Vietnam Journal of Earth Sciences, 43(3), 430-446, https://doi.org/10.15625/2615-9783/17253

Vietnam Academy of Science and Technology
Vietnam Journal of Earth Sciences
http://www.vjs.ac.vn/index.php/jse

Automatic earthquake detection and phase picking in Muong Te, Lai Chau region: an application of machine learning in observational seismology in Vietnam

Nguyen Cong Nghia[1,2], Nguyen Van Duong[3,4*], Ha Thi Giang[3], Dinh Quoc Van[3], Nguyen Le Minh[3,4,5], Bor-Shouh Huang[6], Pham The Truyen[3], Nguyen Tien Hung[3,4], Le Quang Khoi[3], Nguyen Huu Hung[7]

[1]*Taiwan International Graduate Program Earth Sciences System (TIGP-ESS), Academia Sinica, Taiwan*
[2]*Department of Earth Sciences, National Central University, Taiwan*
[3]*Institute of Geophysics, VAST, Hanoi, Vietnam*
[4]*Graduate University of Science and Technology (GUST), VAST, Hanoi, Vietnam*
[5]*Department of International Cooperation, VAST, Hanoi, Vietnam*
[6]*Institute of Earth Sciences, Academia Sinica, Taipei, Taiwan*
[7]*Department of Natural Science Hung Vuong University, Viet Tri, Phu Tho, Vietnam*

Received 04 January 2022; Received in revised form 11 June 2022; Accepted 01 July 2022

ABSTRACT

We applied the automatic detection and picking of P- and S-wave to one-year continuous raw seismic data from 17 seismic stations in the Muong Te area, northwestern Vietnam. The deep learning picker, Earthquake Transformer, has performed automatic picking of P- and S-waves, and phase association, then we located the earthquakes using Hypoinverse and NonLinLoc programs. The newly derived catalog consisted of 893 events, which is significantly higher than the number of events in the manual catalog. From this new catalog, we can observe more earthquakes related to the Muong Te M_L 4.9 earthquake on June 16, 2020, and the earthquake activity in other faults such as the Dien Bien Phu and Muong Nhe faults. The extended catalog can further study the seismogenesis and the seismic velocity structure of the crust beneath northwestern Vietnam.

Keywords: Muong Te earthquake, machine learning, EQ Tranformer, Dien Bien Phu fault, upstream Da river fault, earthquake monitoring.

1. Introduction

In recent years, the rapid development of machine learning algorithms, especially neural networks and deep learning, has allowed for new improvements in various fields of science. Recent advances in machine learning algorithms have also been applied to the seismological aspects, as outlined by Kong et al. (2019). Among the various applications of neural networks and deep learning, automatic earthquake detection and phase picking are significant problems in modern seismology, which are currently attracting attention from seismologists. Earthquake detection and phase picking can be considered one of the first subjects of seismology study since digital seismometers

*Corresponding author, Email: duongnv@igp.vast.vn

were used, with the first algorithm being the well-known short-term average/long-term average (STA/LTA) method (Allen, 1978). This algorithm is still widely used because of its effectiveness and ease of application to seismic data compared to other methods (Trnkoczy, 2009). Other traditional pickers follow a similar approach to detect earthquakes: define a function to represent the features of incoming earthquake signals and use a preexisting threshold to detect and pick the phase (Baer and Kradolfer, 1987; Baillard et al., 2014; Cichowicz, 1993; Saragiotis et al., 2002; Sleeman and Van Eck, 1999).

In contrast to these methods, machine learning pickers, which use deep learning in the algorithm, do not define a specific function but rely on the "training" process from a large amount of data, i.e., hundreds of thousands to millions of waveforms. The three notable deep learning pickers are a general phase generalized seismic phase detection (GPD) (Ross et al., 2018), PhaseNet (Zhu and Beroza, 2019), and the current state-of-the-art Earthquake Transformer (EQT) (Mousavi et al., 2020). These deep-learning-based pickers have improved the quantity and quality of earthquake detection, and phase picks in various studies (Liu et al., 2020; Wang et al., 2020; Xiao et al., 2021; Zhou et al., 2021). Moreover, a new study evaluating the performance of deep learning pickers indicated that the Earthquake Transformer could be considered the most reliable picker for local earthquakes within an epicentral distance of about 350 km (Münchmeyer et al., 2021). In addition to machine learning pickers using global datasets, a neural-network-based picker was developed using the local seismic data in the Lai Chau area, northwestern Vietnam. However, owing to the small number of events, the model shows a relatively low accurate positive accuracy (Wiszniowski et al., 2021).

On 16 June 2020, a moderate-sized earthquake of local magnitude (M_L) 4.9 struck the Muong Te area, located in the northwestern part of Vietnam. According to the compiled seismicity catalog in the Vietnam region (Nguyen et al., 2019), this was one of the strongest earthquakes in the Muong Te block over 100 years ago. The main shock has caused some damage to infrastructure and public panic in the area near the source (Fig. 1). This earthquake was preceded and followed by several foreshocks and aftershocks with magnitudes greater than 1.0. The earthquake sequence was close to the junction of the Upstream Da River fault, Muong Te fault, and several sub-meridian faults (Fig. 1). In the Muong Te block, the most notable fault in the region is the N-S trending Dien Bien Phu fault which is considered a boundary between the Simao and Indochina blocks. In addition, the Muong Te region comprises three other remarkable rank II faults, including the Upstream Da River fault, Muong Te fault, and Muong Nhe fault (Nguyen et al., 2019). Owing to the presence of these faults, a more completed earthquake catalog is needed to deepen the understanding of seismogenic fault zones in the Muong Te block and adjacent areas.

In this study, we applied a deep-learning picker - the EQT phase picker (Mousavi et al., 2020) to build a more completed earthquake catalog for the Muong Te region. The picker was applied to continuous waveform data, and the phase picks were associated with forming an event, followed by locating the event using traditional locating methods. The number of earthquakes in the new catalog has increased by at least 13 times compared to the original catalog, which reveals that the fault system was more active than in previous findings. In a further study, the new catalog provides an opportunity to study in more detail the seismogenic structure, fault geometry, and seismic hazard assessment for northwestern Vietnam.

