Geophysical Journal International

*Geophys. J. Int.* (2018) **212,** 1389–1397
Advance Access publication 2017 November 8
GJI Marine geosciences and applied geophysics

doi: 10.1093/gji/ggx487

# An automatic microseismic or acoustic emission arrival identification scheme with deep recurrent neural networks

Jing Zheng,[1,2] Jiren Lu,[2] Suping Peng[1] and Tianqi Jiang[2]

[1]*State Key Laboratory of Coal Resources and Safe Mining, China University of Mining and Technology (Beijing), Beijing* 100083*, China. E-mail: zhengjing8628@163.com*

[2]*College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing* 100083*, China*

Accepted 2017 November 7. Received 2017 October 30; in original form 2017 January 30

## SUMMARY

The conventional arrival pick-up algorithms cannot avoid the manual modification of the parameters for the simultaneous identification of multiple events under different signal-to-noise ratios (SNRs). Therefore, in order to automatically obtain the arrivals of multiple events with high precision under different SNRs, in this study an algorithm was proposed which had the ability to pick up the arrival of microseismic or acoustic emission events based on deep recurrent neural networks. The arrival identification was performed using two important steps, which included a training phase and a testing phase. The training process was mathematically modelled by deep recurrent neural networks using Long Short-Term Memory architecture. During the testing phase, the learned weights were utilized to identify the arrivals through the microseismic/acoustic emission data sets. The data sets were obtained by rock physics experiments of the acoustic emission. In order to obtain the data sets under different SNRs, this study added random noise to the raw experiments' data sets. The results showed that the outcome of the proposed method was able to attain an above 80 per cent hit-rate at SNR 0 dB, and an approximately 70 per cent hit-rate at SNR −5 dB, with an absolute error in 10 sampling points. These results indicated that the proposed method had high selection precision and robustness.

**Key words:** Time-series analysis; Acoustic properties; Seismic noise; Wave propagation.

## 1 INTRODUCTION

Microseismic or acoustic emission (AE) systems have been frequently used for monitoring the rock mechanics of mining processes, petroleum extraction, stability of geotechnical engineering, and so on (Kendall *et al.* 2011; Vera Rodriguez *et al.* 2012). The locating of microseismic and AE events is the major task taken on by the current monitoring systems (Kiselevitch *et al.* 1991; Chebotareva *et al.* 2008; Kushnir *et al.* 2013; Kushnir *et al.* 2014). Regardless of which method is selected, uncertainties in the *P*- and *S*-waves arrival time selections will induce location errors. The automatic detections of the arrival times of these waves are important in this area of research (Sabbione & Velis 2013). Meanwhile, the reliability of the tomographic velocity models is dependent on the accuracy of the arrival time selections (Diehl *et al.* 2009).

Currently, the most widely used method is based on the short-long time window average energy ratio (STA/LTA), which was proposed by Stevenson (1976). This method is a detection algorithm which was developed based on the differences in the signal and noise energy. Although the STA/LTA method is very effective for data at high SNR, it tends to usually miss or misjudge the effect events for data at low SNR. Meanwhile, there have been difficulties encountered in the selections of the parameters, such as window size and threshold, which are the main weaknesses of the STA/LTA method. In order to improve the performances, the STA/LTA methods are mainly used in conjunction with other methods. For example, Allen (1978) proposed the notion of the use of feature functions to calculate STA/LTA, for the purpose of detecting wave arrival times. Baer & Kradolfer (1987) modified the feature function on the basis of Allen's method, and then proposed a new dynamic threshold method, in which the feature functions could be achieved by using one or more non-linear transformations.

The feature functions which are designed in time-domains are usually based on mathematical statistical methods, such as variance and higher-order statistical characteristics. The kurtosis and skewness are the mainly used in higher-order statistics. When there are no signals in the arrival waves, the changes of these characteristics are known to be steady. Otherwise, significant changes in the arrival times have been observed. Saragiotis *et al.* (2002, 2004), Küperkoch *et al.* (2010), Tselentis *et al.* (2012), Liu *et al.* (2014), Baillard *et al.* (2014) and Li *et al.* (2016) applied higher-order statistics to arrival selections. The feature functions can also be designed in frequency domains. Over the past decade, there have been many time-frequency methods used for the study of seismic and

microseismic signals. These methods have included the following: Hilbert-Huang transform (Wang *et al.* 2012); S transform (Stockwell, 2007; Zheng *et al.* 2013); $\tau - p$ transform (Forghani-Arani *et al.* 2013); apex-shifted parabolic radon transform (Hargreaves *et al.* 2003; Sabbione *et al.* 2015); fractional wavelet transform (Zheng *et al.* 2015), and so on. In recent years, Karamzadeh *et al.* (2013), Bogiatzis & Ishii (2015) and Mousavi *et al.* (2016) have used wavelet transforms to select the arrival times from seismic records.

In addition, model-oriented algorithms have also been proliferating. In recent years, autoregressive (AR) models, along with artificial neural network (ANN) models, have been widely used. The Akaike information statistics can measure the shortcomings of an estimation model (Akaike 1971). Sleeman & Eck (1999) calculated an Akaike information statistic by constructing an AR model (AR-AIC). In their study, they divided a seismic record into two models: a noise model, and a signal model. The boundary between the two models was the minimum AIC point, which was also the point of arrival. Sedlak *et al.* (2013) automatically applied an AR-AIC in the selections of arrival times for AE data sets.

ANN has a strong 'fault-tolerant' and nonlinear fitting ability. Mc Cormack *et al.* (1993) proposed the detection of seismic events based on ANN. Gentili & Michelini (2006) used the kurtosis and skewness characteristics of seismic data as the input to a neural network model, in order to train the model to perform the arrival time selections. Kaur *et al.* (2013) used the neural network to select *P*-wave arrival times. Maity *et al.* (2014) proposed a new neural network model based on the general neural networks to automatically select microseismic signal arrival times. The main differences of the aforementioned methods were that the input data types (such as AIC value, kurtosis, skewness, and so on) of the neural networks, as well as the dimensions of the input data, were different. In theory, the higher the complexity of the parameters for a model was, the greater the 'capacity' of the model was. This suggests that such a model has the ability to learn more complex features in the data. However, in reality, the training of a model with very high complexity is very inefficient, and can easily result in over-fitting. An ANN model which is composed of many hidden layers is difficult to realize. Also, a general ANN model can only learn the simple features of data sets. With the arrival of the era of large data, and the continuous improvements in computer hardware, training efficiency has been significantly improved. Large data play important roles in solving the problems of over-fitting. Therefore, complex neural network models, such as 'Deep Learning', have attracted increasing attention. Deep learning allows computational models to learn representations of data through multiple levels of abstraction. These models are composed of multiple processing layers, in which the input of each layer is the output of the previous layer. In the majority of cases, a back-propagation algorithm is used to indicate the way in which the internal parameters are to be changed. Hinton & Salakhutdinov (2006) used an unsupervised layer-by-layer training approach to perform nonlinear dimensionality reductions of large-dimension data based on the Deep Belief Neural Networks, which allowed the network model to be fully trained. Graves (2012) used a sequence labelling method to supervise the training data with a deep recurrent neural network (RNN). Also, Girshick *et al.* (2014) obtained good training optimization based on the Hebbian rule and multi-scale processing, in order to optimize a 22-layer convolution network model. He *et al.* (2016) established a 152-layer depth residual neural network, in which the recognition performance was found to be much higher than the traditional artificial feature extraction algorithms.

