

On Identification of Anomalies in Multidimensional Hydrogeochemical Data as Earthquake Precursors

Evgueni A. Haroutunian¹, Irina A. Safaryan¹,
Hrachya M. Petrosyan² and Armine R. Gevorkyan²

Institute for Informatics and Automation Problems of NAS RA
Armenian National Survey of Seismic Protection of the Ministry of Emergency Situations of RA
e-mail: eghishe@sci.am

Abstract

We consider two nonparametric methods are discussed for identification of anomalies in multidimensional hydrogeochemical data with the goal of forecasting major seismic events. The first method is based on using the threshold copula function aimed at the definition of change moments in the structure of dependencies between the components of a multidimensional vector of observations. The second method is designed to transform multidimensional data to one-dimensional which allows to use rank score tests.

Keywords: Seismic-forecasting anomalies, Hydrogeochemical precursors of strong earthquake, Rank score test, Threshold copula, Pooled sample method.

1. Introduction

The processes occurring in the upper layers of the Earth's crust prior to large earthquakes, cause abnormal changes in a number of different geophysical and geochemical indicators. An important component of the seismic risk reduction is the continuous monitoring and evaluation of the current seismic hazard. The current seismic hazard assessment (SHA) being carried out at NSSP of RA is based on the data from multiparameter observational network. The processing of monitoring data is presented in monographs of H. Petrosyan [1, 2].

Hydrogeochemical parameters take a special place among the monitored data since the moving parts of the earth's crust-ground waters and gas are considered as informative parameters of monitoring. In particular, it is recognized that the percentage of dissolved chemicals and gases observed in special wells changes prior to strong seismic events. Such changes are observed quite stable for several months before the earthquake, and therefore can serve as its medium-term precursor. The regional geochemical network of NSSP consists of a

number of monitoring stations; probes are taken once a day and analyzed for 14 microcomponents.

On the territory of the Russian Federation the monitoring of hydrogeochemical indicators is conducted by Kamchatka Branch of the Geophysical Survey of Russian Academy of Sciences. More than 20 parameters are measured and analyzed with the interval between the scheduled observations in the range of 3 to 6 days. The statistical analysis of results of multidimensional hydrochemical series is presented in the paper of G. Rjabinin and J. Khatkevich [3] and in the monograph by A. Ljubooshin [4].

In 1997-2000 a computer program implementing a nonparametric algorithm (based on rank score tests) was developed in IPIA which was designed to detect changes in statistical properties of one-dimensional random sequences. The program was applied to treat the hydrogeochemical data for 1985-1995 from the stations Kajaran, Ararat and Surenavan provided by NSSP. Theoretical fundamentals of the algorithm are stated in [5-7], the results of computer processing and the program description are in [8-10].

In this paper several new results presented at the conferences [11-12] in the direction outlined, are devoted to a comparative analysis of time series of helium taken from the three points of observation, as well as to changes of their joint distribution on the eve of major seismic events. The behavior study of the joint distribution function of several indicators in the periods between the strong earthquakes has several objectives:

- forecasting of earthquake parameters such as time coordinates of the epicenter, magnitude, energy class, etc.; one-dimensional time series allow predicting only time;
- derivation of a reliable precursor; for multivariate hydrogeochemical data the structure of time correlation is such a sign indicated in [2];
- working out of a one-dimensional integrated indicator for the seismic hazard.

For detection of changes in multivariable dependence structure in this paper threshold copulas investigated in [13-15] and for obtaining an integrated one-dimensional indicator the method of pooled sample is applied.

2 The Comparative Analysis of Data on Helium from Three Observation Stations

In identification of precursors with statistical methods an anomaly is interpreted as a change of statistical characteristics of the observed series of indicator values which is in some sense conformed to the physical model of a seismic event preparation. The most successful, as it is defined in [1], is the model of consolidation by Dobrovolsky, according to which the cycle of a single earthquake has 4 phases: I – regular state, II – development of heterogeneity (phase of consolidation), i.e. appearing of long-term and medium-term precursors, III – phase of destruction which has two stages (the stage of a short term precursors and for-shocks, namely at this stage the earthquake occurs, and the stage of aftershocks) and IV – restoration of regular state, that is returning to the phase I.

