

FULL PAPER

Open Access



Small earthquake location via machine learning with insufficient data

Ji Zhang^{1,2*}, Aitaro Kato¹, Huiyu Zhu^{2,3} and Jie Zhang²

Abstract

A comprehensive earthquake catalog plays a crucial role in enhancing our understanding of earthquake activity and generation mechanisms. However, due to the low station density and data quality limitations, numerous small earthquakes remain undetected and unlocated. Traditional seismic location methods based on travel time and waveform analysis may be ineffective for such events, while traditional single-station location methods require high signal-to-noise ratio (SNR) data. Overcoming this challenge and improving the detection and location of these small seismic events is crucial. To address the need for locating small seismic events with low SNR, we propose a novel workflow for seismic station networks, leveraging machine learning single-station location method. The method comprises distance and azimuth neural networks. We pre-train the models on a global dataset (STEAD) and fine-tune them using local datasets (INSTANCE, Italy and Sichuan, China). The local datasets are utilized to assess the performance of models, while the Sichuan dataset is also specifically used to evaluate the entire workflow. We evaluate the proposed machine learning methods using the Sichuan, China, testing dataset and achieve a mean absolute error of approximately 3.0 km for epicenter distance and 22.0 degrees for back-azimuth. Extending the application of the models to regions like Yunnan, China, and Italy generates reliable estimates of the spatial distribution of seismic events. Crucially, the workflow incorporates both spatial constraints of the station locations and the constraints from waveforms recorded by the stations, leading to improved location accuracy compared to using less reliable azimuth estimates. The workflow successfully located 6321 seismic events in the Sichuan, China, testing dataset, which is approximately 1.4 times more than the number reported in earthquake catalogs, further complementing seismic activity.

Keywords Machine learning, Epicenter distance, Back azimuth, Single-station location, Location workflow

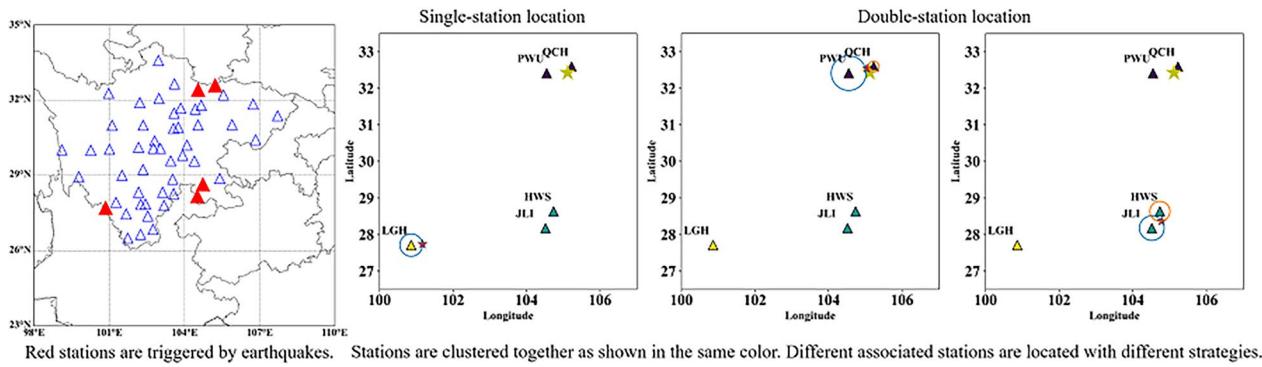
*Correspondence:

Ji Zhang

zhangj@eri.u-tokyo.ac.jp

Full list of author information is available at the end of the article

Graphical Abstract



1 Introduction

Locating earthquakes is a fundamental and critical challenge in seismology, playing a key role in advancing research in the field. Despite significant efforts, it remains a complex task. Seismologists continuously strive to enhance their methods for enhancing earthquake locations. Improving these methods is essential for understanding earthquake processes, assessing seismic hazards, and monitoring regional seismic activity. Over time, notable progress has been made in earthquake location techniques. There exist a range of methods, from traditional approaches based on travel time and waveform analysis to advanced techniques incorporating machine learning algorithms. By refining these methods, the seismological community aims to achieve more precise and reliable earthquake locations, contributing to a deeper understanding of the seismic events' generation mechanisms.

Travel time-based seismic location methods typically involve several steps: seismic event detection, seismic phase picking, phase association, and inversion based on the travel time of associated phases. Common inversion algorithms include the iteratively linearized travel time inversion method (Geiger 1912) and nonlinear methods such as simulated annealing, genetic algorithms (Kennett and Sambridge 1992), and Monte Carlo searches (Sambridge 2002). In the era of artificial intelligence (LeCun et al. 2015; Goodfellow et al. 2016), machine learning techniques are revolutionizing earthquake detection, phase picking, and association. Convolutional neural networks (CNNs) have played a pivotal role in advancing seismic event detection. For instance, Perol et al. (2018) introduced a CNN-based method that enabled the detection of more than 17 times the number of earthquakes compared to traditional catalogs. In phase picking, deep learning models such as PhaseNet (Zhu et al.

2018), which employs U-Net (Ronneberger et al. 2015) architecture, have been developed to accurately pick P- and S-wave phases. Furthermore, global models like the EQTransformer (EQT) by Mousavi et al. (2020) offer simultaneous earthquake detection and phase picking, significantly improving the accuracy and efficiency of seismic data analysis. In phase association, methods like GaMMA (Zhu et al. 2022), the problem is treated as an unsupervised clustering problem, where phases from a network of seismic stations are grouped into clusters corresponding to individual seismic events. Once the associated travel times are determined, conventional seismic location techniques can be employed to estimate the event locations. These advances in machine learning-based seismic event detection have greatly enhanced the data quality for earthquake localization and monitoring.

Waveform-based seismic location methods, on the other hand, do not require prior phase identification or phase picking. These methods can detect and locate events with low SNRs by using migration and stacking techniques commonly employed in exploration seismology (e.g., Beskarde et al. 2018; Cesca and Grigoli 2015; Gajewski et al. 2007; Pesicek et al. 2014; Zhebel and Eisner 2015). The basic principle of waveform-based methods is to locate the source by focusing or reconstructing the source energy into a discrete grid of points using a certain migration or imaging operator. Recently, a neural network-based approach has been developed that directly utilizes the seismic waveforms from multiple stations for was mentioned in the manuscript; however, this was not included in the reference list. As a rule, all mentioned references should be present in the reference list. Please provide the reference details to be inserted in the reference list and ensure that all references are in alphabetical order." location. This network takes as input the three-component multi-station seismic waveforms and

outputs the Gaussian probability distribution of the seismic event location (Zhang et al. 2020).

The worldwide deployment of seismic networks generates vast volumes of high-quality seismograms, containing valuable information on small earthquakes. As earthquake monitoring evolves, the need for more complete earthquake catalogs has grown, increasing the demand for location of small earthquakes. Improved seismic detection capabilities have uncovered many previously undetected events, enhancing the completeness of earthquake catalogs. Many studies suggest that a more complete earthquake catalog could shed light on the mechanisms of seismogenesis and potentially pave the way for earthquake forecasting (Brodsky 2019). Accurate location is essential for studying these small earthquakes. Most undetected seismic events identified by machine learning are small earthquakes or low SNR events, making them difficult to locate using traditional methods. The travel time-based location methods typically require travel times from at least three stations in a seismic network. However, due to the sparsity of seismic networks or weak energy radiated by small earthquakes, seismic signals may often be recorded by one or two nearby stations only. Consequently, travel time-based methods using multiple stations cannot be applied to localize such seismic signals. Similarly, waveform-based location methods, which rely on dense seismic station networks, are also unsuited for locating these small events.

The single-station location method (Magotra and Chael 1987; Frohlich and Pulliam 1999) offers a viable solution for locating seismic events, especially when seismic networks have insufficient station coverage, mitigating the limitations arising from network sparsity. Traditional single-station location methods determine epicentral distances using the differences in P- and S-wave arrival times and obtain azimuths from P-wave polarizations (Magotra and Chael 1987; Frohlich and Pulliam 1999). These methods often require analysis and quality control from an experienced seismologist and rely on high SNR to accurately pick the arrival times of P and S waves and determine the initial motion direction of the P wave. To overcome these challenges, machine learning classifiers have been introduced to solve single-station earthquake location problems without human intervention. Perol et al. (2018) trained a CNN to simultaneously classify events (vs. noise) and group earthquakes into six clusters initially defined by K-means in Oklahoma. Lomax et al. (2019) expanded on the method by classifying seismic waveforms into multiple classes. Both models have been shown to have large errors (Mousavi and Beroza 2020). Mousavi and Beroza (2020) developed another deep learning method for earthquake location, addressing a regression problem using two separate Bayesian neural

networks that provided the predicted result uncertainty estimation. Their model allowed for estimating confidence intervals in the final event location. However, this model required earthquake waveforms with an SNR higher than 25 decibels, which may not always be achievable in cases analyzing small-magnitude events with low SNR.

To address the need for a fast and efficient method to locate seismic events using single-station data, particularly in scenarios with unclear seismic phase arrival times and low SNRs, we propose a neural network-based approach for single-station waveform location. This method provides reasonably accurate event locations even in low SNR conditions. Additionally, we develop a workflow for locating small earthquakes that considers the spatial constraints of seismic station networks. By leveraging machine learning, clustering algorithms, and statistical techniques, we integrate diverse data-driven methods to overcome challenges such as low SNR and limited station coverage, improving the accuracy and reliability of seismic event localization. In machine learning, model training typically requires the initialization of weights, which are usually set randomly. However, transfer learning allows us to use the parameters from a well-trained model as the initial parameters when training models in other regions. Transfer learning in machine learning refers to the process of leveraging a pre-trained model developed for one task as the starting point for a related task. Instead of training a model from scratch, which can be time-consuming and require large amounts of data, transfer learning enables the reuse of knowledge gained from a source domain or task to improve performance and reduce training time in a target domain or task. In this study, we pre-train models on a global dataset (STEAD; Mousavi et al. 2019) containing a large number of records from various regions. This pre-training provides a baseline model for transfer learning, which is then fine-tuned to specific areas, such as Sichuan, Yunan, and Italy. No additional modifications are needed when performing transfer learning.

This paper is organized as follows: first, we introduce the seismic network location workflow and core algorithm for scenarios with insufficient stations. Next, we evaluate the location approach using actual data. Finally, we discuss the performance of our method and the prospects for its further development and application.

2 Methods

The workflow for locating small earthquakes is illustrated in Figure S1. It mainly consists of earthquake detection, clustering, epicentral distance prediction, back-azimuth prediction, location using the least squares method, and magnitude and origin time estimations. The primary

step is to predict the epicentral distance and azimuth of the seismic signals recorded at each seismic station. Then, the seismic records are associated based on spatial clustering, temporal correlation, and waveform similarity. This allows for the accurate identification of seismic records belonging to the same seismic event. Figure 1 presents a real data example demonstrating this process.

The above steps are described in further detail below.

2.1 Earthquake detection

Our methodology is based on an already existing and extensively tested machine-learning-based earthquake detection approach, EQT developed by Mousavi et al. (2020), and we did not design or train a new earthquake detection network. Generally speaking, detection process can be tailored to specific needs by selecting an appropriate algorithm, whether it is a classical approach or a machine learning-based method. The goal is to maximize

the detection capability of earthquake events while minimizing the false alarms.

Here, we employ both the EQT model proposed by Mousavi et al. (2020) and the Short-Term Average/Long-Term Average (STA/LTA) method (Allen 1978), combining traditional algorithms with machine learning to further enhance detection performance.

Based on the transformer mechanism, EQT has achieved excellent results in earthquake detection and phase picking. We only use the detection results from EQT and do not focus on phase picking. The input to the EQT is a three-component seismic waveform with a duration of 60 s, and the output is the probability of an event's existence. The default threshold for event detection is set up at 0.3. For the STA/LTA method, we set the long-time window to 8 s and the short-time window to 4 s. Two thresholds, 1.6 and 1.4, are used simultaneously. Events are identified based on two criteria: STA/LTA values exceeding 1.6, and the duration