The automatic selection and recognition of the arrival times of microseismic and AE events is significant to the realization of the automatic processing of massive microseismic and acoustic data. However, due to the variabilities of the stress waveforms, differences in the trigger source phases of the rupture sources, and the presence of various noise interferences, the research studies regarding automatic recognition and selection methods remain challenging. In the current study, a method of feature extraction was proposed, which was based on sequence labelling with RNNs. The proposed method was tested using AE waveform data from the rock mechanic experiments of coal samples. The results showed that the proposed method had the ability to clearly show the arrival features under different SNR data. It was found that, even when the SNR was −5 dB, the precision did not sharply decrease, and a result of approximately 70 per cent could be attained, with an absolute error in 10 sampling points. Therefore, the results confirmed that this method had good robustness in the selections of arrival times.

## 2 RECURRENT NEURAL NETWORKS

The traditional arrival selection algorithms are greatly influenced by the SNR. In order to ensure the accuracy of the selections, it is required to manually adjust the parameters. In this study, an arrival selection method was proposed based on a deep learning neural network model, for the purpose of solving the problems of parameter adjustments at different SNR using the powerful 'fault tolerance' of the model. In this section, RNNs with Long Short-Term Memory (LSTM) as the hidden layers were first introduced. Then, the training of these networks was discussed.

### 2.1 The theory of recurrent neural networks

As shown in Fig. 1, for the standard RNNs, not only the neurons in the layers were fully connected, but also the neurons in the layers were connected to each other. By assuming that the network model inputs a microseismic record vector $x = (x_1, x_2, ..., x_T)$, a standard RNN computes the hidden vector sequence $h = (h_1, h_2, ..., h_T)$, and the output vector sequence $o = (o_1, o_2, ..., o_T)$, by iterating the following equations from $t = 1$ to $T$:

$$h_t = H(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (1)$$

$$o_t = W_{ho}h_t + b_o, \quad (2)$$

where $W$ denotes the connection weight matrix (for example, $W_{xh}$ represents the connection weight matrix of the input neuron and

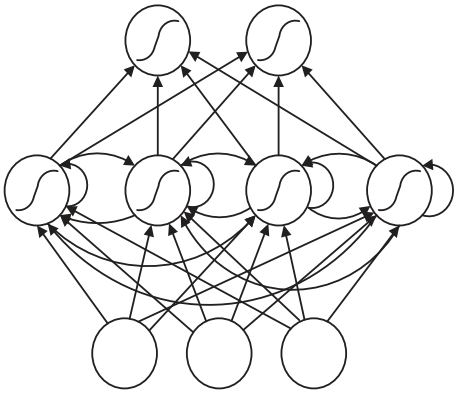

**Figure 1.** Recurrent neural networks.

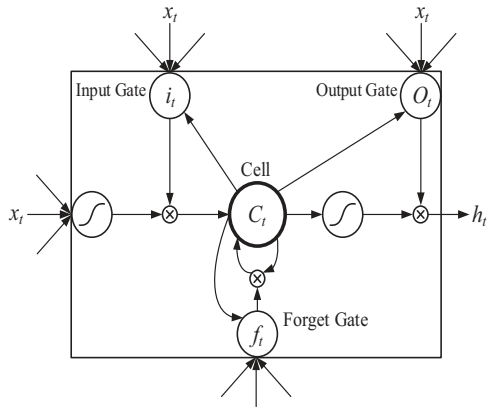


**Figure 2.** Long Short-Term Memory cell.

hidden neuron nodes); $b$ denotes the bias vectors (for example, $b_o$ represents the output bias vector); and $H$ is the hidden layer function. Also, $H$ is usually a logistic function: $H(x) = 1/(1 + e^{-x})$.

Deep RNNs can be created by stacking the multiple RNNs' hidden layers on top of each other, with the output sequence of one layer forming the input sequence for the next. Then, by assuming that the activation functions $H$ of all the hidden layers are the same, the output sequence $h^n$ of the hidden layer can be obtained from the iterated $n = 1$ to $N$, and $t = 1$ to $t = T$ using the following equation:

$$h_t^n = H(W_{h^{n-1}h^n} h_t^{n-1} + W_{h^n h^n} h_{t-1}^n + b_h^n), \tag{3}$$

where $h^0 = x$ is set. Then, the output sequence of the network model can be obtained by the following equation:

$$o_t = W_{ho} h_t^N + b_o, \tag{4}$$

where $N$ is the total number of hidden layers in the network model; $W_{hy}$ represents the connection weight matrix of last hidden neuron and output neuron nodes; and $h^N$ represents the last hidden vector sequence.

Theoretically, the traditional RNNs can handle any length of the sequence. However, in practice, the RNNs can only save the first few moments of the current moment of influence. Therefore, the time after the impact of the previous will be weaker, due to the vanishing gradient problem of the RNNs (Bengio *et al.* 1994). Hochreiter & Schmidhuber (1997) proposed a long short-term memory model in which information can be stored through the memory cells. A complete AE event takes a long period of time in a sequence. Also, the amplitude of an AE event has a basic feature, in which the amplitude gradually becomes larger, and finally becomes gradually smaller, from the beginning of the arrival. In other words, AE events are linked in time. However, the output of traditional RNNs at a given time can only affect the several moments which are adjacent to it. Therefore, a complete sequence of AE events cannot be efficiently extracted in time. The multiplication gate of an LSTM allows cells to store and receive longer information prior to an event, which not only alleviates the problem of gradient dispersion, but also allows for the more efficient extraction of the characteristics of a time series. As shown in Fig. 2, considering that an LSTM is the hidden layers of the RNNs (Graves 2012), and the hidden layer function $H$ of the RNNs, it can be defined by the following functions:

$$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i) \tag{5}$$

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + W_{cf} c_{t-1} + b_f) \tag{6}$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \tag{7}$$

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + W_{co} c_t + b_o) \tag{8}$$

$$h_t = o_t \tanh(c_t), \tag{9}$$

where $\sigma$ represents the logistic function: $\sigma(x) = 1/(1 + e^{-x})$; and $i$, $f$, $o$ and $c$ are the input gate, forget gate, output gate and cell activation vectors, respectively, all of which are the same size as the hidden vector $h$; and tanh() represents a hyperbolic tangent function. The weight matrices from the cell to gate vectors ($W_{ci}$) are diagonal. Therefore, element $m$ in each gate vector only receives input from element $m$ of the cell vector.

