Automatic post processing algorithm for passive seismic monitoring data

K Nepeina
PhD student, Institute of Geosphere Dynamics Russian Academy of Science, Moscow, Russian Federation
E-mail: nepeina.k@mail.ru

Abstract. The problem of monitoring of different types of seismic events - geoacoustic precursors of earthquakes, industrial and field explosions, places fragments fall of separating parts of rockets-carriers, etc. is one of the key in the modern ecology of the environment. The peculiarity of this kind of monitoring is that it is mobile seismic groups, which should be based in the proposed area of occurrence of events. One of the most important steps for solving the problems connected with the detection and identification of recorded data from passive sensors in mobile seismic array (MSA). The task of determining the nature of the source and its’ coordinates lies in the basis of direction, referred to as the geoacoustic location. Using a new approach (not by location but by neural classification of waveform “portraits”) usability of algorithm which based on quantitative parameters of signal will be demonstrated.

1. Introduction
The increase on the demand for Passive Seismic especially in the domains of exploration and production has most recently become a challenging field of opportunity. The method based on uncontrolled sources, requires an understanding and analysis of what we call passive events in a careful manner, so as not to be misinterpreted leading to a possible damage of its credibility [1]. The problem of earthquake occurrence in platform territories, in spite of numerous examples of such seismic events, remains still unsolved in many aspects. This problem is particularly important for densely populated southern European Russia [2].

The first problem of monitoring of different types of seismic events - geoacoustic precursors of earthquakes, industrial and field explosions, places fragments fall of separating parts of rockets-carriers, etc. is one of the key in the modern ecology of the environment. The peculiarity of this kind of monitoring is that it is mobile seismic groups, which should be based in the proposed area of occurrence of events. One of the most important steps for solving the problems connected with the detection and measurement of parameters of sources of seismic events on the basis of registration of seismic signals with the help of spatially distributed sensors mobile seismic group. The task of determining the geographical coordinates of the source lies in the basis of direction, referred to as the geoacoustic location [3].

The second fundamental research focuses on the main differences between processes leading to earthquakes and the seismic rock mass response to mining. In earthquake seismology the driving forces are fairly constant and relatively slow facilitating the processes of self-organization. An important agent in the development of spatial and temporal correlations is seismic activity itself. Mining is not a spontaneous process. It induces stresses at a particular place, at a particular time and at
a particular rate, which are all highly variable compared to the tectonic regime. This process is controlled more by the commodity prices than by the plate tectonics. Consequently, unlike earthquakes, the seismic rock mass response to mining can be controlled to a certain degree [3].

The third objective of seismic monitoring is to detect and locate underground nuclear explosions. Nuclear explosion monitoring is as important today as it was at the dawn of the atomic age. Data resulting from seismic monitoring are used to distinguish between an underground nuclear explosion and the numerous natural and man-made seismic events that occur every day, such as earthquakes and mining explosions [3].

The passive seismic monitoring is a very important method for obtain geodata which is useful for academic science and practical applications. The main targets of passive seismic monitoring are:

- précising the movements of Earth’s crust by understanding frequency of the seismic events
- detecting the character of seismic events and its localization
- zoning the endogenic epicentres and focusing on the territories: understanding nature
- seismic stability for constructions nuclear power stations
- forecasting and statistics of seismic events
- Comprehensive Nuclear-Test-Ban Treaty (CTBT) verification regime: location underground nuclear explosions

The Government has ratified the Federal Program - “Seismic safety of territory of Russia” (2002-2010). The purpose of this Program is the maximal increase of seismic safety of the population, reduction of social, economic, ecological risk in seismically dangerous areas of the Russian Federation, decrease of damages from destructive earthquakes by certification, strengthening and reconstruction of existing buildings and constructions, and also preparation of cities and other settlements, transport, power constructions, pipelines for strong earthquakes. The actual frequency of large earthquakes in Northern Eurasia is three and more times higher than previously assumed. The use of straight plots in past years resulted in significant overestimation of the return time of large earthquakes, hence, in underestimation of seismic hazard practically in all the regions of the former USSR [2].

