Application of adaptive wavelet thresholding to recovery geoacoustic signal pulse waveforms

Olga Lukovenkova[1,*]

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

Abstract. Recorded geoacoustic signals often contain noise and interference. Their appearance is caused by various reasons, e.g. of propagation environment heterogeneity, weather condition influence, human activity, etc. So, geoacoustic emission signals contain a persistent background noise that changes in intensity over time. This noise significantly distorts the geoacoustic pulse waveforms and thus complicates analysis of the signal characteristics. The article presents results of estimating the geoacoustic signal background noise. On the basis of these estimates, a method of adaptive wavelet thresholding is proposed to remove noise from the signal and recovery the single pulse waveforms. In conclusion, the results of a computational experiment are presented. They confirm effectiveness of using the chosen method for the geoacoustic signal preprocessing. The work was carried out as part of the implementation of the state task AAAA-A21-121011290003-0.

1 Introduction

Some of the most common problems faced by researchers in applied science are distortion and noisiness of recorded signals. These problems are typical for different signals, e.g. for acoustic, electrical, electromagnetic ones. In practice, a signal waveform is influenced by natural noise, artificial interference, nonlinearity of receive path, dynamic range limitations, quantization errors, primary hardware processing, etc. At the same time, many promising analysis methods (time-frequency transformations, sparse approximation, classification, clustering, etc.) are sensitive to signal waveform distortions. It negatively affects quality and accuracy of analysis results.

The presented article is devoted to denoising of geoacoustic emission signals (Fig. 1*a*).

It is assumed that the methods of geoacoustic signal secondary processing and analysis are applied to undistorted or slightly distorted pulses [1]. Therefore, it is important to denoise the signal before its processing.

2 Estimation of background noise parameters

The author studied the geoacoustic background noise before choosing a denoising method. For this purpose, fragments without pulses were selected from the geoacoustic signal recorded in good weather conditions (no precipitation and strong wind) at "Karymshina"

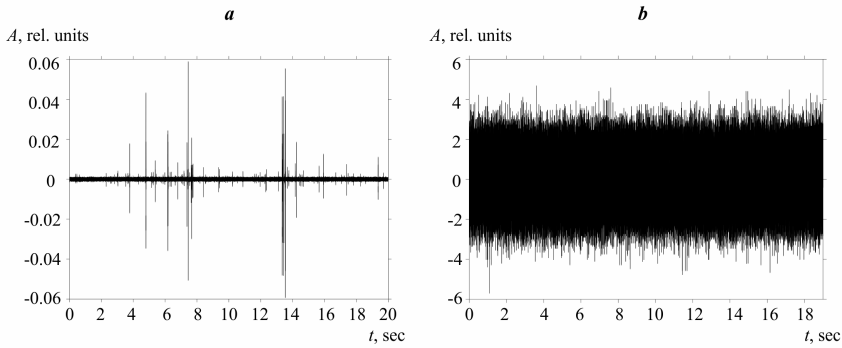
*e-mail: o.o.lukovenkova@yandex.ru

Figure 1. Geoacoustic emission signals: *a* — pulsed signal; *b* — background noise.

site (52.83° N, 158.13° E). Thus, a signal containing only background noise was gener-ated (Fig. 1*b*). This signal was scaled and shifted so that the mean was $\mu = 0$ and the standard deviation was $\sigma = 1$.

There are several ways to estimate the distribution law of a random variable. The simplest ones used at the first stage of the study are a histogram and a quantile-quantile plot (Q-Q plot). Using them, we can visually assess similarity of the distribution with one of the known laws. Figure 2a shows histogram of the noise signal amplitude distribution. The number of intervals k is chosen according to the Sturgess rule

$$k = 1 + \log_2 N,$$

where N is a number of random variable measurements. Figure 2*b* shows Q-Q plot which compared the noise signal amplitudes on the vertical axis to a standard normal distribution values on the horizontal axis.
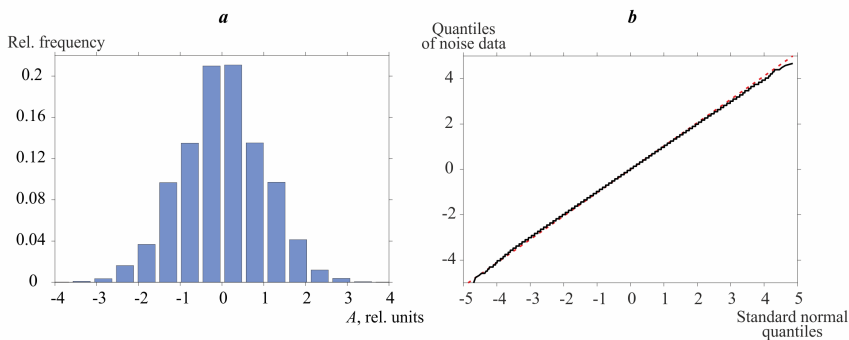

Figure 2. Estimation of noise signal amplitude distribution: *a* — histogram; *b* — Q-Q plot.

