

LSTM-Autoencoder Network for the Detection of Seismic Electric Signals

Jiyan Xue^{ID}, Qinghua Huang^{ID}, Sihong Wu^{ID}, and Toshiyasu Nagao

Abstract—Seismic electric signals (SE斯) are essential short-term precursors of earthquakes. Accurate and efficient detection of SE斯 is significant to short-term predictions of earthquakes. However, SE斯 are usually disturbed by various noises and are thus difficult to recognize. Although conventional techniques have made substantial efforts in improving the SE斯 detection accuracy, the success rates of SE斯 detection at certain stations are still less satisfactory due to the complexity and diversity of noises. In this study, we apply deep learning to extract SE斯 and develop a novel deep learning network based on geoelectric field characteristics by combining the long short-term memory (LSTM) blocks with an autoencoder structure and a time-step attention module. The detection results of both synthetic and real data demonstrate that our proposed network yields superior performance in detecting embedded SE斯 in the presence of severe noise interference compared with traditional methods and several well-known networks. Moreover, our novel network shows the excellent ability of massive data processing, generalization, and migration, which can process one day’s worth of data in only milliseconds, adapt to SE斯 whose durations and amplitudes are different from those of the training set and be easily transferred to newly acquired data. The proposed novel method can provide more efficient and accurate detection results, which will broaden the data availability of hazard mitigation based on SE斯.

Index Terms—Long short-term memory (LSTM)-autoencoder, seismo-electromagnetism, seismic electric signal (SE斯) detection.

I. INTRODUCTION

SHORT-TERM (timescale of a year or shorter) earthquake prediction based on potential precursory phenomena is very important to mitigate earthquake disasters in seismically active regions [1]. Through a large number of retrospective experiments, researchers have explored and summarized a number of precursory phenomena possibly associated with earthquake occurrences [2]. The known short-term earthquake precursors include various geophysical and geochemical

Manuscript received March 20, 2022; revised May 23, 2022; accepted June 6, 2022. Date of publication June 15, 2022; date of current version June 24, 2022. This work was supported by the National Natural Science Foundation of China under Grant 42021003. (Corresponding author: Qinghua Huang.)

Jiyan Xue, Qinghua Huang, and Sihong Wu are with the Department of Geophysics, School of Earth and Space Sciences, Peking University, Beijing 100871, China, and also with the Hebei Hongshan National Observatory on Thick Sediments and Seismic Hazards, Peking University, Beijing 100871, China (e-mail: xjy2001110580@stu.pku.edu.cn; huangq@pku.edu.cn; sihongwu@pku.edu.cn).

Toshiyasu Nagao is with the Institute of Oceanic Research and Development, Tokai University, Shizuoka 424-0902, Japan (e-mail: nagao@scc.u-tokai.ac.jp).

Digital Object Identifier 10.1109/TGRS.2022.3183389

phenomena, e.g., mechanical deformation [3], [4], seismicity change [5]–[9], gas emission [10], [11], variations of groundwater level [12], [13] and ground temperature [14], geoelectric [15]–[17] and geomagnetic [18], [19] anomaly, and fluctuations of ionospheric and magnetospheric parameters [20], [21]. Since the accumulation of stress during earthquake preparation can result in microscopic changes in the lithosphere, which thus disturb the geoelectric signals with distinct characteristics, researchers have become increasingly interested in searching for possible precursors in geoelectric data to predict short-term earthquakes in the last two decades [22].

Varotsos and Alexopoulos [15] systematically summarized the correlation between impending earthquakes and geoelectric field anomalies and defined the observable geoelectric anomaly changes before earthquakes as seismic electric signals (SE斯). According to the observation data accumulated so far, SE斯 are usually approximately rectangular-shaped waveforms in the time domain and occur within a few weeks before earthquakes, with durations ranging from a few minutes to several hours and amplitudes of the order of 10 mV/km [15], [16], [23]–[26]. Varotsos [27] employed years of observations to reveal long-term temporal correlations between SE斯 and seismicity before impending earthquakes. Uyeda and Meguro [28] tested the post-prediction of 15 earthquakes of magnitude 5.0 or greater in Greece between 1985 and 2003 based on the amplitude and duration characteristics of SE斯 and achieved an overall “prediction” success rate of above 70%, which demonstrates that proper identification of SE斯 may be an effective approach to earthquake prediction.

Geoelectric fields are usually disturbed severely by various noises. Natural noises include rainfall, lightning, and the inductions of geomagnetic field variation and the Earth’s ocean tides [24]. The interference from natural noises is relatively weak. However, with rapid urbanization and modernization, large-scale industrial electrical facilities are installed in suburban areas, and rail transit systems are built in large cities. The resulting high-voltage transmission and stray currents bury the SE斯 in noises and make them difficult to recognize [29]–[31].

Particle motion diagram (PMD) is a technique used to extract SE斯 from strongly disturbed data [32]. However, its performance in practical application is unsatisfactory. For example, it fails to detect the SE斯 known to be present on January 17, 1999 in Nagano Prefecture, central Japan. Independent component analysis (ICA) is the most effective method so far to solve the SE斯 detection

problem [24], [33], [34]. The geoelectric field data are decomposed into several independent components to distinguish SESs and railway noise. The ICA-based algorithm FastICA is designed for fixed signal sources [35]. However, long measuring dipoles at a number of stations intersect the railway in Japan, and the noises from moving trains cannot be regarded as fixed sources. Therefore, the detection accuracy by the ICA method at such stations is less desirable [24].

Recently, due to the popularity of many deep learning architectures and increasingly powerful computational hardware, deep learning has become the focus of interest in geophysics [36]. Data-driven deep learning has brought a profound revolution in solving geophysical problems and has been widely applied to various tasks such as fault detection [37], [38], seismic wave arrival-time picking [39], [40], anomaly detection [41], and noise suppression [29], [42], [43]. As the long short-term memory (LSTM) network has excelled in time series applications and the autoencoder network is shown to be more suitable for data compression and feature extraction, the combination of these two networks has been successfully applied to geoelectric field data denoising with excellent performance [29].

In this study, we establish an LSTM-autoencoder network to detect SESs from noise disturbed geoelectric fields. In the remainder of this article, we first introduce the sources and characteristics of the data we use as well as the process of synthesizing the training dataset. Then, we elaborate the structure and implementation details of the LSTM-autoencoder as well as the metrics for evaluating the network performance. The network's performance in detecting synthetic and real SESs is examined and compared with the ICA method. We demonstrate the generalizability and transferability of the network and conclude with future developments.

II. DATA

A. Data Sources and Characteristics

Since 1997, RIKEN has installed three geoelectric field continuous observation stations at Hakuba (HKB), Matsushiro (MTS), and Outaki (OTA) (see Fig. 1) in Nagano Prefecture, central Japan. Data collected by the stations are telemetered once a day to the Earthquake Prediction Research Center (EPRC) of Tokai University with a default sampling frequency of 0.1 Hz. In this study, we use the geoelectric field observations of 1999–2001 at station HKB and 1999–2000 at stations MTS and OTA. Among the three stations, the long measuring dipoles at station HKB intersect the railway track (see Fig. 1). The train leakage current cannot be regarded as a fixed noise source and affects the SES detection accuracy by the ICA method.

On January 17, 1999, anomalous changes with an approximately rectangular shape and a 27-min duration starting from 1:37 A.M. were observed simultaneously at the HKB, MTS, and OTA stations (see Fig. 2). The amplitudes of the anomalies varied between 7 and 23 mV/km. These anomalies are considered to be the SESs of the M4.9 earthquake that occurred 11 days later near the center of the triangle formed by the three stations [16]. In this study, we employ the nine

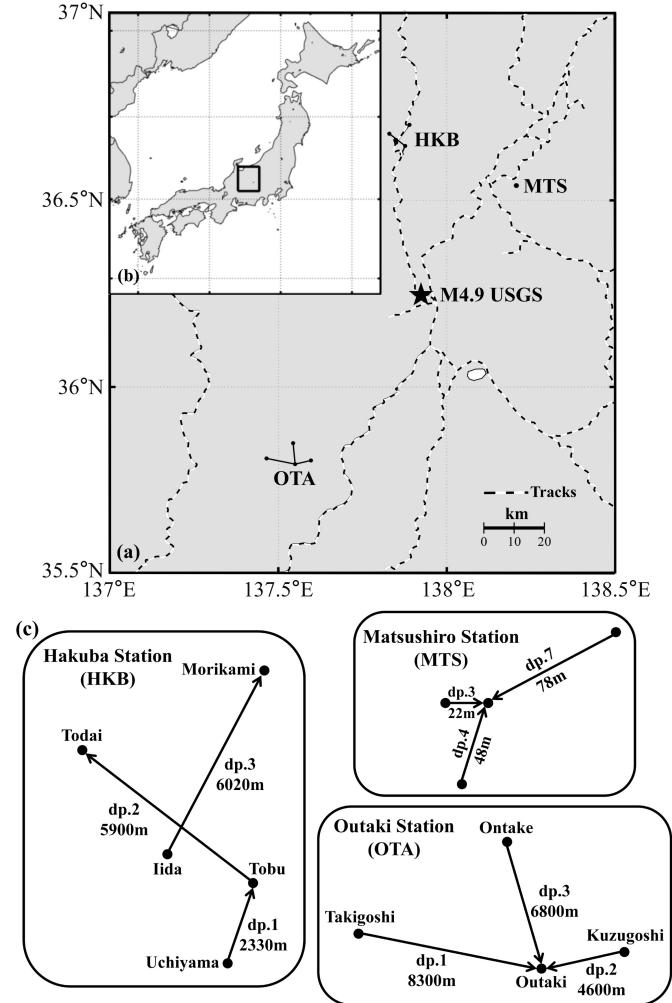


Fig. 1. (a) Zoomed-in view of the study area showing the layout of stations HKB, MTS, and OTA and distributions of local railway tracks. The black star represents the epicenter of the M4.9 earthquake on January 28, 1999. (b) Regional map showing the location of the study area (black box). (c) Configuration map of measuring dipoles at the three geoelectric field stations. Modified from [24].

presumed real SESs to develop our deep learning network for SES detection and examine its performance.

