t_{tg} + \Delta t_i) = \sum_i \sum_j CC(i, j, t_{tg} + \Delta t_i)
\] (2).

This sum is high when the peak CC values in the waveforms of each station and each direction are stacked. In other words, when the hypocenter of the target event is near that of the template event, the stacked peak values make the summed CC high.

MI is defined using the normalized amplitudes of two variables. First, we normalized the waveforms within each time window using the maximum absolute amplitudes. We then assigned the normalized velocities of each time step in each time window, \( \bar{v}_{tp}(t) \) and \( \bar{v}_{tg}(t) \), to the x- and y-axis respectively, as shown in Figure 1. By dividing the normalized velocities to 5 \( \times \) 5 cells, we converted the velocities into integers \( n_{tp} \) and \( n_{tg} \) between 1 and 5 as follows:

\[
\bar{v}_{tp}(t) = \frac{v_{tp}(t)}{\max(|v_{tp}(t)|)} \quad \cdots (3)
\]

\[
\bar{v}_{tg}(t) = \frac{v_{tg}(t)}{\max(|v_{tg}(t)|)} \quad \cdots (4)
\]

\[
n_{tp}(t) = \text{floor} \left( (\bar{v}_{tp}(t) + 1.4) \times 2.5 \right) \quad \text{(When } \bar{v}_{tp}(t) < 1) \cdots (5)
\]
\[ n_{tg}(t) = \text{floor}((\bar{v}_{tg}(t) + 1.4) \times 2.5) \quad \text{(When } \bar{v}_{tg}(t) < 1 \text{)} \]  \hspace{1cm} (6)

\[ n_{tp}(t) = 5 \quad \text{(When } \bar{v}_{tp}(t) = 1) \]  \hspace{1cm} (7)

\[ n_{tg}(t) = 5 \quad \text{(When } \bar{v}_{tg}(t) = 1) \]  \hspace{1cm} (8)

The notation \( | | \) means the absolute function, and \( \text{floor}(\ ) \) means the floor function converting a real number to the largest integer smaller than itself. The constants 1.4 and 2.5 convert the velocity into integers. This calculation means that \( n = 1 \) when the normalized velocity \( \bar{v}(t) \) ranges from −1.0 to −0.6, \( n = 2 \) when \( \bar{v}(t) \) is between −0.6 and −0.2, \( n = 3 \) when \( \bar{v}(t) \) is between −0.2 and 0.2, \( n = 4 \) when \( \bar{v}(t) \) is between 0.2 and 0.6, and \( n = 5 \) when \( \bar{v}(t) \) is between 0.6 and 1.0.

\[ \text{MI}(t) = \sum_{n_{tp}=1}^{5} \sum_{n_{tg}=1}^{5} p(n_{tp}, n_{tg}) \log \frac{p(n_{tp}, n_{tg})}{p(n_{tp})p(n_{tg})} \]  \hspace{1cm} (9)

where \( p(n_{tp}) \) and \( p(n_{tg}) \) are the probabilities of \( n_{tp} \) and \( n_{tg} \), respectively; \( p(n_{tp}, n_{tg}) \) is the joint probability in the cells of \( n_{tp} \) and \( n_{tg} \). As the time window is 8 s and the sampling rate is 25 Hz (i.e., 200 time-steps for each time window), we obtain \( p(n_{tp}) \) and \( p(n_{tg}) \) by dividing the number of points in the bins by 200.

The upper limit of MI depends on the data set according to equation (9). We use the sum of the information entropy \( h_{tp} \) and \( h_{tg} \) to normalize the MI using the following equations (cf. Zhang, 2015):

\[ h_{tp} = \sum -p(n_{tp}) \log p(n_{tp}) \]  \hspace{1cm} (10)

\[ h_{tg} = \sum -p(n_{tg}) \log p(n_{tg}) \]  \hspace{1cm} (11)

\[ \overline{\text{MI}} = \frac{2.0 \times \text{MI}}{h_{tp} + h_{tg}} \]  \hspace{1cm} (12)

Hereafter, we simply describe \( \overline{\text{MI}} \) as MI.

MI is an index that reflects the linear or non-linear relation between two variables. If the relation is linear, both MI and CC are high (Figure 1a). While if the relation is non-linear, MI has a non-zero value (Figure 1b). According to equation (1), CC
gives more weight to data points with high amplitudes in phase, while MI increases when
the data points are concentrated in a small number of cells. In other words, CC are high
when the peaks of two variables match, even if the shapes of the waveforms are different
(Figure 1c).

