Izvestiya, Physics of the Solid Earth, Vol. 40, No. 7, 2004, pp. 587–593. Translated from Fizika Zemli, No. 7, 2004, pp. 50–56.

Application of Wavelet Analysis to the Automatic Classification of Three-Component Seismic Records

A. A. Lyubushin, Jr.*, Z. Kaláb, and N. Častová*****

**Schmidt United Institute of Physics of the Earth, Russian Academy of Sciences, Bol'shaya Gruzinskaya ul. 10, Moscow, 123995 Russia*

***Institute of Geonics (IG), Czech Academy of Sciences (CAS), Ostrava, Czech Republic*

****Ostrava University of Technology, Ostrava, Czech Republic*

Received January 30, 2003

Abstract—A method for automatic classification of three-component seismic records is proposed on the basis of the use of the Donoho–Johnstone wavelet shrinkage level as an informative indicator. The method is exemplified by analysis of a set of three-component records of seismic events in mines of the Silesian coal basin (Czech Republic). The inferred clusters of events supposedly differ in the focal mechanism of an event.

INTRODUCTION

The classification of elements of a large set of seismic records is an important problem in various areas of seismology and seismics. A wide set of statistical approaches based on the ideas of spectral analysis, pattern recognition, syntactical procedures, etc., are traditionally used for its solution [Chen, 1982]. In classification problems, the efficiency critically depends on the choice of the small-dimensional vector of indicators characterizing the object to be classified [Aivazyan *et al.*, 1989; Vapnik and Chervonenkis, 1974; Duda and Hart, 1973]. Therefore, the choice of indicators characterizing a seismic record is crucial to the subdivision of a set of signals into specific subsets, i.e., clusters.

In this paper, we propose using the so-called Donoho–Johnstone wavelet shrinkage level α [Donoho and Johnstone, 1994] as an indicator characterizing a scalar seismic record; the parameter α varies from 0 to 1 and specifies the set of coefficients of the signal expansion in orthogonal finite basis functions (wavelets) that can be rejected without a substantial loss of information on the signal. The α value depends on the choice of the basis of wavelet functions. Therefore, the calculation of the indicator α must be preceded by the determination of the basis in which the signal will be expanded. In this paper, the basis is specified by the method of coherent basis thresholding [Berger *et al.*, 1994; Mallat, 1998], using the criterion of the entropy minimum in the distribution of squared wavelet coefficients in the residual signal obtained by the successive elimination of its most informative components.

At approximately the same noise level in the set of signals analyzed, the indicator α characterizes the "saturation" of the signal with diverse elements of behavior: the nearer the α value to 1, the "simpler" the signal; i.e., it can be adequately described in this case by a smaller number of coefficients of the optimal wavelet basis. Thus, a three-dimensional vector of dimensionless indicators (shrinkage levels of each component) is obtained from each three-component seismic record. The resulting cloud of three-dimensional vectors is then subjected to the standard iterative procedure of cluster analysis minimizing the functional of the division compactness of a set of vectors into a given trial number of clusters [Duda and Hart, 1973]. The optimal number of clusters is determined from the maximum condition of the pseudo-F-statistic [Vogel and Wong, 1978]. Application of this method is exemplified by the classification of a set comprising 111 three-component records of seismic events in mines of the Silesian coal basin in Czechia.

Wavelet analysis [Chui, 1992; Daubechies, 1992; Mallat, 1998] is essentially a more adequate tool for the analysis and classification of nonstationary signals as compared with the traditionally applied Fourier analysis. This is related to the compactness of the basis functions in use, which makes it possible to analyze essentially nonstationary and non-Gaussian signals. In solving many problems of geophysics, the use of wavelets provides a new and fresh insight into the properties and structure of data [Lyubushin, 2000, 2001, 2002; Častová and Kaláb, 1999; Častová *et al.*, 1999].