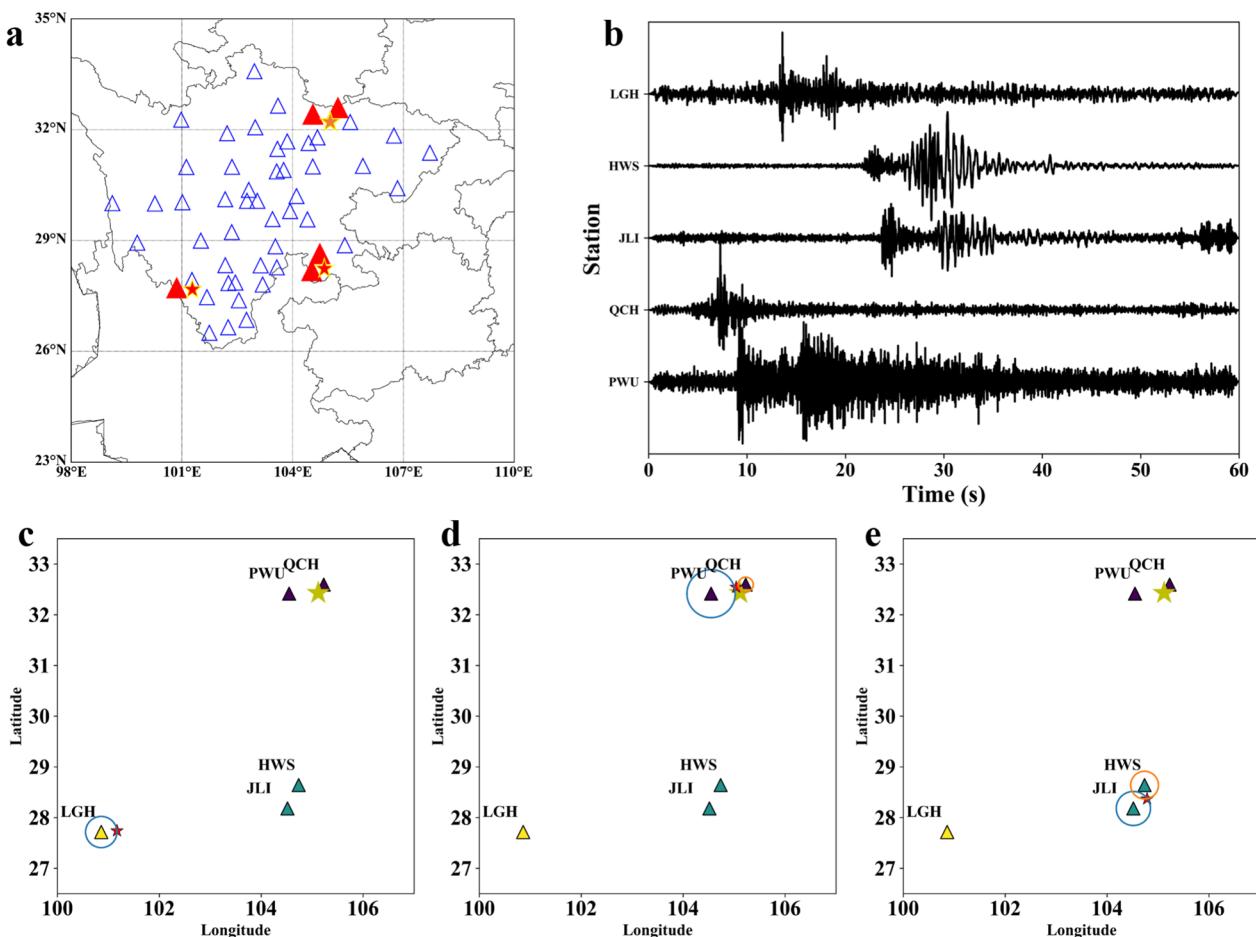


Fig. 1 An example of the location workflow. **a** View of the Sichuan study area and stations. In detection processing, red stations are triggered by selected earthquakes. **b** Event waveforms are detected by the detection scheme. **c–e** In the clustering step, near-field stations of the same event are grouped together and shown in the same color. In the location step, different clustered stations are used for event location under the different strategies, such as **c** single-station and **d** and **e** double-station locations

of STA/LTA values between 1.6 and 1.4 being greater than 2.5 s. These criteria aim to minimize the influence of background noise. Compared with various detection methods, EQT has proven to be one of the better-performing approaches for detection with attention mechanism (Mousavi et al. 2020). EQT can achieve event detection and phase picking at the same time. However, in this study we focus solely on the detection aspect. This is because, in many cases, the EQT network can detect events, but may fail to identify the corresponding P or S phases (Figure S8a) as phase probability values may fall below the threshold. Simply lowering the threshold does not recover previously missed signals and may even introduce more false detections. To assess whether detection performance could be improved by threshold tuning alone, we conducted additional tests by lowering the EQT detection threshold from the default value of 0.3 to 0.1. Although this adjustment led to a slight increase in recall from 82.29 to 83.41%, it also resulted in a decrease in precision from 98.65 to 98.29%. This trade-off highlights a common limitation in threshold-based detection strategies: while lower thresholds can increase sensitivity, they also introduce more false positives. Thus, additionally, to further enhance detection capabilities, we have incorporated the STA/LTA method, which is the most widely used traditional earthquake detection algorithm. Our STA/LTA-based method was applied under strict constraints, including a relatively high trigger threshold and a minimum event duration of 2.5 s. This configuration allowed us to recover additional events missed by EQT while maintaining a high level of precision. The combined approach improved overall detection precision to 98.85% and successfully identified events such as the one shown in Figure S2b, which EQT failed to detect even at low thresholds (0.1). This reflects the inherent limitations of neural networks—even the best-designed network cannot achieve 100% accuracy, in some cases, even for high SNR earthquake signals. However, traditional methods can address this issue effectively, as physical constraints ensure that high SNR signals are always detected. The primary challenge with STA/LTA lies in detecting small events and minimizing false positives. By applying strict constraints, a high threshold and extended duration in the STA/LTA detection method, we can detect only clear and prominent events while suppressing noise-induced triggers and minimizing false detections. This strategy helps to avoid false alarms and complements the limitations of the neural network-based detection.

The detected waveforms and corresponding station locations are saved for subsequent clustering processing. Figure 1a shows an example of the earthquake detection result where the spatial distribution of the seismic stations triggered by the earthquake is highlighted by red color. The corresponding waveforms are shown in Fig. 1b.

2.2 Clustering

The detection step will typically identify many waveforms from different stations that may not originate from the same earthquake. Therefore, we must classify these waveforms to determine which ones belong to the same event. Waveform cross-correlation is one way for associating waveforms (Ashkenazy et al. 2024). However, due to different propagation times, receiving directions, and low SNRs, the cross-correlation method alone may not be sufficient, and additional constraint methods are necessary. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a commonly used density clustering algorithm (Ester et al. 1996). Such algorithms generally assume that categories can be determined by the tightness of the sample distribution. By grouping closely related samples into different categories, we can obtain the result of all clustering categories. The DBSCAN algorithm can be expressed by the following formula:

$$\begin{aligned} D &= \{x_1, x_2, \dots, x_l\} \\ C \in D, d(p, q) &\leq r \forall p, q \in C_i \forall C_i \in C, n(C) \geq E, \end{aligned} \quad (1)$$

where C is a cluster of all set D , x_i is the station location with the detected seismic waveform, $d(p, q)$ is the distance between two points, the distance should be less than the maximum radius r , and the number of C should be greater than the minimum number E .

The DBSCAN clustering method utilizes the recorded station location for spatial clustering. In the Sichuan station network, the minimum and average distances between stations are approximately 20 km and 60 km, respectively. We set the maximum radius r as 110 km to include more nearby stations and the minimum sample number E as 1 to conduct preliminary clustering. Figure 1c–e shows the spatial clustering of triggered stations, with stations of the same color representing the same category.

Although we select seismic waveforms from the same spatial region, they may originate from different earthquakes that occur nearby or at the same location but at different times. Hence, further clustering based on timestamp is necessary. Since we do not pick the phase, we use the timing of the three-component waveform maximum amplitude as the timestamp. Waveforms with a timestamp difference of less than 19 s (estimated by the cluster distance, 110 km) are considered to belong to the same event. We employ subsequent location methods for further processing waveforms belonging to the same event.

2.3 Distance prediction network

We design a network named DisNet for predicting the epicenter distance based on the three-component waveforms, as shown in Fig. 2a. The epicentral distance is calculated in a 2D plane, from the station to the

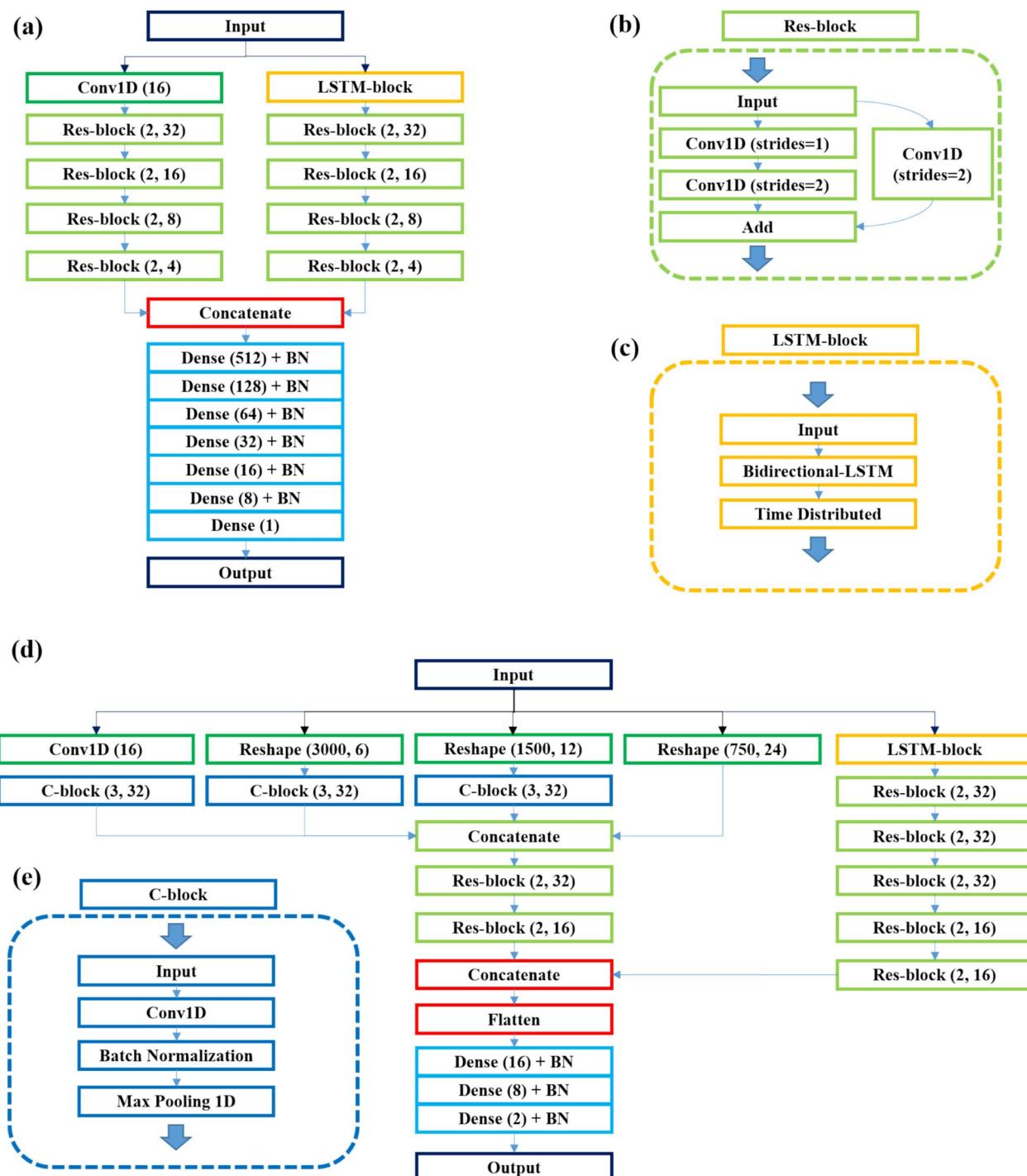


Fig. 2 The architecture of the proposed neural networks. **a** The architecture of DisNet. It mainly consists of Res-block (**b**), LSTM-block (**c**), and full connection layers. Conv1D (16) represents a dimension convolution layer with 16 kernels. The number of the Dense layer represents the number of fully connected layer neurons. BN represents the batch-normalization layer. **b** The structure of Res-block. The two parameters indicate the depth of the ResNet and the number of channels, respectively. **c** The structure of LSTM-block. **d** The architecture of AziNet. **e** The structure of C-block. It consists of Conv1D, BN, and Max Pooling1D layers

epicenter, without considering the effect of depth. The network structure primarily comprises residual blocks (Res-block), Long Short-Term Memory blocks (LSTM-block), and fully connected (FC) layers. The Res-block and LSTM blocks are used to extract waveform features and temporal features, respectively. By combining the extracted waveform and temporal features, we can better analyze the data. There is no optimal number of layers for the network, but we use a deeper 14-layer structure here to better fit the nonlinear relationship between the data and the labels. The Res-block (Fig. 2b) utilizes shortcut connections to remove the same body parts and highlight minor changes. This design helps overcome training convergence challenges caused by increased network layers.

The input to the Res-block is denoted as x , which undergoes two convolution layers represented as F with the activation function being Rectified Linear Unit (ReLU). The function of the Res-block can be expressed as follows:

$$y = \text{ReLU}(F(x) + x). \quad (2)$$

In this processing, $F(x)$ denotes the residual mapping. It is easier to optimize the residual mapping than to optimize the original, direct mapping (He et al. 2016). The Res-block extracts the waveform features and transfers the shallow information to the deep layer to prevent the network from overfitting.

The LSTM-block consists of a bidirectional LSTM (Bi-LSTM) layer and a time-distributed layer. LSTM can solve the long sequence problem effectively by introducing the concepts of the memory cell, input gate, output gate, and forgetting gate. The LSTM (Hochreiter and Schmidhuber 1997) can be described as:

$$\begin{aligned} f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f), \\ i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i), \\ \widetilde{C}_t &= \tanh(W_c[h_{t-1}, x_t] + b_c), \\ C_t &= f_t * C_{t-1} + i_t * \widetilde{C}_t, \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o), \\ h_t &= o_t * \tanh(C_t), \end{aligned} \quad (3)$$

where W and b are weights and biases, σ and \tanh are sigmoid and tanh activation functions, x is the time sequence, h is the hidden state, and C is the memory cell. f_t is the output of the forget gate, x_t is the signal of t time,

h_{t-1} is the hidden state of $t-1$ time. i_t and \widetilde{C}_t are outputs of the input gate, the former represents how often new information is added, and the latter represents new information. After updating the cell C_t , we control how much information needs to be passed to the next state h_t through the output gate o_t .

The memory cell is responsible for preserving important information; the input gate decides whether to write the current input information to the memory cell, the forget gate decides whether to forget the information in the memory cell, and the output gate decides whether to use the information in the memory cell as the current output. The control of these gates can effectively capture important long-term dependencies in the sequence and solve the gradient problem. In this way, LSTM can handle and predict long-time sequence data.

The Bi-LSTM is a combination of a forward LSTM and a backward LSTM so that it can extract information from the above and the following simultaneously. It is often used to model time sequence information in natural language processing tasks. The Bi-LSTM is used to obtain information of the phases. For example, PpkNet, a two-layer bidirectional gated recurrent unit (GRU), is developed to pick P- and S-arrival times (Zhou et al. 2019). The features extracted by the Res-block are concatenated to the time sequence results of the LSTM-block as new features. The epicentral distance is predicted with new features using seven fully connected layers.

