

Search for anomalies in pulse flows of acoustic and electromagnetic emissions

Yury Senkevich^{1*},

¹Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS, Laboratory of Acoustic Research, Mirmaya str., 7, Paratunka, Kamchatskiy kray, 684034, Russia

Abstract. Timely warning of disasters caused by earthquakes ensures life safety. Therefore, the search for markers of pre-seismic events preceding earthquakes remains an important research task. The article presents experimental methods for assessing seismic activity in the Kamchatka region based on the results of processing and analysis of geoacoustic and electromagnetic emission signals. The research is aimed at detecting anomalies in quantitative and qualitative indicators that characterize the pulse streams of acoustic emission of near-surface rocks and electromagnetic emission in the surface layer of the atmosphere. Signal processing and analysis are carried out using special algorithms that take into account the structural features of the variety of pulse shapes and their distribution over time.

1 Introduction

The search for stable markers of seismic events based on the analysis of seismic and electromagnetic signals, despite a fairly long history of study, is constantly in the attention of researchers. The most significant earthquake prediction models were created, for example [1, 2]. These models are mainly based on the analysis of very low frequency acoustic oscillations (up to 100 Hz). In recent years, geophysical signals of a higher frequency range (from 20 Hz to 10–16 kHz and higher) associated with the physical processes of plastic deformation have been studied. The geoacoustic emission (GAE) signals observed in the surface layer of the Earth, as well as electromagnetic radiation (EMR) signals recorded in the surface layer of the atmosphere are considered. [3, 4]. Over the past decade, research on the possibility of creating detectors of anomalies associated with seismic events in these geophysical signals has been carried out at IKIR FEB RAS (Kamchatka). Despite the different physical nature of the considered emission signals, the group behavior of geoacoustic and electromagnetic pulses are similar. The authors [4, 5] show that the change in the pulse variety over time is associated with seismic events. Therefore, there is a special interest in studying the dynamic characteristics of emission geophysical signals. This work presents one of the areas of research based on

*Corresponding author: senkevich@ikir.ru

the information approach to the processing of geophysical signals [6]. The idea of the approach is to transform the stream of pulses of the emission geophysical signal into a code message. This will allow us to apply methods and algorithms of linguistic analysis.

2 Analysis of geophysical pulse signals informativity

To obtain normal and anomalous characteristics of emission geophysical signals, it is necessary to classify the pulses in the stream, and then characterize their distribution by classes on the time axis.

At the first stage, the pulses are clustered in shape according to the values of the “special” points, which largely characterize the signal function. Such points are local extrema of the amplitude-phase pulse pattern, according to a number of methods of mathematical morphology in the recognition and description of graphic objects [7, 8]. The values of the amplitude of local extrema and time intervals between neighboring extrema are fixed. This creates two sets of values that fall under the definition of ordered sets: $\{a_i\} = a_0, a_1, \dots, a_N$ and $\{\tau_i\} = \tau_0, \tau_1, \dots, \tau_{N-1}$. Two square matrices of rank relations of the amplitudes of local extrema and the values of the intervals between them are constructed according to the rule:

$$r_{i,j} = \begin{cases} 1, a_i > a_j \\ 0, a_i \leq a_j \end{cases}, \quad i, j = \overline{1, N} \tag{1}$$

$$\omega_{i,j} = \begin{cases} 1, \tau_i > \tau_j \\ 0, \tau_i \leq \tau_j \end{cases}, \quad i, j = \overline{1, N-1} \tag{2}$$

The resulting binary square matrices with sizes N and $N-1$, respectively, are redundant due to the properties of the inequalities (if $r_{i,j} = 1$, then $r_{j,i} = 0$, similarly for $\omega_{i,j}$). The lower triangular parts of these matrices, including the main diagonal, should be excluded. The relational matrix of time intervals between local extremes is transposed. At the relational matrix of local extrema amplitudes, the first column and the last row containing insignificant zero elements are deleted. The resulting matrices are added. As a result we get a square binary matrix of the form:

$$\mathbf{D} = \begin{pmatrix} r_{0,1} & r_{0,2} & \dots & \dots, -1 & r_{0,N} \\ \omega_{0,1} & r_{1,2} & \dots & \dots, -1 & r_{1,N} \\ \omega_{0,2} & \omega_{1,2} & \dots & \dots, -1 & r_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \omega_{0,N-2} & \omega_{1,N-2} & \dots & \dots, /-1 & r_{N-2,N} \\ \omega_{0,N-1} & \omega_{1,N-1} & \dots & \dots, N-1 & r_{N-1,N} \end{pmatrix} \tag{3}$$