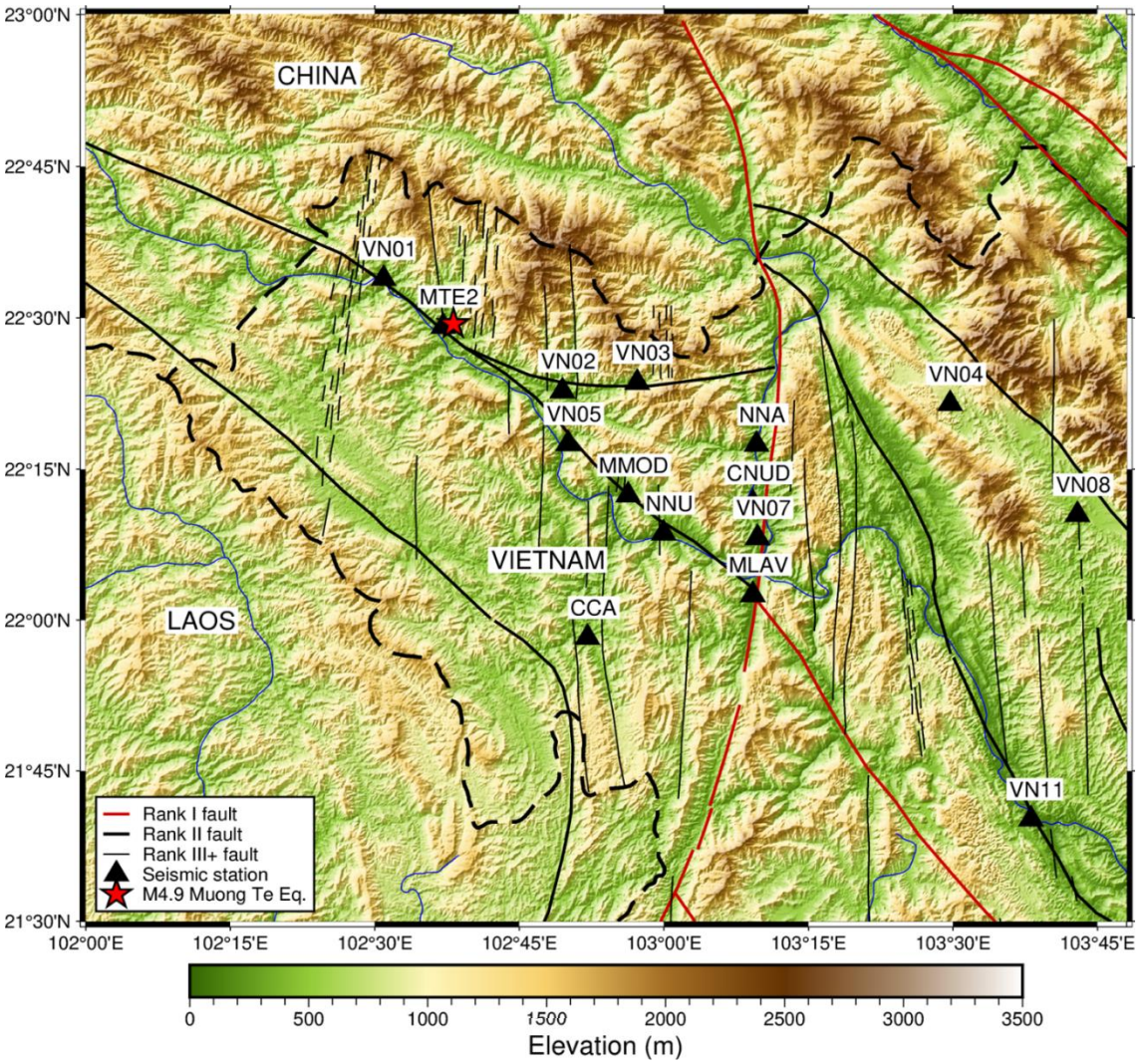


Figure 1. Map showing the study region; the faults and their ranks are from Nguyen et al. (2019); black triangles indicate the seismic stations used in this study

2. Data and Method

2.1. Data

2.1.1. Seismic data

We have collected the 3-component seismic waveforms from 17 Muong Te area seismic stations from January to December 2020 (Fig. 1 and Table 1). Among these seismic stations, three stations are equipped with highly short-period seismometers (CNUD, MMOD, MLAV), while the rest are equipped with broadband seismometers (MTE, MTE2, MLVB, CCA, NNA, NNU, MLVB, and stations start with "VN"). Most seismic stations were permanently or temporarily deployed before the Muong Te earthquake, so they operated for the entire experiment time. In contrast, station MTE2 was deployed after the earthquake by the Institute of Geophysics (IGP) - Vietnam Academy of Science and Technology (VAST). As described by Mousavi et al. (2020), the EQT phase picker

was built using data from various types of seismometers, including those used in our study, so the selector would be expected to work well with our data.

Table 1. List of the seismic stations used in this study

No.	Station name	Seismometer type	Latitude (degree)	Longitude (degree)	Elevation (m)	Total detections (T)	Arrival picks (A)	A/T (%)	Distance from the mainshock (km)
1	CCA	Broadband	21.971	102.868	435	7608	518	6.81	61
2	CNUD	Short period	22.194	103.154	237	765	746	97.52	65
3	MLAV	Short period	22.042	103.154	232	573	212	37.00	74
4	MLVB*	Broadband	22.042	103.154	232	1241	495	39.89	74
5	MMOD	Short period	22.207	102.938	395	1360	266	19.56	46
6	MTE	Broadband	22.379	102.825	349	9617	876	9.11	26
7	MTE2	Broadband	22.486	102.618	244	4115	330	8.02	4
8	NNA	Broadband	22.291	103.16	281	9286	1194	12.86	62
9	NNU	Broadband	22.144	103.00	330	7242	1017	14.04	55
10	VN01	Broadband	22.565	102.515	373	1265	89	7.04	14
11	VN02**	Broadband	22.379	102.824	329	253	1	0.40	26
12	VN03	Broadband	22.393	102.954	544	73	47	64.38	38
13	VN04	Broadband	22.358	103.495	755	1318	282	21.40	94
14	VN05	Broadband	22.292	102.834	38	67	6	8.96	31
15	VN07	Broadband	22.136	103.162	751	2410	475	19.71	69
16	VN08	Broadband	22.175	103.715	499	872	255	29.24	120
17	VN11	Broadband	21.67	103.635	231	895	73	8.16	139
	Total					41807	7006	16.76	

* same location with station MLAV

** same location with station MTE

2.1.2. Earthquake catalog

Before this study, the IGP monitored seismicity in the region and compiled an earthquake catalog. The catalog was built using a routine workflow consisting of three main steps: (1) earthquake detection, (2) manual phase picking and storing the data in the SEISAN format (Havskov and Ottemoller, 1999), and 3) locating the earthquake using traditional locating methods, e. g. Hypoinverse (Klein, 2002) or Hypo71 (Lee and Lahr, 1972). The catalog consists of 67 events with local magnitudes (M_L) of $1.0 \leq M_L \leq 4.9$. To distinguish this catalog from the resulting earthquake catalog from this study, we named this catalog MTMAN, whereas the new catalog was named MTEQT. To analyze the use of the new method in creating the earthquake catalog, it is necessary to compare MTMAN and MTEQT based on the following criteria: the number of earthquakes, the source parameters (location and original time) of the earthquakes that are in both catalogs, and the phase pick. This information and the result of this comparison will be presented in the discussion part.

2.2. Method

2.2.1. The EQT picker

The EQT is a complicated deep learning model for earthquake detection and phase picking with excellent performance when applied to local earthquakes (Mousavi et al., 2020). The model was built using the STanford EArthquake Dataset (STEAD), which contains millions of labeled waveforms that recorded local earthquakes within about 300 km from various instruments (Mousavi et al., 2019). In this study, the network structure of the EQT model is summarized and simplified in Fig. 2, which includes a multilayer encoder and three different decoders. The encoder transforms the three-component waveforms into high-level data

that the model can recognize, while the decoders are used to extract the features and return the output. In this model, the output is expressed in the form of three distinct probabilities: earthquake detection, P-phase, and S-phase. Readers may refer to Ross et al. (2018), Zhu and Beroza (2019), and Mousavi et al. (2020) for further detailed explanations of how the neural network is designed to process the seismic waveform data. In short, the neural network used in these studies is inspired by the neural network commonly used in image processing, e.g., the convolution neural network (CNN) (LeCun and Bengio, 1995), long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) and transformer-attention (Vaswani et al., 2017) because, in computer science, three-component seismic waveforms (vertical and two horizontal components) can be processed similarly to RGB (red, green, blue) images.