The amplitude of the SESs and train leakage currents noise recorded at the OTA station are smaller than those recorded at the HKB and MTS stations (see Fig. 2), probably because of the distance between the OTA station and the epicenter and railway [44]–[46], as well as the electrical structure beneath this station [47], [48]. In addition, the variation in train operation and leakage current energy may exacerbate the differences in the duration and amplitude of the railway noise [24], [33].

The SESs can be easily recognized when no trains were running, as shown in Fig. 2(b) with clear SESs in the quiet background data. However, the geoelectric field data were seriously disturbed by train leakage currents after 6 A.M. [see Fig. 2(a)], and it is difficult to obtain the characteristics of SESs directly from the observations due to the presence of nonstationary noises. Therefore, we propose the

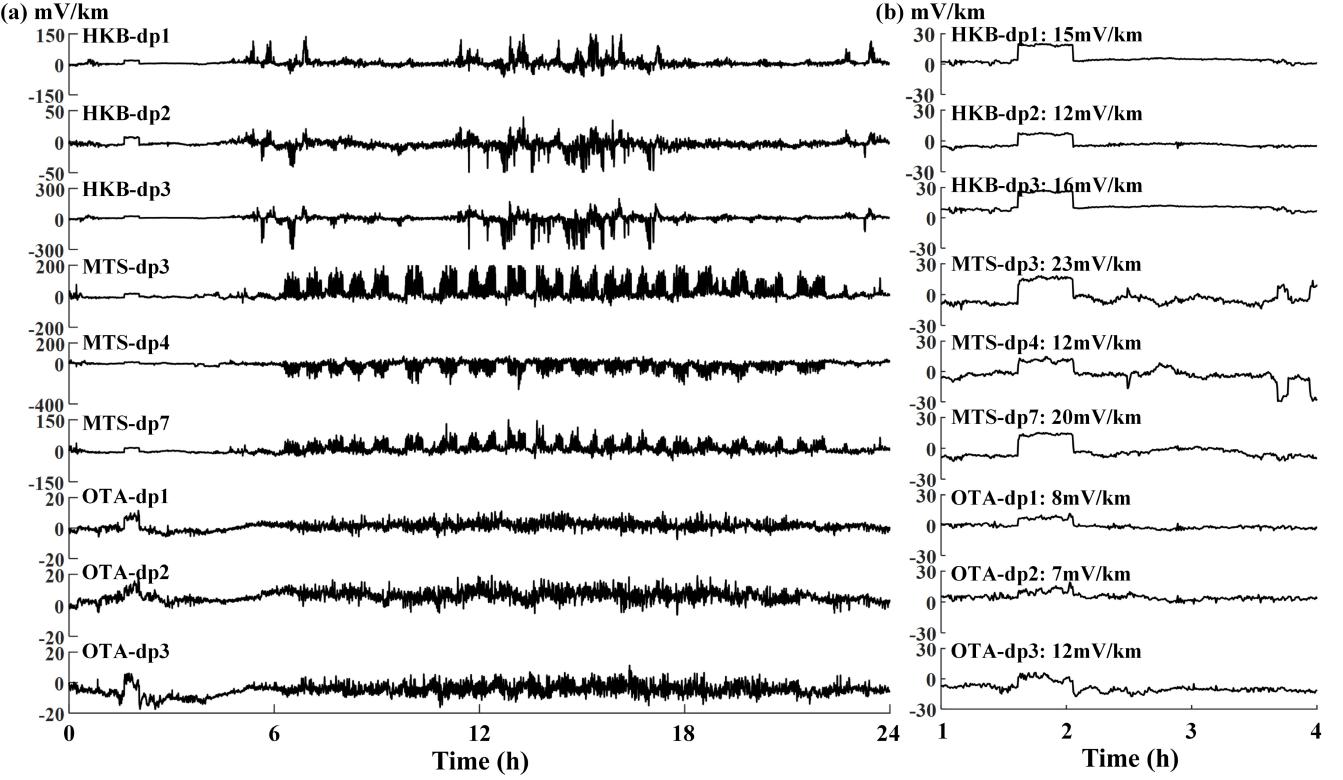


Fig. 2. Time series of the geoelectric field at stations: HKB, MTS, and OTA on January 17, 1999. (a) One-day data. (b) Midnight data from 01:00 to 04:00. SESSs with nearly rectangular shape can be seen clearly starting at 1:37 on the nine dipoles. The amplitudes of the SESSs are given.

LSTM-autoencoder network for the detection of SESSs from the highly disturbed background data.

B. Synthetic Data

The network training requires sufficient data to capture the SESS characteristics. As the known real SESSs are insufficient, we simulate SESSs by rectangular-shaped waveforms to generate the training dataset. The networks of the HKB, MTS, and OTA stations are trained separately using synthetic data for the respective stations. We use the continuous observation at the HKB station from 1999 to 2000 to illustrate the process of data generation. The entire observation (709 days in total after data cleaning) is divided into daily segments. We then repeat the data 30 times and randomly embed the simulated SESSs between 0:00 and 24:00. Based on the duration and amplitude characteristics of the real SESSs at the HKB station, the synthetic SESSs last between 25 and 30 min with amplitudes ranging from 10 to 15 mV/km. We finally obtain a total of 21 270 pairs of training and test datasets embedded with different SESSs.

Fig. 3 shows an example from the training set. We use the Savitzky–Golay and moving-average method to the original data of the HK station to obtain cleaned data. The network's input and output data are obtained by adding the simulated SESS to the observed and cleaned data. The cleaned data are not noise-free but clean enough to easily recognize SESSs, as most of the railway noise in the data is suppressed. If the amplitude difference between two consecutive samples is much smaller

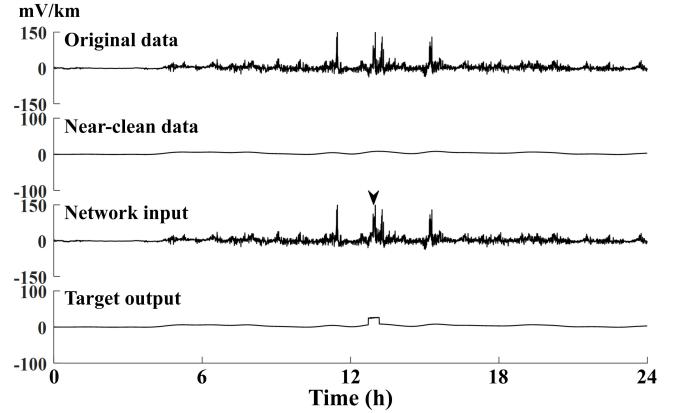


Fig. 3. Sample of synthetic data from the training set.

than that of the embedded simulated SESS, the data are considered cleaned.” The synthetic dataset is randomly shuffled and then divided into the training and test sets by a ratio of 9:1. Furthermore, to ensure the gradient descent direction accurate and reduce oscillation, we adopt the mini-batch [49] training with a batch size of 128.

III. METHOD

A. LSTM-Autoencoder

The basic autoencoder architecture (Appendix A) has an excellent ability to compress signals and extract

information [50], [51]. The LSTM network (Appendix B), developed specifically for time series, has been widely used in electromagnetic signal processing with excellent performance [52], [53]. Moreover, the attention module (Appendix C) has been demonstrated to improve the performance of networks effectively [54]. Therefore, we establish a seven-layer autoencoder network by combining the LSTM blocks with a time-step attention module (Fig. 4). The number of nodes in the input layer is 8640 to match the one-day data dimension with a sampling rate of 0.1 Hz. After the input layer, a convolutional neural network (CNN) layer is used to dilute the data with the time sequence characteristics retained, which significantly reduces the data dimension and the number of trainable network parameters. A bidirectional LSTM is introduced after the CNN layer to make full use of the temporal information both forward and backward to constrain the output of each LSTM block (Appendix B). The useful information in the bidirectional LSTM layer is no longer transmitted in one direction, and the result from the last time step cannot be reasonably selected as its output. Therefore, a time-step attention module is used as a weighted summation operator for all time steps of the bidirectional LSTM layer. Because the bidirectional LSTM already integrates temporal information both forward and backward, unidirectional LSTM networks are used to compress and reconstruct data in subsequent layers of the LSTM-autoencoder. Finally, a dense layer is used to restore the data into the same dimension as the input. To avoid the gradient disappearance and internal covariate shift, layer normalization is introduced into the established LSTM-autoencoder network [55].