**Figure 1** Scatter plots of relations between two variables (top) and their waveforms at the
same time (bottom). MI and CC are given above the plots. Black grid lines correspond to
the borders of the bins used to calculate MI. (a) Two variables with the same sinusoidal
waves. (b) Two variables with sinusoidal waves and a time shift. (c) One variable with a
sinusoidal wave and one with a spike-like wave.

For detection, we use here the MICC index defined as the product of MI and
CC:

$$\text{MICC}(t) = \text{MI}(t) \cdot \text{CC}(t) \quad \cdots (13).$$

The product of two variables contains the characteristics of MI, including information
about the small-amplitude parts, and the characteristics of the CC, which correspond to
the consistency of large-amplitudes parts.

**Synthetic Tests**

We test how the index values of MI, CC, and MICC change in response to a variety
of noise and signals for DLF earthquakes. The analysis uses a synthetic data set
comprising two types of noise (Gaussian and sinusoidal) added to a template waveform
of data from a DLF earthquake recorded by the N.SUKH station at 22:03:56 (JST) on 7
August 2010 (Figure 2).

Figure 2 (a) Distribution of Hi-net stations (blue triangles) and the epicenter (yellow star) of the DLF earthquake (7 August 2010, 22:03:56, JST) used as a template here. The red triangle shows the location of Kirishima volcano. (b) Waveforms of three components of the earthquake recorded at the six stations. The horizontal axis is the time elapsed from the origin. Red parts of the waveforms show the template waveform used in this study.

Gaussian noise

We computed MI, CC, and MICC for the synthetic data. Artificial observation data were made by adding a filtered template waveform (1–8 Hz) with various amplitudes proportional to the signal-to-noise (SN) ratio at the N.SUKH station to Gaussian noise with a variance of 1. The variance of the template waveform is the same as the value of the SN ratio.

When the SN ratio is small, the peaks of each index are unclear; in other words, DLF events are not detectable in the synthetic data (Figure 3a). When the SN ratio is high, MI, CC, and MICC have distinct peaks detectable by setting an appropriate threshold for each index (Figure 3b, c). In the case of Gaussian noise, DLF events are detectable using any of the indices when the SN ratio is over 0.5.
Figure 3 MI, CC, and MICC for a signal comprising the template waveform with added Gaussian noise. The indices are calculated for each time step with the moving time window. The SN ratios are (a) 0.2, (b) 0.5, and (c) 1.0. Red dots show the index corresponding to the template signal.

The test results can verify how the three indices evaluate the similarity of the waveforms. Scatter plots compare the amplitude of the template earthquake with that of the target data (Figure 4a, b). The relationship becomes linear at high SN ratio (Figure 4b). CC rapidly increase when the SN ratio is relatively low, and reach a steady value as the SN ratio becomes large; in contrast, MI gradually increases as the SN ratio rises from 0 to 5 (Figure 4c). MICC gradually increases in any range of SN ratio corresponding to an increase of MI and CC.
Figure 4 (a)–(b) Scatter plots of detection using a waveform comprising a DLF earthquake signal with Gaussian noise at SN ratios of (a) 0.2 and (b) 1.0. Scatter plots (top panels) compare the amplitude of the template waveform with the amplitude of the target data at each time step. Bottom panels show the waveforms of the target data and template waveform at that time. (c) The three indices (MI, CC, and MICC) plotted with respect to the SN ratio of the waveform.

With Gaussian noise, even at low SN ratio, the values of MI, CC, and MICC seem to be large. However, noise associated with microseisms and human activities often overlap with the 1–8 Hz characteristic band of actual DLF earthquakes. The next tested case has sinusoidal noise waves roughly matching the frequency band of DLF earthquakes.
Sinusoidal noise

As in the case of Gaussian noise, the test considers a template waveform with added noise, this time a sinusoidal wave of 1.25 Hz frequency. The variance of the noise is 1. The variance of CC tends to be always high due to the phase similarity with the noise (Figure 5). On the other hand, MI is small when the waveform does not include the template signal, and a distinct peak can be observed compared with the CC. Although CC is relatively large when the noise phases match with those of the template event, the MI remains smaller than CC (Figures 6, 7). MICC has the most-distinct peak corresponding to the signal, and allows the detection of DLF earthquakes at low SN ratio. We therefore suggest that MICC as a proper and sensitive index to detect DLF events.