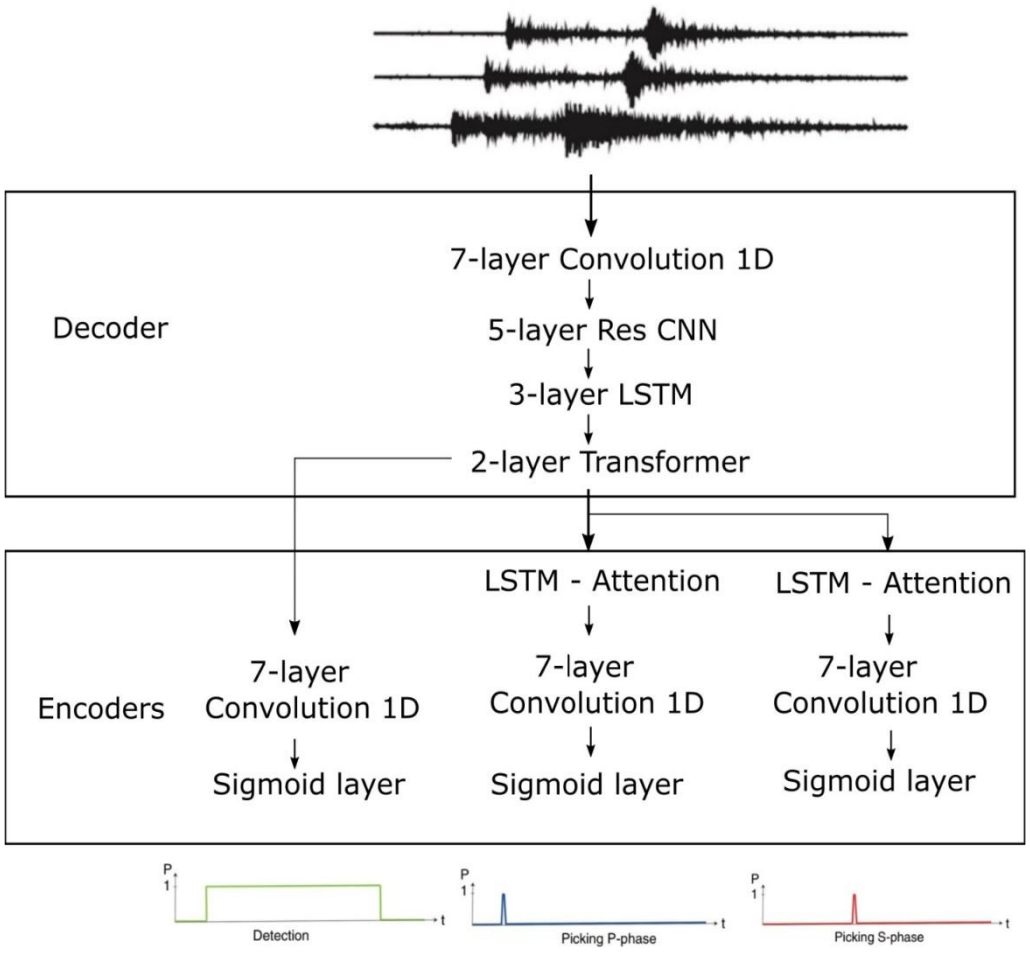


Figure 2. Simplified the network structure of the EQT model; an explanation of the network structure is described in the text

The EQT is specialized in detecting and picking the local earthquakes, therefore is suitable for seismic waveforms recorded close to the earthquakes with a distance to the hypocenters of less than 300 km. The training data of the EQT model is the STEAD dataset, in which the authors have used approximately 1 million earthquake waveforms and 300

thousand noises recorded by seismic stations with a variety of the source mechanisms, tectonic regions, focal depths, magnitude sizes, and seismic instrument types (Mousavi et al., 2019). 113 thousand test waveforms have been used to evaluate the detection ability of the EQT phase picker, which resulted in 1 false positive and 0 false negatives. In addition, the model also showed minimal differences in P- and S-phase picks compared to manual picking (Mousavi et al., 2020). Based on these results and characteristics, we chose the EQT model to detect and pick the seismic phase from the local seismic network in the Muong Te area in one year for this study.

2.2.2. Data processing workflow

To apply the phase picker to detect Muong Te earthquakes, we followed a process illustrated in Fig. 3. The workflow includes 7 steps: (1) collect waveform data; (2) preprocessing; (3) divide the data into smaller sets (monthly data); (4) model parameter calibration; (5) automatic picking; (6) phase association and quality control, and (7) locate the earthquakes and build the catalog. We will briefly describe these steps in the following.

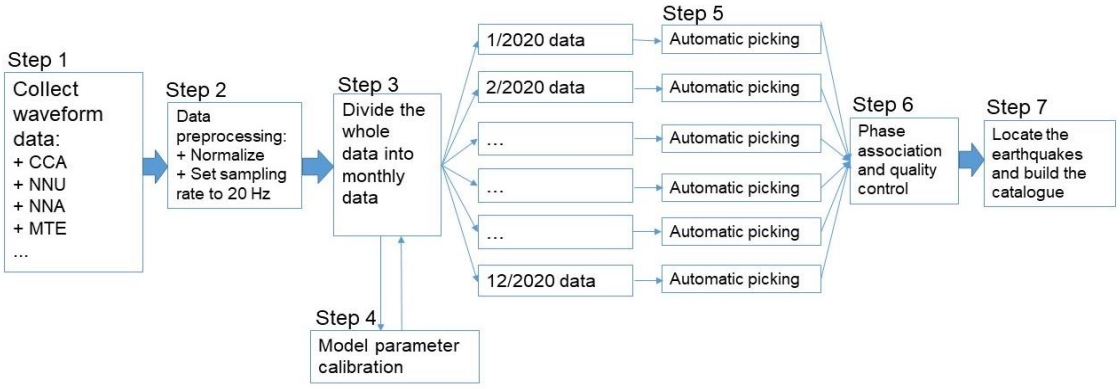


Figure 3. Data processing workflow to apply the phase picker to Muong Te seismic data

The first step is to collect the waveform data from 17 seismic stations from January to December 2020. We collected the 3-component seismic waveforms: the vertical component Z and two horizontal components, N and E. All components in one station (BH*, EH*, and HH*) were collected to have maximum available data. Among 17 seismic stations, two pairs share the same location: station MLAV and station MLVB; station MTE and station VN02 (Table 1 and Fig. 1). Although the stations share the same places, they were deployed at different times and belonged to various projects, therefore, in our seismic processing steps, we need to keep the original name of the stations.

The collected data was in the raw format without any processing. Therefore, we need to preprocess the data (Fig. 4). The preprocessing step includes 3 main techniques: bandpass filter set the sampling rate of 20 Hz and normalization. Figure 4a illustrates a one-day waveform before and after applying a 2 Hz-5 Hz bandpass filter. As shown in the figure, the seismic signals have been enhanced, and more phases can be observed. Figure 4b shows an example of the downsampling rate of a waveform from 100 Hz to 20 Hz; generally, the waveform does not show much difference after downsampling. However, this step is necessary to decrease the data size and fit the phase picker's criteria. Lastly, we normalized the waveform by demean (correct by the average value) and detrended (correct the data trend) techniques. Thus the maximum and minimum value on the y-axis was changed (Fig. 4c). These techniques are very common and frequently used to process seismic data.

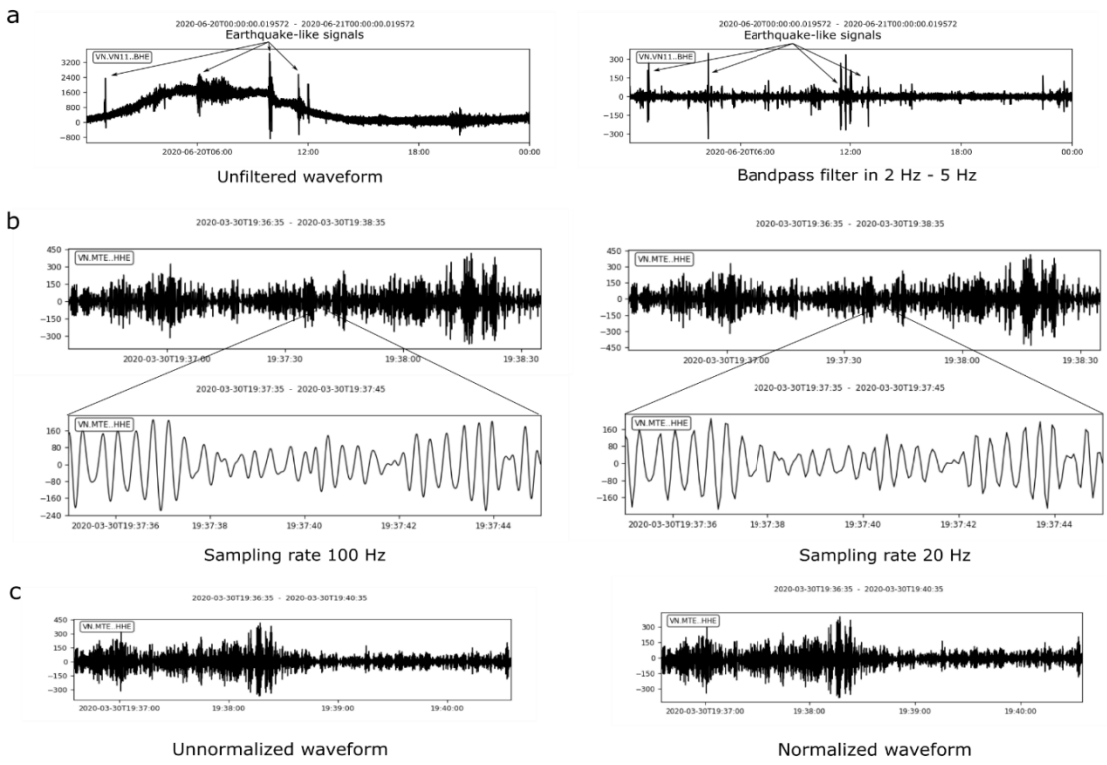


Figure 4. Data preprocessing for the raw seismic waveforms; (a) bandpass filter in 2 Hz-5 Hz; note that the figures show 1-day long seismic waveform, which include both noise and earthquake signals; (b) Downsampling rate to 20 Hz; c) normalization (detrend and demean)