METHOD OF CLASSIFICATION

Let $c_j^{(k)}$ be the wavelet coefficients of the analyzed signal $x(t)$ ($t = 1, \ldots, N$ is the discrete time) expanded in a system of orthogonal finite basis functions. The superscript k is the number of the detail level of the wavelet expansion, and the subscript j indicates the center of the time vicinity. The greatest possible value m of the detail level number depends on the volume of the sample analyzed. The notation used is described in detail in [Lyubushin, 2000, 2001, 2002]. Here, we used a dictionary of 17 wavelets: 10 Daubechies ordinary orthogonal wavelets ranging in order from 2 to 20 (the use of higher orders entails numerical instability) and 7 so-called "symlets"; the latter are modifications of the Daubechies wavelets in which the form of basis functions is more symmetric than in ordinary wavelets [Chui, 1992; Daubechies, 1992; Mallat, 1998]. Symlets possess the same properties of compactness, orthogonality, completeness, and smoothness as wavelets do; however, for orders of 2 to 6, they coincide with the ordinary orthogonal Daubechies basis, while orders of 8 to 20 reveal distinctions in the form of a basis function. For these reasons, we used 17 variants of orthogonal compact basis functions.

In choosing the optimal wavelet basis, the criterion of the entropy minimum in the distribution of the squared magnitudes of the wavelet coefficients

$$E(x) = -\sum_{k=1}^{m}\sum_{j=1}^{2^{(m-k)}} p_j^{(k)}\ln(p_j^{(k)}) \longrightarrow \min, \qquad p_j^{(k)} = \left|c_j^{(k)}\right|^2 / \sum_{l,i}\left|c_i^{(l)}\right|^2 \tag{1}$$

is commonly used. Method (1) selects a basis for the signal $x(t)$ such that the distribution of the signal wavelet coefficients differs most from a uniform distribution. In this case, maximum information concentrates in the minimum number of expansion coefficients. Usually, the application of criterion (1) yields quite satisfactory results. However, a more sophisticated method used below for the choice of the optimal basis has the form of an iterative procedure repeatedly using criterion (1). This was dictated by the desire to reveal the finest distinctions in signal structures by applying one or another basis. This method was proposed by Berger *et al.* [1994] for solving the problem of removing characteristic noises (hissing, cracks, and clicks) from old vocal recordings of opera classics and was called the method of successive coherent basis thresholding. The method can be briefly described as the sequence of the following operations:

(1) Initialization: the initial signal $x_0(t)$ is moved into the working buffer $x(t)$.

(2) The wavelet order is determined from criterion (1) for the signal $x(t)$: $E(x) \longrightarrow \min$.

(3) The wavelet coefficients $c_j^{(k)}$ of the signal $x(t)$ are rearranged in descending order of their magnitudes for testing the basis determined at stage (2), and the coefficients rearranged in this way are designated as d_j, $j = 0, 1, \ldots, (N-1)$. Thus, the coefficient d_0 has the maximum magnitude.

(4) The minimum integer $M = 0, 1, \ldots, (N-1)$ is determined from the inequality

$$\frac{|d_M|^2}{\sum_{j=M+1}^{N-1}|d_j|^2} \le \frac{2\ln(N-M)}{(N-M)}. \tag{2}$$

(5) If condition (2) is immediately fulfilled at $M = 0$, then the optimal order is found and the algorithm stops operating.

(6) If condition (2) is not fulfilled for any $M = 0, 1, \ldots, (N-1)$, then the optimal wavelet order is set to be equal to the value found at stage (2) from the condition of the entropy minimum immediately after the initialization and the algorithm stops operating.

(7) All coefficients $c_j^{(k)}$ for which $\left|c_j^{(k)}\right| \ge |d_M|$ are set at zero, the inverse wavelet transformation is applied to the remaining coefficients, the resulting residual signal is moved into the working buffer $x(t)$, and stage (2) is executed.

The meaning of this procedure reduces to the following. The signal is considered to consist of a desired signal, whose variations are reflected in the values of wavelet coefficients that are fairly large in magnitude, and noise, accounted for by all other coefficients. The problem is to choose the threshold of coefficient moduli above which the coefficients account for the desired signal and below which they account for the noise. Inequality (2) is precisely intended for determining such a threshold. This condition is taken from the formula for the probability of asymptotic maximum deviations of the Gaussian white noise values $B(t)$ (e.g., see [Korolyuk *et al.*, 1985]):

$$\lim_{N\to\infty}\mathrm{Prob}\left\{\max_{0\le t\le(N-1)}|B(t)|^2 / \sum_{j=0}^{N-1}|B(j)|^2 \le \frac{2\ln N}{N}\right\} = 1. \tag{3}$$

The following formula, immediately resulting from (3), is also used below:

$$\lim_{N\to\infty}\mathrm{Prob}\{\max_{0\le t\le(N-1)}|B(t)| \le \sigma\sqrt{2\ln N}\} = 1, \tag{4}$$

where σ is the standard deviation of the Gaussian white noise $B(t)$.