The training parameters in our network include the weighting matrix \mathbf{W} and bias vector \mathbf{b} , which are initialized randomly within the range of $(0, 1]$ and updated iteratively during the network training to minimize the loss function. We apply the mean square error (MSE) metric to penalize the misfit between the target output x^t and the predicted output x^p and add the regularization terms to improve training stability. The target output x^t denotes the manually denoised original data with SESs embedded (“Target output” in Fig. 3), while the predicted output x^p denotes the network-predicted result for the original data embedded with SESs. Thus, the loss function $L(\mathbf{W}, \mathbf{b})$ is defined as [43]

$$L(\mathbf{W}, \mathbf{b}) = \frac{1}{2N} \left[\sum_{i=1}^N (x_i^t - x_i^p)^2 + \lambda_w \sum_{m=1}^M w_m + \lambda_b \sum_{k=1}^K b_k \right] \quad (1)$$

where N , M , and K are the number of training data, matrix \mathbf{W} , and vector \mathbf{b} elements, respectively, w_m and b_k are the elements of the matrix \mathbf{W} and vector \mathbf{b} , and λ_w and λ_b represent the corresponding regularization coefficients.

B. Evaluation Metrics

To evaluate the network convergence, we employ the root-mean-square percentage error (RMSPE) given by

$$\text{RMSPE}(x^o, x^p) = \sqrt{\frac{1}{N} \cdot \sum_{n=1}^N \left(\frac{x_n^o - x_n^p}{x_n^o} \right)^2} \times 100\% \quad (2)$$

where x_n^o and x_n^p are network label and predicted value for the n th sample, respectively, and $N = 8640$ is the number of sampling points.

To evaluate the success rate of SES recognition, we first stipulate that an SES extraction is successful if the following conditions are met:

$$\begin{cases} \text{ST}_N - \text{ST}_S < \frac{1}{2} \cdot \tau_S \\ \text{ET}_N - \text{ET}_S < \frac{1}{2} \cdot \tau_S \end{cases} \quad (3)$$

where ST_N and ET_N represent the start and end times of the network predicted SES, respectively, while ST_S and ET_S represent those of the embedded simulated SES, respectively. τ_S is the duration of the embedded SES. Equation (3) indicates that a successful SES extraction is obtained if the two time differences are both less than 50% of the embedded SES duration. The success rate is then defined as the ratio of the number of SESs successfully extracted (N_2) to the total number of SESs (N_1)

$$\text{Success Rate} = \frac{N_2}{N_1} \times 100\%. \quad (4)$$

C. Implementation

An Adam optimizer [56] with a batch size constant of 128 and a learning rate of 10^{-3} with a decay rate of 8×10^{-6} is implemented in the network training. We conduct 100 training epochs and perform the validation after each training epoch to monitor the training effect. We use open-source libraries TensorFlow and Keras to support all the abovementioned deep learning concepts and optimization algorithms. All computations are carried out with a desktop PC equipped with an NVIDIA GeForce RTX 3070 GPU and an AMD Ryzen 7 3700X CPU with 16-GB memory.

IV. RESULTS

A. Detection of Simulated SESs

We first evaluate the network performance in detecting simulated SESs. The success rates and the RMSPE values of the training and test sets are monitored during training (see Fig. 5). After 90 epochs, the success rates of the training and test sets reach the highest level, and the RMSPE value of the test set no longer decreases with the increase of epoch number, indicating that the network has achieved optimal performance. Therefore, we select the network tuned at the 98th epoch as the final well-trained network. Our network training takes approximately 38 min in total. The RMSPE value and success rate on the test set are approximately 3.3% and 97.6%, respectively. Fig. 6 shows two detection results from the test set.

The LSTM-autoencoder network has an excellent performance in correctly recognizing and extracting SESs from strongly interfered backgrounds. Furthermore, our network takes only 2.1 ms to process one whole day's worth of data, which is highly efficient and suitable for detecting SESs from massive datasets.

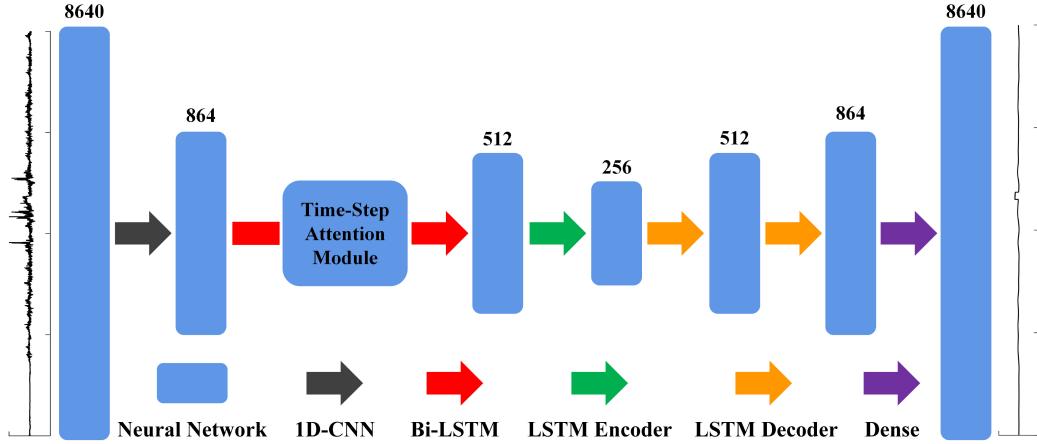


Fig. 4. LSTM-autoencoder network architecture. The network input is the original data with simulated SESs added, with a total of 8640 evenly spaced sampling points in time. The output label is the cleaned data with the same simulated SES added. The blue stripes represent layers inside the neural network, with the numbers on top denoting the layer dimensions. The black arrow represents the 1-D CNN layer with the number of convolution kernels of 1, size, and moving step of 10, and the rectified linear unit (ReLU) activation function. Layer normalization is introduced between adjacent network layers.

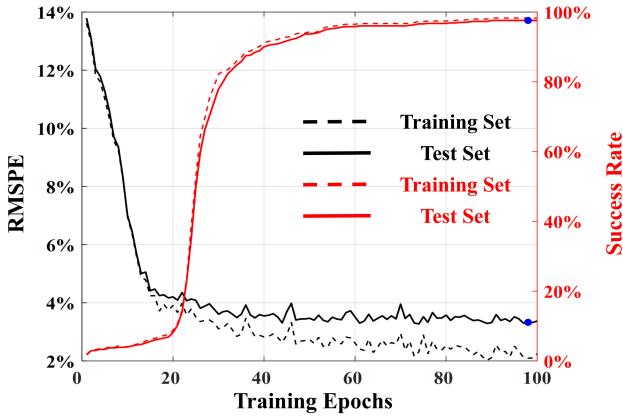


Fig. 5. Variations of the RMSPE values (black curves) and success rates (red curves) with training epochs for the training (dashed lines) and test sets (solid lines). The solid blue dots indicate locations on the curves at the 98th epoch.

B. Detection of Real SESs

Nine SESs suspected generated shortly before the January 28, 1999 M4.9 earthquake in Nagano Prefecture, central Japan, were observed simultaneously at three local stations (see Fig. 2). Due to the lack of known real SES samples during strong disturbances, we adopt the synthesis method of Orihara *et al.* [24] to test the accuracy of our LSTM-autoencoder network in detecting real SESs from data with strong interferences. Data from randomly selected seven days are used, and the SESs starting at 1:37 are embedded at 4:00, 13:00, and 18:00 at each station; 1:37 and 4:00 are when the railway noise interference is absent, while the geoelectric fields at 13:00 and 18:00 are strongly disturbed by trains. A total of 198 experimental datasets are obtained (see Table I). Unlike in the aforementioned synthetic datasets, the embedded SESs are generated by the actual earthquake and directly extracted from the observations, rather than simulated SESs with a rectangular shape.

The real SESs extracted from the geoelectric field data at the HKB, MTS, and OTA stations are recognized by

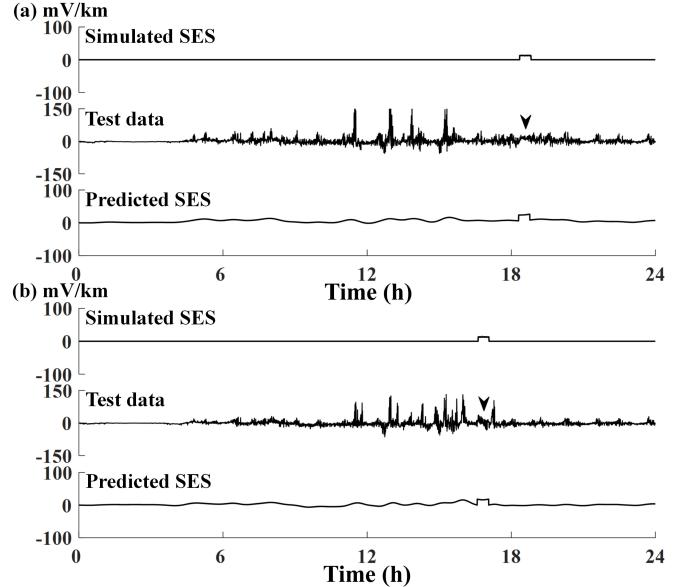


Fig. 6. (a) Sample from the test set, including from top to bottom the simulated SES, the synthetic data embedded with the simulated SES as the input data, and the network output. The simulated SES is embedded in the time range 18:21:20–18:49:20, and the time range determined by the network is 18:18:40–18:46:20. (b) Same as (a) but for another sample. The simulated SES is embedded in the time range 15:16:00–15:44:00, and the time range determined by the network is 15:12:40–15:41:00.

our well-trained LSTM-autoencoder network. The detection results are classified into three categories according to the criteria specified in (3). Detection results that meet the criteria are considered as “success.” If the two time differences specified in (3) are both between 50% and 100% of the SES duration, the detection results are considered as “near success.” Anything else is classified as “failure.” Table I lists the statistical results of the whole test, with several examples shown in Fig. 7.