The FC layers consist of 7 fully connected layers. The fully connected layer function is as follows:

$$y = wx + b, \quad (4)$$

x and y are the input and output, and w and b represent the weight and the bias. The number of neurons is gradually reduced from 512 to 1. After each fully connected layer, ReLU is used as the activation function. Although it is common practice to apply Batch Normalization before activation functions such as ReLU, we found through experimentation that placing the Batch Normalization layer after the ReLU activation led to better performance in our case. This configuration was therefore adopted in our final model as shown in Fig. 2a. The last layer uses one neuron to predict the epicentral distance information.

The input of DisNet is the three-component seismic data x , and the output of DisNet output is an epicentral distance. In short, the DisNet is computed as:

$$\text{DisNet}(x) = \text{ReLU}(\text{FC}(\text{concat}(\text{Resblock}(x), \text{Resblock}(\text{LSTMblock}(x)))). \quad (5)$$

LSTM-block obtains the time sequence features, Res-block obtains the time waveform features, and FC layers combine time sequence and waveform features to make predictions of epicenter distance. Seismic data contain epicentral distance information extracted from waveform and time sequence features. It is possible that our DisNet model can obtain the P and S phase information and build a nonlinear relationship with the epicenter distance based on this information.

To prevent overfitting, we apply early stopping (Prechelt 2002). Training stops when validation loss does not improve for 10 consecutive epochs. In the context of epicentral distance prediction, there may be some outliers that could negatively impact the training performance if given too much attention. Here, we treat these outliers as damaged data and opt for the mean absolute error (MAE) loss function, which treats all errors equally, regardless of their values. The MAE formula is as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N \text{abs}(y_{\text{true}_i} - y_{\text{pred}_i}), \quad (6)$$

$$\text{Fu}(x) = \text{concat}\left(\text{Cblock}(\text{Ra}(x)) + \text{Cblock}\left(\text{Ra}\left(\frac{x}{2}\right)\right) + \text{Cblock}\left(\text{Ra}\left(\frac{x}{4}\right)\right) + \text{Ra}\left(\frac{x}{8}\right)\right). \quad (7)$$

Ra represents the size shape function. Multi-scale feature extraction is similar to extracting effective information from different frequency bands.

The Res-block and LSTM-block extract signal and time sequence features further. Finally, the 3 fully connected layers are used to fit the results. The input of AziNet is the three-component seismic data x , and the output is the sine and cosine values of the back-azimuth (Mousavi and Beroza 2020). The function of the outputs of AziNet is shown as follows:

$$O(\theta) = [\alpha * \sin(\theta), \alpha * \cos(\theta)], \quad (8)$$

where θ is the back-azimuth and α is the scale factor. Here, α is set as 100. In short, the AziNet is computed as:

$$\text{AziNet}(x) = \text{RELU}(\text{FC}(\text{concat}(\text{Resblock}(\text{Fu}(x)), \text{Resblock}(\text{LSTMblock}(x))))). \quad (9)$$

Generally, people only use the waveform near the P wave to predict the azimuth because it is relatively simple and easy to analyze. We think that the whole waveform contains the azimuth information, so we should use the whole waveform as the input and use the powerful nonlinear fitting ability of the neural network to find the internal connection. Considering the complexity of waveforms, we use multi-scale fusion to extract features from waveforms

where y_{true_i} is the label of i th data; y_{pred_i} is the prediction of i th data; a total of N is in one batch.

The optimizer is the Adaptive Moment Estimation (Adam) algorithm (Kingma & Ba 2014).

2.4 Back-azimuth prediction network

The network (AziNet) is proposed to predict the back-azimuth, and its architecture is shown in Fig. 2d. AziNet comprises the convolution block (C-block), LSTM-block, and Res-block. In addition to considering waveform and temporal features, we also incorporate multi-scale feature fusion to further enhance the model's ability to express features. A total of 12 layers are used here, but they include various blocks for fitting nonlinear relationships. The C-block consists of convolution, batch normalization, and max-pooling layers, as shown in Fig. 2e. It is used to extract signal features and reduce the dimension of the data. AziNet uses multi-scale fusion to extract various features. The original data are reshaped to the size of 1500*6, 750*12, and 375*24. The multi-scale information is processed and then concatenated. The multi-scale fusion is computed as:

$$\text{Fu}(x) = \text{concat}\left(\text{Cblock}(\text{Ra}(x)) + \text{Cblock}\left(\text{Ra}\left(\frac{x}{2}\right)\right) + \text{Cblock}\left(\text{Ra}\left(\frac{x}{4}\right)\right) + \text{Ra}\left(\frac{x}{8}\right)\right). \quad (7)$$

at different scales. Additionally, we consider the effect of azimuth changes in the time dimension. Although it is not common to place Batch Normalization after the final layer in regression models, we found that this configuration improved prediction stability and reduced overfitting in our azimuth estimation task. During predicting process, the Batch Normalization layer uses the running statistics computed during training, ensuring deterministic and stable outputs.

To prevent overfitting, we also apply early stopping (Prechelt 2002). Training stops when validation loss does not improve for 10 consecutive epochs. When fitting two parameters simultaneously, larger errors are more likely to occur. Utilizing mean square error (MSE), which is more

sensitive to outliers, can accelerate convergence and lead to a more stable solution. The MSE formula is as follows:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_{\text{true}_i} - y_{\text{pred}_i})^2. \quad (10)$$

Optimizer is also Adam algorithm.

2.5 Earthquake location

Different location strategies are adopted based on the number of received waveforms. If an event is detected by two or more stations (Fig. 1d), both the station location and the predicted epicenter distances are used to locate the event. In this case, the location method is the least square method, and the space average location of these stations is taken as the initial value to iterate until the error meets the requirements or becomes 0:

$$\begin{aligned} L &= \frac{1}{2N} \sum_{i=1}^N (r_i - d_i)^2, \\ r_i &= \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2}, \end{aligned} \quad (11)$$

where L is the objective function. x_i, y_i represents the space coordinates of the i th station. d_i represents the epicentral distance at the i th station. x_0, y_0 represents the source location coordinates. We expect the minimum error between the calculated epicentral distance and the predicted epicentral distance based on the location of the source and the station.

Newton method is used to update the location of the source, with alpha as the step length, the iterative formula is as follows:

$$\begin{aligned} x_0 &= x_0 - \alpha \frac{1}{N} \sum_{i=1}^N (r_i - d_i)(x_0 - x_i)/r_i, \\ y_0 &= y_0 - \alpha \frac{1}{N} \sum_{i=1}^N (r_i - d_i)(y_0 - y_i)/r_i. \end{aligned} \quad (12)$$

Notably, when only two stations receive the waveform, there are two optimal solutions when the sum of predicted epicentral distances is greater than the distance between the two stations (Fig. 1e). After the least square method is adopted, one of the optimal solutions can be obtained, and the other one can be obtained through the symmetric relationship with the station line. Then, the predicted azimuth angle is used to constrain the location of the real source. We perform location estimation separately based on the predicted epicentral distance and azimuth by two stations. We calculate the distance between the two predicted source locations (just using predicted epicentral distances and azimuths) and the two source locations obtained by the least squares method. The source location obtained by the least squares method that is closest to the two predicted locations is taken as the final earthquake location.

(x_1, y_1) and (x_2, y_2) represent the space position of the coordinates of two stations, (x_0, y_0) is one optimal solution calculated by the least square method, and another optimal solution (x_s, y_s) of the formula is:

$$(x_s, y_s) = (x_1 + x_2 - x_0, y_1 + y_2 - y_0). \quad (13)$$

When the sum of predicted epicentral distances is significantly less than the distance between the two stations, it is considered that the two waveforms do not belong to the same event, and a single station location method is used to obtain the locations of the corresponding events.

In case the event is recorded by only one station, we use the three-component waveform of that single station is used to predict the epicentral distance and back-azimuth for obtaining event location.

2.6 Magnitude estimation

In this study, we obtain the maximum amplitude from the S-wave part of each station three-component waveform, along with the epicentral distance and magnitude of the earthquake within the Sichuan region. Utilizing an empirical formula, we fit these parameters to establish an empirical relationship between magnitude, epicentral distance, and maximum amplitude. This empirical relationship assesses the magnitude of the events we detect and locate. To calculate the arrival time of S waves, we use Taup software (Crotwell et al. 1999) with the IASP91 velocity model (Kennett & Engdahl 1991), which considers both the source and station locations. The time location of maximum waveform amplitude is used as an approximation for the S-wave arrival time, from which we deduce the earthquake origin time. As our study primarily focuses on small earthquakes in the Sichuan region of China, we evaluate the empirical relationship for earthquakes ($M < 2$) in this region (Figure S3) and obtain the following result:

$$\text{Mag} = \log(A_{\max}) + 1.09 \times \log(\Delta) + 0.54. \quad (14)$$

A_{\max} is the maximum amplitude from the S-wave part of three-component waveforms, Δ is the epicentral distance.

3 Data

This study uses three datasets: the STanford EArthquake Dataset (STEAD, Mousavi et al. 2019), the Sichuan-Yunnan dataset, and the Italian seismic dataset (INSTANCE, Michelini et al. 2021).

STEAD (Mousavi et al. 2019) is employed as a pre-training dataset. It is a large-scale global dataset that contains a total of 1.2 million signals, comprising local earthquake waveforms from various geographical regions and tectonic settings. The dataset is divided into subsets for training (85%), validation (10%), and testing (5%).

The seismic data from Sichuan and Yunnan were obtained from the China Earthquake Networks Center (CENC) and the earthquake catalog downloaded from the National Earthquake Data Center (NEDC). The

corresponding seismic stations are equipped with three-component broadband seismometers, with a sampling frequency of 100 Hz. In total, there are 109 seismic stations, including 52 in Sichuan and 57 in Yunnan (Figure S8). The Sichuan–Yunnan Dataset contains earthquake recordings from January 2013 to May 2018 in the region. The magnitudes of events range from 0 to 1, and epicentral distances are within 110 km, meeting our requirements for studying small earthquakes at regional distances. There are 28,000 event waveforms in the Sichuan region and 9000 waveforms in the Yunnan region. For the purpose of this study, the entire dataset was partitioned as follows: a training dataset comprising 25,000 event waveforms from the Sichuan region for transfer learning; a validation dataset consisting of 1000 event waveforms from the Yunnan region; and a testing dataset containing 3000 event waveforms from the Sichuan region and 8000 event waveforms from the Yunnan region. The distribution of the epicentral distance and the backazimuth in the training and testing data is shown in Figure S4. The average SNR of the records is 10 dB, and the extracted waveform duration is 60 s, with 10 s before P-wave arrivals.

INSTANCE (Michelini et al. 2021) is a dataset of seismic waveforms and associated metadata suited for analysis based on machine learning. The dataset contains 54,008 earthquakes from 2014 to 2020, totaling 1,159,249 three-channel waveforms, covering the entire region of Italy. Since our method only considers local earthquakes, we have selected a subset from INSTANCE dataset based on an epicentral distance constraint. For this study, we use 96,177 waveforms with epicenters within 240 km from 2014 to 2017 as training data for transfer learning in the Italian region. Additionally, we select 5802 seismic events from 2018 to 2020, each with at least 4 seismic stations' recordings and epicenters less than 240 km away, as the testing dataset.

Although the network model used in this study does not consider depth during the location process, this does not represent a problem as the analyzed earthquake datasets primarily consist of shallow crustal earthquakes (depth < 30 km). Specifically, 88.75% of events in the STEAD testing dataset, 99.88% in the Sichuan testing dataset, and 91.24% in the INSTANCE testing dataset fall within this depth range. Data augmentation techniques are applied to enhance sample diversity. Random time shifting (within 30 s) is performed to obtain data with enhanced network stability. Training machine learning models with a larger and more diverse dataset can significantly enhance their generalization ability and stability. By incorporating

a wide variety of data, the model becomes better equipped to handle unseen scenarios, reducing overfitting to specific patterns in the training set. Preprocessing of the time series from the datasets involves the following steps:

- a. Demeaning, detrending, tapering;
- b. Applying 1–45 Hz band-pass filter;
- c. For each of the three components, the variance was computed, and the highest standard deviation among them was identified. All three components were then standardized using this maximum standard deviation. Following standardization, the absolute maximum amplitude across the three components was calculated and used to normalize the signals.

4 Results

4.1 The results of the DisNet network

We use the training epoch of 100 iterations with a batch size of 512 (Figure S5a). We employed an early stopping strategy with a patience of 10 epochs. After the 7th epoch, the loss did not improve further up to the 17th epoch, at which point the training was terminated. The model with the best performance, obtained at the 7th epoch, was saved. The STEAD training dataset is used to pre-train the model and subsequently evaluate the performance on its testing part. The prediction results of epicentral distance are presented in Fig. 3a, showing true versus predicted epicentral distances. It can be seen that most of the predicted results align well with the actual results. In cases where errors occur, the predicted epicentral distance is generally greater than the actual epicentral distance. The prediction error (true – predicted) distribution is shown in Fig. 3b. The errors are mainly concentrated near 0 km, indicating the reliability of our network. The mean absolute error (MAE) is 11 km, and the standard deviation is 35 km. Additionally, we calculate the relative error, representing the ratio epicentral distance errors to the epicentral distances. As illustrated in Fig. 3c, this relative error decreases as the epicentral distance increases. This indicates that the errors are relatively stable.