Figure 5 1.25 Hz sine wave used as noise combined with template waveforms at SN ratios of (a) 1.0, (b) 2.0, and (c) 3.0. The values of MI, CC, and MICC are calculated at each time. Red dots show the index corresponding to the template signal.
Figure 6 Scatter plots and relationship between the SN ratio of the waveform and the MI, CC, and MICC for the synthetic waveform with sinusoidal noise. The figures are as in Figure 4, but here the noise is a sinusoidal wave of frequency 1.25 Hz.
Figure 7 Scatter plot showing the correspondence between the sinusoidal wave noise and the template waveform, using the data from 12 s before Figures 5 and 6.
Application to field data from DLF earthquakes at Kirishima volcano

Here we apply the new method to field data. We selected 200 waveforms of DLF earthquakes as templates from the JMA catalog from 2004 to 2015 and tested continuous data observed in December 2010 at Kirishima volcano. Only two DLF earthquakes were observed in the catalog of JMA in the month. The waveform data are band-pass filtered at 1–8 Hz. For detection using a single station, we focus on three-component seismograms retrieved at the N.SUKH station (see Figure 2a), which has the highest SN ratio for DLF earthquakes at Kirishima volcano.

First, to determine the threshold value, we examine histograms of MI, CC, and MICC obtained from data for 21 December 2010, when intense DLF earthquakes occurred. An earthquake is counted as occurring when the index value is larger than the threshold. The time series of MI, CC, and MICC vibrate in the same frequency bands of the template waveform. The distribution of maximum values for each time series (Figure 8) is found by simply extracting the value larger than the values occurring both before and after. The distribution of each index differed among the three templates.

Conventional matched filter based on summed CC often has the threshold value determined based on the median absolute deviation (e.g., Shelly et al., 2007, Peng and Zhao, 2009). However, it cannot be used for MI because MI is always positive and the shape of the distribution is different from the approximately Gaussian distribution of CC. We set here the threshold to be 0.35 for MICC, based on visual inspection of waveforms and the histogram distributions of the maximum value of each index (Figure 8). The thresholds for comparative analysis using only MI and CC are 0.45 and 0.85, respectively. A DLF earthquake is detected when the index of any of the three components exceeds the threshold. Some earthquakes are detectable more than once by the different template earthquakes. To prevent double counting of any earthquake, we select only one event with the highest index in any 10 s time window and neglect the other events. After compiling the detection catalogs, results corresponding to the daily artificial signal at 9:00:00 (JST) in the Hi-net data are removed.
Figure 8 Histograms of the maxima of three indices (MI, CC, and MICC) extracted from each time series. Vertical axes correspond to the number of samples in each bin (width, 0.01). The time at the top of each panel is the time of occurrence of the template event (JST). Applied data are for one day, 21 December 2010. Red lines show the threshold value for each index.

Figure 9 gives the calculated time variations of MI, CC, and MICC for each time step with a sampling rate of 25 Hz. MICC has very distinct detection peaks, while CC always vibrates with a large variance. In this data set, the summed CC in the three components of six stations does not show distinct peaks corresponding to some DLF earthquakes detected using the proposed single-station method. For example, the events at 05:48 on December 21 were detected in the single-station method, however, summed CC does not have the peak corresponding to the events. Then, there are some cases in which no peaks are found in the summed CC because the amplitude is small or the hypocentral location is slightly different from those of the template event, suggesting that the single-station method is suitable for comprehensive detection. The comprehensiveness of the method can be also seen from the fact that the peaks of MICC
correspond to large amplitudes in the velocity waveform except some parts corresponding
to the signals of earthquakes occurring in other regions.

MICC identified 354 DLF earthquakes in the test month. Even using data from a
single station achieved a temporal change of the cumulative number of DLF events
similar to that observed from the catalog of events compiled from six stations’ data
(Figure 10). The catalogs include step-like increases in the number of DLF earthquakes
on December 5 and December 12, and the difference between the catalogs over the
month is small. This trend is generally consistent across the results from any of the three
indices. As many of the earthquakes detected by these indices have very small SN
ratios, it is not possible to distinguish between a true signal and a false detection even
by visual inspection of the waveforms. Although the MICC peak is the most distinct and
the detection looks good, the accuracies of the detection of the three single-station
indices are quantitatively evaluated in discussion.
Figure 9 Values of MI, CC, and MICC from 0:00 to 12:00 on December 21, 2010. The lowest panel shows the velocity waveform. The template earthquake is the earthquake shown in Figure 1. Red lines show the threshold of detection. Green triangles in each panel show the detected events by the template using the index. White
triangles show the detection using other templates.

