

Identification method development of focal zones based on seismic monitoring results

Andrey Gladyr^{1*} and Vladimir Miroshnikov¹

¹Mining Institute of Far Eastern Branch of Russian Academy of Sciences, Khabarovsk, Russia

Abstract. Method of seismic active zones control is one of approaches of prediction of dynamic mountain pressure manifestations, which is to identify and analyze formation dynamics of seismo acoustic active zones and predict impact hazard based on revealed patterns of changes in geoacoustic activity of rock massif. The key point of this control method is the zone allocation where a potentially dangerous rock destruction source is formed. In the article, we propose an automated method for damage zone identification, based on the preliminary exclusion of background radiation by nonparametric density estimation method, the seismo acoustic active zone allocation by means of probabilistic cluster analysis and damage zone parametrization by characteristic ellipsoid selection. These tasks are part of complex geomechanical monitoring system development and are solved by upgrading and subsequent integration into the interacted software of the considered seismo acoustic active zone monitoring method. Their solution will allow to fully automate predicting dangerous state process of the controlled rock mass, improving forecast quality and significantly reducing the time spent on obtaining final result.

1 Introduction

The kinetic concept of solid destruction, based on the analysis of existing approaches in the forecast of dynamic mountain pressure manifestations, is the basis of most of methods. Many scientists who adhere to this concept found that several characteristic stages are in rock destruction process in which destruction process begins with crack formation with a size of millimeters to cracks with a size of centimeters and of meters, and can end in gaps from tens to thousands of meters, typical of mountain-tectonic shocks, technogenic earthquakes and etc. [1-4].

At present, a number of methodological approaches to dynamic rock pressure manifestations forecast, using different predictive criteria [1, 5, 6]. The author team from the Institute of mining FEB RAS developed a monitoring method of geomechanical rock mass state – the monitoring method of seismo acoustic active zones. This method consists in identification and analysis of acoustically active zones formation dynamics and impact hazard forecast based on established patterns of changes in geoacoustic rock massif activity. Therefore, the key point of the considered method of seismo acoustic active zones

* Corresponding author: rush3112@mail.ru

monitoring is the controlled zone allocation, which can be qualified as damage zone, within which a potentially dangerous rock destruction source is formed.

2 Filtering background radiation

The practice of using cluster analysis methods of seismic monitoring data shows that final clustering result is significantly affected by original data quality, which necessarily contains a certain amount of faulty data. It is recommended to select a data subset in areas with high point density before beginning of clustering. To achieve this goal, we propose to estimate density distribution at each point using a nonparametric density estimate.

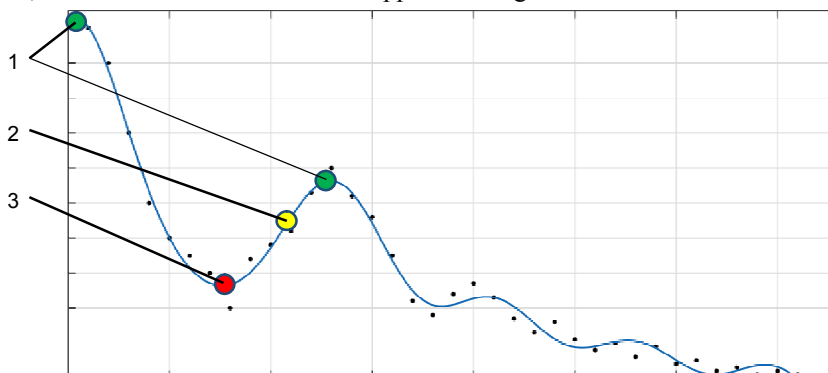
At the same time, nonparametric estimates of distribution numerical random variables density and finite-dimensional random vectors were considered in the early stages. In the 1980s, it was possible to construct such estimates in arbitrary environment spaces [7], and then for specific types of non-numeric data [8].

