DrumCorr program for selecting volcanic earthquake multiplets based on cross-correlation analysis

I A Zavedevkin\textsuperscript{1,2,3}, A A Shakhrova\textsuperscript{1,2} and P P Firstov\textsuperscript{1,2}

\textsuperscript{1} Kamchatka Branch of the Geophysical Survey, Russian Academy of Sciences, Russia
\textsuperscript{2} Kamchatka State University named after Vitus Bering, Petropavlovsk-Kamchatsky, Russia
\textsuperscript{3} E-mail: zed@emsd.ru

Abstract. The DrumCorr program based on cross-correlation detection has been developed to identify multiplets of the volcanic earthquakes. The program is implemented in Python 3 and reads ASCII and MiniSEED seismic data formats. The article presents the algorithm of the program, describing the cross-correlation detector and an example of subsequent processing of seismic data. The program was applied to volcanic earthquakes of the «drumbeats» seismic regime and allowed to identify earthquake multiplets characterized by various wave forms. The article presents the algorithm of the program, describing the cross-correlation detector, the features of the weak volcanic earthquakes selection by the STA/LTA method. And the primary analysis of the values of the correlation coefficients with the calculation of their standard errors depending on different signal-to-noise ratios.

1. Introduction

It is known that earthquakes registered by one seismic station can be characterized by sufficiently close wave forms. This similarity occurs when earthquakes have similar mechanisms of foci and a close hypocentral location \cite{1, 2}. This similarity is characteristic of both tectonic and volcanic earthquakes.

Sometimes, in the process of squeezing out the extrusion, quasi-periodic earthquakes of the «drumbeats» regime are recorded on andesite and dacite volcanoes with magmas of very high viscosity. Only a few cases of the occurrence of this seismic regime are known in the world practice \cite{3}. Sometimes multiplets (families of earthquakes) precede explosive eruptions and, thus, are a sign of an approaching volcanic eruption.

The allocation of earthquake multiplets, including volcanic ones, is based on the cross-correlation of the wave shape. This method is also very common when allocating multiplets of the “drumbeats” regime earthquakes \cite{4-8}.

The cross-correlation of the two signals $x_p$ and $y_p$ can be represented in the following form. The similarity between the events $x_p(t)$ and $y_p(t)$ is quantified using the cross-correlation function $r_{xy}(\tau)$:

$$r_{xy}(\tau) = \int_{-\infty}^{\infty} x_p(t) y_p(t + \tau) d\tau$$

where $\tau$ is the delay between the two signals.

The change in $\tau$ changes the relative position of the signal $x$ relative to the signal $y$. 

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
It should be noted that the correlation function $r_{xy}$ measures only the similarity of the wave form of the signal, and not the amplitude of events. Thus, the amplitudes can vary for events with similar waveforms. This means that the waves moved along an almost identical trajectory, but were not necessarily created by a source with a constant force.

The quantitative correlation parameter is the correlation coefficient $r$. The value of $r$ is in the range of numbers from -1 to 1. The closer $r$ is to 1, the stronger the direct relationship between the variables, the closer $r$ is to -1 – the inverse. When $r=0$, there is no significant relationship between the two variables.

To track changes in the wave forms of numerous earthquakes of the «drumbeats» seismic regime over long observation intervals and divide them into multiplets, the DrumCorr program was developed in the Geophysical Service of the Russian Academy of Sciences, Kamchatka branch (KB GS RAS), which is based on a cross-correlation detector of wave forms.

The DrumCorr program is implemented in the Python 3. The implementation of the cross-correlation function is taken from the ObsPy library for working with seismic data [https://docs.obspy.org/].

The principle of the detector operation is based on the mutual correlation of wave forms. The sliding window, the length of which is equal to the length of the template (reference earthquake), moves along the time series, shifting in increments of one record. For each selected section of the signal with a given pattern, the correlation coefficient is calculated. If the coefficient of the final correlation function exceeds the threshold value, the detector is triggered.

The DrumCorr program allows to work with MiniSEED and ASCII seismic data formats. It is possible to support other time series formats supported by the ObsPy library [https://docs.obspy.org/].

The source code of the program is available on Github: https://github.com/ZiCode0/DrumCorr.

2. Program description
The program starts by reading the configuration file (figure 1), which contains the following parameters: the time series filter, the threshold for detecting correlation, the minimum number of correlated earthquakes in a given file. Also, as additional parameters, the format of the incoming files is specified to determine the file with the signal template and the input time series.

![config_example.json](image)

**Figure 1.** Example of a configuration file.

When working with the «drumbeats» seismic regime, a butterworth filter was applied in the 0.2-10 Hz frequency band for each input file, including the signal template and daily time series. The program also provides using of other filters supported by the ObsPy library [https://docs.obspy.org/].
The result of the cross-correlation detector is recorded in the program environment variables. The result of the detector calculations is a list of found events with the value of the correlation function and the time of the triggered detector (figure 2).