Thus, we present the following working hypothesis. We assume that the data of observation is a sequence of independent random variables which before the earthquake contain at least two change-points (disorder moments), and at least one change-point after the earthquake. The first change-point n_1 corresponds to the beginning of the phase II, the second n_2 – to the moment of coming to some quasi-stable level, i.e. to the middle of the phase II, and the third n_3 – to the start of the phase IV. An identification of precursor consists in obtaining of robust estimates of the moments n_1 and n_2 .

The change in statistical properties is interpreted as a change of the distribution function for the elements of series. The parametric form of distribution and any type of change is not postulated a priori which assumes the application of nonparametric (distribution-free) statistical procedures to identification of these changes. Schematically a number of moments of disorder accompanying the earthquake can be depicted as it is shown on Figure 1.

Thus, as a rule, changes are associated with an increase or decrease in the percentage of micro-component, so it is assumed that $F_2 x = F_1 x - a, a > 0$ where the distribution function $F_1(x)$ reflects the variations of background of the observed indicator. However, not always return to the regular state means that $F_3 x = F_1 x$ in general $F_3 x = F_1 x - b$ and either $b > 0$ or $b < 0$.

In some cases, identification of moments n_1 and n_2 is possible visually. For their identification with statistical methods a sequence of rank score statistics is applied, which is defined as follows:

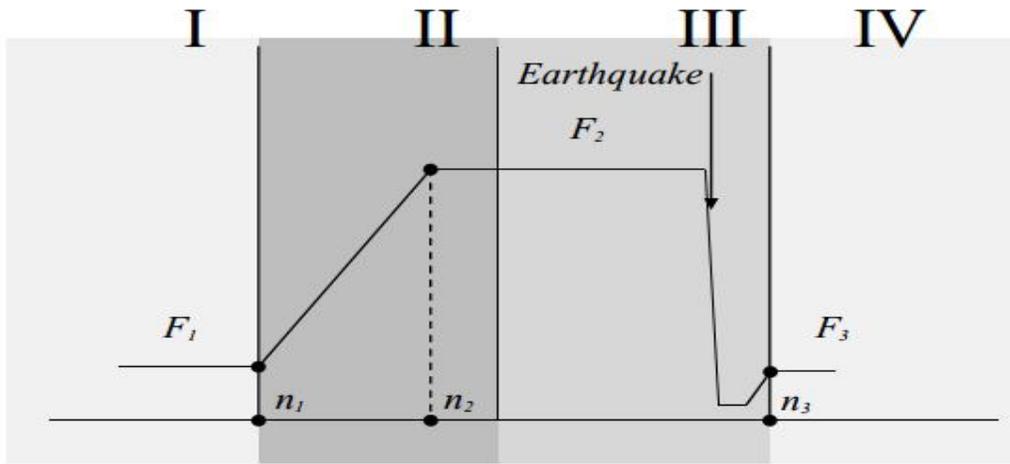


Fig. 1. Disorder moment series before and after a single earthquake

Let $R_i = \# X_n: X_n \leq X_i, n = 1, \dots, N, i = 1, \dots, N$ be ranks of observations X_1, \dots, X_N and

$$T_j^n = \frac{1}{n} \int_0^n \frac{R_i}{N+1} du, \quad A_j = \int_0^1 J u du.$$

The sequence of rank score statistics is defined as follows:

$$W_N^n = \frac{n}{N-n} T_j^n - A_j, \quad n = 1, \dots, N-1.$$

Then under the condition of an appropriate choice of the score function the number of observation defined as $\vartheta = \operatorname{argmin} W_N^n$ is consistent estimate either for the moment n_1 or for n_2 depending on the position of the window in which the sequence of statistics is calculated. Based on theoretical results [5-7] it can be stated that if in the linear plot of the sequence of statistics W_N^n both global and local minimums immediately follow each other and go beyond the critical value then the anomaly is considered as identified.

