JSEE Vol. 23, No. 4, 2021 DOI: 10.48303/jsee.2023.1989647.1050

Research Paper

Evaluation of the Matched-Filter Approach for Detecting Seismic Phases, Case Study on a Local Network

Iman Kahbasi[1], Ehsan Karkooti[2] and Saeed SoltaniMoghadam[2]*

1. Ph.D. Student, International Institute of Earthquake Engineering and Seismology (IIEES), Tehran, Iran
2. Assistant Professor, Seismological Research Center, International Institute of Earthquake Engineering and Seismology (IIEES), Tehran, Iran,
*Corresponding Author; email: ekarkooti@iiees.ac.ir

Received: 14/02/2023
Revised: 05/03/2023
Accepted: 06/03/2023

ABSTRACT

In many seismological studies, the first task is to identify and locate seismic events. While small-magnitude earthquakes may not pose a significant risk to urban areas, they are of great interest to seismologists due to their frequency and abundance. The detection of earthquakes with magnitudes smaller than the magnitude of completeness (Mc) can be a challenging task, as seismic waves tend to attenuate and unavoidable noise levels at seismic stations. Consequently, many earthquakes with magnitudes smaller than the Mc value of a given seismic network may remain undetected. The Matched-Filter technique is an approach based on signal processing, which makes it possible to identify the seismic phases even with very low signal-to-noise ratios by improving the detection capability of the seismic networks in case of repeating events. The goal of this study is to depict the points that must be considered when employing the Matched-Filter approach. As a result, sensitivity tests were performed on each parameter to demonstrate their importance and effectiveness in influencing the outcomes of utilizing this technique. The statistical measurements revealed that selecting incorrect values reduces the quality of the event identifications and may potentially result in mistakes. Finally, depending on the assessments and settings chosen, this method utilized to process 95 days of continuous data from a local temporary seismographic network demonstrates the technique's capability. A comparison between the catalog obtained by this study with a reference one shows an increase in the frequency of good and medium-grade located events in the ILAM's local seismic network, confirming the method's efficiency.

Keywords:
Matched-Filter;
Cross-correlation;
Local network;
Earthquake detection

1. Introduction

The detectability of seismic events in a seismographic network is affected by inter-station spacing (network density) and the level of noise at each station. When seismic waves propagate, the amplitude of the waves decreases over time and distance from the epicenter due to attenuation. This effect, combined with the unavoidable noise levels in the seismic stations caused seismic waves of many earthquakes with magnitudes less than the Mc of the network, buried in the noise. So, the buried seismic phases are undetectable by observation or detectors that rely on signal energy. As a result, many small earthquakes remain undetected in the network. In the two next paragraphs, we explained two solutions to overcome this issue and why we are interested

in the Matched-Filter technique.

To improve earthquake detectability in a seismic network, two approaches could be used. The first strategy involves expanding the seismic network by adding new stations and increasing the quality of existing stations. This strategy, however, is only beneficial for future earthquakes and cannot improve the outcomes of previously recorded earthquakes. Even with a dense network, identifying hidden seismic phases in environmental noise is impossible. However, the high cost of building and maintaining seismic stations and transferring and storing the data of a dense network is very demanding and makes its implementation economically challenging.

The second approach is to use better signal processing techniques to improve the detection and association of seismic phases. Therefore, it is very advantageous to use improved signal processing techniques to increase the efficiency and detectability of a seismological network. The matched-filter method (waveform cross-correlation as an identifier) is a technique based on signal processing methods, which also enables the identification of seismic phases below the noise level in known source conditions (Schaff & Richards, 2014).

Among the many applications of the waveform cross-correlation technique in seismology are improved event location by determining the exact time difference between two comparable signals (Waldhauser & Ellsworth, 2000), detection of fore-shock activity before large earthquakes (Zhang & Wen, 2015), and detection of low-frequency earthquakes within tremors (Chamberlain et al., 2014).

Energy detectors and manual event identification by specialists do not require prior information about the events. Unlike these methods, a priori information is crucial for using the matched-filter technique and identifying seismic phases possible in the case of known sources. Therefore, this approach is one of the most successful ways of finding near-repeating signals in continuous data (Dodge & Walter, 2015; Gibbons & Ringdal, 2006; Schaff & Richards, 2014).

The fundamental capability of the matched-filter technique is determining seismic phases with outstanding accuracy. Also, seismic phases with low signal-to-noise ratios and even phases buried in the noise are detectable by this technique. These seismic phases may belong to earthquakes with a magnitude lower than M_c of the region, or that are outside the region of the seismic network, and whose amplitude has been reduced as they traveled through the earth. When they reach the station, their amplitude is equal to or less than the amplitude of the seismic noise in the recording station. Among the other advantages of this method is the high time accuracy of seismic phase determination in conditions where the reference earthquake phases are accurately determined. As a result, this method can identify the phases of earthquakes with a known focus with good time accuracy and a low signal-to-noise ratio.

The primary goal of this research is to correctly apply the Matched-Filter method for identifying aftershocks and observing its detection capability. To that end, after introducing the necessary parameters for using this method, the effect of each on the outcome was studied and explained. Following the determination of the best values for each of the discussed parameters, this method was applied to 95 consecutive days of data from a temporary seismic network, and the results of the number and quality of the identified earthquakes were compared with previous identifications using the STA/LTA energy detector and then corrected by the user's phase correction.

2. Theory

Correlation is a mathematical operator that calculates the similarity of two signals and returns a number between -1 and 1. The greater the similarity between the two signals makes the absolute value of the correlation coefficient closer to 1. According to research, the correlation result of two identical signals remains meaningful even when one is buried in noise. Considering the ambient vibrations as random or Gaussian seismic wave fields and the unique features of earthquake waves, these waves are distinctive from other waves (natural or artificial). Therefore, any significant similarity in the result of the correlation of seismic waves and recorded earthquake data indicates a seismic wave with similar characteristics

(Schaff & Waldhauser, 2010). In other words, the reference signal utilized for similarity detection acts as a filter on the data, identifying bits of the continuous data that are similar.

The degree of similarity of different earthquake waves decreases with the increase of focal distance and time separation since attenuation and Green's function of the area affect the signal and vary depending on location (Gibbons et al., 2007). Furthermore, features such as the region's tectonic structure, the focal mechanism, and the source time function determine the degree of similarity across earthquake signals and vary over time for earthquakes. Therefore, it is necessary to select the templates for similarity assessments from the region and period under study (Bobrov et al., 2014).

A seismographic station's seismic signal results from the convolution of the instrument response, attenuation factor, Green's function, and source function (Equation 1). These parameters account for the similarity or dissimilarity of two earthquake signals (Zhang & Wen, 2015). Equation (1) summarizes the effect of various parameters on the recorded signal.

$$O = I*A*G*M \tag{1}$$

where the variables O, I, A, G, and M correspond to the recorded seismic signal, instrument response, source time function, Green's function, and source moment tensor, respectively.

As previously stated, identifying seismic phases using the waveform cross-correlation method requires references that contain all the characteristics of the source, environment and network under examination. So, in the matched-filter technique, it is necessary to have a similar path effect and instrument response on the data used for the similarity measurements.

Using earthquake references from the same stations is one of the relevant alternatives for limiting their influence on the degree of similarity of signals. By doing this, although the items above continue to have an impact, the effect is the same between the refe rence and continuous data.

As illustrated in Figure (1), two earthquakes with neighboring epicenters whose waves were

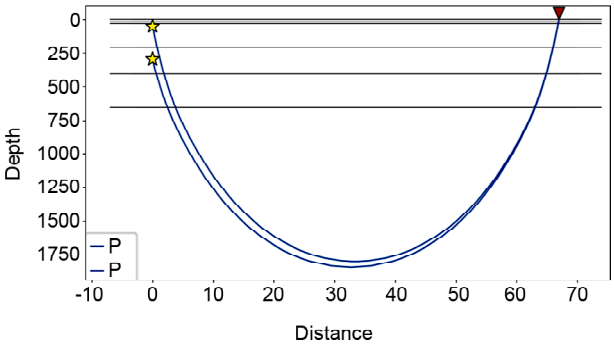

Figure 1. The propagation path of seismic waves caused by two nearby earthquakes to the same station.

detected using the same seismographic stations might be assumed for this purpose. To make things easier, we'll leave out the effect of background noise on the similarity of the two signals. Equation (2) compares the similarity of these two signals.

$$C(O1,O2) = CC\left(I1*A1*G1*M1, I2*A2*G2*M2\right) \tag{2}$$

In this configuration, the effect of attenuation, Green's function, and instrument response will be the same on both signals. Therefore, if one earthquake is identified using the seismic phases of the other, the computed similarities will be independent of the mentioned parameters. As a result, the correlation coefficient (Equation 3) of these two earthquake signals shows the degree of similarity or difference in the features related to the source of the two earthquakes (Zhang & Wen, 2015).

$$CC\left(O1,O2\right) = CC\left(M1,M2\right) \tag{3}$$

By assuming a point source model, the source function will be equal to the multiplication between the wave radiation pattern and the source time function of the source (Equation 4) (Zhang & Wen, 2015).

$$M = F \times S \tag{4}$$

where the variables M, F, and S correspond to the source moment tensor, radiation pattern, and source time function, respectively.

The type of faulting and the tectonic characteristics of a region influence the seismic wave radiation pattern. Because of the slow shift

in tectonic conditions in a region, the focal mechanisms (polarity and radiation pattern) of most earthquakes in that region are nearly identical. As a result, the primary deciding factor in the similarity or difference between seismic waves is the source time function (Zhang & Wen, 2015).

$$CC(O1,O2) = CC(S1,S2) \quad (5)$$

Actually, even if all the mentioned details are the same, due to incoherent environmental noise on the seismic data, the degree of similarity of seismic phases would still be less than 1, the value of the perfect match. Still, the selected references would show good similarity with the seismic phases buried in the noise and retain the ability to identify them (Adushkin et al., 2017).

Studies show that earthquakes in the same region with different source time functions exhibit an acceptable similarity (Zhang & Wen, 2015). Although the detection threshold could decrease to identify earthquakes with various source time functions, at the same time, this issue increases the risk of wrong identifications. Therefore, the detection threshold is a compromise between detecting the maximum possible number of events and the minimum number of false detections (Schaff & Waldhauser, 2010).

The mentioned cases and assumptions are the main foundations of using the matched-filter method in order to identify seismic waves in known source conditions. Therefore, by having the seismic phases of one of the reference earthquakes and calculating its correlation with the continuous data from the same station, it is possible to identify the seismic phases similar to those in the same region and network.

3. Method

This section describes in detail how to utilize the matched-filter technique in multiple steps. In this study, the matched-filter methodology was implemented using the Python programming language, with the Python libraries EQcorrscan (Chamberlain et al., 2018) and ObsPy (Beyreuther et al., 2010). One of the most significant challenges in using the matched-filter technique is the re-quirement of extensive computational resources. The EQcorrscan library addresses this issue by supporting the parallelization of processes, and the central core of the correlation calculation is developed in the C programming language in the frequency domain to increase the computation speed.