## 2.2 Model learning

This study used the method of sequence labelling to mark the raw microseismic/AE signals. As shown in Fig. 3, two events occurred in a record at the same time, and the sequence labelling was used to mark the arrival point as 1, while the others were labelled as zero. It should be noted that the sequence labelling was marked as 1, which included the arrival point and duration of the signal, and the remainder were marked as 0. It was found that the labelling could be designed according to the needs. In this study, $y$ was defined as the target sequence, and the microseismic/AE signal was defined as the input sequence.

Since the tag sequence in this study only contained 'zero' and 'non-zero', a logistical regression function, or soft-max function, could be used in the model. Then, by assuming that a logistical regression model was used, the hypothesis function for the model was as follows:

$$\hat{y}_i = \frac{1}{1 + \exp\left(-(W_{oL})^T o_t\right)} \tag{10}$$

where $W$ is the connection weight matrix between the output layer and the logical regression layer. Then, it was assumed in this study that the target sequence of the model was $y = (y_1, y_2, ..., y_T)$ ,$y_i \in \{0, 1\}$, and $L$ was used to measure the distance between $\hat{y}$ and $y$. An RNN can be used to map an input sequence to an output sequence of the same length. Therefore, the total loss associated with the sequence will be the sum of the losses of all of the time steps. In this study, the loss was defined as:

$$L = -\frac{1}{T}\left[\sum_{i=1}^{T} y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i)\right], \tag{11}$$

where $T$ is the total length of the sequence. It was determined that the loss function for the soft-max was similar to the logistics. Next, the gradient of each network model parameter was obtained using a back-propagation through time (BPTT) algorithm. The derivation of the BPTT is detailed in Appendix A.

Since the parameters of an RNN are the sharing mechanism, the network model parameters were updated by a gradient descent as follows:

$$\hat{\theta} = \theta - \nabla_\theta L \tag{12}$$

where $\hat{\theta}$ represents the updated parameter; $\theta$ represents the parameter before updating; and $\nabla_\theta L$ represents the gradient of the loss $L$ to the parameter $\theta$. The parameters included $b_o$, $b_h$, $W_{ho}$, $W_{oL}$ and $W_{xh}$.

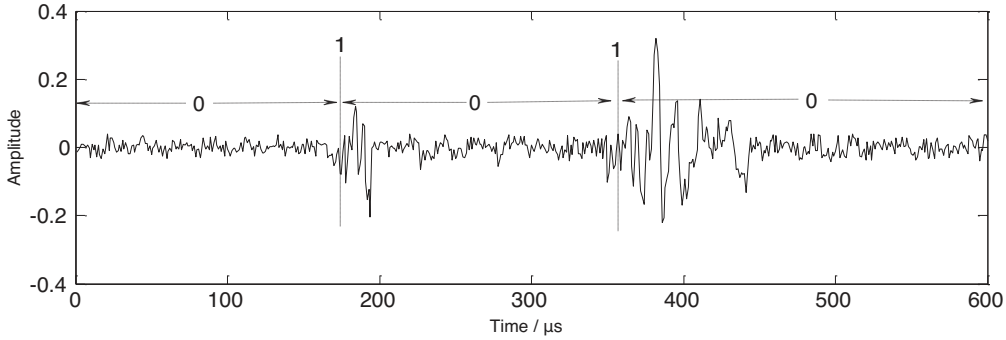


**Figure 3.** Sequence labelling.

## 3 EXPERIMENTAL PROCESS AND ANALYSIS

In this experiment, an RTR high-temperature and high-pressure rock comprehensive test system (GCTS, Geotechnical Consulting and Testing Systems, USA) was used to finish the AE experiment at room temperature. Due to the fact that further ultrasonic testing was required in order to have a regular geometrical profile of the coal sample, the petro physical test standard sample size was used for the coal sample processing. The diameter of the specimen was approximately 50 mm (actual size: 50.04 mm), and the height was approximately 100 mm (actual size: 98.14 mm). The confining pressure was 2 MPa, and the axial compression was increased until the coal produced significant rupture.

### 3.1 Training data

The training iteration time of the RNNs' network model was proportional to the length of the input data. The longer the data length was, the longer the iteration time required. The data length was set as 1024 in order to quickly train the model, and also include a complete AE event in a data sequence. In this study, through the experimental process, a total of 163 248 AE records were collected. Each record included data with lengths of 1024.

In order to obtain the distribution of the total data set, all of the sequences needed to be manually selected, which proved to be undoubtedly very inefficient. Then, for the purpose of improving the experiment's efficiency, this study randomly extracted 10 000 sequences from the 163 248 sequence for wavelet collection. It was assumed that 10 000 sequences roughly reflected the distribution of the overall sequences. For the solution to the problem of data imbalance, different wavelets generated by the AE events were randomly chosen from the 10 000 sequences, and new data sets were generated by random translating. After these steps were completed, one million sequences had been obtained. The corresponding label for each sequence was a binary vector of the same length as the sequences. Then, noise was randomly added to each of the one million sequences , and the SNR of the data was controlled between −5 and 5 dB.

The one million sequences were divided into training data sets and test data sets, and 80 per cent were used as training data sets for training the RNN models, while 20 per cent were used for testing the performances of the RNN models. This study also tried different proportions of the training sets and test set assignments. The results showed that the larger the training data set was, the better the test results would be. However, a small percentage of the test set was found to be not representative. As shown in Fig. 4, the horizontal axis was the ratio of the training data set to the total data set, and the vertical axis was the accuracy rates of the selections. The selection results were considered to be accurate if the differences between the arrivals and the artificial selection results were within ten sampling points.

