

FULL PAPER

Open Access



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

Ryo Kurihara<sup>1\*</sup> , Aitaro Kato<sup>1</sup>, Sumito Kurata<sup>2</sup> and Hiromichi Nagao<sup>1</sup>

## Abstract

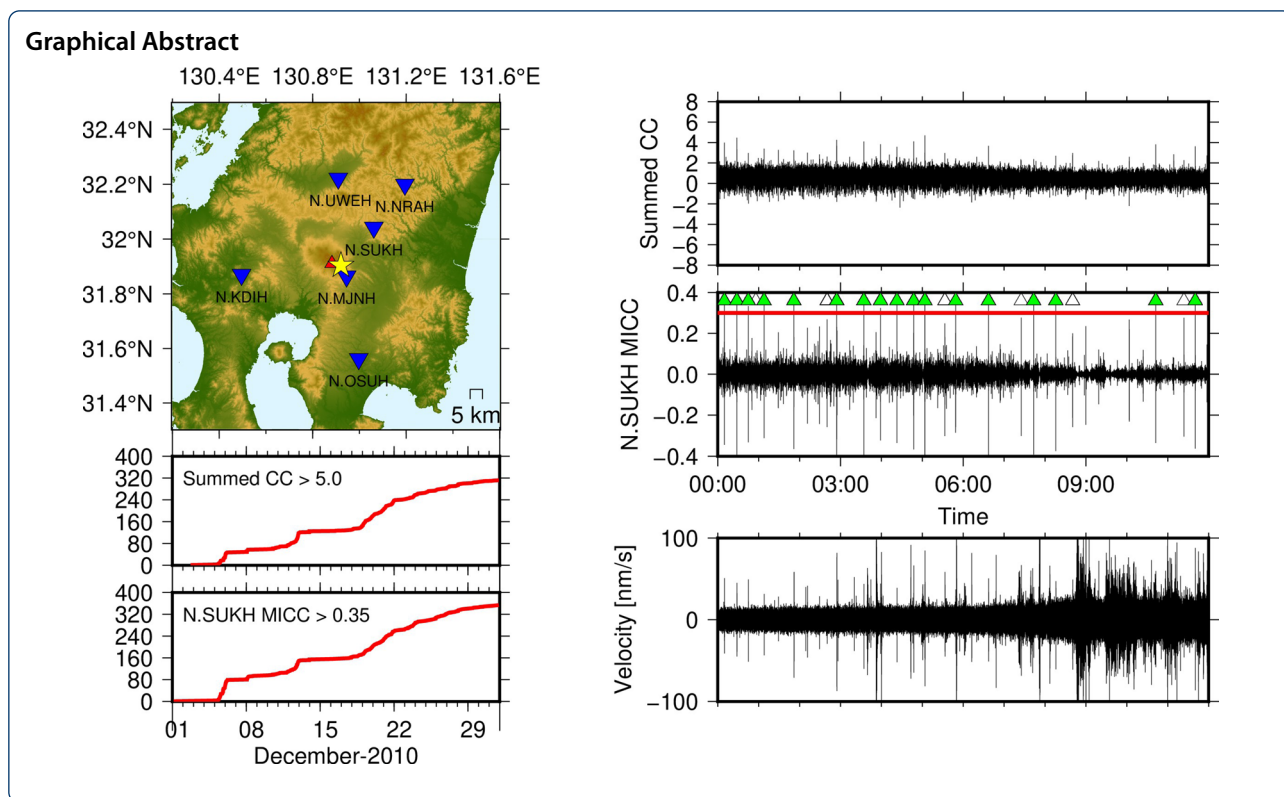
The matched filter technique is often used to detect microearthquakes such as deep low-frequency (DLF) earthquakes. It compares correlation coefficients (CC) between waveforms of template earthquakes and the observed data. Conventionally, the sum of CC at multiple seismic stations is used as an index to detect the DLF earthquakes. A major disadvantage of the conventional method is drastically reduced detection accuracy when there are too few seismic stations. The 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 events from the data in December 2010, whereas the catalog of the Japan Meteorological Agency shows only two. Of the detected events, 314 (89%) are likely DLF earthquakes and other events may be false detections. Most of the false detections correspond to surface-wave arrivals from teleseismic events. The catalog of DLF earthquakes constructed here shows similar temporal behavior to that found by the conventional matched filter method using the sum of the CC of the six stations near the volcano. 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

\*Correspondence: rkuri@eri.u-tokyo.ac.jp

<sup>1</sup> Earthquake Research Institute, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-Ku, Tokyo 113-0032, Japan

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



**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 detected seismic waveforms—and target waveforms (Gibbons and Ringdal 2006). It can detect microearthquakes such as aftershocks (Peng and Zhao 2009; Kato and Obara 2014), low-frequency earthquakes and tremor in the plate subduction zone and volcanic regions (Shelly et al. 2007; Shapiro et al. 2017; Yukutake et al. 2019; Kurihara et al. 2019; Kato and Nakagawa 2020; Kurihara and Obara 2021). The matched filter method is especially useful in situations of continued intense seismic activity, such as aftershocks and seismic swarms, because it is difficult to detect these earthquakes by either visual detection or using conventional methods such as the STA/LTA method. This method has been used to detect many low-frequency earthquakes; the resulting earthquake catalogs provide precise constraints on the spatial and temporal evolution of this intense seismicity. 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 (e.g., Shelly et al. 2007; Shapiro et al. 2017;

Yukutake et al. 2019; Kurihara et al. 2019; Kato and Nakagawa 2020; Kurihara and Obara 2021).

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 matched filter technique usually employs template waveforms with 4- to 5-s time windows for waveforms that have been filtered at either 1–4 or 2–8 Hz to detect low-frequency earthquakes. 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, the quality of a seismic catalog that is constructed using the conventional matched filter technique decreases where there is either an inadequate number of observation stations or a well-determined seismic catalog is lacking. Single-station

matched filter method is potentially 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 large-amplitude 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 into smaller sub-segments, which could effectively distinguish seismic waves generated from different epicenters and focal mechanisms. 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 and unclear P-wave arrivals. Therefore, we tried to improve detection accuracy 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. A similar approach using  $CC|CC|$  (i.e., the product of CC and the absolute value of CC) has been reported recently by Gibbons (2021).

### Data and method

This study considers the waveforms of DLF earthquakes in Kirishima volcano, one of the most active volcanoes in Japan. The observed DLF earthquakes that occurred at 25 km depth during the 2-year period from December 2009 (Kurihara et al. 2019), which includes the January 2011 subplinian eruptions, were activated. Waveform

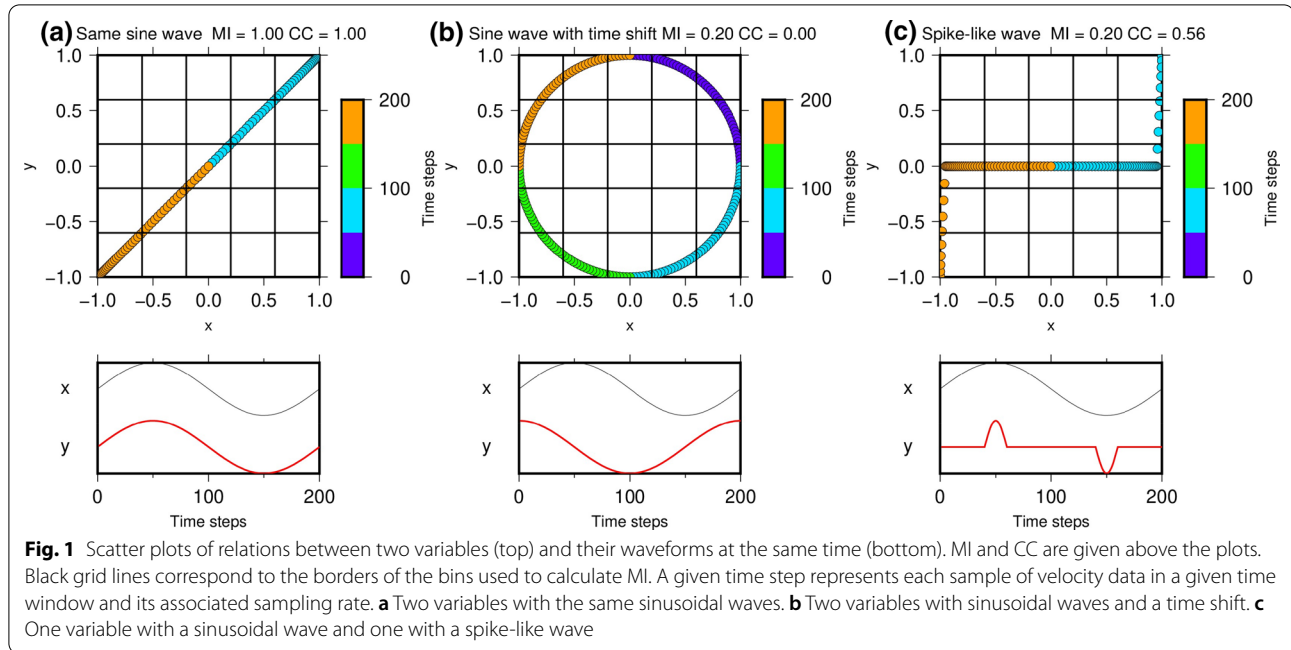
data are from the high-sensitivity seismograph network (Hi-net) of NIED (Okada et al. 2004; National Research Institute for Earth Science and Disaster Resilience 2019). We apply a 1–8 Hz bandpass filter, and decimate the waveform from 100 to 25 Hz sampling before calculations.

