T3.7-P5

CTBT: SCIENCE AND TECHNOLOGY 2017 CONFERENCE

Automatic Classification of Seismic P- and S-Wave Signals Using Multiple Parameters, Frequency Ranges and Artificial Neural Network

Timo Tiira, University of Helsinki, Institute of Seismology, Helsinki, Finland (timo.tiira@helsinki.fi)

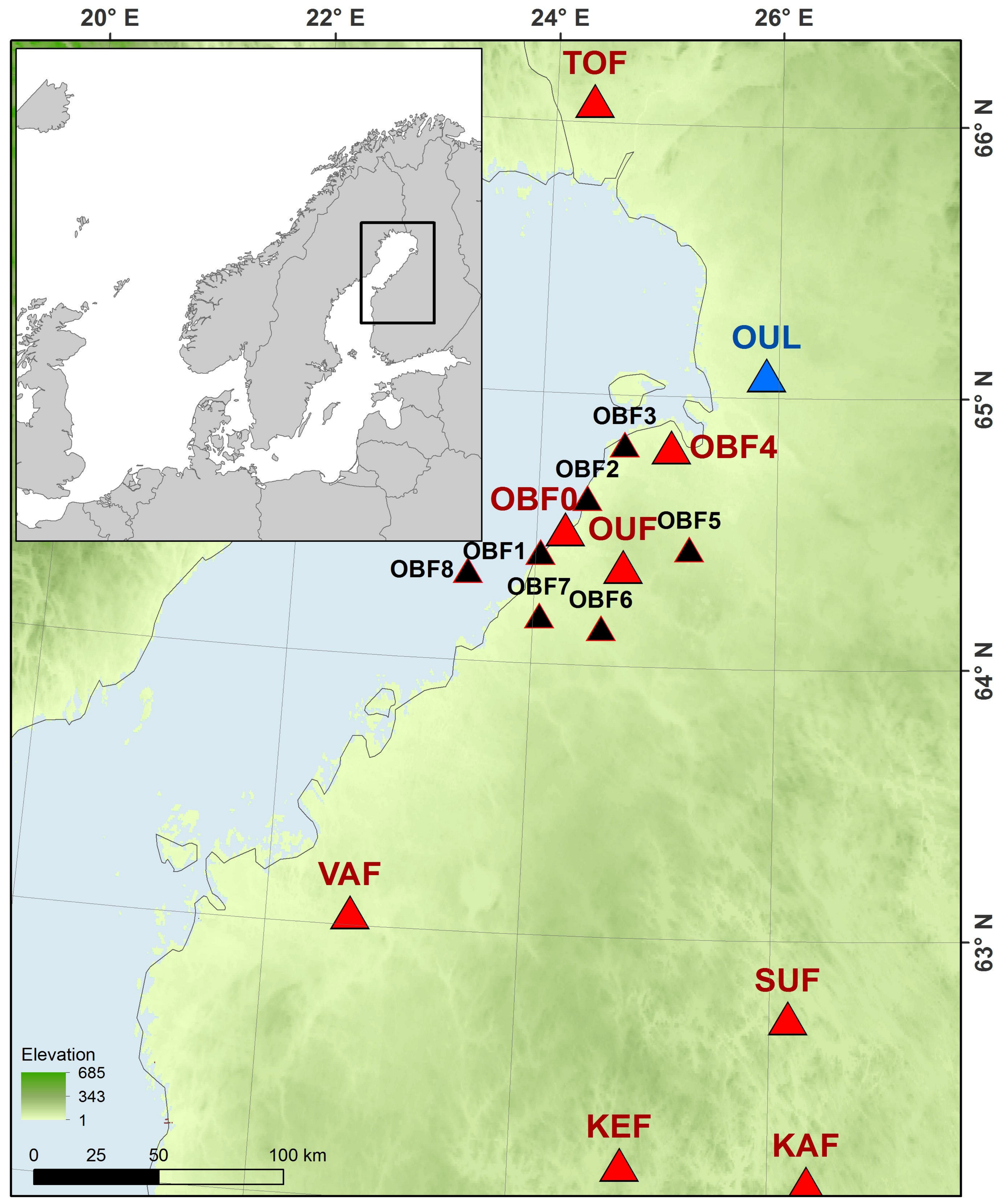

Fig. 1. Location of the Pyhäjoki seismic network. Stations marked with red are part of national seismic network.Station OUL is maintained by Oulu University.The planned power plant site is close to OBF0.

