https://doi.org/10.59849/2219-6641.2024.2.28

MULTIFACTOR MODEL FOR SEISMOLOGICAL RESEARCH

A. A. Bayramov[1], F. N. Abdullayev[1], S. S. Suleymanov[1], R.D. Karimova[1]
H.N. Safarov[1], E. A. Rzayev[2]

azad.bayramov@yahoo.com, fataliabdullayev66@gmail.com,
samir-rpi@rambler.ru, rugijak@gmail.com,
safarov58@rambler.ru, rzayev@rambler.ru

Annotation: The study of risks in seismology or problems of earthquake forecasting are important tasks in modern earth science. The difficulty of solving these problems is that this task is multifactorial, because when assessing risk, many factors that affect the seismicity of an area should be considered. At the same time, to increase the credibility of the assessment, the degree of correlation between factors should be taken into account. Risk assessment depends on various geodynamic factors, which are determined individually for each area of the earth's surface. This paper discusses a multifactor method for implementing the task of assessing the risk of seismicity in an area.

Key words: *risk, seismology, multifactor method, earthquake, model.*

SEYSMOLOJİ TƏDQİQATLAR ÜÇÜN MULTİFAKTOR MODELİ

A A. Bayramov, F. N. Abdullayev,S.S. Süleymanov, R.D. Kərimova
H. N. Səfərov, E.Ə. Rzayev

Xülasə: Seysmologiyada risklərin öyrənilməsi və ya zəlzələnin proqnozlaşdırılması məsələləri müasir Yer haqqında elmin mühüm vəzifələrdir. Bu problemlərin həllinin çətinliyi bu vəzifənin çoxfaktorlu olmasıdır, çünki riskin qiymətləndirilməsi zamanı ərazinin seysmikliyinə təsir edən bir çox amillər nəzərə alınmalıdır. Eyni zamanda, qiymətləndirmənin etibarlılığını artırmaq üçün amillər arasında korrelyasiya dərəcəsi nəzərə alınmalıdır. Riskin qiymətləndirilməsi yer səthinin hər bir sahəsi üçün fərdi olaraq təyin olunan müxtəlif geodinamik amillərdən asılıdır. Bu məqalə bir ərazidə seysmiklik riskinin qiymətləndirilməsi tapşırığını həyata keçirmək üçün çoxfaktorlu metod nəzərdən keçirilir.

Açar sözlər: *risk, seysmologiya, çoxlufaktorlu metodu, zəlzələ, model*

МНОГОФАКТОРНАЯ МОДЕЛЬ ДЛЯ СЕЙСМОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

А.А.Байрамов, А.Н. Абдуллаев, С.С. Сулейманов, Р.Д. Каримова, Г.Н.Сафаров, Э.А.Рзаев

Аннотация: изучение рисков в сейсмологии или вопросы прогнозирования землетрясений являются важными задачами в современной науке о Земле. Сложность решения этих проблем состоит в том, что эта задача является многофакторной, т.к. при оценках риска следует рассматривать очень много факторов, которые воздействуют на сейсмичность района. При этом, для увеличения правдоподобности оценки, следует учесть степень корреляции между факторами. Оценка риска

[1] *Republican Seismic Survey Center of Azerbaijan National Academy of Sciences*
[2] *Strategic Scientific Research Center*

зависит от различных геодинамических факторов, которые определяются в частном порядке для каждого района земной поверхности. В данной статье рассматривается многофакторный метод, для реализации задачи оценки риска сейсмичности района.

Ключевые слова: *риск, сейсмология, многофакторный метод, землетрясение, модель*

Introduction

Seismology is largely an empirical science, since seismological models are based on experimental facts. The goal of empirical sciences is to capture an abstract model of the results of real observations. In order to draw conclusions about the real connection in the set of observations with the highest probability, probability theory methods are used, since no model can cover the entire set of observations. The goal of probabilistic (stochastic) modelling is to recognize the true structure of the system based on observed data.

Modern theories of the structure of the Earth, explaining the occurrence of earthquakes, are based on seismic observations. The main goal of geophysical research is to determine the structure of the environment based on observations of the characteristics of physical fields. The quality of seismic hazard analysis is determined by the attenuation factor of the intensity of seismic impacts.

The characteristics of the seismic impact of an earthquake are determined by the tectonic structure of the region, the depth of the source, the geometry of the source, the direction and course of the process of destruction of rocks, and other parameters. In addition, local geological features play an important role in the manifestation of the seismic effect at points on the surface.

The seismicity model of the earthquake zone determines the quality of the analysis. The seismicity of the region is characterized by:

1. The frequency of earthquakes;
2. The statistical distribution of the strength of tremors by magnitude;
3. The spatial distribution of sources,
4. Macroseismic observation of seismic events, seismic intensity, and the nature of damage.

