

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

Posted Date: August 5th, 2021

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](#)

Version of Record: A version of this preprint was published at Earth, Planets and Space on December 1st, 2021. See the published version at <https://doi.org/10.1186/s40623-021-01534-w>.

1 **Title: Detection of low-frequency earthquakes by matched**
2 **filter technique using the product of mutual information**
3 **and correlation coefficient**

4

5 Author #1: Ryo Kurihara, Earthquake Research Institute, the University of Tokyo,
6 rkuri@eri.u-tokyo.ac.jp

7 Author #2: Aitaro Kato, Earthquake Research Institute, the University of Tokyo,
8 akato@eri.u-tokyo.ac.jp

9 Author #3: Sumito Kurata, Graduate School of Information Science and Technology,
10 the University of Tokyo, kurata@mist.i.u-tokyo.ac.jp

11 Author #4: Hiromichi Nagao, Earthquake Research Institute, the University of Tokyo,
12 nagaoh@eri.u-tokyo.ac.jp

13

14 Indicate the corresponding author: Ryo Kurihara

15

16 **Abstract**

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

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

39
40
41 **Keywords**

42 Matched filter technique, Low-frequency earthquakes, Mutual information, Kirishima
43 volcano

44
45

46 Introduction

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

62 Although matched filter method can be applied to data from only one station (for
63 example, when the seismic network is small) (e.g. Vuan et al., 2018; Wech et al., 2020),
64 the sum of CC between template waveforms and in the three components data measured
65 at multiple observation stations was often used in most cases (e.g. Gibbons and Ringdal,
66 2006; Shelly et al., 2007). The analysis of low-frequency earthquakes usually employs a
67 time window of 4–5 s for filtered waveforms at 1–4 Hz or 2–8 Hz. The similarities of
68 waveforms measured at multiple stations are evaluated by CC stacked over a seismic
69 network. Earthquakes are identified by the summed CC being larger than a threshold
70 value. Using stacked CC from multiple stations is advantageous over using of a single
71 station to detect small-magnitude earthquakes including low-frequency earthquakes,
72 because the CC from a single station always become high, as noise shares the same
73 frequency band as the signal, resulting in substantial false detections. In other words,
74 conventional matched filter technique is applicable only in regions sufficiently covered
75 by observation stations and with well-determined seismic catalogs. Single-station
76 matched filter method could potentially be effective in various regions in which
77 microearthquakes occur, thus allowing detection of much smaller events that are only
78 recorded by a single station (e.g. Vuan et al., 2018). However, much improvements of

79 technique is required to maintain the quality of the catalog.

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

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

101

102 **Data and Method**

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

108 Conventional matched filter usually evaluates detection using stacked CC
109 between observed and template waveforms of three components measured at multiple
110 seismic stations. The template is selected from previously observed earthquakes. The
111 template events used here are DLF earthquakes in the unified catalog of the Japan

112 Metrological Agency (JMA) (e.g. Katsumata and Kamaya, 2003). The CC of
 113 component j at seismic station i are calculated as

$$114 \quad CC(i, j, t_{tg} + \Delta t_i) = \frac{\sum_{\tau}(v_{tp}(i, j, t_{tp} + \Delta t_i + \tau)v_{tg}(i, j, t_{tg} + \Delta t_i + \tau))}{\sqrt{\sum_{\tau}(v_{tp}^2(i, j, t_{tp} + \Delta t_i + \tau))}\sqrt{\sum_{\tau}(v_{tg}^2(t_{tg} + \Delta t_i + \tau))}} \quad (1).$$

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

$$125 \quad \text{Summed } CC(t_{tg} + \Delta t_i) = \sum_i \sum_j CC(i, j, t_{tg} + \Delta t_i) \quad (2).$$

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

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

$$135 \quad \bar{v}_{tp}(t) = \frac{v_{tp}(t)}{\max(|v_{tp}(t)|)} \cdots (3)$$

$$136 \quad \bar{v}_{tg}(t) = \frac{v_{tg}(t)}{\max(|v_{tg}(t)|)} \cdots (4)$$

$$137 \quad n_{tp}(t) = \text{floor} \left((\bar{v}_{tp}(t) + 1.4) * 2.5 \right) \quad (\text{When } \bar{v}_{tp}(t) < 1) \cdots (5)$$

138
$$n_{tg}(t) = \text{floor}((\bar{v}_{tg}(t) + 1.4) * 2.5) \text{ (When } \bar{v}_{tg}(t) < 1 \text{)} \cdots(6)$$

139
$$n_{tp}(t) = 5 \text{ (When } \bar{v}_{tp}(t) = 1 \text{)} \cdots(7)$$

140
$$n_{tg}(t) = 5 \text{ (When } \bar{v}_{tg}(t) = 1 \text{)} \cdots(8)$$

141 The notation $| \cdot |$ means the absolute function, and $\text{floor}(\cdot)$ means the floor
 142 function converting a real number to the largest integer smaller than itself. The constants
 143 1.4 and 2.5 convert the velocity into integers. This calculation means that $n = 1$ when the
 144 normalized velocity $\bar{v}(t)$ ranges from -1.0 to -0.6 , $n = 2$ when $\bar{v}(t)$ is between -0.6
 145 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
 146 0.6 , and $n = 5$ when $\bar{v}(t)$ is between 0.6 and 1.0 .

147 Using the above integers, the MI is calculated as

148
$$MI(t) = \sum_{n_{tp}=1}^5 \sum_{n_{tg}=1}^5 p(n_{tp}, n_{tg}) \log \frac{p(n_{tp}, n_{tg})}{p(n_{tp})p(n_{tg})} \quad (9).$$

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

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

156
$$h_{tp} = \sum -p(n_{tp}) \log p(n_{tp}) \cdots(10)$$

157
$$h_{tg} = \sum -p(n_{tg}) \log p(n_{tg}) \cdots(11)$$

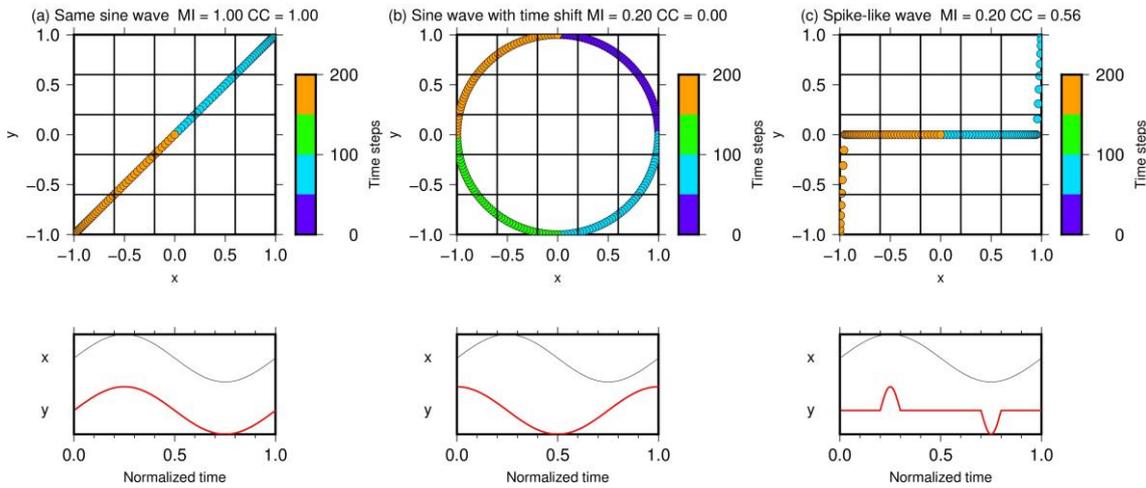
158
$$\overline{MI} = \frac{2.0 * MI}{h_{tp} + h_{tg}} \cdots(12).$$