In Table I, the overall recognition success rate is above 95%, which demonstrates that although rectangular-shaped synthetic

TABLE I
REAL SES DETECTION RESULTS FOR THE OBSERVATIONS AT STATIONS HKB, MTS, AND OTA

Randomly Selected Data	Time for Embedding	HKB dp1	HKB dp2	HKB dp3	MTS dp3	MTS dp4	MTS dp7	OTA dp1	OTA dp2	OTA dp3
1999-01-17	01:37	✓	✓(a)	✓	✓	□	✓	✓	✓	✗
	04:00	✓	✓	✓	✗	□	✓	✓	✓	✓
1999-02-16	13:00	✓	✓	✓	✓	✓	□(b)	✓	✓	✓
	18:00	✓	✓	✓	✓	✓	✓	✓	✓	✓
1999-05-03	04:00	✓	✓	✓	✓	✓	✓	✓	✓	✓
	13:00	✓	✓	✗(c)	✓	✓	✓	✓	✓	✓
	18:00	✓	✓	✓	✓	✓	✓	✓	✓	□
1999-07-18	04:00	✓	✓	✓	✓	✓	✓	✓	✓	✓
	13:00	✓	✓	✗	✓	✓	✓	✓	✓	✓
	18:00	✓	✓	✓	✗	✓	✓	✓	✗	✓(d)
1999-07-29	04:00	✓	✓	✓	✓	✓	✓	✓	✓	✓
	13:00	✓	✓	✗	✓	✓	✓	□(e)	✓	✗
	18:00	✓	✓	✓	✗	✓	✓	✓	✓	✓
2000-03-01	04:00	✓	✓	✓	✓	✓	✓	✓	✓	✓
	13:00	✓	✓	✓	✓	✓	✓	✓	□	✓
	18:00	✓	✓	✓	✓	✓	✓	✓	✓	✓
2000-03-21	04:00	✓	✓	✓	✓	✓	✓	✓	✓	✓
	13:00	✗(f)	✓	✓	✓	✓	✓	✓	✓	✓
	18:00	✓	✓	✓	✓	✓	✓	✓	✓	✓
2000-11-22	04:00	✓	✓	✓	✓	✓	✓	✓	✓	✓
	13:00	✓	✓	✓	✓	✓	✓	✓	✓	✓
	18:00	✓	✓	✓	✓	✓	✓	✓	✓	✓

Note: Real SESs observed at 1:37 are embedded at 4:00, 13:00 and 18:00. Detectability of SESs recovered by the LSTM-autoencoder at each station is marked by “✓”, “□” and “✗” for “success”, “near success” and “failure”, respectively. Results labeled (a)–(f) are shown in Fig. 7

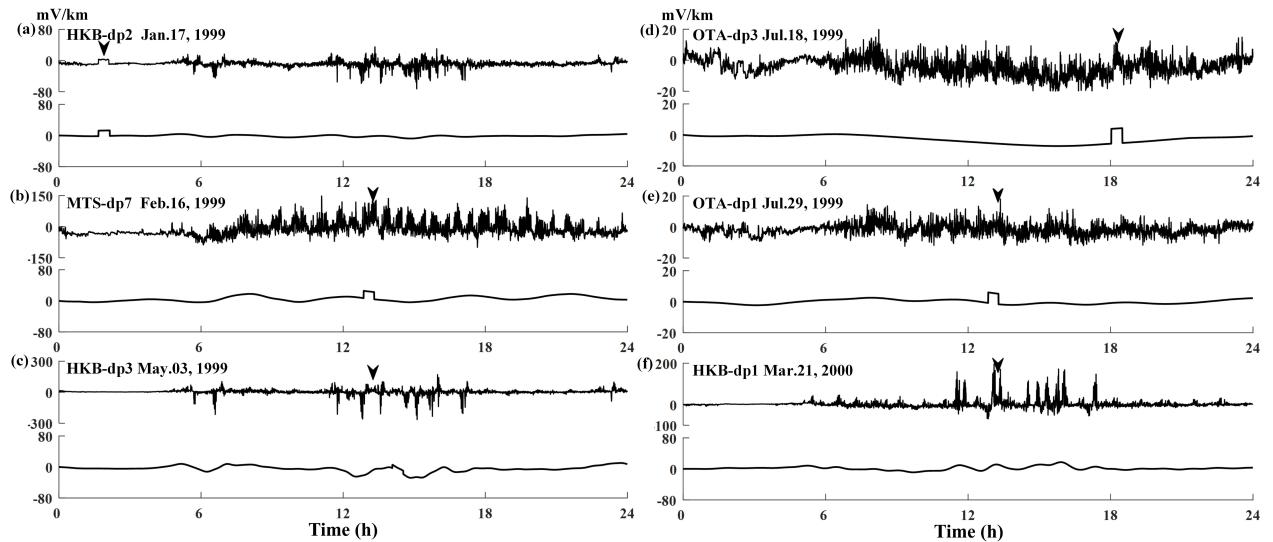


Fig. 7. Data embedded with real SESs and the corresponding network outputs. (a)–(f) are six examples labeled in Table I.

SEs are used in training, the LSTM-autoencoder network can still succeed in detecting real SEs with high accuracy. The

network detection of “failure” may be related to the SNR of data and the difference between the real and synthetic SEs.

TABLE II
COMPARISON OF THE RMSPE VALUES AND SUCCESS RATES OF THE FIVE WELL-TRAINED NETWORKS

Networks	RMSPE	Success Rate
FC-Autoencoder	3.7%	87.7%
FCT-Autoencoder	3.7%	87.9%
LSTM-Autoencoder without time-step attention module	3.6%	92.6%
GRU-Autoencoder	3.4%	95.4%
LSTM-Autoencoder	3.3%	97.6%

In general, the success rates at the HKB, MTS, and OTA stations are about 94%, 96%, and 96%, respectively, indicating an overall high performance.

V. DISCUSSION

A. Comparison With the ICA Method

Orihara *et al.* [24] used the ICA method to detect SESSs from observations collected before the January 28, 1999 earthquake at the HKB, MTS, and OTA stations. The simultaneous SESSs can be recognized even in daytime records at least for the MTS and OTA stations by the ICA method, which was significant progress compared with other conventional methods. However, due to the inherent limitations of the ICA method, the SESS detection accuracy is less desirable for observations collected at the HKB station with dipoles intersecting the railway. This can be seen in the considerably lower SESS detection accuracy at the HKB station than MTS and OTA stations [24].

However, the statistical results in Table I suggest that the LSTM-autoencoder network significantly improves the success rate of the HKB station and achieves equally high detection accuracy at all three stations, indicating that our proposed network can adapt to the dipole arrangement in noisy environments and is thus applicable for various scenarios.

B. Comparison With Several Well-Known Networks

To verify the effectiveness of our proposed network for SESS detection, we conduct a quantitative comparison with several other well-known deep learning algorithms, including the fully convolutional (FC) autoencoder [57], the fully connected (FCT) autoencoder [58], the LSTM autoencoder without the time-step attention module, and the gate recurrent unit (GRU) autoencoder [59] with the attention module. To ensure fairness, the synthetic data acquired to train and test these networks are identical to the data we used previously (Section II-B). Table II shows the final RMSPE values and success rates of the same training and test set using these well-trained networks.

First, comparing the results of the LSTM autoencoder with and without the attention module, the RMSPE value decreases by 0.3%, and the success rate increases by 5.0%, which proves that the time-step attention module can significantly improve the network performance. In addition, compared with the well-known FC, FCT, and GRU autoencoders, the success

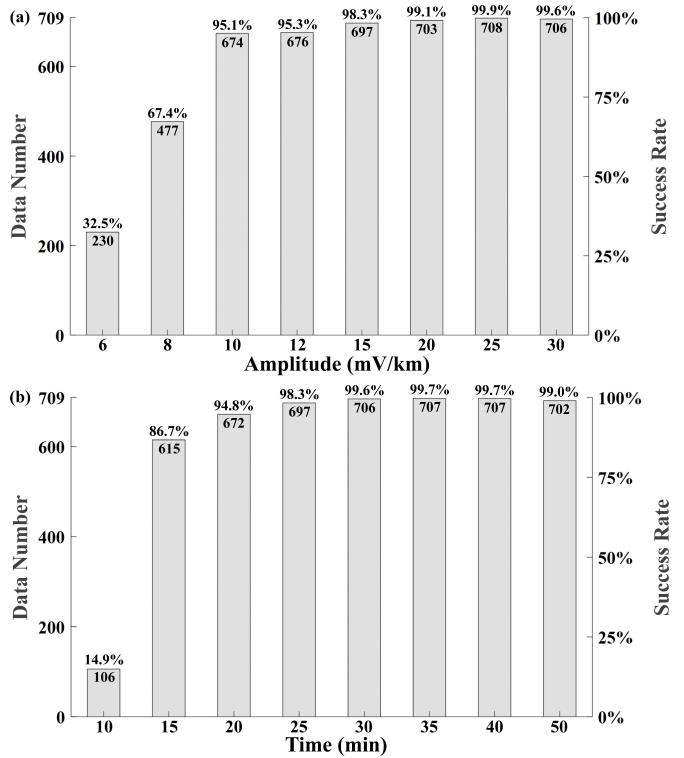


Fig. 8. (a) SES detection number and success rate for the same duration (25 min) but different amplitudes (6–30 mV/km). Each test set contains 709 samples with embedded SESSs. The gray columns indicate the numbers of SESSs successfully recognized (left axis) and the success rates of SESS detection (right axis) for all test sets. (b) Same as (a) but for the same amplitude (15 mV/km) and different durations (10–50 min).

rate of the LSTM autoencoder increases by 9.9%, 9.7%, and 2.2%, respectively. Because the LSTM and GRU blocks are designed specifically for time series, the corresponding detection accuracy is significantly higher than that of other networks. Moreover, the GRU block is simplified based on the LSTM block, which can explain that the accuracy of the LSTM autoencoder is slightly higher than that of the GRU autoencoder. The comparisons demonstrate the advantages of our proposed network.

