Vietnam Journal of Earth Sciences, 48(3), 465–481, https://doi.org/10.15625/2615-9783/24522

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

# Automated earthquake detection and high-resolution catalog construction in the Jailolo volcanic region, Indonesia

Siti Navisa[1*], Titi Anggono[2], Firman Syaifuddin[1], Agnis Triahadini[1], Ade Surya Putra[2]

[1]*Geophysical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
[2]*Research Center for Geological Disaster of the National Research and Innovation Agency, Tangerang, Indonesia*

Received 18 November 2025; Received in revised form 27 March 2026; Accepted 20 April 2026

ABSTRACT

The Jailolo region, located on Halmahera Island, is characterized by intense seismic activity due to its proximity to two subduction zones and the Halmahera volcanic arc. To optimize the analysis of local earthquake activity in the region, we used an automated workflow that combines the deep-learning phase picker EQtransformer with the unsupervised machine learning association algorithm GaMMA to construct a high-resolution microseismic catalog. Phase arrivals were first identified using the EQTransformer and subsequently associated with each other using GaMMA, a Bayesian Gaussian Mixture Model with spatial clustering for earthquake event detection. This workflow produced 6,174 preliminary detections, of which 2,902 high-quality events were retained after relocation and filtering. The earthquakes predominantly occurred at depths shallower than 50 km, with an average local magnitude ($M_L$) of 0.91. The obtained catalog revealed a large number of previously unreported small-magnitude earthquakes in the International Seismological Center (ISC) catalog. The completeness magnitude (Mc = 0.63) indicates a significant improvement in detection capability, enabling detailed analysis of microseismicity. The study highlights the effectiveness of automated machine-learning techniques in improving the resolution of seismicity patterns in a tectonically complex region such as Jailolo.

*Keywords*: Jailolo, EQTransformer, GaMMA, NonLinLoc, HypoDD.

## 1. Introduction

An earthquake catalog is a comprehensive document containing information on earthquakes that have occurred, including location, origin time, magnitude, and hypocenter depth. This information serves as the primary data source for various seismological studies, such as investigations of tectonic activity, active fault structures, earthquake activity patterns, and the development of subsurface seismic wave velocity tomography models (Ammon et al., 2021; McBrearty & Beroza, 2022). The distribution of earthquakes follows a power law; therefore, the more earthquake data are recorded, the higher the spatial and temporal resolutions become (Kagan, 2010). This principle has been demonstrated by Mignan (2014), who analyzed 37 foreshocks of the Haicheng earthquake in China and found that the results depend on the completeness of the catalog used. A similar finding was reported by Naoi et al. (2015), who discovered that seismic activity in certain structures (i.e.,

*Corresponding author, Email: sitinavisabcd@gmailcom

faults) consists almost entirely of small earthquakes with magnitudes less than 2. Therefore, providing detailed earthquake catalogs that include small-magnitude earthquakes is crucial for comprehensive seismic activity analysis.

The development of earthquake catalogs generally proceeds in three main stages: picking the arrival times of P- or S-wave phases, phase association, and earthquake location determination (Kubo et al., 2024). Machine learning (ML) approaches with automated workflows have been implemented in various studies to develop earthquake catalogs. For example, Tamaribuchi et al. (2023) successfully detected hidden earthquakes following the 2011 Japan-Tohoku earthquake, significantly enhancing the previously existing catalog. Their results showed that the conventional catalog (JMA) contained 320,427 events, while the new ML-enhanced catalog identified 922,427 events. The ML approach was also applied to the earthquake sequence in Puerto Rico from 28 December, 2019 to 2020 by Zhu et al. (2023). The authors compared their newly associated catalog to the standard Puerto Rico Seismic Network catalog, producing more detailed information, including small-magnitude events. The studies by Tamaribuchi et al. (2023) and Zhu et al. (2023) demonstrate a promising machine-learning approach for completing conventional catalogs by enhancing the detection of small seismic events. Despite these advances, ML applications have been performed on well-monitored regions with extensive seismic networks. The effectiveness of these approaches in poorly documented areas with limited seismic networks remains untested. Our research addresses this limitation by developing and validating an automated ML workflow specifically designed for sparse seismic monitoring conditions.

A relevant example of poorly documented seismic events is the Jailolo region, located within the tectonically complex Halmahera arc in North Maluku, Indonesia. The regional tectonic setting is characterized by a unique dual subduction system, with the westward subduction zone generating intermediate to deep earthquakes. In contrast, the eastward subduction zone beneath Halmahera produces predominantly shallow earthquakes. The active volcanoes in this arc, including Mount Jailolo, may produce small, shallow seismic events associated with volcanic activity (Morris et al., 1983). This complex geological framework creates ideal conditions for diverse seismic activity patterns, challenging conventional earthquake detection methods.

Based on seismic recordings from November to December 2015 using stations located in nearby municipalities, more than 1,000 earthquake events with magnitudes $< 1$ and 11 earthquakes with magnitudes $> 4.5$ have been detected (Gunawan et al., 2016). However, global, openly accessible earthquake catalogs, such as those from the International Seismological Center (ISC), only record events with magnitudes greater than 3. For example, during 2016–2017, the ISC catalog recorded only a small number of seismic events in Jailolo and has no earthquake swarm records around Mount Jailolo. This Mount Jailolo swarm was previously documented by the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) (BMKG, 2017). The complex tectonic setting, combined with sparse seismic monitoring stations and the predominance of small-magnitude events, creates significant challenges for comprehensive earthquake catalog development. Combined with openly available waveform data in the region, this situation presents an optimal opportunity for ML approaches to improve Jailolo's earthquake catalog.

This study aims to demonstrate that automated machine learning workflows can significantly enhance earthquake catalog completeness in poorly monitored tectonically active regions. To achieve this goal, we developed an earthquake catalog for the Jailolo region using an ML approach and an automated workflow. We used waveform data from the FDSN 7G network, which recorded seismic activity from 2016 to 2017 (Passarelli et al., 2016). The automated processing workflow addresses the challenges of a sparse, monitored region by implementing advanced algorithms that detect events that conventional methods often miss. The collected waveform data were processed through a sequential workflow consisting of: (1) preprocessing, including detrending and bandpass filtering; (2) automated phase detection and arrival-time picking using the deep-learning model EQTransformer; (3) phase association using the unsupervised machine-learning algorithm GaMMA to link P- and S-wave arrivals to individual seismic events; (4) hypocenter determination and relocation using the double-difference method HypoDD; and (5) local magnitude (ML) calculation (Jiang et al., 2022). The resulting catalog enables quantitative evaluation of detection performance, magnitude distribution, and catalog completeness relative to existing regional and global catalogs. The effectiveness of EQTransformer has been demonstrated across various tectonic settings. For example, its application to the 2000 Tottori earthquake sequence in Japan nearly doubled the number of detected events compared to the Japan Meteorological Agency (JMA) catalog, despite using fewer seismic stations (Mousavi et al., 2020). Similarly, Nguyen et al. (2022) applied EQTransformer for small-magnitude earthquakes, highlighting the model's capability to improve catalog completeness in complex tectonic environments. For hypocenter estimation, previous studies have demonstrated the reliability of probabilistic location methods. Wulandari et al. (2017), for instance, applied NonLinLoc to analyze the aftershock sequence of the 27 May 2006 Yogyakarta earthquake and showed that the relocated epicenters were concentrated along the previously hypothesized fault zone parallel to the Opak Fault. This study will produce an enhanced earthquake catalog of the region, with quantitative assessments to improve event detection, magnitude distribution analysis, and catalog completeness evaluation relative to existing regional catalogs.