Simplifying, the registered event with coordinates (x_i, y_i, z_i) will be considered as belonging to the region with high density if point number located inside hypersphere centered at the point with coordinates (x_i, y_i, z_i) and radius h , radius h which greater than the threshold value c .

It is necessary to set the threshold value c and filter the seismo acoustic events related to the background radiation upon calculations completion related to calculation of the distribution density estimation for each registered event.

The following method is proposed for automatic selection of value c :

1. Distribution density value is calculated by original input dataset.
2. Median of the resulting distribution is calculated.
3. Resulting distribution is approximated by a given function. Practice of using this method has shown possibility of application as an approximating function of the polynomial of eighth and higher degrees. However, best results were achieved by using as an approximating function of form $f(x) = a_1 \cdot \sin(b_1 \cdot x + c_1) + \dots + a_i \cdot \sin(b_i \cdot x + c_i)$ with number of terms from six and above.
4. Maximum point's values m_1 and m_2 of approximating function to the right and to the left of the median are determined.
5. Argument value corresponding to the minimum of approximating function on interval $[m_1, m_2]$ is determined. Calculated argument value will be required value c .
6. The described practice is illustrated in Figure 1, where points are density distribution values, and continuous line visualizes approximating function.



- 1 – Maximum points of approximating function in the neighborhood of median;
 2 – Median of density distribution; 3 – Required threshold value.

Fig. 1. Illustration of a method for automatic selection of threshold density values.

The main task in carrying out qualitative calculations of the weight function is the choice of the parameter h . There are several common approaches to choice of this parameter: reference heuristic rules, substitution methods, cross-validation methods, bootstrap methods.

Least-squares cross-validation method is the most suitable of them. It is a fully automatic and data dependent method for selecting parameter h , which is based on principle of choosing parameter h , minimizing integral mean square error of resulting estimate [9-10].

Thus, the developed methodology of background seismo acoustic radiation identification allows, on the one hand, solving problem of data filtering for subsequent separation of seismo acoustic active zones, and on the other hand, to identify recorded seismo acoustic events related to background radiation. It allows assessing background radiation parameters for their further use in predicting dynamic manifestations in the rock massif [11].

3 Damage zones formation

Cluster analysis - a set of mathematical techniques designed for formation of relatively "distant" from each other groups of "close" to each other objects according to distances or relationships (proximity measures) between them [12-14].

Retrospective analysis of considered clustering methods allows us to conclude that use of graph or hierarchical clustering methods seems to be the most appropriate at first glance to solve the problem of damage zones identification based on seismo acoustic monitoring results [15]. This statement is based on idea of physical processes occurring in loading process and subsequent deformation of the rock massif, for example, that destruction process is fractal.

However, these methods of solving clustering problem are related to deterministic mathematical models, as these approaches imply that in clustering process initial set of objects X is divided into several non-intersecting subsets. In other words, these methods are methods of "crisp" clustering.

On the other hand, "fuzzy" clustering methods allow the same object to belong to several (or even all) clusters at the same time, but with different probability. Such probability clustering in many situations is more "natural" than "crisp", because it takes into account random processes inherent in real nature objects and their stochastic description.

Clustering methods are classified by whether the number of clusters is defined or not in advance. In the latter case, clusters number is determined in the process algorithm execution based on distribution of original data. To solve problem of damage zones allocation, C-means algorithm is used, which divides data into known clusters number described by the fuzzy partition matrix. The only difference between the partition matrices for the fuzzy and crisp case is that in the fuzzy partition belonging degree of the object to the cluster takes values from interval $[0, 1]$, and in crisp - from two-element set $\{0, 1\}$. Fuzzy partitioning allows you to simply solve the problem of objects located on the border of two clusters. These objects are assigned belonging degrees equal to 0.5.

The spread criterion is used to evaluate quality of fuzzy splitting [16]. Finding the fuzzy partition matrix with the minimum value of the criterion is a nonlinear optimization problem, which can be solved by different methods. The fuzzy c-means algorithm is the most known and frequently used method for solving this problem, which is based on method of undefined Lagrange multipliers [16].