After being preprocessed, all the waveforms were cut into one-day files and grouped into subsets of one-month data. To choose the appropriate model parameter for the phase picker and check if the picker worked properly, we used a small number of data to do test runs. In this step, several one-day waveforms were chosen to be applied EQT picker, and then we modified the model parameter accordingly. Compared to the traditional pickers, the model parameters in the deep learning model have less impact on the result. The calibrated parameters include the overlap between two waveforms to prevent disruption of the data, detection threshold, and P- and S-phase picking threshold. These thresholds are the minimum probability that the model accepts whether the signal is an earthquake or detected as a P- or S-phase. These minimum probabilities are chosen to be between 0 and 1. After several trials, we selected the model parameters as shown in Table 2. These optimized parameters were similar to the suggestion from the model description (Mousavi et al., 2020), with an overlap of 0.2, an earthquake detection threshold of 0.1, and the P- and S-pick thresholds of 0.2.

Table 2. Final model parameters of the EQT picker

Parameters	Value
Overlap	0.2
Detection threshold	0.1
P pick threshold	0.2
S pick threshold	0.2

The phase picker then processed the waveforms. In this step, the EQT model reads 3-component waveforms and outputs the probability of earthquake detection, P-pick, and S-pick every 15 seconds. When a signal has a higher probability than the set

threshold, the time of the seismic phase will be recorded and plotted in spectrogram images. Figure 5 shows examples of detected earthquakes, P- and S-phase pick with their corresponding spectrogram images.

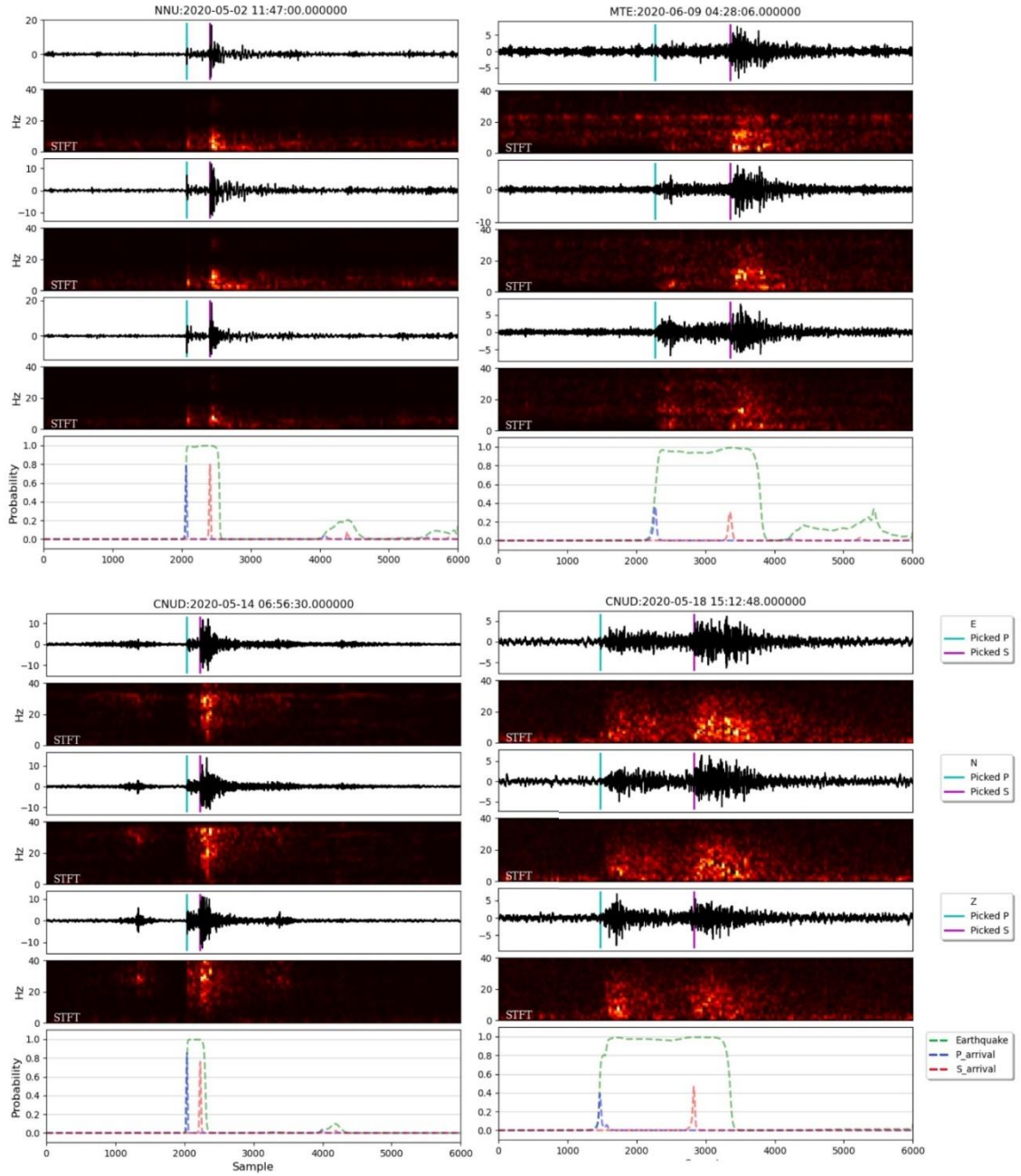


Figure 5. Examples of earthquake detection and phase picking using the EQT model. The raw waveforms are plotted without a bandpass filter. The spectrograms result from applying the Short Time Fourier Transform (STFT) for the time-series waveform. Cyan lines mark the P arrivals, while purple lines mark the S arrivals

The phase picks were then associated and grouped into events that were consequently located using traditional earthquake location methods: Hypoinverse (Klein, 2002) and NonLinLoc (Lomax et al., 2009). We also checked the quality of the phase picks and detection by inspecting the visualization of phase picks and the earthquake locations. Figure 6 illustrates the performance of earthquake detection on the day of the Muong Te earthquake (16/6/2020) for stations VN04 showing a side-by-side comparison between the manually picked events and automatically detected events. We can see that the automatic method could detect all the events included in the manual catalog, plus a significant number of newly detected events. The seismic signals have been identified accurately; we will present more analysis based on a comparison with the original catalog MTMAN to examine the quality of the method. We also determined the local magnitude of newly earthquake catalog by applying the equation from Hutton and Boore (1987):

$$M_L = \log A_{W-A} + 1.11 \log_{10}(r) + 0.00189r - 2.09$$

In which, the $A_{W\text{-}A}$ is the Wood-Anderson zero-to-peak amplitude (mm) of the horizontal seismogram; r is the epicentral distance (km); and M_L is the local magnitude of the event.