Figure 10 Cumulative number of detected DLF earthquakes based on each index during December 2010. From the top, the indices are summed CC of the six stations and MI, CC, and MICC of the N.SUKH station.
To confirm which index performs best, we compared the catalog obtained using each of three indices with one based on conventional matched filter method using summed CC (e.g. Gibbons and Ringdal, 2006; Kato and Nakagawa, 2020; Kurihara et al., 2019; Shapiro et al., 2017; Shelly et al., 2007; Yukutake et al., 2019). Using a threshold of 5.0 for conventional detection using the summed CC of three components of the six seismic stations detected 312 events throughout December 2010. We assume that the conventional method provides a true catalog against which the performance of other methods can be compared using the threat score (also called the critical success index), which is often used in the weather forecasting (Japan Meteorological Agency 2019). The score is defined as follows:

\[
\text{Threat score} = \frac{TP}{TP+FP+FN} \quad \cdots (14),
\]

where TP, FP, and FN are the numbers of true positive, false positive, and false negative events, respectively. TP events are those detected in the catalogs from both the conventional multiple-station method and the single-station method. FP events are those detected by the single-station method but not the conventional method. FN events are those detected conventionally but not by the single-station method. The threat score is 0.366 for MI, 0.261 for CC, and 0.441 for MICC applied to single-station data. As MICC gives the highest threat score, it appears to give the best performance among the three indices.

Next assessed is the effect of changing the detection threshold for each catalog. The threat score is maximized at 0.393 for the MI catalog using a threshold of 0.430, at 0.354 for CC with a threshold of 0.800, and at 0.441 for MICC with a threshold of 0.350 (Figure 11). The threat score for the MICC catalog remains similar within a broad range of threshold values of 0.3 to 0.4, indicating that the same accuracy can be obtained regardless of which threshold is selected within this range.
Figure 11 The threat score and the number of TP, FP, and FN events.

The upper panels show the threat score for three indices. Horizontal axis corresponding to the threshold values. Red, blue, and green circles in the lower panels show the number of TP, FN, and FP events for each threshold value.

We compare the maximum value of summed CC in the 5 s before and after the time with that of each index at the single station (Figure 12). The summed CC is positively correlated with each of MI, CC, and MICC applied to a single station’s data. Especially, MICC shows the best positive relationship with the summed CC.

Figure 12 Relations between summed CC (horizontal axis) and three indices (vertical axis) of a component of the data from N.SUKH station. Black dots correspond to each event. Red broken lines show the threshold of detection.

The threat score for the MICC catalog is the highest among the three catalogs; however, the value seems low. We next assess the quality of the catalog compiled using MICC. Of the 88 events detected with summed CC over 6.0, 80 are also included in the MICC catalog. On the other hand, of the 354 events detected using MICC, 85 % of the events (302 events) are also in the catalog of summed CC with the threshold of 4.0. In other words, the FP and FN events included those which are around the threshold value,
and the threat score of 0.441 does not mean that only 44.1% of DLF earthquakes are
detected. The single-station catalog probably includes DLF earthquakes with very small
magnitudes that are only observed at the single station (N.SUKH); however, this
comparison cannot distinguish FP detections from actual small DLF earthquakes.

This study calculates MI by dividing the points of velocity-seismograms into
5 × 5 cells. The number of divisions was determined by evaluating the clarity of the peaks
corresponding to the signal. Comparing calculations for each number of divisions for the
continuous waveform showed that odd-numbered divisions such as 3 × 3 and 5 × 5 give
sharper peaks than even-numbered division such as 4 × 4 when the number of divisions
is small (Figure 13). This is because the points near the origin in even-numbered divisions.
belong to different cells due to slight differences of the seismograms. This leads to a
decrease in the MI when considering the distribution diagram as shown in Figure 4. As
3 × 3 division has a large variance for parts not including signals of DLF earthquakes,
5 × 5 is considered suitable for detection in the three data sets. On the other hand, as the
number of divisions is increased, the baseline of MI rises and the peaks become generally
less sharp (Figure 13) because there are too many cells relative to the number of the points.
Therefore, for our dataset with 25 Hz sampling and an 8 s time window, 5 × 5 division is
optimal.