Visually, the distribution of noise signal amplitudes is similar to the normal one. How-ever, the normality hypothesis H_0 requires confirmation. The results of testing the statistical hypothesis H_0 (the noise signal amplitudes have the normal distribution $N(\mu, \sigma)$, where μ and σ are estimated from the tested data) using Pearson's chi-squared test [2], Anderson–Darling test [3] and Lilliefors test [4] are given in Table 1. H_1 is an alternative hypothesis.

Table 1. Results of applying various statistical tests.

Test	Significance level, α	Sample size, N	Accepted hypothesis, H_0 or H_1
Pearson	0.05	500	H_0
	0.05	1000	H_0
Anderson–Darling	0.05	500	H_0
	0.05	1000	H_1
Lilliefors	0.05	500	H_0
	0.05	1000	H_1

When the sample size increases, most of the statistical tests reject the hypothesis H_0. This effect can be explained by quantization errors of the geoacoustic signal. The analog-to-digital converter used in the signal registration system has a resolution of 16 bits (65536 levels, values from -1 to 1). According to calculations, dynamic range of the background noise signal contains 112 levels (values from -0.00189 to 0.001526). It is supported by the Q-Q plot (Fig. 2b), its points form a stepped line.

To test this assumption, a Gaussian noise signal was generated. The signal was digitized so that its dynamic range contains 112 levels. The distributions of amplitudes of the real and model signals were compared using the Wilcoxon–Mann–Whitney test [5].

At significance level $\alpha = 0.05$ for sample sizes $N = 500$ and $N = 1000$, the Wilcoxon–Mann–Whitney test confirmed the hypothesis H_0 (the distributions of both samples are equal). This result is invariant to changes of the Gaussian noise generator seed. The conducted research shows that the background noise of geoacoustic signal is similar in structure to the roughly digitized Gaussian noise.

3 Wavelet thresholding as denoising method

A wavelet thresholding is one of the popular methods for removing noise, including Gaussian one, from a signal and separating an informative signal component. There is a large number of studies showing effectiveness of the wavelet thresholding method to remove high-frequency noise from signals of different nature [6–9]. The wavelet thresholding process consists of three stages: forward wavelet transform, thresholding, and inverse wavelet transform.

Denoising consists in thresholding of detail wavelet coefficients. The most common are soft and hard thresholds [10]. There are various strategies for choosing a threshold value: the Universal threshold [10], the Stein Unbiased Risk Estimator (SURE) [11], the Block James-Stein estimator (BlockJS) [12], the False Discovery Rate threshold (FDR) [13], the Empirical Bayes method [14], the Minimax principle [10], etc.

To select the most appropriate algorithm for calculating a threshold value, a computational experiment was carried out on a model geoacoustic signal. The signal containing 100 pulses with amplitudes from 0.05 to 1 relative units was generated. Gaussian noise of various amplitudes was artificially superimposed on the pulses so that the signal-to-noise ratio (SNR) was varied from -10 to 40 dB. The average relative error of wavelet thresholding was calculated for each SNR value by the formula

$$\varepsilon = \frac{1}{N} \sum_{i=1}^{N} \frac{\|s_i(t) - \hat{s}_i(t)\|}{\|s_i(t)\|}, \tag{1}$$

where $s_i(t)$ is the undistorted pulse; $\hat{s}_i(t)$ is the pulse processed by wavelet thresholding; N is the total number of processed pulses, $N = 100$.

Dependences of the error ε on SNR for the considered threshold calculation methods are shown in Fig. 3.
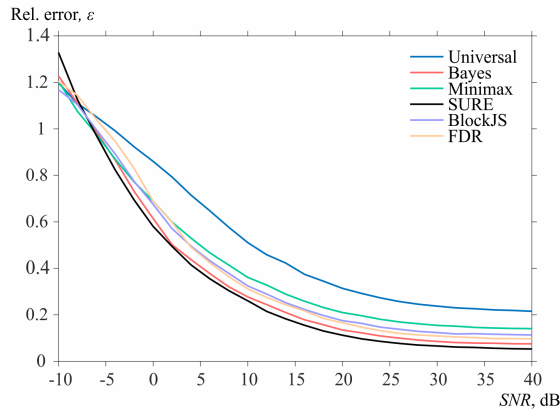

Figure 3. $\varepsilon(SNR)$ plots for different threshold calculation methods.

The best results were obtained using the Empirical Bayes and SURE methods. The Bayes method works more accurately on highly noisy signals (SNR less than -5 dB).