159 Hereafter, we simply describe \overline{MI} as MI.

160 MI is an index that reflects the linear or non-linear relation between two
 161 variables. If the relation is linear, both MI and CC are high (Figure 1a). While if the
 162 relation is non-linear, MI has a non-zero value (Figure 1b). According to equation (1), CC

163 gives more weight to data points with high amplitudes in phase, while MI increases when
 164 the data points are concentrated in a small number of cells. In other words, CC are high
 165 when the peaks of two variables match, even if the shapes of the waveforms are different
 166 (Figure 1c).

167



168

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

174

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

177
$$\text{MICC}(t) = \text{MI}(t)\text{CC}(t) \quad \dots(13).$$

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

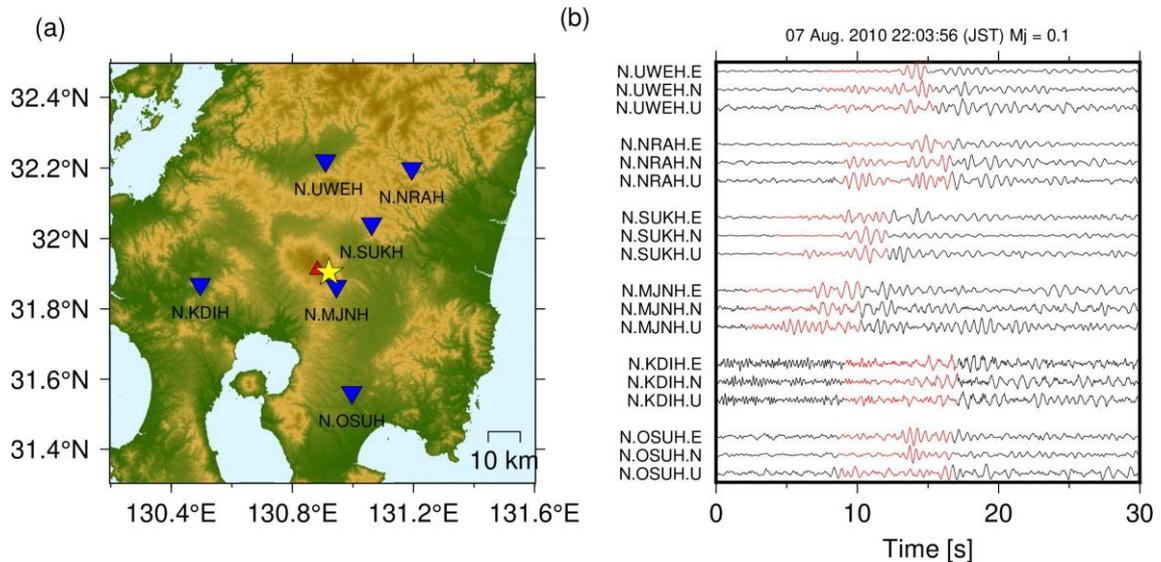
181

182 Synthetic Tests

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

187 August 2010 (Figure 2).

188



189

190

191

192

193

194

195

196

197

Gaussian noise

198

199

200

201

202

203

204

205

206

207

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)

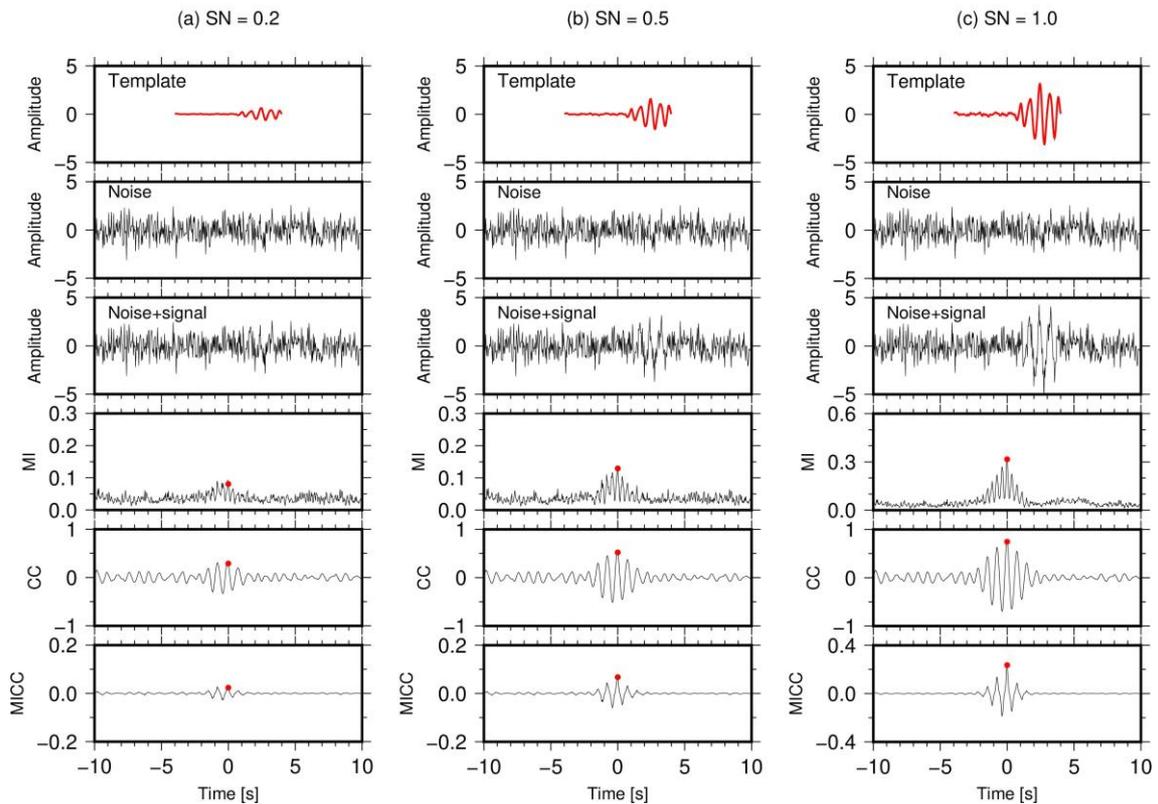
Waveforms of three components of the earthquake recorded at the six stations. The

horizontal axis is the time elapsed from the origin. Red parts of the waveforms show

the template waveform used in this study.