Therefore, the meaning of condition (2) consists in the division of wavelet coefficients into noisy and useful. Coefficients responsible for noise have rather small absolute values (the lower limit of the sum in (2) is $M + 1$) lying within the asymptotic limits for the white noise (formula (3)). However, such an extraction of noise from the signal depends on the basis used (thus, noise in one basis does not necessarily satisfy criterion (2) in another). Therefore, upon choosing the basis from the entropy minimum condition (stage 2), noise is determined in terms of this (stage 7); an optimal basis is then again determined, this time for the residual (depleted in information) signal (stage 2); and so on, until the residual signal becomes the noise, even with respect to its own optimal basis (stage 5). The last optimal basis determined in this way is considered as optimal because it is capable of recovering at least something from the most depleted residual signal. This approach makes sense because information can be retrieved from the initial signal $x_0(t)$ with the use of any basis, whereas only the best basis is effective in the case of a depleted residual signal. Note that, as regards the seismic records treated below, the basis found from such an iterative procedure coincides in 90% of cases with the basis determined immediately at the initialization stage; however, this cannot be stated a priori for any given signal.

After the optimal wavelet basis is determined for a given signal, we define the set of wavelet coefficients that are smallest in modulus and can be rejected in the inverse wavelet transformation because they account for noise. For this purpose, we assume that noise concentrates mostly in variations at the first, highest-frequency detail level with the exception of a small number of points at which high-frequency features of the desired signal behavior concentrate and to which, consequently, large values of the first-level wavelet coefficients correspond. Due to the wavelet transform orthogonality, the variance of wavelet coefficients is equal to the variance of the initial signal. Therefore, we estimate the standard deviation of the noise σ as the standard deviation of wavelet coefficients at the first detail level. This estimation must be robust, i.e., insensitive to outliers in desired values of wavelet coefficients at the first level. For example, we can use the robust median estimate of the standard deviation for a normal random value [Huber, 1981]

$$\sigma = \mathrm{med}\{|c_j^{(1)}|, j = 1, \dots, N/2\}/0.6745. \tag{5}$$

Now, the estimate σ being found from (5), we can use (4) for estimating the threshold of wavelet coefficient moduli below which they can be set at zero because they are carriers of noise variations. This threshold is equal to $\sigma\sqrt{2\ln N}$. As a result, one can readily determine the Donoho–Johnstone level α for the signal shrinkage, namely, the ratio of the number of coefficients for which the condition $|c_j^{(k)}| \le \sigma\sqrt{2\ln N}$ is fulfilled to their total number N. It is precisely this dimensionless parameter α ($0 < \alpha < 1$) that is used below as an integral characteristic of the seismic signal scalar component, subjected to the procedure of classification.

The calculation of α for each component of three-component seismic records yields a cloud of 3-D vectors ξ inside the unit cube in the positive octant of the space. This set of vectors was subjected to cluster analysis in order to reveal compact groups of points. We applied the ISODATA method [Aivazyan *et al.*, 1989; Duda and Hart, 1973], widespread in cluster analysis. The centers of trial clusters are randomly distributed within the minimum parallelepiped containing the points to be classified ξ; the number $q \ge 2$ of the trial clusters is fixed. The initial random distribution of these clusters is denoted as Γ. For a given distribution of the cluster centers, the set of points is tentatively divided into groups of points nearest to a cluster center. Let c_k, $k = 1, \dots, q$, be the vectors of the cluster centers; n_k, the number of points in the kth cluster; $\sum_{k=1}^{q} n_k = M$, the total number of points in the set; and B_k, the set of vectors belonging to the kth cluster. We calculate the vectors of centers of gravity for the clusters obtained: $r_k = \sum_{\xi \in B_k} \xi / n_k$. If $c_k = r_k$ for all k, the division procedure is stopped. Otherwise, the vectors of cluster centers are moved to the centers of gravity r_k, the set of points is again divided into clusters, new centers of gravity are calculated, the condition of completion of the division is checked, and so on. The procedure converges rather rapidly. However, the division into clusters obtained by this iterative procedure depends on the random distribution of the centers of trial clusters Γ adopted at the very beginning of iterations. The quality of the final division is estimated by the cluster compactness criterion

$$J(q|\Gamma) = \sum_{k=1}^{q} \sum_{\xi \in B_k} |\xi - c_k|^2. \tag{6}$$

It is natural to try to find a random initial distribution Γ minimizing quantity (6) for a given number of clusters q. This can be achieved by the Monte Carlo method: experiments with random insertions of tentative cluster centers into the cloud of points are repeated many times (below, in the analysis of concrete data, we used 10^4 attempts); as a result, a distribution Γ minimizing (6) is selected.

