ГЕОФИЗИЧЕСКИЕ ИССЛЕДОВАНИЯ, 2010, том 11, № 4, с.29-34

550.834

METHOD FOR CLASSIFYING THREE-COMPONENT SEISMIC RECORDS BASED ON WAVELET ANALYSIS

© 2010 г. Alexey A. Lyubushin[1], Baldev Raj Arora[2], Sh. Naresh Kumar[2]

[1] *Schmidt Institute of Physics of the Earth, Russian Academy of Sciences, Moscow, Russia*
[2] *Wadia Institute of Himalayan Geology, Dehra Dun, India*

Automatic classification of earthquake sources is necessary to understand the day to day activity in the regions of high seismicity (high seismic activity) zones. Himalaya is formed due to the continent-continent collision and still it is in the deformation stage. The complex geotectonic setup along with neo-tectonic activity in the region predetermines existence of different types of the earthquake source mechanisms in Himalaya. The improved event detection and location of a regional network can be achieved by developing and incorporating new concept for seismic data analysis.

We present a new method for reliable characterization of seismic sources using local seismic events, which can be useful for automatic classification of local seismic records. Spectral features of the seismic records and the parameterized attributes of their waveform have been used to classify the seismic records in NW Himalaya. The approach is based on automatic classification of three-component seismic records through Donoho–Johnstone wavelet shrinkage level as an informative indicator.

Keywords: three-component seismic records, classification, wavelets.

Introduction

The regions of high seismic activity have the large set of seismic records recorded at different seismic stations of the seismic array. The analysis of these records take time and the classification of these records on the basis of seismic sources, the types of source characterization and the distance between source and the recording station is a big problem. This problem can be solved if the classification is done automatically using three component waveform seismic records. This classification is possible on the basis of spectral analysis, pattern recognition, syntactical procedures etc [*Chen*, 1982].

During current study, we have used the wavelet shrinkage method of Donoho–Johnstone [1994] to characterize the pattern of seismic records. The methodology is used to automatically classify the seismic records of Kangra region of Western part of Himalaya.

Geotectonic information and seismic data

Himalaya is the largest and youngest mountain range in the Central Asia formed due to the continent-continent collision. Geologically the western part of Himalaya comprises the rock types from Palaeoproterozoic to the Quaternary age with high degree of deformation and complex composition due to collision. Mainly the region can be divided into four zones with different geological setup. Tectonically the region is very complex containing thrust faults, normal faults and strike-slip faults. We have analyzed three-component seismic records of 407 small energy earthquakes obtained from the local networks in Western part of Himalaya (199 records with 100 Hz sampling rate and 208 – with 20 Hz sampling rate) with a purpose to split them into a number of clusters (groups) which follows from properties of their waveforms.

Description of method

Let $Z(t)$ be a three-component seismic record which contains the record of seismic event, t is a discrete time index, numerating the successive samples. The method of data processing could be presented as the following sequence of points:

1. Coming to principal components $Z(t) \to P(t)$ by calculating covariance matrix of initial three-component data $Z(t)$ and projecting then onto its normalized eigenvectors.

2. Seeking for the P-wave onset time moment τ_P using the 1st principal component $P_1(t)$.

3. Extracting $Q(t)$ – the main three-component part of S-waveform within some vicinity of maximum modulus amplitude time moment of the 1st principal component $P_1(t)$.

4. Seeking for the best orthogonal wavelet basis [*Daubechies,* 1992; *Mallat,* 1998] for each component $Q_j(t)$ using coherent basis thresholding method [*Berger et al.,* 1994; *Mallat,* 1998]. The vocabulary of tested wavelets consists of 17 basic functions: 10 usual Daubechies wavelets with number of vanishing moments from 1 to 10 and 7 symlets with number of vanishing moments from 4 to 10.

Thus, each main S-waveform component $Q_j(t)$ is characterized by some number m_j of vanishing moments for the best wavelet. The value of m_j could vary from 1 to 10 (one vanishing moment means Haar's wavelet) and describes the degree of smoothness of the S-waveform component $Q_j(t)$.