The pre-trained DisNet is tested on the Sichuan testing dataset, with results shown in Fig. 3d. Large errors are observed for data with epicentral distances under 20 km and beyond 60 km (Fig. 3e). Although the MAE can reach 8 km, the relative error for distances near the epicenter around 1 (Fig. 3f), indicating that the model is unsuitable for such data as the one from Sichuan. Further training, including the Sichuan data, becomes imperative. To address this, transfer training is conducted using the Sichuan training dataset from 2013 to 2017 on the

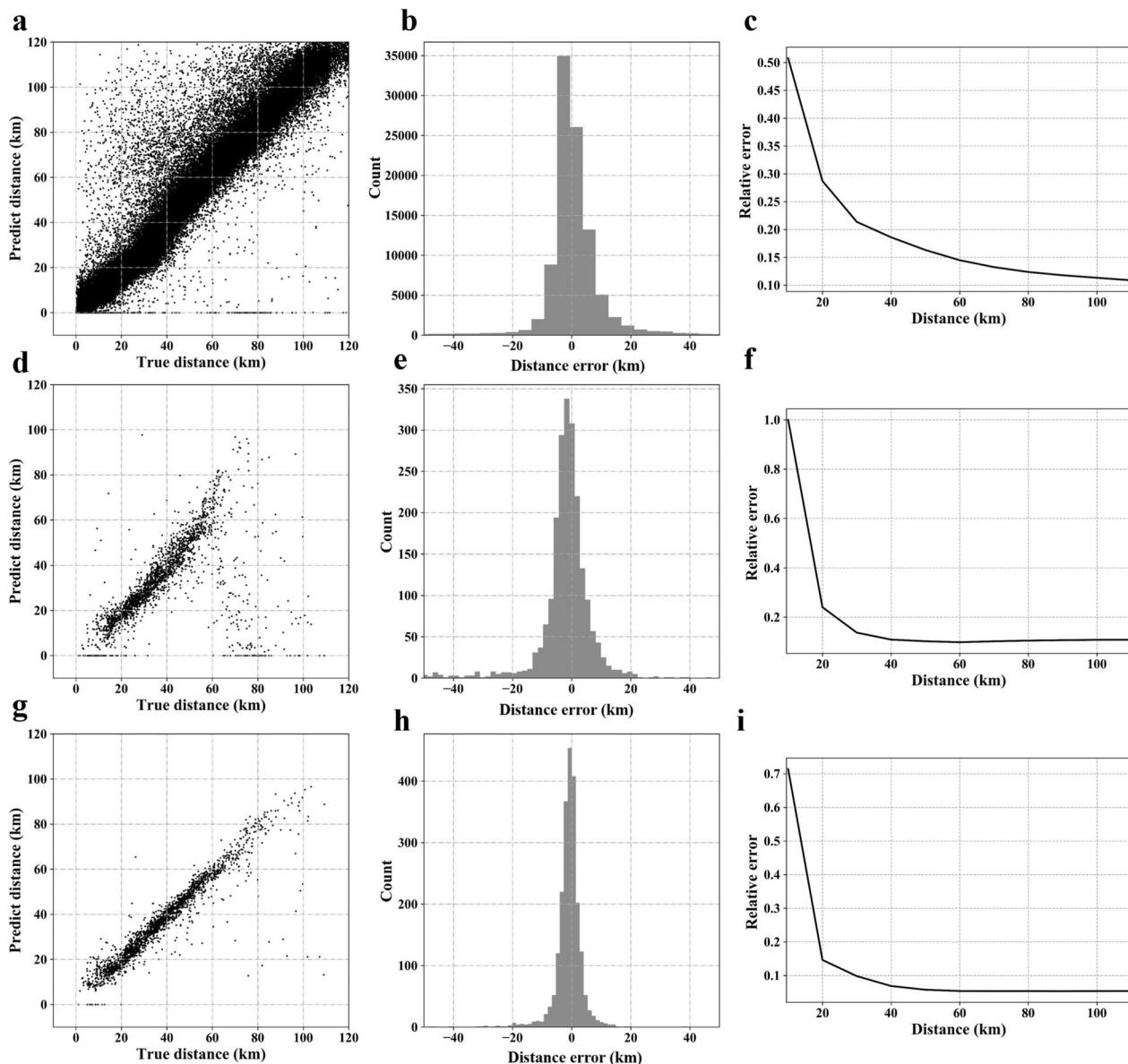


Fig. 3 The results of DisNet. **a–c** STEAD testing data; **d–f** Sichuan testing data, and **g–i** Sichuan testing data with transfer-learning model. **a, d, g** True versus predicted epicentral distances. **b, e, h** The distribution of distance errors. **c, f, i** The relative errors with distances

pre-trained model. The results of the retrained model on the Sichuan testing dataset are displayed in Fig. 3g. The results show that the DisNet after retraining can accurately predict the epicentral distance, fitting the predicted and actual values well. The MAE is estimated as 3 km, and the standard deviation is 4 km (Fig. 3h). The relative error for distances near the epicenter is 0.7 (Fig. 3i), marking an improvement over the pre-training model's performance.

In real data processing, P waves can occur anytime within the signal. To test the prediction capability of

time-shifted data, we apply DisNet to 1000 seismic data samples with varying time shifts, as shown in Fig. 4a. The results for non-time-shifted and time-shifted data are presented in Fig. 4b and c, respectively. The MAE for time-shifted data is around 4 km, slightly larger than the error (3 km) of non-time-shifted data. This suggests that the network can effectively process time-shifted data, as significant signals may emerge at any time in the data. Moreover, we further demonstrated the results of time shifts (about 30 s; Figure S6a), indicating that time shifts have minimal impact on epicentral distance prediction.

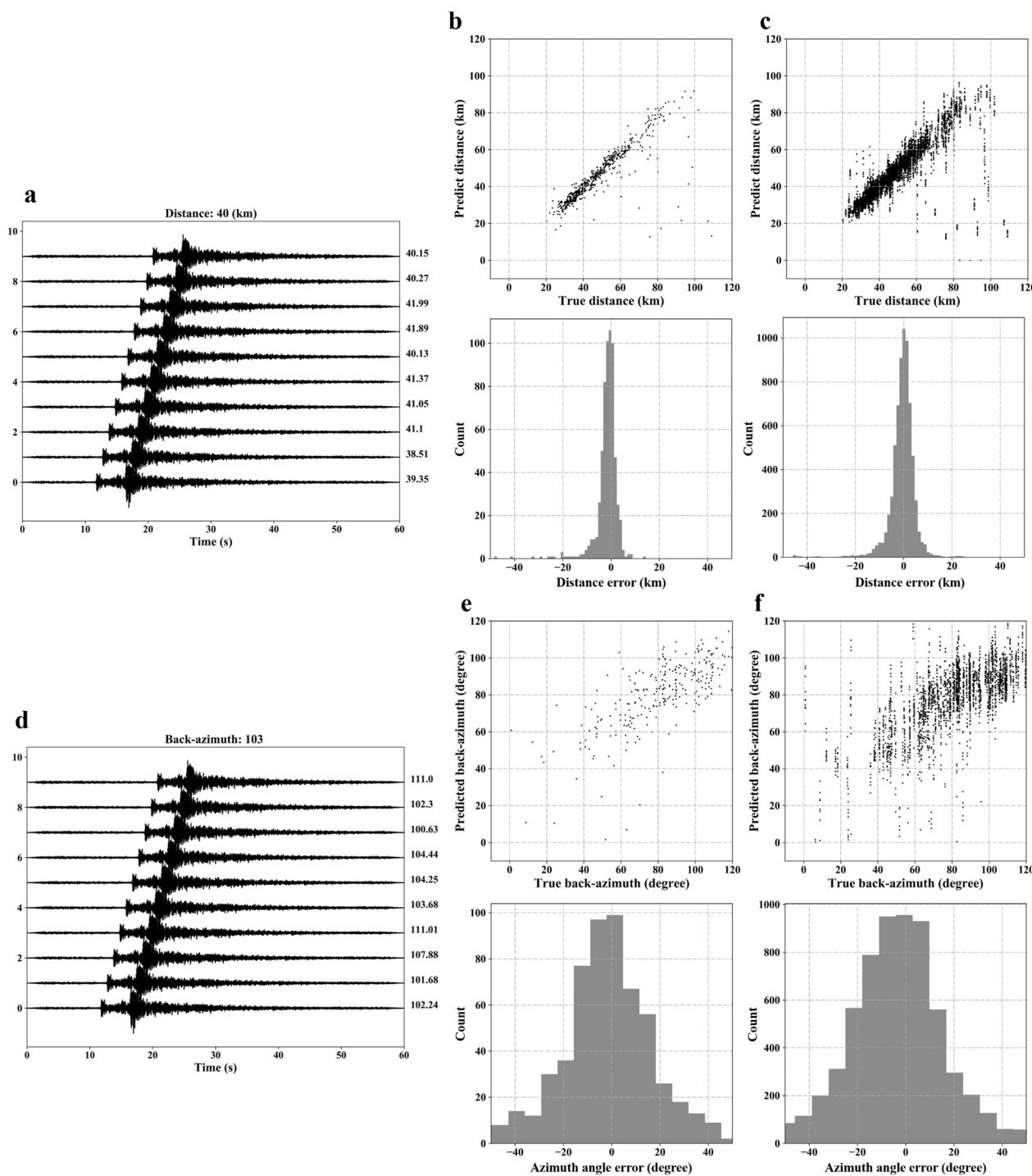


Fig. 4 The DisNet and AziNet results for time-shifted data. The different time-shifted waveforms of **a** DisNet and **d** AziNet. The predicted results of no time-shifted data of **b** DisNet and **e** AziNet. The predicted results of time-shifted data **c** DisNet and **f** AziNet

4.2 The results of AziNet network

The STEAD training dataset is further used to pre-train the AziNet network (Figure S4b), and the pre-trained network is then tested on the STEAD testing dataset. Accurate prediction of epicentral distance and back-azimuth

for small earthquakes poses inherent challenges. Despite considerable fluctuations observed in the training loss curve (Figure S5a and b), the overall downward trend suggests acceptable model convergence. When using global data to train the model, the prediction results of

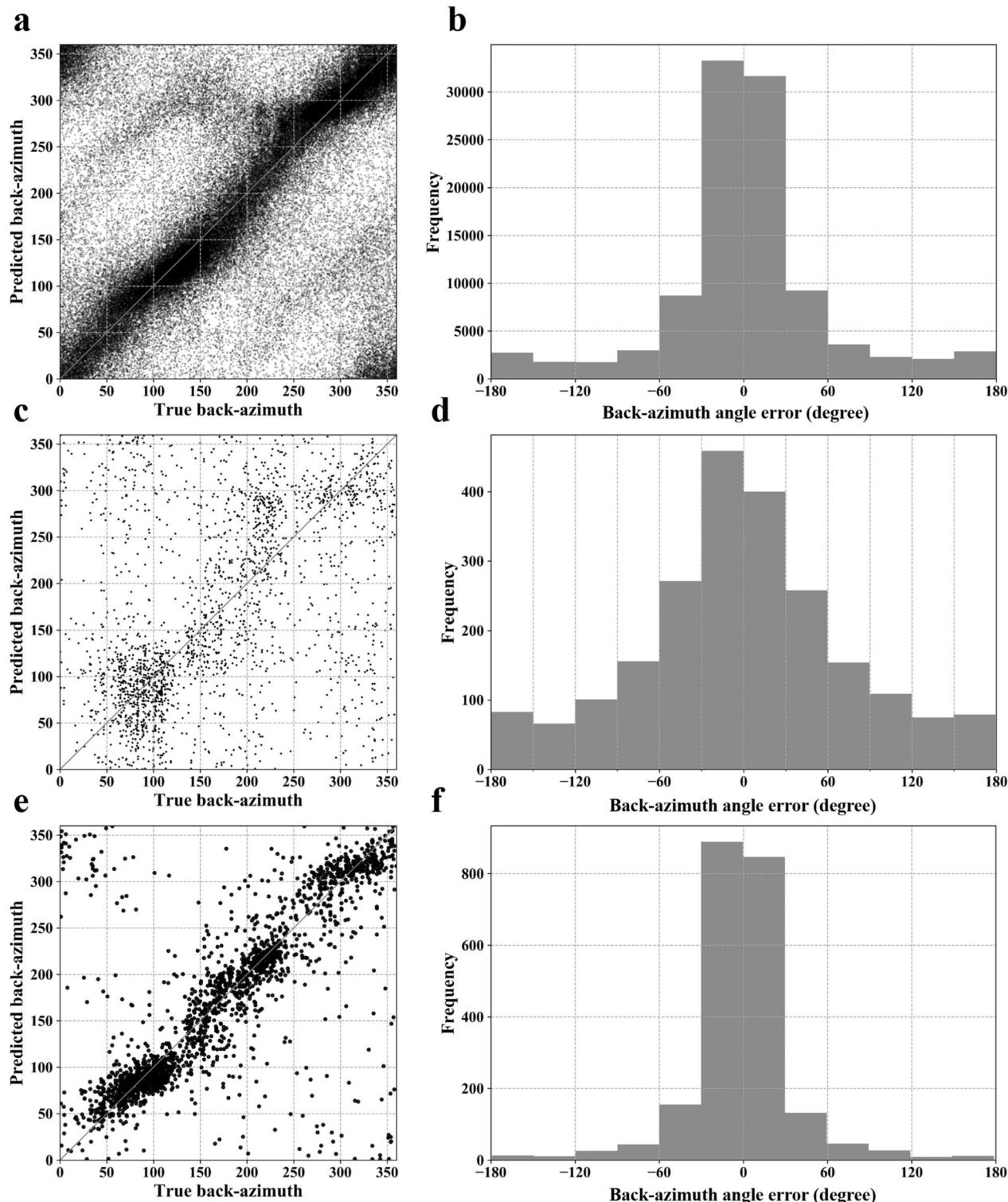


Fig. 5 The results of AziNet. **a, b** STEAD testing data; **c, d** Sichuan testing data, and **e, f** Sichuan testing data with transfer-learning model. **a, c, e** True versus predicted back azimuth. **b, d, f** The distribution of back-azimuth errors

back-azimuth are shown in Fig. 5a. It can be observed that the results of the testing dataset mostly align with the diagonal line, but scattered points are distributed around it, indicating many data points with large errors in azimuth prediction. Predicting azimuth accurately is challenging due to the complexity of waveform data. The distribution of prediction errors (true – predicted) is shown in Fig. 5b. The mean absolute error (MAE) is 38 degrees, and the standard deviation is 44 degrees.

The pre-trained AziNet is then tested on the Sichuan testing dataset, with results displayed in Fig. 5c. The back-azimuth data exhibit large errors, with an MAE reaching 57 degrees and a standard deviation of 47 degrees (Fig. 5d), indicating that the network is unsuitable for the Sichuan data or that the data in Sichuan are more complex. To address this, transfer training is conducted using the Sichuan training dataset from 2013 to 2017 on the pre-trained model. The results of the retrained model on the Sichuan testing dataset are shown in Fig. 5e. The predicted and actual values align well, with an MAE of 22 degrees and a standard deviation of 28 degrees.