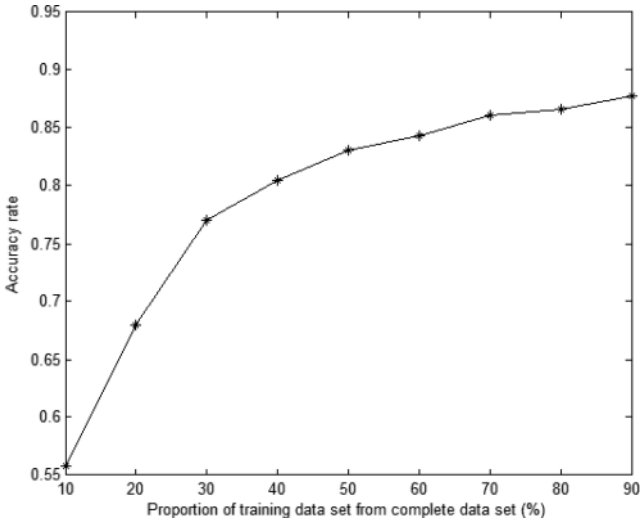


**Figure 4.** The relationship between proportion of training data set and picking accuracy.

### 3.2 Training of the RNNs models

In this study's experiment, seven LSTM layers were stacked on top of each other, which formed a model which was capable of learning high-level temporal representations. The length of the input sequence, number of hidden layer nodes and length of the output sequence were all 1024. The optimization was a stochastic gradient descent (Srivastava *et al.* 2014), and the dropout rate was 0.5.

The output of the RNN model was obtained after the model training was completed. Then, in order to verify the performance of the proposed method, an STA/LTA method was chosen for comparison purposes. The basic principle of an STA/LTA is as follows: the changes in the amplitude of waveform data are reflected by the ratio between the average of the energy in the short-time window, and the average of the energy in the long-time window. The value of the STA is much larger than that of the LTA when an arrival is coming. When the ratio is greater than a pre-set threshold, it can be assumed that the point is the arrival. Then, by choosing two sequences, the output features of the different selection methods were obtained, as shown in Fig. 5. Fig. 5(a) contains only one event, and Fig. 5(b) contains multiple events.

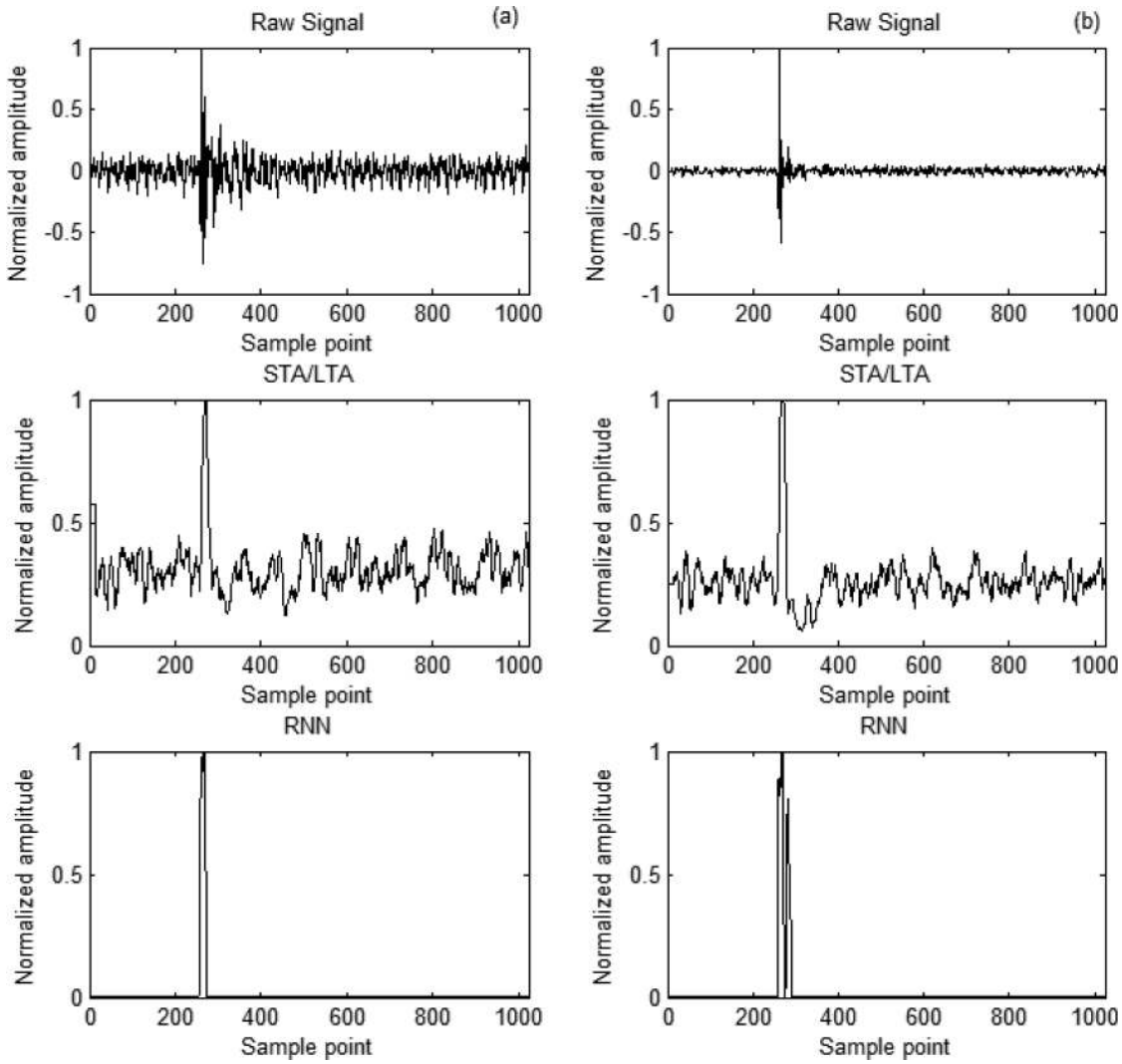


**Figure 5.** The arrival feature curve of different pick-up methods. (a) The different feature curves of single event; (b) the different feature curves of multiple events.

In this study, by using an STA/LTA feature function, the obtained feature curves were found to effectively highlight the arrival. Also, even in the multievent selection, an improved highlight of the characteristics of the arrival was achieved. However, when multiple events occurred with large amplitude differences and small intervals in the waveform record, the events with smaller amplitudes and following occurrences were correspondingly smaller or even undetectable in the feature curve, as detailed in Fig. 5(b). Therefore, in order to show the results more clearly, these events are detailed in Figs 5(b) and 6. The curve on the bottom is the output of the RNNs. The curve highlights the characteristics of the arrival, and the characteristics of the arrivals were observed to have better anti-interference abilities.

### 3.3 Experimental comparison

In the following section, traditional arrival selection methods (STA/LTA), along with this study's proposed method, were used to obtain the arrivals with different SNR data. Six of the AE sequences which were not included in the training and test data set mentioned above were selected from different time periods, in order to verify the performances of the RNNs models. The six AE sequences are shown in Fig. 7. The arrivals were both manually selected, and selected by auto-pickers, as shown in Table 1. The numbers 1–3 in Table 1 correspond to the three events detailed in Figs 7(a)–(f), and the selections are given in the samples.