4 Wavelet family selection

Orthogonal and biorthogonal wavelets are most often used in noise reduction tasks. In the presented article, the following wavelet families are considered [15]: the Daubechies wavelets (Fig. 4a), the symlets (Fig. 4b), and the coiflets (Fig. 4c).

Experiment on model data show that the most accurate results are obtained using the fourth-order coiflets, and the high-order Daubechies wavelets and symlets (from the eighth order and higher). Figure 5 shows the $\varepsilon(SNR)$ plots obtained by the Empirical Bayes method using various wavelet families.

5 Approbation of adaptive wavelet thresholding on real data

According to the results obtained on model data, the Empirical Bayes method (the detail wavelet coefficients are estimated by the posterior mean value) and the fourth-order coiflets were chosen for geoacoustic signal recovering and denoising. Figure 6a–e shows the processing results obtained for the geoacoustic pulses of various waveforms (all ones were recorded in January, 2017). The processed signals are smoothed and denoised. The degree of denoising depends on the noise level. For example, the processing of highly noisy pulses is shown in Fig. 6a,c. The processing of undistorted one is shown in Fig. 6b. And Fig. 6d–f shows denoising of pulses with average noise level.

6 Conclusion

Registered geoacoustic signals are often highly distorted and noisy, while the noise level is constantly changing. It significantly complicates the processes of identification, analysis and classification of these signals. Study of the background noise shows that it is similar
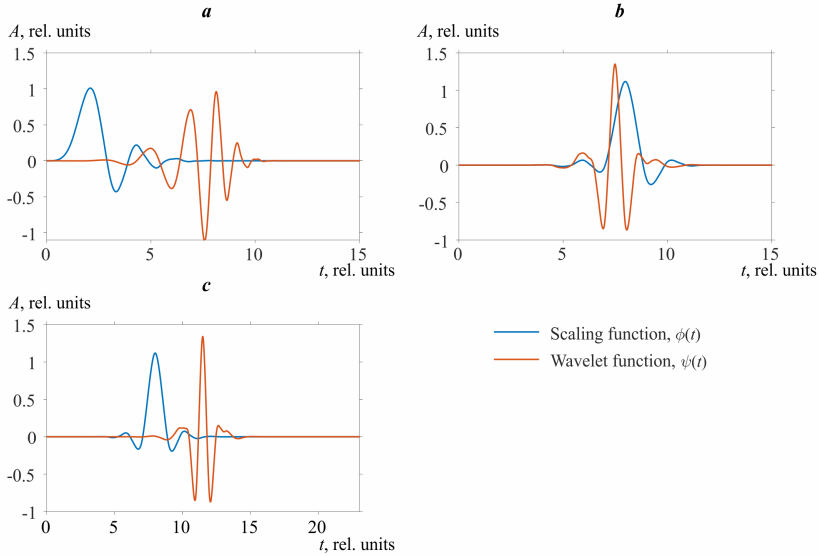
Figure 4. Mother wavelets $\psi(t)$ and scaling functions $\phi(t)$: *a* — eighth-order Daubechies wavelet, db_8; *b* — eighth-order symlet, sym_8; *c* — forth-order coiflet, $coif_4$.
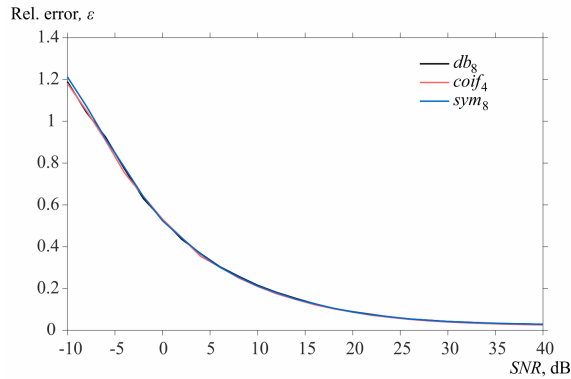

Figure 5. $\varepsilon(SNR)$ plots obtained by the Empirical Bayes method using various wavelet families.