To use the similarity filter, quality control, and pre-processing should be performed on the continuous data, and the portions of data that cause problems in the similarity measurement process should be removed. Since the identification of events is conducted on the total correlation signal (a result of stacking the calculated correlation signals), the sampling rate of all templates and continuous data should be the same among all stations (Adushkin et al., 2017).

Following the necessary data pre-processing, the first stage is to identify reference signals from the catalog of primary earthquakes. Based on the suitable parameters of the reference length and pre-phase time, the seismic phases of each earthquake are sliced and utilized as a similarity measurement reference. Selecting appropriate reference earthquakes with Complementary features and a diverse distribution in the study area increases the probability of detecting subsequent aftershocks using the matched filter method. Choosing references with similar properties from the same location can lead to duplicated detections and longer processing times. However, duplicated detections are not dangerous because will be removed in post-processing. The decision to select all or a subset of the earthquakes in the primary catalog as a reference depends on the catalog size, data processing limitations, and the ability to differentiate earthquakes based on their characteristics.

For each reference event, seismic phase similarity is calculated using continuous data from the same station and component. The resulting correlation signals will have a random and Gaussian trend due to the Gaussian and random behavior of the external noise. Otherwise, If there are identical earthquake phases to the reference, the amplitude of the correlation signal rises and approaches 1 or -1 (Warren-Smith et al., 2017).

Next, the 24-hour correlation signals are obtained

for each reference earthquake and shifted based on the relative time difference between the phases of the same reference, and their sum is computed. When the waveforms are plotted beneath each other, the reference earthquake's phases are positioned on the same vertical line due to the time shift. As a result, if an earthquake happens near the reference earthquake, its seismic phases arrive at the stations with the same time difference as the reference earthquake, and by superimposing them, the peak of similarity appears in the stacked correlation signal that indicates a new event.

The number of superimposed phases (amplitude of the total correlation signal) is directly proportional to the focal distance of the detected a nd the reference earthquake. Due to the increasing distance, these similarities picks do not superimpose and vanish in random fluctuations, so the earthquake cannot be recognized by the reference earthquake (Frank & Abercrombie, 2018). The identification process is performed on the overall correlation signal (stacked correlation for each reference event) by defining a similarity threshold, and the discovered phases are classified as seismic events with a position near the reference.

The outcome of this process for all references on each day is the identification of the time of all possible earthquakes on that day based on the identifying reference earthquake and their total correlation value; however, due to the number of reference earthquakes, an earthquake can be detected many times by references close to it. As a result, by defining a time window (the shortest time interval between two consecutive identifications) and keeping the identification with the highest total correlation value, only references with the highest total correlation value (the nearest reference earthquake with the best linearization) are considered as identification factors (Bobrov et al., 2014; Vuan et al., 2018).

Because the new detections have a minor geographic variation from their identifying reference, the times of their seismic phases will be slightly different from the reference earthquake phases. to determine the precise onset of these phases, the correlation of the reference earthquake phases recalculates 0.5 s around the detection time separately. This accurate onset can be obtained by finding the moment of the maximum similarity that exceeds the acceptance threshold, as shown in Figure (2) (Warren-Smith et al., 2017). The times acquired after correcting the pre-phase time used during reference selection are regarded as the actual time of the identified phase, and their type is regarded as equivalent to the corresponding seismic phase in the reference (Figure 3).

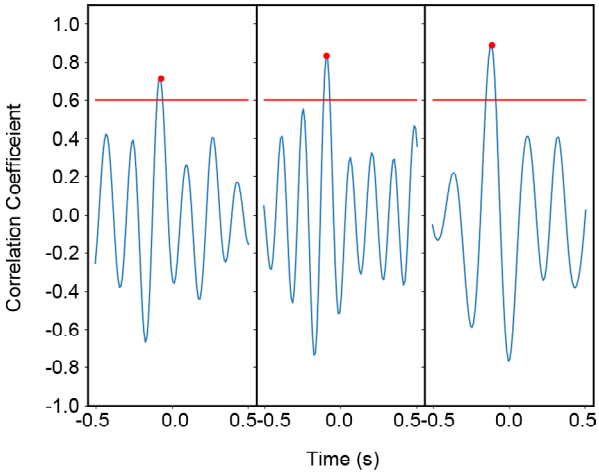

Figure 2. The phase correlation signals for each component and station are used for accurate phase onset identification. By correlating the reference seismic phases within half a second before and after the calculated detection time, a correlation signal of one-second length is obtained. A strict similarity threshold is applied to identify all seismic phases, as shown by the red line in the figure.

After obtaining the precise onset of the seismic phases for each earthquake, the geographical location of these events may be estimated using the velocity model calculated for the region and any location software. In this study, the identified aftershocks were located using the HYPOELLIPSE software (Lahr, 1999).

4. Data and Parameter Optimization

There are several important parameters with a considerable impact on the detection capability and quality of phases in the matched-filter technique. It is crucial to understand the effect of these parameters on the performance of the matched-filter approach. Among these parameters is the length of the window of the reference phases, the amount of pre-phase time in these references, the data sampling rate, the frequency ranges of

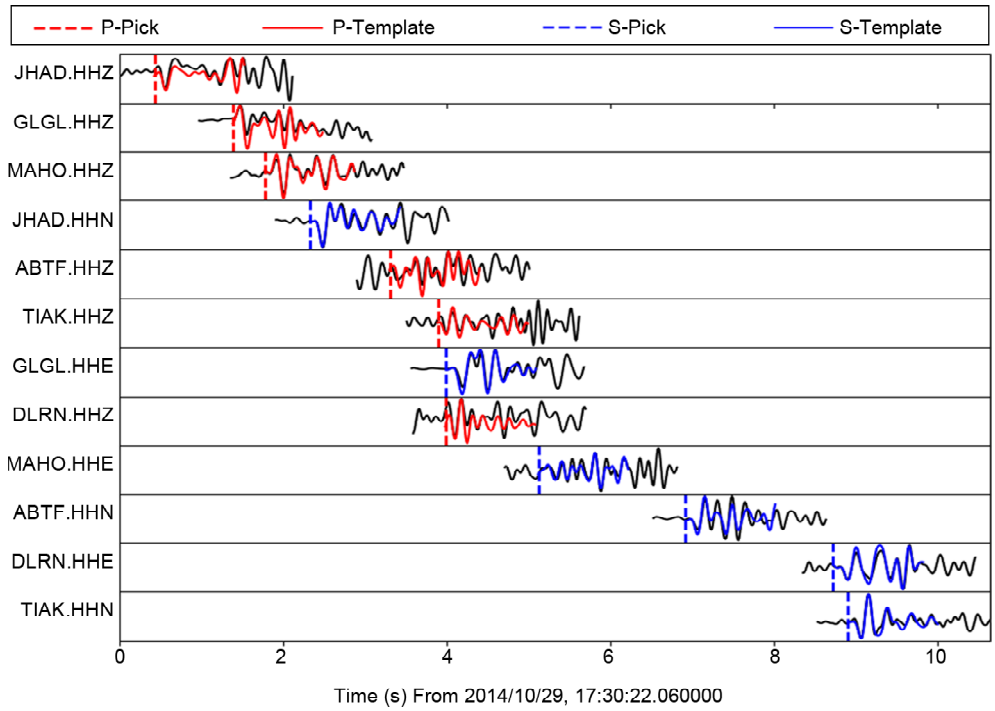

Figure 3. A phase determination example for high-quality identification. The red and blue signals represent the P and S seismic phases of the reference earthquake, respectively, and the black signals are part of the observed environmental noises. The amplitude of the reference signals is scaled in order to see the overall trend in comparison to the detected parts. The identification of seismic phases beneath noise level is indicated by the black signal having the same amplitude throughout its whole length.

the applied filter, the detection threshold of the correlation coefficient in the total correlation signal, the correlation threshold of the seismic phases, and the minimum time interval between consecutive identifications. These parameters need to be designated according to the features of the network and the earthquakes in the region. Therefore, various examinations were conducted to select the optimum parameter set for the network under study.

It should be noted that the seismic phases selected for the templates are derived from the pre-processed continuous data from the previous step involved in pre-processing and filtering the continuous and reference data are identical.

This study investigates the recorded data from the temporary seismographic network installed after the 2014 earthquake in the Ilam region in Iran. This network has included 13 seismograph stations, which recorded seismicity activity for about three months. Although, in the first half of this period, the network was active with only 6 to 7 stations, and most of the aftershocks were detected and located with a low number of seismic phases. There are 838 detected aftershocks (Figure 4) in the reference catalog obtained by the energy detector and revision of the phase by the manual expert. Figure (5) provides a summary of the location parameters for these aftershocks, which are further categorized into four groups based on these parameters (see Table 1). The continuous data were grouped into daily sections (24 hours) with a sampling rate of 100 samples/second, and the DC offset was removed from each segment. Due to the extensive amount of computation in the matched-filter technique, these investigations were carried out using a 24-hour subset of the data, containing only seven earthquakes in the reference catalog. However, the similarity measurements were conducted employing all of the 838 template events.

4.1. Template Length Selection

The time length of the windows separated from the continuous data as reference earthquakes (called templates) is an important component in the performance of the similarity filter approach. Because local earthquakes have a high-frequency content, a shorter data duration can cover the entire P or S seismic phases. So, selecting a

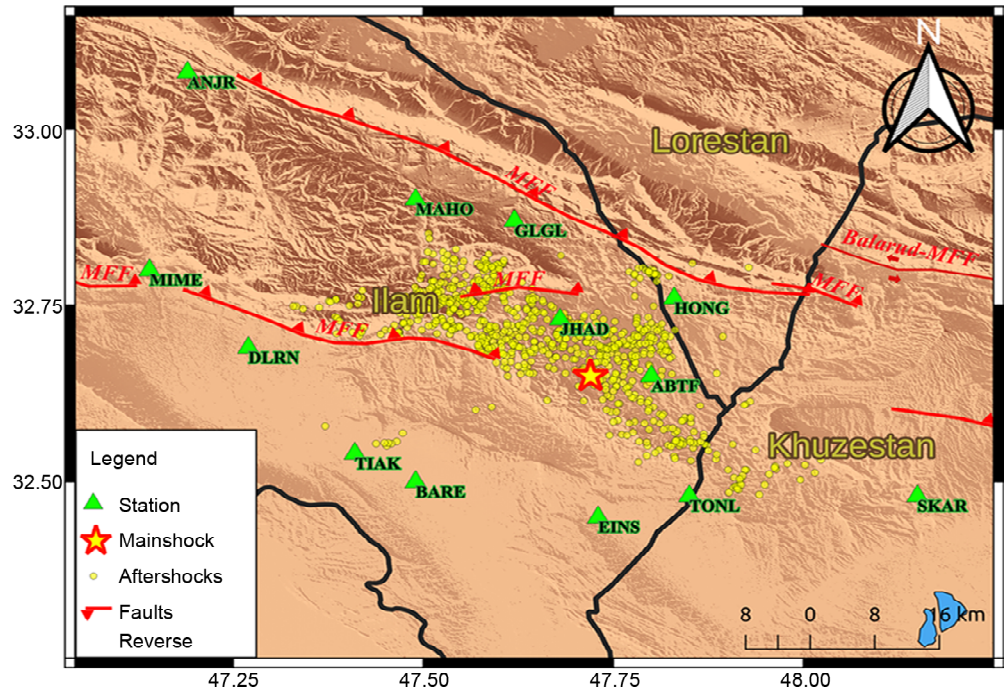

Figure 4. The map shows the locations and distribution of 838 aftershocks, local faults, temporary monitoring network stations, and the main shock.