## 2. Tectonic setting

Following Macpherson & Hall (2002), we summarize the regional tectonic background in our study area, Jailolo. Jailolo is a region on Halmahera Island located in a highly dynamic tectonic zone characterized by the convergence of oceanic plates and island arcs. This region lies within the Molucca Sea Collision Zone, characterized by the contact between the Halmahera, Molucca Sea, and Sangihe plates (Fig. 1). The Maluku Sea Plate converges and creates double subduction in this region, westward beneath the Sangihe Arc and eastward beneath the Halmahera Arc, with a movement rate of about 1.5 cm/year. The subduction process creates a U-shaped or inverted V-shaped configuration, resulting in two Benioff Zones. The westward subduction zone generates intermediate to deep earthquakes, while the eastward subduction zone beneath Halmahera produces shallow earthquakes (Morris et al., 1983).

The Sangihe and Halmahera Arcs are two active volcanic arcs bounding the Jailolo region to its west and east, respectively. These arcs are separated by approximately 250 km. The Halmahera Arc consists of a chain of Quaternary volcanoes stretching over 300 km, underlain by a Benioff zone at

depths of 90-125 km. The active volcanoes in this arc are spaced roughly 60 km apart and are composed of volcanic rocks overlying oceanic crust (Morris et al., 1983). The Halmahera Arc is intersected by numerous minor faults formed by upper-crustal deformation associated with subduction and magma intrusion. Continuous volcanic activity near Mount Jailolo further contributes to shallow seismicity, which has been linked to stress changes driven by a combination of tectonic forces and deep magma activity (Supendi et al., 2018). The presence of older volcanic rocks that have undergone extensive weathering and fracturing creates zones of weakness, allowing stress to be more easily released and potentially triggering microearthquakes.

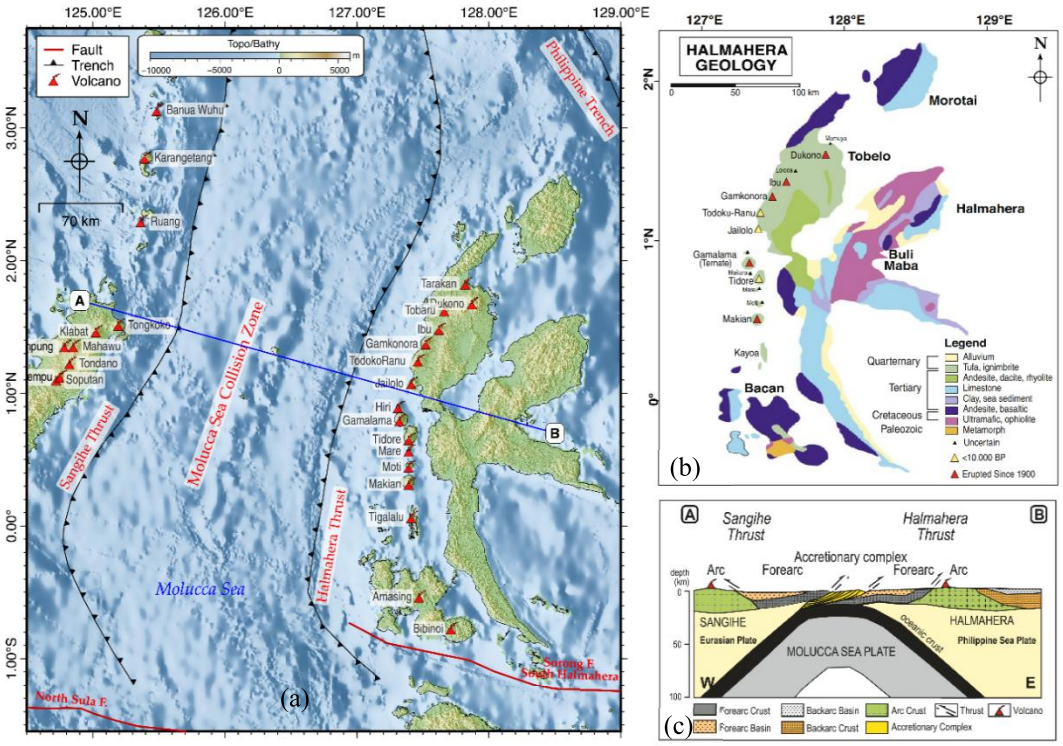


*Figure 1*. (a) Tectonic setting of the Jailolo region (modified from Macpherson and Hall, 2002); (b) Geological setting of the Jailolo region (modified from Apandi and Sudana, 1980); (c) Cross section of divergent double subduction Molucca Sea Collision Zone (modified from Zhang et al., 2017)

## 3. Data and Method

Continuous waveform data were obtained from the seismic temporary network 7G, deployed around the Jailolo volcanic region from August 2017 to July 2017, as shown in Fig. 2 (Passarelli et al., 2016). The network consisted of broadband three-component seismometers (HH* channels). Waveforms were downloaded from the GEOFON GFZ Helmholtz Center for Geoscience in miniSEED format, along with their station metadata. All data were sampled at 100 Hz. The station codes and their location are shown in Supplementary Table T1. We resampled the data to 100 Hz to ensure consistency across all stations. Preprocessing steps included bandpass filtering the data to 0.1−10 Hz and removal of the mean and linear trend.

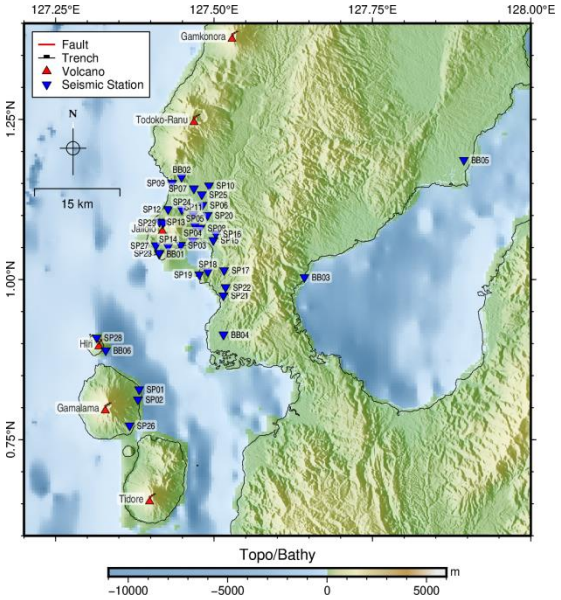


*Figure 2*. Distribution of seismic stations from the temporary 7G network deployed around the Jailolo area. The network consisted of multiple stations equipped with broadband sensors (HH* channels) that recorded three-component seismic data (E, N, Z) continuously from August 2016 to July 2017

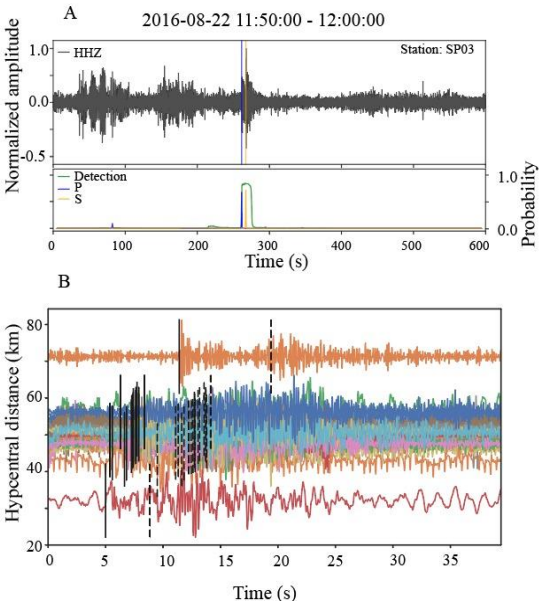


*Figure 3*. (A) Example of phase picking using EQTransformer at station SP03, with the waveform filtered using a 0.1–10 Hz bandpass filter; (B) phase association using GaMMA for the corresponding event.