The appropriate choice of the score function means that for distributions differing in shift $J(u) = u$, then $W_N^{n, N}$ is called a Wilcoxon statistics, for distributions differing in scale $J(u) = (u - \frac{1}{2})^2$ and $W_N^{n, N}$ is called a Mood statistics. In general case distributions $F_1(x)$ and $F_2(x)$ differ both in shift and scale, while statistics $W_N^{n, N}$ is a linear combination of statistics of Wilcoxon and Mood, the coefficients of which can be obtained as sample estimates in retrospective analysis. An algorithm for producing such estimates will be elaborated and included in the new version of the program.

Time series of helium on the three monitoring stations Ararat, Kajaran, Surenavan over the period 1985-1989 and behavior of sequences of Wilcoxon statistics, computed in a window including several seismic events, are presented on Figure 2. Dates of earthquakes occurring in the region in that period are marked with vertical lines. On the left vertical axis values of indicator of helium and on the right axis the values of Wilcoxon statistics are scaled.

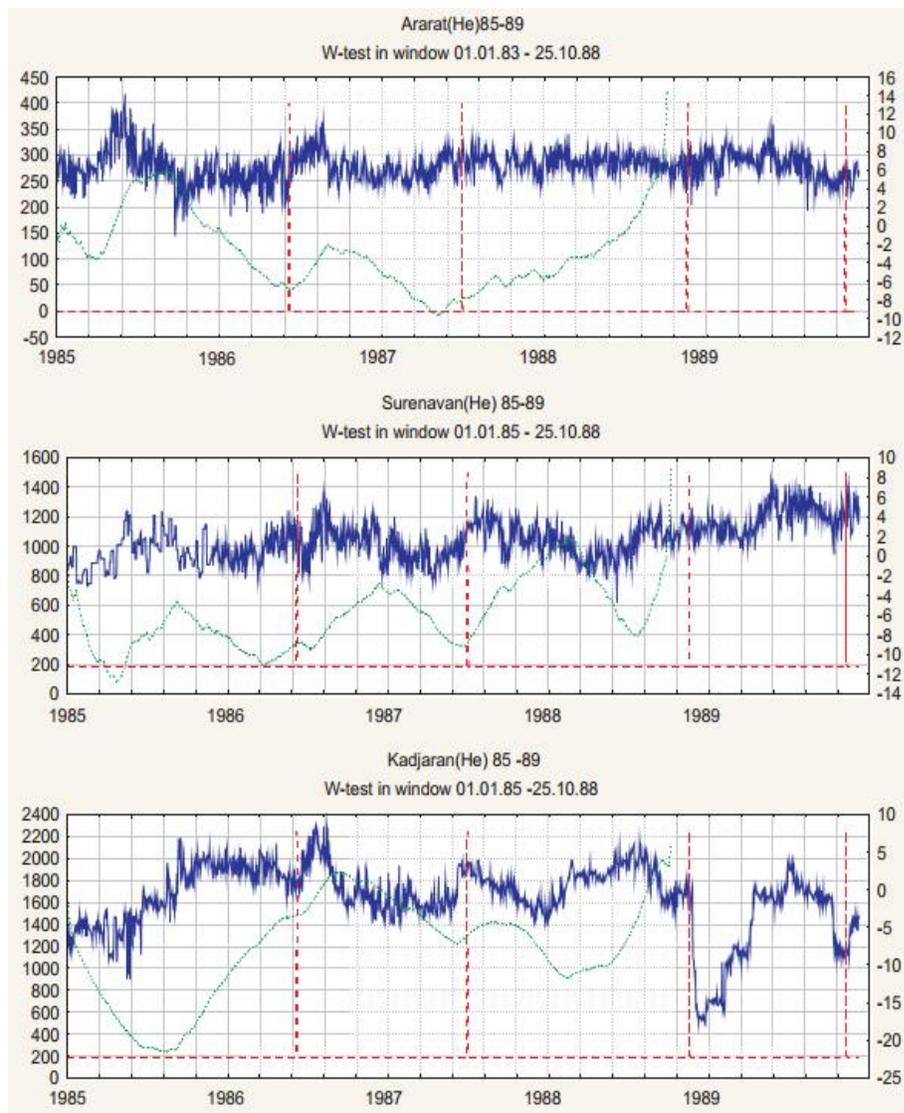


Fig. 2. Behavior of the sequence of Wilcoxon statistics in window 01.01.85-25.10.88 and time series of helium over the period 1985-1989 for the stations Ararat, Surenavan, Kajaran.

The comparative analysis showed that each observation point (Ararat, Kajaran, Surenavan) is characterized by its precursors anomaly which differs by its appearance term. All the three monitoring stations respond to the earthquakes of the period 1985-1989. The Wilcoxon test allows detecting seismic-forecasting anomalies (i.e. change-points n_1 and n_2 marking the periods of growing increase of the average value of helium indicator) before each earthquake. The station Ararat most significantly reacted to the earthquake of 1988 (Spitak), the stations Kajaran and Surenavan - to the earthquake in 1986, though a delay was detected on the station Surenavan. On the base of exclusively visual analyses of the time series it is not possible to obtain these conclusions.

3. Threshold Copula Models for Identification of Seismic-Forecasting Anomalies