C. Generalization Capability of the Network

We have demonstrated that the LSTM-autoencoder network can successfully detect SESSs with similar amplitudes and durations to the training set. In this section, we use the network trained by synthesizing data from the HKB station as an example and investigate the ability of our network in recognizing SESSs with different amplitudes and durations from those of the training set.

We generate 16 test sets with SESSs of different amplitudes and durations and examine the success rate of SESS detection for each test set. The statistical results for different amplitudes shown in Fig. 8(a) indicate that the detection accuracy of SESSs decreases rapidly when the SESS amplitude in the test set (6–8 mV/km) becomes lower than those in the training set. However, the detection accuracy still maintains a satisfactory level when the SESS amplitudes in the test set are higher than

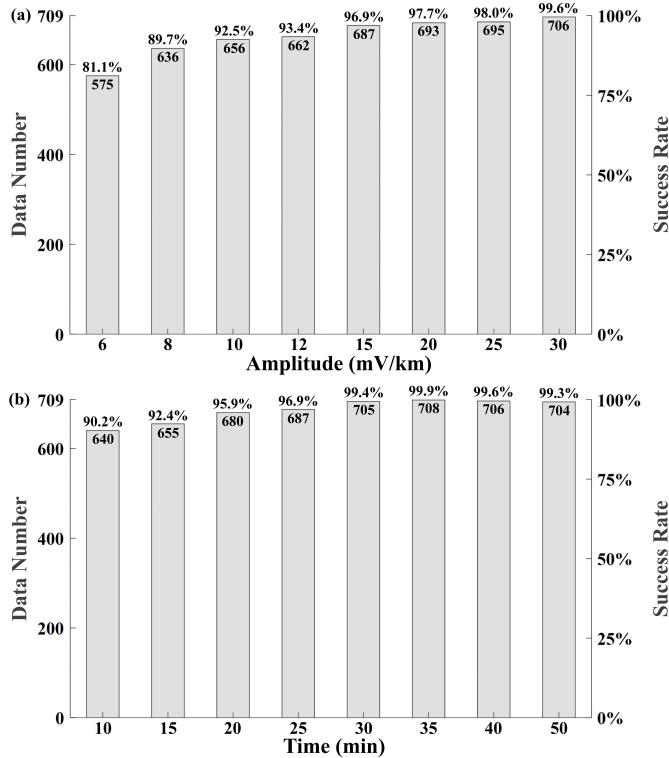


Fig. 9. (a) and (b) are the same as Fig. 8 but for the newly-trained network. The amplitudes of the simulated SESs embedded in the new training set range of 6–8 mV/km and the durations vary from 10 to 15 min.

the minimum amplitude of SESs in the training set. Similarly, the statistical results for different durations in Fig. 8(b) lead to almost the same conclusion. The detectable ranges of amplitude and duration depend heavily on the training samples. Our network can be used to detect SESs with the same or greater amplitudes and durations than the training set.

The statistical results shown in Fig. 8 indicate that the current network is less satisfactory for detecting SESs with amplitudes below 10 mV/km and durations below 15 min. We reembed SESs with amplitudes of 6–8 mV/km and durations of 10–15 min into the data to generate a new training set to retrain the network. The newly trained network is applied to the aforementioned test sets. Statistical results show that the newly trained network increases the recognition success rates from 32.5% and 67.4% to 81.1% and 89.7% for the SES amplitudes of 6 and 8 mV/km, respectively, and the recognition success rate from 14.9% to 90.2% for SES with a duration of 10 min (see Fig. 9). The newly trained network using decreased SES amplitude and duration improves the detection accuracy of SESs with smaller amplitudes and shorter durations.

D. Migration Capability of the Network

To improve the LSTM-autoencoder network performance in processing new data collected from the HKB station, we take the data collected by the station in 2001 as an example to conduct transfer learning based on the network trained by the data collected in the previous two years. The observations at

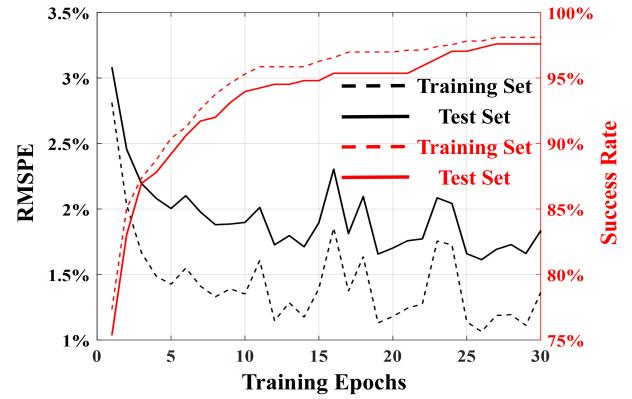


Fig. 10. Same as Fig. 5 but for the transfer training result.

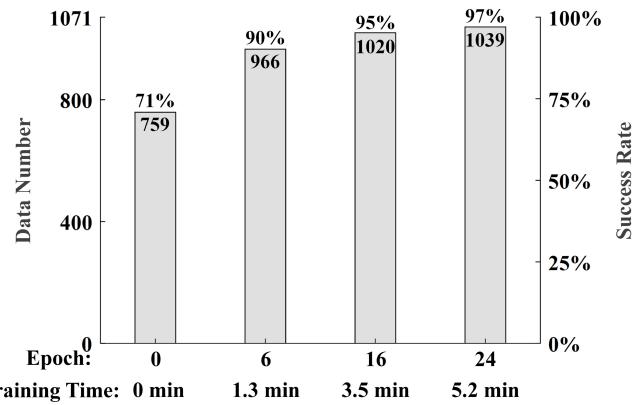


Fig. 11. Same as Fig. 8 but for different numbers of training epochs. Epoch 0 indicates that the LSTM-autoencoder network is used directly without transfer training.

the HKB station in 2001 are processed in the same way as described in Section II-B. We obtain 9639 pairs of training datasets and 1071 pairs of test datasets. The RMSPE values and the success rates of the training and test sets are monitored throughout the training process as before (see Fig. 10). The success rate for the test set stabilizes at above 97% after 24 epochs of transfer learning.

Fig. 11 shows the statistical results and success rates for the test set at four selected epochs. If the network without transfer training is used directly to process the test set, corresponding to the zeroth epoch, the success rate is about 71%. At the 6th, 16th, and 24th training epochs, the success rate reaches 90%, 95%, and 97%, with the transfer training times of 1.3, 3.5, and 5.2 min, respectively. Our LSTM-autoencoder network is easily adapted to the newly collected data by transfer training in minutes, significantly improving the SES detection performance.

E. Future Developments

In this study, we simulate SESs with rectangular-shaped waveforms based on the accumulated observation data and retrospective studies [15], [16], [24]–[26], [33], in particular the observations collected at the HKB, MTS, and OTA stations

before the January 28, 1999 M4.9 earthquake. Our well-trained LSTM-autoencoder network recognizes the specific SESs with approximate rectangular-shaped waveform characteristics. However, with the growing amount of observational data and more in-depth analyses, other types of SESs have been reported, such as high-frequency disturbances and intermittent pulses [25]. In future research, we may implement SESs with other waveform characteristics in the training set to achieve a more adaptable recognition ability.

VI. CONCLUSION

In this study, we combine deep learning with geoelectric field characteristics to develop a novel autoencoder network by employing the time-step attention module and LSTM memory blocks. We demonstrate that, by both synthetic and real data, the LSTM-autoencoder network can effectively detect the simulated SESs in the presence of severe railway noise interference, with a success rate of above 95%. Although our method needs a long training period to learn the implicit relationships between inputs and outputs, once trained, the LSTM-autoencoder network can efficiently process one whole day's worth of data in milliseconds.

The LSTM-autoencoder network has more general applicability than the conventional ICA method and can significantly improve the success rate for stations whose dipoles intersect a railway. Compared with other commonly used networks such as the FC-autoencoder, FCT-autoencoder, GRU-autoencoder, and LSTM-autoencoder without attention module, our proposed network obtains better accuracy and higher SES detection success rate. In addition, our network's excellent generalization and migration capabilities make it more practical. Results presented in this study suggest that the LSTM-autoencoder network can be an efficient and accurate novel means to detect SESs from massive geoelectric field data.