5. Having a sufficiently large set of three-component seismic records from different sources and at different stations we can obtain the correspondent set of m_j values. Thus, it is possible to calculate the histograms $N_j(m), m = 1,...,10$ – the number of cases when $m_j = m$. The local maximums of $N_j(m)$ with respect to m could extract different types of sources. The main interest should be applied to the histogram of the 1st principal component $Q_1(t)$ as the most informative.

Below we give the descriptions of each point except the 1st one – just because of its simplicity and give some results of applications. It should be mentioned first of all that some initial seismic data contains rather intensive low-frequency components which should be suppressed before the classification procedure. For this purpose we perform direct orthogonal wavelet transform of records using Daub04 wavelet (2 vanishing moments), setting to zero all wavelet coefficients above some maximum working detail level $\alpha_{\max}$ and making an inverse wavelet transform. We use the value $\alpha_{\max} = 4$ what means that we retain variation with scales from $2\Delta t$ up to $32\Delta t$ where Δt is the sampling time interval.

Seeking for P-wave onset time. Let $P_1(t)$ be the 1st principal component. For detecting P-wave onset time moment we have applied STA/LTA method. Let L be the length of "long" moving time window and S be the length of "short" time window which compose the most left-hand part of the long time window. We have taken the values $L = 8 \cdot 2^{(\alpha_{\max}+1)}$, $S = L/8$. Let us compute sample estimates $\sigma_L^2(\tau)$ and $\sigma_S^2(\tau)$ of variance within long and short time windows in dependence on their common right-hand time moment τ. For stationary behavior of the signal $P_1(t)$ the ratio $r(\tau)=\sigma_S^2(\tau)/\sigma_L^2(\tau)\approx 1$. When short moving time window came across the event this ratio became increase. Thus, the P-wave onset time moment could be defined as the first time moment when $r(\tau)\ge\gamma>1\Rightarrow\tau=\tau_P$. The value of the parameter γ controls the sensitivity and reliability of the detector. The usual values of $\gamma\approx 4$.

Extracting the main S-waveform. Let us consider the behavior of $P_1(t)$ on the time interval τ_P+M where M is some sufficiently large number of samples which a priory covers time interval of S-wave observation. For our data we have taken M=2400 samples with 20 Hz sampling rate. Let ξ_S be the time moment corresponding to maximum of $|P_1(t)|$ for $\tau_P<t<\tau_P+M$. Let ξ_1 be the minimum value of time moments for which the value of $|P_1(t)|$ exceeds a 75%-quantile of $|P_1(t)|$-distribution for $\tau_P<t<\tau_P+M$ and ξ_2 – the maximum value of such time moments.

Let us calculate $t_1=\max(\xi_1,\xi_S-8\cdot 2^{(\alpha_{\max}+1)})$, $t_2=\min(\xi_2,\xi_S+8\cdot 2^{(\alpha_{\max}+1)})$. The main part of S-waveform $Q(t)$ presents three principal components for time moments $t_1\le t\le t_2$.

Seeking the best wavelet for each component $Q_j(t)$, j=1, 2, 3. The method of coherent basis thresholding is described in details in [*Berger et al.,* 1994; *Mallat,* 1998], see also Appendix. In the paper [*Lyubushin et al*., 2004] it was applied for classification of the set of three-component seismic records of weak mine events in Silesian coal basin, Czech Republic. The method is iterative. Each iteration consists of seeking the best wavelet within given vocabulary from minimum entropy of squared wavelet coefficients distribution; setting to zero all “sufficiently large” (by absolute values) wavelet coefficients (the threshold for defining “large” coefficients is calculated from asymptotical distribution of large values for Gaussian white noise) and performing the inverse wavelet transform – calculating a so called residual signal. These iterations are continuing till the moment when neither one wavelet coefficient would not be “sufficiently large” – this means that the whole residual signal is the noise from the point of view of all basis functions from given vocabulary. The last best wavelet which was found before this final iteration from minimum of entropy of residual signal considered to be the best for initial signal.

After determining the best wavelet for each component of main S-waveform $Q_j(t)$ we know their values of m_j – number of vanishing moments or degree of smoothness. Thus, we are able to calculate histograms $N_j(m), m=1,...,10$ which are the main tools for classifying seismic events waveforms.