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.



208

209

210

211

212

213

214

215

216

217

218

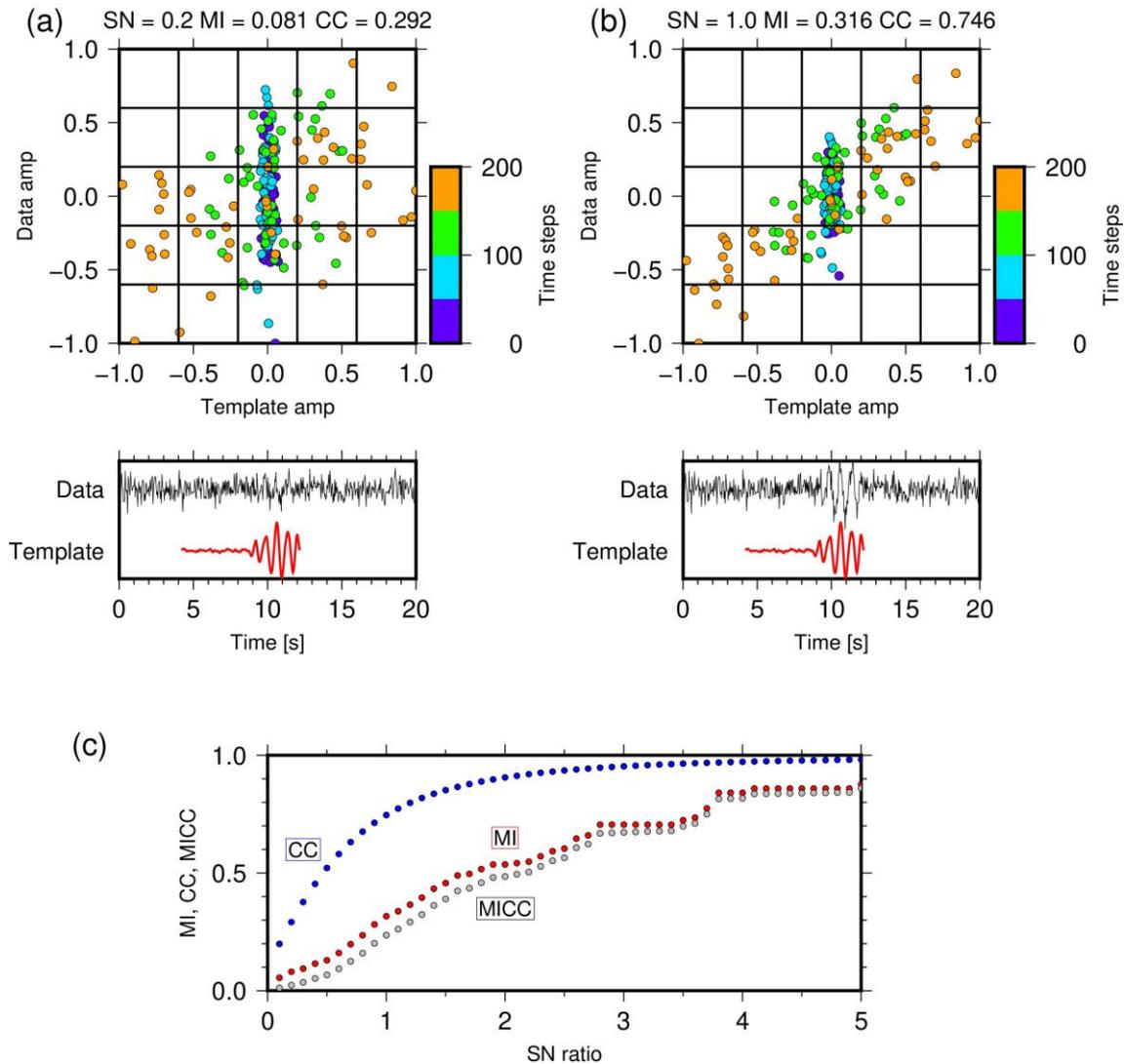
219

220

221

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.



222

223

224

225

226

227

228

229

230

231

232

233

234

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.

235

Sinusoidal noise

236

237

238

239

240

241

242

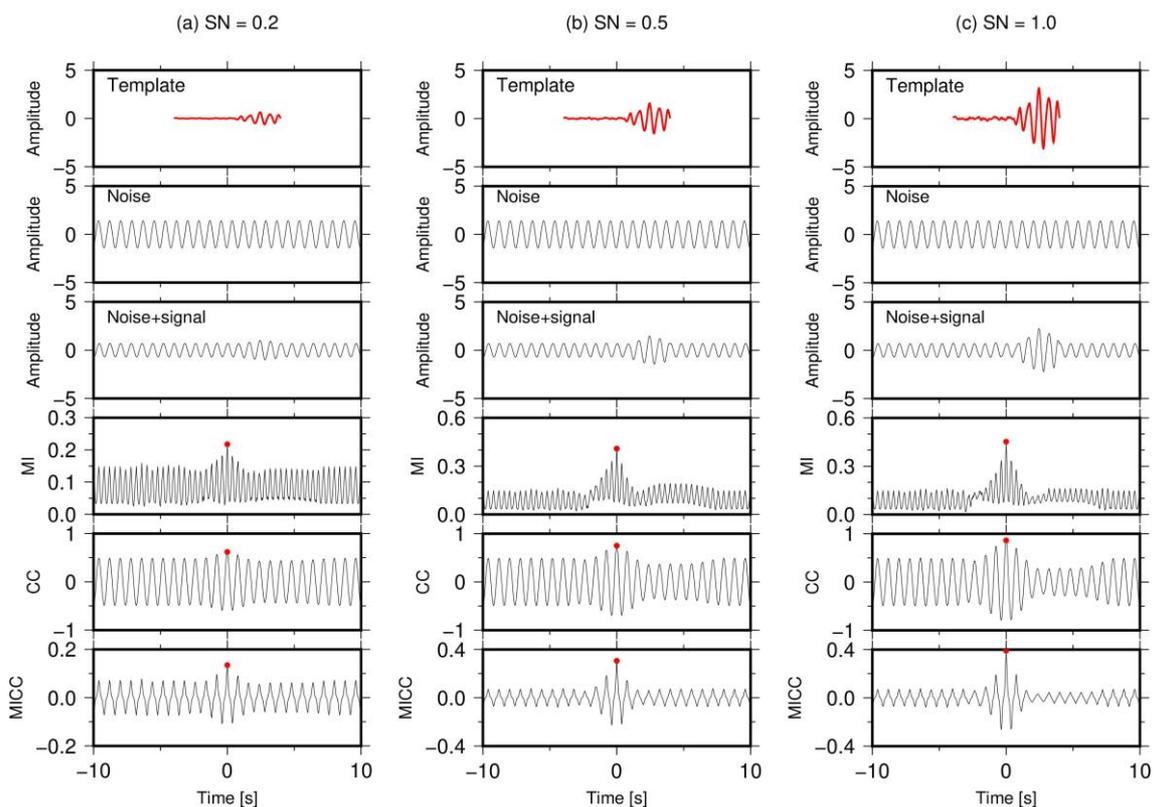
243

244

245

246

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.