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 Meteorological Agency (JMA) (e.g., Katsumata and Kamaya 2003). The CC of component  $j$  at seismic station  $i$  is calculated as:

$$CC_{ij}(t_{tg} + \Delta t_i) = \frac{\sum_{\tau} \left( v_{ij}^{tp}(t_{tp} + \Delta t_i + \tau) v_{ij}^{tg}(t_{tg} + \Delta t_i + \tau) \right)}{\sqrt{\sum_{\tau} \left( v_{ij}^{tp^2}(t_{tp} + \Delta t_i + \tau) \right)} \sqrt{\sum_{\tau} \left( v_{ij}^{tg^2}(t_{tg} + \Delta t_i + \tau) \right)}}, \quad (1)$$

where  $v_{ij}^{tp}(t)$  and  $v_{ij}^{tg}(t)$  are the velocities of the template earthquake and target data at time  $t$  of component  $j$  at station  $i$ , respectively;  $t_{tp}$  is the occurrence time of the template earthquake;  $t_{tg}$  is the time of the target event, which is determined by the search method;  $\Delta t_i$  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. The mean of the template and data in each time window should be removed during the calculation of the true statistical correlation coefficient. However, we used the bandpass-filtered waveform data after removing its mean and trend. Therefore, the mean of the template and data in the target window is very small (0–2% of the maximum value for most of the template waveforms in this study), such that we can ignore the mean to calculate CC in this study. This assumption does not affect the detection capability (e.g., Beaucé et al. 2018). 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 origin 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:

$$\text{Summed } CC(t_{tg}) = \sum_i \sum_j CC_{ij}(t_{tg} + \Delta t_i). \quad (2)$$



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 and the two events have similar source properties, the stacked peak values yield a large summed CC.

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 Fig. 1. MI is calculated by dividing the points of the velocity seismograms into discrete cells in the  $x, y$  domain. Here, we selected  $5 \times 5$  cells for the division process; the details of this division selection are presented in the discussion section. 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:

$$\bar{v}_{tp}(t) = \frac{v_{tp}(t)}{\max(|v_{tp}(t)|)}, \quad (3)$$

$$\bar{v}_{tg}(t) = \frac{v_{tg}(t)}{\max(|v_{tg}(t)|)}, \quad (4)$$

$$n_{tp}(t) = \lfloor ((\bar{v}_{tp}(t) + 1.4) * 2.5) \rfloor \text{ (when } \bar{v}_{tp}(t) < 1), \quad (5)$$

$$n_{tg}(t) = \lfloor ((\bar{v}_{tg}(t) + 1.4) * 2.5) \rfloor \text{ (when } \bar{v}_{tg}(t) < 1) \quad (6)$$

$$n_{tp}(t) = 5 \text{ (when } \bar{v}_{tp}(t) = 1), \quad (7)$$

$$n_{tg}(t) = 5 \text{ (when } \bar{v}_{tg}(t) = 1). \quad (8)$$

The notation  $\lfloor \rfloor$  means the absolute function, and  $\lfloor \rfloor$  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  $1.0$ .

Using the above integers, the MI is calculated as:

$$MI(t) = \sum_{n_{tp}=1}^5 \sum_{n_{tg}=1}^5 p(n_{tp}, n_{tg}) \log \frac{p(n_{tp}, n_{tg})}{p(n_{tp})p(n_{tg})}, \quad (9)$$

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  $p(n_{tp})$ ,  $p(n_{tg})$ , and  $p(n_{tp}, n_{tg})$  by dividing the number of points in the bins by 200.

The upper limit of MI depends on the data set according to Eq. (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):



$$h_{tp} = - \sum p(n_{tp}) \log p(n_{tp}), \tag{10}$$

$$h_{tg} = - \sum p(n_{tg}) \log p(n_{tg}), \tag{11}$$

$$\overline{MI} = \frac{2.0 * MI}{h_{tp} + h_{tg}} \tag{12}$$

Hereafter, we simply describe  $\overline{MI}$  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 (Fig. 1a). While if the relation is non-linear, MI has a non-zero value (Fig. 1b). According to Eq. (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 (Fig. 1c).

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

$$MICC(t) = MI(t)CC(t). \tag{13}$$

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

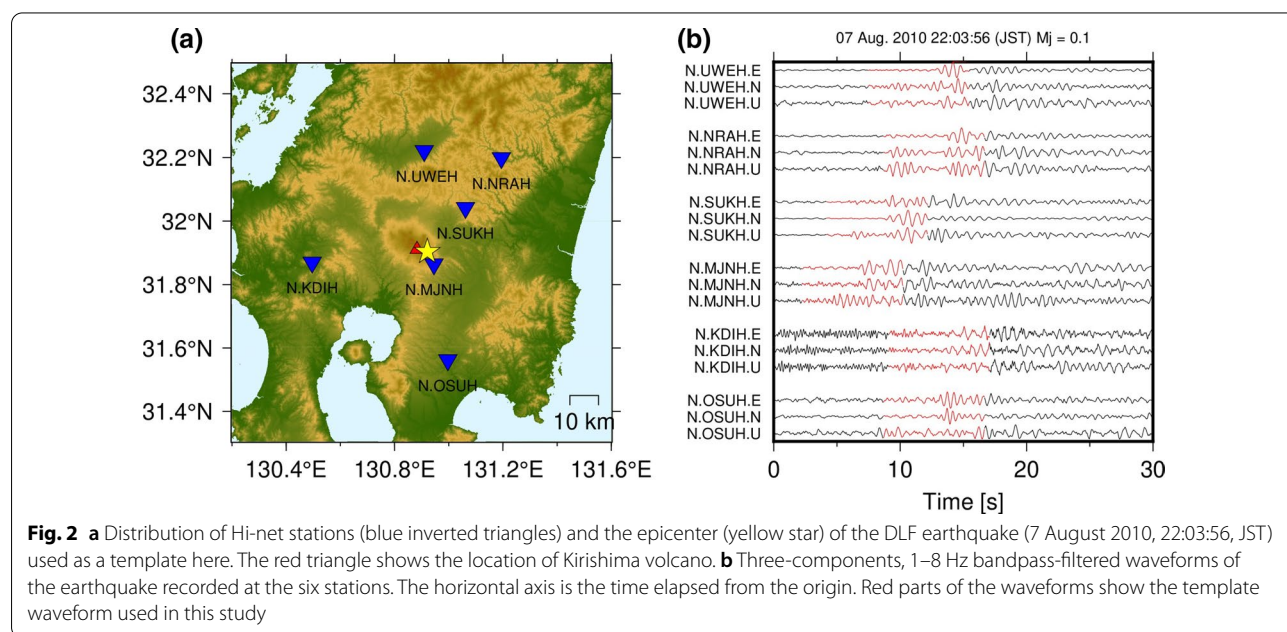
DLF earthquakes. The analysis uses a synthetic data set comprising three types of noise (Gaussian, sinusoidal and random phase noise) added to a template waveform of data from a DLF earthquake recorded by station N.SUKH at 22:03:56 (JST) on 7 August 2010 (Fig. 2).