Further, one should determine the optimal number of clusters into which the set of indicators should be divided. We denote $J_0(q) = \min_\Gamma J(q|\Gamma)$. If the number of trial clusters q is successively decreased from a certain large value to the minimum $q = 2$, the quantity $J_0(q)$ will monotonically increase but will display a kink at the optimal number of clusters (if such does exist). We applied another, more effective method for finding the optimal number of clusters based on the use of the pseudo-F-statistic [Vogel and Wong, 1978]:

$$\mathrm{PFS}(q) = (M - q)\sum_{k=1}^{q} |c_k - r_0|^2 / ((q-1)J_0(q)), \tag{7}$$

where $r_0 = \sum \xi / M$ is the center of gravity of the entire set of points classified. The maximum of function (7) defines the optimal number of clusters.

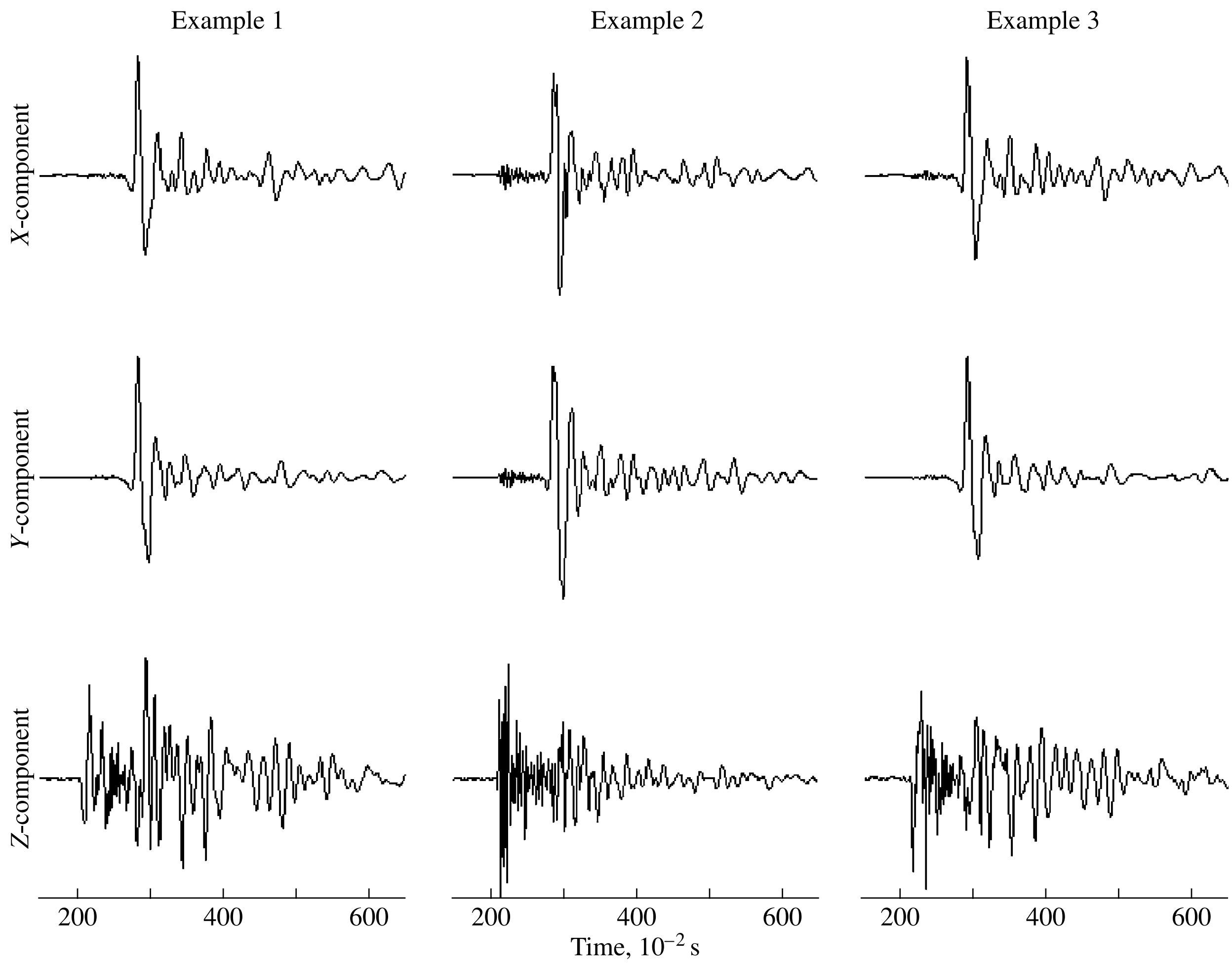

Fig. 1. Fragments of initial records of seismic events in mines of the Silesian coal basin (Czech Republic) from the total set of 111 three-component records. Shown are the records most typical of the three clusters derived from the subsequent analysis (Fig. 4).

DATA ANALYZED AND RESULTS

We analyzed a set of data comprising 111 three-component records of seismic events in mines of the Silesian coal basin obtained by IG CAS researchers [Kaláb and Knejzlik, 2002; Knejzlik and Kaláb, 2002]. The sampling rate of these records is 100 Hz, and their lengths range from 1080 to 883 samples, including a background segment before each event. The recording was conducted from May to July 2000 by a seismic network consisting of four three-component geophones. If an event was recorded by more than one seismometer, only one record, corresponding to the best signal/noise ratio, was selected.