In addition, 1000 seismic data samples are time-shifted, and the AziNet is used to test the prediction capability of time-shifted data, as shown in Fig. 4d. The title of the figure indicates the true azimuth, while the right side displays the azimuth predicted by the network for different time shifts. The predicted results of non-time-shifted and time-shifted data are shown in Fig. 4e and f, respectively. The MAE for time-shifted data is 29 degrees, slightly larger than the error of non-time-shifted data (27 degrees). This indicates that the network can effectively handle azimuth predictions for data with smaller time shifts (10 s). When time shifts (about 30 s) are applied, the error in azimuth prediction increases as the time shift grows (Figure S6b). Azimuth prediction is inherently challenging, and time shifts may cause the loss of some effective waveforms, further amplifying the error.

4.3 The results of the single-station location for Yunnan dataset

We apply the trained DisNet and AziNet models to the testing datasets in the Yunnan area, and the results are presented in Fig. 6. In this process, we assume that all events can only be recorded by a single station. All stations with detectable signals are evaluated independently under this assumption, allowing us to assess the network's performance at the single-station level. The prediction results for epicentral distances by DisNet are displayed in Fig. 6a. It can be observed that the distance prediction results are poor for epicentral distances smaller than 20 km, with the predicted distance generally greater than the actual distance. Conversely, for predicted epicenter distances greater than 20 km, the predicted distance is

generally smaller than the actual distance. This indicates the necessity of transfer learning and training for new areas. The error analysis of epicentral distance is shown in Fig. 6b, with the error concentrated around –4.0 km. The mean absolute error (MAE) is 5.0 km, and the standard deviation is 7.0 km. Figure 6c illustrates the variance of relative error as a function of epicentral distance. As the epicentral distance increases, the relative error gradually decreases and stabilizes. When the epicentral distance is less than 20 km, the relative error is greater than 25%. When the epicentral distance is greater than 20 km, the relative error is around 20%. The prediction results for back-azimuth are presented in Fig. 6d, showing a poor fitting effect. There are two possible reasons for this. First, it could be due to small earthquakes' low SNRs. Second, fitting the nonlinear relationship between back-azimuth and seismic data could be challenging. The estimated MAE of the back-azimuth is 55 degrees, and the standard deviation is 46 degrees (Fig. 6e). The location error using the deep learning method with single station data is approximately 23 km (Fig. 6f).

The influence of back-azimuths on the final location results varies depending on the epicentral distances of the events. While they have minimal impact on events with small epicentral distances, their significance becomes pronounced for events with larger epicentral distances. Figure 6g and h displays the earthquake locations in the earthquake catalog and the locations calculated using the network models, respectively. A comparative analysis reveals that the spatial distribution patterns of locations derived from machine learning with single station data closely match those in the earthquake catalog. The differences in details are mainly attributed to the large error in back-azimuth. Earthquakes are distributed in circular patterns in Fig. 6h. Figure 6i illustrates the earthquake locations in the catalog relative to the locations calculated using neural networks. Overall, the testing in the Yunnan region demonstrates that our method can be used for location analysis with single-station data.

4.4 The results for the INSTANCE dataset

Similar to the procedure in the Sichuan-Yunnan region, we conduct transfer learning on the pre-trained model and apply it to the Italian region dataset—INSTANCE. Training data from 2014 to 2017 in Italy are utilized for transfer training on both DisNet and AziNet models. Subsequently, these trained models are employed on testing data spanning from 2018 to 2020 in Italy. Figure 7a and b presents the predicted epicenter distance and back azimuth results, respectively. Error analysis of epicentral distance, shown in Fig. 7c, highlights a concentration of errors around –3.5 km. The mean absolute error (MAE) is reported as 6.4 km, with a standard

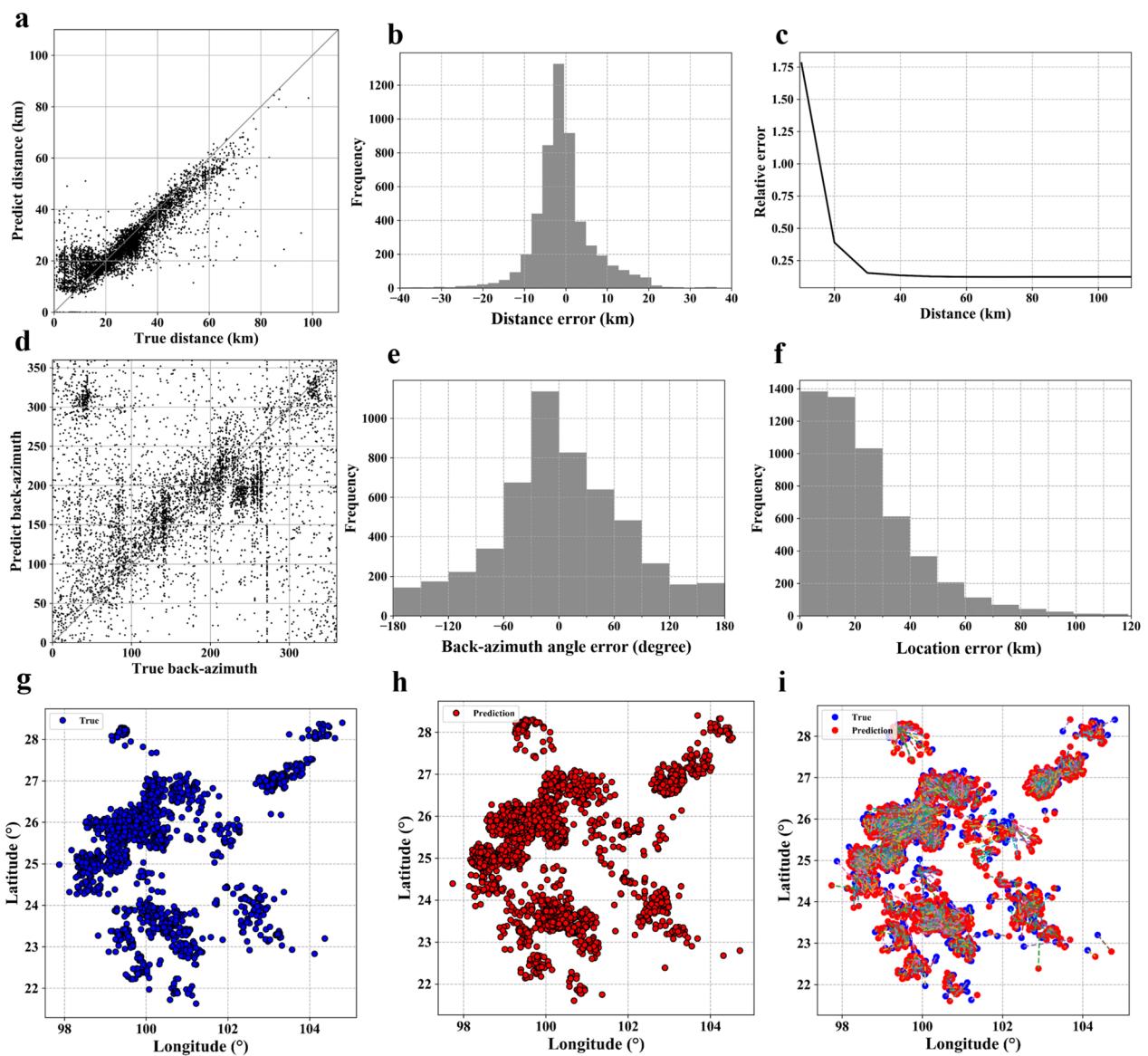


Fig. 6 The results for the Yunnan area. **a** The predicted distance results. **b** The distribution of distance errors. **c** The relative errors with distances. **d** The back-azimuth predicted results. **e** The distribution of back-azimuth errors. **f** The location errors with distances and back-azimuths. **g** The event locations provided by the catalog. **h** The location predicted by DisNet and AziNet. **i** The relative location relationship between provided and predicted events

deviation of 14.4 km. For back-azimuth, the MAE is 36 degrees, with a standard deviation of 55 degrees (Fig. 7d). In our location workflow, we randomly select one station for single-station location and two or three stations for multi-station locations. As there is no association and clustering process for multi-station locations since the same event's waveform is selected, Fig. 7e shows event location (black), single-station location only using one station of one selected station per event (green), and multiple-station location (red). The mean location errors for multi-station and single-station locations are 14 km

and 18 km, respectively. Notably, the effectiveness of multi-station locations surpasses that of single-station locations, as evident from the distribution of locations. Regardless of whether it is a single station or multiple stations, the predicted earthquake distribution closely aligns with the ground truth.

5 Application

We test the proposed location workflow in the Sichuan region, using 2756 seismic events with magnitudes ranging from 0 to 1 recorded from January to May

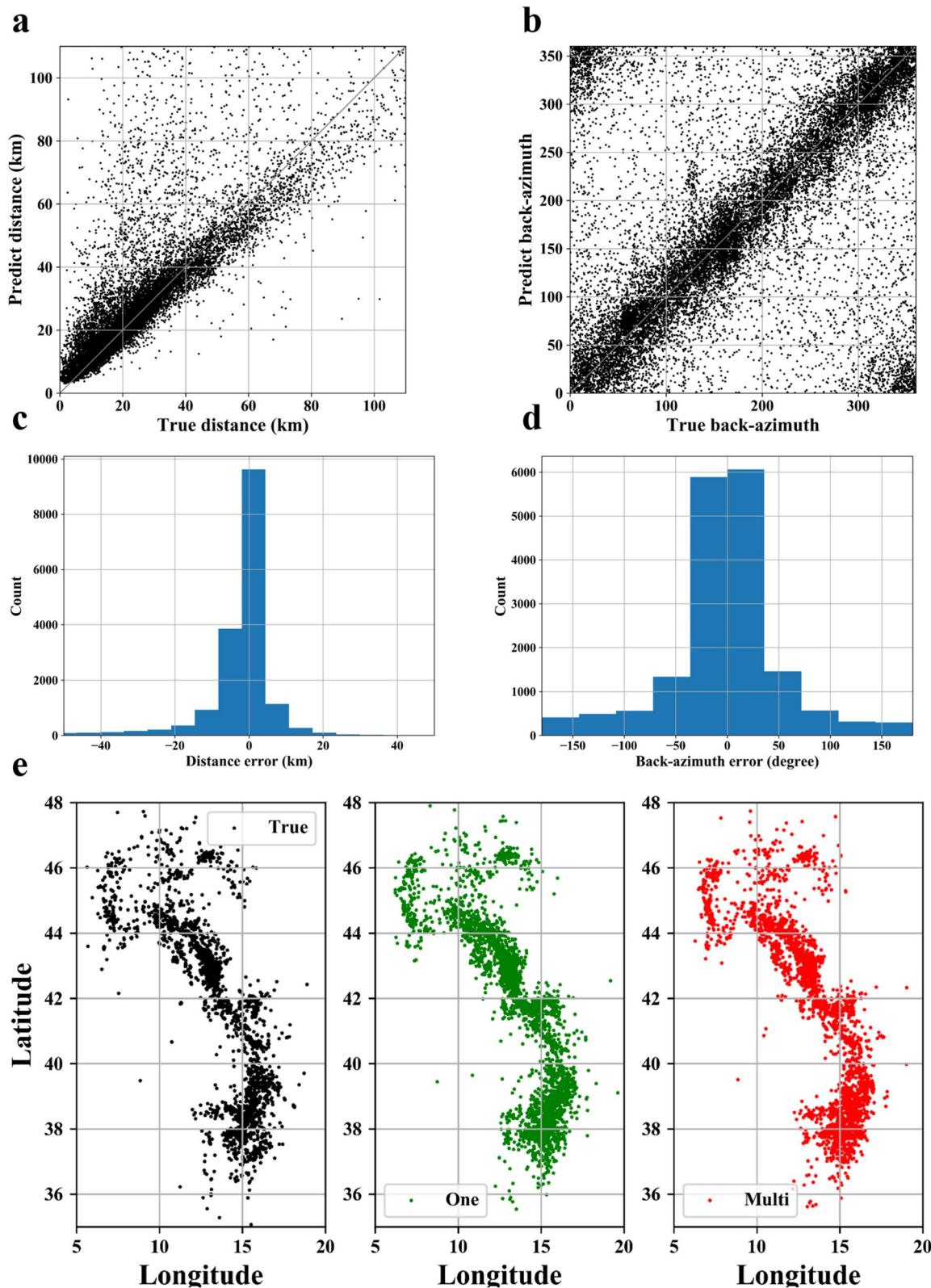


Fig. 7 The results for the Italian area. **a** The predicted distance results. **b** The back-azimuth predicted results. **c** The distribution of distance errors of single station. **d** The distribution of back-azimuth errors of single station. **e** The locations of ground truth (black), single-station location (green), and multi-station location (red)

2018. For each event, 60-s waveforms with a 0.01-s sampling interval are extracted from the continuous seismic records at all stations, starting from the event origin time (t_0) to $t_0 + 60$ s. Among the total events, our approach successfully detects and locates 2170 events, while 502 events remain undetected, and 84 events could not be processed due to insufficient data. The mean absolute error in location for the detected events is 18 km (Figure S7a). In addition, we detect and locate 5378 seismic events that are not present in the earthquake catalog, which is about 1.4 times the number of events in the catalog. We manually inspected 100 events and found a false positive rate of 14%. The mean

absolute error between the predicted magnitudes and the magnitudes provided by the earthquake catalog is 0.25, with a variance of 0.04 (Figure S7b). Generally, the predicted magnitudes are slightly smaller than those in the Sichuan catalog, indicating a certain degree of underestimation. In Fig. 8a and b, we show the earthquake distribution according to both the earthquake catalog and the results of our location workflow, respectively. The predicted earthquake locations match well with the actual spatial distribution. However, there are discrepancies in specific details, notably in the distribution of seismic events represented by the black circles. The seismic event distribution provided by the earthquake catalog follows a northwest-to-southeast pattern (Fig. 8a), whereas the occurrences identified using our