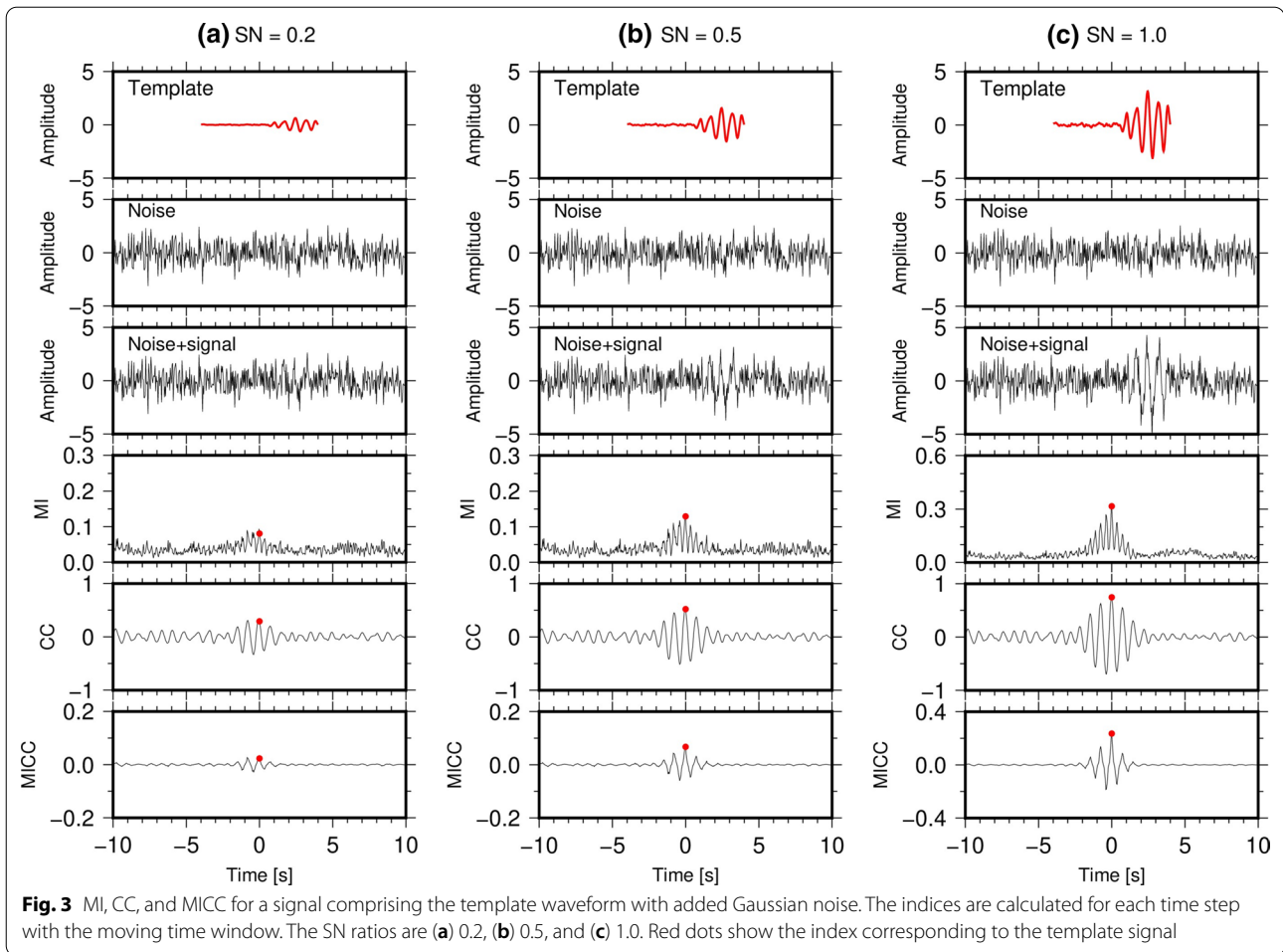
### Gaussian noise

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 station N.SUKH to Gaussian noise with a variance of one. 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). For high SN ratio examples, event detection is possible if the appropriate threshold is set, since the signal possesses larger MI, CC, and MICC indices than the surrounding noise. However, it is difficult to set such an appropriate threshold, since the range above the noise level and below the peak level, which corresponds to the target signal, is extremely narrow for CC. In the case of Gaussian noise, DLF events are detectable using any of the indices when the SN ratio is over 0.5.

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 (Fig. 4a, b). The relationship becomes linear at high SN ratio (Fig. 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 (Fig. 4c). MICC gradually increases in any range of SN ratio corresponding to an increase of MI and CC.

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**

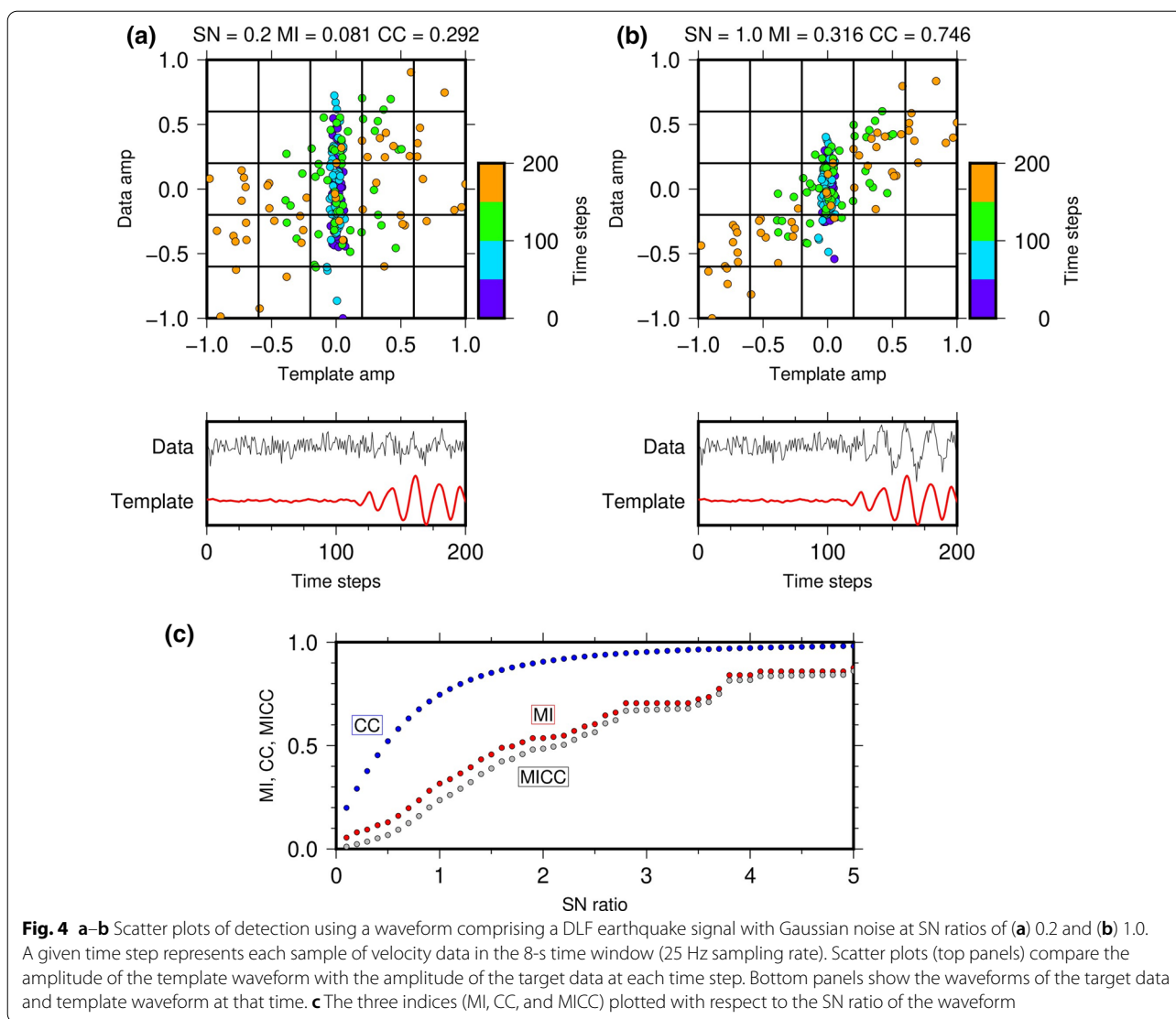
As in the case of Gaussian noise, the test considers a template waveform with added noise, this time a sinusoidal wave of 1.25 Hz frequency. The variance of the noise is 1. The variance of CC tends to be always high due to the phase similarity with the noise (Fig. 5). On the other hand, MI is small when the waveform does not include

the template signal, and a distinct peak can be observed compared with the CC. While CC is relatively large when the noise phases match with those of the template event, the MI remains smaller than CC (Figs. 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.