247

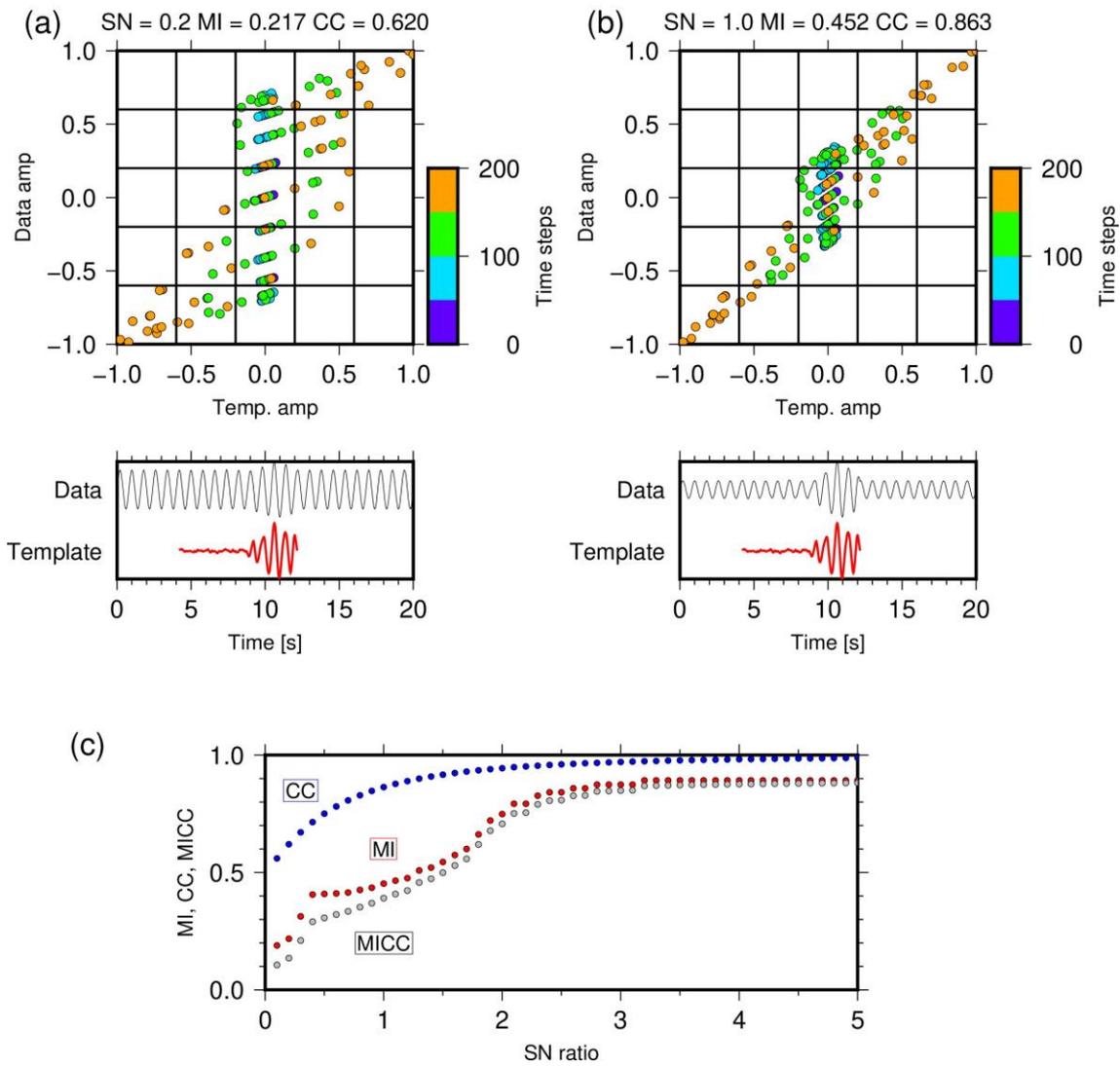
248

249

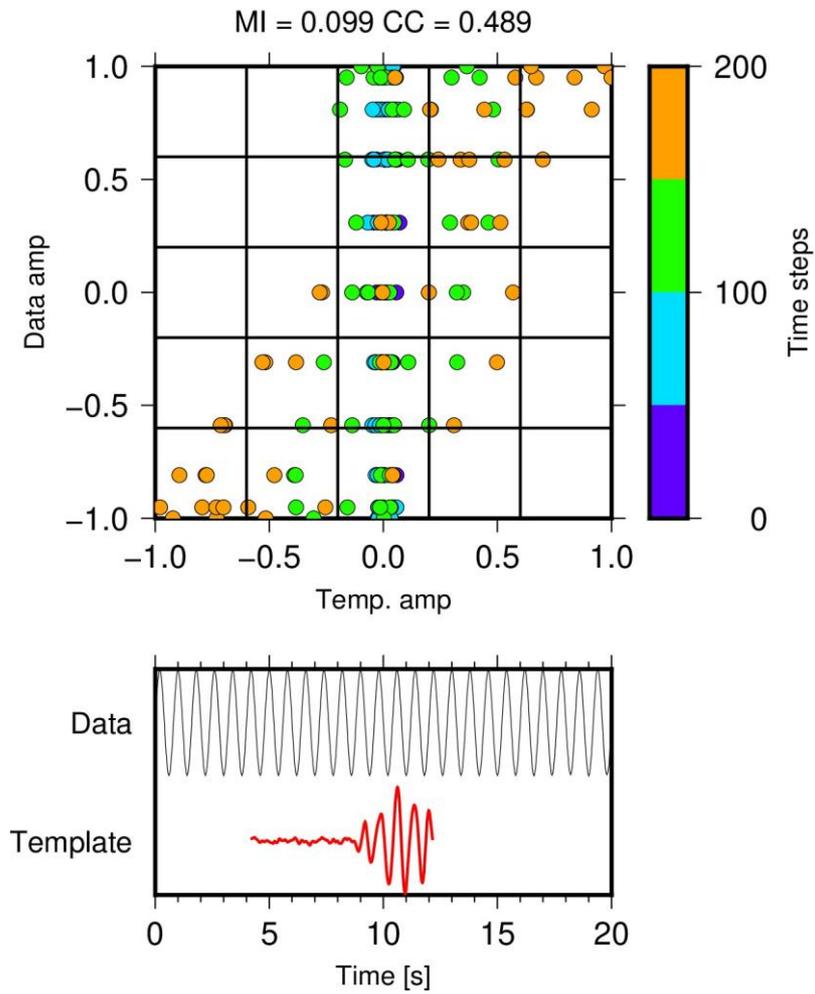
250

251

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.



