An Attempt to Pick Teleseismic P Wave Arrival Using Envelope and Artificial Neural Network Algorithm

| Item Type       | Conference Paper |
|-----------------|------------------|
| Authors         | Lobo, R. K.; Gaol, Y. H. L.; Fatimah, D. Y.; Abdullah, A.; Zaky, D. A.; Suhardja, S. K.; Nugraha, A. D.; Zulfakriza, Z.; Widiyantoro, S.; Ramdhan, M.; Palgunadi, K. H.; Mujihardi, B. |
| Citation        | Lobo, R. K., Gaol, Y. H. L., Fatimah, D. Y., Abdullah, A., Zaky, D. A., Suhardja, S. K., ... Mujihardi, B. (2021). An Attempt to Pick Teleseismic P Wave Arrival Using Envelope and Artificial Neural Network Algorithm. IOP Conference Series: Earth and Environmental Science, 873(1), 012059. doi:10.1088/1755-1315/873/1/012059 |
| Eprint version  | Publisher's Version/PDF |
| DOI             | 10.1088/1755-1315/873/1/012059 |
| Publisher       | IOP Publishing |
| Rights          | This is an author-created, un-copyedited version of an article accepted for publication/published in [JournalTitle]. IOP Publishing Ltd is not responsible for any errors or omissions in this version of the manuscript or any version derived from it. The Version of Record is available online at http://doi.org/10.1088/1755-1315/873/1/012059 |
| Download date   | 2023-11-01 16:38:12 |
| Item License    | http://creativecommons.org/licenses/by/3.0/ |
| Link to Item    | http://hdl.handle.net/10754/676124 |
An Attempt to Pick Teleseismic P Wave Arrival Using Envelope and Artificial Neural Network Algorithm

To cite this article: R K Lobo et al 2021 IOP Conf. Ser.: Earth Environ. Sci. 873 012059

View the article online for updates and enhancements.

You may also like

- STUDY OF SOLAR ENERGETIC PARTICLE ASSOCIATIONS WITH CORONAL EXTREME-ULTRAVIOLET WAVES
  Jinhye Park, D. E. Innes, R. Bucik et al.

- Microseismic event location using an inverse method of joint P-S phase arrival difference and P-wave arrival difference in a borehole system
  Wen Zhou, Liangshu Wang, Luping Guan et al.

- How proton pulse characteristics influence protoacoustic determination of proton-beam range: simulation studies
  Kevin C Jones, Chandra M Seghal and Stephen Avery
An Attempt to Pick Teleseismic P Wave Arrival Using Envelope and Artificial Neural Network Algorithm

R K Lobo\textsuperscript{1}, Y H L Gaol\textsuperscript{1}, D Y Fatimah\textsuperscript{1}, A Abdulla\textsuperscript{h1}, D A Zaky\textsuperscript{1}, S K Suhardja\textsuperscript{1}, A D Nugraha\textsuperscript{2}, Z Zulfakriza\textsuperscript{2}, S Widiyantoro\textsuperscript{2,3}, M Ramdhan\textsuperscript{4}, K H Palgunadi\textsuperscript{5}, B Mujihardi\textsuperscript{6}

\textsuperscript{1}Faculty of Exploration and Production Technology, University of Pertamina, Jl. Teuku Nyak Arief, Simprug, Kebayoran Lama, Jakarta 12220, Indonesia
\textsuperscript{2}Global Geophysics Research Group, Faculty of Mining and Petroleum Engineering Institute of Technology Bandung, Jalan Ganesa No. 10, Bandung 40132, Indonesia
\textsuperscript{3}Faculty of Engineering, Maranahta Christian University Bandung, Jalan Surya Sumantri No 65, Bandung 40164, Indonesia
\textsuperscript{4}Agency for Meteorology, Climatology and Geophysics, Jalan Angkasa I, No. 2, Kemayoran, Jakarta, Indonesia
\textsuperscript{5}Physical Science and Engineering, King Abdullah University of Science and Technology, Saudi Arabia.
\textsuperscript{6}Geothermal Research – Innovation & New Ventures, PT. Pertamina, Jl. Mega Kuningan Barat III, Jakarta Selatan, Indonesia