The matrix \mathbf{D} describes the shape of the pulse from the position of amplitude-phase relations and it is the basic pulse characteristic. Hereinafter, this matrix will be called a pulse image matrix. Due to the properties of inequalities, the image matrix acquires the property of invariance to the time and amplitude transformation of the signal.

Invariance is also preserved if there is a monotone trend in the signal. This means that each image matrix describes a variety of pulse shapes. The invariance property reflects the physical processes of signal attenuation and stretching in dense media, and also makes it possible to construct algorithms for identifying and clustering the fixed pulse variety.

Clusters are formed from impulses close in shape taking into account an empirically specified similarity coefficient. The similarity coefficient of the pulses is calculated by the number of matching elements of their image matrices. Each cluster is characterized by a certain image matrix, which can be associated with a symbolic notation. The set of all symbols, ranked by the number of pulses assigned to a particular cluster, makes up the alphabet. The alphabet characterizes the impulse variety in the time interval T . An example of a 15-minute GAE signal and its alphabet are shown in Fig. 1. At the bottom of the Fig. 1, the numbers on the alphabet chart indicate the sizes of the characters. The number of pulses included in a cluster of a particular symbol is measured on the vertical axis.

Fig. 2 shows the results of applying the clustering procedure for various similarity factors. The clustering procedure reduces the redundancy of the symbol composition of alphabets. Redundancy occurs because of distortion of the pulse pattern under the influence of noise, on the one hand, and due to the nonlinearity of the propagation medium, on the other hand. Distortions of pulse patterns change image matrices and generate redundant alphabet characters.

In order to identify anomalies in the signals, it is possible to conduct a quantitative and qualitative assessment of changes in the distribution of impulses and the assessment of the dynamics of the variety of shapes over time, comparing the alphabets that are found in equal successive episodes of observation.

Writing characters in the order of appearance of the corresponding pulses of the real signal generates a conditional text, which is used in the further analysis as an encoded message. To “decipher” such a message, that is, to search for hidden connections, it is convenient to use methods of linguistic analysis.

In this paper, the analysis of messages is carried out using the following indicators and criteria.

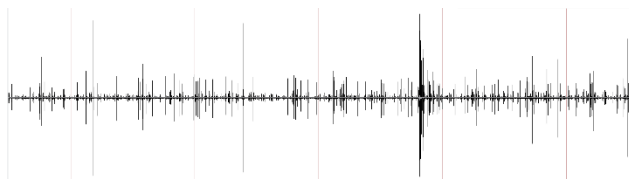


Fig. 1. Example of the GAE signal (on the top) and its alphabet (at the bottom). Numeric values indicate the size of the symbol. The volume of clusters is displayed on a logarithmic scale. The minimum significant repetition frequency of a symbol in a file is 2.

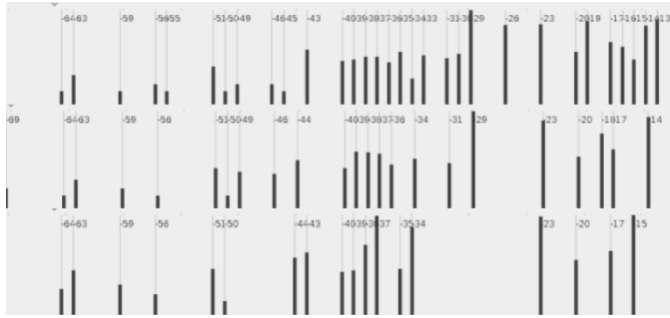


Fig. 2. An example of the clustering of a pulse stream shown in Fig. 1. From top to bottom, the similarity coefficients are 0.9, 0.8, and 0.7, respectively.