254 **Figure 6** Scatter plots and relationship between the SN ratio of the waveform and
 255 the MI, CC, and MICC for the synthetic waveform with sinusoidal noise. The figures are
 256 as in Figure 4, but here the noise is a sinusoidal wave of frequency 1.25 Hz.



258

259

260

261

262

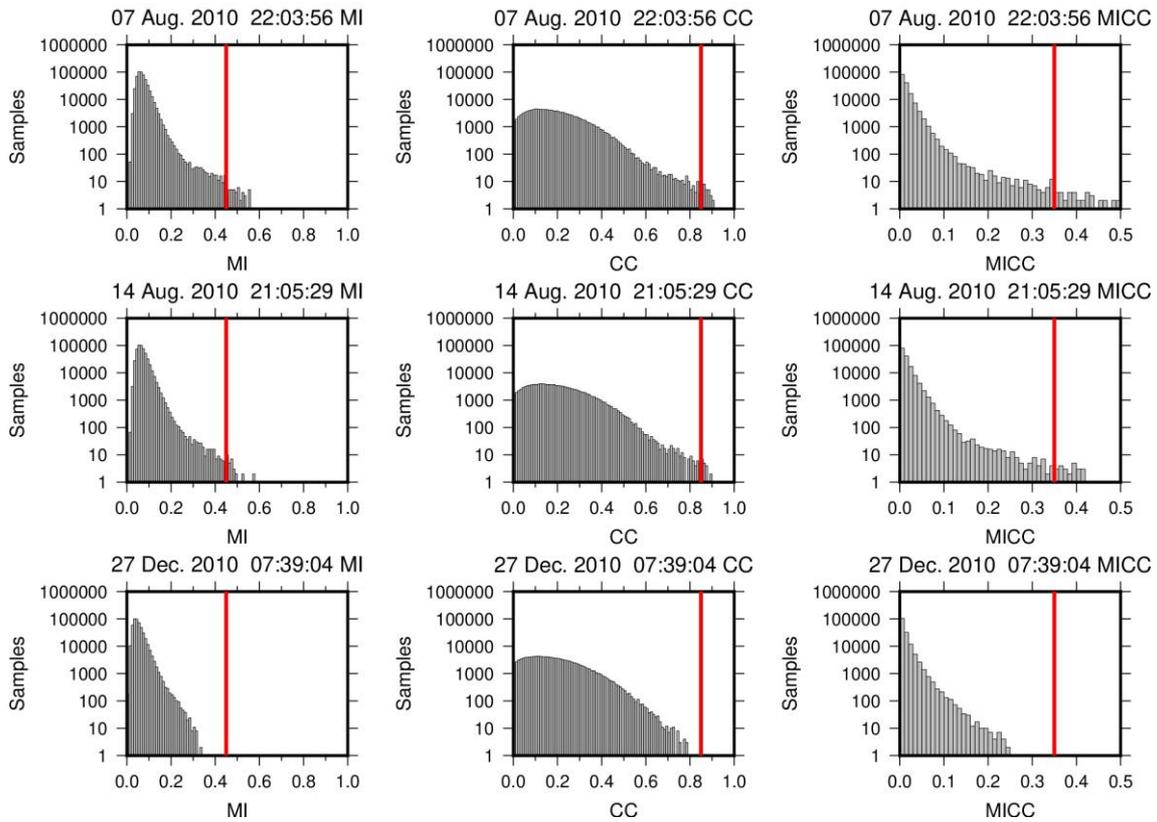
263

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.



294

295 **Figure 8** Histograms of the maxima of three indices (MI, CC, and MICC) extracted
 296 from each time series. Vertical axes correspond to the number of samples in each bin
 297 (width, 0.01). The time at the top of each panel is the time of occurrence of the template
 298 event (JST). Applied data are for one day, 21 December 2010. Red lines show the
 299 threshold value for each index.