2. Prerequisites

Array stations allow for a better estimation of the azimuth of incoming signals, i.e. they identify the direction from which the signal arrived. The spatial distribution of the sensors also permits an estimation of the seismic waves’ speed. Information on both the direction and speed of the incoming seismic waves is crucial when identifying the source of a particular event. Three-component stations employ only one seismic sensor that measures the three spatial components of seismic waves—up-down, East-West and North-South. This set-up provides information on the depth and strength of an event. In order to eliminate the influence of seismic noise, seismic stations are usually built in remote areas, preferably on the outcrops of geological hard rock and as far away as possible from human activity [4].

These were difficult to search when seeking past events for comparison with new events and expensive to maintain. Catalogs devised by many individual observatories around the planet were inconsistent in format and difficult to collate and reconcile manually. Automatic catalog reconciliation is now a reality for scores of formats. Databases are used to integrate and index waveform observations from thousands of stations in many formats into structured collections of uniform data amenable to sophisticated data mining approaches. Increasingly the construction of databases will be automated with autonomous agents scouring the Web for new catalog results and waveform data collected across the Internet in continuous streams. The convergence of communication technology, storage technology, sophisticated software tool development and database technology will put terabytes of organized information in the hands of researchers through largely automated processes [5].
2.1. Research object

Modern seismological monitoring of the EEP is carried out by the Geophysical Survey of Russian Academy of Sciences (GS RAS) in close contact with different regional organizations, including academic and university research units. Local networks and processing centers are set up near the Russian cities of Voronezh, Perm, Saint-Petersburg, and Arkhangelsk. Some of these stations use an automatic signal detecting and identification application module. But identification criteria are different in various territories. Thus we should consider territorial features (depth of the sediment layers, body wave velocities etc.). The small aperture seismic array (MSA) Mikhnevo is situated in Moscow area on East-European platform. An advantage of this array is the noise reduction by coupling the receivers. Using SM3-KV instruments we collected data from 2004 year (Table 1). The interpreters made catalogs of seismic events and produced database of seismic events: among others there were earthquakes with magnitudes about 2.5-3.0.

| SM3-KV                          | Dimension | Value      |
|---------------------------------|-----------|------------|
| Measurement frequency domain    | Hz        | 0.5 – 100  |
| Conversion coefficient          | V × sec / m | 135 ± 20  |
| Threshold of sensivity          | m / sec  | 3 × 10-9   |
| Self-noise                      | m / sec  | 3 × 10-10  |
| Natural period                  | sec      | 2.0 ± 0.1  |
| T                               | °C       | -20 … +40  |

Table 1. Technical characteristics of seismic receivers SM3-KV.

The objective of this presentation is to look at one of the methodologies that benefit from passive sources, and justify its use with case studies from MSA Mikhnevo. The targets of MSA are the low-magnitude events (M > 1.5-4.5). The data of the Mikhnevo 3-C station from the geofon network are added to these observations for some earthquakes. The average value of microseismic noise spectral density for displacements is equal to 2 nm²/Hz at the frequency of 1 Hz, which is comparable to array conditions at other sites [Harjes, 1990; Nevskiy et al., 2003]. The array configuration is similar to the NORES type arrays. The low-frequency acoustic/seismic background noise (between 0.5-10.0 Hz), which is actively emitted by the earth, lies near the central frequency of wanted signal (body waves: 4-8 Hz – P-waves, 2-3 Hz – S-waves, 0.5-0.8 Hz – Surface waves). In 2007, MHVAR independently recorded 623 local events with magnitudes ($M_L = 0.79$ – $3.24$) which are identified as quarry blasts. The seismic monitoring of the territory of the East European Platform (EEP) is a part of the complex study of dynamics of tectonic processes, manmade influence and geo-ecological effects. The main feature of MHVAR is that the constantly operating small aperture array is set up in the region with the presence of sedimentary cover which is 1.5–2 km thick [6].