In this study, we calculate MI by dividing the points of velocity-seismograms into the
5 × 5 cells. The number of divisions was determined by evaluating the clarity of the peaks
for the continuous waveform, the peak of the odd-numbered division such as 3 × 3 and 5
× 5 is sharper than that of the even-numbered division such as 4 × 4 when the number of
divisions is small (Figure 13). This is because the points near the origin belong to different
cells due to slight differences of the seismograms. It leads to a decrease in the MI when
considering the distribution diagram as shown in Figure 4. In addition, 3 × 3 has a large
variance for the parts which do not include signals of DLF earthquakes, therefore, 5 × 5
is suitable for the detection in the three data sets. On the other hand, as the number of
divisions is increased, the baseline of MI rises and the peaks become duller in general
(Figure 13) because the number of cells is too much compared to the number of the points.
Therefore, for our dataset with 25 Hz sampling and an 8-second time window, we
determined that a 5 × 5 split was optimal.
Using MICC with a single station’s data is a potentially powerful tool, especially for monitoring shallow volcanic earthquakes occurring beneath the crater because there cannot be many seismic stations in that region due to the risk and low accessibility. In addition, the single-station method will improve the completeness of small-magnitude volcanic seismicity studies, deepening understanding about volcanic activity beneath the crater. As in previous cases of single-station analysis (Vuan et al. 2018; Wech et al. 2020), this method will also contribute to the monitoring of other low-SN ratio earthquakes, such as DLF earthquakes and swarm earthquakes that can be observed only with part of an observation network.

Figure 13 Time series of MI using 10 different divisions. The time window is one hour from 00:00:00, 21 December 2010 (JST). The template earthquake and observation station are as in Figure 9.
Conclusions

This study developed a new matched filter method of earthquake detection using MICC applied to a single station’s data. Tests using synthetic waveforms revealed that using MICC gave more-distinct peaks than MI or CC. Applying this method to DLF earthquake data from Kirishima volcano in December 2010 detected 354 DLF earthquakes. Comparison with conventional matched filter applied to multiple stations’ data showed the large detection accuracy of the proposed method. Overall, the proposed single-station matched filter technique could be useful in various regions where observations from multiple stations are not possible, as it can detect microearthquakes using only a small number of stations and templates.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

List of abbreviations

MFT: Matched filter technique
DLF: Deep Low-Frequency
CC: Correlation coefficients
MI: Mutual information
MICC: Product of mutual information and correlation coefficients
Hi-net: High-sensitivity seismograph network
NIED: National Research Institute for Earth Science and Disaster Resilience
JMA: Japan meteorological agency
MAD: Median absolute deviation
TP: True positive
FP: False positive
Availability of data and materials
We used the Hi-net seismic observation data from NIED (National Research Institute for Earth Science and Disaster Resilience 2019) and JMA’s unified earthquake catalog. Those are available from the web page of NIED Hi-net (http://www.hinet.bosai.go.jp).

Competing interests
The authors declare that they have no competing interests.

Funding
This work was supported by Japan Science and Technology Agency (JST) CREST Grant Number JPMJCR1763, Japan.

Authors' contributions
RK mainly design and analyze the data and write the paper. AK advises and discusses the contents of this paper. HN and SK advise about indices in statics and discuss the paper.

Acknowledgments
We used Generic Mapping Tools for drawing figures (Wessel and Smith 1998) and collated the Hi-net seismic observation data (http://www.hinet.bosai.go.jp) from NIED (National Research Institute for Earth Science and Disaster Resilience 2019). We used JMA’s unified earthquake catalog (http://www.jma.go.jp) and the computer systems of the Earthquake and Volcano Information Center of the Earthquake Research Institute, the University of Tokyo. This work was supported by the JST CREST (grant number JPMJCR1763) and partially supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, under its The Second Earthquake and
Volcano Hazards Observation and Research Program (Earthquake and Volcano Hazard Reduction Research).

