# Overview of processing and analysis methods for pulse geophysical signals

Olga Lukovenkova<sup>1</sup>, Yury Senkevich<sup>1</sup>, Alexandra Solodchuk<sup>1\*</sup>, and Albert Shcherbina<sup>1</sup>

<sup>1</sup>Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS, Laboratory of Acoustic Research, Mirnaya str., 7, Paratunka, Kamchatskiy kray, 684034, Russia

Abstract. The paper discusses the processing and analysis methods for the geoacoustic and electromagnetic emission pulse signals recorded for more than 20 years at the IKIR FEB RAS geodynamic proving ground (Kamchatka Peninsula). The methods for pulse detection, waveform reconstruction, pulse time-frequency analysis using adaptive sparse approximation, structural description of pulse waveforms and pulse classification are proposed. To detect pulses, the adaptive threshold scheme is used. It adjusts to the noise level of a processed signal. To analyze time-frequency structure of the pulses, the adaptive matching pursuit algorithm is used. To identify pulse waveform, the structural description method is proposed. It encodes pulses with special image matrices. The method of the identified pulses classification is considered. Since the methods for pulse structure analysis are sensitive to noise and distortions, the authors propose the method for pulse waveform reconstruction based on wavelet filtering. The geophysical signal information features determined during the analysis can be used to search for anomalies in the data, and then establish a relationship between these anomalies and deformation process dynamics, in particular, with earthquake development processes.

# **1** Introduction

Continuous monitoring of geophysical signals at various stages of the deformation process has been carried out for more than 20 years at the IKIR FEB RAS geodynamic polygon in the seismically active region (Kamchatka Peninsula). The dynamics of pulse flow parameters of the recorded geophysical signals can be considered as an indicator of the stress-strain state of the geophysical system that generates these signals [1-4].

A variety of waveforms and a wide dynamic range of frequencies and amplitudes of signals, as well as strong noise from both natural and industrial noises significantly complicate the analysis of pulse geophysical signals using well-known statistical, spectral, energy, and other methods. Therefore, it is proposed to analyze geophysical signals in two directions. The first is to classify pulses by amplitude-time form in order to reduce pulse diversity and identify characteristic and anomalous forms; the second is to use sparse approximation to study the features of the time-frequency structure of geophysical signal

<sup>\*</sup> Corresponding author: aleksandra@ikir.ru

<sup>©</sup> The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

pulses. The methods considered in the paper are used for the analysis of geoacoustic and electromagnetic (electrical component of sferics) signals.

## 2 Signal model

In a general case, the studied geophysical signals can be mathematically described as a sum of noise  $\varepsilon(t)$  and some function s(t), the analytical expression of which is unknown:

$$\mathbf{x}(t) = \mathbf{s}(t) + \mathbf{\varepsilon}(t), \quad \|\mathbf{\varepsilon}(t)\| \ll \|\mathbf{s}(t)\|, \tag{1}$$

where  $\varepsilon(t)$  describes nonstationary background noise containing noises of natural and artificial origin and s(t) is a «useful» signal. Signal s(t) is a flux of pulses different waveform, amplitude and frequency. Moreover, we accept the following constraint  $||s(t)|| > ||\varepsilon(t)||$ . Therefore, s(t) can be represented as:

$$s(t) = \sum_{i} A_i \cdot g_i(t - \tau_i), \qquad (2)$$

where  $g_i(t)$  is a function describing the *i*-th pulse;  $A_i$  is the amplitude of the *i*-th pulse;  $\tau_i$  is the generation time of the *i*-th pulse. Then the recorded geophysical signal is represented by an additive model (3):

$$x(t) = \sum_{i} A_{i} \cdot g_{i}(t - \tau_{i}) + \varepsilon(t).$$
(3)

Fig. 1 shows the graph of a geoacoustic signal fragment.



Fig. 1. Geoacoustic emission signal fragment: (a) – noise, (b), (c) – pulses.

#### **3 Pulse waveform reconstruction**

In practice, a signal waveform is affected by natural noises, artificial interference, nonlinearity of the receiving path, limited dynamic range, etc. Distortions of geophysical signal waveforms negatively affect the quality of signal processing and analysis, so they must be preliminarily eliminated.