in structure to the roughly digitized Gaussian noise. Therefore, the author proposes to use an adaptive wavelet thresholding as a signal preprocessing method. Experiments on model data show that the best noise reduction quality is achieved using the Empirical Bayes method and the fourth-order coiflets. The results of real data processing confirm effectiveness of the selected method for the geoacoustic signal preprocessing.
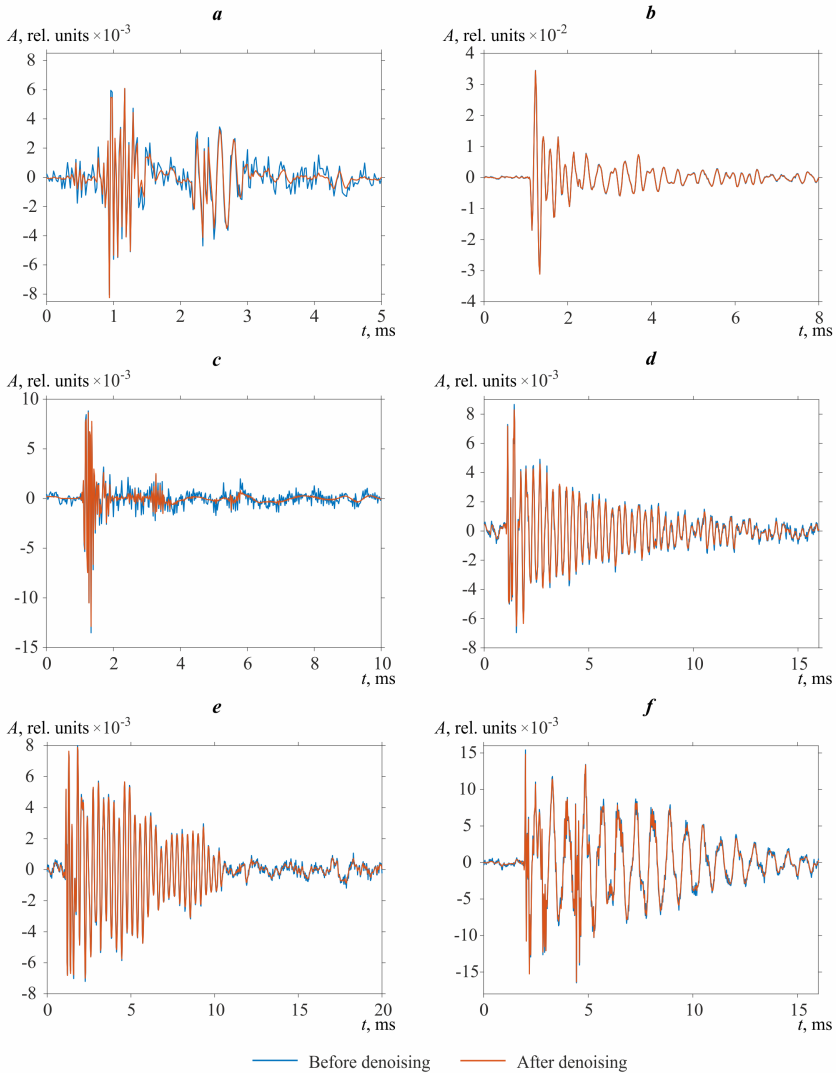
Figure 6. Applying Empirical Bayes method to geoacoustic pulses.

References

[1] Yu. V. Marapulets, Yu. I. Senkevich, O. O. Lukovenkova, A. A. Solodchuk, I. A. Larionov, M. A. Mishchenko, E. I. Malkin, A. O. Shcherbina, M. I. Gapeev, *Complex analysis of acoustic and electromagnetic signals to assess seismic hazard level* (Dalnauka, Vladivostok, 2020) 120 p.

[2] P. E. Greenwood, M. S. Nikulin, *A guide to chi-squared testing* (Wiley, New York, 1996) 280 p.

[3] T. W. Anderson, D. A. Darling, Ann. Math. Stat., **23**, 193–212 (1952)

[4] H. W. Lilliefors, J. Am. Stat. Assoc., **62**, 399–402 (1967)

[5] M. Hollander, D. A. Wolfe, *Nonparametric Statistical Methods* (John Wiley & Sons, Inc., Hoboken, NJ, 1999) 779 p.

[6] T. W. Anderson, D. A. Darling, IEEE Trans. Image Process., **9**, 1532–1546 (2000)

[7] M. Alfaouri, K. Daqrouq, Am. J. Appl. Sci., **5**, 276–281 (2008)

[8] M. Srivastava, C. L. Anderson, J. H. Freed, IEEE Access, **4**, 3862–3877 (2016)

[9] G. Srikar, Ch. R. Prasad, Adv. Comput. Sci. Tech., **10**, 3111–3121 (2017)

[10] G. Luo, D. Zhang, *Wavelet Denoising, Advances in Wavelet Theory and Their Applications in Engineering, Physics and Technology* (IntechOpen, 2012) 24 p.

[11] D. L. Donoho, IEEE Trans. Inf. Theory, **41**, 613–627 (1995)

[12] T. T. Cai, Stat. Sin., **12**, 1241–1273 (2002)

[13] F. Abramovich, Y. Benjamini, D. L. Donoho, I. M. Johnstone, Ann. Stat., **34**, 584–653 (2006)

[14] I. M. Johnstone, B. W. Silverman, Ann. Stat., **32**, 1594–1649 (2004)

[15] I. Daubechies, *Ten lectures on wavelets* (SIAM, Philadelphia, Pa, 1992) 377 p.