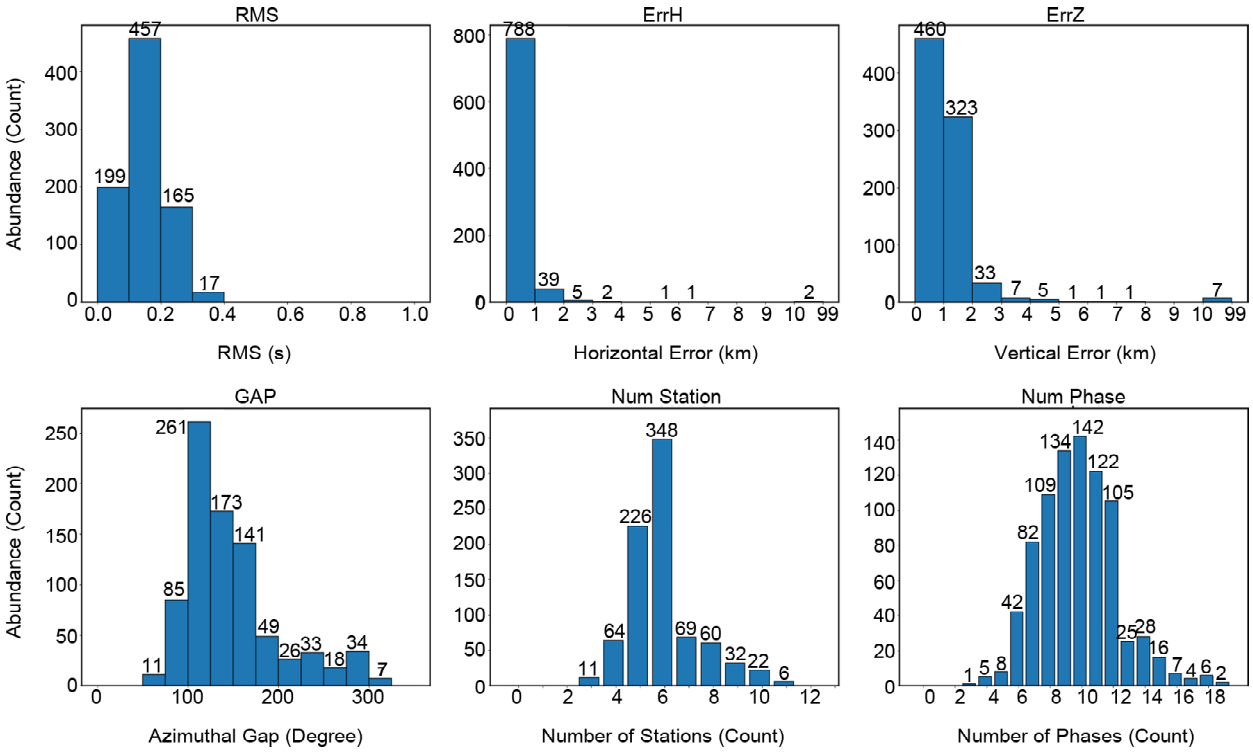

Figure 5. The frequency distribution diagrams of various parameters of 838 aftershocks that were used as a reference in the adaptation-filter method, including time residual values (RMS), horizontal error (ErrH), vertical error (ErrZ), azimuthal gap (GAP), number of stations (Num station), and number of phases (Num phase), are arranged from left to right and top to bottom.

Table 1. Classification of 838 reference earthquakes according to localization quality.

Group	RMS	GAP	Num Stations	ErrH	ErrZ	Rate
A	rms ≤ 0.3	gap ≤ 180	6 ≤ n-st	ErrH ≤ 3	ErrZ ≤ 3	458
B	rms ≤ 0.4	gap ≤ 200	5 ≤ n-st	ErrH ≤ 5	ErrZ ≤ 5	662
C	rms ≤ 0.5	gap ≤ 250	5 ≤ n-st	ErrH ≤ 7	ErrZ ≤ 7	714
Others	rms > 0.5	gap > 250	3 ≤ n-st	ErrH > 7	ErrZ > 7	124

longer data length than the earthquake duration introduces environmental noise into the references and reduces the correlation value with other earthquake data. On the other hand, in the case of distant earthquakes, the duration of the earthquake wave is longer, and considering only a few seconds of seismic phases is too short and doesn't include even a few whole periods of the P or S seismic phase. Failure to capture the seismic aspects of the event in the templates results in a loss of similarity or inaccurate identification of similarity in parts of the data that lack a seismic phase.

In addition, to consider the spatial distance of the detected aftershocks compared to the reference agent, selecting disjointed templates of P and S phases is essential to pick the onset of seismic phases accurately in the phase-picking stage. This study proposes developing a histogram of the S-P time differences of the earthquakes in the reference catalog (Figure 6) to determine the distribution of the distance of earthquakes from network stations. Therefore, using the minimum observed value in the S-P frequency distribution can help limit the chance of combined phases in templates.

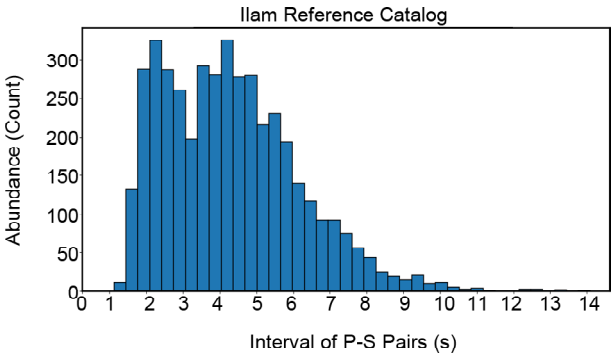

Figure 6. Frequency distribution of time difference of P and S phase pairs from the same station in the catalog of 838 reference earthquakes. The horizontal axis in this diagram represents the time difference between the pairs of phases in seconds, and the vertical axis represents their frequency in different intervals. The smallest observed phase difference is 1.12 seconds..

4.2. Pre-phase Time Selection

According to Warren-Smith et al. (2017), the pre-phase assumed in the references contributes significantly to selecting the onset of seismic phases by reducing uncertainty. Because the data has a small amplitude in the moments preceding the arrival time of the seismic phases and considering the pre-phase placed some small amplitudes in the initial part of templates. Therefore, the templates match better at the moment of entering the seismic phases and cause a higher value of similarity compared to placing in the next period cycle of seismic fluctuations.

Therefore, choosing the optimum quantity for the pre-phase will influence the identification power. In order to evaluate the effect of this parameter on the identification power, the references were selected in four groups of pre-phase duration, including 0.1, 0.3, 0.5, and 1 second of pre-phases and 1 second of after-phase time.

Based on the findings, as the pre-phase time increases, the ability to identify P phases decreases, while the ability to identify S phases increases (Figure 7). This is because the pre-phase of the S phase is also an earthquake wave, meaning that the selected reference has more seismic wave characteristics that aid in identifying and distinguishing the S phase. However, the pre-phase of the P phase includes random fluctuations of external noises, reducing the degree of similarity. Also, it is found that the rate of decline in the detection of p phases is significantly faster than the rate of increase in the detection of S phases.

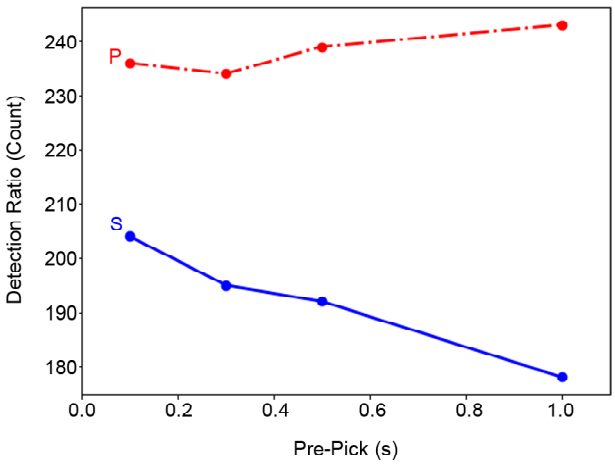

Figure 7. The frequency of P and S identified by references with pre-phase times of 0.1, 0.3, 0.5, and 1 second.

4.3. Choosing Detection Threshold

The fluctuation range of correlation signals and their superposition is unpredictable due to employing multiple references with varying phases and the daily variation in station noise level.

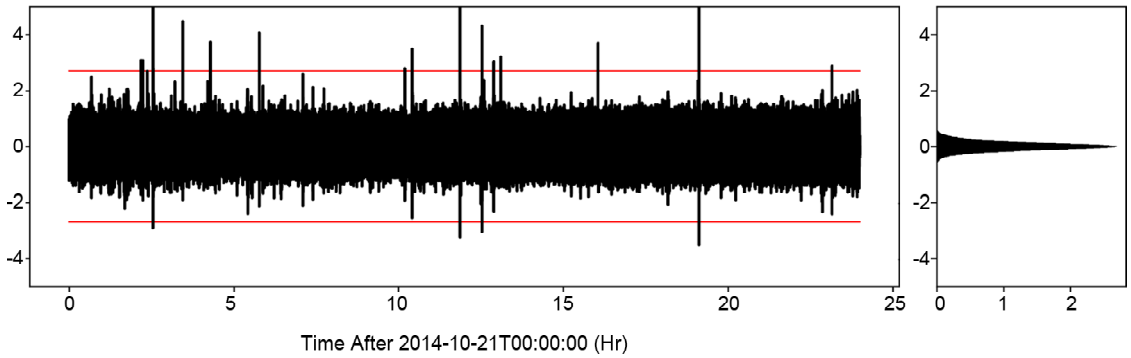

Figure 8. An example of a reference's total correlation signal in one day and setting the detection threshold using a coefficient of the Median Absolute Deviation (MAD). The black signal in the diagram on the left represents the overall correlation signal, and the red lines correspond to the identification threshold, which is a coefficient of the MAD. The frequency distribution of the standard deviation of the overall correlation signal is given in the graph to the right.

Therefore, a constant threshold cannot be utilized (Shelly & Hill, 2011).