The distance between the object and the cluster center is calculated by standard Euclidean norm, which allows allocating clusters in the form of hyperspheres, in basic algorithm of fuzzy c-means. Diagonal norm allows us to allocate clusters in form of

hyperellipsoids oriented along coordinate axes. The Mahalanobis norm allows to allocate clusters in the form of hyperellipsoids. Their axes can be oriented in any direction.

The given clustering algorithm allows us to allocate data clusters for some data sets in form of different geometric shapes: spheres, ellipsoids with different orientations, chains etc. Form of all clusters is the same for clustering algorithm result with a fixed rate. Clustering algorithms how-to impose a structure that is unusual for data, which leads not only to suboptimal, but sometimes to fundamentally wrong results. There are several methods to remove this weakness. The authors recommend using the Gustafson-Kessel algorithm [17].

The Gustafson-Kessel algorithm uses an adaptive norm for each cluster; i. e. there is a norm-generating matrix B_i for each i -th cluster. This clustering algorithm is optimized not only cluster centers coordinates and the fuzzy partitioning matrix, but also the norm-generating matrices for all clusters. This allows you to select clusters of different geometric shapes.

The number of clusters (c) is the most important parameter in the algorithm. It is quite difficult to choose the right number of clusters for real tasks without any a priori information about structures in data. There are two formal approaches to choosing the number of clusters.

The first approach is based on criteria of compactness and separability of obtained clusters. It is logical to assume that data will be divided into compact and well-separated groups if the number of clusters is chosen correctly. Otherwise, clusters will probably be not compact and well separated. There are several criteria for assessing the compactness of clusters. However, the issue is remained unresolved how to formally and reliably determine the correct choice of the number of clusters for an arbitrary data set. It is recommended to use the index Xie-Beni [19] for fuzzy c-means algorithm [18].

The second approach is suggested to start clustering with a sufficiently large number of clusters, and then sequentially to combine similar adjacent clusters. Various formal similarity criteria for clusters are used.

However, the subtractive clustering method can be considered the most accurate method of determining the number of clusters. One of its advantages is that you do not have to specify the initial number of clusters. The method was proposed By R. Yager and D. Filev in 1993 [20].

The following complex algorithm of damage zones formation is proposed based on results seismo acoustic monitoring of based on above algorithms:

1. Regions with high seismic activity are selected from entire recorded seismic events set using nonparametric density estimation.
2. The number of clusters is determined using the subtractive clustering algorithm.
3. The process of multiple clustering repetition is performed using the Gustafson-Kessel algorithm to obtain a stable result of splitting the initial objects set into the optimal number of clusters by estimating of distribution probability. The highest probability of an event belonging to each cluster is calculated after the process is completed for each case.

The developed algorithm makes it possible to assess the quality of clustering. At the third step of the algorithm, as described above, there is a direct process of objects distribution into clusters. This process is performed as follows: cluster split procedure is performed a large number of times, belonging probability to a given cluster, which it belonged, in the process for each event is saved. Greatest probability (in frequency ratio) of belonging to each cluster is selected for each event are completed after the iterations. Thus, the result is the division of events into probability of belonging to a particular cluster and clustering characteristic, i.e. reference reliability of each object (observation) to each of clusters.

4 Calculation of characteristic ellipsoid parameters

Seismic and acoustic signals location of allows to identify the defects cracking area and to identify the propagation direction of cracks in space. It can be concluded with some assumption that this area will have some ellipsoid form with its main diagonals in space, which characterize length, width and thickness ratio of destroyed layer and orientation of this layer in space. Construction of ellipsoid is based on statistical distribution of crack centers according by location. The coordinate table of recorded events is taken as source data.