Reviewing datasets we can observe some different seismic events. They could have nature from: natural hazard (landslides, earthquakes etc.) or anthropogenic activity (mining exploration, munitions disposal etc.). We need to produce an automatic algorithm which can indicate a seismic event nature. After this procedure operator will use results for next processing (such seismic event location). This is of particular importance in passive seismic systems where $STA/LTA$ algorithms require an input signal. Typical casing failure events have large relative amplitude, a body waves $P/S$ ratio close to unity.
Then we estimate P/S ratios, Hilbert transform, signal lasting and Fourier spectra – these are waveform “portraits”.

3. Methods
Trends in instrumentation are providing us with progressively more and better data, and at lower cost. Mathematics was translated to assembler code in the early days of the computer and through the evolution of the computer the research and business world has seen FORTRAN, BASIC, PASCAL, C++ and many other languages. Automatic catalog reconciliation is now a reality for scores of formats. Databases are used to integrate and index waveform observations from thousands of stations in many formats into structured collections of uniform data amenable to sophisticated data mining approaches [5].

The object of this work is to construct a post-processing algorithm for waveform identification. At first raw data is a rather large massive. For programming and quick automatic identification we need to describe waveform by limited set of parameters. We provide an application CASE (Classification Algorithm for Seismic Events) employing Matlab 6.1. and internal «Fast Artificial Neural Network Library» (Figure 1). CASE can make solutions like an expert system providing an interpretator choice. Neural algorithms are the most powerful for classification and identification. The complex database includes waveform “portraits” (2004-2013 years) is used for network training. This algorithm contains three computational blocks. The first one is raw data processing and detection of main parameters. Then we could observe data limits and make decisions about difference between close events. After this we construct a simple neural network with weight matrice for machine teaching. The third block is made for identification a text file with some set of parameters and it is presented as a neural network too.

Modeling this network is rather difficult task. At the end of this work we expect to have one of the answers: noise, known event and unknown event (Figure 2).
Fig. 2. Possible three simpled results after neural network response. Here vectors $\overline{X}$ and $\overline{W}$ means set of parameters and weight function correspondingly.

3.1. A signal detecting.

The algorithm $STA/LTA$ is processing already filtered signal with the help of two moving time window («moving average»): a Short Time Average window $(STA)$ and a Long Time Average $(LTA)$.

$$STA_i = \frac{1}{NS} \sum_{j=i-NS}^{i} |X_j|, \quad LTA_i = \frac{1}{NL} \sum_{j=i-NL}^{i} |X_j|$$

$$r_i = \frac{STA_i}{LTA_i},$$

where $NS$ и $NL$ – number of digitizing points in a short and in a long time “windows”? consequently $(NS \ll NL)$, $r_i$ – ratio $STA/LTA$. In this work after some investigations the ratio was equal 1.4. The time of start is the time of primary wave (P). This fact fits well with the article Plenkers, Ritter & Schindler [2012].

3.2. Waveform parameters for neural network system.

At the beginning we should install rules of signal selection and rejection. This question is close to clasterization problem, because it is important to classify variety of patterns. The most intense parameter for cluster determination is a time delay between $S$- and $P$-wave: close or far event. Then there is a possibility to define a surface wave’s existence. If yes – it should be an explosion, otherwise it is an earthquake. For more accurate definition a body wave’s spectrum amplitude ratio and envelope function need to be employed. For distinct separation scientist should have a data statistics, which is laid in neural network operation. Hence the waveform database from 2004 is so required. Thus four main parameters and three ongoing parameters were selected for description image (Table 2).