*Corresponding author: s4ndy104@gmail.com

Abstract. Seismic events detection and phase picking play an essential role in earthquake studies. Typical event detection is done visually or manually on recorded seismogram by choosing a series of higher amplitude signals recorded on at least 4 stations. More sophisticated methods have been used in event detection and picking with additional attributes such as Short Time Average over Long Time Average (STA/LTA). This method is based on average number sampled at multiple predefined windows. However, STA/LTA is dependent on the window size which becomes its drawback. In this study, we explore one derivative attribute, popularly known as envelope or instantaneous amplitude. It has been extensively used in seismic reflection and refraction method. In principle, this method uses the Hilbert Transform to calculate complex seismic trace and take the magnitude of complex seismic trace as envelope amplitude that can be used to analyze P wave arrival time. We employed one of the machine learning methods, Artificial Neural Network (ANN). The ANN method works by analyzing various inputs and training them to recognize patterns in P wave arrival signals. We started our study by applying envelope attribute to synthetic data with noise addition. We found that with noisy data the envelope attribute still gives a clear signal for first-time arrival. Next, we trained 300 seismograms of teleseismic events recorded on IRIS-US networks and tested our trained program on 20 seismograms as a blind test. To compare performance between the two methods, we calculated the difference between the results of automatic picking and manual picking. The final calculation shows an average deviation of 0.355 seconds. Twenty-five percent of testing data (5 samples) has a deviation above 0.5 seconds, and 75\% of the remainder (15 samples) already had a deviation under 0.5 seconds. The more significant deviations of the P wave picks are likely due to noisy signals in the data set and...
complex arrival signals. This study shows that the combination of envelope attribute and machine learning method is promising to distinguish teleseismic P wave arrival and automatically pick them.

1. Introduction.

Arrival times picking of seismic phases plays a significant role in earthquake data since it can provide information about the Earth’s interior structure. Seismic waveforms and its arrival time can be utilized with further advance methods such as Receiver Function, Shear Wave Splitting and Seismic Tomography to image Earth sub-surface structures. For the last 10 years, the quantity of seismic stations around the world are quickly increasing, thus did the seismic data volumes. Traditionally, an observer will do arrival time picking by visual observation on the seismogram. The result of arrival time picking depends on the experience of the observer. The manual picking process gives good results, yet this process is time-consuming since the massive volumes of seismic data. Moreover, the manual picking process sometimes has some trouble when it has a low signal-to-noise ratio (SNR).

To perform arrival time picking efficiently, numerous studies have proposed an automatic arrival time picking algorithms and one of the most broadly utilized methods is the Short-Term Averaging/Long-Term Averaging (STA/LTA) algorithm [1], that simply the ratio of the two averages calculated on sliding/moving windows over the trace. Coppens method, Modified Energy Ratio (MER) method, statistical attributes such as skewness and kurtosis, maximum-likelihood method, and cross-correlation methods are some of the commonly used methods in first arrival time picking algorithm. Most of this method is only effective with high SNR data. For low SNR data, there are possibilities that the first arrival did not detect by those methods [2].

With the target to improve time data processing, we used one of the machine learning technique named neural networks to analyze the first arrival on teleseismic data. STA/LTA (Short Term Averaging over Long Term Averaging) and envelope (instantaneous amplitude) are used as input parameters for neural networks data training. STA/LTA is utilized to clarify the contrast between seismic event signal and noise signal. The purpose of envelope usage is to perceive amplitude fluctuation when an earthquake event is coming so that the patterns of the first arrival on the seismogram could be distinguished. The envelope attribute uses Hilbert transform on a vertical component trace to obtain the complex seismic trace that has two parts, the real and imaginary parts of the trace. This research's primary purpose is to develop an auto-picking algorithm first arrival picking on teleseismic events with an accuracy that is similar with human expert performance. The application of the neural network's method is expected to give good picking results on low SNR data, where the first arrival patterns are not clearly seen.