We performed phase detection and picking using EQTransformer, a deep learning model that estimates event-detection and phase-arrival probabilities for P- and S-phases (Mousavi et al., 2020). We applied the Instance dataset (Italian Seismic Dataset for Machine Learning) (Michelini et al., 2021) and a pretrained model available through the SeisBench framework (Woollam et al., 2022). We adopted a probability threshold of 0.1 for both P- and S-phases to enhance sensitivity to low-magnitude events, which we expected to dominate microseismicity in the Jailolo region (Mousavi et al., 2020; Nguyen et al., 2022). This relatively low threshold may increase false detection; therefore, subsequent phase association was used to ensure that detected phases were from the same event. Our processing produced about 2.16 million P-phase picks and 1.26 million S-phase picks. An example of event detection using EQTransformer is shown in Fig. 3.

We performed phase association using GaMMA (Gaussian Mixture Model Association), an unsupervised association framework that combines Bayesian Gaussian mixture modeling with density-based clustering to group individual P- and S-picks into a single event (Zhu et al., 2022; Zhu et al., 2023). In GaMMA, candidate picks are clustered using DBSCAN, and association quality is filtered based on the minimum number of picks required for an event to be considered a valid earthquake. We set the GaMMA configuration in the spatial dimension (kilometers) and adopted a constant-velocity model with Vp = 7.5 km/s and Vs = 4.2 km/s. DBSCAN clustering used a maximum time separation of 10 s for the linked phase and a minimum of 3 phases to form a cluster. We set the oversampling factor of 4 to refine location accuracy. Event filtering required ≥ 7 phase picks per earthquake. Amplitude data were excluded, focusing solely on travel-time and spatial

correlations. All of these GaMMA workflows yielded preliminary hypocenter locations for 6,174 events (Fig. 4a).

We estimated absolute hypocenter locations using NonLinLoc, which uses a grid-search framework with an oct-tree importance-sampling algorithm (Lomax and Curtis, 2001). We computed travel times on a three-dimensional grid with 1 km spacing in all directions, and we used the global AK135 velocity model to generate P- and S-wave travel times, which were also used in the HypoDD relocation stage (Kennett et al., 1995). For each event, NonLinLoc evaluates a location probability density function and returns a maximum-likelihood hypocenter and associated uncertainty derived from the posterior distribution. Figure 4b shows the result of the hypocenter location using the NonLinLoc algorithm. The output from the hypocenter estimation was evaluated using root-mean-square (RMS) residuals and lateral errors. The obtained RMS and error outputs from NonLinLoc are shown in Supplementary Figure S1. Earthquake hypocenters with lateral errors > 20 km and RMS residuals exceeding 3 were excluded from the analysis (i.e., Jiang et al., 2022). This procedure eliminated 3,272 events, accounting for 53% of the total GaMMA events, leaving 2,902 remaining (Fig. 4c).

Event relocation was subsequently performed using the HypoDD double-difference algorithm, which minimizes the residual differential travel time between pairs of nearby earthquakes (Waldhauser and Ellsworth, 2000). HypoDD compares two or more earthquake hypocenters based on the Geiger method. In this study, the relocation was performed using catalog differential travel times without waveform cross-correlation data and solved using an LSQR least-squares inversion. The complete parameters are summarized in Supplementary Table T2. From 2,902 identified events, 281,905 P-phases and 134,082 S-phases were recorded, with outliers constituting 6% of the total phases. The chosen P- and S- phase pairs exceeded 85%, comprising 242,939 and 113,904 phases, respectively. A total of 38,725 event pairs were acquired, with an average of 9 events per pair and an average offset of 5.47 km between pairs. The HypoDD relocation resulted in a substantial improvement in relative location accuracy, as indicated by a reduction of the weighted RMS of catalog differential travel times (RMSCT) from 0.244 s in the initial iteration 0.0006 s by iteration 20, corresponding to a reduction of approximately 99.8%. Results from the HypoDD showed favorable RMSCT (Root Mean Square of Cross-Temporal residuals) values below 1 ms, indicating accurate positions. Concurrently, the mean absolute RMS decreased from 0.356 s to 0.0012 s. Location uncertainties also improved markedly, with the mean horizontal 2σ error reduced from approximately 1.6−1.8 km in the initial iteration to less than 0.5 km by iteration 10, and to < 0.01 km in the final iterations. The mean hypocentral shifts decreased progressively from initial values of ~600−1400 m to < 1 m in the final iteration, indicating stable convergence of the relocation. This improvement was further supported by an average quality (AQ) value of 2,086 on the 18$^{th}$ iteration, with a condition number (CND) of 42, which lies within the recommended stability range of 40−80 ($40 \leq CND \leq 80$) (Waldhauser, 2001). Overall, 2,427 events were relocated by this step (Fig. 4d), while the residual events with no corresponding match were excluded.

We estimate the earthquake local magnitude ($M_L$) using the following equation.

$$M_L = \mathrm{Log}_{10}(\mathrm{amp} \cdot 1000 \frac{\mathrm{mm}}{\mathrm{m}}) + a \cdot d_{epi} + b \quad (1)$$

where amp represents the maximum Wood-Anderson amplitude (mm), and $d_{epi}$ is epicentral distance (km). The $M_L$ formulation follows the classical local magnitude definition introduced by Richter (1935),

which is based on Wood-Anderson instrument responses. The distance correction coefficients a and b were adopted from the Swiss local magnitude calibration, with values of a = 0.018, b = 2.17 for $d_{epi} \leq 60$ km, and a = 0.0038 and b = 3.02 for $d_{epi} > 60$ km, as implemented in the Swiss Seismic Network magnitude determination procedure (Nanjo et al., 2010). We downloaded the event waveforms using a 90 s window starting from 15 s before the origin time. We then instrument-corrected the waveform to a Wood-Anderson seismogram using ObsPy, and station magnitudes were sorted by median to obtain the final event $M_L$ (Beyreuther et al., 2010).

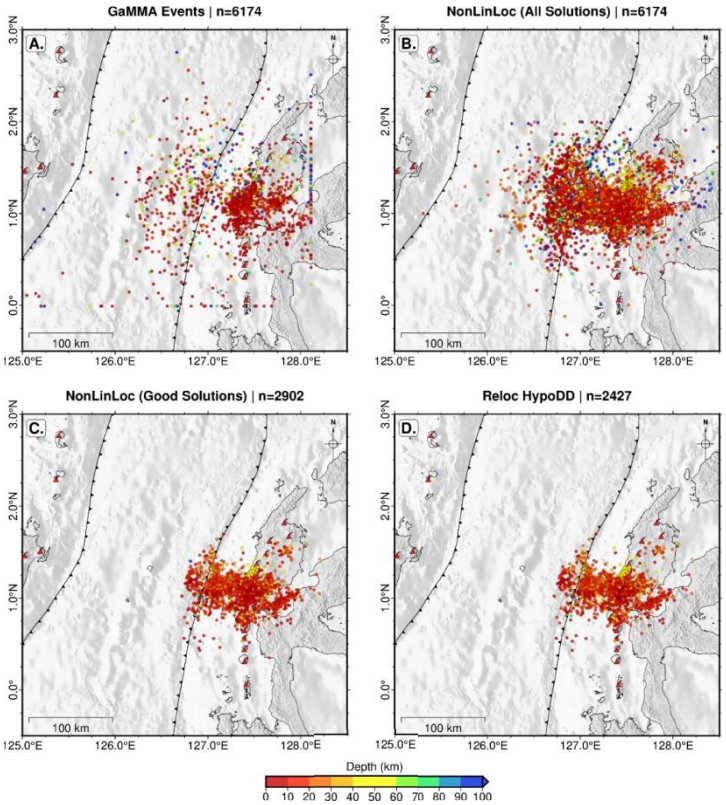


*Figure 4*. (A) Event detections from GaMMA; (B) Hypocenter locations from NonLinLoc (all events); (C) NonLinLoc results after removing poor-quality solutions; (D) Relocated events using HypoDD