![Figure 2. A block diagram of the Drumcore program.](image)

2.1. Extraction of the maximum earthquake amplitude
The catalog for the earthquake wave forms analysis was created using the STA/LTA (short-time average/long-time average) algorithm [9, 10] (figure 3). To date, this algorithm is most widely used in the isolation of volcanic earthquakes [11-14]. It continuously calculates the average values of the absolute amplitude of the seismic signal in two consecutive time windows. A short time window (STA) is sensitive to seismic events, whereas a long time window (LTE) provides information about the seismic noise amplitude. A block diagram of finding the maximum amplitude of each event is shown in figure 3.

![Figure 3. A block diagram of finding the maximum amplitude of each event.](image)

The ObsPy library has several implementations for allocating signals using the STA/LTA algorithm. Our goal was volcanic earthquakes with a low signal-to-noise ratio (SNR) in some periods. Also,
technogenic noise leads to the appearance of several local maxima of the event amplitude values on the graph and the difficulties of automating $A_{\text{max}}$ allocation. Therefore, the Delayed STA/LTA method is the most suitable variation of the algorithm for selecting and finding the $A_{\text{max}}$ signal [2]. Its peculiarity lies in the cumulative effect of the final function, which allows most clearly allocate the maximum of the signal, unlike the classical implementation of the method.

The DrumCorr program includes the calculation of the Delayed STA/LTA averaged function, based on a sample of the first 20 signals found. If less than 20 of them were found in the input time series, the maximum possible number is used to calculate the averaged function. The resulting normalized function is used to find the $A_{\text{max}}$ of all events in the time series (figure 4).

![Figure 4. Flowchart for calculating the averaged function STA/LTA 1/2.](image)

To process the time series, need to configure the STA and LTA parameters. The choice of these parameters is set by the user based on the analysis of the seismic record and depends on the research tasks. As a rule, the duration of the STA of local earthquakes is taken to be equal to 0.5-0.3 s [2], sometimes it is noted that the STA should correspond to the duration of the P-phase of the earthquake. The duration of the LTA is taken much longer than the duration of the STA [2].

![Figure 5. The lengths of the STA and LTA time windows on the example of the «drumbeats» regime earthquake 5.10.2011 at 14: 48 (a) and an example of the allocation of its signal by the STA/LTA method (b).](image)
To extract earthquakes from continuous seismic data recorded at the KZV seismic station of the KB GS RAS, located 2.6 km from the Kizimen volcano (Kamchatka, Russia), the parameters STA=1 s, LTA=4 s were selected (figure 5a).

For each event detected by the cross-correlation detector, a time fragment with a length of 50 seconds is allocated (25 seconds before and 25 seconds after the detector is triggered). Due to the cumulative feature of the STA/LTA algorithm, where 1/3 are zero values, this part was excluded from finding the maximum. The final length of the time series was a 30 s fragment of the recording, in which the value of the maximum amplitude was found based on the maximum of the STA/LTA function (figure 6).

**Figure 6.** Flowchart for calculating the averaged function STA/LTA 2/2.

After selecting the signals and writing them to the program environment variables, a classical cross-correlation detector is applied. Then get the Pearson correlation coefficient at the output for each correlated earthquake.

The result of calculating the cross-correlation detector for each file is a report file, which displays the total number of selected signals (N), the specified correlation coefficient (r_s), the average and maximum values of the correlation coefficients (r, r_max), as well as a list of signals with time stamps and corresponding r_i. An example of a report file is shown in figure 7c.

3. Practical application of the program

The analysis of earthquake wave forms was carried out on continuous seismic data recorded by the KZV seismic station. We consider two periods of earthquake generation of the “drumbeats” regime, corresponding in time to the movement of the main “language” of the lava flow (in 2011) and the second “language” of the lava flow (2012).

For the first period, a template earthquake (a) and a fragment of a seismic record with the corresponding calculated correlation coefficients are shown in Fig. 7b. As can be seen from the fragment of the DrumCorr program file report for 2011/09/27 (figure 7c), almost all earthquakes registered on 2011/09/27 have a correlation coefficient with the template earthquake selected on 2011/09/06 at 00:00 more than 0.6. During the day, 830 earthquakes were correlated with r=0.63, which indicates a similarity between the earthquakes.

Consider a template earthquake selected for the second period, registered on 2012/04/29 at 02:52 (figure 8a). A fragment of the seismic record with the corresponding calculated correlation coefficients is shown in Fig. 8b. Correlated earthquakes registered in 2012/03/10 have 0.54≤r≤0.84. During these days, 236 earthquakes were correlated with r =0.63.
Figure 7. A template earthquake of the “drumbeats” regime 2011/09/06 at 00:00, KZV, channel SHZ (a); a fragment of the seismic record, KZV, channel SHZ, 2011/09/27 with the start at 05:11:41 with the corresponding correlation coefficients relative to the template earthquake (b) and the corresponding fragment of the DrumCorr program file report (c).