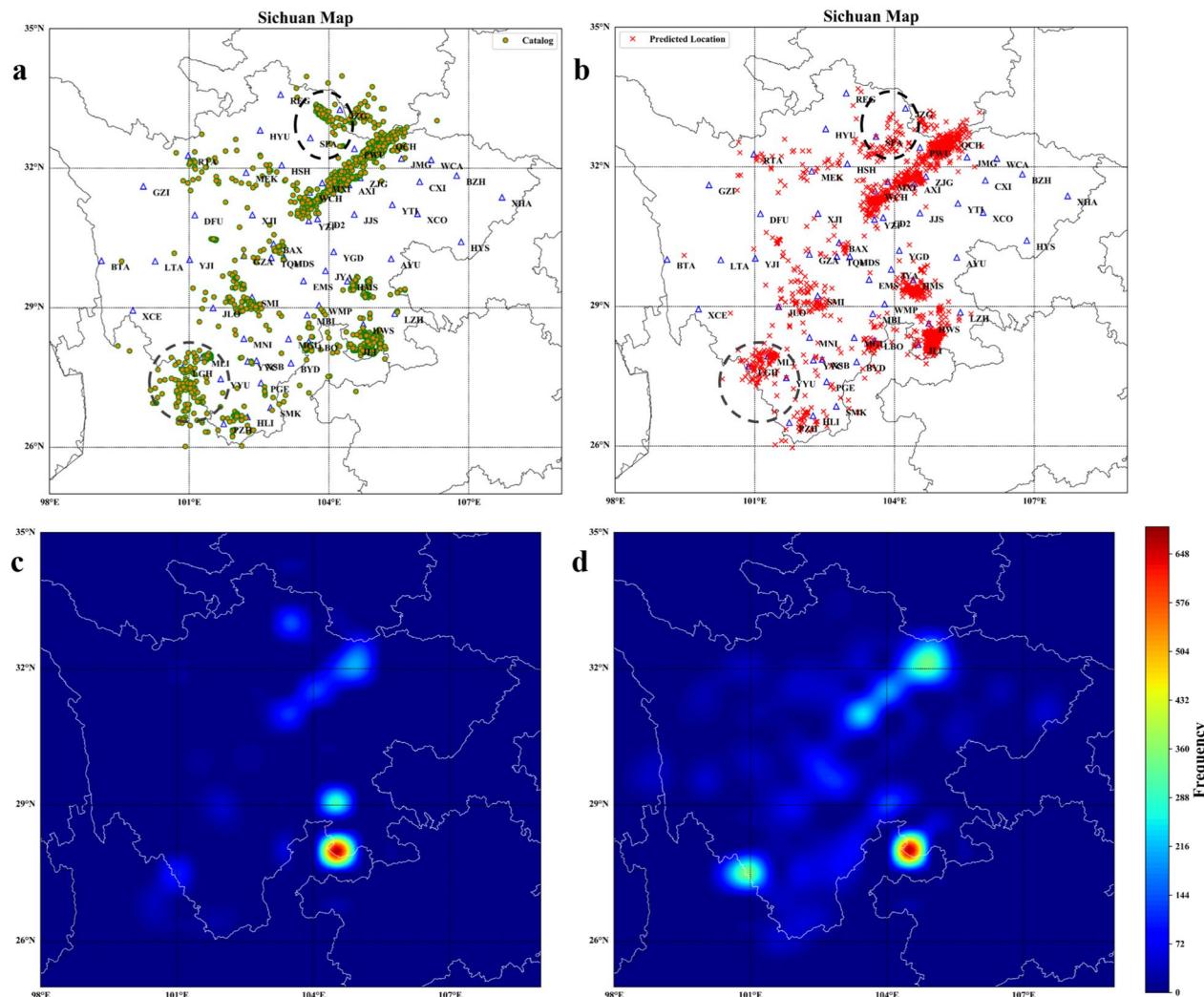


Fig. 8 The results for the Sichuan area. **a** The event locations provided by the catalog from NEDC. **b** The location predicted by location workflow. **c** The frequency of seismic activities provided by the catalog. **d** The frequency of seismic activities by location workflow

single-station location method show limited activity in the northwest-to-southeast direction. Instead, they exhibit a prevailing distribution trend from northeast to southwest (Fig. 8b). For the earthquake distribution represented by the gray circle, while the earthquake catalog shows a relatively scattered distribution, our method reveals a more concentrated distribution of earthquakes. Furthermore, Fig. 8c and d illustrates the occurrence frequency of events in the catalog and those obtained through our method, respectively. Brighter colors indicate a higher frequency of occurrences in the area. The results show the potential of our method in enhancing the earthquake catalog by providing more comprehensive coverage of seismic events.

6 Discussion

There are numerous hyperparameters to consider when designing a neural network, including the network depth, the size (e.g., 3, 5, 7) and number of convolutional kernels, the choice of activation function, and the placement of batch normalization layers. In this study, we focused on evaluating the impact of different activation functions (e.g., tanh, ReLU) and the position of batch normalization layers. Experimental results show that the combination of the tanh activation function followed by batch normalization yields the lowest loss in testing dataset and overall best model performance, and the best model architecture is as shown in Fig. 2. Of course, our tests may not be exhaustive, and it is

possible that others could further improve the model by adjusting parameters based on our framework.

For neural networks, the general rule is that better data lead to better predictions. The data quality is closely related to the effectiveness of the predictions. We analyze epicenter distance and azimuth prediction performance under different SNRs, as shown in Fig. 9. Since the SNR of the data is primarily concentrated in the range of 5–15 dB, and there is less data available in other ranges, there may be statistical errors. Therefore, our primary focus is the 5–15 dB data. Analyzing the error variation within this range reveals that a gradual decrease in epicenter distance error as SNR increases (Fig. 9a). Similarly, back azimuth error also shows a gradual decrease (Fig. 9b). These findings indicate that higher SNR values result in better prediction performance. It is indeed a challenging task to locate small events with low SNR.

We further discuss the errors of single-station location and highlight the advantages of the proposed workflow. By leveraging predictions of epicentral distance and azimuth angle, we can determine earthquake locations within the seismic network, incorporating relevant station constraint conditions. When only one station receives the seismic signal, it is equivalent to performing a single-station location. Figure 10 illustrates a single-station location in the Sichuan station network.

Figure 10a demonstrates an accurate prediction of epicentral distance and azimuth angle, resulting in a location error of approximately 4 km. Although the first

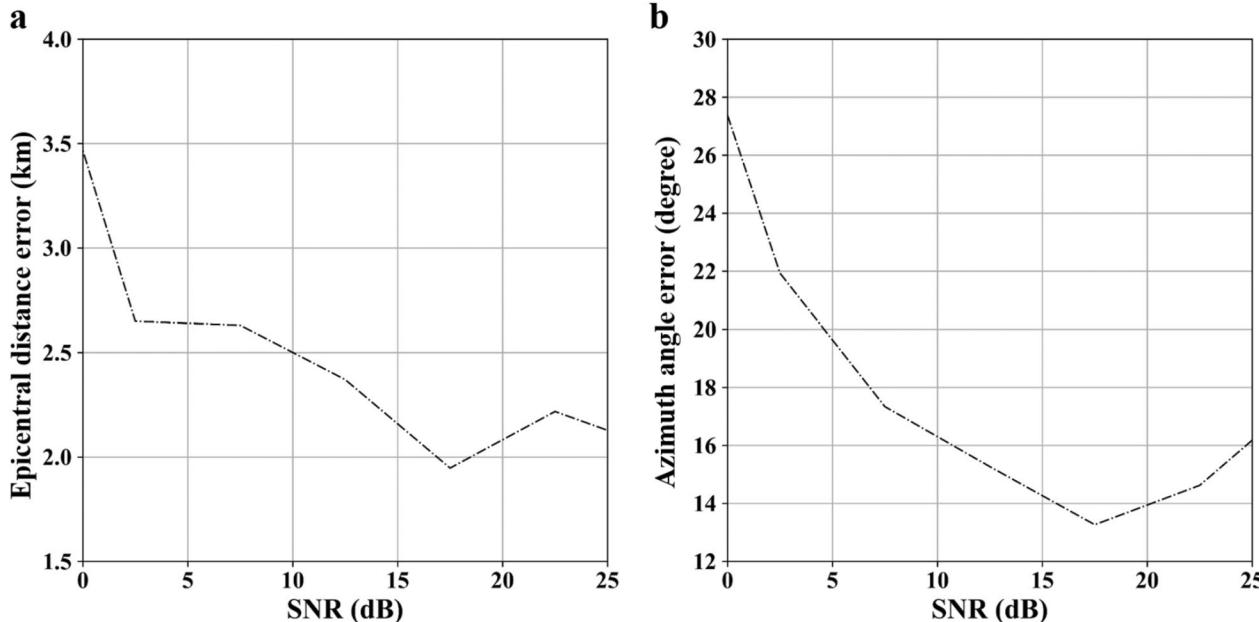


Fig. 9 The relationship predicted results with SNR. **a** Epicentral distance with SNR. **b** Azimuth angle with SNR

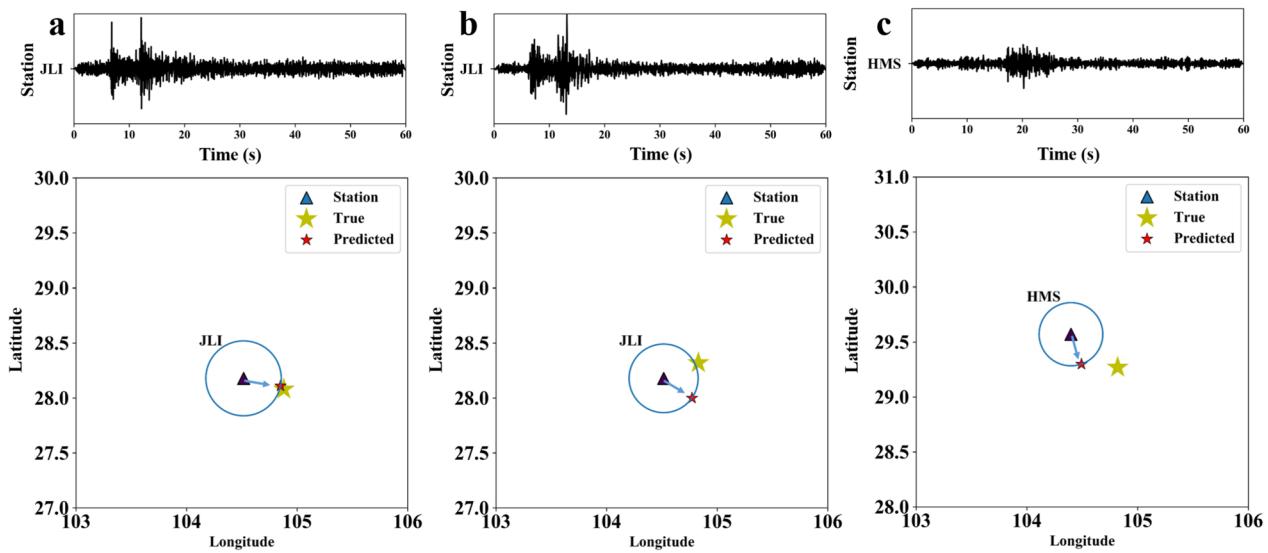


Fig. 10 Example of the single station location. Three event waveforms and single-station locations. The yellow star represents the actual earthquake location, and the red star represents the predicted earthquake location. The arrow shows the direction of azimuth angles. **a** An example of distance and azimuth predicted well (location error: 4 km). **b** An example of distance predicted well, but azimuth predicted not well (location error: 36 km). **c** An example of distance and azimuth predicted not well (location error: 36 km)

arrival quality of the corresponding waveform is not high, the P and S waves are discernible. In Fig. 10b, the epicentral distance is predicted accurately, but the azimuth angle is not, leading to a location error of about 36 km. The P and S waves of the corresponding waveform can still provide a good prediction of the epicentral distance. Figure 10c shows a scenario where the prediction of epicentral distance and azimuth angle are inaccurate, resulting in a location error of approximately 36 km. In this case, the P wave of the corresponding waveform is not evident. Although the S-wave is clear, the predicted distance is too large, likely due to an underestimated P-wave arrival time. The traditional method of calculating the azimuth is to determine the azimuth angle according to the polarization direction of the P wave by using the initial motion of the three-component P wave. This method is generally suitable for large earthquakes or seismic waveforms with high SNR. However, it is no longer applicable for low SNR signals. Instead of separately extracting the waveform near the P wave for analysis and prediction, our approach focuses on the information in the entire waveform, aiming to capture the relevant azimuth angle. Although azimuth errors may be relatively large, it is inevitable given the circumstances.

The analysis of results reveals a notable discrepancy: the error in predicting earthquake epicenter distance is considerably smaller than that in predicting the azimuth angle. Many inaccurate location situations arise due to azimuth errors (Fig. 10). Evidently, the epicenter distance in these examples is relatively accurate, deviations in

azimuth angles lead to significant location errors. Moreover, it is worth noting that azimuth errors have a greater impact on larger epicenter distances. The test results above demonstrate that the prediction accuracy of the seismic distance is superior to that of the azimuth angle. By leveraging the station without relying on the azimuth data, we can enhance the accuracy of the earthquake location.