In this study, it was assumed that the six AE sequences were noise-free signals, and Gaussian white noise was added using Matlab mathematical processing. After the noise was added, the SNR of the new sequences was controlled the levels of 10, 5, 0, and −5 dB. It should be noted that the signals were not noise-free, and the SNR was lower. The data for each SNR level were obtained by randomly adding the noise of the corresponding intensity to the original data.

In order to make this stochastic process statistically significant, the data of each signal-to-noise ratio were obtained by randomly adding the noise to the original data 100 times using a Monte Carlo method. The 100 sequences were picked up by the STA/LTA method (with the short time window, long window, and threshold of the STA/LTA method set in advance), and RNNs method (with the RNNs model pre-trained as described in Section 2). Then, the AE sequences outside the training data set were input into the trained

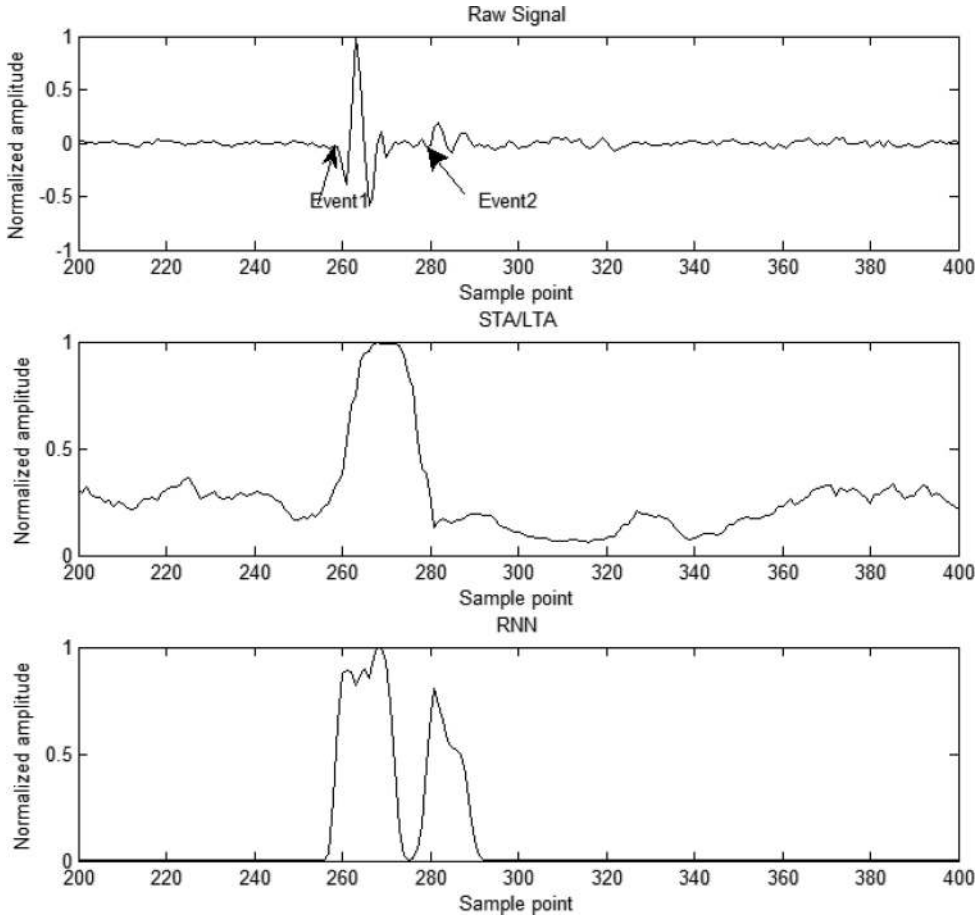


**Figure 6.** The details for Fig. 5(b).

model. The model output the corresponding feature of the sequence, and finally obtained the arrival from the feature curve by setting the threshold and minimum intervals between different events. These 100 sequences were selected by the methods, and the results of the selection and artificial picking were compared.

In this study, in order to evaluate the performance of the proposed method, the hit-rate and false-rate of the proposed method were compared with the STA/LTA method. The hit-rate and false-rate were calculated using the following equations:

$$\begin{aligned} \text{Hit} - \text{rate} &= N/M \\ \text{False} - \text{rate} &= (K - N)\big/M, \end{aligned} \tag{13}$$

where $N$ represents the number which had been correctly selected by the auto-pickers, $M$ is the total number of events, and $K$ is the total number which had been selected by autopickers. It should be noted that the events recognized by the autopickers were treated as being correct only if the differences between the picked-arrivals and manual results were no larger than 10 sample points. The hit-rate and false-rate are recorded in Tables 2 and 3.

As shown in Table 2, with the decrease of the SNR, the hit-rates of both methods were decreased. However, the proposed method was able to achieve an accuracy of approximately 70 per cent, even when the SNR was −5 dB. The hit-rate was observed to be higher than the STA/LTA at all of the SNRs, and was more robust.

## 4 DISCUSSION AND CONCLUSIONS

The results of the experiments detailed in Section 3.2 showed that the characteristics of the arrivals using the RNNs had better anti-interference abilities. However, the RNNs' model training required a large number of data sequences with labels, which meant that it required a certain amount of time to select the artificial waves in order to obtain accurate arrival labels. It was found that the most important point in identifying a series of waveforms was that, when there was not an AE event, the RNNs model could easily misjudge the two points with large amplitude differences. In order to solve this problem, this study modified the label vector to give duration to the AEs. In the study, the duration was set as 13. For example, if the original only marked the arrival point 210 as 1, this was now extended to the point within the range of 204 to 216, which was equivalent to the addition of some further priori information.