## 4. Results and Discussion

### *4.1. Seismicity*

Figure 4 (a-d) illustrates the distribution of earthquake events at each analysis stage, as explained in the previous chapter. Figure 4a illustrates the spatial distribution of earthquake events identified by GaMMA, which still shows some location inconsistencies, as several events appear to follow the predefined search boundaries. In contrast, Fig. 4b presents the distribution of all events after hypocenter determination using NonLinLoc, revealing a more concentrated cluster of seismicity aligned with the 7G station network and extending toward the Halmahera Trench. This result is consistent with the HypoDD relocation (Fig. 4d), which successfully relocated 2,427 events, whereas 475 events could not be relocated due to limitations in the HypoDD parameters, such as the minimum inter-event

distance and station-pairing requirements. These unrelocated events were still used to ensure the catalog's completeness. For relocated events, the average differences between the two solutions were 1.8 km in horizontal position, 1.5 km in vertical position, and 2.65 km in overall three-dimensional displacement. Figure 5 shows the horizontal and vertical distribution of the final earthquake dataset, which combines the 2,427 events relocated by HypoDD and the 475 events by NonLinLoc. Our catalog consists of 2,902 events recorded across 7G network stations, with depths ranging from near-surface to >100 km, and includes both land and oceanic earthquakes. Based on the depth classification of earthquakes proposed by Shearer (1999), 81% of the events in our catalog are categorized as shallow (0–50 km), whereas the remaining 19% are classified as intermediate-depth events (50–200 km).

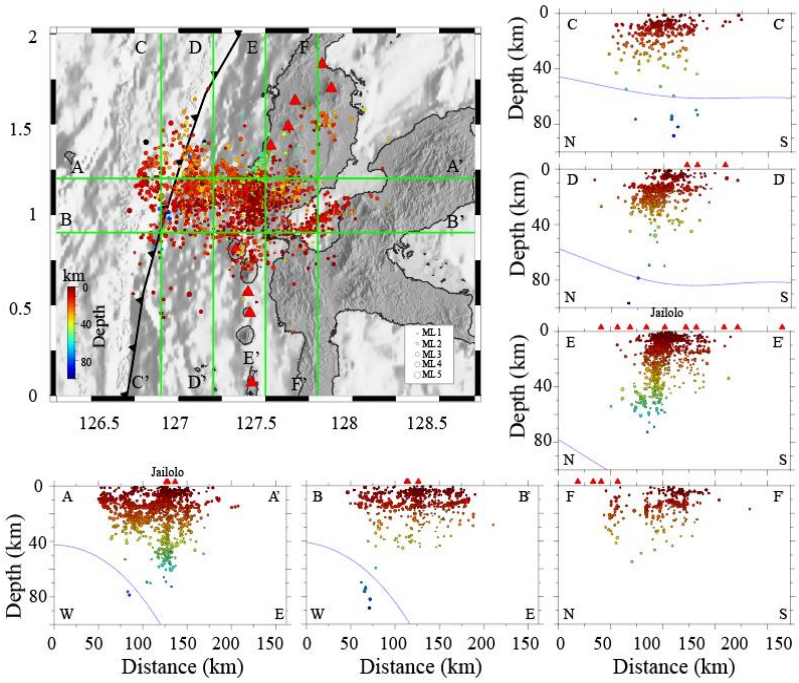


*Figure 5*. Vertical cross-section of earthquake map distribution along line A-A′; B-B′; C-C′; D-D′; E-E', & F-F'

The number of earthquake occurrences (event count) varied significantly throughout the recording period, which began on 22 August, 2016 (Fig. 6a). The number of daily events generally ranged from 5 to 20 and alternated between periods of low activity with only a few events and intervals of increased seismic activity (Fig. 6b). Periods with incomplete or poor-quality waveform recordings indicated data limitation rather than a lack of seismic activity, and were represented as gaps in Fig. 6a. In Fig. 6a, the $M_L$ in our catalog ranged from 0 to over 4, with 92% of events having $M_L < 2$ and 99% $M_L < 3$. The average $M_L$ for the entire catalog was 0.91. The increases in earthquake occurrence on day 40, with approximately 70 events per day, and around day 310, with about 50 events per day, were dominated by $M_L < 2$ earthquakes. The magnitude scale applied in this study is the local magnitude ($M_L$), which generally underestimates the magnitudes of deeper events as it is primarily based on epicentral distance.

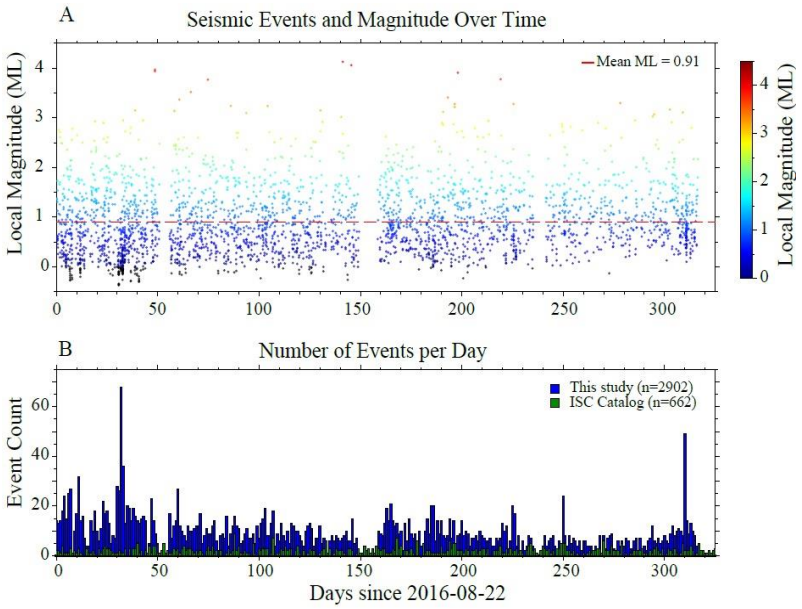


*Figure 6*. (A) Magnitude variation of seismic events; (B) Daily variation of earthquake occurrences

The integration of the deep learning model EQTransformer for automatic detection and the unsupervised machine learning algorithm GaMMA for event association proved effective in identifying earthquake events, particularly those of small magnitude that are frequently excluded from conventional catalogs. The key advantage of this approach lies in the increased number of phase picks generated by EQTransformer, which enables GaMMA to form more clusters and subsequently identify more events within its processing constraints. In this study, a relatively low probability threshold of 0.1 was applied to maximize the number of detected phase picks, thereby enhancing sensitivity to weak and low-magnitude events. This threshold value has also been adopted in previous studies to increase pick completeness (Nguyen et al., 2022; Pita-Slim et al., 2023). While a low threshold inevitably increases the likelihood of false picks, this limitation was mitigated by the GaMMA association process, which identifies earthquake events only when phase picks are spatially and temporally consistent across three or more stations, thereby filtering out spurious detections.

Implementing EQTransformer produced fewer S-picks than P-picks. Testing across several waveform time windows showed that reducing the probability threshold increased in the number of picks. Although probability does not serve as a perfect predictor at the individual-pick level, it nevertheless remains an invaluable parameter for quality control. Pita-Sllim et al. (2023) reported a non-linear relationship between probability and accuracy, particularly in recall and false-positive rates. We tested this relationship in our dataset by evaluating the correspondence between EQTransformer probability values and picking accuracy for selected earthquake events. Picking accuracy was quantified as the time difference (Δ) between automatic and manual phase arrival times, with smaller Δ values indicating closer agreement with manual picks. The distributions shown in Supplementary Figure S2a and the summary statistics presented in Supplementary Table T3 confirm that EQTransformer probability values should not be interpreted as direct indicators of picking accuracy at the individual phase level. Instead, probability

serves as a useful quality-control parameter at the catalog scale, supporting previous findings that low-probability picks can still be temporally accurate, whereas high-probability picks may exhibit larger timing uncertainties.