Many tasks of seismology are associated with the possibility of calculating the probabilities of some events associated with the seismic process. The main task of seismology is to learn to predict the strength, time and place of earthquakes.

Seismic processes occur and develop in time and space under the influence of the internal determinism of global tectonics. Uncertainties associated with the interplay of the Earth's internal physical fields and the gravitational force of celestial bodies, as well as their influence on global tectonics, introduce an element of randomness into seismicity models [Burtiyev, 2017].

Multifactor analysis in seismology

Seismic processes are diverse, since their formation is caused by various geological and geophysical processes occurring in the depths of the Earth. These processes are characterized by many different parameters, which are characterized by multifactor random variables. In the study of such multiparameter processes, multivariate analysis is used [Tarasov, 2011].

Multifactor analysis is based on determining the minimum number of factors that make up the largest share of data variance. In studying the complex nature of seismicity, factor analysis helps to better understand the essence of seismic processes. Factor analysis allows us to study the structure of the relationship of variables, where each group of variables will be determined by the factor for which these variables have maximum loads.

The result of factor analysis is the transition from a set of initial variables to a smaller number of new variable factors. Factor analysis solves the problem of reducing the number of variables with minimal loss of initial information. If we assume that the correlations between variables are explained by the influence of hidden factors, then the main task of factor analysis is to analyze the correlations of a set of parameters.

Most factor analysis methods are based on principal component analysis, which transforms a group of correlated original variables into another group of uncorrelated variables. Principal component analysis solves the problem of reducing the number of variables while maintaining the maximum proportion of variance in observations by selecting only the main components that explain the majority of variance.

To improve the efficiency of the seismicity risk assessment process in a given area, it is necessary to be able to conduct forecasting based on multifactor models.

The main stages of forecasting in a multifactor model are:

1) selection of factors that are not correlated with each other;

2) determination of the parameter values of the multifactor forecasting model;

3) verification of the adequacy of the resulting multifactor forecasting model;

4) application of the multifactor forecasting model.

The main stages of factor analysis: calculation of the correlation matrix; extraction of factors; selection and rotation factors; interpretation of factors; calculation of coefficient values; assessment of the quality of the model. For example, in [Ioane, 2021], an array of seismic data is analyzed, consisting of parameter values characterizing the mechanism and geometry of the Vrancea seismic zone in Romania. It is assumed that the values of these parameters are due to hidden factors that cannot be observed directly. Identification of these factors is the task of factor analysis.

The task of factor analysis is to determine the minimum number of factors that contribute most to data dispersion. The use of factor analysis begins with the calculation of descriptive statistics of the array of observed values for 16-dimensional random variables characterizing the seismicity mechanism and geometry of the Vrancea source. Factor analysis allows us to identify the relationships that determine the correlation interdependence between seismic parameters. It is assumed that the cause of a significant correlation between the parameters are the sought-after factors. The task of factor analysis is to determine the minimum number of factors that contribute most to data dispersion. The values of the statistical characteristics of the parameters in this work have a large spread, that is, their values are measured in different units that are incompatible with each other, so they are scaled using standardization (1):

$$X_{ij}^{s} = \frac{x_{ij} - \bar{X}_j}{\sigma_j} \qquad (1)$$

where: *Xij* is the value of the *j* parameter in the *i* observation, *Xi* is the value and *σi* is the standard deviation of the *j* parameter.

The values of the statistical characteristics of the parameters in the experiment [Ioane, 2021] have a large scatter, and their values are measured in different units that are incomparable with each other, so they should be brought to a single scale using standardization. The first idea of the presence of dependent parameters can be obtained from the correlation matrix, which characterizes the degree of correlation between the parameters of the original array. The higher the proportion of high correlations, the better the data are suitable for factor analysis. For example, the value of the correlation coefficient between the parameters np_{1stk} (direction of nodal plane 1) and np_{2stk} (direction of nodal plane 2) is 0.726 [Burtiyev, 2017]. The value of the correlation coefficient between the parameter p_{az} (azimuth of the compression axis) and np_{1dp} (angle of incidence of the nodal plane - plane of zero displacements 1) is 0.884 [Burtiyev, 2017], which indicates a sufficiently high degree of dependence between the parameters and is the basis for including them in one group. In factor analysis, the correlation matrix is transformed, after which all non-diagonal elements of the correlation matrix become zero, and the diagonal elements change their values. This means that the parameters become independent of each other. Most factor analysis methods are based on principal component analysis, which transforms a set of correlated input parameters into a set of uncorrelated factors. In this method, after calculating the correlation matrix, it is orthogonally transformed. The factor loadings are determined by the values of the matrix elements. Recall that the factor loading is the value of the correlation coefficient of each of the original features with each identified factor. The closer the connection of a given feature with the factor in question, the higher the value of the factor loading. The most common rotation method is Varimax [Athar, 2023] (uncorrelated factors). The method is based on finding the eigenvalues and eigenvectors of the correlation matrix from the solution of the equation:

$$R = AA' \quad (2)$$

where: *R* is the original correlation matrix; *A* is the matrix whose elements are the loading factors a_{ij} of the parameter *i* by the component *j*; *A'* is the transposed matrix.