To reduce noise and to detect signal useful component, wavelet filtration was applied. This technology has proven itself well in geophysics and it is quite effective in the analysis of highly noisy signals [5-7]. Denoising is performed in three stages:

1) forward wavelet transform, as a result of which a signal is represented as the sum of one approximating and N detailing components;

2) modifying each detailing component based on the selected threshold scheme;

3) inverse wavelet transform, as a result of which the signal is reconstructed from the approximating component of the level N and the modified detailing components of levels 1 to N.

To select a wavelet family, a threshold scheme, and an algorithm for calculating the threshold value, a computational experiment was carried out on model and real geophysical signals. The best results were obtained using fourth-order Symlet wavelets (maximum possible level of decomposition), the posterior median rule, and the Empirical Bayes method [8] to calculate the threshold value. An example of noisy signal waveform reconstruction is shown in Fig. 2.



Fig. 2. Waveform reconstruction of noisy geoacoustic signal.

#### 4 Pulse detector

To detect geophysical signal pulses, an algorithm based on calculating an adaptive threshold in disjoint windows of length n is proposed. The current threshold value is calculated by averaging the values of the previous n samples using the formula:

$$S_k = \overline{M_{k-1}} + B \cdot \sigma_{k-1},\tag{4}$$

where  $S_k$  is the current window threshold,  $\overline{M_{k-1}}$  and  $\sigma_{k-1}$  are the mean value and standard deviation of the previous *n* sample amplitudes, *B* is the experimentally determined parameter.

The sequence of n samples is non-contiguous; the samples that are part of the pulse are excluded from it. This ensures that the threshold only depends on the background level. It was found that for the signals under study, the parameter n value lies in the range from 200 to 400 samples. A smaller value of the n parameter increases the number of detector false responses (type I errors); a larger value of n increases the number of missed targets (type II errors).

It was experimentally established that the parameter B optimal value is in the range from 2.1 to 2.5. As the parameter B decreases, false responses become more frequent; when parameter B increases, it misses the target. To increase the number of detected pulses, it is

advisable to slightly lower this parameter by entering an additional check of the pulse waveform. To do this, the pulse is divided into 3 equal parts, within which the average amplitude is determined. The signal under study is considered a pulse if the average amplitude of one of the parts exceeds the others by more than 1.2 times. This check is based on the fact that the pulses have a sharp and very short edge. In this case, the maximum pulse amplitude must exceed the  $S_k$  threshold value by at least 1.8 times, and the minimum pulse duration is 0.1 ms.

Examples of threshold adaptation are shown in Fig 3. The algorithm adapts the threshold value on both the geoacoustic emission signal and the electromagnetic radiation signal.



**Fig. 3.** Geoacoustic (a) and electromagnetic (b) emission signals. The dotted line indicates the adaptive pulse detection threshold.

For more precise pulse detection, a "logical filtering" procedure is used [9]. A signal fragment identified by the detector threshold scheme as a pulse is checked for the integrity of its local maxima and minima sequence. If the length of the continuous sequence of local extrema found in this fragment belongs to a specified interval from  $N_{min}$  to  $N_{max}$  (determined empirically, the minimum value is  $N_{min} = 3$ , in practice the length of the local extrema sequence ranges from 8 to 200 or more), then this signal fragment is taken as the required pulse. Conversely, if the number of continuously following one after another local extrema goes beyond the specified interval, the counts of the signal fragment are set equal to zero. An example of detected geophysical signal pulses is shown in Fig. 4.



Fig. 4. Geoacoustic (a) and electromagnetic (b) pulses detected by the algorithm.

#### 5 Analysis of pulse time-frequency composition

To analyze time-frequency structure of geophysical signal pulses, it is proposed to use a sparse approximation [10]. According to the main idea of this approach, a signal is represented in the form of a finite linear combination of elementary functions selected from a large, generally linearly dependent set of functions. The difference from a simple approximation is that not all functions are included in the decomposition, but only some.

In contrast to the classical time-frequency analysis methods, which generate redundant signal decompositions that include all basis functions, the sparse approximation builds compact signal representations without losing accuracy.