The frequency of the occurrence of a symbol in the alphabet is different and indirectly indicates a likely change in the behavior of a dynamic system. A reflection of the functioning of this system is observed through the dynamics of the pulse stream. Therefore, the analysis uses the probability of occurrence of the symbol a_i of the alphabet A , equal to the ratio of the number of occurrences of a symbol in a message (m) to the total number of symbols in this message (N): $P(a_i) = m / N$. According to information theory, this indicator is an estimate of the information load per symbol of the alphabet. In accordance with the recommendations of the authors of symbolic dynamics [9, p. 220], one should try to detect connections at the level of the Markov chain of higher orders, calculating the conditional probabilities of the occurrence of symbol combinations:

$$P(b | a) = P(ab) / P(a); P(c | ab) = P(abc) / P(ab), \text{ etc.}, \quad (4)$$

where a, b, c are symbols from the alphabet $\mathbf{A} = \{a, b, c, \dots, z\}$ extracted from the signal, z is the “null symbol”, a sign that complements the alphabet to the full. This symbol designates all occurring pulses in the stream, the image matrices of which were not repeated in the era of analysis.

The rate of production of new symbols (v) is equal to the ratio of the number of new symbols $\Delta N = N_{t+1} - N_t$ produced by the system during a given measurement interval to the length of the interval Δt : $v = \Delta N / \Delta t$; in this case, if $v > 0$ – the processes are excited in the medium generating the signal (exit from the stable state); if $v < 0$ – stabilization of processes in the medium generating the signal is observed (entrance to the stable state).

Alphabet saturation criterion: the rate of production of new symbols when approaching the moment of detection of a finite number of alphabet A symbols during the observation time $T \gg t_0$ (t_0 is the established observation interval) tends to zero,

$$\Omega = \lim_{\Delta t \rightarrow T} \Delta N / \Delta t = 0 \quad (5)$$

With long-term observation, there may come a time when new symbols do not occur. This criterion is used for training working algorithms for determining the stationary states of the signal-generating medium.

The criterion for alphabets association is calculated as the degree of intersection. The coefficient of symbolic overlap of the alphabets k_{AB} for alphabets \mathbf{A} and \mathbf{B} with sizes N and M , respectively, is determined by the formula:

$$k_{AB} = |\mathbf{A} \cap \mathbf{B}|^2 / (N^2 + M^2) \tag{6}$$

Introduce the following definitions and operations for manipulating symbols and alphabets. Each element of the alphabet is an object with attributes: the size of the symbol (s), the frequency of the symbol occurrence in the message (p), the image matrix of the pulse (\mathbf{D}). In other words, when considering alphabets, we are dealing with sets consisting of objects, each of which also represents an independent nonuniform ordered set – a tuple $a = \{s, p, \mathbf{D}\}$. Alphabet: $\mathbf{A} = \{a = \{s, p, \mathbf{D}\}, s \in \mathbf{S}, p \in \mathbf{P}, \mathbf{D} \in \check{\mathbf{D}}\}$. Consider actions with alphabets.

Definition 1. The union of the two alphabets \mathbf{A}_i and \mathbf{A}_j is the alphabet \mathbf{A}_k , constructed by the rule:

$$\mathbf{A}_k = \mathbf{A}_i \cup \mathbf{A}_j = \{a_k: a_k \in \mathbf{A}_i \text{ or } a_k \in \mathbf{A}_j\} \tag{7}$$

where $a_k = \{s_i, p_i, \mathbf{D}_i\}$, if $a_k \in \mathbf{A}_i$ and $a_k \notin \mathbf{A}_j$; $a_k = \{s_j, p_j, \mathbf{D}_j\}$, if $a_k \notin \mathbf{A}_i$ and $a_k \in \mathbf{A}_j$; $a_k = \{s_i, p_i + p_j, \mathbf{D}_i\}$, if $a_k \in \mathbf{A}$ and $a_k \in \mathbf{A}_j$.

Definition 2. The intersection of the two alphabets \mathbf{A}_i and \mathbf{A}_j is the alphabet \mathbf{A}_k , constructed by the rule:

$$\mathbf{A}_k = \mathbf{A}_i \cap \mathbf{A}_j = \{a_k: a_k \in \mathbf{A}_i \text{ or } a_k \in \mathbf{A}_j\} \tag{8}$$

where $a_k = \{s_i, p_i + p_j, \mathbf{D}_i\}$.