We choose select the direction in which the coordinate distribution will be maximal. A projection of event location points on the normal direction will have the maximum dispersion in this case. This normal characterize the length direction of damage zone. In the same way, we determine zone thickness in orthogonal plane to normal, as a direction with a minimum dispersion in the plane. Then we will consider the dispersion of scatter points as a characteristic of zone width in the remaining third orthogonal direction to two previous.

The following practice of ellipsoid construction is proposed. The selected cluster from recorded seismic events set is considered as input data. For further analysis, an origin is moved to damage zone center. The standard event coordinates offsets from the center have different meanings for different directions. We use procedure of successive approximations to determine ellipsoid axes direction. The maximum value of the standard deviation of the coordinate projection on the coordinate axes is revealed. The axis with the maximum deviation is variable. All points of the array in new coordinate system are rotated sequentially around other two axes at some small angle α in positive and negative directions.

We determine such direction of the axis by varying angle α , for which standard deviations of event coordinates will be the maximum. If this value corresponds to a successful variation (rotation), it is taken as the next value of the center point from which the following variations are made. At the same time, new value of coordinate system transformation matrix is determined for each step using same rotation operators.

If central point is the maximum, then we consider the goal reached accurate to angle α . Angle α is reduced to determine a more accurate solution, and then the procedure is repeated until desired accuracy is achieved. The value $\alpha = 10^{-3}$ rad is assumed as a satisfactory accuracy in practical calculations.

Next step second main direction of distribution of seismo acoustic events coordinates is determined. First, the nearest coordinate axis is selected, on which coordinate projections have a minimum spread. Then, rotation operator is calculated around previously inclined axis until the minimum value of mean square spread on the selected coordinate axis is reached. Search procedure is similar to the one described only with the difference that only two variations near center point are determined by rotation operator now. The procedure is performed until the required angle accuracy is achieved. Inverse coordinate system transformation matrix consists of three columns. They are unit vectors of three main orthogonal directions of an ellipsoid corresponding to main standard deviations in original coordinate system.

Main ellipsoid axes are vectors with lengths proportional to the standard deviation. It is believed that no more than 0.27% of values exceeding 3σ , (σ – standard deviation) in the classical Gaussian distribution, however, these values may differ significantly in the case of real data. The region fracture volume estimated as ellipsoid volume with dimensions σz , where z - parameter takes into account crack distribution density in selected area.

5 Approbation of the identification method of damage zones based on seismic monitoring results

The seismic monitoring data of the "Prognoz ADS" automated system of mountain pressure control at the "Antey" field for 2011, during which whole range of multi-scale seismic events was registered, were analyzed to check of quality of the developed damage zones identification method.

Was located 14065 seismo acoustic events only in 2011. 38% of the events were related to background radiation based on the results of background radiation filtering. 3 clusters were allocated for the remaining events as a result of cluster analysis. 373 events (21%) were assigned to the largest cluster. The results of characteristic ellipsoid formation describing the selected damage zone are shown in figure 2 (the center of coordinate system is added to the center of selected damage zone). 28 events (7.5 per cent) fall outside 3σ , 69 events (18%) – 2σ , 209 events (56%) – 1σ in the figure.

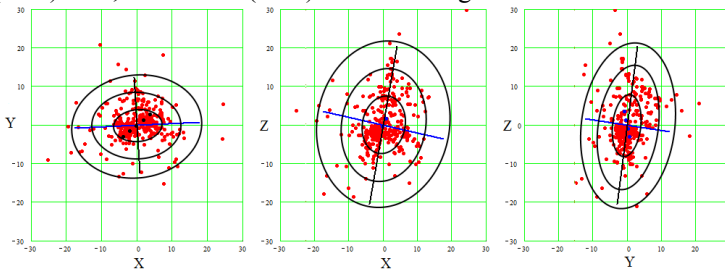


Fig. 2. The result of the damage zone selection algorithm according to seismic monitoring in the projection on orthogonal planes.

Thus, use of the characteristic ellipsoid has allowed reducing number of parameters describing the selected damage zone more than 100 times, while providing ability to control change in volume and shape of a damage zone, both in static representation and in dynamics.