The sparse approximation problem is reduced to the signal representation in the linear combination form of the minimum possible number of basis functions:

$$s(t) = \sum_{m=0}^{N-1} a_m g_m(t) + R_N(t),$$
  

$$\|R_N\| \to \min,$$
  

$$\|a\|_0 \to \min.$$
(5)

where s(t) is the signal,  $g_m(t)$  are basis functions from a preselected dictionary  $D = \{g_m(t): \|g_m\| = 1\}$ ;  $a_m$  are coefficients of decomposition; N is a number of components;  $R_N$  is an error of decomposition;  $\|\cdot\|_0$  is the pseudo-norm ( $L_0$ -norm) that is equal to the number of nonzero elements of the coefficient vector.

To solve the sparse approximation problem, it is proposed to use the matching pursuit (MP) algorithm [11].

To assess the constructed approximations accuracy, the error indicator  $ERR_N$  is calculated using the formula (6):

$$ERR_N = \frac{\left\|R_N\right\|}{\left\|s\right\|} \cdot 100\%.$$
(6)

To approximate signals, the authors propose to use parametric dictionaries (sets of functions into which a signal is decomposed), consisting of functions described by a general analytical expression and a small set of parameters. The main advantages of using parametric dictionaries are:

- the ability to describe signals based on the physical prerequisites for their generation;

- saving spatial computing resources: a small set of parameters can be stored for a parametric dictionary, in contrast to numerically defined dictionaries, for which it is required to store the waveform of each basis function.

Various analytical functions were used to form parametric dictionaries. As a result of the experiment [12], it was found that for the studied geophysical signals containing pulses similar in waveform, it is required to use dictionaries composed of asymmetric functions that allow frequency variation. The Gauss and Berlage functions provide the greatest accuracy.

The main disadvantages of the MP algorithm are, firstly, the need to use huge dictionaries to ensure sufficient accuracy of the decompositions, and secondly, the "rough" discretization of the basis functions parameters. Therefore, it was proposed to improve the classical algorithm so that it was possible to construct decompositions of the required accuracy on dictionaries of limited size. Since at each iteration of the algorithm the parameters of the atom with the largest scalar product by the signal are determined, the matching pursuit iteration can be described as the problem of finding the maximum of a function of many variables:

$$F(\tau, \mathbf{p}) = \left\langle g(t - \tau, \mathbf{p}), R_i(t) \right\rangle \to \max_{\mathbf{p}}.$$
(7)

The MP algorithm was supplemented with a procedure for setting the parameters p of the function that has the greatest correlation with the signal. The developed algorithm was named Adaptive Matching Pursuit (AMP). Refinement can be done in different ways. During a series of experiments, it was proved that using of the grid search method with a separate frequency refinement allows the authors to construct the most accurate approximations [13].

To visualize the pulse time-frequency structure, the Wigner-Ville distribution is used, because its time-frequency resolution minimally depends on the duration of the analyzed signal. However, since the transformation is non-linear, the representation shows a strong interference component for additive signals. Therefore, it was decided to calculate the Wigner-Ville transforms separately for each of the basis functions included in the decomposition (5) and to superimpose their representations on the time-frequency plane on top of each other.

The results of sparse approximation of electromagnetic and geoacoustic pulses by the AMP algorithm are shown in Fig. 5.



Fig. 5. Time-frequency structure of electromagnetic (a) and geoacoustic (b) pulses.

#### 6 Waveform analysis

In addition to the different time-frequency composition, the geophysical signal pulses are characterized by a wide variety of waveforms. To identify the waveform and to classify pulses (in order to reduce their diversity), a method of structural description is proposed. It consists in coding pulses with special matrices [9]. According to the method, the pulse waveform is described through the relative position of its singular points (local extrema). It is assumed that the intermediate values of the signal function lying between the singular points are not informative enough, and during the analysis they can be ignored.

As an example, we consider a typical geoacoustic pulse. We fix the values of the pulse local extrema amplitudes  $\{a_i\} = a_0, a_1, ..., a_N$  and the time intervals between neighboring extrema  $\{\tau_i\} = \tau_0, \tau_1, ..., \tau_{N-1}$  (Fig. 6).



Fig. 6. Pulse local extrema: local maxima are marked with red dots, local minima are marked with blue dots.