Computer programs have been developed for methods for extracting pulses and composing alphabets of messages used to process and analyze emission geophysical signals. The sensitivity and noise immunity of the implemented algorithms were estimated using a numerical experiment [10]. We use the introduced definition of the alphabets intersection to assess changes in the pulse stream from one time episode of observation to another.

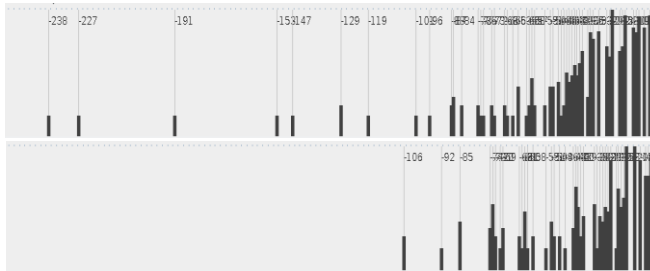


Fig. 3. Changing the composition of alphabets with decreasing soil moisture. Wet soil at the beginning of the day (on the top) and less moist soil at the end of the day (at the bottom).

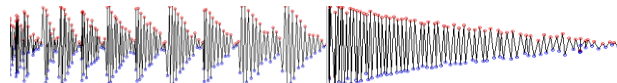


Fig. 4. Chain of overlapping GAE pulses (on the left). GAE pulse with 130 local extrema (on the right).

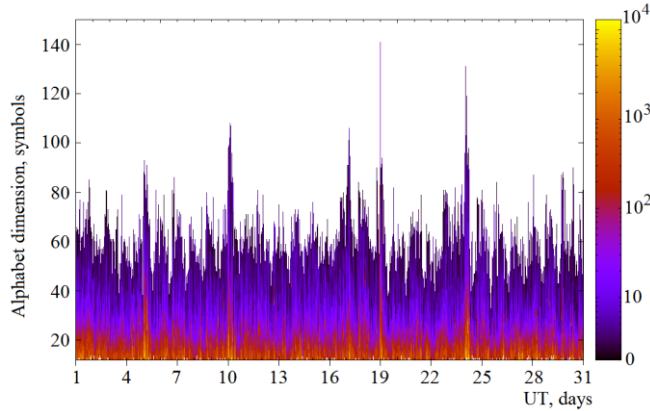


Fig 5. 3D-representation of alphabets of EMR signals received during July 2018. The x -axis marks the measurement date (the measurement step is 15 minutes); along the y axis is the dimension of the alphabet; along the z axis (color bar) is the frequency of pulses occurrence in the 15-minute measurement interval.

It should be expected that if there is a connection between the processes inside the medium that generate and / or affect the generation of pulses with quantitative and qualitative indicators of the pulse stream, the symbolic composition of the alphabets for the observation episode before the occurrence of an event and for the observation episode associated with the event should noticeably change. An example of this change when soil moisture decreases is shown in Fig.3. When the soil dries, the number of pulses decreases. Symbols larger than 100, formed by chains of overlapping impulses (Fig. 4, on the left) due to leakage of soil gas bubbles, as well as long high-amplitude pulses due to cracking of the soil (Fig. 4, on the right) disappear.

The developed algorithms for converting pulse streams into messages allow us to observe anomalies in the 3D-dimensional representation of sequentially measured alphabets. An example of the dynamics of the EMR signal alphabets is shown in Fig. 5. It shows anomalous changes in the dimension of the alphabets of EMR signals received at the Karymshina station of the IKIR FEB RAS (Kamchatka) on July 4, 9, 16, 17, 18, 22 and 23, 2018. Anomalies are observed against the background of diurnal fluctuations in the alphabets dimensions and are associated with recorded earthquakes. For comparison Fig. 6 shows a graph of the dynamics of the EMR signal dispersion. Obviously, the absence of energetic signs of the seismic events occurrence.

In the course of the analysis, it was important to obtain an estimate of the emission signal streams informativity. This estimate can be obtained by calculating the relative information entropy of the resulting alphabets. The results of the calculation of the EMR signal dispersion in July 2018.