The development of a more complex and adequate attenuation model is the subject of study by many seismologists. Thus, the analysis of macroseismic and instrumental data of the Vrancea mid-depth earthquakes revealed some features of the earthquake impact:

- impact on large areas with a predominant North-East and South-West orientation;

- a high degree of dependence of the amplitude of seismic ground displacements on local and regional geological conditions compared to the magnitude and distance from the source;

- high variability of the parameters of strong ground movements; reflection of the earth's surface relief by isolines [Ismail-zade, 2007].

In the near zone, where $r \sim h$ (*r* is the distance to the earthquake epicenter, *h* is the depth of the earthquake source), the geometry of the earthquake source has a decisive influence on the configuration of the macroseismic field [Shebalin, 1961].

From the values of descriptive statistics it is evident that the values of statistical characteristics of seismic parameters have a large spread, and their values are initially measured in different units that are incompatible with each other, so they should be brought to a single scale. using standardization (1).

The values of the regression coefficients are estimated by the least squares method, and for consistency, biaslessness and efficiency of estimates, the Gauss-Markov conditions must be met:

$$E(\varepsilon_i) = 0, \ (i=1,n) \quad (3)$$

The variance of the random component in all observations must be constant (homoscedasticity), equal to zero, i.e.

$$D(\varepsilon_i) = \sigma^2, \ (i=1,n) \quad (4)$$

Recall that homoscedasticity is a property that denotes the constancy of the conditional variance of a vector or sequence of random variables. It is assumed that the random component has a normal distribution, i.e.

$$E(\varepsilon_i \varepsilon_j) = 0, \ \varepsilon_i \sim N(0;\ \sigma^2), \ (i \neq j) \quad (5)$$

The multivariate linear regression model is expressed by the following expression:

$$y_i = \alpha + \beta x_1 + \beta x_2 + \ldots + \beta x_{18} + \varepsilon_i \quad (i=1,n) \quad (6)$$

where: *y* is a dependent random variable, $x_1 \ldots, x_j$ are regression variables; $\beta_1, \beta_2, \ldots, \beta_j$ are regression coefficients; ε is a random variable. The appearance of a random component:

$$\varepsilon_i = y_i - y^\wedge \quad (7)$$

where: y_i is the real value, $y\hat{}$ are the values of the dependent variable calculated using the regression equation. They are associated with the influence of unaccounted factors on the dependent variable y, the inadequacy of the selected model and observation errors.

To explain 100% of the variance in seismic data from the experiment [Ioane, 2021], 16 components are required. However, using the factor extraction procedure, it was possible to extract only 4 factors out of 16, which explains 84,996% of the variance in the data set. That is, a factor model consisting of 4 factors preserves 84.996% of the original information.

When grouping the original array of parameters, information loss is inevitable. Saving information by only 60% is considered a fairly good indicator. Typically, when conducting factor analysis, the first principal components are used, the total share of variance of which exceeds 60%. Considering that during factor analysis the number of parameters is reduced several times, even with a large loss of information, for example, by 40%, it is possible to use a factor model.

The graph of eigenvalues is informative in determining the number of factors that begin to decline sharply. After a given number of factors, the dependence under study is close to the horizontal line, that is, the decrease in eigenvalues slows down. The inflection point is on the 4th factor, i.e. no more than 4 factors are identified. This method is used to determine the number of factors before rotation, the purpose of which is to determine a simple model. In this model, each parameter corresponds to a large value of factor loading for one factor and a small value for all the others.

The factor load is determined by the correlation coefficient of each parameter with each of the identified factors. The sum of the squares of the column loads is the dispersion of the coefficient and serves as a measure of its informativeness. In the principal component method, the dispersion of the principal components is equal to the eigenvalues of the correlation matrix of the original variables. When implementing the stage of selecting factors that do not correlate with each other, the sample values of the paired correlation coefficient (r_{ij}) between the factors x_i and x_j are used. If the condition is met [Tikhomirov, 2003]

$$|r_{ij}| > (0{,}7\text{–}0{,}8), \quad (1)$$

then there is a significant connection between these factors. In this case, one of the factors should be excluded so that the same factor is not taken into account twice. Let us denote the risk as R, and $x_1,\ldots,x_n$ as geodynamic factors influencing the risk of seismicity, and the values of the pair correlation coefficient (r_{ij}) between factors xi and x

$$|r_{ij}| < 0{,}6 \quad (2)$$

A multiple linear regression model is proposed as a predictive multifactorial model.