Pulse amplitude-phase structure can be described by a square matrix of the form (8):

$$\mathbf{D} = \begin{pmatrix} r_{0,1} & r_{0,2} & \cdots & & & r_{0,N} \\ \omega_{0,1} & r_{1,2} & \cdots & & & r_{1,N} \\ \omega_{0,2} & \omega_{1,2} & \cdots & & & r_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \omega_{0,N-2} & \omega_{1,N-2} & \cdots & & & r_{N-2,N} \\ \omega_{0,N-1} & \omega_{1,N-1} & \cdots & & & r_{N-1,N} \end{pmatrix},$$
(8)

$$r_{i,j} = \begin{cases} 1, \ a_i > a_j \\ 0, \ a_i \le a_j \end{cases}, \qquad \omega_{i,j} = \begin{cases} 1, \ \tau_i > \tau_j \\ 0, \ \tau_i \le \tau_j \end{cases}$$
(9)

where  $r_{i,j}$  is the result of comparison of the *i*-th and *j*-th values of the extreme amplitudes;  $\omega_{i,j}$  is the result of comparison of the *i*-th and *j*-th values of the intervals between the extrema.

We call the matrix  $\mathbf{D}$  (8) a pulse descriptive matrix. Examples of a pulse and its descriptive matrix are illustrated in Fig. 7.



Fig. 7. Representation of a pulse by descriptive matrix: (a) - pulse with detected extrema; (b) - its descriptive matrix.

Descriptive matrices have invariant character under time and amplitude transformations that is proved by the following properties:

- invariance under time compression or expansion (if k > 0 and  $\tau_i > \tau_i$ , then  $\tau_i \cdot k > \tau_i \cdot k$ );

- invariance under amplitude compression or expansion (if k > 0 and  $a_i > a_j$ , then  $a_i \cdot k > a_j \cdot k$ ).

Pulses are classified accorded to the results of calculation of descriptive matrix similarity coefficient. First, we restrict ourselves to descriptive matrices of the same order. In order to do that we divide descriptive matrices into groups  $\mathbf{D}^{(Z)}$  of matrices of the same order Z.

For the two descriptive matrices of the same order  $\mathbf{D}_1^{(Z)}$  and  $\mathbf{D}_2^{(Z)}$  the similarity coefficient  $g(\mathbf{D}_1^{(Z)}, \mathbf{D}_2^{(Z)})$  is determined as a ratio of elements, coinciding in value and taking the same positions, to the total number of elements in the matrix. In case of exact identity, the similarity coefficient is 1.

Classification is the comparison of descriptive matrices of the same size and selection of the empirical threshold  $G_0$ ,  $0 < G_0 \le 1$  when calculating the similarity coefficient (10).

$$g(\mathbf{D}_{1}^{(Z)}, \mathbf{D}_{2}^{(Z)}) = \frac{\#(d_{1ij} = d_{2ij})}{Z^{2}} > G_{0},$$
(10)

where #(*condition*) is the operator for calculation of the number of elements satisfying the condition,  $d_{1ij}$  are the elements of the matrix  $\mathbf{D}_1^{(Z)}$ ,  $d_{2ij}$  are the elements of the matrix  $\mathbf{D}_2^{(Z)}$ ,  $G_0$  is the empirically defined threshold.

When calculating the similarity coefficient of descriptive matrices of different orders, it should be taken into account that their erratic comparison can lead to random absorption of smaller matrices by bigger matrices. We put bounds on the possibility for absorption of descriptive matrices of smaller orders by the threshold coefficient  $S_0$ :

$$N_L/N_M \ge S_0, \quad 0 < S_0 \le 1,$$
 (11)

(11)

where  $N_L$  and  $N_M$  are the orders of smaller and larger descriptive matrices, respectively,  $S_0$  is the empirically defined threshold.