**Random phase noise**

We next consider random phase noise as more realistic noise. Random phase noise was produced following the method proposed by Chamberlain et al. (2014). We first obtained one day of continuous data (N–S component) at station N.SUKH on 21 December 2010, when many DLF earthquakes were observed. These one-day data were converted from the time domain to the frequency domain by Fourier transform.

We then changed the phase randomly in the frequency domain. Finally, we produced random phase noise by an



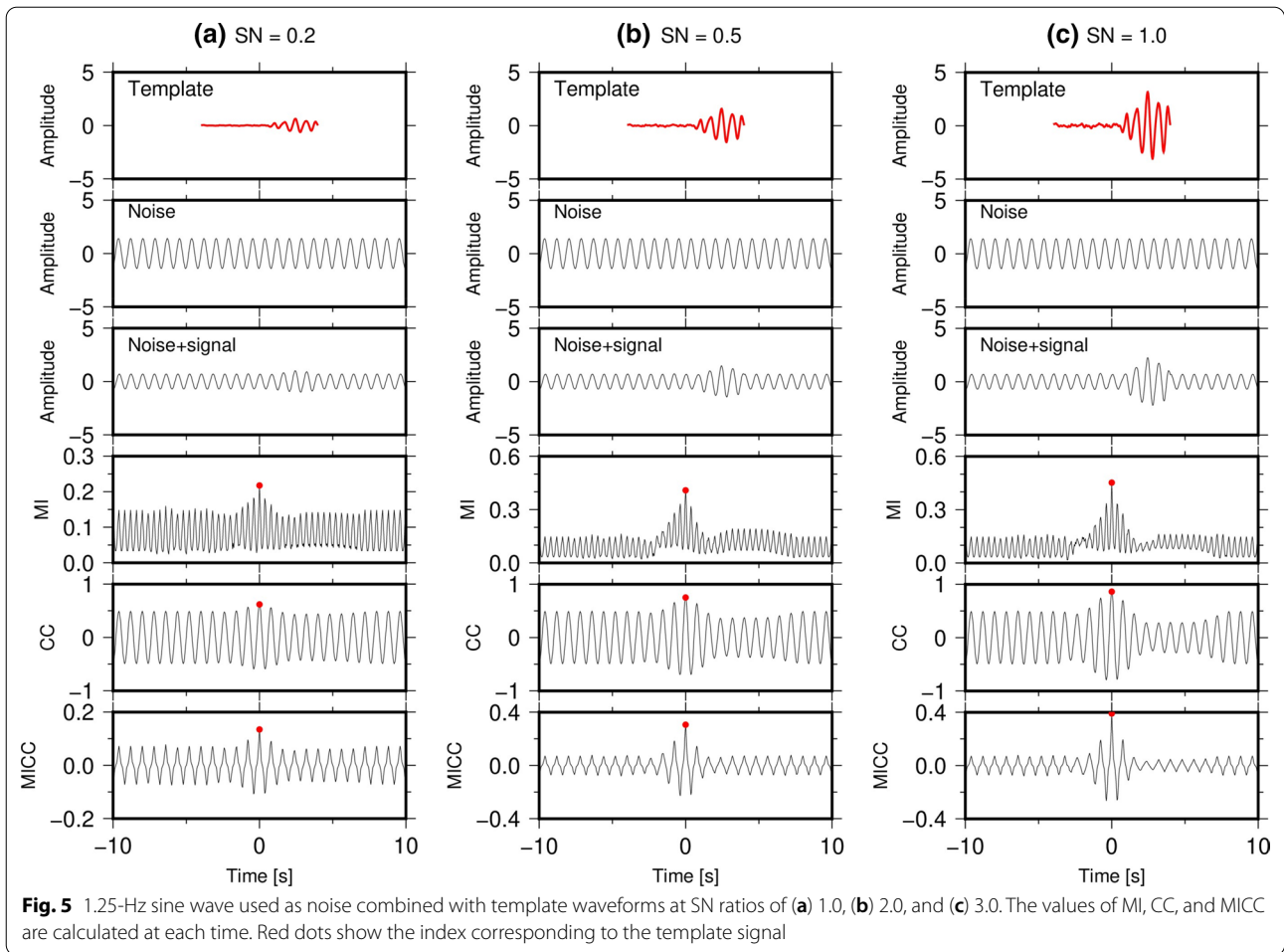
**Fig. 4** a–b Scatter plots of detection using a waveform comprising a DLF earthquake signal with Gaussian noise at SN ratios of (a) 0.2 and (b) 1.0. A given time step represents each sample of velocity data in the 8-s time window (25 Hz sampling rate). Scatter plots (top panels) compare the amplitude 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

inverse Fourier transform of the spectrum data with the randomized phase. We generated a synthetic waveform after adding template waveforms to the random noise every 400 s with various SN ratios.

We calculated MI, CC, and MICC between the template and synthetic waveforms (Fig. 8). MI and MICC show a very distinct peak that corresponds to the added template waveform when the SN ratio is above 1.0. Although we can detect DLF earthquakes using every index with the appropriate threshold, MICC shows the most distinct peak, which allows us to easily select the appropriate threshold.

### Application to field data from DLF earthquakes at Kirishima volcano

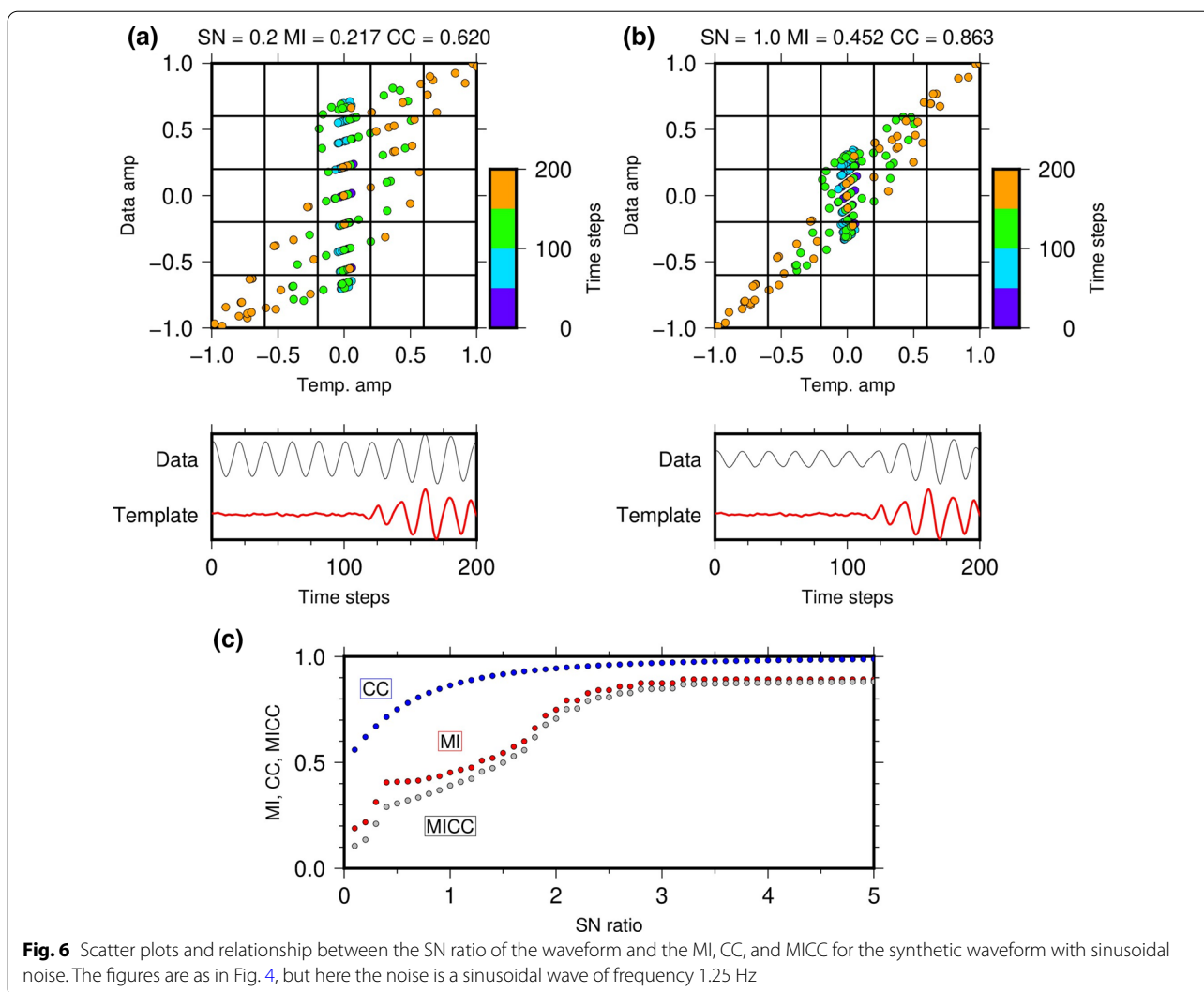
Here, we apply the new method to field data. We selected 200 waveforms of DLF earthquakes as templates from the JMA catalog from 2004 to 2015 and tested continuous data observed in December 2010 at Kirishima volcano. The DLF earthquakes that occurred during this month were characterized by lower dominant frequencies than those that occurred during other periods (Kurihara et al. 2019). Only two DLF earthquakes were observed in the catalog of JMA in the month. The waveform data are bandpass filtered at 1–8 Hz. For detection