APPENDIX

A. Autoencoder

Autoencoder is a special kind of neural network architecture that can obtain an abbreviated representation of the input [50], [51], [58]. The basic autoencoder generally consists of two parts, encoder and decoder, and contains three layers: the input, hidden, and output layers. The number of nodes in the input layer is the same as that in the output layer, while the number of nodes in the hidden layer is usually the least, which is convenient for feature extraction. The computation involved is expressed as follows:

$$\text{Encoder : } \mathbf{h} = f_e(\mathbf{W}_e \mathbf{x} + \mathbf{b}_e) \quad (\text{A1})$$

$$\text{Decoder : } \mathbf{y} = f_d(\mathbf{W}_d \mathbf{h} + \mathbf{b}_d) \quad (\text{A2})$$

where \mathbf{x} , \mathbf{h} , and \mathbf{y} stand for the input, hidden, and output layers, respectively; f_e and f_d are nonlinear functions of the encoder and decoder, respectively; and \mathbf{W}_e and \mathbf{W}_d are weighting matrices, while \mathbf{b}_e and \mathbf{b}_d are the associated bias vectors.

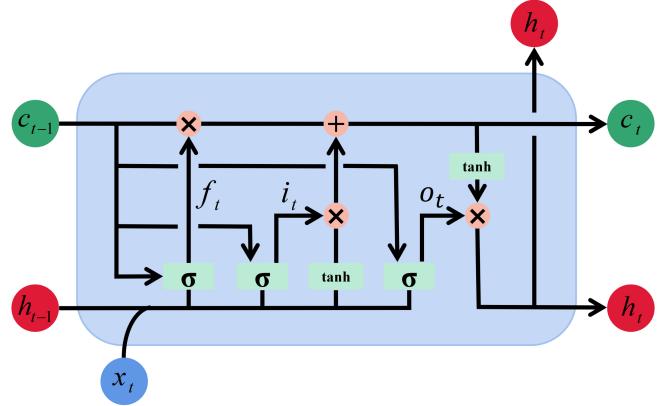


Fig. 12. Long short-term memory (LSTM) block.

B. LSTM Network

The LSTM network is a specific subcategory of recurrent neural networks (RNNs) with memory blocks [52]. The internal structure of each memory block (see Fig. 12) consists of a memory unit and three control gates (the input, forget, and output gates). The three control gates update the information flow through the memory cells, which can be summarized as follows [53]:

$$i_t = \sigma(\mathbf{W}_{xi} \mathbf{x}_t + \mathbf{W}_{hi} \mathbf{h}_{t-1} + \mathbf{w}_{ci} \circ \mathbf{c}_{t-1} + \mathbf{b}_i) \quad (\text{B1})$$

$$f_t = \sigma(\mathbf{W}_{xf} \mathbf{x}_t + \mathbf{W}_{hf} \mathbf{h}_{t-1} + \mathbf{w}_{cf} \circ \mathbf{c}_{t-1} + \mathbf{b}_f) \quad (\text{B2})$$

$$o_t = \sigma(\mathbf{W}_{xo} \mathbf{x}_t + \mathbf{W}_{ho} \mathbf{h}_{t-1} + \mathbf{w}_{co} \circ \mathbf{c}_{t-1} + \mathbf{b}_o) \quad (\text{B3})$$

$$\mathbf{c}_t = f_t \circ \mathbf{c}_{t-1} + i_t \tanh(\mathbf{W}_{xc} \mathbf{x}_t + \mathbf{W}_{hc} \mathbf{h}_{t-1} + \mathbf{b}_c) \quad (\text{B4})$$

$$\mathbf{h}_t = o_t \circ \tanh(\mathbf{c}_t) \quad (\text{B5})$$

where \mathbf{c}_t represents the memory unit and i_t , f_t , and o_t stand for the input, forget, and output gates, which determine the information to be stored, discarded, and output in the unit, respectively. \circ denotes element-wise multiplication, and $\sigma(\cdot)$ and $\tanh(\cdot)$ are the element-wise sigmoid and hyperbolic tangent functions, respectively. \mathbf{W}_{xi} , \mathbf{W}_{xf} , \mathbf{W}_{xo} , and \mathbf{W}_{xc} are weighting matrices of the input data, and \mathbf{W}_{hi} , \mathbf{W}_{hf} , \mathbf{W}_{ho} , and \mathbf{W}_{hc} are weighting matrices of the recurrent data. \mathbf{w}_{ci} , \mathbf{w}_{cf} , and \mathbf{w}_{co} are weighting vectors for the memory cells, while \mathbf{b}_i , \mathbf{b}_f , \mathbf{b}_o , and \mathbf{b}_c are the corresponding bias vectors.

The LSTM network can be regarded as consisting of multiple memory blocks connected in series, and these blocks share weights to reduce the parameters to be trained [60]. The unidirectional LSTM network, as shown in Fig. 13, divides the input time series into 24 time steps and transmits them to the corresponding memory blocks. Each block receives the input information from the current input vector \mathbf{x}_t , the previously hidden state \mathbf{h}_{t-1} , and the previous memory cell state \mathbf{c}_{t-1} , then calculates the output \mathbf{h}_t of each time step, and updates the memory cell state \mathbf{c}_t . In the LSTM-autoencoder architecture, the output of the unidirectional LSTM network is the output data at the last time step because the output has already integrated the useful information of the whole time series. In this study, the first layer of the LSTM-autoencoder network

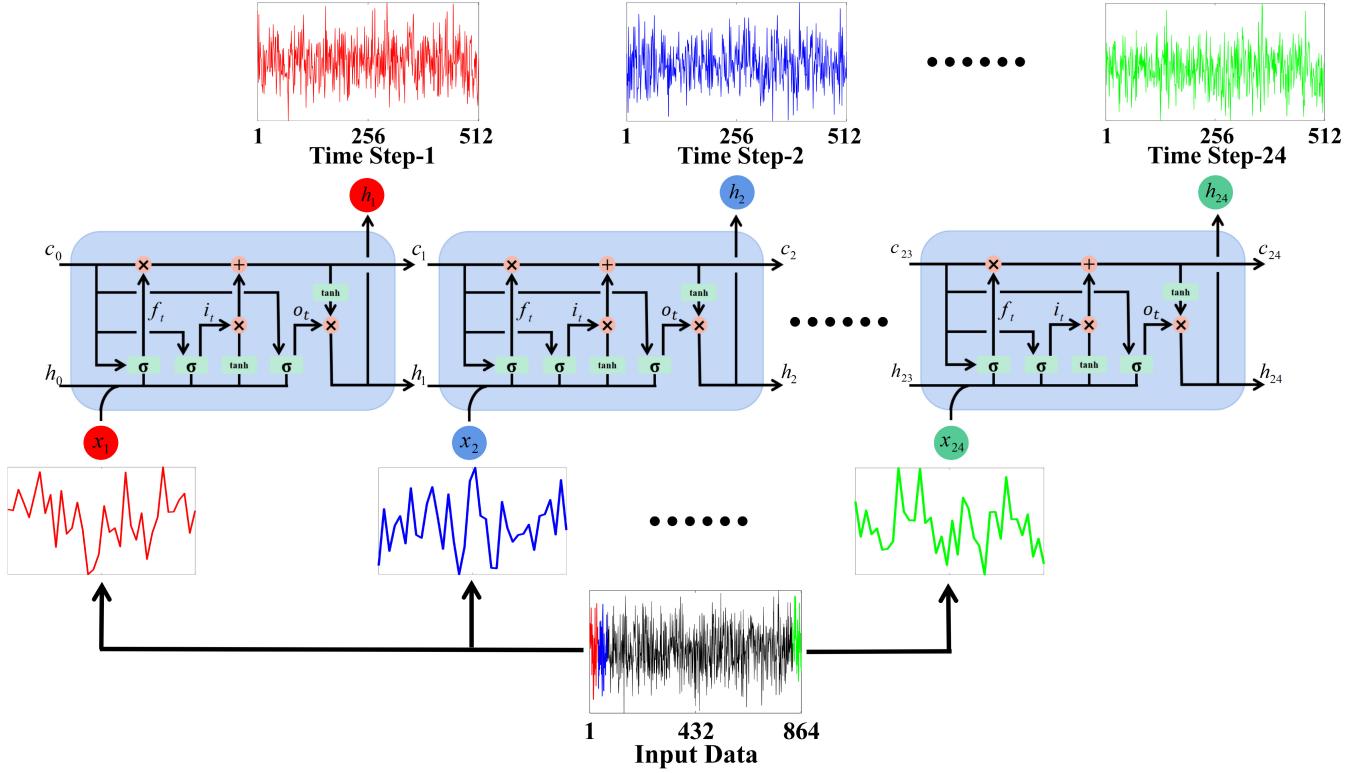


Fig. 13. Example of the unidirectional LSTM network with input and output data dimensions of 864 and 512, respectively. The input time series is divided into 24 time steps, and the dimension of the corresponding output data at each time step is 512. In the LSTM-autoencoder network, the output of the unidirectional LSTM network is only the output result corresponding to the last time step.