Analysis of results

The main result of the application the method to three-component seismic records is presented on Fig. 1. We can notice that the number of vanishing moments forms three clusters with maximums at 5, 8 and 10 moments (Fig. 1(a), 1st and 2nd principal components). The similar result is for classification of 100 Hz seismic records of swarm at Fig. 1(b) although maximum of 5 vanishing moments migrated to 4.

Fig. 2 presents waveforms of S-waves for most typical records with 20 Hz sampling rate for all three principal components. We have taken the records for which all three components have the same number of vanishing moments (this is typical also). We can see the really different smoothness of signals, corresponding to histogram peaks at Fig. 1.

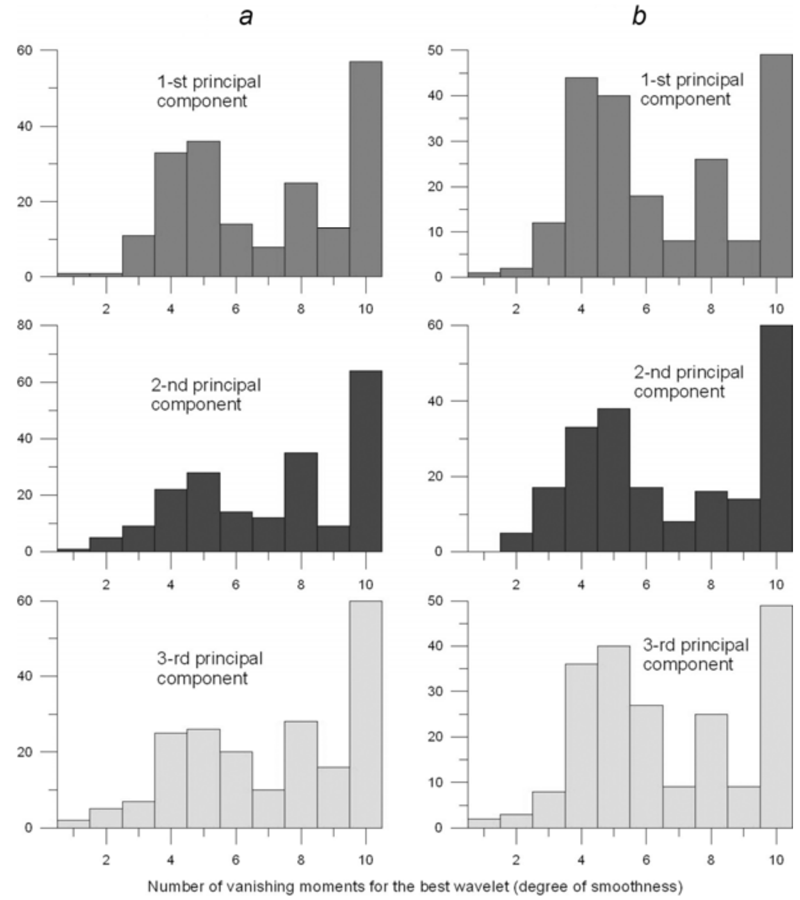

Fig. 1. Histograms of distribution of seismic events principal components waveforms in dependence on their degree of smoothness. *a* – results for classification of 370 three-component seismic records, 20 Hz sampling interval; *b* – results for classification of 208 three-component seismic records during earthquake swarm 02–05 March 2005