To avoid false detection, this threshold should always be larger than random fluctuations but low enough to recognize occurrences with poor correlation coefficients. So, for each reference and 24-hour data segment, a threshold needs to be determined, which can be calculated by multiplying a constant coefficient with the Median Absolute Deviation (MAD) (Equation 6) of the sum of correlation signals for each day and template, as shown in Figure (8) (Warren-Smith et al., 2017).

$$Threshold = coefficient \times MAD = coefficient \times (1.4826 \times variance) \quad (6)$$

Selecting the coefficient value is a trade-off between maximizing the number of detections and minimizing the rate of false detections. So, the value depends on the goal of the study. The threshold selected is equal to 10 times the median absolute value of the deviation (MAD) (6.745 times the standard deviation) of the total correlation signal (Equation 7).

$$Threshold = 10 \times MAD = 6.745 \times \sigma \quad (7)$$

The chance of incorrect identification can be computed based on the chosen identification threshold. This probability can be computed using the data sampling rate, the period of the investigated data, and the number of earthquake references. In the absence of seismic waves, the likelihood of surpassing this threshold for each sample of the total correlation signal and false identification is calculated using Equation (8) (Warren-Smith et al., 2017).

$$Error_{Onesample} = 1 - erf\left(\frac{6.745}{\sqrt{2}}\right) = 1.53 \times 10^{-11} \quad (8)$$

As a result, by having the likelihood that a sample of the stacked correlation signal is greater than the specified threshold, the overall false detection may be computed simply by multiplying by the total number of samples of the stacked correlation signal arising from the entire process (Equation 9).

$$TotalError = Error_{Onesample} \times npts \quad (9)$$

The npts represents the number of total correlation signal samples collected from each reference earthquake; the total number of samples (npts) is simply calculated from (Equation 10).

$$npts = sps \times days \times hour_{day} \times second_{hour} \times N_{templates} \quad (10)$$

In which, the variables *npts*, *sps*, *days*, $hour_{day}$, $second_{hour}$, and $N_{templates}$ correspond to the total number of points, sampling frequency, the overall duration of the processed data, hours per day, seconds per hour, and the number of reference earthquakes, respectively.

$$npts = 100 \times (95 \times 24 \times 3600) \times 838$$

$$pts = 687,830,400,000$$

As a result, based on the chosen threshold, the calculated chance of incorrect identification for the entire computation leads to 10.5 wrong

detections over the entire period of 95 days of processed data.

4.4. Choosing Minimum Interval between Detections

The presence of duplicated identifications by different references leads to considering the minimum time interval between the adjacent detections to remove duplicate ones. This value is a trade-off between maximizing the number of detections and minimizing duplicate detections. Thus, increasing this value makes nothing wrong, but maybe some detections were lost. In contrast, the remaining earthquakes become more reliable. The duration of this window was chosen based on the almost maximum value present in the S-P frequency distribution in the reference catalog (Figure 6) and set to 10 s.

4.5. Choosing the Filters

The continuous data are filtered within an appropriate frequency range to enhance signal quality, minimize the impact of environmental noise, and improve the similarity of seismic phases between small and large earthquakes (Warren-Smith et al., 2017). This was accomplished by filling the data gaps with zero, filtering the data in the investigated frequency range, and again replacing the previous locations of gaps with zero to remove artificial fluctuations caused by the filter (Chamberlain et al., 2017). The estimated values for the correlation rate will suffer if these low-amplitude oscillations are not addressed.

The process of selecting an efficient filter is highly influenced by the region's surface area, the frequency content of the seismic sources, and, most importantly, the distance of the target earthquakes from the station location. As (McNamara & Buland, 2004) demonstrated, distant earthquakes have low-frequency content, while local earthquakes have high-frequency content. Moreover, the noise level of each station can vary according to its specific environmental factors.

To demonstrate the influence of the employed filter on the identifications, the templates, and continuous data were filtered in several different frequency ranges, and the effect of each of the selected filter intervals on the number and quality of the identifications was explored and compared. To select an appropriate filter for this study, we tested all possible combinations of low and high-frequency limits using four different frequencies for the low limit (0.1, 0.5, 1, and 2 Hz) and four frequencies for the high limit (8, 10, 12, and 15 Hz). In total, we tested 16 different frequencies, and the results were analyzed to identify the most effective filter for the study. The phase detections employing the filters incorporating a lower limit of 0.1 Hz or upper limit of 15 Hz, were of inferior quality. Therefore, the seven filter intervals that contained these two frequencies are discarded. The nine remaining filter intervals (namely 0.5-8, 0.5-10, 0.5-12, 1-8, 1-10, 1-12, 2-8, 2-10, and 2-12 Hz) will be discussed in the subsequent analysis.

1) The correlation value of total identifications (Figure 9): The stronger the total correlation at the detection time, the greater the similarity of the detected earthquake with the phases of the reference earthquake. As a result, the total correlation value can help select an appropriate filter; consequently, according to this parameter, a desirable filter can identify more earthquakes with a higher correlation level. Given that the amount of total correlation varies with the number of phases, the sum of correlation values is normalized by the number of identifying reference phases to ensure that the comparison between references with different numbers of seismic phases is valid. The presence of phase identifications with a total correlation of 1 in all frequency ranges indicates self-identification by the templates. Because the signal of these earthquakes remains identical after filtering and changing the filter parameters, the self-identification of a reference is considered independent of the filter type and should be removed from this analysis. Figure (9) shows that increasing the high corner frequency of the filter reduces the number of identifications with high correlation values while increasing the number of phase identifications with low correlation values. This demonstrates the predominance of noise in the signal at these frequencies, which causes some phases of earthquakes to go undetected and the amount of total correlation to decrease.

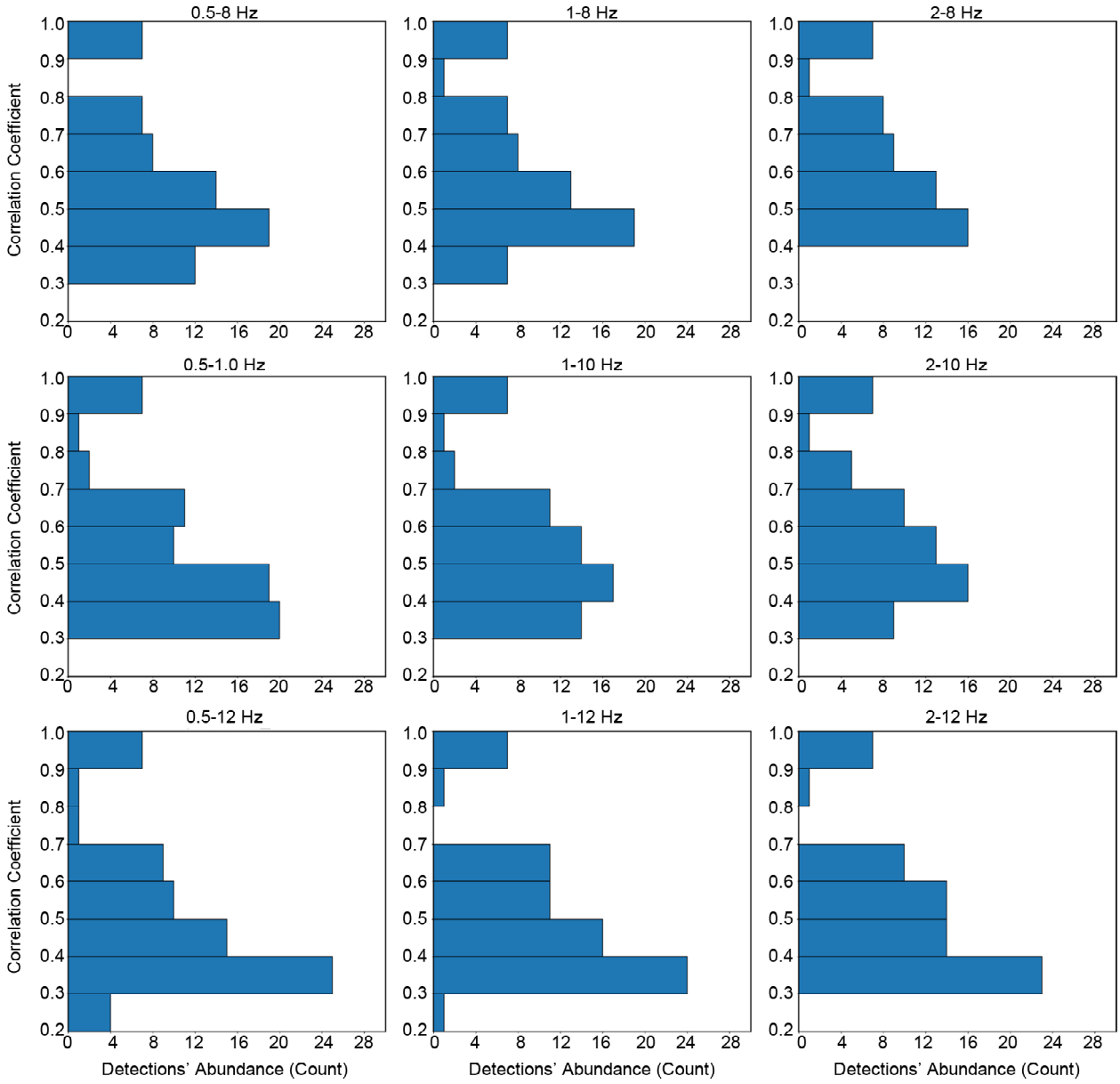

Figure 9. The frequency distribution of the sum of correlation values for phase identifications in the nine frequency intervals is used to select the appropriate filters.

This decrease in correlation can also be seen by lowering the low corner frequency of the used filter.

2) The sum of the correlation of phases identified during the phase determination stage (Figure 10): Because of the spatial difference between the epicenter and the reference earthquake, it is expected that the phases of the detected earthquake will not be completely simultaneous, and the total correlation will decrease during the detection phase. As a result, the sum of the correlations of the detected phases can be used as an additional criterion. Figure (10) shows that many identifications are made employing frequency ranges with an upper limit of 12 Hz and a lower limit of 0.5 Hz. However, they were identified with lower similarity values. The main competition for identifying the seismic phases with the greatest similarity is between the 2 to 8 Hz and 2 to 10 Hz filters, with the frequency of identification occurring in other filters with lower similarities. Furthermore, at this stage, increasing the filter's upper limit and decreasing the filter's lower limit results in a drop in the value of similarities.