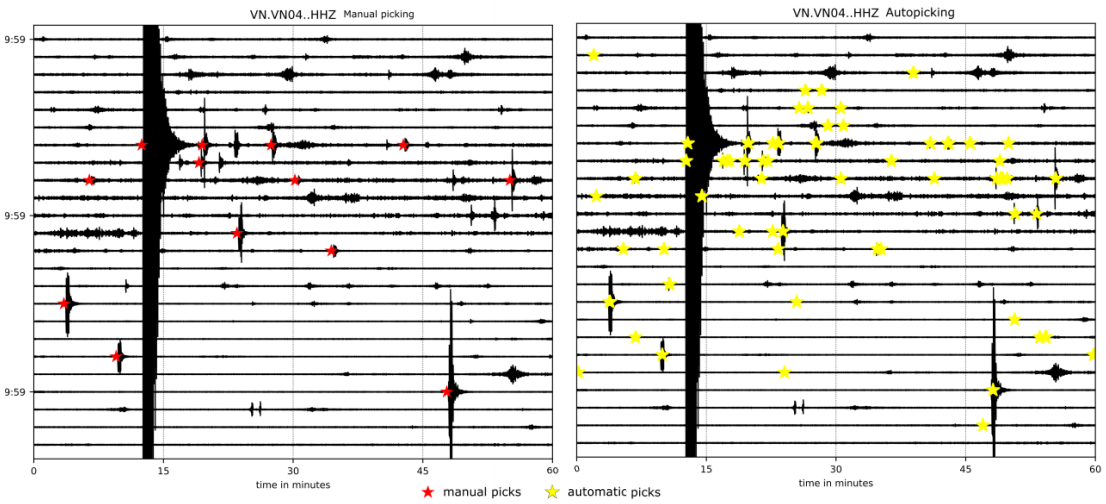


Figure 6. The Z-component seismic waveform of station VN04, when the M_L4.9 Muong Te earthquake occurred with the phase picks from the manual catalog (left) and automatic catalog (right). The red stars denote the manually detected events; the yellow stars denote the events detected by the EQT model. This figure shows that the automatic picker could detect all the events in the manual catalog, plus several new events

3. Result

3.1. Result of the earthquake detection and phase picking

After the automatic picking step and phase association, we have obtained a total of 1198 events with 4983 P picks and 4761 S picks. Figure 7 shows examples of phase picking for 2 events (the Muong Te mainshock and an earthquake with M_L 2.5) after we have associated the phases. Fig. 7a, c shows a typical case of the earthquake in which the phases are pretty clear. In this case, the picks obtained from automatic and manual picking are very close, with the different times in most stations being less than 0.5 seconds. On the other hand, Fig. 7b, d shows a case not detected in the manual catalog MTMAN, mostly due to the waveforms being very unclear with a high noise level. In fact, in the

case of the Muong Te earthquake, most earthquakes are quite small, with magnitudes smaller than 4. Therefore, the signals are often weak and missed under manual processing while they can be detected using the EQT. This proves that the automatic deep-learning picker is necessary and valuable to extend the earthquake catalog.

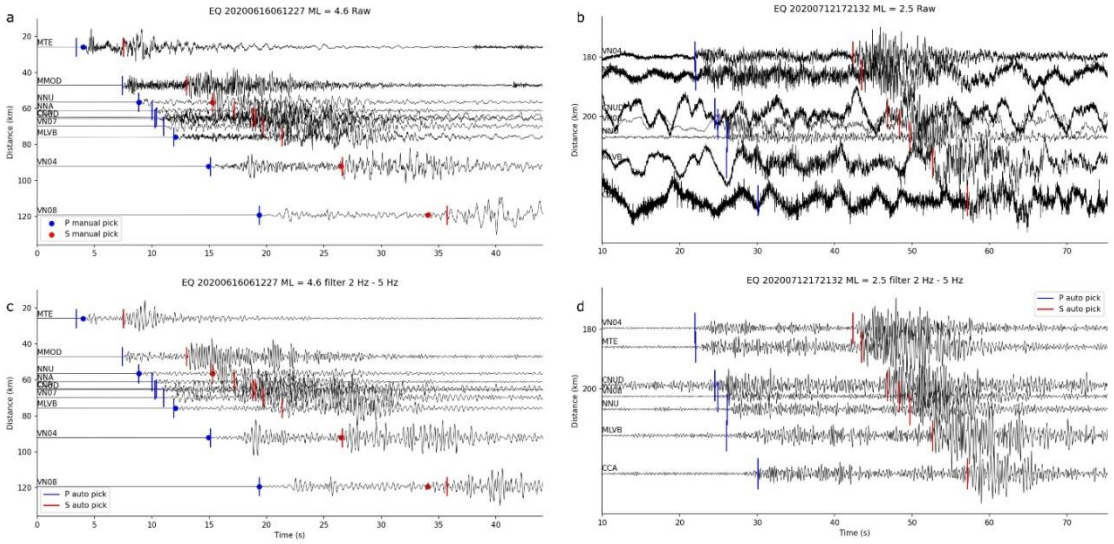


Figure 7. Examples of the phase picks for two events: the Muong Te earthquake (a) and an earthquake with M_L = 2.5 (b, d). Figure a) and b) show the waveforms before applying the bandpass filter, while figures c and d show the waveforms under the bandpass filter of 2 Hz. In the first event, the manual picks are marked with circles

The number of detections, which indicates the event detection for each station, is plotted in Fig. 8. From the figure, we can observe 4 stations with a high number of detections: NNU, CCA, NNA, and MTE, in which each station detects more than 8000 detections. These four stations are located quite close to the Muong Te earthquake, thus suggesting the high number of detections relating to the Muong Te earthquake sequence. This number of detections is much lower than the number of events due to an increased number of detections that cannot be associated with other phases from other stations, thus preventing the formation of an event. The lower number of detections in other stations might be due to the further distance from the event. Therefore, the signals would be too weak, the incomplete record of the data, and a high noise level.

To validate our results, we applied a strategy used by Cianetti et al. (2021) to estimate the number of false detections indirectly: compare the total number of the phase picks (T) versus the number of phases picks that can be associated with arrivals (A). The ratio of A/T is a proxy to evaluate the model prediction, with A/T close to 1 meaning that most of the picks can be associated with an earthquake event. In contrast, A/T close to 0 might indicate a high number of mispredicted signals. However, interpretation of low A/T needs to consider that the picks which are not associated could be true earthquake signals of unlocated small events detected near too few seismic stations.

The results of the estimation of A/T are shown in Table 1 and Fig. 8b, which indicate a total A/T for the whole data set of 16.76%.

For each station, the A/T ranges from 0.4% to 97.52%, with most stations in the range of 8.02% to 39.89%. Compared with the distance from the mainshock, except for three stations, VN02, VN03, and CNUD, we can see a linear trend of A/T with the distance (Fig. 4), which can be a proof that unlocated small earthquakes are occurring near the mainshock. The A/T drop for station VN11 could be interpreted as small events occurring near this particular station that belong to another fault (note the location of this station in Figure 1 of the manuscript). From this result, we can expect average minimum positive accuracy of 16.65%, which is comparable with those reported in Cianetti et al. (2021) and higher than the accurate positive accuracy report in Wiszniowski et al. (2021).

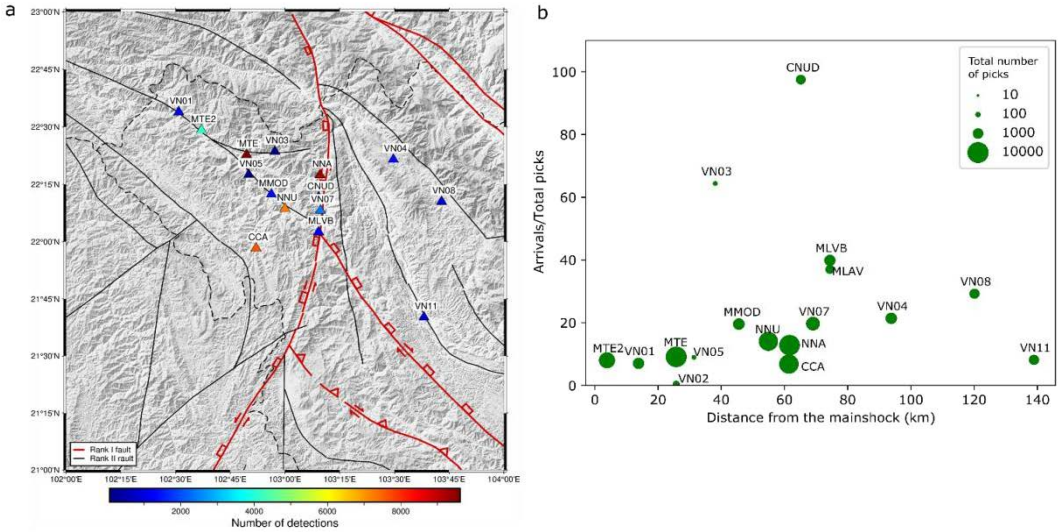


Figure 8. (a) The number of detections for each station in the map view; (b) The graph showing the ratio of associated picks (arrivals) versus the number of total picks by distance from the mainshock