The results of the experiments detailed in Section 3.3 showed that the hit-rate of the RNNs was higher than those of the STA/LTA method. However, the false-rate was also higher than the STA/LTA, especially at a low SNR. This was determined to be due to the fact that the data sets were not large enough. Also, the wavelets which were required to generate the training data were chosen manually, which resulted in the wavelets being incomplete. All of the aforementioned reasons caused the model to overfit. In order to solve these problems, this study required more experience, along with the collection a larger amount of data, and more in-depth training of the models. However, through this study's analysis and testing

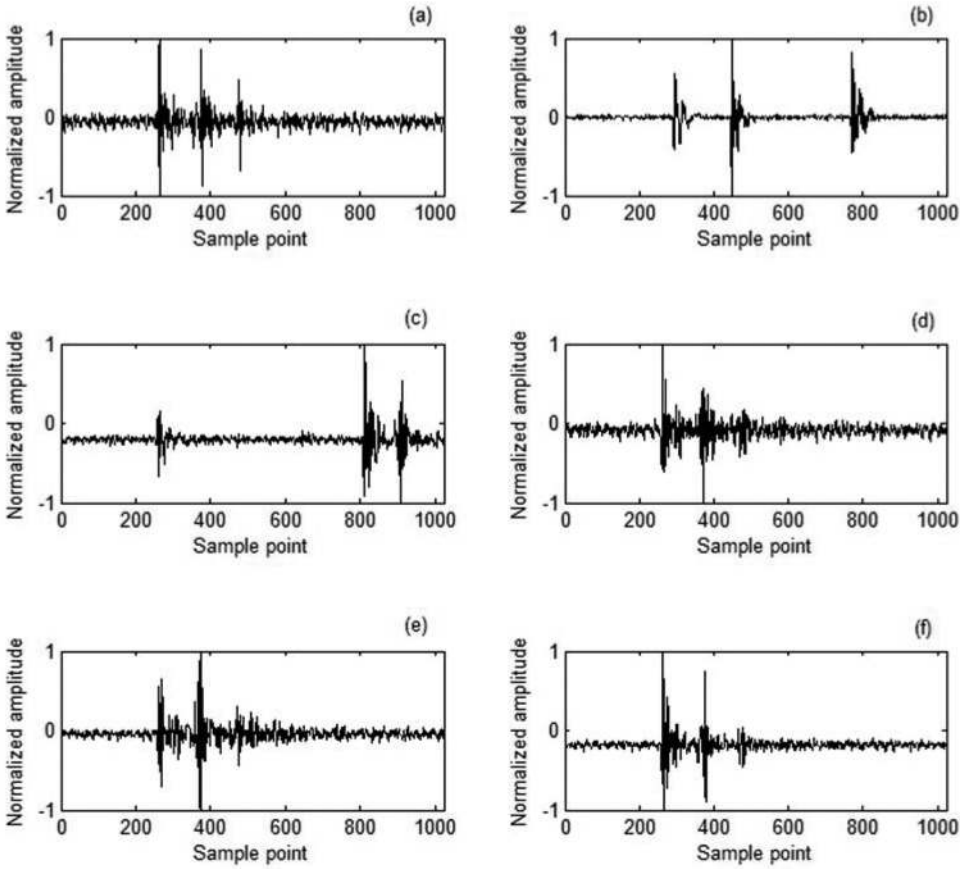


**Figure 7.** Testing sequences. (a) Test sequence 1; (b) test sequence 2; (c) test sequence 3; (d) test sequence 4; (e) test sequence 5; (f) test sequence 6.

**Table 1.** Arrivals picked by different methods.

| | | Artificial Arrival point | | | STA/LTA Arrival point | | | The proposed method Arrival point | | |
|---|---|---|---|---|---|---|---|---|---|---|
| NO. | SNR (dB) | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| 1 | raw | 260 | 370 | 474 | 257 | 369 | 472 | 260 | 370 | 473 |
| 2 | raw | 290 | 445 | 766 | 286 | 441 | 763 | 288 | 444 | 767 |
| 3 | raw | 257 | 808 | 904 | 254 | 804 | 899 | 257 | 808 | 905 |
| 4 | raw | 254 | 360 | 470 | 255 | 363 | 474 | 258 | 362 | 476 |
| 5 | raw | 256 | 363 | 470 | 255 | 362 | 474 | 258 | 364 | 469 |
| 6 | raw | 259 | 367 | 478 | 255 | 363 | 470 | 261 | 369 | 479 |

**Table 2.** Hit-rate and False-rate evaluated by different methods.

| | | Hit-Rate | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Method | SNR (dB) | sequence 1 | sequence 2 | sequence 3 | sequence 4 | sequence 5 | sequence 6 | mean |
| STA/LTA | 10 | 1.00 | 1.00 | 0.98 | 0.94 | 0.90 | 0.98 | 0.97 |
| | 5 | 1.00 | 0.99 | 0.99 | 0.78 | 0.80 | 0.80 | 0.89 |
| | 0 | 0.88 | 0.99 | 0.93 | 0.66 | 0.70 | 0.67 | 0.81 |
| | −5 | 0.37 | 0.56 | 0.42 | 0.23 | 0.41 | 0.46 | 0.41 |
| The proposed method | 10 | 1.00 | 0.99 | 1.00 | 0.97 | 1.00 | 0.85 | 0.97 |
| | 5 | 1.00 | 0.93 | 1.00 | 0.90 | 1.00 | 0.79 | 0.94 |
| | 0 | 0.94 | 0.84 | 0.95 | 0.81 | 0.90 | 0.74 | 0.86 |
| | −5 | 0.81 | 0.62 | 0.68 | 0.69 | 0.72 | 0.63 | 0.69 |

experiences, it was found that deep learning methods such as RNNs can be effectively used to recognize the AEs. Therefore, in the future, researchers should continue the study of deep learning methods such as RNNs.

In the current research study, an algorithm to select the arrivals of microseismic and AE events based on deep RNNs was proposed. This novel method contained three major steps for arrival selection as follows: (1) the conversion of the arrival selection tasks into a

**Table 3.** False-rate evaluated by different methods.

| Method | SNR (dB) | False-rate | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | sequence 1 | sequence 2 | sequence 3 | sequence 4 | sequence 5 | sequence 6 | mean |
| STA/LTA | 10 | 0.00 | 0.00 | 0.03 | 0.01 | 0.01 | 0.06 | 0.018 |
| | 5 | 0.00 | 0.01 | 0.01 | 0.00 | 0.00 | 0.01 | 0.005 |
| | 0 | 0.01 | 0.00 | 0.00 | 0.02 | 0.01 | 0.00 | 0.007 |
| | −5 | 0.01 | 0.03 | 0.01 | 0.19 | 0.04 | 0.02 | 0.005 |
| The proposed method | 10 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.003 |
| | 5 | 0.00 | 0.07 | 0.00 | 0.02 | 0.00 | 0.00 | 0.015 |
| | 0 | 0.01 | 0.16 | 0.00 | 0.03 | 0.03 | 0.00 | 0.038 |
| | −5 | 0.03 | 0.28 | 0.13 | 0.05 | 0.10 | 0.09 | 0.113 |

sequence labelling task; (2) the training of the model to update the weight parameters using an error back-propagation through time algorithm. The entire network parameter could then be optimized with a gradient descent, and a serial annotation model could be built using an RNN; (3) finally, the raw waveform signal was input into a trained network model, and then the model output a feature vector with an obvious arrival time.