3) The ratio of the number of phases to the total number of identifications (Figures 11 and 12): The number and relative position of seismic stations involved in determining an earthquake's location

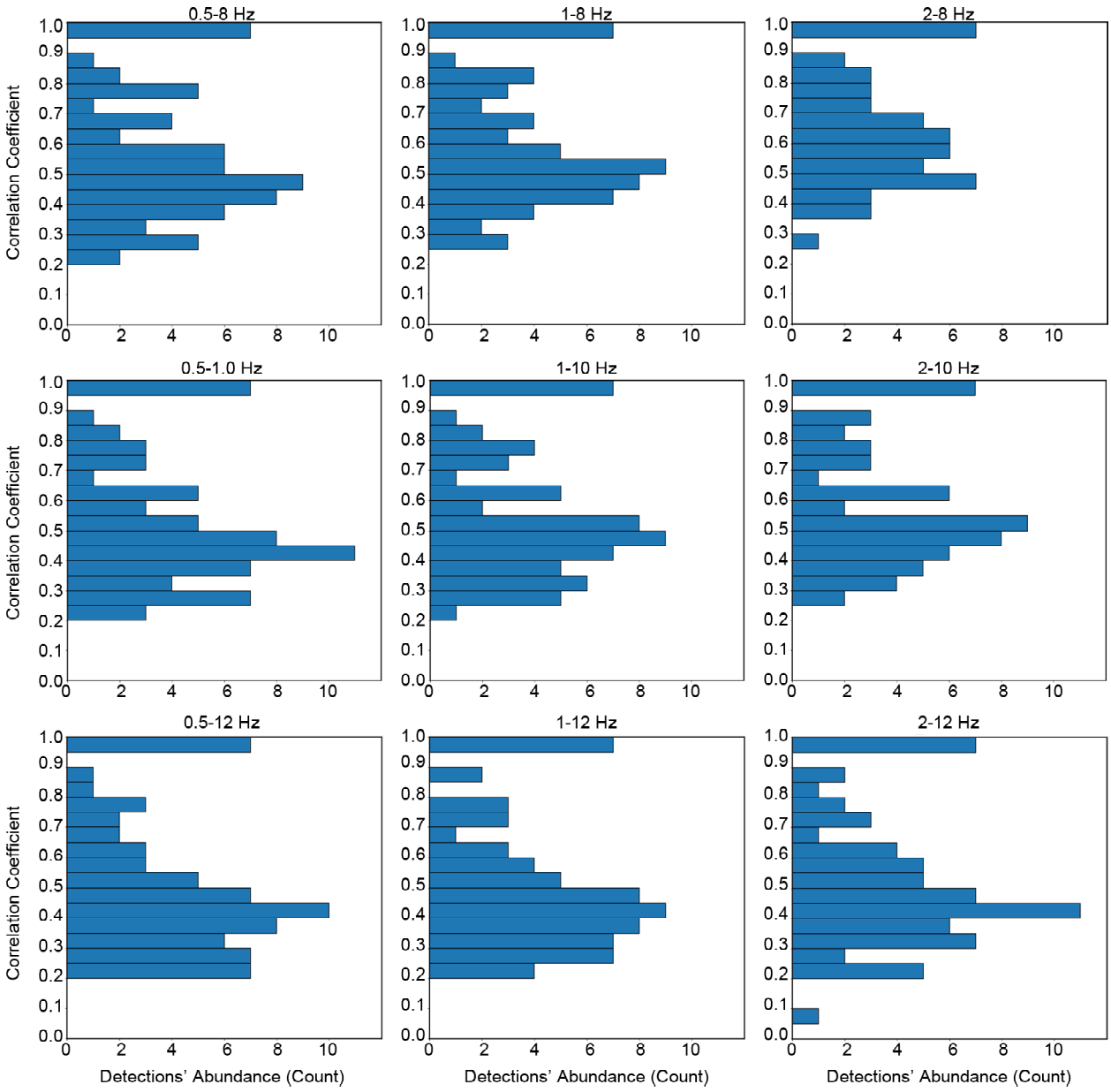

Figure 10. The frequency distribution of normalized correlation coefficients for the sum of seismic phases was identified using the nine frequency intervals. The vertical axis in these graphs represents the normalized value ranges for the correlation of the sum of seismic phases, while the horizontal axis represents the frequency of these values in the respective intervals.

are essential factors in earthquake location quality. It is obvious that increasing the number of seismic phases for an earthquake improves localization accuracy. The number of phases employed for detecting earthquakes was not considered in the investigation for determining the best filter settings by maximizing the total correlation of the detections. An earthquake detected with a high correlation sum in one filter period could have been identified by a small number of high correlation phases or a large number of low correlation phases.

To investigate this issue, bar charts of the frequency distribution of the number of earthquakes versus the number of observed phases are pre- sented, as illustrated in Figure (11). These graphs also show the average number of P, S, and total phases to the number of identifications, with horizontal red, blue, and green lines, respectively. It is observed that by employing filters with frequency limits of 0.5 and 12 Hz, the number of phase identifications has increased (436 total phases), though the events have been detected with fewer phases (ph/ev ratio equal to 6.32). As shown in the frequency distribution plots, the phases detected by employing the 2 to 8 Hz filter have the highest phases pre-events (ph/ev ratio

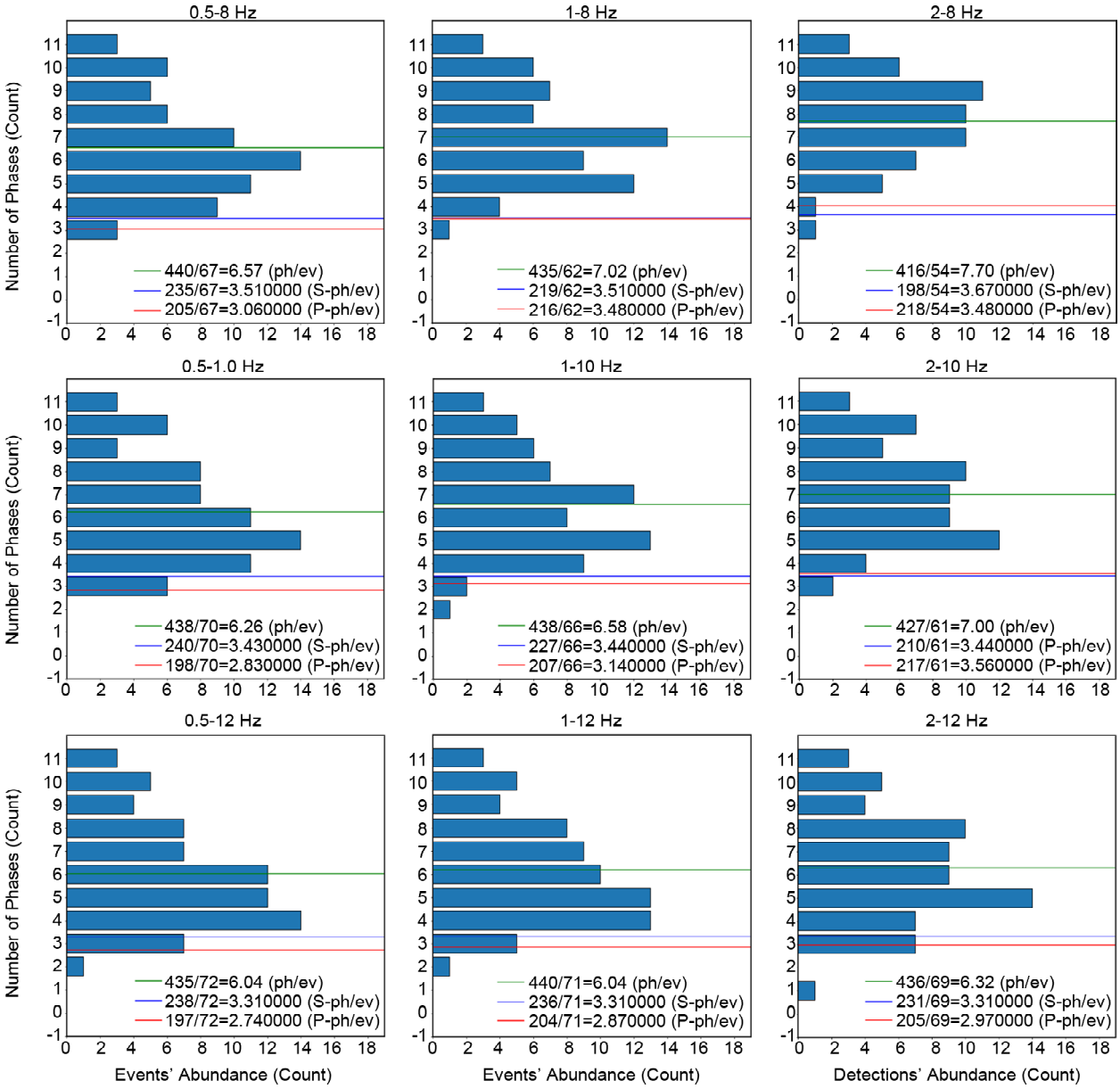

Figure 11. The frequency distribution of detected earthquakes based on the number of phases, divided into frequency filter periods. The vertical axis represents the number of phases per earthquake, while the horizontal axis shows the number of earthquakes with a given phase count. The green, red, and blue lines represent the average total number of phases, P-phases, and S-phases per event, respectively. The boxes on the right and bottom of each panel show the total number of detected earthquakes, P-phases, and S-phases, along with their respective averages.

equal to 7.7). This assessment was repeated after determining the location of earthquakes and deleting erroneous seismic phases and earthquakes with fewer than three P phases and one S phase. The final results are presented in Figure (12). The number of earthquakes removed in each filter period indicates the bad quality or few numbers of the recognized phases. As a result, the frequency range with the most remaining detected events can be considered an acceptable filter.

For example, the comparison of Figures (11) and (12) reveals that the 2 to 8 Hz filter interval had 54 identifications before localization. However, after localization and elimination of inaccurate identifications, the number was reduced to 26, indicating a decrease of approximately 48% in identifications. On the other hand, the 0.5-12 Hz filter interval experienced the greatest reduction in earthquake detections after localization, with a removal of around 71% and a retention of only 29% of the original detections. It is discovered that the 2-8 Hz and 2-10 Hz filters perform better in earthquake identification with a greater number of seismic phases and lower error.