3.2. Result of earthquake location and catalog compilation

After locating the earthquake, we applied quality control of the earthquakes by discarding duplicate events. We set the quality control as follows: two events with the original time difference of less than 5 s and a distance of less than 10 km are considered one event. This step is necessary because, in the automatic detection process, the EQT model will try different combinations in linking the phases into one event. The result of the quality control process is that the number of events decreased from 1198 events to 893 events, meaning that 305 events were discarded. Finally, we have obtained a total of 893 events with 3615 P phase and 3391 S-phase. The local magnitude of the earthquakes was also calculated in this part. The locations of the earthquakes from the new catalog are shown in Fig. 9a, while Fig. 9b shows the locations of the earthquakes from the manual catalog for comparison. From the figure, we can see that most of the earthquakes are located in the Muong Te earthquake source zone. Overall, most earthquakes occurred at a shallow depth between 5 to 20 km, which is not clear in the manual catalog. In the cross-section plot cutting perpendicular to the Upstream Da River fault, we can see that the earthquakes are concentrated near the fault.

Figure 10 shows the distribution of earthquakes with time in the new catalog. A significant increase in the number of

earthquakes can be seen in 6/2020, which corresponds to the time of the Muong Te earthquake, followed by a slight increase of events in 7/2020. The data emphasize the increase in seismic activity related to the Muong Te earthquake.

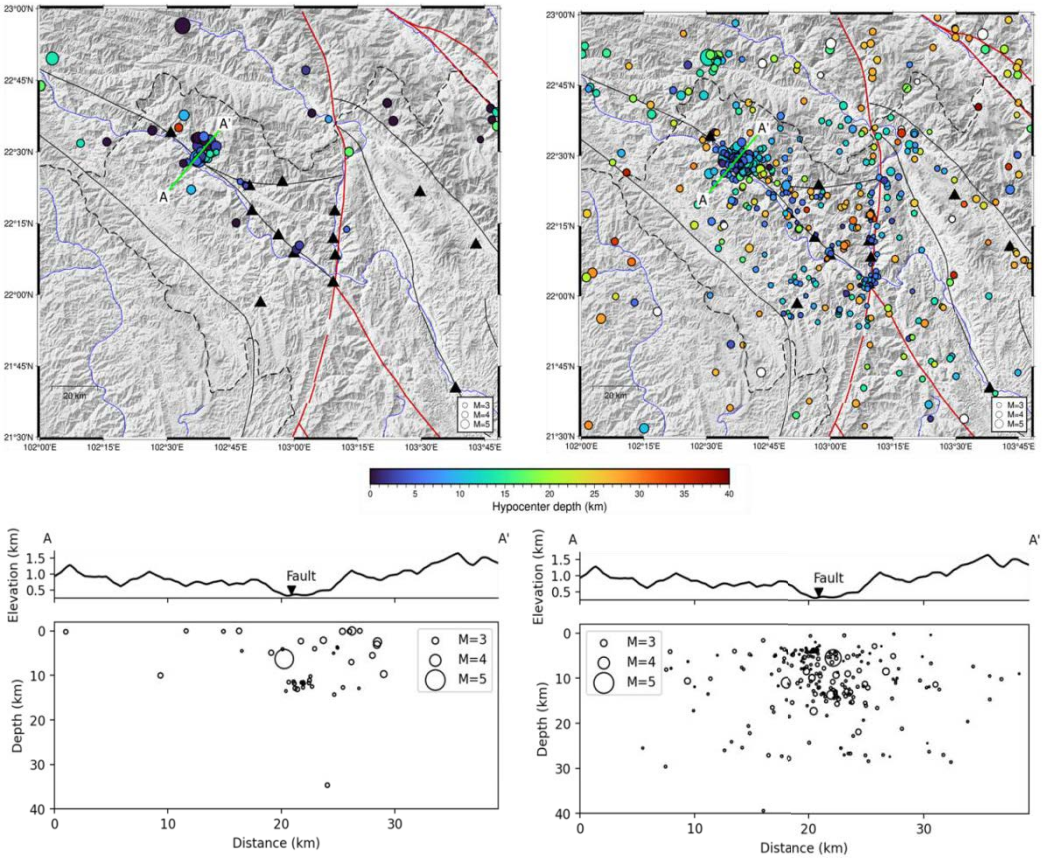
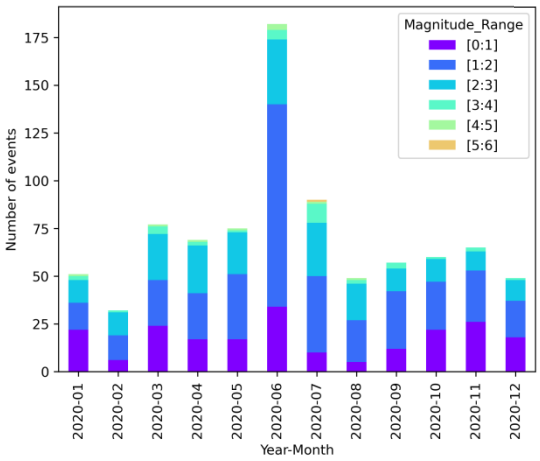


Figure 9. The earthquakes with location from the manual catalog MTMAN (a) and the new catalog MTEQT (b). The cross-section is perpendicular to the Upstream Da River AA' in the manual catalog (c) and the latest catalog (d)



←*Figure 10*. The number of earthquakes with their local magnitude in each month. Note the increase in the number of detected events in June 2020 and July 2020

4. Discussions

4.1. Comparison between automatic and manual catalog

We first compare the automatic catalog MTEQT and the manual catalog MTMAN on quantity and quality. First of all, the number of the events from automated picking is

approximately 13 times that in the manual catalog (893 events compared with 67 events), with the minimum local magnitude of -0.7 in MTEQT, while the minimum local magnitude of MTEQT is 1.0. Furthermore, the processing time using the EQT phase picker was much shortened compared to the manual way. The total processing time of the building catalog using the deep learning phase picker was estimated to be about one day. In contrast, the manual-processing time for 1-year data with 17 stations could probably take a month.

On the other hand, we compared the quality of picks between the two methods. Fig. 11 shows the histogram of the time difference for the P and S-picks, respectively. In general, the result shows a remarkable concordance; the average difference and standard deviation of P picks are 0.02 s and 0.18s, respectively, while the average difference and standard deviation of S picks are -0.04 s and 0.05 s. These measurements of picking quality are quite similar to those reported for the model in the EQT paper (Mousavi et al., 2020), which implies that the picker performance for our data is consistent with the data used to make the EQT picker.