Figure 8. Template earthquake of the “drumbeats” mode 2012/04/29 at 02:52, KZV, channel SHZ (a); a fragment of the seismic record, KZV, channel SHZ, 2012/03/10 with the start at 00:00:00 with the corresponding correlation coefficients relative to the template earthquake (b) and the corresponding fragment of the DrumCorr program file report (c).

The confidence interval of the calculated correlation coefficient was estimated using the formula derived from the work of Xu et al. [15]. The standard deviation for $r_i$ was estimated using the formula:

$$\hat{\delta}(r_i) = \frac{\delta}{\sqrt{\sum_i x_i^2}}$$

(2)

where $r_i$ - correlation coefficient, $n$ is the number of samples, $x_i$ - is the amplitude.

The standard deviation of the noise was estimated from the recording area before the detected earthquake using the formula:
\[
\delta_z = \sqrt{\frac{1}{(1-m)} \sum_{i=1}^{m} x_i^2}
\]  

(3)

where \(\delta_z\) is the estimate of the standard deviation of the noise \(z\), \(m\) - is the number of samples, \(x_i\) - the amplitude.

In cases where earthquakes occurred with a frequency of more than 10 seconds, sections with noise amplitudes were recorded (figure 8b). The initial analysis of the standard deviation showed that in these periods \(\delta(r_i)\) did not exceed 0.08 (figure 9b). If the earthquakes occurred often enough, the wave form of the subsequent earthquake was superimposed on the code wave of the previous one (figure 7b). Then \(\delta(r_i)\) exceeded 0.08, sometimes reaching values of 0.18 (figure 9a). This suggests that when the SNR is less than 3, the reliability of the obtained correlation coefficient decreases.

Figure 9. The calculated correlation coefficients of the sequence of the “drumbeats” regime earthquakes, shown in figure 7 and 8, and their standard deviation.

4. Conclusion

The DrumCorr program allows to identify earthquake multiplets based on a cross-correlation waveform detector. The program is universal and can be applied to seismic data in the MiniSEED and ASCI formats. It is possible to support other formats supported by the ObsPy library and containing time series. The program was tested on the “drumbeats” seismic regime earthquakes and allowed to identify earthquake multiplets characterized by special wave forms. The source code of the program is available on Github: https://github.com/ZiCode0/DrumCorr.

The authors want to express gratitude to Iskandar Abubakirov, Ph.D., for consulting and assistance in data analysis.

Acknowledgements

The authors want to honor the memory of Pavel Firstov, Dr. Sc., under whose strict guidance this study was conducted.

Funding

The work was supported by Ministry of Science and Higher Education of the Russian Federation (№ 075-00576-21). The data used in the work were obtained with large-scale research facilities «Seismic infrasound array for monitoring Arctic cryolitozone and continuous seismic monitoring of the Russian Federation, neighbouring territories and the world» (https://ckp-rf.ru/usu/507436/, http://www.gsras.ru/unu/). The work was carried out within the framework of the research project of the Vitus Bering Kamchatka State University Natural disasters of Kamchatka earthquakes and volcanic eruptions (monitoring forecast, study, psychological support of the population) No. AAAA-A19-119072290002-9.
References

[1] Geller R and Mueller C 1980 *Geophys. Res. Lett.* 7 821-4
[2] Trnkoczy A 2002 Understanding and parameter setting of STA/LTA trigger algorithm *New Manual of Seismological Observatory Practice, IASPEI, GFZ German Research Centre for Geosciences* 8 1-20
[3] Shakirova A and Firstov P 2019 *E3S Web of Conferences* 127 1-6
[4] Stephens C and Chouet B 1989 *J. Volcanol. Geotherm. Res.* 109 133-48
[5] Green D and Neuberg J 2006 *J. Volcanol. Geotherm. Res.* 153 51-63
[6] Petersen T J 2007 *Volcanol. Geotherm. Res.* 166 177-92
[7] Ottemoller L 2008 *J. Volcanol. Geotherm. Res.* 177 1-25
[8] Matoza R and Chouet B 2010 *J. Geophys. Res.* 115 1-26
[9] Allen R 1978 *Bulletin of the Seismological Society of America* 68 1521-32
[10] Allen R 1982 *Bulletin of the Seismological Society of America* 72 225-42
[11] Matoza R, Milton A, Chouet B, D’Auria L, Hedlin M, Groot-Hedlin C and Waite P 2009 *J. Geophys. Res.* 114 1-38
[12] Thelen W, West M and Senyukov S 2010 *J. Volcanol. Geotherm. Res.* 194 201-13
[13] De Angelis S and Henton S 2011 *Geophys. Res. Lett.* 38 1-5
[14] Ketner D and Power J 2013 *J. Volcanol. Geotherm. Res.* 259 45-62
[15] Xu W, Changrun C and Jisheng D 2019 *Signal Processing* 161 165-79