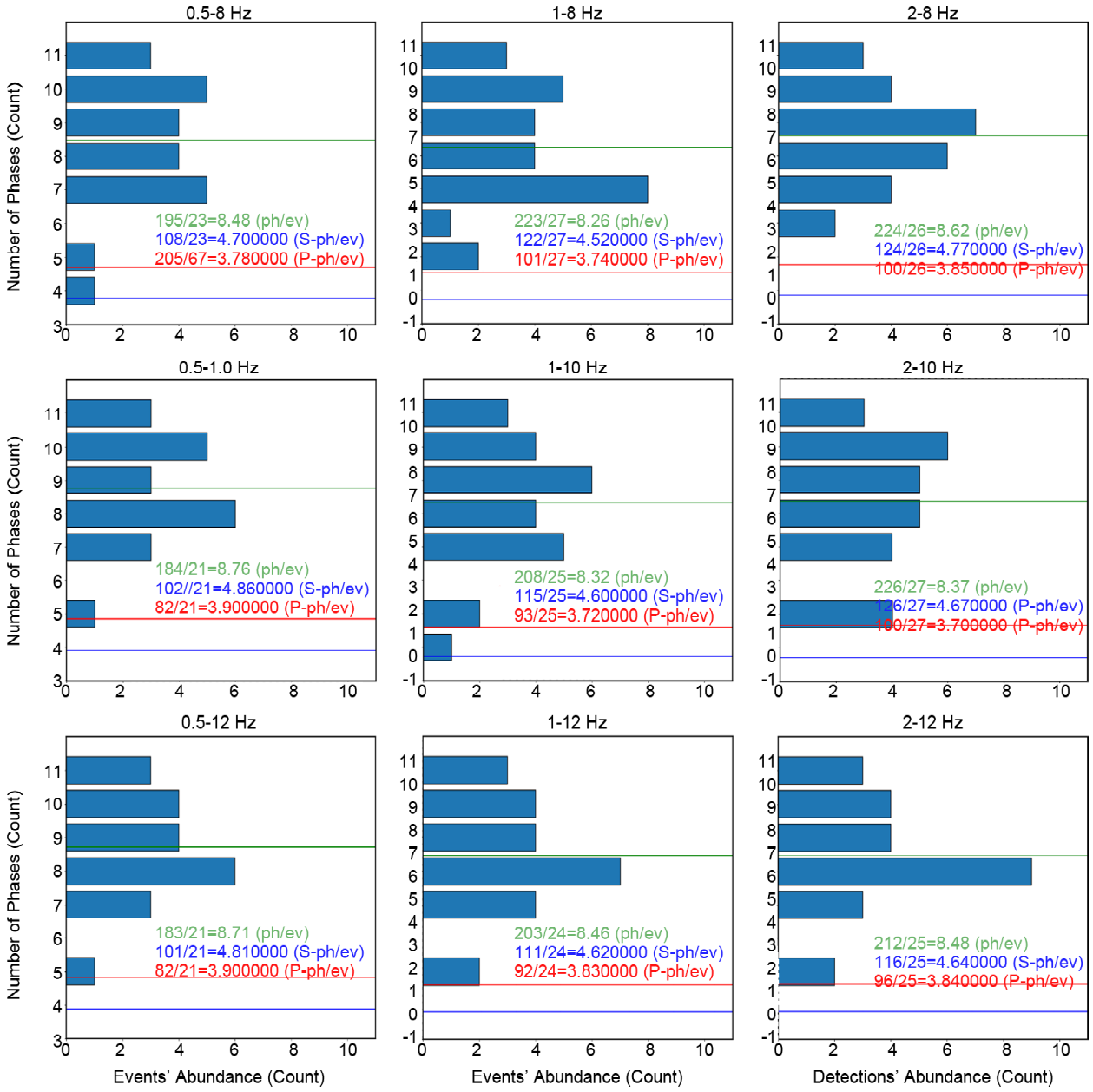

Figure 12. The frequency distribution of the number of detected earthquakes versus the number of their phases after locating and removing outlier seismic phases from the catalog. The vertical axis represents the number of phases of each earthquake, while the horizontal axis represents the number of earthquakes with the desired phase number. In addition, the total number of earthquakes, phases, P phases, and detected S phases are displayed in the boxes on the right and bottom of each frequency period.

4) The Frequency distribution of time errors of identified phases (Figure 13): After the stages of phase detection and location determination, a good criterion for evaluating the quality of the detections is the time residuals of the phases in the detected earthquakes. Figure (13) shows the frequency distribution diagrams of the phase time residuals in each frequency interval. In the frequency range of 2 to 8 Hz, many seismic phases are detected with a residual between -1 to 1 second. Furthermore, the number of seismic phases with considerable time residuals has decreased compared to results obtained using other filter ranges.

5) The Frequency distribution of location error (Figures 14 and 15): The location error of the detected earthquakes is an important parameter since the correct representation of seismic trends is affected by the earthquake location accuracy. Therefore, the frequency distributions of horizontal and vertical location errors for the detections employing the nine frequency intervals are investigated. Estimating the horizontal location error of

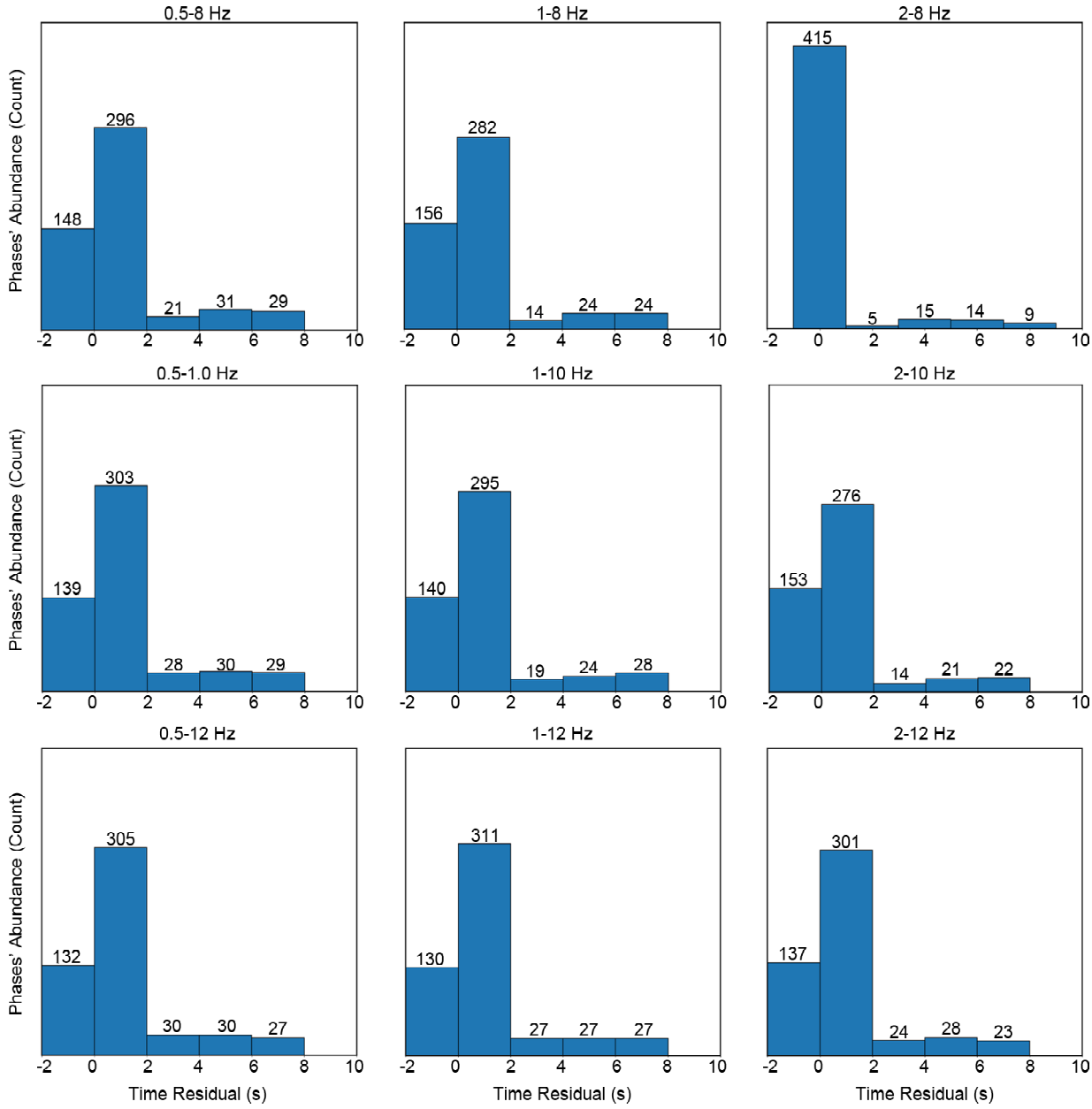

Figure 13. The frequency distribution of the time residuals of the phases of all observed earthquakes in each of the nine analyzed frequency intervals.

the earthquakes reveals a similar quality for various frequency ranges (1-8, 2-8, 1-10, and 2-10 Hz).

However, the number of earthquakes with vertical and horizontal location error of more than 2 km is less in the frequency range of 2 to 8 Hz, and using this filter seems suitable (Figures 14 and 15).

Based on the findings of these experiments, the 2 to 8 Hz filter was selected for this study, and all continuous data and references were filtered in this frequency range. However, as previously indicated, the selection of appropriate filter parameters is dependent on network configuration and the relative location of the network and earthquakes. Further investigation reveals that using frequencies less than 1 Hz in a reference with a length of one second is naturally meaningless because the length of the selected references is smaller than the period of the waves, explaining why frequencies between 0.1 and 0.5 Hz must be filtered. Also, due to the high level of ambient noise in the region, the high-frequency content of the signal has a destructive influence on performance, and it was discovered that the filter with a high limit of 8 Hz performs better on the analyzed data.

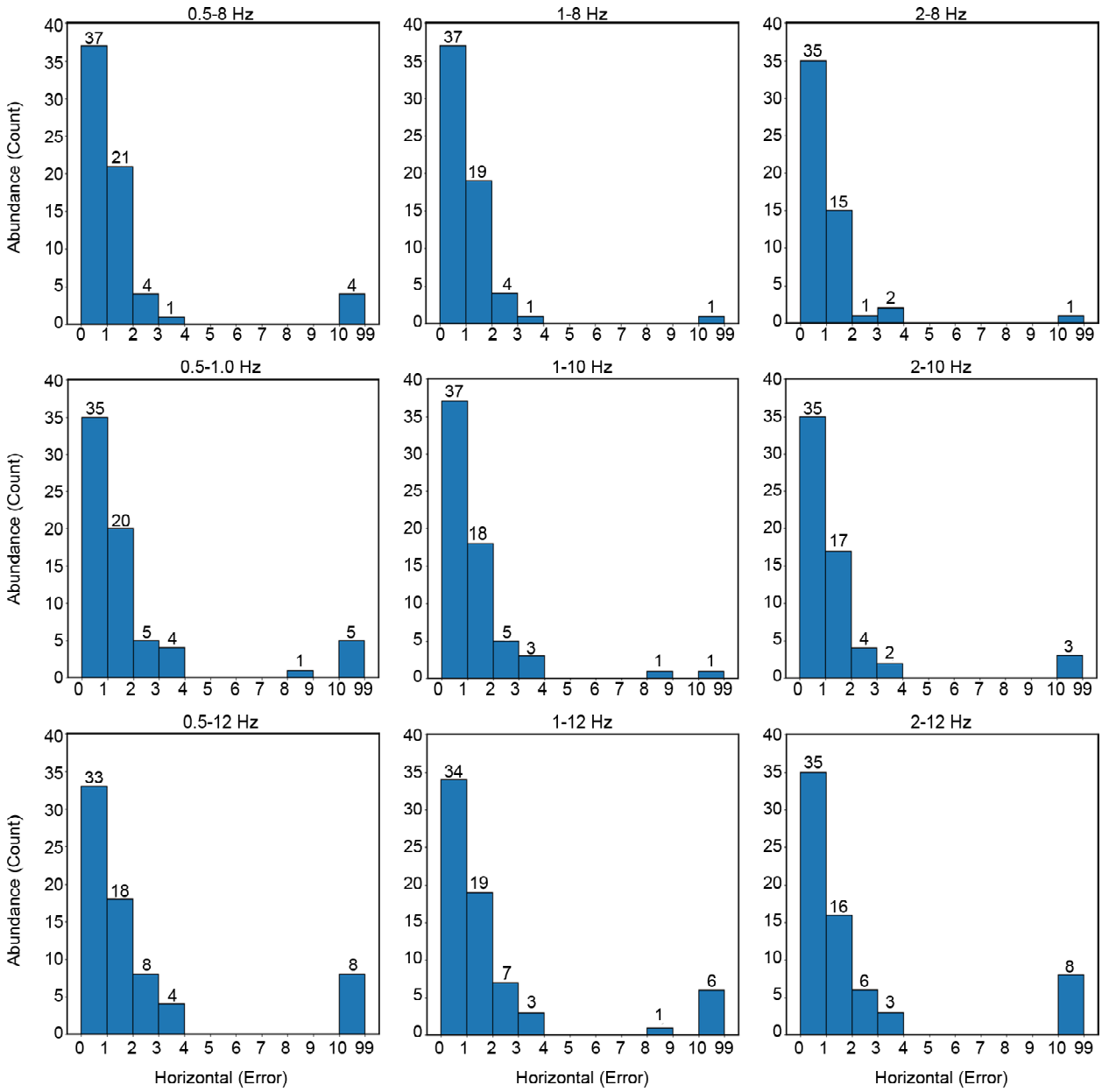

Figure 14. The frequency distribution of horizontal positioning error of identifications in the nine frequency intervals. The horizontal axis in these figures represents distinct ranges of surface error, whereas the vertical axis represents the frequency of horizontal error in this range.