It should also be noted that the proposed method of the damage zones identification is fully customizable and requires a necessary pre-commissioning to adjust parameters in transition to permanent practical use in industrial facilities.

6 Conclusion

It is established by results of existing approaches analysis to dynamic manifestations prediction of rock pressure that the kinetic concept of solids destruction is the basis of most methods. Many scientists who adhere to this concept established that in the process of rocks destruction are several characteristic stages in which destruction process begins with microcracks formation with sizes from millimeters to centimeters and meters, and can be end with breaks from tens to thousands of meters, typical of mountain-tectonic shocks, technogenic earthquakes, etc.

Allocation of the controlled zone, which can be qualified as a damage zone, within which a potentially dangerous source of rock destruction is formed, is the key point of monitoring seismic active zones method, as one of monitoring the geomechanical state methods of rock massif.

Development of mathematical methods capable of statistically sound identification of seismic active zones based on geomechanical monitoring results, as well as the development of a technique for the formation of a characteristic ellipsoid capable of describing geometric properties of identified damage zones with a given accuracy and

subsequent integration into associated software of considered monitoring method of seismic active zones will fully automate the process of forecast dangerous state of controlled rock massif, by improving forecast quality, while significantly reducing time spent on obtaining the final result.

References

1. I. Yu. Rasskazov, A. Yu. Iskra, G. A. Kalinov, P. A. Anikin, A. V. Gladyr, M. I. Rasskazov, A. V. Sydlyr, *Mining Information and Analytical Bulletin*, **8**, 212 (2011).
2. Marcak Henryk, Mutke Grzegorz, *J. Seismol*, **17**, 4, 1139 (2013).
3. Paige E. Snelling, Laurent Godin, Stephen D. McKinnon, *International Journal of Rock Mechanics and Mining Sciences*, **58**, 166 (2013).
4. T. Li, M. F. Cai, M. Cai, *International Journal of Rock Mechanics and Mining Sciences*, **44**, 8, 1149 (2007).
5. V. I. Miroshnikov, A. V. Gladyr, *Problems of Complex Development of Georesources V International Scientific Conference*, 84 (2013).
6. I. Yu. Rasskazov, S. V. Cirel, A. O. Rozanov, A. A. Tereshkin, A. V. Gladyr, *Physical and technical problems of mineral development*, **2**, 29 (2017).
7. A. I. Orlov, *A nonparametric estimate of the density in topological spaces* (Science, 1983).
8. A. I. Orlov, *Nuclear density estimates in spaces of arbitrary nature* (Perm State University, 1996).
9. A. W. Bowman, *Biometrika*, **71**, 353 (1984).
10. M. Rudemo, *Scandinavian Journal of Statistics*, **9**, 65 (1982).
11. A. V. Gladyr, *Mining Information and Analytical Bulletin*, **21**, 51 (2016).
12. I. S. Enyukov. *Methods, algorithms, programs of multivariate statistical analysis: PPCA package* (Finance and Statistics, 1986).
13. M. Jambu, *Hierarchical cluster - analysis and compliance*, Finance and Statistics, (1988)
14. E. Patrik, *Fundamentals of pattern recognition theory* Sov. Radio, (1980).
15. V. S. Berikov, G. S. Lbov, *All-Russian Competitive Selection of Review and Analytical Articles on The Priority Direction «Information and Telecommunication Systems»*, 26 (2008).
16. J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function*, Plenum Press (1981)
17. D. E. Gustafson, W. C. Kessel, *Proc. of IEEE CDC*, 761 (1978).
18. R. Babuska. *Fuzzy Modeling for Control*, Kluwer Academic Publishers (1998).
19. X. L. Xei, G. A. Beni, *IEEE Trans. on Pattern Anal. and Machine Intell*, **8**, 841 (1991).
20. R. Yager, D. Filev, *Essentials of Fuzzy Modeling and Control*, John Wiley & Sons, (1984).