using a single station, we focus on three-component seismograms retrieved at station N.SUKH (see Fig. 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 is found by simply extracting the local maximum value larger than the values before and after. The distribution of local maximum values for each index differed among the templates; three representative examples are shown in Fig. 9. Conventional matched filter based on summed CC often has the threshold value determined based on the median absolute deviation (MAD) (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. Furthermore, the upper limit for both MI and CC is 1.0. If we define the threshold using eight or nine times the MAD, then the threshold values for some templates are above the maximum CC value (i.e., 1.0). Therefore, we use a fixed threshold value in this study. 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 (Fig. 9). The thresholds for a comparative analysis using only MI and CC are similarly set as 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



**Fig. 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 Fig. 4, but here the noise is a sinusoidal wave of frequency 1.25 Hz

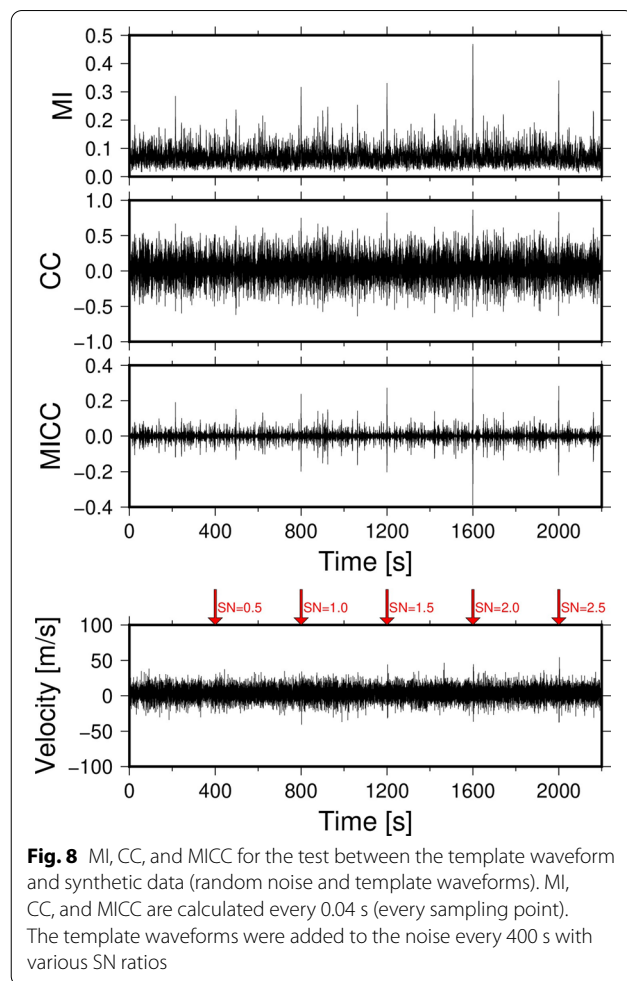
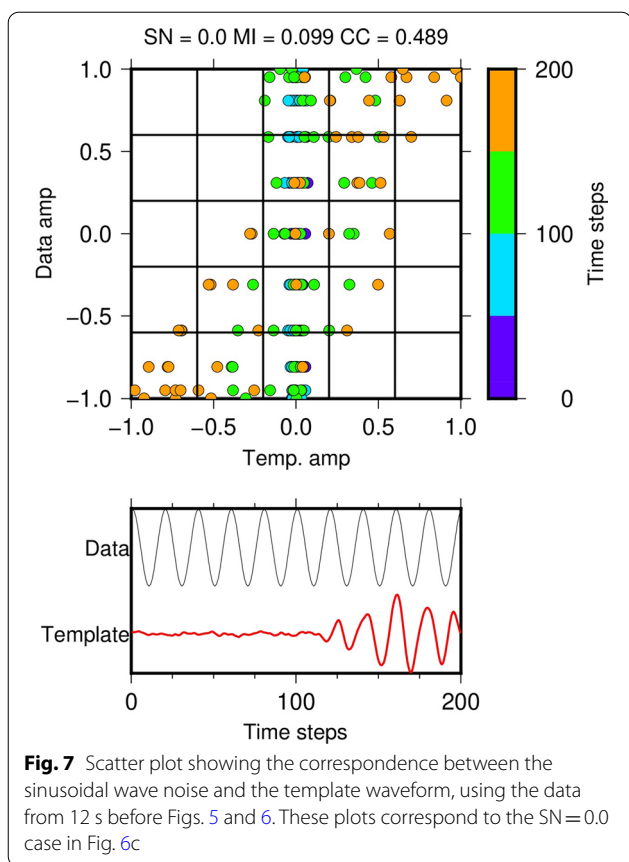
events. After compiling the detection catalogs, results corresponding to the daily artificial signal at 9:00:00 (JST) in the Hi-net data are removed.

Figure 10 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. Furthermore, MICC has a larger SN ratio and more distinct peaks than CC|CC|, which is an index that has been proposed by Gibbons (2021) (Additional file 1: Figure S1). In this data set, the summed CC in the three components of six stations (see Fig. 2) 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 21 December 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 detection without misdetection. 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. Although false detections, such as other earthquakes, may be included, they can be removed from the seismic catalog by applying signal processing that targets these undesired signals.