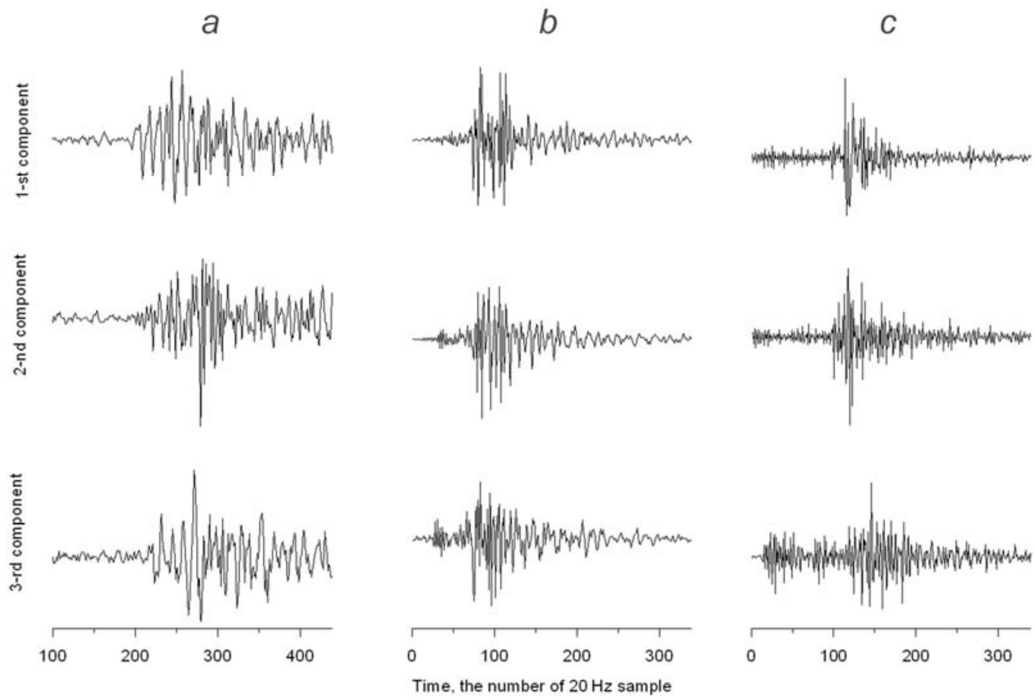

Fig. 2. Waveforms of S-waves after coming to 20 Hz sampling rate for the most typical records having number of vanishing moments equal to 10 (*a*), 8 (*b*) and 5 (*c*) for all three principal components. The duration of presented fragments is the same and equals to 340 samples

Conclusion

The result of occurring three stable clusters of waveform types could be connected with many factors such as combination of focal mechanisms and conditions of propagation. Just because the number of clusters equals to three, we propose that the peaks of histograms at Fig. 1 could be due to three main types of source mechanisms: strike-slip and normal displacement as the most typical sliding motions on discontinuities in a fault and pull-apart movement. To confirm this hypothesis or propose another, one needs detailed information about source mechanisms.

Appendix: Method for seeking the best orthogonal wavelet

Let $B = \{g_n, n = 1,..., N\} = \{g_n(t), n, t = 1,..., N\}$ – a set of real orthonormal N-dimensional vectors which form (as columns) a matrix of discrete wavelet transform of N-dimensional data vectors (usually $N = 2^m$).

Let Z be an N-dimensional vector with initial time series; Y be the working buffer for time series processing. The method could be presented as the sequence of following operations (numbered points).

Point 1. Initialization: $Y = Z$.

Point 2. Perform wavelet transform of Y for the number of permissible bases and find wavelet basis B from the condition $C(Y, B) \to \min\limits_{B}$.

Here:

$$C(Y, B) = -\sum_{n=1}^{N} P_n^{(Y)} \cdot \ln(P_n^{(Y)})$$

– entropy of distribution of squared wavelet coefficients:

$$P_n^{(Y)} = d_n^{(Y)} \Big/ \left\| d^{(Y)} \right\|^2, \ \left\| d^{(Y)} \right\|^2 = \sum_{n=1}^{N} \left| d_n^{(Y)} \right|^2 .$$

Point 3. For the current wavelet basis find $M = M(Y, B)$ from the condition:

$$\frac{\left| d_{n_{M+1}}^{(Y)} \right|^2}{\sum\limits_{k=M+1}^{N} \left| d_{n_k}^{(Y)} \right|^2} \le \frac{2\ln(N - M)}{(N - M)},$$

where $d_{n_k}^{(Z)}$, k=1, …, N is the sequence of wavelet coefficient sorted in the order of decreasing of their absolute values: $|| d_{n_k}^{(Z)} | \ge | d_{n_{k+1}}^{(Z)} |$ (the 1st member is the wavelet coefficient with the maximum absolute value).