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_4 \cdot x_4 + a_9 \cdot x_9 + a_{10} \cdot x_{10} \quad (3)$$

To estimate the parameters of this model (3), the least squares method was chosen.

$$A = (E^T E)^{-1}(E^T Y), \quad (4)$$

where E is the matrix of factor values, Y is the vector of statistical data (time series), T is the transposition operation, (–1) is the operation of finding the inverse matrix. When checking the adequacy of the obtained models, it is recommended to conduct an additional statistical analysis, namely: to check the significance of the coefficients of the forecast models.

$$H0: bj = 0,\ H1: bj \neq 0,\ j=1, 2, 4, 9, 10.$$

The t-statistic is recommended as a criterion [Lemeshko, 2008]

$$t_j = \frac{a_j}{s(a_j)} \rightarrow t(\nu_0) \quad (5)$$

where: a_j – estimates of the coefficients b_j obtained by the least squares method; $s(a_j)$ – estimates of the standard deviation of the coefficients a_j; the value (5) under known restrictions [Lemeshko, 2008] has a *t*-distribution with the number of degrees of freedom v_0, in our case $v_0 = 7$.

The following were chosen as adequacy criteria: Fisher's F-criterion, the Durbin-Watson criterion, the criterion of "ascending" and "descending" series, and the criterion of peaks. The Durbin-Watson criterion checks the most important premise of regression analysis – the absence of autocorrelation in the sequence of residuals e_t.

The value of the criterion is calculated using the formula

$$d = \frac{\sum_{t=2}^{m}(e_t - e_{t-1})^2}{\sum_{t=1}^{m} e_t^2}$$

This value is compared with two table levels: the lower d_1 and the upper d_2. If the obtained value d is greater than two, then before comparison it must be transformed:

$$d' = 4 - d.$$

If d (or d □) is in the interval from zero to 1 d, then the residual values are strongly autocorrelated. If the value of the d -criterion falls in the interval from d_2 to 2, then there is no autocorrelation. If d_1□□d□□d_2, an unambiguous conclusion about the absence or presence of autocorrelation cannot be made and it is necessary to use another criterion, for example, the first-order autocorrelation coefficient.

Conclusion

The study of risks in seismology or earthquake forecasting issues are important tasks in modern Earth science. The complexity of solving these problems is that this task is multifactorial, since when assessing risk, it is necessary to consider many factors that affect the seismicity of the area. At the same time, to increase the credibility of the assessment, it is necessary to take into account the degree of correlation between the factors. And in some cases, it is also necessary to study the synergy of processes, when the simultaneous impact of several factors has more than the total impact. Risk assessment depends on various geodynamic factors, which are determined individually for each area of the earth's surface. This article discusses a multifactor method for implementing the task of assessing the risk of seismicity of the area.

References

1. Athar Mahmood, Manisha Seth The Dynamics of Cognition Process of Micro, Small, and
2. Burtiev Rashid, Seismic Hazard Assessment Method Based on the Stochastic Models of Seismicity, Bulletin of the International Institute of Seismology and Earthquake Engineering, Vol.51, 2017, Tsukuba, Japan, 22-38.
3. D. Ioane, I. Stanciu Vrancea seismic zone – a new geophysical model based on wrench tectonics, volcanism and regional geodynamics. Rev. Roum. GÉOPHYSIQUE, 65, p. 3–47, 2021
4. Ismail-Zadeh, A., Sokolov, V. and Bonjer, K., Tectonic stress, seismicity, and seismic hazard in the southeastern Carpathians, Nat. Hazards, Springer, 2007, 1-22.
5. Lemeshko B.Yu., Lemeshko S.B. On the stability and power of criteria for testing the homogeneity of averages. Measurement technology, 2008, No. 9. – P.23-28 (original in Russian)
6. Medium Enterprises. International Journal of Asian Business and Information Management. 2023. Vol. 14, Is. 1. pp. 1-12.
7. Shebalin N.V. The intensity, magnitude and depth of the earthquake source. Earthquakes in the USSR, Moscow, Science, 1961, 126–138.
8. Tarasov, V. E. Fractional Dynamics: Applications of Fractional Calculus to Dynamics of Particles, Fields and Media, New York: Springer, 2011, 505 p
9. Tikhomirov N.P., Dorokhina E.Yu. (2003). Econometrics. Moscow: Exam. 512 p. (original in Russian).