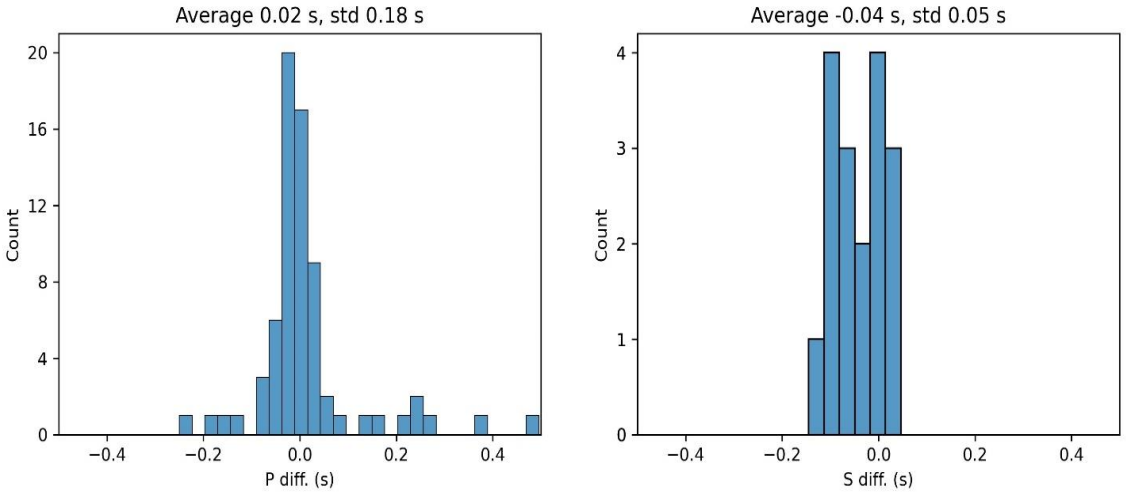


Figure 11. Histogram comparison of the phase picks between automatic catalog MTEQT and manual catalog MTMAN. The x-axis shows the time difference between automated picking and manual picking. The positive value means the picking by EQT is later than the manual picking and vice versa

4.2. Extended catalog reveals hidden earthquake

As we can see in Fig. 9 and discuss in 4.1., the number of earthquakes detected has significantly increased in the area of the Muong Te earthquake. The main reason for this increase in the number of events is the ability of the EQT model to detect small hidden earthquakes, which are very unclear to detect manually. Take Fig. 7b, d as an example. We can see that, before applying a bandpass filter, the waveform is complicated to recognize as a seismic event. The improvement of the detection rate results demonstrated that the new MTEQT catalog provides a far more detailed picture of how seismicity evolves in the Muong Te block in 2020.

Another worth mentioning point is the earthquake occurrence of other faults in the study area. As shown in Fig. 9a, the epicenters concentrated almost near the Muong Te earthquake region, and very few earthquakes were found outside the source zone in the manual catalog. At the same time, a high number of events were detected not only in the Muong Te earthquake area but also in other places in the automatic catalog (Fig. 9b). Small earthquakes below magnitude 2.0 occur more often, in which people cannot feel the shaking, and instruments can hardly detect them. The MTEQT catalog revealed the small earthquake activity in the study area, especially the connected or blind faults associated with the mainshock. This suggests that there can be interactions between one earthquake and another (Brodsky, 2019; Ross et al., 2019).

4.3. Prospect study based on the extended catalog

In seismological studies and especially research about the seismic source, the identification of earthquakes is the most critical task. Using the deep-learning phase picker EQT, we have identified more events than manual picking and saved time and human resources on processing seismic data. On top of that, the high-quality phase picking can allow for a better constraint of earthquake locations using absolute locating methods (Hypoinverse, Nonlinloc). On the other hand, a more completed earthquake catalog is crucial for successfully applying relative earthquake relocation, hypoDD (Waldhauser, 2001) or GrowClust (Trugman and Shearer, 2017) because these methods, earthquakes are divided into clusters based on how close the distance between each event. More events in each cluster may contribute to better relocation, while a single earthquake cannot be relocated. The relative relocation methods are very effective in tracing the fault from seismicity, which reveals the seismogenesis and quantifies the seismic hazards (Trugman and Shearer, 2017; Waldhauser and Schaff, 2008). The extended catalog in our study region with preliminary earthquake location can now be used to apply earthquake relocation, thus may provide a better constraint of the earthquake process and mechanism.

Another prospective study from our extended catalog is studying the structure in the Muong Te area, particularly by applying seismic tomographic methods. We take travel time seismic tomography as an example; in this method, travel times from earthquake-seismic stations are used to invert the velocity structure of the region where ray paths cross (Zhao, 2015). Fig. 12 shows the crossing ray paths from earthquakes to the seismic stations in the two catalogs. In this figure, a significant increase in the number of ray paths for both P- and S-wave in the MTEQT catalog allows for seismic tomographic studies for the Northwestern part of Vietnam. A worth mentioning point in the extended catalog is that the number of S-wave ray paths (3391 S-phase) is almost the same as with the P-wave ray path (3615 P phase). This can improve the resolution of S-wave tomography compared to seismic tomographic inversion based on manual picking because the number of S picks are often notably less than the number of P picks due to inevitable large errors in picking the S phase manually.

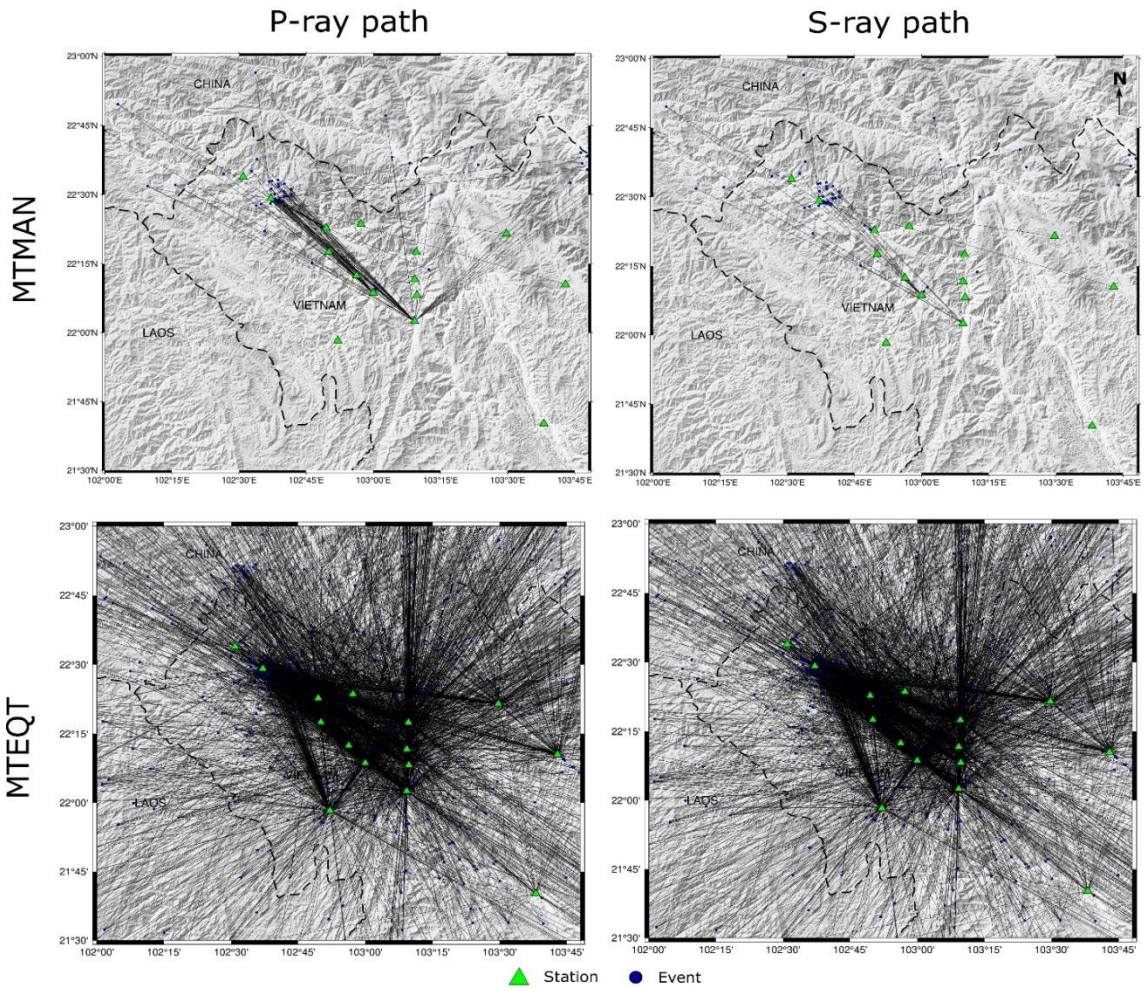


Figure 12. Crossing ray paths of the P- and S-wave from the two catalogs. Note the significant improvement of the ray paths in the automatic catalog compared MTEQT to the manual catalog MTMAN

5. Conclusions

In this study, we have applied the deep-learning phase picker EQT to the 1-year seismic data from 17 seismic stations in the Muong Te region. The phase picks were then grouped as events and located using Hypoinverse and NonLinLoc, calculated for the local magnitude, and compiled into an extended catalog MTEQT. The new catalog consists of 893 events, approximately 13 times the number of events compared to the manual catalog. This new method has allowed us a more completed catalog with a dramatic increase in the number of phase picks and time-saving. This improved catalog has opened up for subsequent studies on the earthquake genesis and structure of the study area.