References

Gao D, Kao H (2020) Optimization of the match-filtering method for robust repeating earthquake detection: The multi segment cross-correlation approach. J Geophys Res: Solid Earth 125(7):1–19. https://doi.org/10.1029/2020JB019714

Gibbons SJ, Ringdal F (2006) The detection of low magnitude seismic events using array-based waveform correlation. Geophys J Int 165(1):149–166. https://doi.org/10.1111/j.1365-246X.2006.02865.x

Japan Meteorological Agency (2019) Outline of the operational numerical weather prediction of the Japan Meteorological Agency. Japan Meteorological Agency. Retrieved from http://www.jma.go.jp/jma/jma-eng/jma-center/nwp/outline2013-nwp/pdf/outline2013_03.pdf

Kato A, Nakagawa S (2020) Detection of deep low-frequency earthquakes in the Nankai subduction zone over 11 years using a matched filter technique. Earth Planets Space 72(1):128. https://doi.org/10.1186/s40623-020-01257-4

Katsumata A, Kamaya N (2003) Low-frequency continuous tremor around the Moho discontinuity away from volcanoes in the southwest Japan. Geophys Res Lett 30(1):20-1–204. https://doi.org/10.1029/2002GL015981

Kurihara R, Obara K, Takeo A, Tanaka Y (2019) Deep low-frequency earthquakes associated with the eruptions of Shinmoe-dake in Kirishima volcanoes. J Geophys Res: Solid Earth 124(12):13079–13095. https://doi.org/10.1029/2019JB018032

National Research Institute for Earth Science and Disaster Resilience (2019) NIED Hi-net. National Research Institute for Earth Science and Disaster Resilience. https://doi.org/10.17598/nied.0003

Okada Y, Kasahara K, Hori S, Obara K, Sekiguchi S, Fujiwara H, Yamamoto A (2004) Recent progress of seismic observation networks in Japan —Hi-net, F-net, K-NET and KiK-net—. Earth Planets Space 56(8):xv–xxviii. https://doi.org/10.1186/BF03353076

Peng Z, Zhao P (2009) Migration of early aftershocks following the 2004 Parkfield earthquake. Nat Geosci 2:877–881. https://doi.org/10.1038/ngeo697

Pluim JPW, Maintz JBA, Viergever MA (2003) Mutual-information-based registration
of medical images: a survey. IEEE Trans Med Imaging 22(8):986–1004.
https://doi.org/10.1109/TMI.2003.815867

Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, et al.
(2011) Detecting novel associations in large data sets. Science 334(6062):1518–
1524. https://doi.org/10.1126/science.1205438

Reshef YA, Reshef DN, Finucane HK, Sabeti PC, Mitzenmacher M (2016) Measuring
dependence powerfully and equitably. J Mach Learn Res 17:1–63.

Sagar RP, Guevara NL (2005) Mutual information and correlation measures in atomic
systems. J Chem Phys 123(4):044108. https://doi.org/10.1063/1.1953327

Shapiro NM, Droznin DV, Droznina SY, Senyukov SL, Gusev AA, Gordeev EI (2017)
Deep and shallow long-period volcanic seismicity linked by fluid-pressure transfer.

Ueno H, Hatakeyama S, Aketagawa T, Funasaki J, Hamada N (2002). Improvement of
hypocenter determination procedures in the Japan Meteorology Agency. Quar J
Seismol 65:123–131. Retrieved from
https://www.jma.go.jp/jma/kishou/books/kenshin/vol65p123.pdf

Vuan A, Sugan M, Amati G, Kato A (2018) Improving the detection of low-magnitude
seismicity preceding the Mw 6.3 L’Aquila Earthquake: Development of a scalable
code based on the cross correlation of template earthquakes. Bull Seismol Soc Am
108(1):471–480. https://doi.org/10.1785/0120170106

Wech AG, Thelen WA, Thomas AM (2020) Deep long-period earthquakes generated by
second boiling beneath Mauna Kea volcano. Science 368(6492):775–779.
https://doi.org/10.1126/science.aba4798

Wessel P, Smith WHF (1998) New, improved version of generic mapping tools
released. Eos, Trans Am Geophys Union 79(47):579–579.
https://doi.org/10.1029/98EO00426

Yukutake Y, Abe Y, Doke R (2019) Deep low-frequency earthquakes beneath the
Hakone Volcano, Central Japan, and their relation to volcanic activity. Geophys
Res Lett 2019GL084357. https://doi.org/10.1029/2019GL084357

Zhang P (2015) Evaluating accuracy of community detection using the relative
normalized mutual information. J Stat Mech: Theory Exp 2015(11):P11006.
Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- graphicalabst.jpg