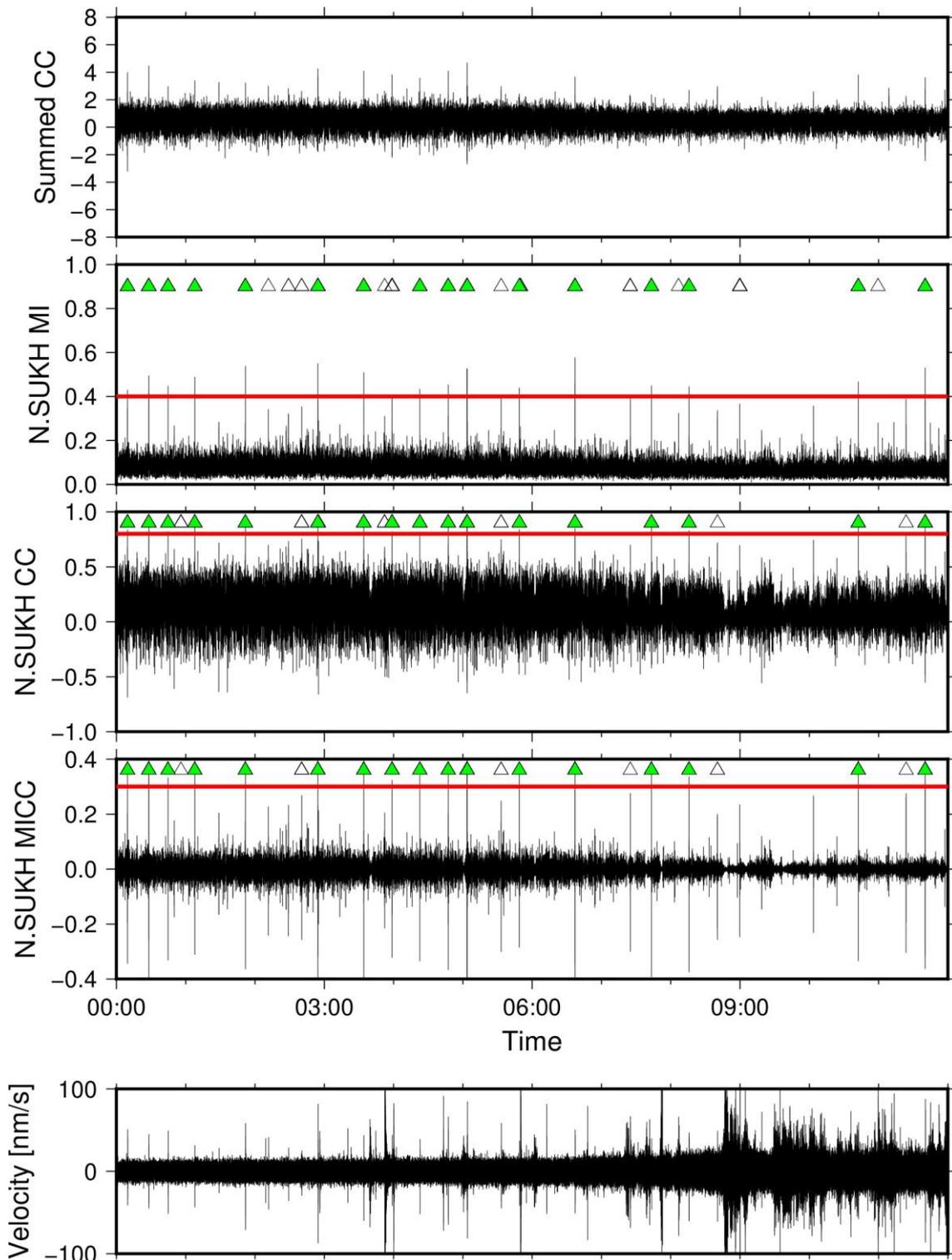
300

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

312 correspond to large amplitudes in the velocity waveform except some parts corresponding
313 to the signals of earthquakes occurring in other regions.

314 MICC identified 354 DLF earthquakes in the test month. Even using data from a
315 single station achieved a temporal change of the cumulative number of DLF events
316 similar to that observed from the catalog of events compiled from six stations' data
317 (Figure 10). The catalogs include step-like increases in the number of DLF earthquakes
318 on December 5 and December 12, and the difference between the catalogs over the
319 month is small. This trend is generally consistent across the results from any of the three
320 indices. As many of the earthquakes detected by these indices have very small SN
321 ratios, it is not possible to distinguish between a true signal and a false detection even
322 by visual inspection of the waveforms. Although the MICC peak is the most distinct and
323 the detection looks good, the accuracies of the detection of the three single-station
324 indices are quantitatively evaluated in discussion.

325



326

327

328

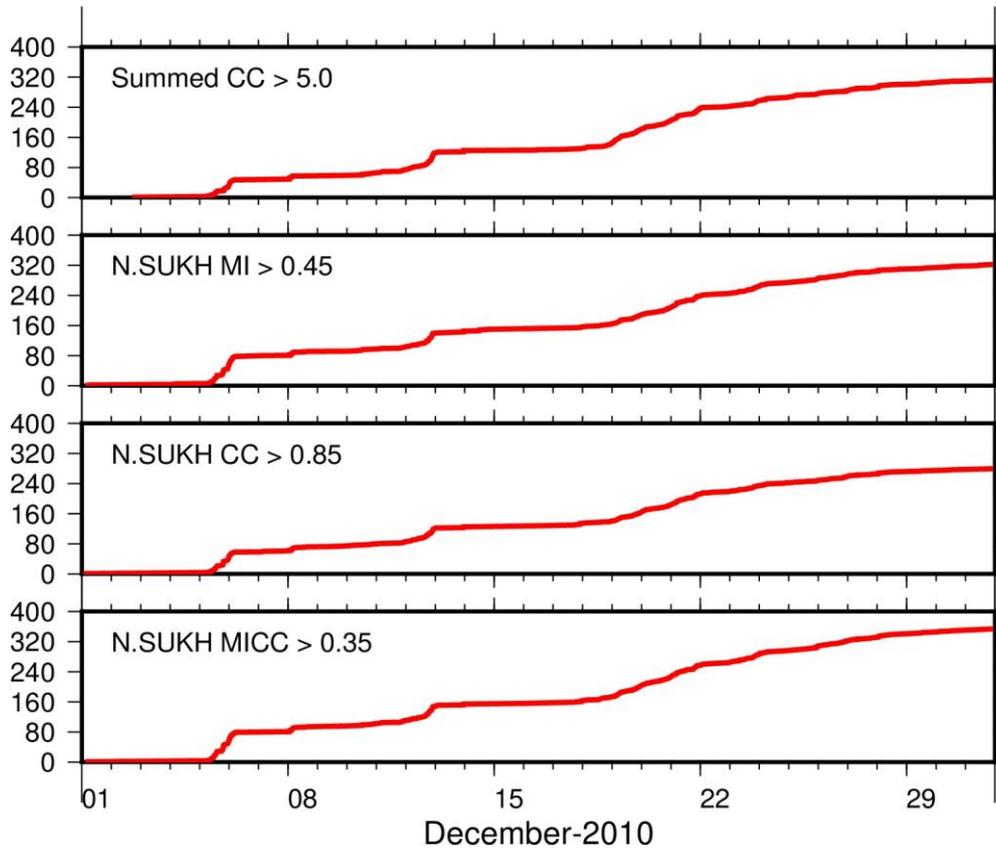
329

330

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

331 triangles show the detection using other templates.

332



333

334

335

336

337

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.

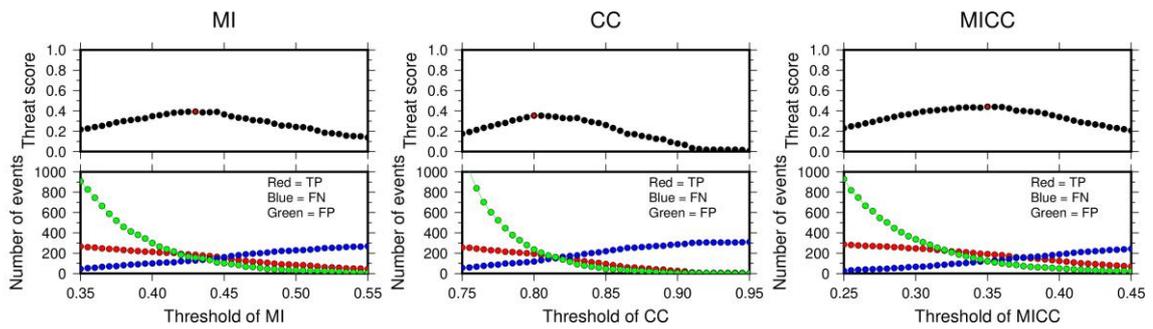
Discussion

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:

$$\text{Threat score} = \frac{TP}{TP+FP+FN} \quad \dots(14),$$

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.



364

365

Figure 11 The threat score and the number of TP, FP, and FN events.

366

The upper panels show the threat score for three indices. Horizontal axis corresponding

367

to the threshold values. Red, blue, and green circles in the lower panels show the

368

number of TP, FN, and FP events for each threshold value.