A direct spatial comparison between the ISC catalog and the final relocated catalog is presented in Fig. 7. The ISC catalog shows earthquake epicenters distributed widely across the study region, including areas along the Mayu-Talaud ridge and offshore zones. In contrast, the final catalog derived in this study exhibits a more concentrated spatial distribution, primarily within the coverage of the 7G seismic network. The difference is substantial, indicating that the ISC catalog includes events detected by broader regional and global networks. In contrast, our catalog is strongly influenced by the geometry and density of the local 7G stations. A progressive reduction in spatial extent is observed throughout the processing stages (Fig. 4a-d). The initial GaMMA-associated events (Fig. 4a) display a spatial pattern that is relatively similar to the ISC distribution, including events extending toward the ridge area. After absolute location using NonLinLoc (Fig. 4b), the overall pattern remains similar, although some peripheral events exhibit increased location uncertainty. However, after applying quality control by removing events with large lateral errors (> 20 km) and high RMS residuals (> 3 s), the resulting "good solutions" dataset (Fig. 4c) shows a markedly reduced spatial extent. The final HypoDD relocation (Fig. 4d) further refines the clustered seismicity pattern, concentrating events within the region of dense station coverage. This reduction is primarily attributed to the quality filtering applied to the NonLinLoc solutions. Events located far from the 7G network tend to exhibit large azimuthal gaps and poor station geometry, resulting in higher RMS residuals and larger location uncertainties. Because NonLinLoc estimates hypocenters probabilistically, solutions in areas with sparse station coverage exhibit broader probability density functions and consequently larger errors. These events were excluded during the quality-control stage.

An additional evaluation of seismic station performance was carried out by quantifying the number of earthquakes and phase picks contributed by each station to the final catalog, as shown in Supplementary Figure S3. From the associated phase picks, a total of 72,787 P-phase and 31,075 S-phase arrivals were obtained. Stations such as SP22, SP04, SP17, and SP20 contributed the most associated events and phase picks, whereas SP26 and SP07 contributed relatively little, likely due to differences in waveform signal quality, particularly in amplitude levels and signal variance.

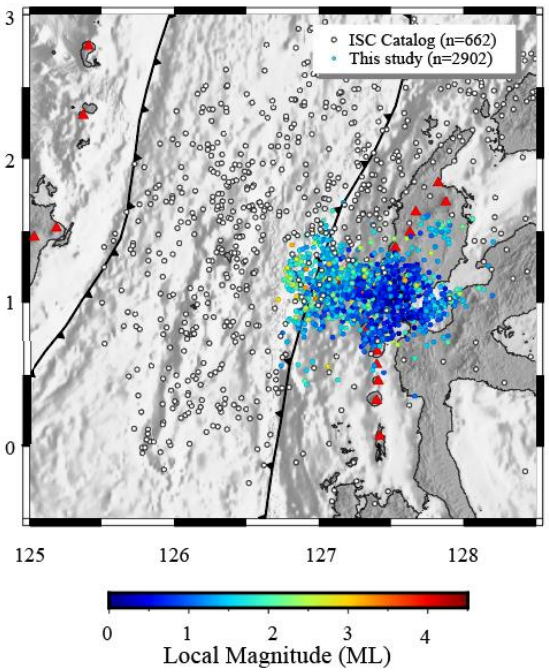


*Figure 7*. Comparison of the Event Distribution from ISC and the obtained catalog in this study

In this study, automatic picks from EQTransformer closely matched manual picks, with only small time differences. While S-phase showed a larger over-time difference,

GaMMA effectively handled these differences by assessing the consistency of phase arrivals across multiple stations during the association process (Zhu et al., 2022). The quality of the waveform from each station also played a critical role in the final results. As noted by Mousavi et al. (2020), variations in the Signal-to-Noise Ratio (SNR) significantly affect the number and reliability of picks, with low SNR conditions leading to lower confidence in automatic phase detections. Compared with previous regional studies, Nguyen et al. (2022) applied EQTransformer with a probability threshold of 0.1 to improve the detection of weak seismic phases, particularly for small-magnitude events. Using the same threshold, our results show that EQTransformer similarly generates a large number of phase picks for low-magnitude earthquakes. Although this approach increases the number of uncertain picks, the dense phase detections provide robust input to GaMMA, enabling the recovery of seismicity commonly absent from conventional earthquake catalogs.

Overall, this pipeline provides an efficient and systematic approach to detecting and clustering earthquake events in subduction zones, such as Jailolo, By combining through preprocessing steps (detrending, resampling, bandpass filtering), deep learning models trained on representative data (Instance dataset) for automatic picking, and probabilistic model-based association methods (BGMM in GaMMA), the workflow achieved strong performance despite limited phase data (either P or S) The high match rate with the ISC catalog, demonstrates that the pre-trained machine learning models EQTransformer and GaMMA can significantly expand earthquake detection capabilities. These results align with previous studies by Zhu et al. (2022) in Hawaii and Puerto Rico, and by Tamaribuchi et al. (2023) for the 2011 Tohoku earthquake, which also reported detecting many more earthquakes than conventional catalogs.

### *4.2. Magnitude completeness*

Figure 8 shows that the estimated Magnitude of Completeness (Mc) from this study is 0.63, which is considerably lower than the Mc values observed for other active subduction zones. Nishikawa & Ide (2014) and White et al. (2019) reported Mc values of approximately 3.0 in the Mexican, Izu-Mariana, and Sumatra-Sunda subduction zones. The low Mc value observed in this study indicates that the earthquake catalog produced here is highly complete, even for small-magnitude earthquakes. This result is consistent with the findings of Wiemer & Wyss (2000), who stated that Mc tends to be lower in areas with dense seismic station coverage and high detection quality. The use of automatic picking techniques and a well-distributed station network in this study enables the detection of microearthquakes that are currently not included in the global catalog.

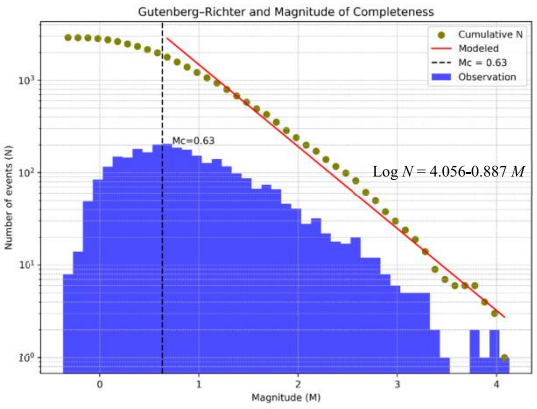


*Figure 8*. Magnitude of completeness (Mc) estimated using the Gutenberg-Richter relation

The frequency-magnitude distribution follows the Gutenberg-Richter relationship, expressed as log N = 4.056–0.887 M, with an estimated b-value of 0.89 (Fig. 8). This b-value is slightly below 1, which is generally interpreted as an indicator of elevated stress levels in the study region (Ghosh et al., 2008). These results align with the b-values documented by Nishikawa & Ide (2014) and

Jiang et al. (2022) for the Sumatra-Sunda subduction system.

### *4.3. Implications for the Jailolo volcanic system*

The findings suggest the potential for magmatic activity beneath Mount Jailolo. As shown in Fig. 9, most detected earthquake events had ML values below 2.0. This pattern is commonly associated with a swarm earthquake, a series of events of similar magnitude without a distinct mainshock, or with micro-earthquakes generated by magmatic fluid pressure or rock fracturing within the volcanic system. Passarelli et al. (2018) reported that a swarm earthquake occurred in Jailolo between November 2015 and February 2016, which they interpreted as the result of a lateral dike intrusion approximately 20 km in length, with a hypocenter migration rate of roughly 10 km/day. This intrusion produced earthquakes characterized by non-double-couple mechanisms (tensile opening) and led to the formation of surface cracks observed on the southern flank of Mount Jailolo. The earthquake catalog developed in this study, based on data collected from August 2016 to July 2017, revealed a notable increase in seismic activity during specific periods, including September -October 2016, December 2016, and February-March 2017. These increases may represent a pre-swarm phase or a continuation of the magmatic disturbance that began in 2015. Seismic activity reached its peak on 29 September, 2017, coinciding with a new seismic swarm event in the Jailolo region, as documented in the BMKG and further analyzed by Rachman (2023).