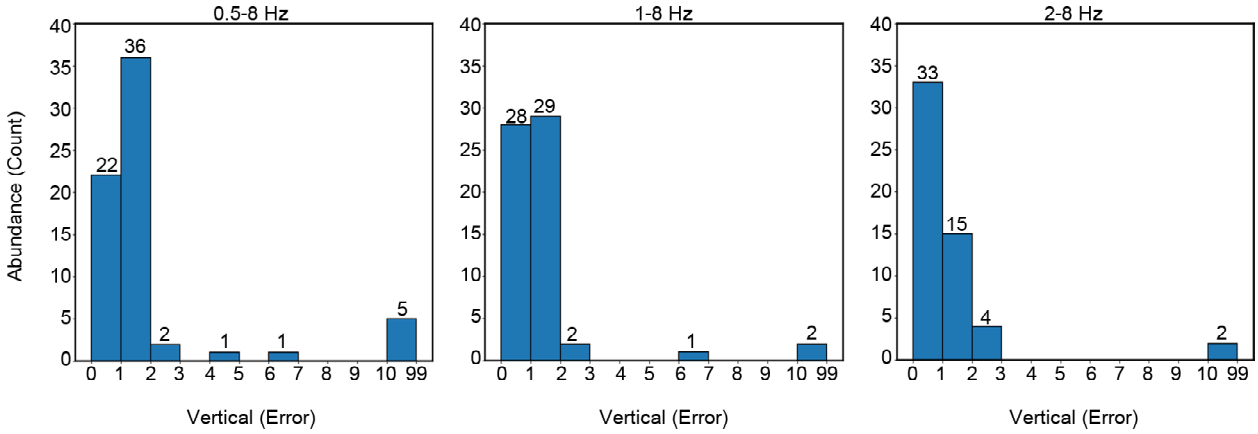

Figure 15. The frequency distribution of the vertical positioning error of identifications in the nine frequency bands. The horizontal axis in these figures represents different ranges of vertical errors, while the vertical axis represents its frequency within this range.

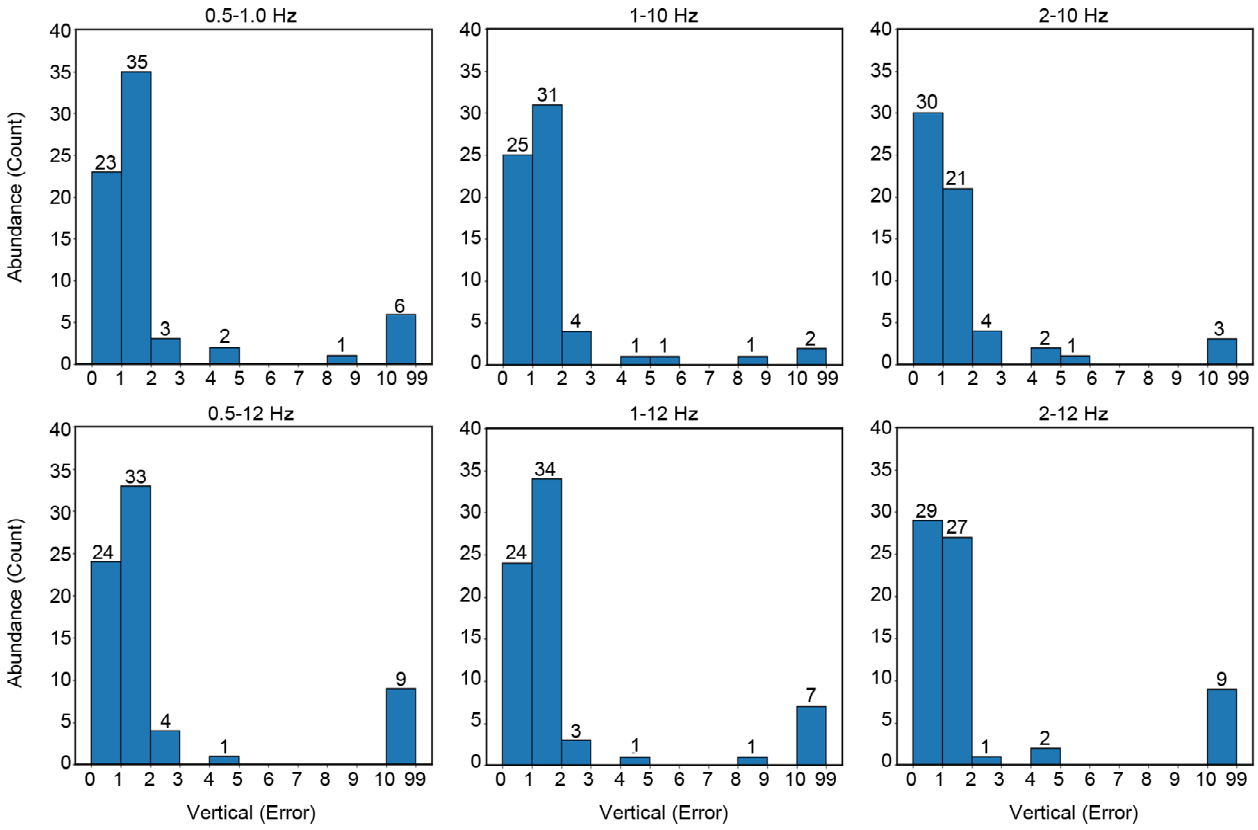

Figure 15. Continue.

5. Result for Aftershocks Detection in Ilam Region

In this study, the optimal values of the influencing parameters to use the Matched Filter technique were evaluated and selected in the case of the local seismograph network of Ilam. The template duration was chosen based on the idea of obtaining sparse P and S reference phases and so set to 1 second according to interevent station distances. This value seems appropriate considering the frequency content of the local seismic waves and the presence of more than one period of the seismic phase in one data frame.

The pre-phase parameter of earthquake templates is necessary to decrease the uncertainty of the phase unset and is set to 0.1 second in order not to lose the desirable ability to identify the P phases compression of S phases.

The detection stage is performed on the stacked correlation signals using the threshold equal to the multiplication of a constant and MAD of their stacked correlations, specific for each template and day. Selecting the coefficient is a trade-off between maximizing the number of detections and minimizing the rate of false detections. So, the value depends on the target of the study. In this study, ten times MAD was used as the threshold. Statistical investigation showed that this threshold could lead to about 10 or 11 false detections in 95 days which is an acceptable rate.

The occurrence of numerous identifications by unique references is eliminated if there is a minimal time gap between adjacent detections. Based on the most commonly observed value in the S-P time pairs, this parameter was set to 10 s.

Because of the extensive range of selectable low-cut and high-cut frequencies and the influence of selecting each pair on the result, selecting an appropriate filter is the most challenging element of parameter selection. Statistical evaluation was performed on the correlation value of detection, the sum of correlation values in phase picking, and location parameters like time residual and epicentral error to evaluate the functionality of each frequency pair. As described, the best results are obtained using the 2 to 8 Hz and 2 to 10 Hz filters with slight differences, but because of the superior performance of the 2-8Hz filter in phase detection, it was chosen as the best filter range.

Finally, after finding the optimum parameters for using the matched-filter technique, 3575 aftershocks were identified (4.27 times the references) in the 95 days of continuous data of the temporary seismographic network by 838 reference earthquakes. Figure (16) shows the location and distribution of the detected aftershocks by matched-filter techniques. Figure (17) provides a summary of the location parameters for these aftershocks and demonstrates that the majority of detections were made using 3 to 6 stations and 5 to 10 identified seismic phases, with time

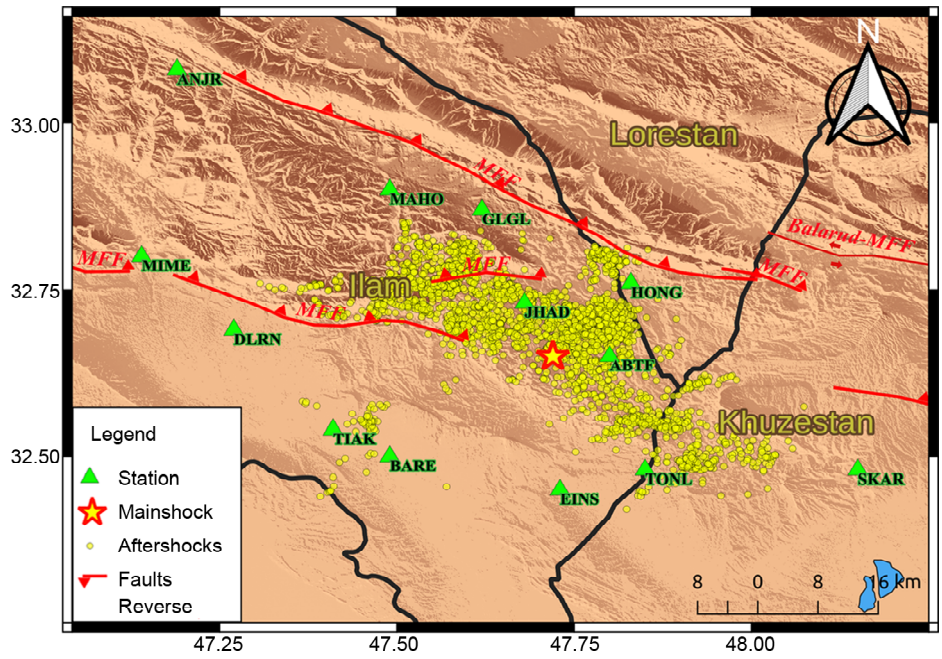

Figure 16. The map shows the locations of 3575 aftershocks detected by the Matched-Filter technique, local faults, temporary monitoring network stations, and the main shock.