From the detection of the hidden earthquake events, we can determine the location of the M_L 4.9 Muong Te earthquake and its aftershocks, which supports that the earthquake is likely related to the activation of the Upstream Da River fault. Moreover, we have also identified small earthquakes along the area's major faults, most notably the Dien Bien Phu fault. Most of the events in

northwestern Vietnam are determined to be 5 to 30 km in depth, suggesting that the fault system in the area is active at the crustal level depth.

Acknowledgments

This study has been funded by the Vietnam Academy of Science and Technology (VAST) under grant number: CT0000.02/22-23 and Vietnam National Foundation for Science and Technology Development (NAFOSTED) under the grant number: 01/2021/ĐX. We want to thank two anonymous reviewers for their suggestions and comments on earlier drafts of the manuscript.

References

Allen R.V., 1978. Automatic earthquake recognition and timing from single traces. Bulletin of the seismological society of America, 68, 1521-1532.

Baer M., Kradolfer U., 1987. An automatic phase picker for local and teleseismic events. Bulletin of the Seismological Society of America, 77, 1437-1445.

Baillard C., Crawford W.C., Ballu V., Hibert C., Mangeney A., 2014. An automatic kurtosis-based P- and S-phase picker designed for local seismic networks. Bulletin of the Seismological Society of America, 104, 394-409.

Brodsky E.E., 2019. The importance of studying small earthquakes. Science, 364(6442), 736-737. Doi: 10.1126/science.aax2490.

Cichowicz A., 1993. An automatic S-phase picker. Bulletin of the Seismological Society of America, 83, 180-189.

Cianetti S., Bruni R., Gaviano S., Keir D., Piccinini D., Saccorotti G., Giunchi C., 2021. Comparison of deep learning techniques for the investigation of a seismic sequence: an application to the 2019, Mw 4.5 Mugello (Italy) earthquake. Journal of Geophysical Research: Solid Earth, p.e 2021JB023405.Havskov, J., Ottemoller, L., 1999. SEISAN earthquake analysis software. Seismological Research Letters, 70, 532-534.

Hochreiter S., Schmidhuber J., 1997. Long short-term memory. Neural Computation, 9, 1735-1780.

Hutton L., Boore D.M., 1987. The ML scale in southern California. Bulletin of the Seismological Society of America, 77, 2074-2094.

Klein F.W., 2002. User's guide to HYPOINVERSE-2000, a Fortran program to solve for earthquake locations and magnitudes. Open-File Report, US Geological Survey. Doi: 10.3133/ofr02171.

Kong Q., Trugman D.T., Ross Z.E., Bianco M.J., Meade B.J., Gerstoft P., 2019. Machine learning in seismology: Turning data into insights. Seismological Research Letters, 90, 3-14.

LeCun Y., Bengio Y., 1995. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks. MIT Press, 3361(10), 1995.

Lee W.H.K., Lahr J.C., 1972. HYPO71: a computer program for determining hypocenter, magnitude, and first motion pattern of local earthquakes, Open-File Report, US Geological Survey. Doi: 10.3133/ofr72224.

Liu M., Zhang M., Zhu W., Ellsworth W.L., Li H., 2020. Rapid Characterization of the July 2019 Ridgecrest, California, Earthquake Sequence From Raw Seismic Data Using Machine-Learning Phase Picker. Geophysical Research Letters, 47, e2019GL086189.

Lomax A., Michelini A., Curtis A., Meyers R., 2009. Earthquake location, direct, global-search methods. Encyclopedia of Complexity and Systems Science, 5, 2449-2473.

Mousavi S.M., Ellsworth W.L., Zhu W., Chuang L.Y., Beroza G.C., 2020. Earthquake transformer an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nature Communications, 11, 1-12.

Mousavi S.M., Sheng Y., Zhu W., Beroza G.C., 2019. STanford EArthquake Dataset (STEAD): A global data set of seismic signals for AI. IEEE Access, 7, 179464-179476.

Münchmeyer J., Woollam J., Rietbrock A., Tilmann F., Lange D., Bornstein T., Diehl T., Giunchi C., Haslinger F., Jozinović D., 2022. Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. Journal of Geophysical Research: Solid Earth, p.e2021JB023499.

Nguyen H.P., Pham T.T., Nguyen T.N., 2019. Investigation of long-term and short-term seismicity in Vietnam. Journal of Seismology, 23, 951-966.

Ross Z.E., Meier M.A., Hauksson E., Heaton T.H., 2018. Generalized seismic phase detection with deep learning. Bulletin of the Seismological Society of America, 108, 2894-2901.

Ross Z.E., D.T. Trugman, E. Hauksson, P.M. Shearer, 2019. Searching for hidden earthquakes in Southern California. Science, 364, 767-771. Doi: 10.1126/science.aaw6888.

Saragiotis C.D., Hadjileontiadis L.J., Panas S.M., 2002. PAI-S/K: A robust automatic seismic P phase arrival identification scheme. IEEE Transactions on Geoscience and Remote Sensing, 40, 1395-1404.

Sleeman R., Van Eck T., 1999. Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. Physics of the Earth and Planetary Interiors, 113, 265-275.

Trnkoczy A., 2009. Understanding and parameter setting of STA/LTA trigger algorithm, New Manual of Seismological Observatory Practice (NMSOP). Deutsches GeoForschungsZentrum GFZ, 1-20.

Trugman D.T., Shearer P.M., 2017. GrowClust: A hierarchical clustering algorithm for relative earthquake relocation, with application to the Spanish Springs and Sheldon, Nevada, earthquake sequences. Seismological Research Letters, 88, 379-391.

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I., 2017. Attention is all you need. Advances in Neural Information Processing Systems, 5998-6008.

Waldhauser F., 2001. hypoDD-A Program to Compute Double-Difference Hypocenter Locations, Open-File Report. US Geological Survey. Doi: 10.3133/ofr01113.

Waldhauser F., Schaff D.P., 2008. Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods. Journal of Geophysical Research: Solid Earth, 113(B8). Doi: https://doi.org/10.1029/2007JB005479.

Wang J., Li T., Gu Y.J., Schultz R., Yusifbayov J., Zhang M., 2020. Sequential Fault Reactivation and Secondary Triggering in the March 2019 Red Deer Induced Earthquake Swarm. Geophysical Research Letters, 47, e2020GL090219.

Wiszniowski J., Plesiewicz B., Lizurek G., 2021. Machine learning applied to anthropogenic seismic events detection in Lai Chau reservoir area, Vietnam. Computers & Geosciences, 146, 104628.

Xiao Z., Wang J., Liu C., Li J., Zhao L., Yao Z., 2021. Siamese Earthquake Transformer: A pair-input deep-learning model for earthquake detection and phase picking on a seismic array. Journal of Geophysical Research: Solid Earth, 126, e2020JB021444.

Zhao D., 2015. Multiscale seismic tomography. Springer Geophysics Book Series. Doi: 10.1007/978-4-431-55360-1.

Zhou L., Zhao C., Zhang M., Xu L., Cui R., Zhao C., Duan M., Luo J., 2021. Machine-learning-based earthquake locations reveal the seismogenesis of the 2020 Mw 5.0 Qiaojia, Yunnan earthquake. Geophysical Journal International, 228, 1637-1647.

Zhu W., Beroza G.C., 2019. PhaseNet: a deep-neural-network-based seismic arrival-time picking method. Geophysical Journal International, 216, 261-273.