In this investigation an algorithm with machine teaching was applied. Feed-back neural networks - the input information defines the initial activity state of a system. The first output of the system is taken as the new input, which produces a new output and so on. The first input pattern is presented to an initially randomized network, and the weights and thresholds adjusted in all the links. Other patterns are then presented in succession, and the weights and thresholds adjusted from the previously determined values. This process continues until all patterns in the training set are exhausted. It is generally accepted that this procedure is independent of the order in which the example patterns are
presented. However, a final check can be performed by looking at the pattern error which is defined as the square of difference between desired output and neural network output for each pattern and the system error which is defined as the average of all pattern errors, to determine whether the final network solution satisfies all of the patterns presented to it within a certain error. The set of weights and thresholds in the network are now specifically tailored to “remember” each input and output pattern, and can consequently be used to recognize or generate new patterns given an unknown input [7].

| Parameters                                   | Nature etalons                                                                 | Ranges                  |
|----------------------------------------------|-------------------------------------------------------------------------------|-------------------------|
| Existence P-wave                             | ✅ 50 career blasts                                                           | yes/no                  |
| Time P-wave                                  | ✅ noise:                                                                     | hh:mm:ss                |
| Time S-wave                                  | • Landslides                                                                 | hh:mm:ss                |
| Delta Time S-P-waves                         | • Thunderstorms                                                              | 00:00:ss                |
| Signal Length                                | • Avalanches                                                                 | 00:mm:ss                |
| Existence L-wave                             | • Falling meteorites                                                         | yes/no                  |
| Polynomial coefficients for Hilbert transform| • Anthropogenic noise                                                        | A1…A10                  |
| Spectral amplitude ratio P/S waves function  | from transport, building etc.                                               | 0-1,5                   |
| Extremes of spectral amplitude ratio P/S waves| ✅ endogenic processes                                                        | B1…B3                  |

Table 2. Used parameters for neural network teaching and work

During comparison author decided use Euclidian measure (Figure 3) – the simplest way for definition and is computed as an equation (3).

$$\rho(x_i, x_j) = \sqrt{(x_i^1 - x_j^1)^2 + (x_i^2 - x_j^2)^2 + \ldots + (x_i^p - x_j^p)^2}$$  \hspace{1cm} (3).

Referred to a monograph [7] for the group of “per sample” it is more suitable to apply Euclidian measure for computing and k-means method or LVQ (Linear Vector Quantization). The first one and the simplest is the k-means algorithm. This algorithm clusters observations into k groups, where k is provided as an input parameter. It then assigns each observation to clusters based upon the observation’s proximity to the mean of the cluster. The cluster’s mean is then recomputed and the process begins again. Here’s how the algorithm works:

1. The algorithm arbitrarily selects k points as the initial cluster centers (“means”).
2. Each point in the dataset is assigned to the closed cluster, based upon the Euclidean distance between each point and each cluster center.
3. Each cluster center is recomputed as the average of the points in that cluster.
4. Steps 2 and 3 repeat until the clusters converge. Convergence may be defined differently depending upon the implementation, but it normally means that either no observations change clusters when steps 2 and 3 are repeated or that the changes do not make a material difference in the definition of the clusters.

LVQ directly define class boundaries based on prototypes, a nearest-neighbour rule and a winner-takes-it-all paradigm. The main idea is to cover the input space of samples with ‘codebook vectors’ (CVs), each representing a region labelled with a class. A CV can be seen as a prototype of a class member, localized in the centre of a class or decision region (‘Voronoï cell’) in the input space. Note that every set of parameters (waveform portraits) should be normalized by etalon characteristics (certain average object). Whereupon minimize error function is fitted.

The most efficient for solving the main problem is k-means algorithm, because we can predict a number of clusters. Moreover during the processing we obtain different values, such is available for means estimation. So Euclidian measure is a distances $L_{1,0}$ and $L_{2,0}$ (Figure 3) either between cluster cores, or average waveforms portraits and between, also or new unknown events and etalon (a standard vector line).
Fig. 3. Euclidian measure for waveform “portraits” comparison in N-space. There is a simple example where one (grey colour) point is an etalon used as a basis for comparison; points with question-mark are unknown events. $L_{1-0}$ and $L_{2-0}$ are the distances between observed set of parameters $I_k(x_{11}, x_{12}, \ldots, x_{1N})$ and $I_m(x_{21}, x_{22}, \ldots, x_{2N})$.