It is important to note that the study region is located near an active volcanic system, where volcanic-related seismicity may co-occur with tectonic earthquakes. EQTransformer was originally trained on the Stanford Earthquake Dataset (STEAD), which primarily consists of tectonic earthquakes and does not explicitly include volcanic long-period events. Recent studies have shown that deep-learning phase pickers trained on tectonic datasets may perform poorly when applied to volcanic seismicity, particularly for low-frequency and emergent waveforms characteristic of long-period events (Zhong & Tan, 2024).

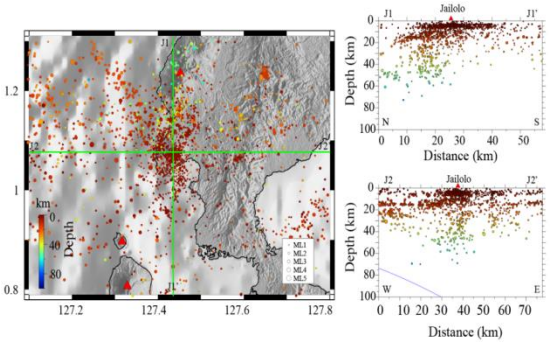


*Figure 9*. Horizontal and vertical distribution of seismic events around Mount Jailolo

## 6. Conclusions

This study used an automated workflow by combining EQTransformer for phase detection and GaMMA for event association, enabling efficient processing of continuous seismic data from the global network. While the development of phase pickers specifically trained on volcanic datasets represents an important direction for future research, the successful implementation of the EQTransformer, GaMMA workflow in this study demonstrates its effectiveness for enhancing earthquake catalogs in regions with complex tectonic settings and sparse monitoring networks, including volcanic environments, across Indonesia and similar regions worldwide. The workflow initially identified 6,174 events, After relocation and quality assessment, the final catalog contained 2,902 events with magnitudes ranging from $< 0.5$ to $> 4$. Depth distribution analysis showed that 81% of the events were shallow (0–50 km), while the remaining 19% were

intermediate-depth (50–200 km). Notably, numerous oceanic earthquakes were detected, despite the station network being primarily land-based.

The resulting earthquake catalog showed enhanced completeness and resolution, with a magnitude of completeness (Mc) of 0.63, significantly lower than values reported in other active subduction zones. This improvement reflects the effectiveness of automated picking and dense station coverage, enabling the detection of microearthquakes that are often missed in global catalogs. The Gutenberg-Richter catalog yielded a b-value of 0.89, indicating elevated stress in the study region.

Analysis of the temporal and spatial distribution of earthquakes indicated potential magmatic activity beneath Mount Jailolo. Most recorded events had a Local Magnitude ($M_L$) below 2.0, a pattern characteristic of earthquake swarms or microearthquakes related to magmatic fluid movement and rock fracturing.

## Acknowledgements

Seismic data were obtained from the GEOFON data center of the GFZ Helmholtz Center for Geosciences.

The author(s) would like to express sincere gratitude to ASP through the Postdoctoral Fellowship program of the National Research and Innovation Agency for valuable support in data curation and research analysis, which greatly contributed to the completion of this study.

## Author contributions

SN: conceptualization, methodology, formal analysis, visualization, writing - original draft, and writing. TA: methodology, supervision, validation, review & editing. FS: data curation, supervision, validation. AT: supervision, review & editing. ASP: data curation, supervision.

## Data availability

The seismic waveform data used in this study were obtained from the 7G FDSN Seismic Network (available at: https://www.fdsn.org/networks/detail/7G_2016/) and cover the period from August 2016 to July 2017. The earthquake catalog was sourced from the International Seismological Center (ISC) (available at: https://www.isc.ac.uk) for the period 2016–2017. The SeisBench codebase used for automatic phase picking and associating is openly available at: https://github.com/seisbench/seisbench.

## References

Ammon C.J., Velasco A.A., Lay T., Wallace T.C., 2021. Earthquake size & descriptive earthquake statistics. In C.J. Ammon, A.A. Velasco T. Lay, T.C. Wallace (Ed.), Foundations of Modern Global Seismology (Second Edition). Academic Press, 197–222. https://doi.org/10.1016/B978-0-12-815679-7.00014-8.

Apandi T., Sudana D., 1980. Peta geologi lembar Ternate, Maluku Utara. Geological Research and Development Centre, Bandung, Indonesia, 9.

Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), 2017. Siaran pers gempabumi swarm di Jailolo Prov. Maluku Utara, 29 September 2017 [Press release on earthquake swarm in Jailolo, North Maluku]. Pusat Gempa Nasional. https://www.bmkg.go.id/siaran-pers/siaran-pers-gempabumi-swarm-di-jailolo-prov-maluku-utara (accessed 8 September 2025).

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. https://doi.org/10.1785/gssrl.81.3.530.

Ghosh A., Newman A., Thomas A., Farmer G., 2008. Interface locking along the subduction megathrust from b-value mapping near Nicoya Peninsula, Costa Rica. Geophysical Research Letters, 35(1), L01301. https://doi.org/10.1029/2007GL031617.

Gunawan E., Ghozalba F., Syauqi, Widiastomo Y., Meilano I., Hanifa N.R., Daryono Hidayati S.,

2016. Field Investigation of the November to December 2015 Earthquake Swarm in West Halmahera, Indonesia. Geotechnical and Geological Engineering, 35(1), 425–432. https://doi.org/10.1007/s10706-016-0117-4.

Jiang C., Zhang P., White M.C.A., Pickle R., Miller M.S., 2022. A Detailed Earthquake Catalog for Banda Arc-Australian Plate Collision Zone Using Machine-Learning Phase Picker and an Automated Workflow, The Seismic Record, 2(1), 1–10. https://doi.org/10.1785/0320210041.

Kagan Y.Y., 2010. Earthquake size distribution: Power-law with exponent?. In Y. Y. Kagan (Ed.), Tectonophysics, Elsevier, 490(1–2), 103–114. https://doi.org/10.1016/j.tecto.2010.04.034.

Kennett B.L.N., Engdahl E.R., Buland R., 1995. Constraints on seismic velocities in the Earth from travel times. Geophysical Journal International, 122(1), 108–124. https://doi.org/10.1111/j.1365-246X.1995.tb03540.x.

Kubo H., Naoi M., Kano M, 2024. Recent advances in earthquake seismology using machine learning. Earth, Planets and Space, 76(1), 36. https://doi.org/10.1186/s40623-024-01982-0.

Lomax A., Curtis A., 2001. Fast, probabilistic earthquake location in 3D models using oct-tree importance sampling. Geophys. Res. Abstr., 3(955), 10–1007.

Macpherson C.G., Hall R., 2002. Timing and tectonic controls in the evolving orogen of SE Asia and the western Pacific and some implications for ore generation. In C. G. Macpherson & R. Hall (Eds.), Geological Society, London, Special Publications. Geological Society of London, 204(1), 49–67. https://doi.org/10.1144/GSL.SP.2002.204.01.04.

McBrearty I.W., Beroza G.C., 2023. Earthquake phase association with graph neural networks. Bulletin of the Seismological Society of America, 113(2), 524–547. https://doi.org/10.1785/0120220182.

Mignan A., 2014. The debate on the prognostic value of earthquake foreshocks: a meta-analysis. Scientific Reports, 4, 4099. https://doi.org/10.1038/srep04099.

Morris J.D., Jezek P.A., Hart S.R., Hill J.B., 1983. The Halmahera Island Arc, Molucca Sea collision zone, Indonesia: A geochemical survey. The Tectonic and Geologic Evolution of Southeast Asian Seas and Islands: Part 2. Geophysical Monograph Series, 27, 373–387. https://doi.org/10.1029/GM027p0373.

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), 3952. https://doi.org/10.1038/s41467-020-17591-w.