An important index of upcoming seismic event is a change in the structure of dependence between different indicators of one station or one indicator for different stations. In [2] a graph of correlation dependencies between various hydrogeochemical indicators of different stations of the NSSP is presented. In particular it is confirmed that significant correlation between different stations in data on helium is absent.

A contemporary approach to assessment of the structure of dependence is based on the use of the copula function. For theoretical and practical applications of the copula function we refer to the monograph of Nelsen [16]. One threshold copula, introduced in [13-15], can be used in case when the dependence between the components of two-dimensional data arises in time as a precursor of upcoming seismic event.

We call a random variable $RV Y$ homogeneous with respect to $RV X$ if for all pairs x, y on the plane the following conditional probabilities are equal: $\Pr Y \leq y/X \leq x = \Pr Y \leq y/X > x$.

The concept of homogeneity is equivalent to the notion of independence. If there exists a unique value of $x = \mu$ such that for all $y \in \mathbb{R}$, $\Pr Y \leq y/X \leq x = \Pr Y \leq y/X \leq \mu$ for $x \leq \mu$,

$$\Pr Y \leq y/X > x = \Pr Y \leq y/X > \mu \text{ for } x > \mu,$$

and

$$\Pr Y \leq y/X \leq \mu \neq \Pr Y \leq y/X > \mu,$$

then the statistical dependence between X and Y is called one-threshold and the value μ is called a threshold.

Hypotheses on the homogeneity indicator of helium on stations Ararat and Surenavan with respect to that on the station Kajaran for the period 01.01.85 - 03.02.86 were tested. The indicators on stations Surenavan and Kajaran are independent. Meanwhile when comparing the stations Ararat and Kajaran we found that the value of helium threshold, which was equal to 1571, registered at the station Kajaran dated 15.08.85 proves to be separating for the indicator at the station Ararat. The two-dimensional histograms built for copulas, depicted on Figure 4 present evidence that the point 15.08.85 is actually the moment of change in the structure of dependence.