369

370

We compare the maximum value of summed CC in the 5 s before and after

371

the time with that of each index at the single station (Figure 12). The summed CC is

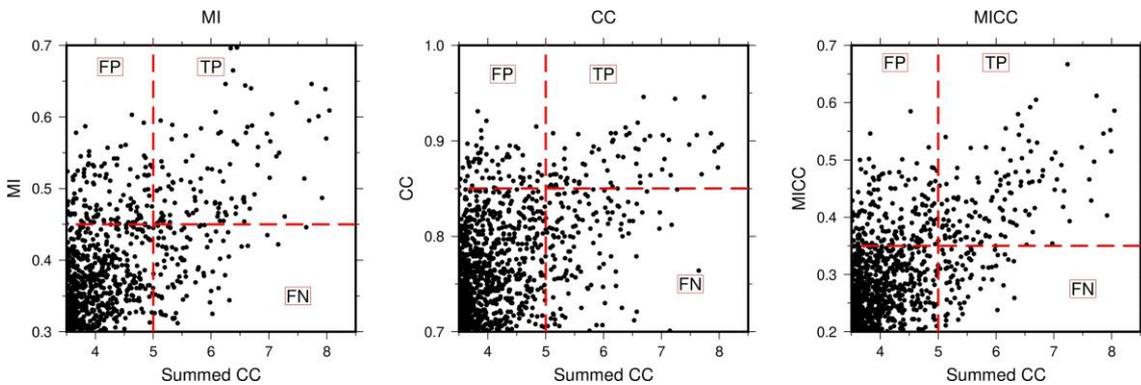
372

positively correlated with each of MI, CC, and MICC applied to a single station's data.

373

Especially, MICC shows the best positive relationship with the summed CC.

374



375

376

Figure 12 Relations between summed CC (horizontal axis) and three indices

377

(vertical axis) of a component of the data from N.SUKH station. Black dots

378

correspond to each event. Red broken lines show the threshold of detection.

379

380

The threat score for the MICC catalog is the highest among the three catalogs;

381

however, the value seems low. We next assess the quality of the catalog compiled using

382

MICC. Of the 88 events detected with summed CC over 6.0, 80 are also included in the

383

MICC catalog. On the other hand, of the 354 events detected using MICC, 85 % of the

384

events (302 events) are also in the catalog of summed CC with the threshold of 4.0. In

385

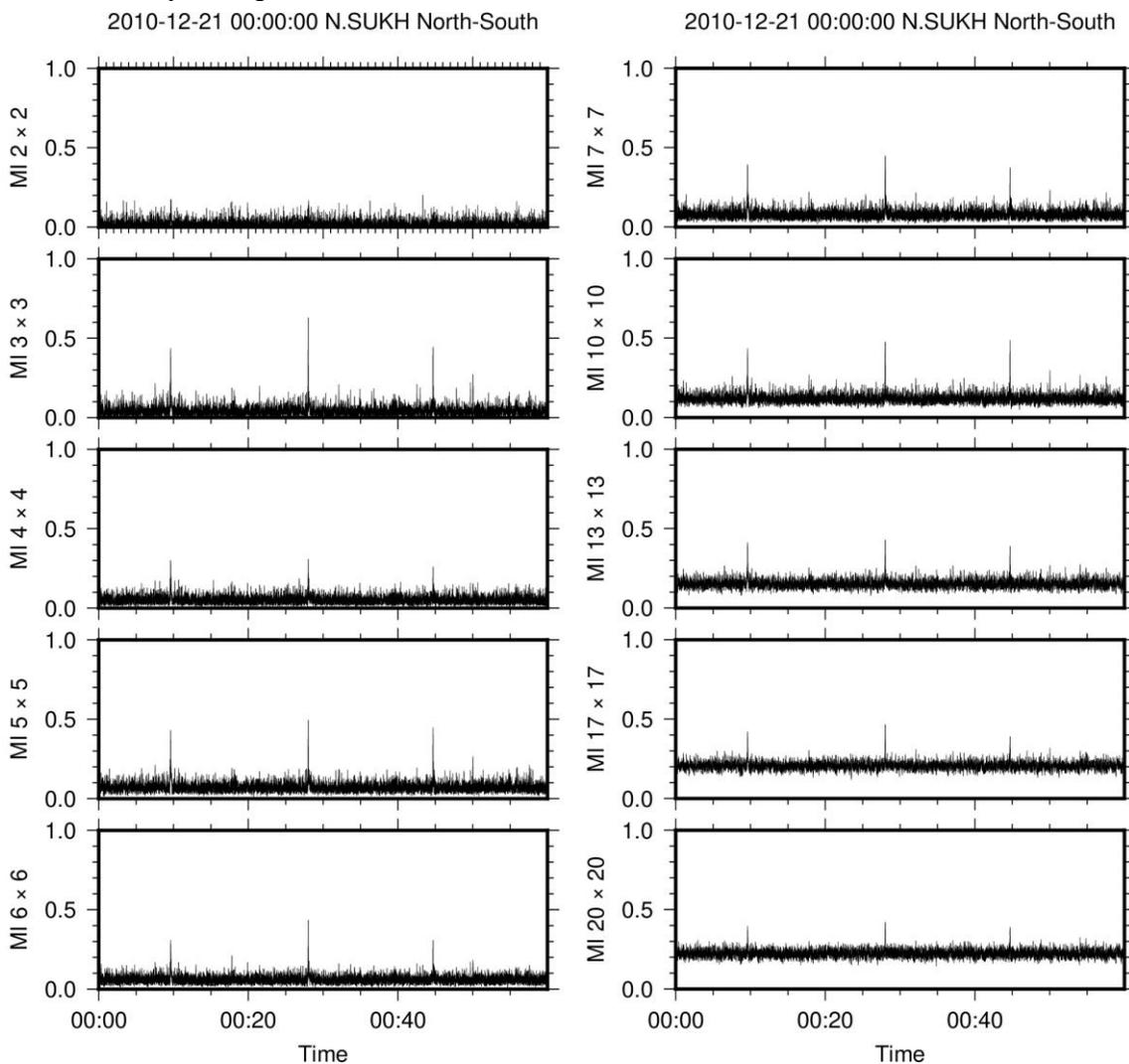
other words, the FP and FN events included those which are around the threshold value,

386 and the threat score of 0.441 does not mean that only 44.1% of DLF earthquakes are
387 detected. The single-station catalog probably includes DLF earthquakes with very small
388 magnitudes that are only observed at the single station (N.SUKH); however, this
389 comparison cannot distinguish FP detections from actual small DLF earthquakes.