Next, we highlight the advantages of our location workflow through examples of separate and joint locations using multiple stations. In Fig. 11a, the waveforms recorded by the detected stations are depicted, while Fig. 11b illustrates the clustering of these stations and their corresponding location results. Stations sharing the same color belong to the same category. In this instance, PWU, QCH, and JMG are clustered together, while the YJI station forms a separate category. Two distinct cluster groups indicate the occurrence of two seismic events during this period. To underscore the benefits of network constraints, we separately calculate the location errors of PWU, QCH, and JMG stations and collectively analyze their location errors. When comparing the joint location results with the earthquake catalog, the location error is 1 km. Conversely, the single-station location errors for these three stations are 17 km, 20 km, and 49 km, respectively (Fig. 11c). Despite the large errors observed in single-station locations based on predicted epicenter distance and azimuth angle, combining the stations and solely utilizing the epicenter distance for location enables us to reduce the location error significantly. This

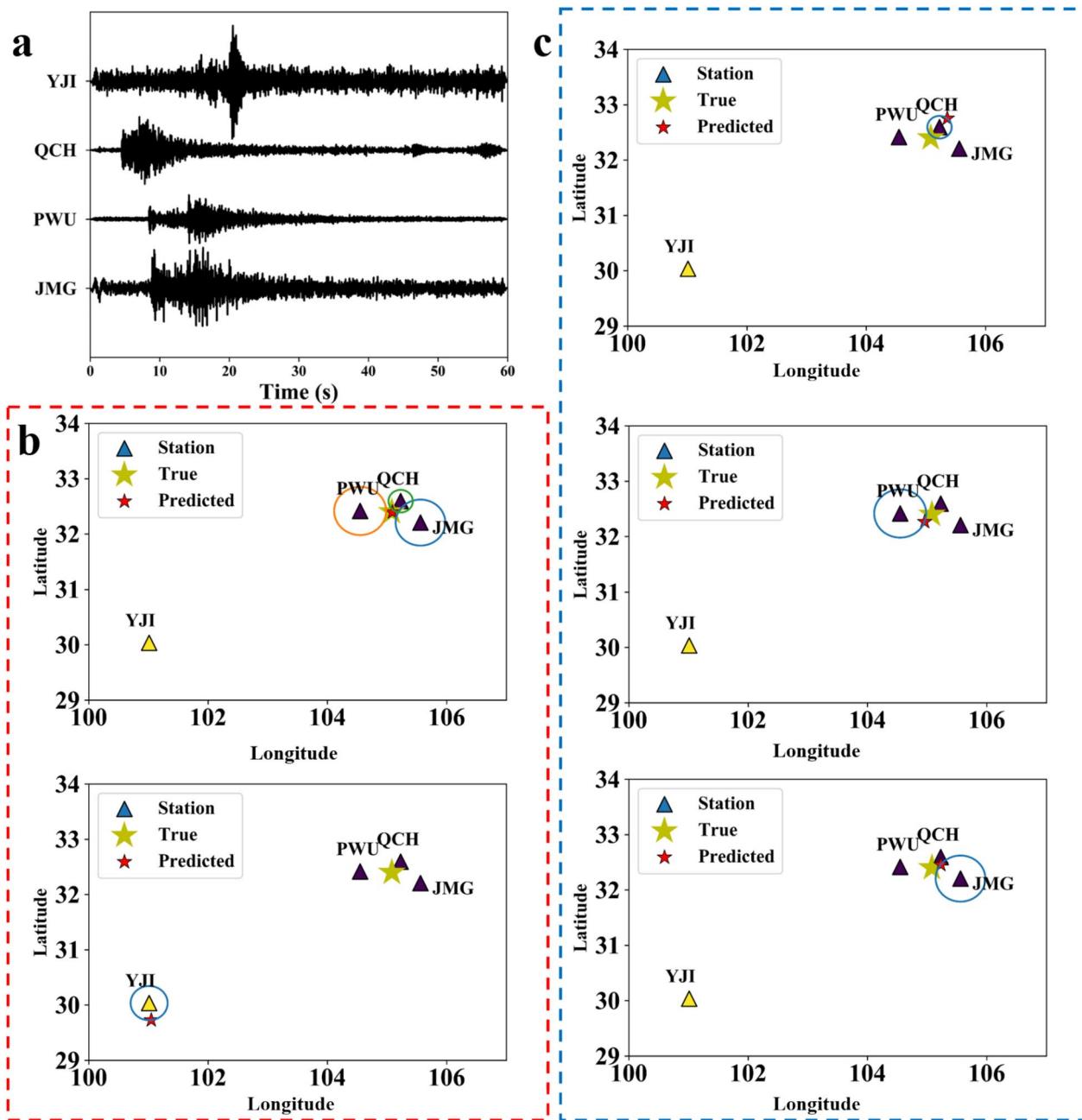


Fig. 11 Example of the multi-station location. **a** The detected waveforms. **b** Location of different clusters. Different triangle colors represent different clusters. **c** Single-station location of cluster stations (QCH, PWU, JMG). The triangle represents the seismic station, the yellow star represents the actual earthquake location, and the red star represents the predicted earthquake location in **b** and **c**

demonstrates that the accuracy of locating small earthquakes can be improved by incorporating more precise epicentral distances and station locations.

The example of when only two stations record the same seismic event, with the station distribution and the corresponding recorded waveform is shown in Fig. 12a and b. In this scenario, with two associated stations, we can

calculate one intersection point using the least square method (Fig. 12b left) and determine the other focal point through geometric relations (Fig. 12b right). To demonstrate the location results for the two stations, WCH and MXI, we employ the single-station location method, as shown in Fig. 12c. The results indicate that their locations are proximate to one of the two sources

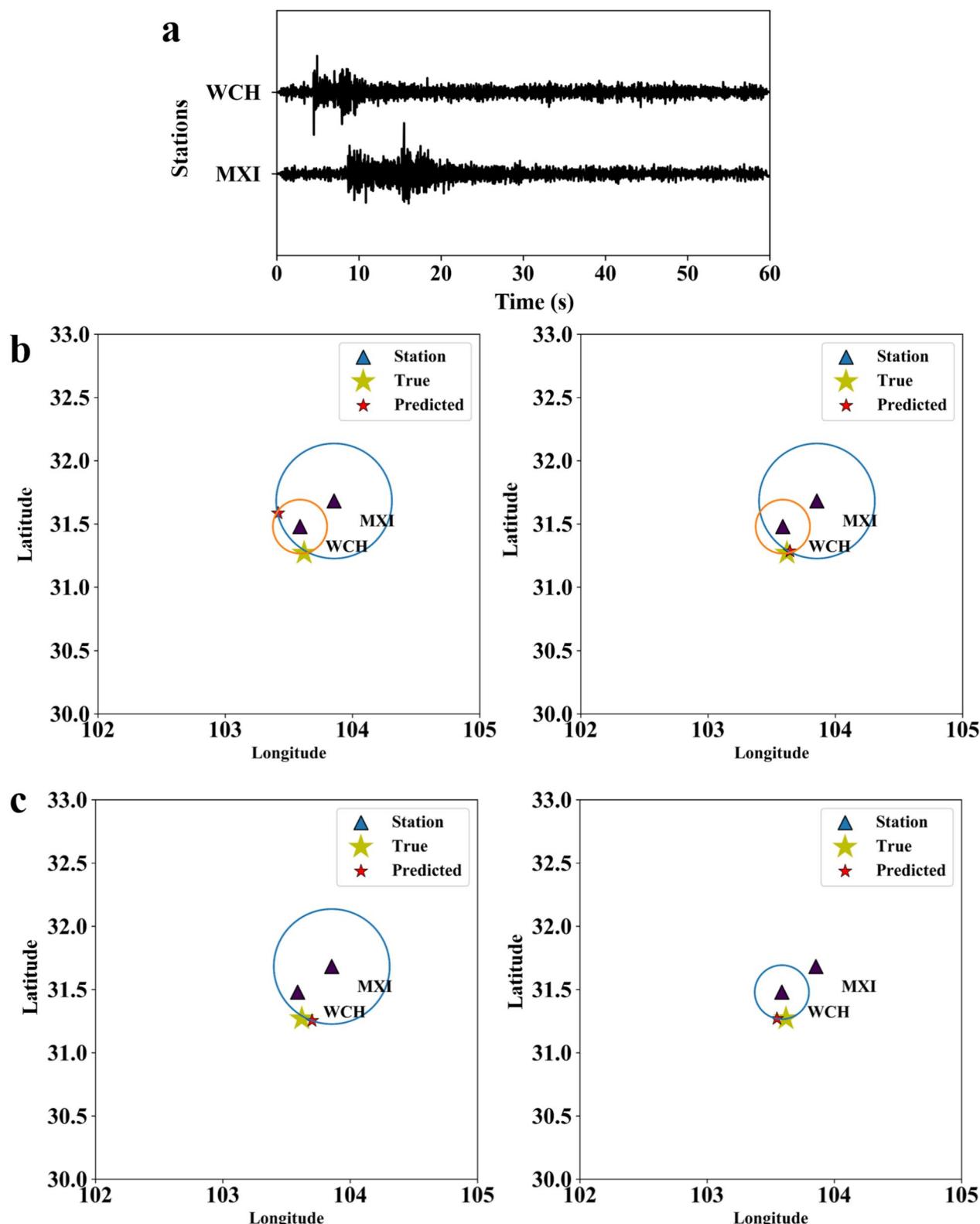


Fig. 12 Example of the two-station location. **a** The detected waveforms. **b** Two-station location. LSM (Least Squares Method) can yield a single solution (left), which is then utilized to derive the final location by employing geometry and azimuth calculations (right). **c** Single-station location of cluster stations. The triangle represents the seismic station, the yellow star represents the actual earthquake location, and the red star represents the predicted earthquake location

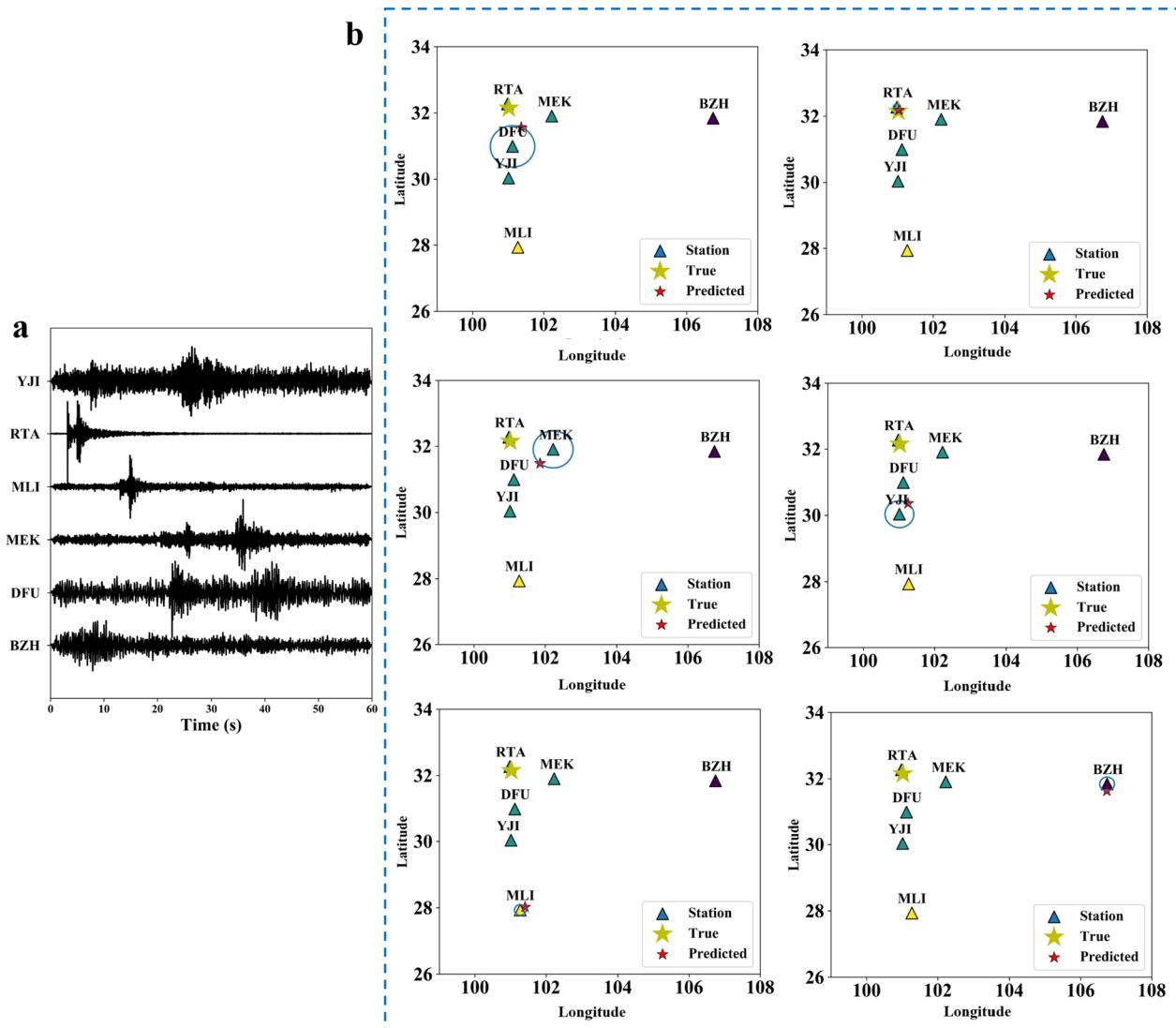


Fig. 13 Example of the problematic case for the multi-station location. **a** The detected waveform records. **b** Location result of each triggered station. If multiple waveform records of an event are not correlated or associated, they are regarded as separate or distinct events (RTA, DFU, and MEK). The triangle represents the seismic station, the yellow star represents the actual earthquake location, and the red star represents the predicted earthquake location

(Fig. 12b right), contributing to the final location result. The red star is the final location using the LSM and the azimuth information in Fig. 12b right. Specifically, the single-station location errors are 9 km and 8 km, respectively, while the joint location error is reduced to 3 km.