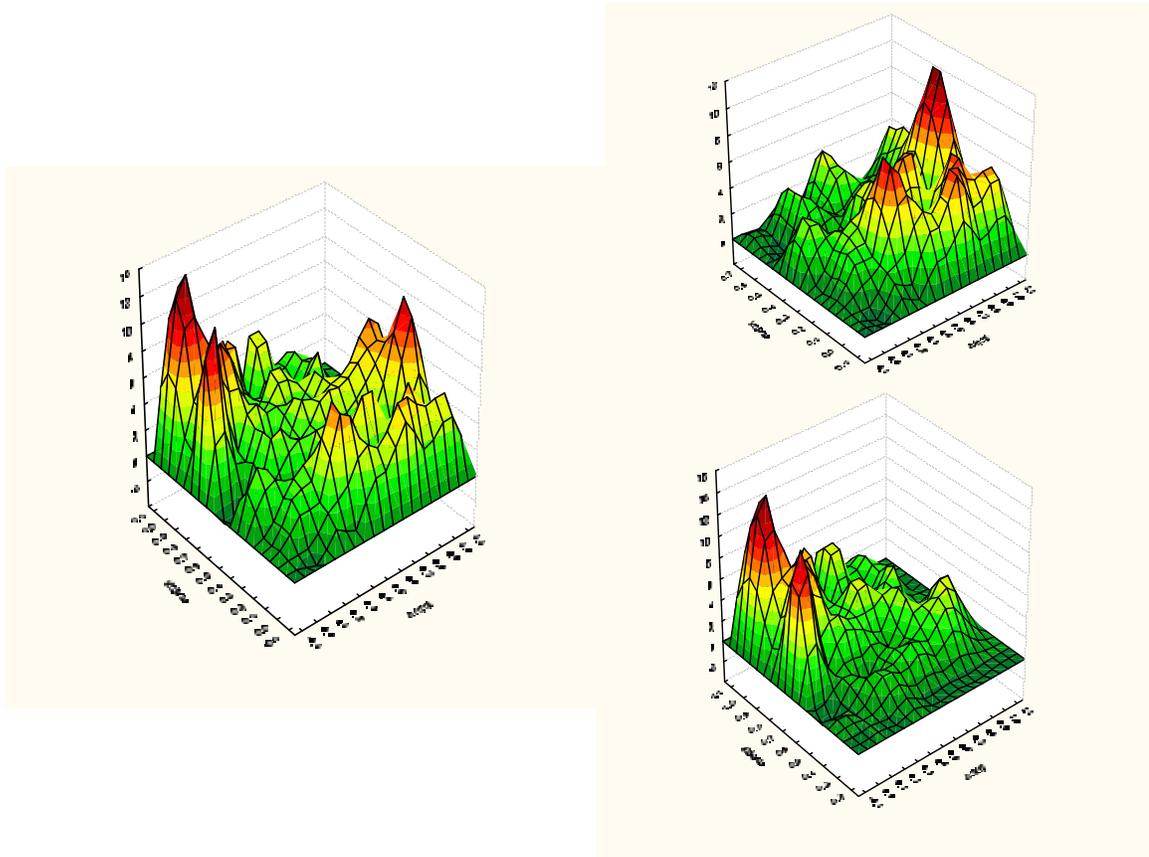


Fig. 4. Copula density functions: left - for period 01.01.85 - 03.02.86, right - for periods 01.01.85-15.08.85 and 16.08.85 – 03.02.86; data of earthquake: 13.05.86, change-point moment data 15.08.85.

4. Pooled Sample Method

If the observed series refer to the same physical quantity the method of mixed samples can be used as well to detect the anomalies preceding the earthquake. There exist different ways of mixing.

We joined the data on helium over 14 months preceding to Spitak earthquake in the following way: the odd numbers correspond to observations at the station Kajaran, while for the first sample the even observations registered at the station Surenavan and for the second sample the even observations multiplied to six corresponded to the ones registered at the station Ararat that day. The pooled samples are presented on Figure 5.

Then, using the Wilcoxon statistics to the combined samples, we obtained that for the first sample a global minimum statistic falls at 541, which corresponds to the date 26.06.88 and for the second - at 244, which corresponds to the date 11.02.88. Thus, by using a mixed sample you can get a reliable estimate of the beginning of the accumulation of changes than by one-dimensional samples.