Assume that there are two descriptive matrices  $\mathbf{D}_1$  and  $\mathbf{D}_2$  of the orders  $N_1$  and  $N_2$ , respectively  $(N_1 \ge N_2)$ . If  $N_2/N_1 \ge S_0$   $(N = S_0 \cdot N_1)$ , then the matrices are comparable. For the meaningful comparison, the matrices  $\mathbf{D}_1$  and  $\mathbf{D}_2$  are matched so that the latest N elements of the main diagonal  $\mathbf{D}_2$  are overlapped on the first N elements of the main diagonal  $\mathbf{D}_1$ . Then we calculate the similarity coefficient (10) of matrices of the order N in the domain bounded by the intersection. Elements are further compared after  $\mathbf{D}_2$  shift along the main diagonal  $\mathbf{D}_1$ . The similarity coefficient (10) is calculated for each subsequent shift so far as the matrices in the domain bounded by the intersection have the order of not less than N (Fig. 8). The comparison result is the maximum value of the similarity coefficient.

$\mathbf{D}_1: N_1 = 6$							<b>D</b> <sub>2</sub> : $N_2 = 5$										
1	0	1	0	1	1			1	1	1	1	0	5	$S_0 =$	0.7		
0	1	1	1	1	0	1		0	0	1	0	1	1	$V_2 / .$	$N_{1} >$	0.7	
0	0	0	0	0	1			0	1	1	1	0	1	V =	[0.7	·6]	= 4
0	0	1	1	1	1			0	1	0	0	0					
0	0	1	0	0	1			0	0	0	1	1					
0	0	0	0	1	1												
1	1	1	1	0				1		1			1				
0					1	1		0		1		1	0				
0		1	1		1	0	1	0				0	1				
0		0	0	0	0	1		0					1				
0	0	0	1	1	1	1		0	0				1				
	0	0	1	0	0	1		0	0	0	0	1	1				
	$0 0 0 1 1 1$ $g_2 = 10/25 = 0.4$																
$g_1 = 9/16 = 0.5625$																	
1	0	1	0	1	1	]		1	0	1	0	1	1				
0	1	1	1	1	0			0	1	1	1	1	0	1			
0	0	0		0	1			0	0				1	0			
0	0	1	1	1				0	0			1		1			
0	0	1	0	0				0	0				1	0			
0	0	0	0	1	1			0	0	0				0			
_	$g_3 =$	22	/ 25	= (	.88					0	0	0	1	1			
	- /							,	$z_{4} =$	4/	16 =	= 0.2	25				

Fig. 8. Comparison of descriptive matrices of orders 6 and 5.

In the result of classification many classes are formed. Each of them unites the pulses close in amplitude-phase structure with the accuracy up to the values determined by the coefficients  $G_0$  and  $S_0$ . Efficiency of the proposed classification method can be estimated by a contraction coefficient

$$k = \frac{\left|A_{input}\right|}{\left|A_{output}\right|},\tag{12}$$

where  $|A_{input}|$  is the alphabet cardinality (number of pulses) before classification,  $|A_{output}|$  is the cardinality of the alphabet obtained as the result of classification.

Processing of the data archive of IKIR FEB RAS for 2016–2019 showed that contraction coefficient (12) varies from 50 to 200.

The results of a computational experiment for evaluating the classification algorithm noise immunity show that the classification method is nonlinearly sensitive to the choice of threshold functions of  $G_0$  and  $S_0$ : decrease of  $S_0$  causes pulse redistribution into classes with larger-order descriptive matrices and decrease of  $G_0$  makes it possible to compact classes. Based on the experiment results, the recommended values of the thresholds  $S_0$  (from 0.6 to 0.9) and  $G_0$  (from 0.7 to 0.9) were selected.

Fig. 9 illustrates the results of classification of model signal pulses with the signal/noise ratio of 8.9 dB. The signal contains 50 pulses of three different waveforms (in the figure, the pulses of one waveform are marked with the same color). Before being classified, each

detected pulse was denoised by the selected method of wavelet filtering. Classification algorithm thresholds were selected as  $S_0 = 0.6$  and  $G_0 = 0.7$ . For a noisy signal (Fig. 9a) we succeeded to define correctly the number of classes and to structure the pulses into classes (Fig. 9b). In the course of multiple repetition of the experiment it was discovered that the obtained structuring into the classes is stable to the changes of initial conditions of white noise generator.



**Fig. 9**. Classification results: (a) – signal with overlapped white noise and initial structuring into classes; (b) – classification of  $S_0 = 0.6$ ,  $G_0 = 0.7$ ; three classes were defined.