390 This study calculates MI by dividing the points of velocity-seismograms into
391 5×5 cells. The number of divisions was determined by evaluating the clarity of the peaks
392 corresponding to the signal. Comparing calculations for each number of divisions for the
393 continuous waveform showed that odd-numbered divisions such as 3×3 and 5×5 give
394 sharper peaks than even-numbered division such as 4×4 when the number of divisions
395 is small (Figure 13). This is because the points near the origin in even-numbered divisions.
396 belong to different cells due to slight differences of the seismograms. This leads to a
397 decrease in the MI when considering the distribution diagram as shown in Figure 4. As
398 3×3 division has a large variance for parts not including signals of DLF earthquakes,
399 5×5 is considered suitable for detection in the three data sets. On the other hand, as the
400 number of divisions is increased, the baseline of MI rises and the peaks become generally
401 less sharp (Figure 13) because there are too many cells relative to the number of the points.
402 Therefore, for our dataset with 25 Hz sampling and an 8 s time window, 5×5 division is
403 optimal.

404 In this study, we calculate MI by dividing the points of velocity-seismograms into the
405 5×5 cells. The number of divisions was determined by evaluating the clarity of the peaks
406 corresponding to the signal. As a result of the calculation for each number of divisions
407 for the continuous waveform, the peak of the odd-numbered division such as 3×3 and 5
408 $\times 5$ is sharper than that of the even-numbered division such as 4×4 when the number of
409 divisions is small (Figure 13). This is because the points near the origin belong to different
410 cells due to slight differences of the seismograms. It leads to a decrease in the MI when
411 considering the distribution diagram as shown in Figure 4. In addition, 3×3 has a large
412 variance for the parts which do not include signals of DLF earthquakes, therefore, 5×5
413 is suitable for the detection in the three data sets. On the other hand, as the number of
414 divisions is increased, the baseline of MI rises and the peaks become duller in general
415 (Figure 13) because the number of cells is too much compared to the number of the points.
416 Therefore, for our dataset with 25 Hz sampling and an 8-second time window, we
417 determined that a 5×5 split was optimal.

418 Using MICC with a single station's data is a potentially powerful tool,
 419 especially for monitoring shallow volcanic earthquakes occurring beneath the crater
 420 because there cannot be many seismic stations in that region due to the risk and low
 421 accessibility. In addition, the single-station method will improve the completeness of
 422 small-magnitude volcanic seismicity studies, deepening understanding about volcanic
 423 activity beneath the crater. As in previous cases of single-station analysis (Vuan et al.
 424 2018; Wech et al. 2020), this method will also contribute to the monitoring of other low-
 425 SN ratio earthquakes, such as DLF earthquakes and swarm earthquakes that can be
 426 observed only with part of an observation network.



427
 428 **Figure 13** Time series of MI using 10 different divisions. The time window is one hour
 429 from 00:00:00, 21 December 2010 (JST). The template earthquake and observation
 430 station are as in Figure 9.

431

432 **Conclusions**

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

442

443 **Declarations**

444 **Ethics approval and consent to participate**

445 Not applicable

446 **Consent for publication**

447 Not applicable

448

449 **List of abbreviations**

450

451 MFT: Matched filter technique

452 DLF: Deep Low-Frequency

453 CC: Correlation coefficients

454 MI: Mutual information

455 MICC: Product of mutual information and correlation coefficients

456 Hi-net: High-sensitivity seismograph network

457 NIED: National Research Institute for Earth Science and Disaster Resilience

458 JMA: Japan meteorological agency

459 MAD: Median absolute deviation

460 TP: True positive

461 FP: False positive

462 FN: False negative

463 JST: Japan Science and Technology Agency

464

465

466 **Availability of data and materials**

467 We used the Hi-net seismic observation data from NIED (National Research Institute
468 for Earth Science and Disaster Resilience 2019) and JMA's unified earthquake catalog.

469 Those are available from the web page of NIED Hi-net (<http://www.hinet.bosai.go.jp>).

470

471

472

473 **Competing interests**

474 The authors declare that they have no competing interests.

475

476

477 **Funding**

478 This work was supported by Japan Science and Technology Agency (JST) CREST

479 Grant Number JPMJCR1763, Japan.

480

481 **Authors' contributions**

482 RK mainly design and analyze the data and write the paper. AK advises and discusses

483 the contents of this paper. HN and SK advise about indices in statics and discuss the

484 paper.

485

486 **Acknowledgments**

487 We used Generic Mapping Tools for drawing figures (Wessel and Smith 1998) and

488 collated the Hi-net seismic observation data (<http://www.hinet.bosai.go.jp>) from NIED

489 (National Research Institute for Earth Science and Disaster Resilience 2019). We used

490 JMA's unified earthquake catalog (<http://www.jma.go.jp>) and the computer systems of

491 the Earthquake and Volcano Information Center of the Earthquake Research Institute,

492 the University of Tokyo. This work was supported by the JST CREST (grant number

493 JPMJCR1763) and partially supported by the Ministry of Education, Culture, Sports,

494 Science and Technology (MEXT) of Japan, under its The Second Earthquake and

495 Volcano Hazards Observation and Research Program (Earthquake and Volcano Hazard
496 Reduction Research).

497

498 **References**

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

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

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

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

512 Katsumata A, Kamaya N (2003) Low-frequency continuous tremor around the Moho
513 discontinuity away from volcanoes in the southwest Japan. *Geophys Res Lett*
514 30(1):20-1-20–4. <https://doi.org/10.1029/2002GL015981>

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

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

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

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

527 Plum JPW, Maintz JBA, Viergever MA (2003) Mutual-information-based registration

528 of medical images: a survey. *IEEE Trans Med Imaging* 22(8):986–1004.
529 <https://doi.org/10.1109/TMI.2003.815867>

530 Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, et al.
531 (2011) Detecting novel associations in large data sets. *Science* 334(6062):1518–
532 1524. <https://doi.org/10.1126/science.1205438>

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

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

537 Shapiro NM, Droznin DV, Droznina SY, Senyukov SL, Gusev AA, Gordeev EI (2017)
538 Deep and shallow long-period volcanic seismicity linked by fluid-pressure transfer.
539 *Nat Geosci* 10(6):442–445. <https://doi.org/10.1038/ngeo2952>

540 Shelly DR, Beroza GC, Ide S (2007) Non-volcanic tremor and low-frequency
541 earthquake swarms. *Nature* (7133):305–307. <https://doi.org/10.1038/nature05666>

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

546 Vuan A, Sukan M, Amati G, Kato A (2018) Improving the detection of low-magnitude
547 seismicity preceding the Mw 6.3 L’Aquila Earthquake: Development of a scalable
548 code based on the cross correlation of template earthquakes. *Bull Seismol Soc Am*
549 108(1):471–480. <https://doi.org/10.1785/0120170106>

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

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

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

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

561

<https://doi.org/10.1088/1742-5468/2015/11/P11006>

562

Supplementary Files

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

- [graphicalabst.jpg](#)