4. Results

For numerical calculations was used Matlab 6.1 and its Neural Networks Toolbox. The multi layer perceptron is the most employed, because it holds a good computational power and capability to learning. It is referred to feed-forward back propagation algorithm. For the learning rule was applied trainlm function. For adaptation that leaerngdm. In accordance with chosen network the activation function was tansig and performance criterion – mse (mean-square error). All description you may find in Matlab manuals. It was decided provide two layers for this construction. One of the main decision rule was delta time S-P-waves: in case it is less than 0,20 sec, 0,20-0,30 and more than 0,30 sec.

After some manipulation with input and output vectors errors were estimated. And we obtained three big clusters. As a result of this conception it was possible to classify different rock response from the region seismic activity. It is very crucial to correctly find a set of parameters for each etalon. However it remains a problem of retraining. It is worth to specify deviations from the standard. After all procedures classes were derived among which were allocated by the characteristics as career explosions but earthquakes were simply discarded. The most recognizable were carrier blasts from Novogurovskyi, Kovrov, Shurovskoy ore mining department. They differ in general view (Figure 4) and consequently in portrait parameters (Table 3). Thus various portraits “points” got into different clusters. This information is the newest one for near Moscow region. More than 10 events were referred to unknown. They strongly attracted to the cluster with Kovrov explosions.

| Name        | $T_{signal}$ sec | $T_s$-$T_p$, sec | $maxA_p/maxA_s$ |
|-------------|------------------|------------------|-----------------|
| Novogurovskyi | 28-35            | 6-7              | 1,02-1,88       |
| Shurovskoy   | 40-60            | 8-9              | 0,8-1,12        |
| Kovrov       | 14-20            | 29-30            | 0,9-1,20        |

Table 3. Some resulted main parameters for carrier explosions.
5. Conclusions
A new approach without procedure of epicenter location is innovative method and is capable to be programmed. This algorithm optimizes human work and help to observe a large data more quickly. Automatic detection facilitates processing. Other researchers could use this software in future for bulletins and catalogs of events. Basing on these results scientists could make forecast using probability and statistics of events timing.

It is widely known that neural networks are ideally suited to situations where standard algorithms cannot be used as the mathematical relationship is uncertain, provided a supply of many examples of what is required can be generated. The performance depends on the training data and its ability to generate solutions cannot lie too far outside its experience-in our case. Although, in general, extra training sets should help with this problem, for this application it is too time-consuming to train the network to recognize more complicated portraits, and pre-processing for the source function appears necessary. Consequently, application of a network can often be used as a tool to highlight relationships which exist between certain quantities, and help gain a greater insight into the physical process [8].

Acknowledgements
I would like to thank IDG RAS for collected data and my academic tutor Sanina I.A. and Institute of Geosphere Dynamics for seismic data.

References
[1] Kapotas S, Martakis N, Tselentis A 2006 The Place of Passive Seismic in E&P Passive Seismic: Exploration and Monitoring Applications A21 (Dubai)
[2] Ulomov V I 2007 Seismogeodynamics and seismic hazard prediction National report to the International Association of Seismology and Physics of the Earth’s Interior of the International Union of Geodesy and Geophysics 2003 – 2006 (Moscow) pp 45-49
[3] Tkhak K 2013 Past, Present and Future Passive Seismic Tasks for Sustainable Development Sustainable Earth Sciences (DOI: 10.3997/2214-4609.20131639)
[4] http://www.ctbto.org/verification-regime/monitoring-technologies-how-they-work/seismic-monitoring/

[5] Anderson D N et al 26th Seismic Research review – trends in nuclear explosion monitoring (www.ees.lanl.gov/pdfs/01-01.pdf)

[6] Sanina I A et al 2011 The Mikhnevo small aperture array enhances the resolution property of seismological observations on the East European Platform J. Seismol. 15 (DOI 10.1007/s10950-010-9211-x) pp 545–556

[7] Ovchinnikov P E 2012 An artificial neural networks application for signals processing Nizhnii Novgorod p 32 (http://window.edu.ru/resource/846/77846/files/Ovch_ANN.pdf)

[8] Dai H, MacBeth C 1994 Split shear-wave analysis using an artificial neural network? First break Vol 12 12 pp 605-613