Three of these records are shown in Fig. 1. They were selected after the classification of data according to the principle of the smallest distances to the centers of the three inferred clusters in the space of indicators. In this sense, they are the most typical representatives of their clusters.

Since we are interested in the classification of signals after an event, the initial segments of records containing only background seismic oscillations were rejected. As a result, the number of samples decreased, varying from 748 to 414. Subsequently, in order to account for the spatial structure of seismic records, we passed to the orthogonal principal components. The covariance 3×3 matrices were estimated, their eigenvalues and eigenvectors were determined, and the projections of 3-D seismic signals onto the eigenvectors of the covariance matrix were calculated. Thus, after the rejection of the initial segments, we obtained the first, second, and third principal components (in descending order of the corresponding eigenvalues) of the initial data, which were then subjected to the procedure of cluster analysis.

First, the optimal orders of orthogonal Daubechies wavelets were determined by the method of successive coherent basis thresholding with the use of the entropy minimum criterion (1) for each of the $3 \times 111 = 333$ principal components. The range of the resulting optimal orders of wavelets is very wide. Now, for each number of vanishing moments of basis functions, we determine the number of cases in which the corre-

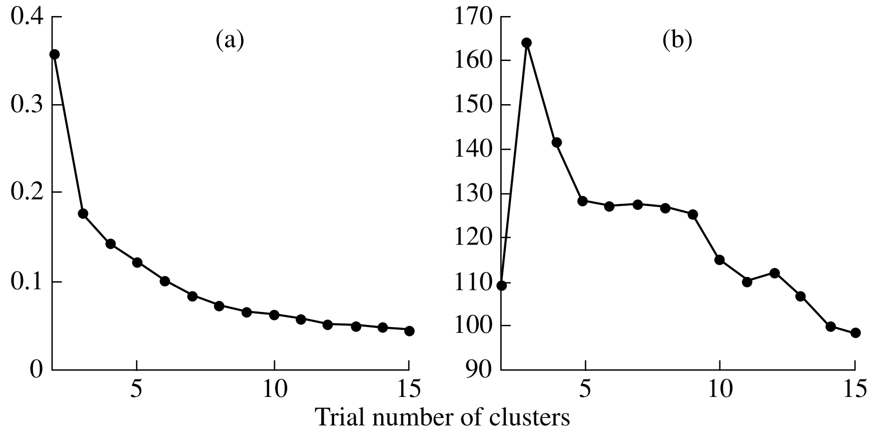

Fig. 2. Dependences of the compactness (a) and pseudo-F-statistic (b) on the trial number of clusters. The best number of clusters, corresponding to the maximum in plot (b), is equal to 3.