Nanjo K.Z., Schorlemmer D., Woessner J., Wiemer S., Giardini D., 2010. Earthquake detection capability of the Swiss Seismic Network. Geophysical Journal International, 181(3), 1713–1724. https://doi.org/10.1111/j.1365-246X.2010.04593.x.

Naoi M., Nakatani M., Otsuki K., et al., 2015. Steady activity of microfractures on geological faults loaded by mining stress. Tectonophysics, 649, 100–114. https://doi.org/10.1016/j.tecto.2015.02.025.

Nguyen C.N., Van D.N., Thi G.H., Le M.N., Huang B.S., Tien H.N., Le Quang K., Huu H.N., 2022. Automatic earthquake detection and phase picking in Muong Te, Lai Chau region: an application of machine learning in observational seismology in Vietnam. Vietnam Journal of Earth Sciences, 44(3), 430–446. https://doi.org/10.15625/2615-9783/17253.

Nishikawa T., Ide S., 2014. Earthquake size distribution in subduction zones linked to slab buoyancy. Nature Geoscience, 7(12), 904–908. https://doi.org/10.1038/ngeo2279.

Nugraha A.D., Supendi P., Widiyantoro S., Daryono, Wiyono S., 2018. Hypocenter relocation of earthquake swarm around Jailolo volcano, North Molucca, Indonesia using the BMKG network data: time periods of September 27-October 10, 2017. In AIP conference proceedings. AIP Publishing LLC, 1987(1), 020093. https://doi.org/10.1063/1.5047378.

Passarelli L., Heryandoko N., Cesca S., Rivalta E., Rasmid, Rohadi S., Dahm T., Milkereit C., 2018.

Magmatic or Not Magmatic? The 2015–2016 Seismic Swarm at the Long-Dormant Jailolo Volcano, West Halmahera, Indonesia. Frontiers in Earth Science, 6, 79. https://doi.org/10.3389/feart.2018.00079.

Passarelli L., Heryandoko N., Muzli, Rasmid, Zimmer M., Cesca S., Rivalta E., Rohadi S., Merdianto U., Dahm T., Milkereit C., 2016. Jailolo network, West Halmahera - Indonesia 2016/2018 [Data set]. GFZ Data Services. https://doi.org/10.14470/2T7561654577.

Pita-Slim O., Chamberlain C.J., Townend J., Warren-Smith E., 2023. Parametric testing of EQTransformer's performance against a high-quality, manually picked catalog for reliable and accurate seismic phase picking. The Seismic Record, 3(4), 332–341. https://doi.org/10.1785/0320230024.

Rachman G., 2023. Tomografi Seismik Waktu Tempuh Untuk Penentuan Struktur 3-D Vp, Vs, Dan Rasio Vp/Vs Di Zona Tumbukan Laut Maluku Dan Implikasinya Terhadap Aktivitas Tektonik Dan Gunungapi Di Halmahera (Disertasi Doktoral, Institut Teknologi Sepuluh Nopember).

Richter C.F., 1935. An Instrumental Earthquake Magnitude Scale. Bulletin of the Seismological Society of America, 25, 1–32.

Shearer P.M., 1999. Introduction to Seismology. Cambridge University Press, Cambridge.

Tamaribuchi K., Kudo S., Shimojo K., Hirose F., 2023. Detection of hidden earthquakes after the 2011 Tohoku earthquake by automatic hypocenter determination combined with machine learning. Earth, Planets and Space, 75(1), 155. https://doi.org/10.1186/s40623-023-01915-3.

Waldhauser F., 2001. HypoDD-A program to compute double-difference hypocenter locations. U.S. Geological Survey Open-File Report 2001-113. https://doi.org/10.3133/ofr01113.

Waldhauser, F., 2000, A Double-Difference Earthquake Location Algorithm: Method and Application to the Northern Hayward Fault, California: Bulletin of the Seismological Society of America, 90, 1353–1368. Doi: 10.1785/0120000006.

White M.C.A., Ben-Zion Y., Vernon F.L., 2019. A Detailed Earthquake Catalog for the San Jacinto Fault-Zone Region in Southern California: Journal of Geophysical Research: Solid Earth, 124, 6908–6930. Doi: 10.1029/2019JB017641.

Wiemer S., Wyss M., 2000. Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the Western United States, and Japan. Bulletin of the Seismological Society of America, 90(4), 859–869. https://doi.org/10.1785/0119990114.

Woollam J., Münchmeyer J., Tilmann F., Rietbrock A., Lange D., Bornstein T., Diehl T., Giunchi C., Haslinger F., Jozinović D., Michelini A., Saul J., Soto H., 2022. SeisBench-A toolbox for machine learning in seismology. Seismological Research Letters, 93(3), 1695–1709. https://doi.org/10.1785/0220210324.

Wulandari A., Anggraini A., Suryanto W., 2018, Hypocenter Analysis of Aftershocks Data of the Mw 6.3, 27 May 2006 Yogyakarta Earthquake Using Oct-Tree Importance Sampling Method: Applied Mechanics and Materials, 881, 89–97. Doi: 10.4028/www.scientific.net/AMM.881.89.

Zhang Q., Guo F., Zhao L., Wu Y., 2017. Geodynamics of divergent double subduction: 3-D numerical modeling of a Cenozoic example in the Molucca Sea region, Indonesia. Journal of Geophysical Research: Solid Earth, 122(5), 3977–3998. https://doi.org/10.1002/2017JB013991.

Zhong Y., Tan Y.J., 2024. Deep learning-based phase picking for volcano tectonic and long-period earthquakes. Geophysical Research Letters, 51(12), e2024GL108438. https://doi.org/10.1029/2024GL108438.

Zhu W., Hou A.B., Yang R., Datta A., Mousavi S.M., Ellsworth W.L., Beroza G.C., 2023. QuakeFlow: a scalable machine-learning-based earthquake monitoring workflow with cloud computing. Geophysical Journal International, 232(1), 684–693. https://doi.org/10.1093/gji/ggac355.

Zhu W., McBrearty I.W., Mousavi S.M., Ellsworth W.L., Beroza G.C., 2022. Earthquake phase association using a Bayesian Gaussian mixture model. Journal of Geophysical Research: Solid Earth, 127(5), e2021JB023249. https://doi.org/10.1029/2021JB023249.

SUPPLEMENTARY FIGURE

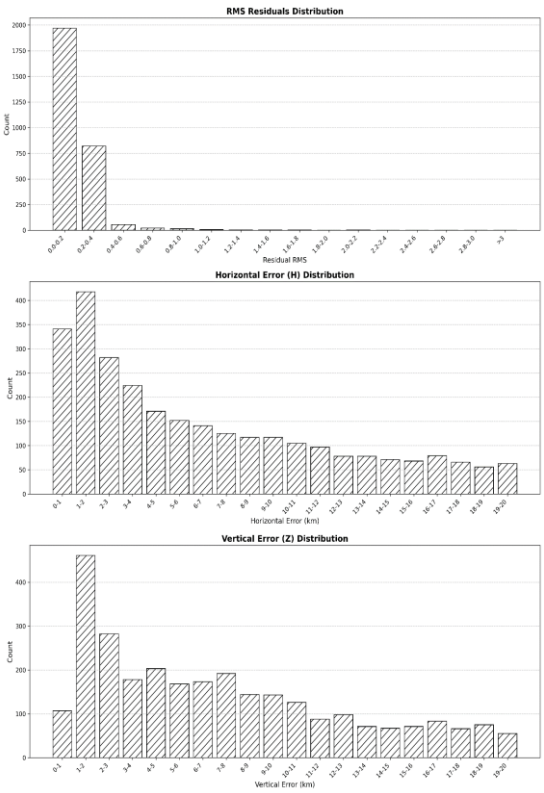


*Figure S1.* Lateral error and RMS distribution results from NonLinLoc Processing