## 7 Conclusions

The paper discusses the processing and analysis methods for the geoacoustic and electromagnetic emission pulse signals recorded for more than 20 years at the IKIR FEB RAS geodynamic proving ground (Kamchatka Peninsula). The geophysical signal pulses are characterized by a wide variety of waveforms and different time-frequency structures.

Waveforms of the recorded signals changes under the influence of natural noise, artificial interference, nonlinearity of the receiving path, etc. Therefore, to reduce noise and to detect signal useful component, wavelet filtration is applied. It was established that the best results of pulse waveform reconstruction are obtained using fourth-order Symlet wavelets.

To detect geophysical signal pulses, the algorithm based on calculating the adaptive threshold is used. For more precise pulse detection, a "logical filtering" procedure is used. It consists in checking the integrity of the local maxima and minima sequence of the detected pulse.

The analysis of geophysical signals is carried out in two directions. The first direction is the analysis of their time-frequency structure. To do this the Adaptive Matching Pursuit algorithm is used. As the result of its application, geophysical signal is represented in the form of a finite linear combination of elementary functions. It was found that for sparse approximation of the studied geophysical signals, it is required to use asymmetric functions that allow frequency variation, such as the Gauss and Berlage functions.

The second direction is the analysis of pulse waveforms. To identify the waveform, the method of structural decomposition is used. It consists in coding pulses with special descriptive matrices. In order to reduce the variety of waveforms, pulses are classified by comparing their descriptive matrices. The classification procedure makes it possible to reduce the pulse diversity by at least 50 times.

The geophysical signal informational features determined during the analysis can be used to search for anomalies in the data, and then establish a relationship between these anomalies and deformation process dynamics, in particular, with earthquake development processes.

The work was carried out according to the Subject AAAA-A17-117080110043-4 "Dynamics of physical processes in the active zones of near space and geospheres".

### References

- 1. G.I. Drugin, Yu.V. Marapulets, N.V. Cherneva, A.Yu. Isaev, A.A. Solodchuk, Dokl. Earth Sci., **472**(2), 215-219 (2017)
- 2. O.P. Rulenko, Yu.V. Marapulets, Yu.D. Kuzmin, A.A. Solodchuk, Izv.-Phys. Solid Earth, **55**(5), 766-776 (2019)
- O. Lukovenkova, A. Solodchuk, A. Tristanov, E. Malkin, E3S Web. Conf., 127, 03001 (2019)
- S.A. Pulinets, V.G. Bondur, M.N. Tsidilina, M.V. Gaponova, Geomagn. Aeronomy, 50(2), 231-242 (2010)
- L. Alperovich, L. Eppelbaum, V. Zheludev, J. Dumoulin, F. Soldovieri, M. Proto, M. Bavusi, A. Loperte, J. Geophys. Eng., 10(2), 1-17 (2013)
- A.A. Solodchuk, Vestnik KRAUNTs, Ser. Fiziko-matematicheskie Nauki, 2, 69-73 (2012)
- O.V. Mandrikova, I.S. Solovjev, S.Yu. Khomutov, D.G. Balishev, V.V. Geppener, D.M. Klionskiy, Pattern Recognition and Image Analysis, 26(4), 773-781 (2016)
- 8. I.M. Johnstone, B.W. Silverman, Ann. Statist., **32**(4), 1594-1649 (2004)
- M.I. Gapeev, Yu.I. Senkevich, O.O. Lukovenkova, A.A. Solodchuk, J. Phys.: Conf. Ser., 1368, 052034 (2019)
- Y.V. Marapulets, O.O. Lukovenkova, A.B. Tristanov et al., *Methods for recording and for time-frequency analysis of geoacoustic emission signals* (Dalnauka, Vladivostok, 2017)
- S. Mallat, Z. Zhang, IEEE Transactions on Signal Processing, 41(12), 3397-3415 (1993)
- A. Tristanov, O. Lukovenkova, Yu. Marapulets, A. Kim, *Improvement of methods for* sparse model identification of pulsed geophysical signals, in Proceedings of the conference Signal Processing: Algorithms, Architectures, Arrangements, and Applications, SPA-2019, 18-20 September 2019, Poznan, Poland (2019)
- O. Lukovenkova, Yu. Marapulets, A. Tristanov, A. Kim, E3S Web of Conf., 62, 02012 (2018)