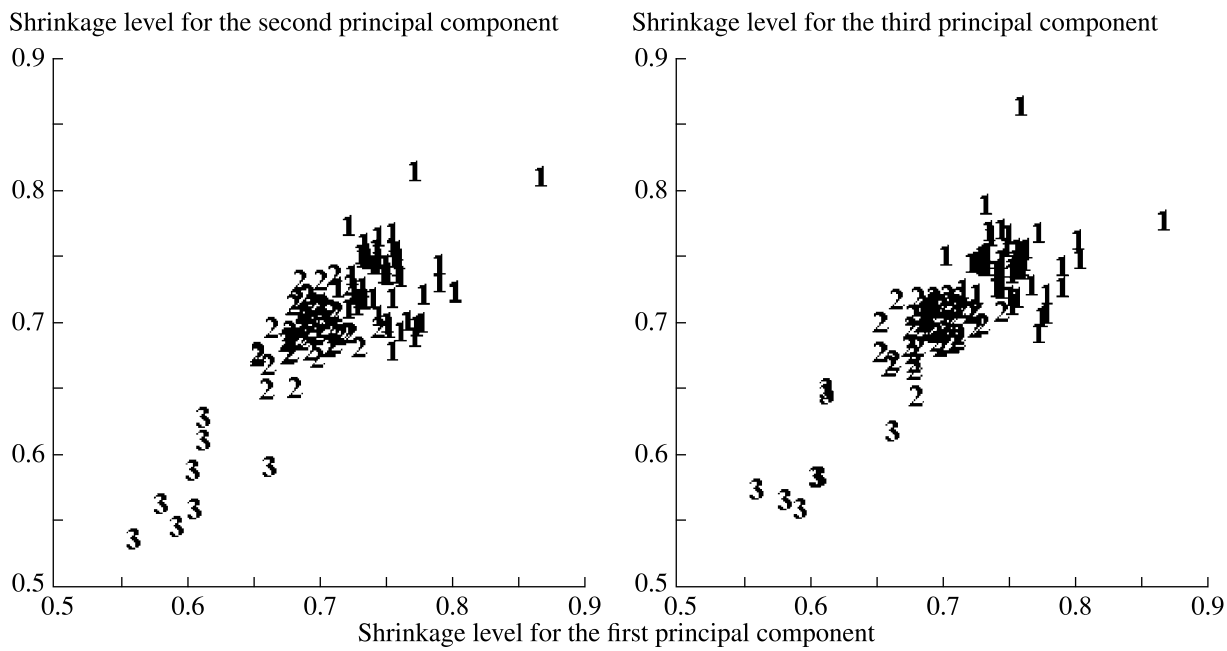

Fig. 3. Cloud of the 3-D vectors of shrinkage thresholds for the principal components of 3-D seismic records projected on two coordinate planes. The numbers at the points indicate clusters of the best division (into three clusters).

sponding wavelet is optimal. We do not discriminate between ordinary Daubechies wavelets and symlets because they coincide in smoothness and length of support. A single vanishing moment (the Haar wavelet) was encountered in only 1 case; two vanishing moments, in 15 cases; three, in 16; four, in 33; five, in 61; six, in 19; seven, in 33; eight, in 44; nine, in 55; and ten, in 56 cases. Therefore, the histogram of the numbers of vanishing moments exhibits two maximums at the numbers 5 and 10 separated by a rather deep minimum at the number of moments 6.

Figure 2 illustrates the dependences of the compactness $J_0(q)$ and pseudo-F-statistic PFS(q) on the number of trial clusters q. It is seen that $q = 3$ is an optimal value; another local maximum corresponding to the possible division into seven clusters is seen in the plot PFS(q), but this variant is statistically much less significant. The projections of three-dimensional vectors of indicators (the shrinkage thresholds for the principal components) on two coordinate planes are presented in Fig. 3, where each point is labeled with the number of the cluster to which it belongs. The first, second, and third clusters contain 51, 52, and 8 vectors, respectively.