In this study, experiments regarding AE signals were carried out using both the traditional and proposed methods based on the measured AE data. The results revealed the following: (1) in the conventional method, the STA/LTA had the highest selection accuracy in the raw signal. However, it relied heavily on the manual adjustment of the parameters; (2) the misjudgments and hit-rates of the proposed method were larger than those of the STA/LTA method, which indicated that the precision of the proposed method was higher; (3) the false-rate of the proposed method was larger than that of the STA/LTA method, which indicated that the data set was not complete enough, and that there was a possibility that the model was overfit.

## ACKNOWLEDGEMENTS

This research is financially supported by National Natural Science Foundation of China (Grant No. 41504041), State Key Laboratory of Coal Resources and Safe Mining, China University of Mining & Technology (No. SKLCRSM14KFA05) and the Fundamental Research Funds for the Central Universities. The authors also thank anonymous reviewers and editors for promotion of this paper.

## REFERENCES

Akaike, H., 1971. Autoregressive model fitting for control, *Ann. Inst. Stat. Math.,* **23**(1), 163–180.

Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces, *Bull. seism. Soc. Am.,* **68**(5), 1521–1532.

Baer, M. & Kradolfer, U., 1987. An automatic phase picker for local and teleseismic events, *Bull. seism. Soc. Am.,* **77**(4), 1437–1445.

Baillard, C., Crawford, W.C., Ballu, V., Hibert, C. & Mangeney, A., 2014. An automatic kurtosis-based *P*- and *S*-Phase picker designed for local seismic networks, *Bull. seism. Soc. Am.,* **104**(1), 394–409.

Bengio, Y., Simard, P. & Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult, *IEEE Trans. Neural Netw.,* **5**(2), 157–166.

Bogiatzis, P. & Ishii, M., 2015. Continuous wavelet decomposition algorithms for automatic detection of compressional-and Shear-wave arrival times, *Bull. seism. Soc. Am.,* **105**(3), 1628–1641.

Chebotareva, I.Ya., Kushnir, A.F. & Rozhkov, M.V., 2008. Elimination of high-amplitude noise during passive monitoring of hydrocarbon deposits by the emission tomography method, *Izv. Phys. Solid Earth,* **44**(12), 1002–1017.

Diehl, T., Kissling, E., Husen, S. & Aldersons, F., 2009. Consistent phase picking for regional tomography models: application to the greater Alpine region, *Geophys. J. Int.,* **176**(2), 542–554.

Forghani-Arani, F., Willis, M., Haines, S.S., Batzle, M., Behura, J. & Davidson, M., 2013. An effective noise-suppression technique for surface microseismic data, *Geophysics,* **78**(6), KS85–KS95.

Gentili, S. & Michelini, A., 2006. Automatic picking of P and S phases using a neural tree, *J. Seismol.,* **10**(1), 39–63.

Girshick, R., Donahue, J., Darrell, T. & Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587.

Graves, A., 2012. *Supervised Sequence Labeling with Recurrent Neural Networks,* Springer, pp. 5–45.

Hargreaves, N., verWest, B., Wombell, R. & Trad, D., 2003. Multiple attenuation using an apex-shifted Radon transform, in *2003 SEG Annual Meeting,* Society of Exploration Geophysicists.

He, K., Zhang, X., Ren, S. & Sun, J., 2016. Deep residual learning for image recognition, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.

Hinton, G.E. & Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks, *Science,* **313**(5786), 504–507.

Hochreiter, S. & Schmidhuber, J., 1997. Long short-term memory, *Neural Comput.,* **9**(8), 1735–1780.

Karamzadeh, N., Doloei, G.J. & Reza, A.M., 2013. Automatic earthquake signal onset picking based on the continuous wavelet transform, *IEEE Trans. Geosci. Remote Sens.,* **51**(5), 2666–2674.

Kaur, K., Wadhwa, M. & Park, E.K., 2013. Detection and identification of seismic P-Waves using Artificial Neural Networks, in *The 2013 International Joint Conference on Neural Networks (IJCNN),* IEEE, Dallas, TX, USA, pp. 1–6.

Kendall, M., Maxwell, S., Foulger, G., Eisner, L. & Lawrence, Z., 2011. Microseismicity: beyond dots in a box – introduction, *Geophysics,* **76**(6), WC1–WC3.

Kiselevitch, V.L., Nikolaev, A.V., Troitskiy, P.A. & Shubik, B.M., 1991. Emission tomography: main ideas, results, and prospects, in *61st Annual International Meeting*, *SEG, Expanded Abstracts,* 1602.

Kushnir, A., Rozhkov, N. & Varypaev, A., 2013. Statistically-based approach for monitoring of micro-seismic events, *Int. J. Geomath.,* **4**(2), 201–225.

Kushnir, A., Varypaev, A., Dricker, I., Dricker, I. & Rozhkov, M., 2014. Passive surface microseismic monitoring as a statistical problem: location of weak microseismic signals in the presence of strongly correlated noise, *Geophys. Prospect.,* **62**(4), 819–833.

Küperkoch, L., Meier, T., Lee, J., Friederich, W., & EGELADOS Working Group, 2010. Automated determination of P-phase arrival times at regional and local distances using higher order statistics, *Geophys. J. Int.,* **181**(2), 1159–1170.

Li, X., Shang, X., Wang, Z., Dong, L. & Weng, L., 2016. Identifying *P*-phase arrivals with noise: an improved Kurtosis method based on DWT and STA/LTA, *J. Appl. Geophys.,* **133,** 50–61.

Liu, X.Q., Cai, Y., Zhao, R., Zhao, Y.-G., Qu, B.-A., Feng, Z.-J. & Li, H., 2014. An automatic seismic signal detection method based on fourth-order statistics and applications, *Appl. Geophys.,* **11**(2), 128–138.

Maity, D., Aminzadeh, F. & Karrenbach, M., 2014. Novel hybrid artificial neural network based autopicking workflow for passive seismic data, *Geophys. Prospect.,* **62**(4), 834–847.

Mccormack, M.D., Zaucha, D.E. & Dushek, D.W., 1993. First-break refraction event picking and seismic data trace editing using neural networks, *Geophysics,* **58**(1), 67–78.

Mousavi, S.M., Langston, C.A. & Horton, S.P., 2016. Automatic microseismic denoising and onset detection using the synchrosqueezed continuous wavelet transform, *Geophysics,* **81**(4), V341–V355.

Sabbione, J.I. & Velis, D.R., 2013. A robust method for microseismic event detection based on automatic phase pickers, *J. Appl. Geophys.,* **99**(12), 42–50.