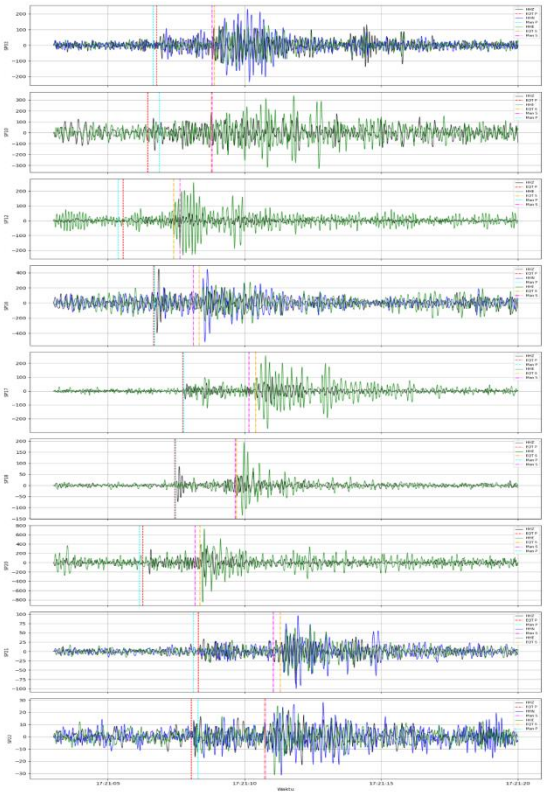

*Figure S2.* Comparison of picking quality between EQTransformer and manual methods, sampling of phase picks for event 2016/08/29-17:21:03

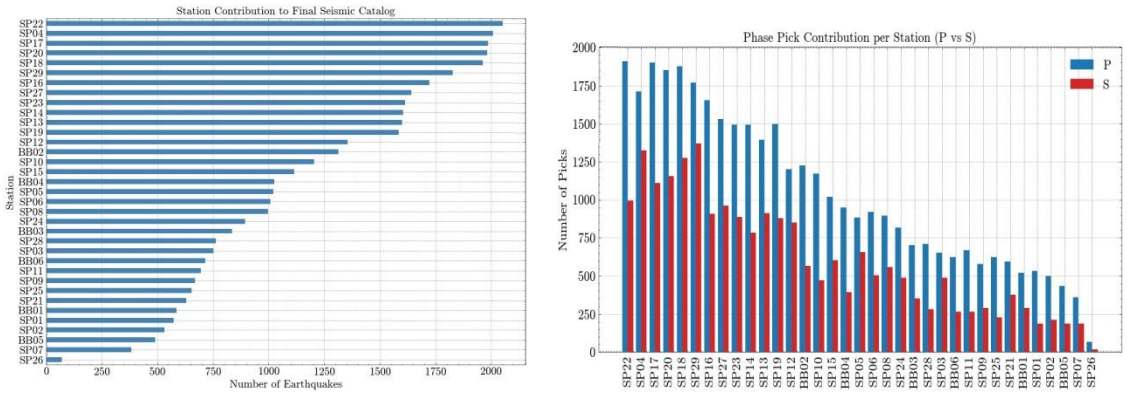


*Figure S3.* Number of associated phase picks (P and S) and associated earthquake events contributed by each seismic station

SUPPLEMENTARY TABLE

*Table T1.* Metadata of Seismic Observation Stations in the Study Area

| | ID | Longitude (°E) | Latitude (°N) | Elevation(m) |
|---|---|---|---|---|
| 0 | 7G.BB01 | 127.4133 | 1.0399 | 5 |
| 1 | 7G.BB02 | 127.4488 | 1.1599 | 83 |
| 2 | 7G.BB03 | 127.6427 | 1.0036 | 22 |
| 3 | 7G.BB04 | 127.515 | 0.9140 | 31 |
| 4 | 7G.BB05 | 127.8942 | 1.1865 | 18 |
| 5 | 7G.BB06 | 127.3291 | 0.8894 | 31 |
| 6 | 7G.SP01 | 127.3817 | 0.8284 | 29 |
| 7 | 7G.SP02 | 127.3799 | 0.8126 | 0 |
| 8 | 7G.SP03 | 127.4476 | 1.0538 | 43 |
| 9 | 7G.SP04 | 127.4666 | 1.0613 | 20 |
| 10 | 7G.SP05 | 127.4699 | 1.0840 | 24 |
| 11 | 7G.SP06 | 127.4818 | 1.1164 | 26 |
| 12 | 7G.SP07 | 127.4683 | 1.1420 | 10 |
| 13 | 7G.SP08 | 127.4785 | 1.0805 | 11 |
| 14 | 7G.SP09 | 127.4339 | 1.1510 | 19 |
| 15 | 7G.SP10 | 127.4921 | 1.1467 | 39 |
| 16 | 7G.SP11 | 127.4671 | 1.1014 | 7 |
| 17 | 7G.SP12 | 127.4275 | 1.1098 | 26 |
| 18 | 7G.SP13 | 127.4141 | 1.0902 | 124 |
| 19 | 7G.SP14 | 127.4276 | 1.0515 | 32 |
| 20 | 7G.SP15 | 127.4989 | 1.0613 | 8 |
| 21 | 7G.SP16 | 127.5028 | 1.0713 | 31 |
| 22 | 7G.SP17 | 127.5160 | 1.0144 | 172 |
| 23 | 7G.SP18 | 127.4902 | 1.0124 | 24 |
| 24 | 7G.SP19 | 127.4768 | 1.0078 | 16 |
| 25 | 7G.SP20 | 127.4899 | 1.1004 | 27 |
| 26 | 7G.SP21 | 127.5148 | 0.9750 | 71 |
| 27 | 7G.SP22 | 127.5182 | 0.9880 | 111 |
| 28 | 7G.SP23 | 127.4134 | 1.0414 | 27 |
| 29 | 7G.SP24 | 127.4492 | 1.1094 | 24 |
| 30 | 7G.SP25 | 127.4806 | 1.1327 | 27 |
| 31 | 7G.SP26 | 127.3668 | 0.7720 | 63 |
| 32 | 7G.SP27 | 127.4082 | 1.0532 | 11 |
| 33 | 7G.SP28 | 127.3148 | 0.9093 | 44 |
| 34 | 7G.SP29 | 127.4199 | 1.0888 | 208 |

*Table T2.* HypoDD Parameters

| Parameters | Setting Value |
|---|---|
| Inversion method | LSQR least-squares solver |
| Number of iteration sets | 6 |
| Catalog phase weights | P = 1, S = 1 |
| Residual cutoff | Reduced from 4 s to 1 s |
| Damping factor | 100 |
| Clustering parameters | Maximum station distance = 150 km |
| | Maximum event separation = 30 km |
| | Maximum neighbors per event = 20 |
| | Minimum links to define neighbors = 3 |
| | Minimum links per event pair = 3 |
| | Maximum links per event pair = 20 |

*Table T3.* **A**rrival-time differences (s) between EQTransformer (EQT) automatic picks and manual picks

| Event 29 August 2016 (17:21:03) | | | | |
|---|---|---|---|---|
| **Stasiun** | **Delta P (s)** | **Probability P** | **Delta S (s)** | **Probability S** |
| SP03 | 0.134 | 0.9091 | 0.093 | 0.1575 |
| SP10 | 0.413 | 0.9078 | 0.019 | 0.4068 |
| SP12 | 0.167 | 0.9430 | 0.238 | 0.1685 |
| SP16 | 0.030 | 0.6189 | 0.214 | 0.9220 |
| SP17 | 0.036 | 0.5922 | 0.244 | 0.8595 |
| SP18 | 0.044 | 0.5042 | 0.048 | 0.6116 |
| SP20 | 0.118 | 0.6265 | 0.176 | 0.1398 |
| SP21 | 0.187 | 0.7323 | 0.261 | 0.6804 |
| SP22 | 0.245 | 0.8960 | 0.043 | 0.5306 |
| **Mean** | **0.018** | | **0.081** | |