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





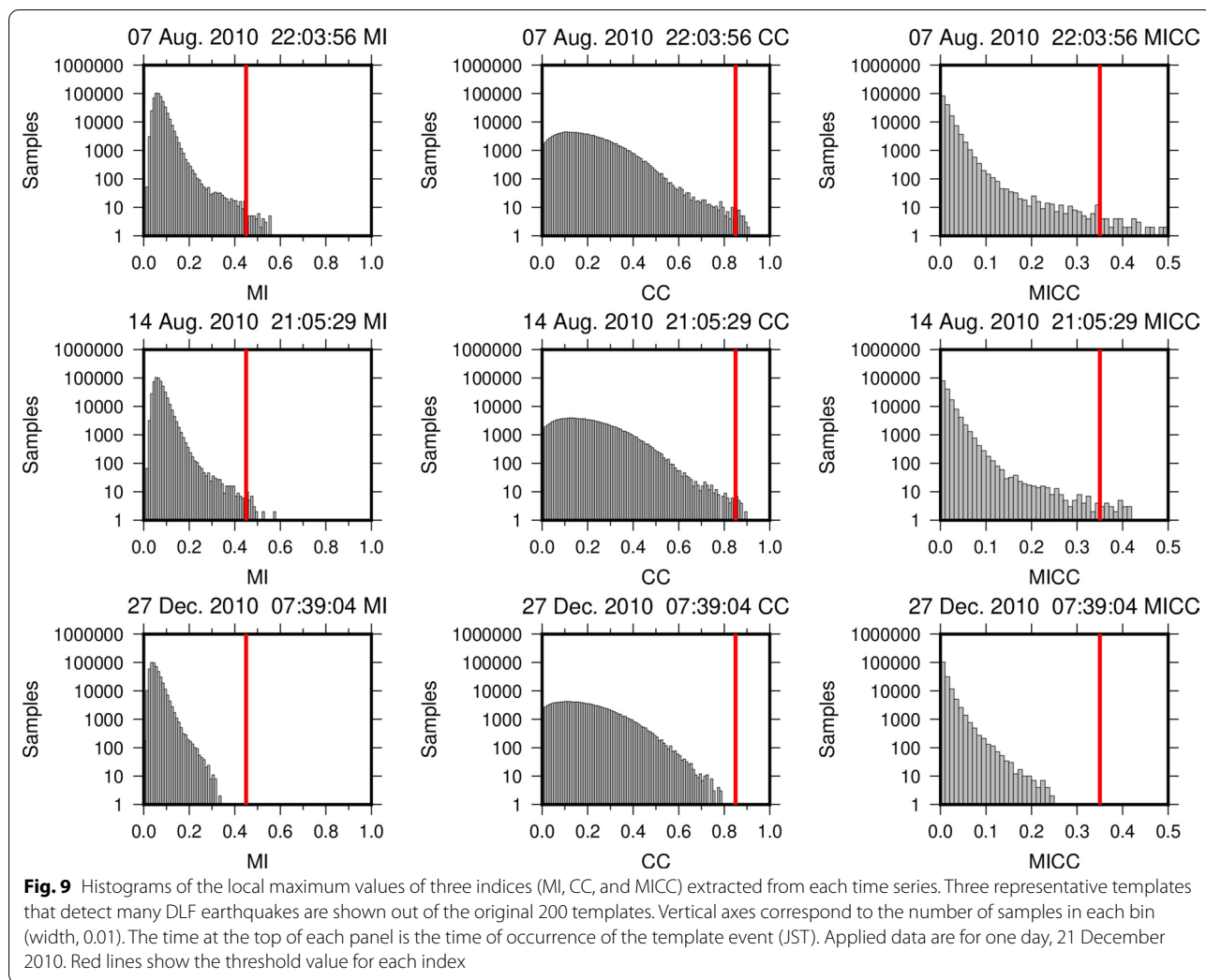
events compiled from six stations' data (Fig. 11). The catalogs include step-like increases in the number of DLF earthquakes on 5 and 12 December, and the difference between the catalogs over the month is small. This trend is generally consistent across the results from any of the three indices. As many 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.

**Discussion**

We first visually inspected the events that were detected by either the summed CC or MICC approaches to assess the accuracy of the detected events. We focused on the ratio of the amplitudes for each component that were recorded at the six stations (Additional file 1: Figures S2–S4). Although some events were classified as misdetections owing to their surface-wave arrivals, most of the events were identified as DLF earthquakes (Table 1). Most of the misdetections due to surface waves could be

automatically removed by calculating the amplitude ratio between the 0–10 s and 10–20 s intervals after the detection onset in the 2–4 Hz frequency band. If the amplitude ratio was smaller than one, then the detection was classified as a misdetection. The 10-s waveform duration appears to be appropriate for this analysis, since most of the DLF earthquakes have durations of < 10 s.

We then confirmed the index with the best statistical performance via a numerical comparison of the catalogs obtained using each of the three indices with one based on the conventional matched filter method using the 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:

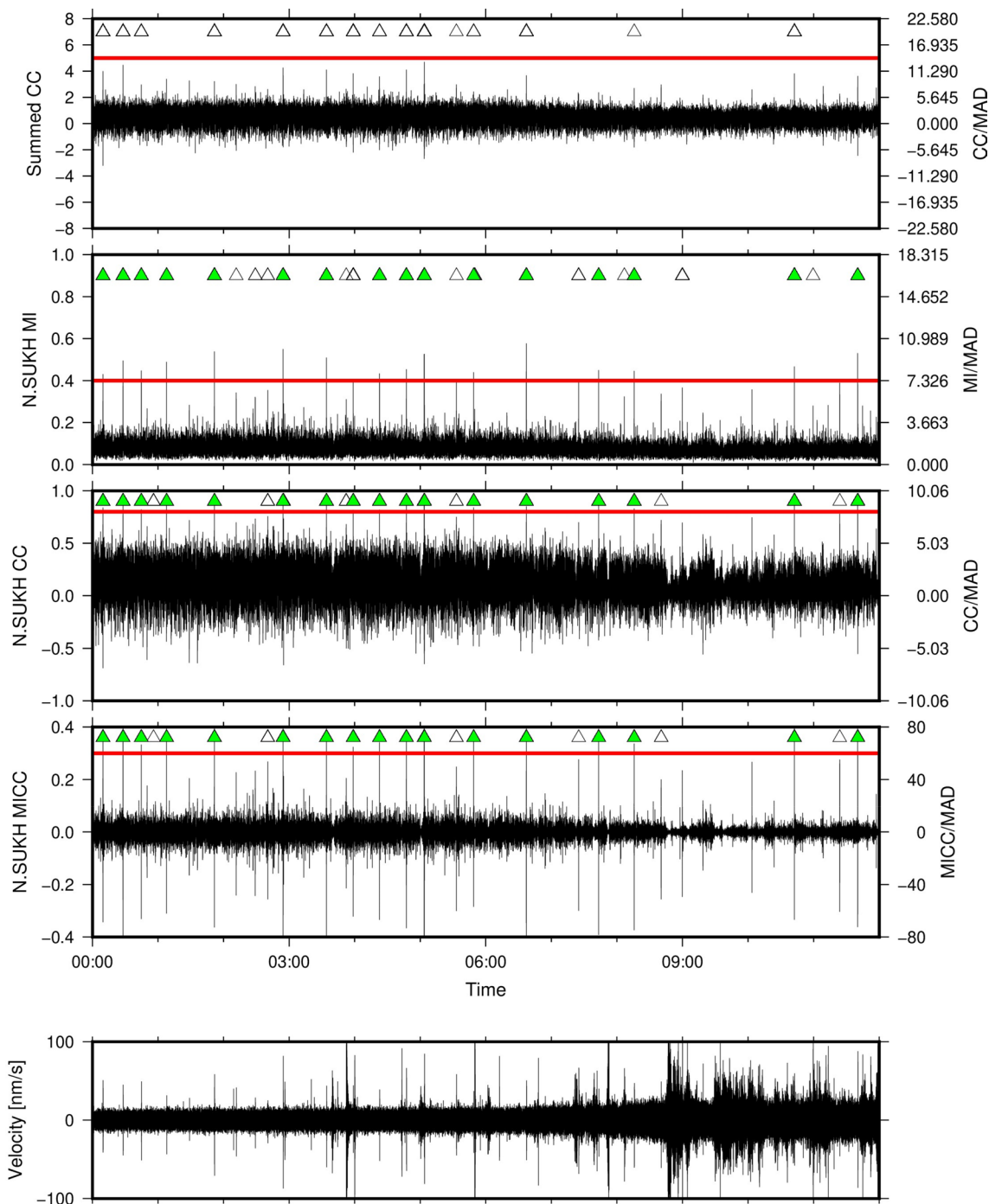
$$\text{Threatscore} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \tag{14}$$

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. To assess the effect of changing the detection threshold for each catalog, we change the threshold and calculated the threat score for each threshold of the three indices. The threat score is maximized

at 0.437 for the MI catalog using a threshold of 0.445, at 0.451 for CC with a threshold of 0.850, and at 0.461 for MICC with a threshold of 0.360 (Fig. 12; Table 2). 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 similar accuracy can be obtained regardless of which threshold is selected within this range.

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 (Fig. 13). The summed CC is positively correlated with each of MI, CC, and MICC applied to a single station’s data. In particular, MICC shows the best positive relationship with the summed CC, whereas large CC values are limited because the maximum value is one.