The complexity of seismic events' generation sometimes leads to the situation when multiple stations are recording seismograms during the same time-interval, as illustrated in Fig. 13. However, those may not originate from the same event. The recordings exhibit differences in waveform, spatial location, and onset time. For instance, the RTA and MLI stations display distinct

seismic wave shapes indicative of nearby earthquakes. Although both stations are triggered, they are spatially distant and thus cannot be attributed to the same seismic event. Similarly, the BZH station, significantly distant from the others, recorded a separate seismic event. The low SNR observed at the YJI station may be the indication of the coda from the previous seismic event. In this scenario, we consider the waveform records of the DFU and MEK stations as separate seismic events, although they are more likely associated with the seismic events near the RTA station. However, due to substantial spatial distance, long time intervals exceeding our time difference

threshold, and very low SNRs, we consider them distinct events. Due to the spatial, temporal, and waveform constraints imposed by our clustering method, an inherent limitation of the workflow is that a single seismic event may occasionally be split into multiple clusters. Therefore, this workflow is primarily suitable for identifying small earthquakes with low SNRs or sparse station coverage. Large earthquakes with strong energy sustained, long durations, and wide influence ranges may violate the assumptions of our method and thus render it less effective. Such large events can be reliably handled using a conventional travel-time-based location method, which also helps mitigate the risk of a single big event being mistakenly split into multiple clusters during the association process. In scenarios where multiple earthquakes occur nearly simultaneously at proximate locations, particularly under sparse station coverage, distinguishing individual events becomes inherently challenging, even in the absence of clustering algorithms. To mitigate the risk of erroneously merging adjacent but distinct events, we carefully determined the spatial threshold for clustering. Based on the average inter-station spacing (~ 60 km), a clustering radius of 110 km was selected. This threshold offers a reasonable trade-off: it is sufficiently broad to encompass most stations likely triggered by the same event, while reducing the likelihood of including distant, unrelated events within the same cluster. Nonetheless, this approach may still result in the merging of closely spaced events, which remains a potential limitation to be addressed in future work. Despite these limitations, for small earthquakes or seismic events with insufficient records, when no other information is available, our workflow can provide reasonable location estimations.

The main disadvantages of the workflow are summarized below:

- When earthquakes occur in similar areas at similar times, distinguishing them in time or space becomes challenging. This can result in grouping multiple earthquakes into a single event within a similar area.
- Using a fixed time window of 60 s may lead to inaccuracies, especially when multiple seismic events occur simultaneously. To address this, we use a 20-s window for continuous data processing, ensuring that small seismic events are not overlooked. For local events (< 120 km), 20 s is sufficient to capture the effective P and S phases. Based on the previous time shift analysis, we know that the azimuth prediction performs better with smaller time shifts. During the processing of continuous data, each time an event is detected, we record the location of the maximum value and extract 30 s of waveform data before and

after this point as the network's input. This approach helps to maximize the prediction accuracy.

- This method is unsuitable for large earthquakes (3 M or above), particularly those with long durations. Such events may be fragmented into multiple seismic events during spatial clustering, potentially complicating analysis.
- Our approach primarily focuses on the distribution of earthquakes in two-dimensional space, overlooking depth influence. We primarily discuss shallow-source seismic events within 20 km.

Throughout the entire process of single-station location or the location workflow systems, determining the back-azimuth is challenging and prone to significant errors. Compared to the traditional polarization direction method, the back-azimuth prediction network does not require picking up the initial motion phase. Instead, it utilizes the entire waveform to predict the back-azimuth. We evaluate the azimuth using the traditional polarization method and AziNet on the 2018 Sichuan testing dataset. For small earthquake events with low SNR in the Sichuan region, the average absolute error of the back-azimuth angle using the traditional method is 77.5 degrees, while the average absolute error predicted by AziNet is 22.7 degrees. These results demonstrate that our method outperforms the traditional method based on the initial dynamic polarization direction despite exhibiting a larger error in back-azimuth prediction compared to the true value. The traditional method has stringent requirements on data SNR and is often ineffective for small seismic events with low SNR. In contrast, our network model can compensate for this limitation and provide a reliable back-azimuth evaluation.

When the seismic station network is not dense, traditional methods are often inadequate for locating detected waveforms. In such scenarios, we employ machine learning to predict the epicentral distance and azimuth using single-station waveforms. By incorporating the station locations and seismic waveforms, we apply spatial clustering constraints and temporal and waveform constraints to enhance the location capabilities, especially for networks with limited station coverage. An advantage of this approach is the elimination of the picking process, which proves particularly beneficial for locating low SNR events. This method enables us to provide event locations with insufficient stations quickly. In addition, when the station network is dense, traditional methods are initially employed for association and location. Subsequently, any remaining records that are not associated can still be processed using the single-station location method to further refine the earthquake catalog.

7 Conclusions

We propose a seismic event location workflow for seismic networks specifically designed for small earthquakes. This method uses machine learning to predict epicentral distance and azimuth with single station data. Training the network with global data and fine-tuning it with local seismic records through transfer learning can significantly enhance prediction accuracy. Notably, the prediction of epicentral distance demonstrates relatively high accuracy, as indicated by promising results obtained from both global and Sichuan datasets. For Sichuan, the absolute mean error is reported at 3 km, with a variance of 4 km. However, predicting the azimuth angle proves more challenging, with an absolute mean error of 22 degrees for Sichuan and a variance of 28 degrees. Despite this, it still serves as a valuable reference range in the absence of additional information. In the Yunnan region, conducting single-station location tests and evaluating epicentral distances and azimuth angles allow us to capture and reflect the distribution of seismic events.

By integrating neural network predictions of epicentral distances and azimuth angles from one station with location and waveform data from corresponding stations, we can correlate different seismic records to determine seismic events and conduct location analysis. We adopt different location strategies based on the number of stations recording seismic events. When more than two stations record seismic events, we rely solely on epicentral distances for location analysis. Conversely, when fewer than three stations record seismic events, we incorporate both epicentral distances and azimuth angles in the location analysis. The location workflow is also applied to the Italian region, providing the mean location errors for multi-station and single-station locations are 14 km and 18 km, respectively. Remarkably, whether utilizing data from a single station or multiple stations, the predicted earthquake distribution closely mirrors the ground truth. Applying this location workflow to the Sichuan region allows us to identify 78.7% of recorded seismic events and discover numerous small ones that previously went unnoticed. This established process for locating small earthquakes is expected to enhance the earthquake catalog, fill in missing seismic events, and lower the limit of magnitude of completeness, thereby enriching our understanding and analysis of the study area. Looking ahead, this workflow holds potential for application in detecting lunar quakes and other planetary vibrations.

Abbreviations

SNR	Signal-to-noise ratio
EQT	EarthQuake Transformer
CNNs	Convolutional neural networks
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DisNet	Distance Prediction Network
Res-block	Residual blocks

LSTM-block	LSTM blocks
FC	Fully connected layer
GRU	Gated recurrent unit
AziNet	Azimuth Prediction Network
C-block	Convolution block
MSE	Mean square error
STEAD	STanford EArthquake Dataset
INSTANCE	Italian Seismic Dataset
MAE	Mean absolute error
CENC	China Earthquake Networks Center
NEDC	The National Earthquake Data Center

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40623-025-02258-x>.

Supplementary Material 1.

Acknowledgements

We thank the open datasets from the STEAD data (<https://github.com/smousavi05/STEAD>) and INSTANCE data (<https://github.com/INGV/instance>). We also thank the China Seismic Experimental Site, China Earthquake Networks Center (<https://news.ceic.ac.cn/>), and the National Earthquake Data Center (https://data.earthquake.cn/dashshare/report.shtml?PAGEID=earthquake_zhengshi) for providing the Sichuan and Yunnan datasets and corresponding catalogs.

Author contributions

Ji Zhang analyzed the data and designed the research. Aitaro Kato, Huiyu Zhu, and Jie Zhang acquired the data and contributed to the discussions. All authors read and approved of the final manuscript.

Funding

The authors thank the National Key R&D Program of China (Grant Number 2021YFC3000701) for the financial support. This research is also supported by the "Establishment of a Research and Education Complex for Developing Disaster-resilient Societies—MARTEST" promoted by the Japan International Cooperation Agency (JICA) and Japan Science and Technology Agency (JST).

Data availability

The pre-training and pre-testing data are available from the STanford EArthquake Dataset (STEAD) (Mousavi et al. 2019) at <https://github.com/smousavi05/STEAD> (last accessed May 2024). INSTANCE data (Michelini et al. 2021) are available at <https://github.com/INGV/instance> (last accessed May 2024). The Sichuan and Yunnan datasets are not publicly available and require an application for access (https://data.earthquake.cn/dashshare/website/data/contwave_csn.jsp). The single-station location using machine learning code is available at <https://github.com/LolitaZJ/DISAZI>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Earthquake Research Institute, University of Tokyo, Tokyo, Japan. ²School of Earth and Space Sciences, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China. ³Institute of Engineering Mechanics, China Earthquake Administration, Harbin, People's Republic of China.

Received: 8 October 2024 Accepted: 16 July 2025
Published online: 11 August 2025

References

- Allen RV (1978) Automatic earthquake recognition and timing from single traces. *Bull Seismol Soc Am* 68:1521–1532. <https://doi.org/10.1785/bssa0680051521>
- Ashkenazy Y, Kurzon I, Asher EE (2024) Earthquake activity as captured using the network approach. *Chaos Solitons Fractals* 186:115290. <https://doi.org/10.1016/j.chaos.2024.115290>
- Beskardes GD, Hole JA, Wang K et al (2018) A comparison of earthquake backprojection imaging methods for dense local arrays. *Geophys J Int* 212:1986–2002. <https://doi.org/10.1093/gji/ggx520>
- Brodsky EE (2019) The importance of studying small earthquakes. *Science* 364:736–737. <https://doi.org/10.1126/science.aax2490>
- Cesca S, Grigoli F (2015) Full waveform seismological advances for microseismic monitoring. *Adv Geophys* 56:169–228. <https://doi.org/10.1016/bs.agph.2014.12.002>
- Crotwell HP, Owens TJ, Ritsema J (1999) The TauP toolkit: flexible seismic travel-time and ray-path utilities. *Seismol Res Lett* 70:154–160. <https://doi.org/10.1785/gssrl.70.2.154>
- Ester M, Kriegel HP, Sander J, Xu X (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In: *kdd*, Vol. 96, No. 34, pp 226–31.
- Frohlich CA, Pulliam J (1999) Single-station location of seismic events: a review and a plea for more research. *Phys Earth Planet Inter* 113:277–291. [https://doi.org/10.1016/s0031-9201\(99\)00055-2](https://doi.org/10.1016/s0031-9201(99)00055-2)
- Gajewski D, Anikiev D, Kashtan B, et al (2007) Localization of seismic events by diffraction stacking. In: 2007 SEG annual meeting
- Geiger L (1912) Probability method for the determination of earthquake epicentres from the arrival time only. *Bull St Louis Univ, London*
- Goodfellow I, Bengio Y, Courville A (2016) Deep learning. The Mit Press, Cambridge
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
- Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. *Neural Comput* 9:1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Kennett BLN, Engdahl ER (1991) Traveltimes for global earthquake location and phase identification. *Geophys J Int* 105:429–465. <https://doi.org/10.1111/j.1365-246x.1991.tb06724.x>
- Kennett N, Cambridge M (1992) Earthquake location — genetic algorithms for teleseisms. *Phys Earth Planet Inter* 75:103–110. [https://doi.org/10.1016/0031-9201\(92\)90121-b](https://doi.org/10.1016/0031-9201(92)90121-b)
- Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. In: arXiv.org. <https://arxiv.org/abs/1412.6980>
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
- Lomax A, Michelini A, Jozinović D (2019) An investigation of rapid earthquake characterization using single-station waveforms and a convolutional neural network. *Seismol Res Lett* 90:517–529. <https://doi.org/10.1785/0220180311>
- Magotra N, Ahmed N, Chael EP (1987) Seismic event detection and source location using single-station (three-component) data. *Bull Seismol Soc Am* 77:958–971. <https://doi.org/10.1785/bssa0770030958>
- Michelini A, Cianetti S, Gaviano S, et al (2021) INSTANCE – the Italian seismic dataset for machine learning. *Earth System Science Data* 13:5509–5544. <https://doi.org/10.5194/essd-13-5509-2021>
- Miglani A, Kumar N (2019) Deep learning models for traffic flow prediction in autonomous vehicles: a review, solutions, and challenges. *Vehicular Commun* 20:100184. <https://doi.org/10.1016/j.vehcom.2019.100184>
- Mousavi SM, Beroza GC (2020) Bayesian-deep-learning estimation of earthquake location from single-station observations. *IEEE Trans Geosci Remote Sens* 58:8211–8224. <https://doi.org/10.1109/tgrs.2020.2988770>
- Mousavi SM, Sheng Y, Zhu W, Beroza GC (2019) STanford earthquake dataset (STEAD): a global data set of seismic signals for AI. *IEEE Access* 7:179464–179476. <https://doi.org/10.1109/access.2019.2947848>
- Mousavi SM, Ellsworth WL, Zhu W et al (2020) Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat Commun*. <https://doi.org/10.1038/s41467-020-17591-w>
- Perol T, Gharbi M, Denolle M (2018) Convolutional neural network for earthquake detection and location. *Sci Adv* 4:e1700578. <https://doi.org/10.1126/sciadv.1700578>
- Pesicek JD, Child D, Artman B, Ciešlik K (2014) Picking versus stacking in a modern microearthquake location: comparison of results from a surface passive seismic monitoring array in Oklahoma. *Geophysics* 79:KS61–KS68. <https://doi.org/10.1190/geo2013-0404.1>
- Prechelt L (2002) Early stopping-but when? *Neural Networks: tricks of the trade*
- Ronneberger O, Fischer P, Brox T, (2015) U-Net: convolutional networks for biomedical image segmentation, *Miccai*, pp 234–241.
- Cambridge M (2002) Monte Carlo methods in geophysical inverse problems. *Rev Geophys*. <https://doi.org/10.1029/2000rg000089>
- Zhang X, Zhang J, Yuan C et al (2020) Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method. *Sci Rep*. <https://doi.org/10.1038/s41598-020-58908-5>
- Zhebel O, Eisner L (2015) Simultaneous microseismic event localization and source mechanism determination. *Geophysics* 80:KS1–KS9. <https://doi.org/10.1190/geo2014-0055.1>
- Zhou Y, Yue H, Kong Q, Zhou S (2019) Hybrid event detection and phase-picking algorithm using convolutional and recurrent neural networks. *Seismol Res Lett* 90:1079–1087. <https://doi.org/10.1785/0220180319>
- Zhu W, Beroza GC (2018) Phasenet: a deep-neural-network-based seismic arrival time picking method. *Geophys J Int* 216:261–273. <https://doi.org/10.1093/gji/ggy423>
- Zhu W, McBrearty IW, Mousavi SM et al (2022) Earthquake phase association using a Bayesian Gaussian mixture model. *J Geophys Res Solid Earth*. <https://doi.org/10.1029/2021jb023249>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.