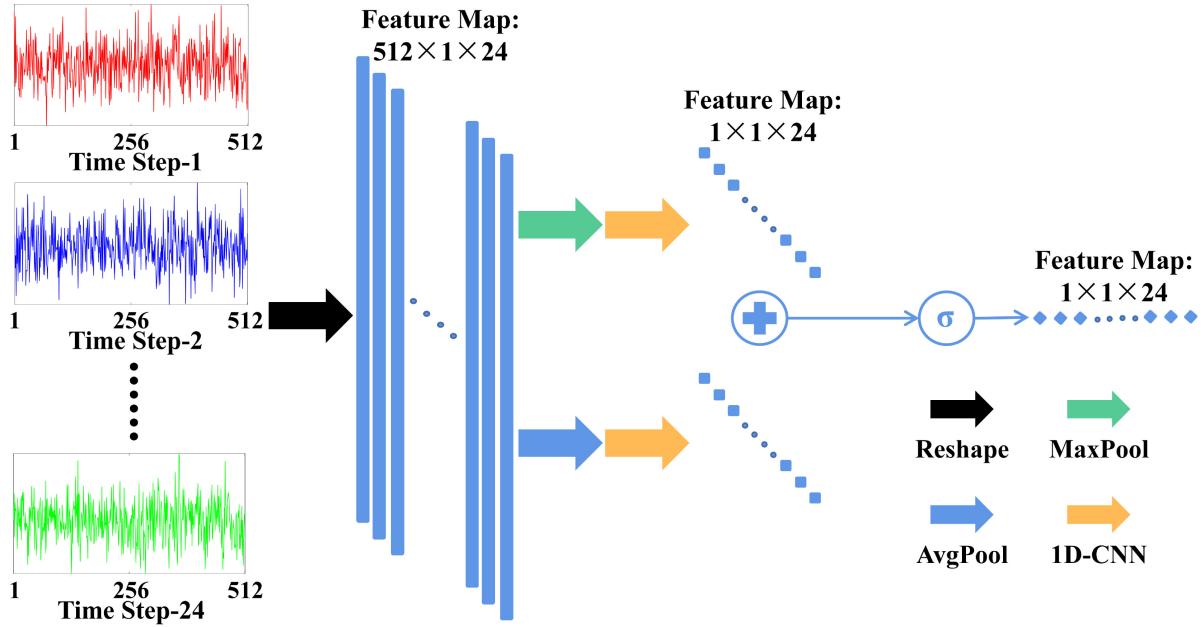


Fig. 14. Example of the time-step attention module with input data dimension $512 \times 1 \times 24$. The input of this module is the output of the bidirectional LSTM network, and the output is the corresponding weight of each time step. The black arrow indicates the reshaping of the input data. The green and blue arrows represent the maximum and average pooling layers, respectively. The orange arrow represents two 1-D convolution layers with the numbers of convolution kernels of 24 and 4, the size and moving step of 1, and the rectified linear unit (ReLU) function. + represents a sum operation, and σ denotes the sigmoid function.

uses a bidirectional LSTM network. In the bidirectional LSTM network, the time series is fed into the unidirectional LSTM

network in positive and negative directions, and then, the two outputs of each time step are averaged. Since the output

of the bidirectional LSTM network shuffles the sequential relationship in time, it is no longer reasonable to select the output at a special time step as the final output of the network. Instead, the time-step attention module is introduced to perform a weighted summation of the network output.

C. Time-Step Attention Module

The time-step attention module is modified from the channel attention module in the CNN [54]. The module modifies the weighted target from each channel in the CNN to each time step in the LSTM network and adaptively adjusts the weight of each time step according to the training set. The time steps contributing significantly to the extraction of SES are given larger weights.

The time-step attention module, as shown in Fig. 14, passes the output of the LSTM network through the maximum and average pooling layers and then through the two 1-D convolution layers. Finally, the outputs are added up, and the corresponding weight of the data at each time step is obtained by the sigmoid function. In short, the time-step attention is computed as

$$\text{Weight}(\mathbf{F}) = \sigma [\text{CNN1}(\text{CNN2}(\mathbf{F}_{\max})) + \text{CNN1}(\text{CNN2}(\mathbf{F}_{\text{avg}}))] \quad (\text{C1})$$

where $\text{Weight}(\cdot)$ is the weight given to each time step. \mathbf{F} represents the output data of the bidirectional LSTM layer. $\sigma(\cdot)$ represents the sigmoid function. CNN1 and CNN2 represent two 1-D convolution layers. \mathbf{F}_{\max} and \mathbf{F}_{avg} are the outputs after passing through the maximum and average pooling layers, respectively. In the LSTM-autoencoder network, the output of the bidirectional LSTM network is the weighted sum of the outputs of all time steps.

ACKNOWLEDGMENT

The authors acknowledge the course “English Presentation for Geophysical Research” of Peking University (Course #01201110) for help in improving the manuscript.

REFERENCES

- [1] S. Uyeda, T. Nagao, and M. Kamogawa, “Short-term earthquake prediction: Current status of seismo-electromagnetics,” *Tectonophysics*, vol. 470, nos. 3–4, pp. 205–213, May 2009.
- [2] L. Conti, P. Picozza, and A. Sotgiu, “A critical review of ground based observations of earthquake precursors,” *Frontiers Earth Sci.*, vol. 9, p. 528, Jul. 2021.
- [3] F. Niu, P. G. Silver, T. M. Daley, X. Cheng, and E. L. Major, “Preseismic velocity changes observed from active source monitoring at the Parkfield SAFOD drill site,” *Nature*, vol. 454, no. 7201, pp. 204–208, Jul. 2008.
- [4] J. Langbein, “Borehole strainmeter measurements spanning the 2014 $M_{W}6.0$ South Napa earthquake, California: The effect from instrument calibration,” *J. Geophys. Res., Solid Earth*, vol. 120, no. 10, pp. 7190–7202, Oct. 2015.
- [5] Q. Huang, “Seismic quiescence before the 2000M= 7.3 Tottori earthquake,” *Geophys. Res. Lett.*, vol. 29, no. 12, pp. 191–194, 2002.
- [6] G. A. Sobolev, Q. Huang, and T. Nagao, “Phases of earthquake’s preparation and by chance test of seismic quiescence anomaly,” *J. Geodynamics*, vol. 33, nos. 4–5, pp. 413–424, May 2002.
- [7] Q. Huang and X. Ding, “Spatiotemporal variations of seismic quiescence prior to the 2011 $M 9.0$ Tohoku earthquake revealed by an improved region-time-length algorithm,” *Bull. Seismol. Soc. Amer.*, vol. 102, no. 4, pp. 1878–1883, Aug. 2012.
- [8] P. A. Varotsos, N. V. Sarlis, E. S. Skordas, S. Uyeda, and M. Kamogawa, “Natural-time analysis of critical phenomena: The case of seismicity,” *Europhys. Lett.*, vol. 92, no. 2, p. 29002, Oct. 2010.
- [9] S.-R.-G. Christopoulos, E. S. Skordas, and N. V. Sarlis, “On the statistical significance of the variability minima of the order parameter of seismicity by means of event coincidence analysis,” *Appl. Sci.*, vol. 10, no. 2, p. 662, Jan. 2020.
- [10] L. F. Khilyuk, J. O. Robertson, B. Endres, and G. Chilingarian, *Gas Migration: Events Preceding Earthquakes*. Houston, TX, USA: Elsevier, 2000.
- [11] S. A. Pulinets, A. D. Legen’ka, T. V. Gaivoronskaya, and V. K. Depuev, “Main phenomenological features of ionospheric precursors of strong earthquakes,” *J. Atmos. Solar-Terrestrial Phys.*, vol. 65, nos. 16–18, pp. 1337–1347, Nov. 2003.
- [12] N. Koizumi, E. Tsukuda, O. Kamigaichi, N. Matsumoto, M. Takahashi, and T. Sato, “Preseismic changes in groundwater level and volumetric strain associated with earthquake swarms off the east coast of the Izu Peninsula, Japan,” *Geophys. Res. Lett.*, vol. 26, no. 23, pp. 3509–3512, Dec. 1999.
- [13] R. K. Chadha, A. P. Pandey, and H. J. Kuempel, “Search for earthquake precursors in well water levels in a localized seismically active area of reservoir triggered earthquakes in India,” *Geophys. Res. Lett.*, vol. 30, no. 7, pp. 61–69, Apr. 2003.
- [14] V. Tramutoli, V. Cuomo, C. Filizzola, N. Pergola, and C. Pietrapertosa, “Assessing the potential of thermal infrared satellite surveys for monitoring seismically active areas: The case of Kocaeli (Izmit) earthquake, August 17, 1999,” *Remote Sens. Environ.*, vol. 96, nos. 3–4, pp. 409–426, 2005.
- [15] P. Varotsos and K. Alexopoulos, “Physical properties of the variations of the electric field of the Earth preceding earthquakes, I,” *Tectonophysics*, vol. 110, nos. 1–2, pp. 73–98, Dec. 1984.
- [16] S. Uyeda, T. Nagao, Y. Orihara, T. Yamaguchi, and I. Takahashi, “Geoelectric potential changes: Possible precursors to earthquakes in Japan,” *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 9, pp. 4561–4566, Apr. 2000.
- [17] Q. Huang, H. Ren, D. Zhang, and Y. J. Chen, “Medium effect on the characteristics of the coupled seismic and electromagnetic signals,” *Proc. Jpn. Acad., B*, vol. 91, no. 1, pp. 17–24, 2015.
- [18] Y. Rong, Q. Wang, X. Ding, and Q. Huang, “Non-uniform scaling behavior in ultra-low-frequency (ULF) geomagnetic signals possibly associated with the 2011 $M_9.0$ Tohoku earthquake,” *Chin. J. Geophys.*, vol. 55, no. 11, pp. 3709–3717, 2012.
- [19] G. Xu, P. Han, Q. Huang, K. Hattori, F. Febriani, and H. Yamaguchi, “Anomalous behaviors of geomagnetic diurnal variations prior to the 2011 off the Pacific coast of Tohoku earthquake ($Mw9.0$),” *J. Asian Earth Sci.*, vol. 77, pp. 59–65, Nov. 2013.
- [20] M. Hayakawa, R. Kawate, O. A. Molchanov, and K. Yumoto, “Results of ultra-low-frequency magnetic field measurements during the Guam earthquake of 8 August 1993,” *Geophys. Res. Lett.*, vol. 23, no. 3, pp. 241–244, Feb. 1996.
- [21] V. M. Sorokin, V. M. Chmyrev, and M. Hayakawa, “A review on electrodynamic influence of atmospheric processes to the ionosphere,” *Open J. Earthq. Res.*, vol. 9, no. 2, pp. 113–141, 2020.
- [22] M. Hayakawa and Y. Hobara, “Current status of seismo-electromagnetics for short-term earthquake prediction,” *Geomatics, Natural Hazards Risk*, vol. 1, no. 2, pp. 115–155, Jun. 2010.
- [23] P. A. Varotsos, “What happened before the last five strong earthquakes in Greece: Facts and open questions,” *Proc. Jpn. Acad., B*, vol. 82, no. 2, pp. 86–91, 2006.
- [24] Y. Orihara, M. Kamogawa, T. Nagao, and S. Uyeda, “Independent component analysis of geoelectric field data in the northern Nagano, Japan,” *Proc. Jpn. Acad., B*, vol. 85, no. 9, pp. 435–442, 2009.
- [25] J. Xi et al., “Dynamic changes of second sampling geoelectric field observation data before and after Qinghai Madoi $M_57.4$ earthquake,” *Chin. J. Geophys.*, vol. 65, no. 2, pp. 580–593, 2022.
- [26] N. V. Sarlis et al., “Seismic electric signals in seismic prone areas,” *Earthq. Sci.*, vol. 31, pp. 44–51, Apr. 2018.
- [27] P. A. Varotsos, *The Physics of Seismic Electric Signals*. Tokyo, Japan: TerraPub, 2005.
- [28] S. Uyeda and K. Meguro, “Earthquake prediction, seismic hazard and vulnerability,” *Geophys. Monogr.*, vol. 150, pp. 349–358, Jan. 2004.
- [29] K. Wang, Q. Huang, and S. Wu, “Application of long short-term memory neural network in geoelectric field data processing,” *Chin. J. Geophys.*, vol. 63, no. 8, pp. 3015–3024, 2020.
- [30] E. S. Skordas, N. V. Sarlis, and P. A. Varotsos, “Effect of significant data loss on identifying electric signals that precede rupture estimated by detrended fluctuation analysis in natural time,” *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 20, no. 3, Sep. 2010, Art. no. 033111.