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

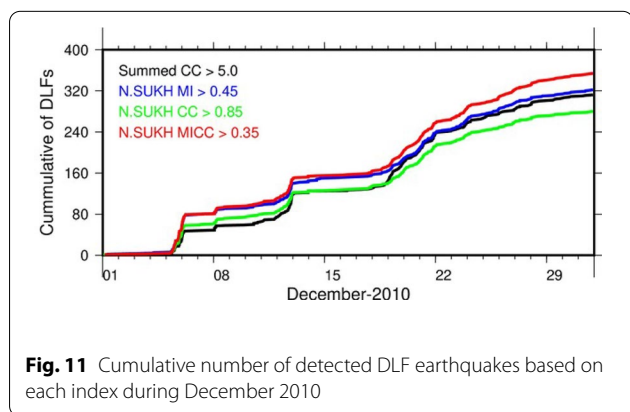


**Fig. 10** Values of MI, CC, and MICC from 0:00 to 12:00 on 21 December 2010. The right axis shows each index after normalization by the MAD. The MAD is calculated using the half-day waveform data. The lowest panel shows the velocity waveform. The template earthquake is the earthquake shown in Fig. 2. 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. All of the events in this time window that were detected by the summed CC and MICC at station N.SUKH were identified as DLF earthquakes by visual inspection

**Table 1** Detection results by visual inspection

Number of events (not removed by amplitude ratio)	Summed CC > 5.0	N.SUKH MICC > 0.35
DLF earthquake	297 (295)	314 (309)
Surface wave	12 (10)	28 (9)
Noise	2 (1)	8 (6)
Regular earthquake, shallow tremor, etc	1 (1)	4 (4)
Total	312 (307)	354 (328)

\* Numbers in brackets are the total numbers of detected events after removing the surface-wave misdetections (i.e., the amplitude ratio between the 0- to 10-s and 10- to 20-s intervals after the detection onset is larger than one)



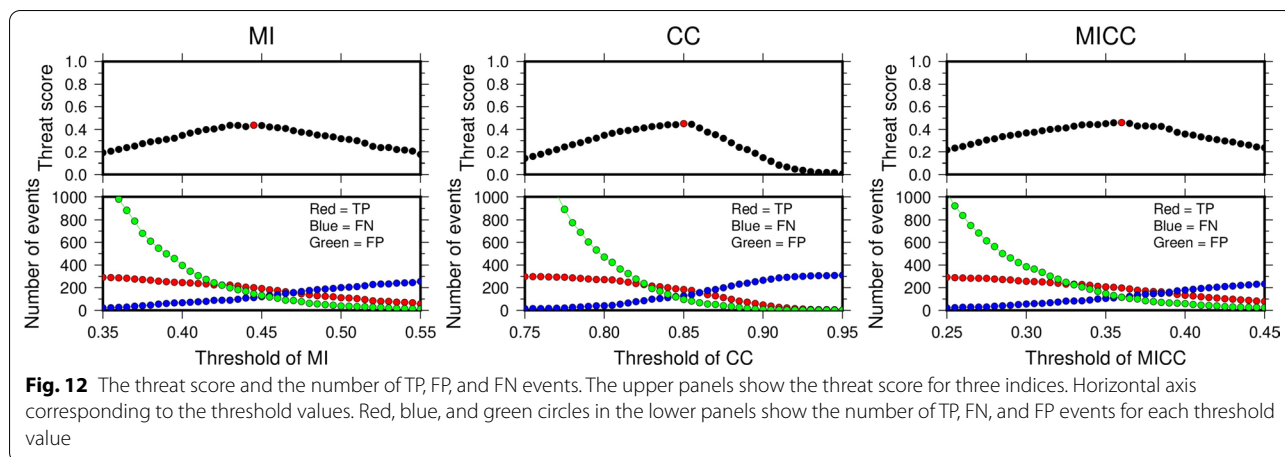
**Fig. 11** Cumulative number of detected DLF earthquakes based on each index during December 2010

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 around the threshold value, and the threat score of 0.461 does not mean that only 46.1% 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.

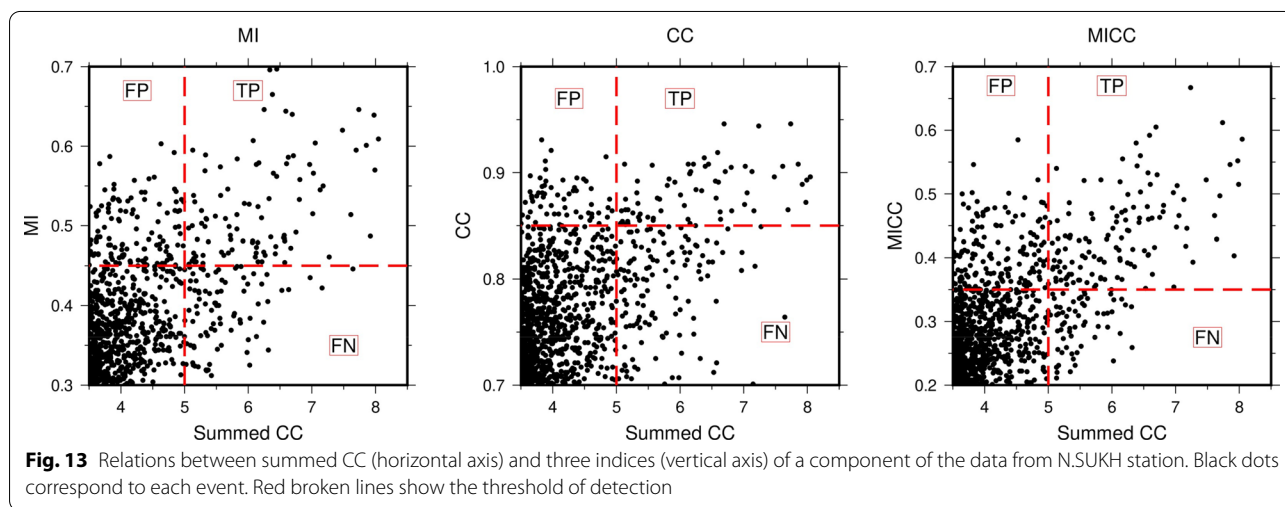
This study calculates MI by dividing the points of velocity-seismograms into  $5 \times 5$  cells. The number of divisions was determined by evaluating the clarity of the peaks corresponding to the signal. Comparing calculations for each number of divisions for the observed continuous

waveform data in the Kirishima region showed that odd-numbered divisions such as  $3 \times 3$  and  $5 \times 5$  give sharper peaks than even-numbered division such as  $4 \times 4$  when the number of divisions is small (Fig. 14). 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 Fig. 4. As  $3 \times 3$  division has a large variance for parts not including signals of DLF earthquakes,  $5 \times 5$  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 (Fig. 14) 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,  $5 \times 5$  division is 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. Even if a micro-earthquake hypocenter cannot be determined due to a weak observation network, the method can at least count the number of microearthquakes. 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.