The data used in this study are vertical component seismograms from teleseismic events from IRIS stations in the United States region. Waveform data were downloaded from the IRIS database using the ISC earthquake catalogue. The teleseismic earthquake event used has a magnitude above 6 (M> 6) with a recording period ranging from 2012-2017. The recording length of each seismogram is two minutes containing the arrival event of the P wave. The seismogram has a sampling rate of 40 Hz, so the data length of each seismogram is 4800 samples. There are 33 teleseismic earthquake events recorded by 61 stations on the IRIS seismological network in the Americas. Most of the earthquake events used are located on the South American Continent with a distance of more than 1000 km from the receiving station (Figure 1). In total, we collected 300 teleseismic events used for the training data process. For the validation process and blind test, 20 earthquake events were used. The blind test process is P-phase picking process on a new set of seismograms.
Seismogram pre-processing (filtering and normalization) aims to eliminate the effect of noise and uniform the amplitude value as seen on Figure 2. The filtering process was carried out using a third order low-pass filter with a cutoff frequency of 1 Hz. Less variation of amplitude or a uniform amplitude value will make it easier for the ANN learning process.

Figure 1. Map of US IRIS seismic network stations (red) and teleseismic event distribution (yellow).

Figure 2. (a) Sample seismogram after pre-processing, (b) classic STA/LTA, (c) recursive STA/LTA, (d) envelope attribute.
2. Methods

**STA/LTA**

The short time average/long time average (STA/LTA) has been widely reliable for detecting earthquake events and can be used with other methods to improve picking accuracy. The primary purpose of this method is to clarify the contrast between the earthquake signal and noise. STA/LTA measure the average of the absolute value from a series of seismic signal using two moving time windows – a short time average window (STA) and a long-time average window (LTA). The STA window will be more sensitive to seismic events because it measures the instant amplitude of the seismic signal and keeps track of earthquake events. Meanwhile, the LTA window functions to obtain background noise information, which usually has a longer period [3]. Figure 3 is an example of events and STA/LTA calculation. When clear signal arrived, a spike in STA/LTA attributes is obvious and able to differentiate between real arrival and not. In this research, we used classic STA/LTA [4] and recursive STA/LTA [5]. An example of STA/LTA attributes calculation can be seen on Figure 3.

![Figure 3](image-url)

**Figure 3.** The 10th real data set along with the attributes; (a) Sample seismogram after pre-processing, (b) classic STA/LTA, (c) recursive STA/LTA, (d) envelope attribute.

**Envelope Attribute (Instantaneous Amplitude)**

Hilbert transform process can used to remove the negative frequency part and double the magnitude of positive frequency. It can be seen as a band pass filter. The outputs of this process are instantaneous amplitude, frequency, and phase attribute information from seismic signal data. A seismic signal is a real part of a complex seismic trace. A complex seismic trace \( z(t) \) has real \( x(t) \) and imaginary parts \( y(t) \) as defined by the equation below:

\[
z(t) = x(t) + iy(t)
\]  

(1)

The real part \( x(t) \) is the real seismic trace, and \( y(t) \) is the imaginary seismic trace obtained from the Hilbert transformation of \( x(t) \). The envelope represents the total instantaneous energy of a seismic signal with an amplitude range that varies from zero to the maximum amplitude of a seismic signal. This method is commonly used to detect first breaks in seismic data. A spike or increase in the
amplitude value of the envelope can indicate the arrival time of the waves on the seismogram (Figure 3).

**Artificial neural Network**

Artificial Neural Network (ANN) is one of the machine learning method based on the working principle of human biological neural networks. The machine-learning algorithm focused on studying patterns from a set of training data. Models generated from machine learning have characteristics derived from the training data. After the model studies the patterns in the training dataset, the model can predict some desired output from other data.