Figure 4 shows the principal components of records from clusters whose vectors of indicators are nearest to

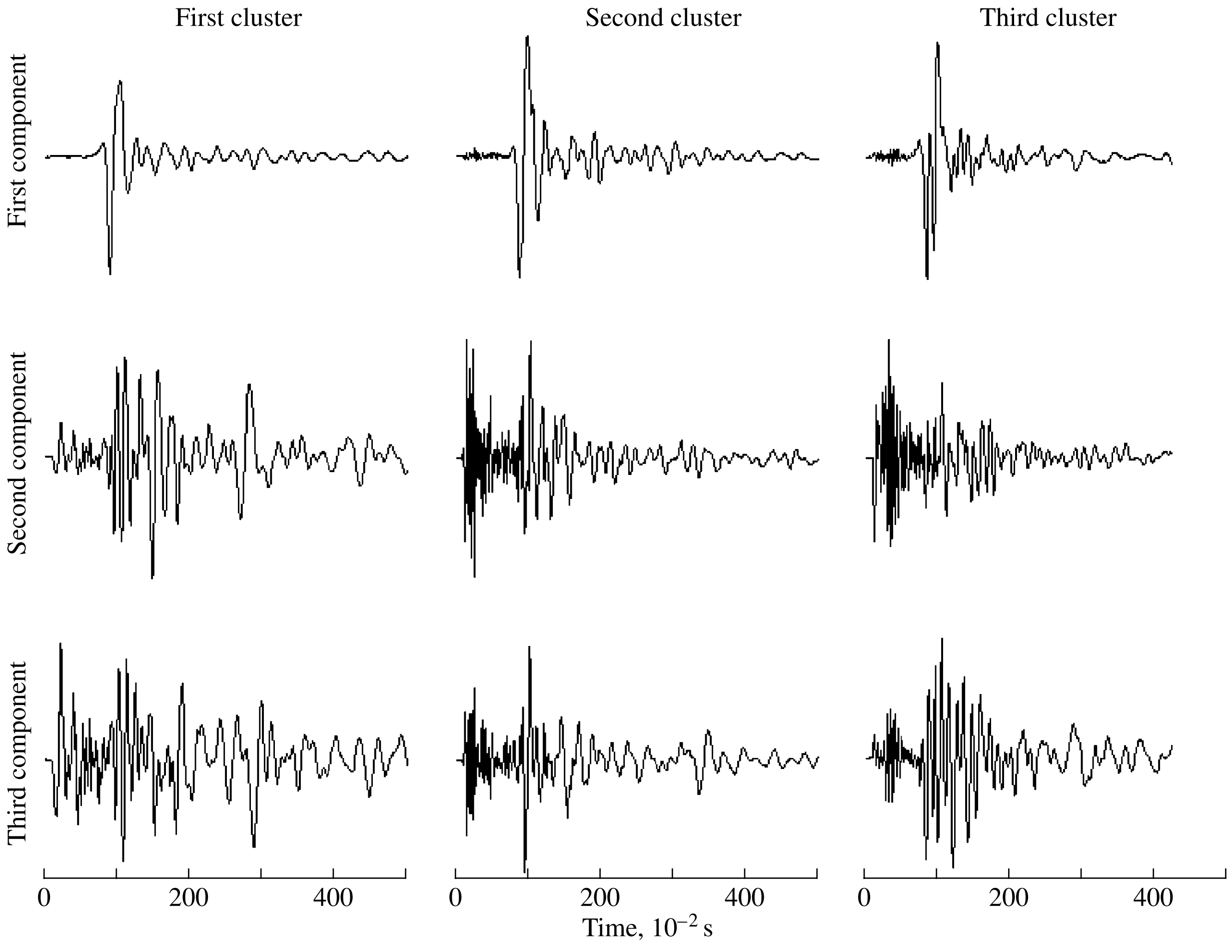

Fig. 4. Central records of clusters in projections on the principal axes (the first 500 samples after first arrivals).

centers of clusters. As mentioned above, their corresponding initial records are presented in Fig. 1. The plots in Fig. 4 clearly display significant distinctions in the behavior of signals. For example, the behavior of principal components of the first cluster is visually the simplest and is dominated by lower frequencies. Therefore, it is not surprising that the first cluster is characterized by maximum shrinkage levels: the retention of a relatively low percentage of the wavelet coefficients largest in modulus is sufficient for describing simple behavior. Similarly, small shrinkage levels seem natural for the elements of the third cluster, which is characterized by the largest diversity of behavior elements in the most typical representative of this cluster.