Sabbione, J.I., Sacchi, M.D. & Velis, D.R., 2015. Radon transform-based microseismic event detection and signal-to-noise ratio enhancement, *J. Appl. Geophys.,* **113,** 51–63.

Saragiotis, C.D., Hadjileontiadis, L.J. & Panas, S.M., 2002. PAI-S/K: a robust automatic seismic *P* phase arrival identification scheme, *IEEE Trans. Geosci. Remote Sens.,* **40**(6), 1395–1404.

Saragiotis, C.D., Hadjileontiadis, L.J., Rekanos, I.T. & Panas, S.M., 2004. Automatic P phase picking using maximum kurtosis and kappa-statistics criteria, *IEEE Geosci. Remote Sens. Lett.,* **1**(3), 147–151.

Sedlak, P., Hirose, Y. & Enoki, M., 2013. Acoustic emission localization in thin multi-layer plates using first-arrival determination, *Mech. Syst. Signal Process.,* **36**(2), 636–649.

Sleeman, R. & Eck, T., 1999. Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Phys. Earth planet. Inter.,* **113**(1–4), 265–275.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R., 2014. Dropout: a simple way to preve-nt neural networks from overfitting, *J. Mach. Learn. Res.,* **15**(1), 1929–1958.

Stevenson, P.R., 1976. Microearthquakes at flathead lake, montana: a study using automatic earthquake processing, *Bull. seism. Soc. Am.,* **66**(1), 61–80.

Stockwell, R.G., 2007. A basis for efficient representation of the S-transform, *Digit. Signal Process.,* **17**(1), 371–393.

Tselentis, G.A., Martakis, N., Paraskevopoulos, P., Lois, A. & Sokos, E., 2012. Strategy for automated analysis of passive microseismic data based on S-transform, Otsu's thresholding, and higher order statistics, *Geophysics,* **77**(6), KS43–KS54.

Vera Rodriguez, I., Bonar, D. & Sacchi, M., 2012. Microseismic data denoising using a 3C group sparsity constrained time-frequency transform, *Geophysics,* **77**(2), V21–V29.

Wang, T., Zhang, M., Yu, Q. & Zhang, H., 2012. Comparing the applications of EMD and EEMD on time–frequency analysis of seismic signal, *J. Appl. Geophys.,* **83,** 29–34.

Zheng, J., Peng, S.P., Liu, M.C. & Lianga, Z., 2013. A novel seismic wavelet estimation method, *J. Appl. Geophys.,* **90,** 92–95.

Zheng, J., Zhu, G. & Liu, M., 2015. Vibrator data denoising based on fractional wavelet transform, *Acta Geophys.,* **63**(3), 776–788.

## APPENDIX A

First, recursion begins at the node where the final loss occurs:

$$\frac{\partial L}{\partial L_t} = 1. \tag{A1}$$

The output $o_t$ is taken as a parameter of the logistic regression function, and the gradient $\nabla_{o_t} L$ on the outputs at time step $t$ is as follows:

$$\nabla_{o_t} L = \frac{\partial L}{\partial o_t} = \frac{\partial L}{\partial L_t}\frac{\partial L_t}{\partial o_t} = \hat{y}_t - 1\{y_t = \hat{y}_t\}, \tag{A2}$$

where $1\{y_t = \hat{y}_t\}$ is the indication function, the expression in braces is true, then the value is 1, otherwise it is 0. Next, we start from the end of the sequence. At the last time step $\tau$, $h_\tau$ has only $o_\tau$ as the successor node, so its gradient is calculated as follows:

$$\nabla_{h_\tau} L = (W_{ho})^T \nabla_{o_\tau} L, \tag{A3}$$

where $(W_{ho})^T$ represents the transposed matrix of the connection weight matrix between the hidden layer node and the output layer node. Next, iterate from $t = \tau - 1$ to $t = 1$ and back-propagate the gradient through the time step. Since $h_t (t < \tau)$ has two subsequent nodes, $o_t$ and $h_{t+1}$, its gradient is calculated as follows:

$$\begin{aligned}\nabla_{h_t} L &= \left(\frac{\partial h_{t+1}}{\partial h_t}\right)^T (\nabla_{h_{t+1}} L) + \left(\frac{\partial o_t}{\partial h_t}\right)^T (\nabla_{o_t} L) \\ &= (W_{oL})^T (\nabla_{h_{t+1}} L)\, diag(1 - (h_{t+1})^2) + (W_{ho})^T (\nabla_{o_t} L),\end{aligned} \tag{A4}$$

where $diag(1 - (h_{t+1})^2)$ represents the diagonal matrix containing the elements $1 - (h_{t+1})^2$. By gradient back propagation, the remaining parameters are given by

$$\nabla_{b_o} L = \sum_t \left(\frac{\partial o_t}{\partial b_o}\right)^T \nabla_{o_t} L = \sum_t \nabla_{o_t} L \tag{A5}$$

$$\nabla_{b_h} L = \sum_t \left(\frac{\partial h_t}{\partial b_h}\right)^T \nabla_{h_t} L = \sum_t diag(1 - (h_t)^2) \nabla_{h_t} L \tag{A6}$$

$$\nabla_{W_{ho}} L = \sum_t \left(\frac{\partial L}{\partial o_t}\right) \nabla_{W_{ho}} o_t = \sum_t (\nabla_{o_t} L)(h_t)^T \tag{A7}$$

$$\begin{aligned}\nabla_{W_{oL}} L &= \sum_t \left(\frac{\partial L}{\partial h_t}\right) \nabla_{W_{oL}} h_t \\ &= \sum_t diag(1 - (h_t)^2)(\nabla_{h_t} L)(h_{t-1})^T\end{aligned} \tag{A8}$$

$$\nabla_{W_{xh}} L = \sum_t \left(\frac{\partial L}{\partial h_t}\right) \nabla_{W_{xh}} h_t = \sum_t diag(1 - (h_t)^2)(\nabla_{h_t} L)(x_t)^T, \tag{A9}$$

where $\nabla_{b_o} L$ is the gradient of the loss $L$ to the output layer bias vector $b_o$; $\nabla_{b_h} L$ is the gradient of the loss $L$ to the hidden layer bias vector $b_h$; $\nabla_{W_{ho}} L$ represents the gradient of the loss $L$ to the connection weight matrix $W_{ho}$ of the hidden layer and the output layer; $\nabla_{W_{oL}} L$ represents the gradient of the loss$L$ to the connection weight matrix $W_{oL}$ of the output layer and the logical regression layer; $\nabla_{W_{xh}} L$ represents the gradient of the loss $L$ to the connection weight matrix $W_{xh}$ of the input layer and the hidden layer.