The concept of the ANN learning process is learning from the wrong output values and will be updated as iterations increase. At the beginning of the learning step, the output model will have an error value which is the difference between the desired output value and the predicted output value. Changes in output values are based on learning experiences. The data training step aims to get the predicted output value that is close to the actual desired output.

The purpose of using the ANN method is to predict where on seismogram is the P-wave first arrival. The ANN model learns from the dataset input, which consists of the STA/LTA attributes, envelope attribute, and the label. Labels are guidance to differentiate between the first arrival zone and the non-first arrival zone on the seismogram.

The STA/LTA and envelope attribute were generated from the seismogram, these attributes will be used as a guide for the manual picking. From these attributes, the P-wave arrival pattern will be seen from the presence of an amplitude spike. Consequently, manual P-arrival picking is performed on the STA/LTA and envelope attributes. This manual pick intends to guide where the P-wave “arrivals” and “non-arrivals” are located. P-wave “arrivals” will be labeled 1 (true) and the “non-arrivals” zone will be labeled 0 (false). Therefore, this process is also known as labeling. The ANN auto-picker was then trained using the STA/LTA, envelope, and the label as the input parameters. The model generated from the training process is then used to predict the P-wave arrival from new data. Model validation is executed by predicting the P-wave arrival on synthetic data with some given noise. The model must be able to predict P-wave arrival on synthetic data with results that are close to manual picks by the observer.

3. Results and Discussions.

This section shows the results of predicted P-wave arrival on the real teleseismic data set. ANN auto-picker results will be compared with the manual arrival time picking by the observer. The ANN model was applied to predict arrival time from the blind data set. The model created can recognize the P-wave arrival pattern from the attributes used. The P-wave arrival can be seen at the beginning of the first spike of each attribute. This is following the theory that P-wave is the fastest wave recorded in a seismogram. The ANN model can distinguish between the highest first spike (arrival of the P-wave) and the next spikes, which are the other wave type arrivals.

Figure 4 shows the comparative results between the ANN auto-picker and the manual pick results. Five data samples (25% of the total data) has a deviation of more than 0.5 seconds. Meanwhile, the remaining 15 samples (75% of the total data) already had a deviation of below 0.5 seconds. The average deviation from 20 data samples is 0.355 seconds. This result is quite good because it is quite close to the manual pick results. The greater the deviation is likely due to the more complex attribute patterns from the data and has not yet been identified by the ANN model.

In some situations where the data is a bit noisy (figure 4), the ANN auto-picker will have picking results that are less close to the manual picks. This can occur due to the observer's manual picking errors or the complexity of the P-wave arrival patterns so that the ANN auto-picker cannot appropriately recognize it. More training datasets with more varied attribute patterns such as kurtosis and skewness are needed to make the ANN model better at identifying the correct P-wave arrival.
Figure 4. Comparison of manual picking results (red line) and predicted pick results on real seismograms (blue line).
References

[1] Allen R V 1978 Automatic Earthquake Recognition and Timing from Single Traces Bulletin of the Seismological Society of America 68(5) 1521-1532

[2] Aminzadeh F, Maity D, Tafti T A, Brouwer F 2011 Artificial neural network based autopicker for micro-earthquake data In SEG Technical Program Expanded Abstracts pp 1623-1626

[3] Trnkoczy A 2012 Understanding and parameter setting of STA/LTA trigger algorithm New Manual of Seismological Observatory Practice (NMSOP) pp. 1-20 Deutsches GeoForschungsZentrum GFZ

[4] Agius M R 2007 Automatic Earthquake Detection and Localisation from a Three-Component Single-Station M Sc diss University of Malta

[5] Grigoli F et al. 2014 Automated seismic event location by waveform coherence analysis Geophysical Journal International 196(3) 1742-1753