- [31] P. A. Varotsos, N. V. Sarlis, and E. S. Skordas, "Identifying long-range correlated signals upon significant periodic data loss," *Tectonophysics*, vol. 503, nos. 3–4, pp. 189–194, Apr. 2011.
- [32] T. Yamaguchi *et al.*, "A regionally observed simultaneous pre-seismic geoelectric potential change in Nagano prefecture, Japan," *Bull. Inst. Ocean. Res. Develop., Tokai Univ.*, vol. 21, pp. 11–22, Mar. 2000.
- [33] M. Koganezawa, S. Sawa, H. Shouno, T. Nagao, and K. Joe, "An effective evaluation function for ICA to separate train noise from telluric current data," in *Proc. 4th Int. Symp. Independ. Compon. Anal. Blind Signal Separat. (ICA)*, Nara, Japan, 2003, pp. 837–842.
- [34] C. Ishikawa, C. Watanabe, T. Nagao, and K. Joe, "Extracting seismic electronic signals from the telluric current data for the Niigata Chuetsu earthquake by ICA," in *Proc. Int. Conf. Parallel Distrib. Process. Tech. Appl.*, Las Vegas, NV, USA, 2005, pp. 488–494.
- [35] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492, Jul. 1997.
- [36] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geosci. Frontiers*, vol. 7, no. 1, pp. 3–10, Jan. 2016.
- [37] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.
- [38] X. Wu, L. Liang, Y. Shi, and S. Fomel, "FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation," *Geophysics*, vol. 84, no. 3, pp. IM35–IM45, May 2019.
- [39] W. Zhu and G. C. Beroza, "PhaseNet: A deep-neural-network-based seismic arrival-time picking method," *Geophys. J. Int.*, vol. 216, no. 1, pp. 261–273, 2019.
- [40] R. M. H. Dokht, H. Kao, R. Visser, and B. Smith, "Seismic event and phase detection using time-frequency representation and convolutional neural networks," *Seismol. Res. Lett.*, vol. 90, no. 2A, pp. 481–490, Mar. 2019.
- [41] S. Kanarachos, S.-R. G. Christopoulos, A. Chroneos, and M. E. Fitzpatrick, "Detecting anomalies in time series data via a deep learning algorithm combining wavelets, neural networks and Hilbert transform," *Expert Syst. Appl.*, vol. 85, pp. 292–304, Nov. 2017.
- [42] X. Wu, G. Xue, Y. He, and J. Xue, "Removal of multisource noise in airborne electromagnetic data based on deep learning," *Geophysics*, vol. 85, no. 6, pp. B207–B222, Nov. 2020.
- [43] S. Wu, Q. Huang, and L. Zhao, "De-noising of transient electromagnetic data based on the long short-term memory-autoencoder," *Geophys. J. Int.*, vol. 224, no. 1, pp. 669–681, Nov. 2020.
- [44] P. Han *et al.*, "Assessing the potential earthquake precursory information in ULF magnetic data recorded in Kanto, Japan during 2000–2010: Distance and magnitude dependences," *Entropy*, vol. 22, no. 8, p. 859, Aug. 2020.
- [45] K. Hattori, "ULF geomagnetic changes associated with large earthquakes," *Terr. Atmos. Ocean. Sci.*, vol. 15, no. 3, pp. 329–360, 2004.
- [46] Y. Zhou *et al.*, "Application of reference-based blind source separation method in the reduction of near-field noise of geomagnetic measurements," *Chin. J. Geophys.*, vol. 62, no. 2, pp. 572–586, 2019.
- [47] Q. Huang and Y. Lin, "Selectivity of seismic electric signal (SES) of the 2000 Izu earthquake swarm: A 3D FEM numerical simulation model," *Proc. Jpn. Acad. B*, vol. 86, no. 3, pp. 257–264, 2010.
- [48] F. Jiang, X. Chen, C.-C. Chen, and H.-J. Chen, "Relationship between seismic electric signals and tectonics derived from dense geoelectric observations in Taiwan," *Pure Appl. Geophys.*, vol. 177, no. 1, pp. 441–454, Jan. 2020.
- [49] G. Hinton, N. Srivastava, and K. Swersky, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent," *Cited*, vol. 14, no. 8, p. 2, 2012.
- [50] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986.
- [51] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn. (ICML)*, New York, NY, USA, 2008, pp. 1096–1103.
- [52] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [53] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000.
- [54] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 3–19.
- [55] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," 2016, *arXiv:1607.06450*.
- [56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [57] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2016.
- [58] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, Dec. 2010.
- [59] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*.
- [60] Y. LeCun, "Generalization and network design strategies," in *Connectionism in Perspective*. Amsterdam, The Netherlands: Elsevier, 1989, pp. 143–155.



Jiyian Xue received the B.S. degree in geophysics from the China University of Geosciences, Wuhan, China, in 2020. He is currently pursuing the Ph.D. degree with the Department of Geophysics, Peking University, Beijing, China, under the supervision of Prof. Q. Huang.

His research interests include geoelectric field data processing and deep learning.



Qinghua Huang received the Ph.D. degree in physics from Osaka University, Osaka, Japan, in 1999.

He is currently a Professor of geophysics with the School of Earth and Space Sciences, Peking University, Beijing, China. His research interests include the development of new algorithms, methodologies, and models for geoscience applications. He has coauthored more than 120 articles in national and international journals.

Prof. Huang was a recipient of the First Prize for Science and Technology of Chinese Geophysics by the Chinese Geophysical Society (CGS) in 2016. He is the Vice-President of the Seismological Society of China and an Associate Secretary-General of the Chinese Geophysical Society. He served as the managing guest editor for various journals. He is the Deputy Editors-in-Chief of *Chinese Journal of Geophysics* and an Associate Editor of the *Journal of Geophysical Research: Solid Earth*.



Sihong Wu received the B.S. degree in geophysics from the University of Science and Technology of China, Anhui, China, in 2016, and the Ph.D. degree in solid geophysics from Peking University, Beijing, China, in 2021.

She is currently a Post-Doctoral Fellow with the Department of Geophysics, Peking University. Her research interests include airborne electromagnetic data processing, forward modeling and inversion, and deep learning.



Toshiyasu Nagao received the Ph.D. degree in geophysics from The University of Tokyo, Tokyo, Japan, in 1987.

He is currently an Invited Professor with Tokai University, Tokyo, and the University of Shizuoka, Shizuoka, Japan. His research interests are short and intermediate-term earthquake prediction studies, volcano eruption prediction studies especially by using electromagnetic methods, and hazard mitigation. He has coauthored more than 200 articles in national and international journals.

Prof. Nagao is currently the Chairperson of IUGG Inter-Association EMSEV (Electromagnetic Studies of Earthquake and Volcanoes) and the President of the Earthquake Prediction Society of Japan.