Background

Sites of nuclear power plants must be evaluated for seismic risk and monitored for seismicity. A new nuclear power plant is planned to be constructed at Hanhikivi, Pyhäjoki, Northern Ostrobothnia in Finland. Pyhäjoki is situated in the Central part of the Fennoscandian Shield, a region characterized by low intraplate seismicity. According to IAEA's guidelines a network of sensitive seismographs having a recording capability for micro-earthquakes should be installed to acquire more detailed information on potential seismic sources when a nuclear power plant site is evaluated. The operation period of the seismograph network should be long enough to obtain data for seismotectonic interpretation. A network of 10 broad-band stations has been installed within 50 km radius from the planned nuclear power plant site.

A seismic event (e.g. earthquake, chemical or nuclear explosion, landslide, volcanic eruption) exitates compressional primary waves (P-waves) and secondary waves (S-waves) which are shear waves. P- and S-waves propagate at different velocities. Seismic events are located using arrival times of P- and S-wave signals from source to seismic stations. The seismic signal detection and event location are done automatically in data centers. In automatic location systems reliable definition of the type of the incoming signal is crucial. To improve the current automatic seismic event location system at Pyhäjoki network a new classification method for P- and S-wave signals is needed. The automatic detection and location system sometimes discards or mislocates an events due to wrong identification the signal type.

The problem is solved using multiple parameters which describe different aspects of differences between P- and S-waves. The information the parameters contain is combined and processed with an Artificial neural Network (ANN) to produce reliable classification of the signal types.

Abstract

Automatic classification of seismic P- and S-wave signals is essential in automatic seismic event detection and location systems. The problem is tackled by utilizing multiple signals of different types each in several frequency bands. Artificial Neural Networks (ANN) are a robust and efficient tool in classification using large amount of input parameters. P- and S-wave signals have fundamentally different polarization properties. The input parameters depending on signal polarization in this study included rectilinearity, principal ellipticity, global polarization parameter, eigenresultants, quadratic resultant and predicted coherency. Several statistical parameters were used also. They included skewness, kurtosis and Jarque-Bera test. Instead of or in addition to several parameters their variances in time were added to the input database. Different amplitude ratios were used also. Many of the parameters were computed separately from vertical and horizontal channels. All parameters were computed at 6 different frequency range and time window combinations resulting 210 input parameters. The parameters were computed from 9946 seismic traces of local events creating 2088660 new time-series. Independent training, testing and validation datasets picked from these time-series consisted ~2.0M, ~2.0M and ~1.0M inputs correspondingly. Using a deep ANN with 4 hidden layers over 98% of the signals of validation data were classified correctly.

Method

The particle motions of P- and S-waves are fundamentally different. Thus parameters describing polarization are most important in classification. Since location of the event is not known when the classification is needed, all the parameters are independent of the direction of the incoming signals. The polarization parameters are computed from three-component seismograms mostly using eigenvalues and eigenvectors which describe principal axis of rotational motion. Other parameters include variance of skewness, kurtosis (average, variance and peak), Jarque-Bera test and STA/LTA. These additional parameters can be computed separately from each of the three components though all parameter-component combinations were not used. Total number of the different parameter types including parameters from different components was 35.

There are also differences in frequency contents of the signals. So all the parameters are computed in several frequency ranges. Since station to event distance varies and this affects length of the signals and time difference between P-wave and S-wave two different time window lengths were used. The time window and frequency range combinations are shown in Table 1. The total number of input parameters is 210. All parameters were computed in time domain after filtering the data.