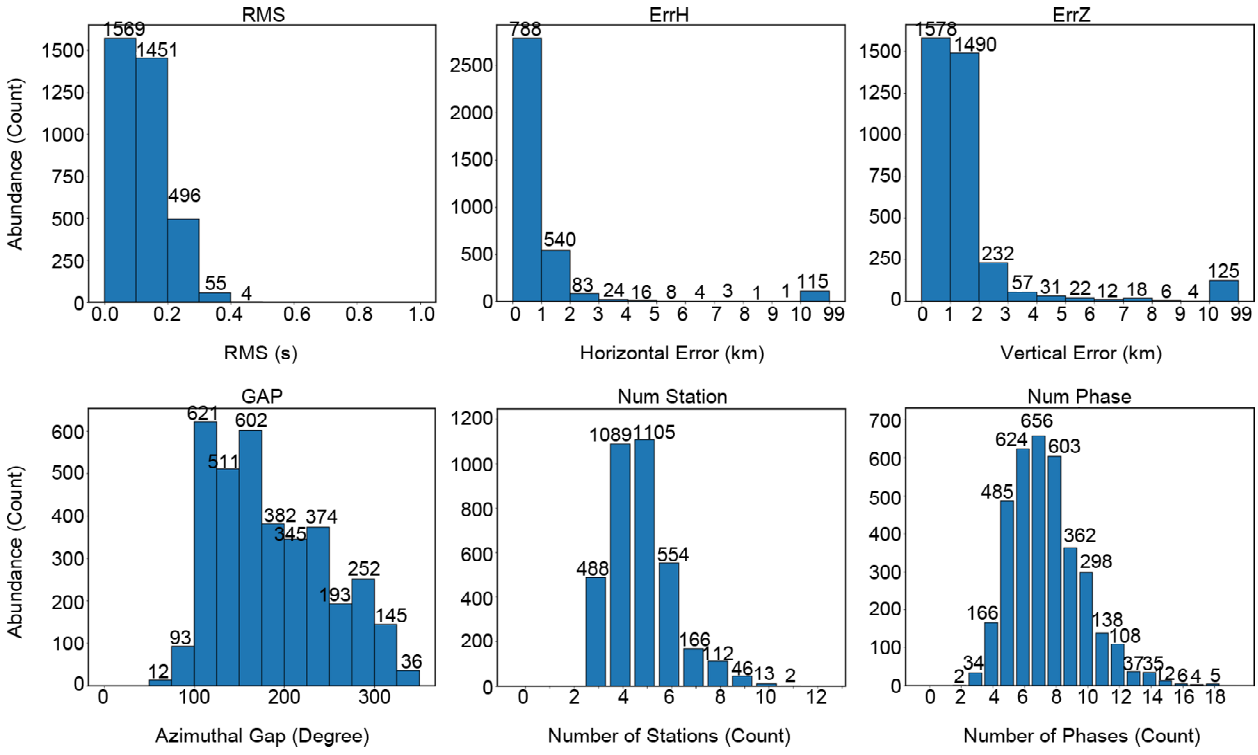

Figure 17. Frequency distribution of time residual values, horizontal error, vertical error, azimuth gap, number of stations, and number of phases of 3575 aftershocks identified by adaptive filter approach.

Table 2. The classification of 3575 detected aftershocks into four groups based on location quality.

Group	RMS	GAP	Num Stations	ErrH	ErrZ	Rate
A	rms ≤ 0.3	gap ≤ 180	6 ≤ n-st	ErrH ≤ 3	ErrZ ≤ 3	637
B	rms ≤ 0.4	gap ≤ 200	5 ≤ n-st	ErrH ≤ 5	ErrZ ≤ 5	1436
C	rms ≤ 0.5	gap ≤ 250	5 ≤ n-st	ErrH ≤ 7	ErrZ ≤ 7	1768
Others	rms > 0.5	gap > 250	3 ≤ n-st	ErrH > 7	ErrZ > 7	1807

residuals of less than 0.2 s, lateral errors of less than 1 km, and depth errors of less than 2 km. Furthermore, the azimuth gap of these identifications ranges from 50 to 300 degrees, with more than half relating to an azimuth gap of fewer than 180 degrees.

The location quality of the detected aftershocks is evaluated using the effect of the Azimuthal gap, Number of stations and phases, and Nearest distance station (Bondar et al., 2004). The results are shown in Table (2) with the same criteria used in Table (1) to evaluate the performance of the matched-filter method. It is discovered that, despite the low values of time residual, epicentral error, and depth error, fewer earthquakes were placed in group A, because of the number of phases in templates earthquakes. However, as compared to the reference catalog, there is an increase in the frequency of good and medium-grade located events, confirming the method's efficiency. The matched-filter identifier outperforms the other identifiers in identical conditions, and when an appropriate network is installed, the majority of identifications will be performed with the required high quality.

6. Discussions and Conclusions

The strength of the similarity filter was tested in this study, and it was discovered that more earthquakes may always be detected using this method. The obtained results show that the matched-filter detector has a high detection power in detecting aftershocks with a low signal-to-noise ratio, and the obtained results show a significant increase in the number of identifications made while the network is entirely unchanged. Using the matched-filter detector and 838 reference earthquakes, 3575 aftershocks with at least three P phases and one S phase were discovered in this investigation. This figure is 4.27 times the number of aftershocks identified on this temporary seismographic network using energy detectors or by the user. This was accomplished by calculating 838 total correlation signals per day and 79610 total correlation signals in 95 days of continuous data. Considering the parameters used in this study, such as the number of stations, the number of references, and the data sampling rate, the time required to use the matched filter using a computer with ten processing cores and 32 GB of RAM is less than 48 hours. Using this technology is much more cost-effective than increasing the number of stations; hence, it should be prioritized.

The ability and performance of the waveform cross-correlation detector compared to the energy detectors were demonstrated in this study, and the capacity of this technique to reliably recognize tiny earthquakes and investigate the seismicity of the region is highly remarkable. The various parameters that influence the use of this method were described, as well as their efficiency and optimal selection. The length of seismic references necessary for similarity measurement in continuous signals was investigated, and it was theoretically explained that this length is dependent on the frequency content of the recorded signal, or in other words, by the distance between the earthquake and the seismic station. The importance of independence between the P and S references at the exact moment of the seismic phase was also emphasized. Next, the use of the seismic pre-phase for reference selection was investigated to reduce uncertainty in phase detection, and its effect on identification power was assessed. It was revealed that an increase in the duration of the seismic pre-phase in references collected from the P and S seismic phases affects the identification power due to the different nature of the pre-phase signal. It was also demonstrated that increasing the pre-phase time decreases P-phase detection power while increasing S wave detection power. It is also clear

that the quality of the phases identified by this method is comparable to the seismic phases of the reference earthquakes utilized, and an incorrect seismic phase in the template will result in similar errors in the new detections. As a result, the significance of the seismic phase database's completeness and quality for future investigations, as well as the development and improvement of new algorithms.

Another parameter that was investigated is the minimum time interval between identifications. If the minimum time interval between identifications is not taken into consideration, more than 15 thousand events would be detected, emphasizing the importance of successfully taking this parameter into account in the matched-filter technique for correct event identifications.

Furthermore, the role of filter selection was investigated, and it was discovered that when an inappropriate filter interval is employed, multiple identifications are obtained, but the majority of these identifications have a limited number of phases, making determining the location of the event impossible. Events with the required number of phases, strong correlation, and minimal localization error are detected by selecting a suitable filter interval.

Acknowledgments

The authors would like to express their gratitude to the International Institute of Earthquake Engineering and Seismology for providing the seismic network data set. We would also like to thank Dr. Mohammad Tatar and Ms. Mahbobeh Siyamack for supplying the manual earthquake catalog and the velocity model of the study area.

References

Adushkin, V., Bobrov, D., Kitov, I., Rozhkov, M., & Sanina, I. (2017). Remote detection of aftershock activity as a new method of seismic monitoring. Doklady Earth Sciences,

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A Python toolbox for seismology. *Seismological Research Letters*, *81*(3), 530-533.

Bobrov, D., Kitov, I., & Zerbo, L. (2014). Perspectives of cross-correlation in seismic monitoring at the international data centre. *Pure and Applied Geophysics*, *171*, 439-468.

Bondar, I., Myers, S.C., Engdahl, E.R., & Bergman, E.A. (2004). Epicentre accuracy based on seismic network criteria. *Geophysical Journal International*, *156*(3), 483-496.

Chamberlain, C.J., Boese, C.M., & Townend, J. (2017). Cross-correlation-based detection and characterisation of microseismicity adjacent to the locked, late-interseismic Alpine Fault, South Westland, New Zealand. *Earth and Planetary Science Letters*, *457*, 63-72.

Chamberlain, C.J., Hopp, C.J., Boese, C.M., Warren-Smith, E., Chambers, D., Chu, S.X., Michailos, K., & Townend, J. (2018). EQcorrscan: Repeating and near-repeating earthquake detection and analysis in Python. *Seismological Research Letters*, *89*(1), 173-181.

Chamberlain, C.J., Shelly, D.R., Townend, J., & Stern, T.A. (2014). Low-frequency earthquakes reveal punctuated slow slip on the deep extent of the Alpine Fault, New Zealand. *Geochemistry, Geophysics, Geosystems*, *15*(7), 2984-2999.

Dodge, D. & Walter, W. (2015). Initial global seismic cross?correlation results: Implications for empirical signal detectors. *Bulletin of the Seismological Society of America*, *105*(1), 240-256.

Frank, W.B. & Abercrombie, R.E. (2018). Adapting the matched-filter search to a wide-aperture network: an aftershock sequence and an earthquake swarm in connecticut short note. *Bulletin of the Seismological Society of America*, *108*(1), 524-532.

Gibbons, S.J. & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, *165*(1), 149-166.

Gibbons, S.J., Sorensen, M.B., Harris, D.B., & Ringdal, F. (2007). The detection and location of low magnitude earthquakes in northern Norway using multi-channel waveform correlation at regional distances. *Physics of the Earth and Planetary Interiors*, *160*(3-4), 285-309.

Lahr, J. (1999). *Revised 2012, HYPOELLIPSE: A*

Computer Program for Determining Location Earthquake Hypocentral Parameters, Magnitude, and First-Motion Pattern. US Geological Survey Open-File Report, 99-23.

McNamara, D.E. & Buland, R.P. (2004). Ambient noise levels in the continental United States. *Bulletin of the Seismological Society of America*, *94*(4), 1517-1527.

Schaff, D.P. & Richards, P.G. (2014). Improvements in magnitude precision, using the statistics of relative amplitudes measured by cross correlation. *Geophysical Journal International*, *197*(1), 335-350.

Schaff, D.P. & Waldhauser, F. (2010). One magnitude unit reduction in detection threshold by cross correlation applied to Parkfield (California) and China seismicity. *Bulletin of the Seismological Society of America*, *100*(6), 3224-3238.

Shelly, D.R. & Hill, D.P. (2011). Migrating swarms of brittle-failure earthquakes in the lower crust beneath Mammoth Mountain, California. G*eophysical Research Letters*, *38*(20).

Vuan, A., Sugan, M., Amati, G., & Kato, A. (2018). Improving the detection of low-magnitude seismicity preceding the Mw 6.3 L'Aquila Earthquake: development of a scalable code based on the cross correlation of template earthquakes improving the detection of low-magnitude seismicity preceding the Mw 6.3 L'Aquila Earthquake. *Bulletin of the Seismological Society of America*, *108*(1), 471-480.

Waldhauser, F. & Ellsworth, W. L. (2000). A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California. *Bulletin of the Seismological Society of America*, *90*(6), 1353-1368.

Warren-Smith, E., Chamberlain, C.J., Lamb, S., & Townend, J. (2017). High-precision analysis of an aftershock sequence using matched-filter detection: The 4 May 2015 ML 6 Wanaka earthquake, Southern Alps, New Zealand. *Seismological Research Letters*, *88*(4), 1065-1077.

Zhang, M., & Wen, L. (2015). An effective method for small event detection: Match and locate (M & L). *Geophysical Journal International*, *200*(3), 1523-1537.