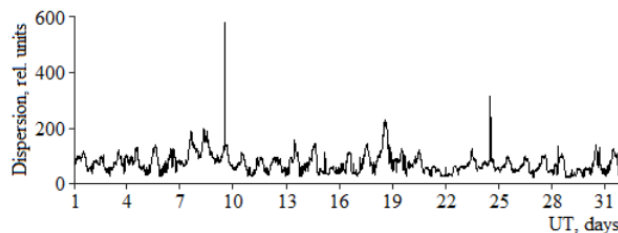


Fig 6. The results of the calculation of the EMR signal dispersion in July 2018.

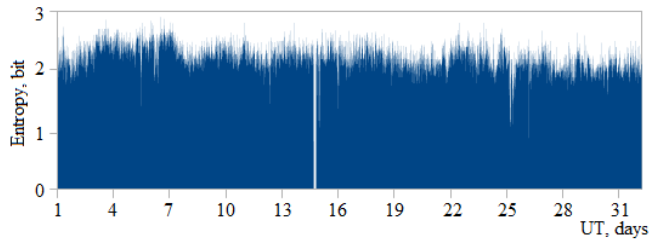


Fig 7. The results of calculating the informational entropy values of the EMR signal alphabets in July 2018.

An example of the information entropy graph for EMR data registered in July 2018 is shown in Fig. 7. The values of the relative information entropy indicator are calculated as the ratio of information entropy in the current 15-minute measurement interval to the entropy value obtained if the alphabet has a uniform distribution of the symbol occurrence frequency along the entire axis of the alphabet dimensions in the same time interval. The results of the information entropy calculation show insignificant increases in the level during the period of increased seismic activity in the Kamchatka region in July 2018.

3 Conclusion

The symbolic description of impulse signals allows us to transfer processing and analysis from classical numerical computation to the field of processing and analysis of code sequences, making available the applicability of linguistics, information and meaning search in the form of identifying hidden rules and grammatical forms of metalanguage. The method of symbolic description of pulses of geophysical signals expands the technical base of data preprocessing tools for analyzing the state of dynamic systems based on quantitative and qualitative characteristics of signals. The results of an experimental verification of the method of symbolic description of pulsed signals showed that for the correct operation of the algorithms it is necessary to ensure the minimum influence of noise interference as a distortion factor of the initial pulse shapes. Therefore, at the first stages of processing it is important to ensure the elimination of errors of the second kind, which will lead to the appearance of false clusters. Further research involves the development of methods for detecting the anomalous behavior of emission geophysical signals, including an attempt to relate the identified anomalies to seismic events, as well as the use of symbolic description and entropy analysis apparatuses as a basic tool with verification of their reliability and significance.

The research was supported by Russian Science Foundation (project No. 18-11-00087).

References

1. V.N Chebrov, V.A. Saltykov, Yu.K. Serafimova, Earthquake Prediction in Kamchatka, 304 (2011)
2. G.A. Sobolev, A.V. Ponomarev, Earthquake Physics and Precursors, 270 (2003)
3. Yu.V. Marapulets, B.M. Shevtsov, Mesoscale acoustic emission, 126, (2012)
4. V.N. Shuman, Yu.A. Bogdanov, Geophys. J., **30**(6), 39-50 (2008)

5. V.A. Gordienko, T.V. Gordienko, N.V. Krasnopistsev, A.V. Kuptsov, I.A. Larionov, Yu.V. Marapulets, A.N. Rutenko, B.M. Shevtsov, *Acoustical Physics*, **54**(1), 82-93 (2008)
6. Yu.I. Senkevich, V.A. Duke, M.A. Mishchenko, A.A. Solodchuk, *E3S Web of Conf.*, **20**, 02012 (2017)
7. J. Serra, *Image Analysis and Mathematical Morphology*, 610 (1982)
8. Yu.P. Pytiev, *Tasks of morphological image analysis*, in *Digest Mathematical Methods for Studying the Earth's Natural Resources from Space* (1984)
9. G. Nikolis, I. Prigogine, *Exploring Complexity*, 512, (1990)
10. Yu.I. Senkevich, Yu.V. Marapulets, O.O. Lukovenkova, A.A. Solodchuk, *SPIIRAS Proceedings*, **18**(5), 1066-1092, (2019)