Point 4. If $M = 0$, then exit,
else $Y = R^M Y$; go to the point 2.

Let

$$R^M Y = Y - \sum_{k=1}^{M} d_{n_k}^{(Y)} \cdot g_{n_k} = \sum_{k=M+1}^{N} d_{n_k}^{(Y)} \cdot g_{n_k}$$

be the residual after removing the first M absolute values maximum wavelet coefficients.

The decision rule in the point 3 is based on the following fact from asymptotic distribution of independent Gaussian random values:

$$\text{if} \quad B(t) \sim N(0,1) \Rightarrow \lim_{N\to\infty} \Pr\left\{ \max_{1 \le t \le N} |B(t)|^2 \Big/ \|B\|^2 \le \frac{2\ln N}{N} \right\} = 1 .$$

Reference

Berger J., Coifman R., Goldberg M. Removing noise from music using local trigonometric bases and wavelet packets // J. Audio Eng. Soci. 1994. V. 42, N. 10. P.808–818.

Chen C.H. (ed.) Seismic signal analysis and discrimination // Methods in Geochemistry and Geophysics. Amsterdam: Elsevier, 1982.

Daubechies I. Ten Lectures on Wavelets // CBMS-NSF Series in Applied Mathematics. Philadelphia: SIAM, 1992. V. 61.

Lyubushin A.A., Kalab Z., Castova N. Application of Wavelet Analysis to the Automatic Classification of Three-Component Seismic Records // Izvestiya. Physics of the Solid Earth. 2004. V. 40, N 7. P.587–593.

Mallat S. A wavelet tour of signal processing. San Diego, London, Boston, N.Y., Sydney, Tokyo, Toronto: Academic Press, 1998. 577 p.

Authors:

LYUBUSHIN Alexey Alexandrovich – Dr. Sci., Chief research scientist, Institute of Physics of the Earth, Russian Academy of Sciences, 123995, Russia, Moscow, B. Gruzinskaya Str., 10. Phone: +7(499)254-23-50. E-mail: lyubushin@yandex.ru

ARORA Baldev Raj – Director of Wadia Institute of Himalayan Geology; 33, Gen. Mahadev Singh Road Dehra Dun - 248001. Phone No.: 0135-2625952, Fax No.: 0135-2625212.

KUMAR Naresh – Senior Scientist of Wadia Institute of Himalayan Geology; 33, Gen. Mahadev Singh Road Dehra Dun - 248001. Phone No.: 0135-2625952, Fax No.: 0135-2625212.

УДК 550.834

МЕТОД КЛАССИФИКАЦИИ ТРЕХКОМПОНЕНТНЫХ СЕЙСМИЧЕСКИХ ЗАПИСЕЙ НА ОСНОВЕ ВЕЙВЛЕТ-АНАЛИЗА

А.А. Любушин[1], Б.Р. Арора[2], Н. Кумар[2]

[1] *Институт физики Земли им. О.Ю. Шмидта РАН, г. Москва, Россия*
[2] *Институт геологии Гималаев им. Вадиа, Дехрадун, Индия*

Аннотация. Автоматическая классификация источников сейсмических событий является операцией, необходимой для понимания природы сейсмического процесса в регионах его высокой интенсивности. Ниже в качестве такого региона рассматриваются Гималаи, которые образовались в результате столкновения континентов и сейчас находятся в стадии интенсивных деформационных процессов. Большое разнообразие тектонических условий в этом регионе влечет за собой наличие различных типы механизмов источников сейсмических событий.

Процедуры автоматической классификации сейсмических записей для региональных сетей могут быть модифицированы и улучшены путем применения новых концепций теории сигналов. Ниже представлен один из новых методов, который может быть полезен для классификации локальных событий и основан на параметризации спектральных свойств трехкомпонентных волновых форм источников на северо-западе Гималаев с использованием в качестве информативной характеристики вейвлетного порога сжатия Донохо–Джонстона для оптимального ортогонального вейвлета.

Ключевые слова: трехкомпонентные сейсмические записи, классификация, вейвлеты.