CONCLUSIONS

The procedure of cluster analysis (classification) of three-component seismic records based on the use of an informative indicator of the signal (the Donoho–Johnstone wavelet shrinkage level) is proposed. An advantage of this approach is complete automation of the search for the optimal division into clusters. The use of wavelets as basis functions makes it possible to take into account fine distinctions in the nonstationary behavior of signals. When applied, as a case study, to the classification of records of seismic events, this method yielded a division of data into three clusters differing in the degree of complexity of their behavior, likely pointing to three possible types of displacements in the sources of the seismic events. The first and second (the largest) clusters correspond, most probably, to strike-slip and normal faults as the most typical sliding motions on discontinuities in a rock mass. As regards the third cluster (only 7% of the total number of events), it is supposedly related to the rarest focal mechanism, namely, the pull-apart fault.

REFERENCES

1. S. A. Aivazyan, V. M. Bukhshtaber, I. S. Enyukov, and L. D. Meshalkin, *Apllied Statistics* (Finansy i Statistika, Moscow, 1989) [in Russian].
2. J. Berger, R. Coifman, and M. Goldberg, "Removing Noise from Music Using Local Trigonometric Bases and Wavelet Packets," J. Audio Eng. Soc. **42** (10), 808 (1994).

3. N. Častová and Z. Kaláb, "Processing of Seismological Signals Using Wavelet Transform," in *Tools for Mathematical Modelling. Mathematical Research* (St. Petersburg, 1999), Vol. 4, pp. 31–39.
4. N. Častová, Z. Kaláb, and R. Kučera, R., "Wavelet Transform: Presentation of Time–Frequency Decomposition for a Mining Induced Seismic Event," in *Publications of the Institute of Geophysics, Polish Academy of Sciences, M-22(310)* (Warszawa, 1999), pp. 147–151.
5. C. H. Chen, Ed., *Seismic Signal Analysis and Discrimination. Methods in Geochemistry and Geophysics, 17* (Elsevier, Amsterdam, 1982; Mir, Moscow, 1986).
6. C. K. Chui, *An Introduction to Wavelets* (Academic, San Diego, 1992; Mir, Moscow, 2001).
7. I. Daubechies, *Ten Lectures on Wavelets* (CBMS-NSF Series in Applied Mathematics) (SIAM, Philadelphia, 1992; NITs "Regulyarnaya i Khaoticheskaya Dinamika," Izhevsk, 2001), vol. 61.
8. D. Donoho and I. Johnstone "Ideal Spatial Adaptation Via Wavelet Shrinkage," Biometrika **81**, 425 (1994).
9. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis* (Wiley, New York, 1973; Mir, Moscow, 1976).
10. P. J. Huber, *Robust Statistics* (Wiley, New York, 1981; Mir, Moscow, 1984).
11. Z. Kaláb and J. Knejzlik, *Systematic Measurement and Preliminary Evaluation of Seismic Vibrations Evoked by Mining Induced Seismicity in Karviná Area,* Publications of the Institute of Geophysics, Polish Academy of Sciences, M-24(340) (Warszawa, 2002).
12. J. Knejzlik and Z. Kaláb, *Seismic Recording Apparatus PCM3-EPC*, Publications of the Institute of Geophysics, Polish Academy of Sciences, M-24(340) (Warszawa, 2002).
13. V. S. Korolyuk, N. I. Portenko, A. V. Skorokhod, and A. F. Turbin, *Handbook on Probability Theory and Mathematical Statistics* (Nauka, Moscow, 1985) [in Russian].
14. A. A. Lyubushin, Jr., "Multidimensional Wavelet Analysis of Geophysical Monitoring Time Series," Fiz. Zemli, No. 6, 41–51 (2001) [Izvestiya, Phys. Solid Earth **37**, 474–483 (2001)].
15. A. A. Lyubushin, Jr., "Robust Wavelet-Aggregated Signal for Geophysical Monitoring Problems," Fiz. Zemli, No. 9, 37–48 (2002) [Izvestiya, Phys. Solid Earth **38**, 745 (2002)].
16. A. A. Lyubushin, Jr., "Wavelet-Aggregated Signal and Synchronous Peaked Fluctuations in Problems of Geophysical Monitoring and Earthquake Prediction," Fiz. Zemli, No. 3, 20–30 (2000) [Izvestiya, Phys. Solid Earth **36**, 204 (2000)].
17. S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, San Diego, 1998).
18. V. Vapnik and A. Ya. Chervonenkis, *Pattern Recognition Theory* (Nauka, Moscow, 1974) [in Russian].
19. M. A. Vogel and A. K. C. Wong, "PFS Clustering Method," in *IEEE Trans. Pattern Analysis. Mach. Intell. PAMI-1* (1978), pp. 237–245.