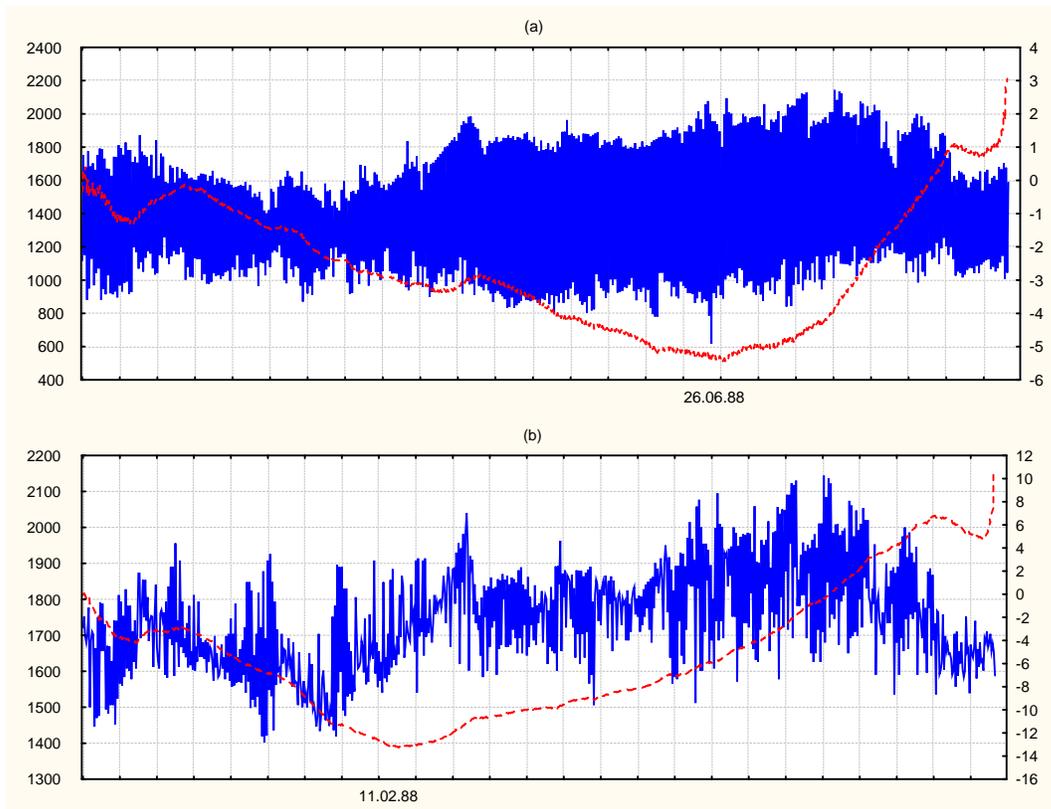


Fig. 5. Surenavan-Kajaran pooled sample – (a), Ararat-Kajaran pooled sample – (b).

5. Conclusion

The presented results show the prospects of application of threshold copula methods and mixed sampling to determination of anomalies in multidimensional hydrogeochemical data occurring prior to earthquakes. Further theoretical elaboration of algorithms and implementation of programs for their realization is planned.

Acknowledgement

This work was supported in part by SCS of MES of RA under Thematic Program No SCS 13-1A295.

References

- [1] H. M. Petrosyan, *Earthquake testing and prognosis*, Yerevan, 2004.
- [2] H. M. Petrosyan, *Precursors and Prognosis of Earthquakes on the Territory of the Republic of Armenia*, Yerevan, 2009.
- [3] G. D. Rjabinin and Yu. M. Khatkevich, “On the issue of possibility of the earthquake prediction with hydrogeochemical methods”, *Materials of the IV All- Russia Symposium on Volcanology and Paleo-volcanology*, pp. 660-663, 2009.