There are many machine learning methods which can be used for classification. We use basic ANN known also as multi-layer-perceptron. In large number of input parameters there is bound to be significant amount of overlap but because we need all the information we can get, we use all parameters and let the ANN to utilize the useful part of the information. The ANN was used for this work because it is very error tolerant. The training data set can be very noisy, it can even include total outliers and the network can still converge. The ANN has also a high ability to generalize, to interpolate output for input vectors, which lie between vectors of the training set. The ANN consists of simple processing units called neurons or nodes and one-directional links between the nodes of consecutive layers. In the ANN the neurons are assembled in layers. The net consists of an input layer, an output layer and possibly one or more hidden layers. The topology is symmetric. Each node is connected to each node of the next layer. Each link between two nodes has a weight value. The memory of ANN lies in the weight matrixes between layers. The weights are changed during training until the optimal weight matrix is found. The neural net is trained using a training data set, which consists of known events. During training each input vector of the training database is fed to the network in turn. The output values are compared with wanted outputs and the error is propagated backwards through each layer. Using the error, the weights are corrected in each layer. The procedure is repeated until the net converges and the normalized error of the outputs is minimized. It is a good practice to use one part of the data for training, another for testing the performance of the net during training and a third for final validation of the trained net. The ANN is trained to give elevated output value for P- or S-wave signal and low value for noise samples and S- or P-wave signal correspondingly. Separate ANN's were trained for P-wave and S-wave signals.

Training of ANN and the data

The input parameters were computed for each sample in 2 second time window of P-wave signal and similarily for S-wave signal and for noise sample. The noise window had equal length and started 5 s before the P-wave onset. In short distances P-wave to S-wave time difference was so small that shorter windows were used to prevent overlap between P- and S-waves. In that case all three windows were equally shortened. The sampling rate was 100 hz. This resulted to 600 input vectors for most traces. Size of training and testing databases was 1534657 input vectors. Size of validation database was 786601 input vectors. For validation database the input vectors were computed in only 1 second long long time windows in order to simulate classification of unknown signals from unknown sources with unknown source to station distance. Number of traces used for testing and training data was 6488 and for validation data 3458. The traces were selected from 4906 seismic events. The magnitude range was from -0.46 to 4.1 and the median magnotude 1.3. The events were registered at local distances. The input data of each station was scaled and the scaling parameters were stored so that they can be used with new data.

ANN processing was done with program package Keras (Chollet 2015) and Python programming laguage. The ANN had 4 hidden layers each with 10 nodes. Sigmoid activation function (f(x)=1/(1+e**(-x))) was used in output layer and rectified linear unit (ReLU) (f(x)=ln(1+e**x)) in other layers. The ANN was trained using the training database. During training the testing database was used observe the convergence of the training process. After training the validation dataset was used to evaluate the performance of the trained ANN.

References
F. Chollet, Keras, (2015), GitHub repository, https://github.com//fchollet/keras
Python Software Foundation. Python Language Reference, version 2.7. Available at http://www.python.org

Conclusions

Machine learning tools, e.g. ANN, are powerful in classifying seismic P-wave and S-wave signals. The success rates are 98.1% and 98.9% for P- and S-waves. The false alarm rates are 1.1% and 2.0% correspondingly. The higher false alarm rate for S-waves reflects the contents of short period seismic noise which composes more of shear type of waves than compressional waves. Improving any automatic seismic detection and location system is possible with machine learning based classification if the performance of the system is deteriorated due to wrong identification of the signal detections.

TABLE 1

Time window [s]	Time window of LTA [s]	frequency range [hz]
1.0	5.0	18.0-30.0
1.0	5.0	27.0-40.0
2.0	10.0	1.5-5.0
2.0	10.0	3.0-8.0
2.0	10.0	5.0-12.0
2.0	10.0	11.0-20.0

Results

The evaluation of the method was done using the independent validation data from 2650 traces. One second of data from both P- and S-wave signal onsets of each trace were processed for ANN input. The inputs started at offset of the signals. Separate ANN's were used to recognize P- and S-wave signals, ANNP and ANNS. For each signal absolute maximum of the difference between ANNP and ANNS outputs was used to classify the signal. The system classified 98.1% of the P-wave signals and 98.9 of the S-wave signals correctly. The false alarm rates were low. The system classified only 1.1% of the noise samples to be a P-wave signal and only 2.0% to be a S-wave signal.