**Fig. 12** 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



**Fig. 13** 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

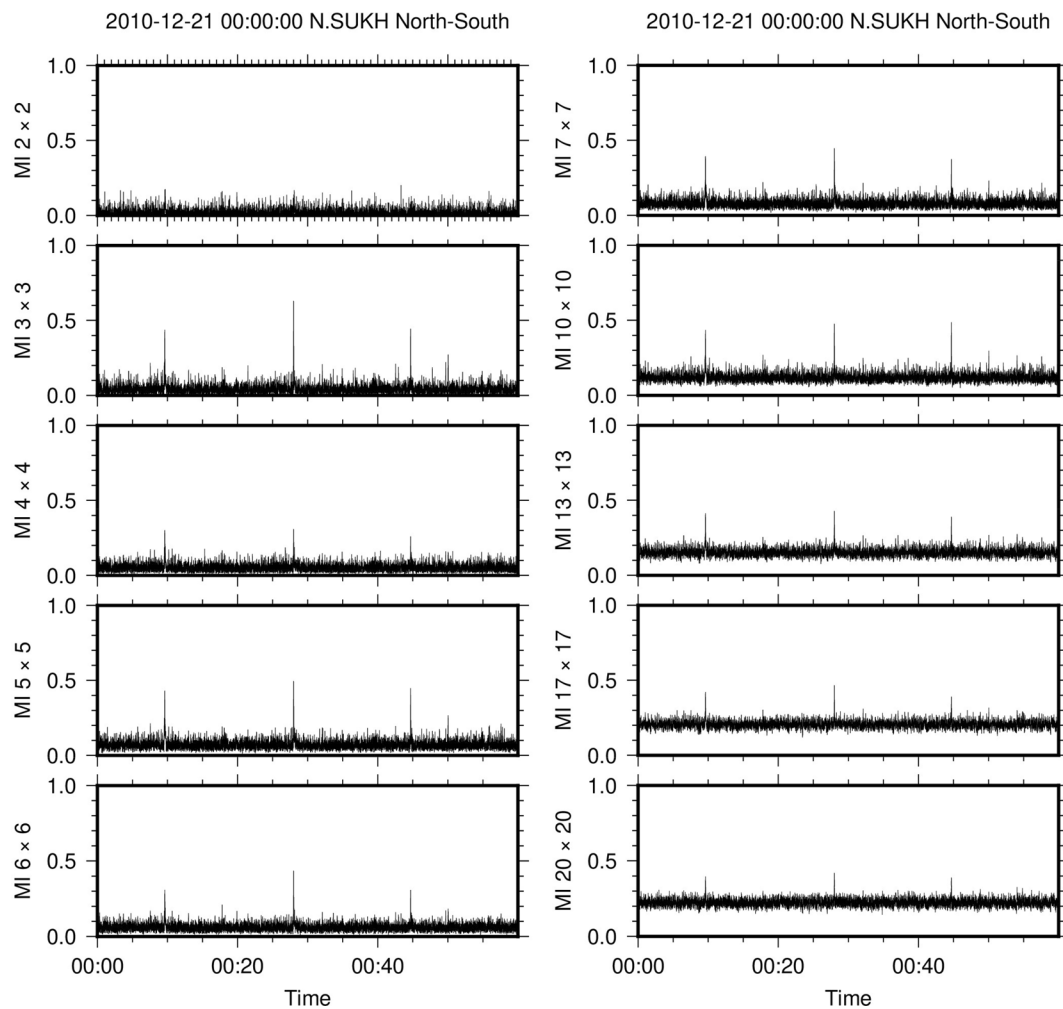
**Table 2** Relationships between the summed CC (horizontal axis; six stations) and the three indices (vertical axis) from a single station (N.SUKH)

One station/ multi-station		Summed CC (six stations)	
		Detected	Not detected
N.SUKH MI	Detected	192	130
	Not detected	120	–
N.SUKH CC	Detected	184	96
	Not detected	128	–
N.SUKH MICC	Detected	207	147
	Not detected	121	–

**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 314 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.





**Fig. 14** 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 (N.SUKH) are as in Fig. 2

#### 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; FN: False negative; JST: Japan Science and Technology Agency.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40623-021-01534-w>.

**Additional file 1: Figure S1** Values of CC[CC] (Gibbons, 2021) and MICC from 0:00 to 12:00 on December 21, 2010. The right axis shows each index after normalization by the MAD. The MAD is calculated using the half-day waveform data. **Figure S2.** Three-components, 1–8 Hz bandpass-filtered waveforms of the earthquake recorded at the six stations. Three waveforms E-W, N-S, U-D components are plotted from top. Each waveform is normalized by the maximum absolute amplitude, which is written in the right of the Figure. **Figure S3.** Three-components,

1–8 Hz bandpass-filtered waveforms of a mis-detected event of surface wave recorded at the six stations. **Figure S4.** Three-components, 1–8 Hz bandpass-filtered waveforms of a false negative event, which is detected by summed CC and not detected by MICC at N.SUKH, at six stations

#### Acknowledgements

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).

#### Authors' contributions

RK mainly designed and analyzed the data and wrote the paper. AK advised on and discussed the contents of this paper. HN and SK advised about indices in statistics and discussed the paper. All authors read and approved the final manuscript.

**Funding**

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

**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>).

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Earthquake Research Institute, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-Ku, Tokyo 113-0032, Japan. <sup>2</sup>Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-Ku, Tokyo 113-8656, Japan.

Received: 30 July 2021 Accepted: 18 October 2021

Published online: 14 December 2021

**References**

- Beaucé E, Frank WB, Romanenko A (2018) Fast matched filter (FMF): an efficient seismic matched-filter search for both CPU and GPU architectures. *Seismol Res Lett* 89:165–172. <https://doi.org/10.1785/0220170181>
- Chamberlain CJ, Shelly DR, Townend J, Stern TA (2014) Low-frequency earthquakes reveal punctuated slow slip on the deep extent of the Alpine Fault, New Zealand. *Geochem Geophys Geosyst* 15:2984–2999. <https://doi.org/10.1002/2014GC005436>
- Gao D, Kao H (2020) Optimization of the match-filtering method for robust repeating earthquake detection: the multisegment cross-correlation approach. *J Geophys Res Solid Earth* 125:1–19. <https://doi.org/10.1029/2020JB019714>
- Gibbons SJ (2021) The optimal correlation detector? *Geophys J Int* 228:355–365. <https://doi.org/10.1093/gji/ggab344>
- Gibbons SJ, Ringdal F (2006) The detection of low magnitude seismic events using array-based waveform correlation. *Geophys J Int* 165:149–166. <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 Meteorol Agency.
- 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. <https://doi.org/10.1186/s40623-020-01257-4>
- Kato A, Obara K (2014) Step-like migration of early aftershocks following the 2007 M w 6.7 Noto-Hanto earthquake, Japan. *Geophys Res Lett* 41:3864–3869. <https://doi.org/10.1002/2014GL060427>
- Katsumata A, Kamaya N (2003) Low-frequency continuous tremor around the Moho discontinuity away from volcanoes in the southwest Japan. *Geophys Res Lett* 30:20 1–20 4. <https://doi.org/10.1029/2002GL015981>
- Kurihara R, Obara K (2021) Spatiotemporal characteristics of relocated deep low-frequency earthquakes beneath 52 volcanic regions in Japan over an analysis period of 14 years and 9 months. *J Geophys Res Solid Earth*. <https://doi.org/10.1029/2021JB022173>
- 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* 124:13079–13095. <https://doi.org/10.1029/2019JB018032>

- National Research Institute for Earth Science and Disaster Resilience (2019) NIED Hi-net. *Natl Res Inst Earth Sci Disaster Resil*. <https://doi.org/10.17598/nied.0003>.
- Okada Y, Kasahara K, Hori S et al (2004) Recent progress of seismic observation networks in Japan—Hi-net, F-net, K-NET and KiK-net—. *Earth Planets Space*. 56:xxv–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/ngeo0697>
- Pluim JPW, Maintz JBA, Viergever MA (2003) Mutual-information-based registration of medical images: a survey. *IEEE Trans Med Imaging* 22:986–1004. <https://doi.org/10.1109/TMI.2003.815867>
- Reshef DN, Reshef YA, Finucane HK et al (2011) Detecting novel associations in large data sets. *Science* 334:1518–1524. <https://doi.org/10.1126/science.1205438>
- Reshef YA, Reshef DN, Finucane HK et al (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* 123:044108. <https://doi.org/10.1063/1.1953327>
- Shapiro NM, Droznin DV, Droznina SY et al (2017) Deep and shallow long-period volcanic seismicity linked by fluid-pressure transfer. *Nat Geosci* 10:442–445. <https://doi.org/10.1038/ngeo2952>
- Shelly DR, Beroza GC, Ide S (2007) Non-volcanic tremor and low-frequency earthquake swarms. *Nature* 446:305–307. <https://doi.org/10.1038/nature05666>
- Ueno H, Hatakeyama S, Aketagawa T et al (2002) Improvement of hypocenter determination procedures in the Japan Meteorology Agency. *Q J Seismol* 65:123–131
- Vuan A, Sagan M, Amati G, Kato A (2018) Improving the detection of low-magnitude seismicity preceding the Mw 6.3 LAquila earthquake: development of a scalable code based on the cross correlation of template earthquakes. *Bull Seismol Soc Am* 108:471–480. <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* 368:775–779. <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* 79:579–579. <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* 2019(46):11305–11043. <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:P11006. <https://doi.org/10.1088/1742-5468/2015/11/P11006>

**Publisher's Note**

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

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)