- [4] A. A. Ljubooshin, *Analysis of Data of Geophysical and Ecological Monitoring*, Moscow, Nauka, 2007.
- [5] D. G. Asatryan and I. A. Safaryan. “Nonparametric methods for detecting changes in the properties of random sequences”, In: *Detecting Changes in Random Processes*. Ed. by L. Telksnys, New York, pp.1-13, 1986.
- [6] E. A. Haroutunian and I.A. Safaryan. “Nonparametric consistent assessment of moments of property changes of random sequences”, *Mathematical Problems of Computer Science*, vol. 17, pp. 76-85, 1997.
- [7] I. A. Safaryan, “Nonparametric estimation under gradual change of the random sequence”, *Statistical Problems of Control*, vol. 83, pp.121-126, 1988.
- [8] E. A. Haroutunian, I. A. Safaryan, P. A. Petrossian and H. V. Nersessian, “Earthquake precursor identification on the base of statistical analysis of hydrogeochemical time series”, *Mathematical Problems of Computer Science*, vol. 18, pp. 33-39, 1997.
- [9] H .V. Nersessian and E. A. Haroutunian, “Statistical nonparametric analysis of hydrogeochemical data of the Ararat station HSSP of Armenia”, *Mathematical Problems Computer Science*, vol. 19, pp. 40-44, 1988.
- [10] P. A. Petrossian, “Software implementation of the algorithm of detection of moments in the state changes of time series”, *Mathematical Problems of Computer Science*, vol. 19, pp. 32-39, 1988.
- [11] E. A. Haroutunian, I. A. Safaryan and A.O. Yesayan, “Application of special statistical algorithms of the hydrogeochemical time series monitoring with the goal of the earthquake prognosis”, *Proceedings of the Conference dedicated to the 23-d anniversary of the Spitak Earthquake*, MES, Yerevan, 2011.
- [12] I. A. Safaryan and A. O. Yesayan, “Copula-model applications to seismic-forecasting anomalies detection”, *Multivariate Statistical Analysis and Econometrics, Proceedings of VIII International School-Seminar*, Armenia – Tsakhkadzor. M., CEMI RAS, p. 189, 2012.
- [13] E. A. Haroutunian and I. A. Safaryan. “Copulas of two-dimensional threshold models”, *Mathematical Problems of Computer Science*, vol .31, pp. 40-48, 2008.
- [14] E. A. Haroutunian and I. A. Safaryan. “On certain threshold copulas estimators”, *Abstracts of Internatioal Conference on Computer Science and Information Technologies*, Yerevan, pp. 145-147, 2009.
- [15] E. A. Haroutunian and I. A. Safaryan. “On estimation of threshold parameter in three-dimensional copulas model”, *Abstracts of Internatioal Conference on Computer Science and Information Technologies*, Yerevan, pp. 129-131, 2011.
- [16] R.V. Nelsen, *An Introduction to Copulas*, Springer, New York, 2006.

Submitted 10.09.2013, accepted 16.10.2013.

Երկրաշարժերին նախորդող բազմաչափ հիդրոերկրաքիմիական տվյալներում անկայունության նույնականացման մասին

Ե. Հարությունյան, Ի. Սաֆարյան, Ի. Պետրոսյան և Ա. Գևորգյան

Ամփոփում

Քննարկվում են հզոր սեյսմիկ իրադարձությունների կանխատեսման նպատակով բազմաչափ հիդրոերկրաքիմիական տվյալներում անկայունության նույնականացման երկու ոչ պարամետրիկական մեթոդներ: Առաջինը հիմնված է դիտարկումների բազմաչափ վեկտորի բաղադրիչների միջև կախվածությունների կառուցվածքի փոփոխության պահի որոշման նպատակով շեմային կապակցիչի օգտագործման վրա: Երկրորդ եղանակը կառուցված է ըստ բազմաչափ տվյալների միաչափերի վերածման միտման, որը թույլ է տալիս կարգային նշումների տեստերի օգտագործումը: