

# Methods for Analyzing Heterogeneous Data in the Tasks of Assessing Territorial Risks

Olga V. Taseiko<sup>a,b</sup>, Uliana. S. Postnikova<sup>a,c</sup>, Margarita Georgieva<sup>d</sup>, Hranislav Milosevic<sup>e</sup> and Stefan Panic<sup>e</sup>

<sup>a</sup> Krasnoyarsk Branch of the Federal Research Center for Information and Computational Technologies Krasnyarsk, 660049, Russia

<sup>b</sup> Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochoy Av, Krasnyarsk, 660037, Russia

<sup>c</sup> Siberian Federal University, 79 Svobodny pr., Krasnyarsk, 660041, Russia

<sup>d</sup> Forest Research Institute, 132 "St. Kliment Ohridski" Blvd., Sofia, 1756, Bulgaria

<sup>e</sup> University of Pristina, Filipa Visnjica bb, 38220, Kosovska Mitrovica, Serbia

## Abstract

The safety of territorial entities is characterized by the presence of facilities with functionally complex and highly hazardous production systems, which carry obvious threats to the life and health of the population (risks of accidents and disasters) and concealed ones (environmental pollution). The presence of a large number potentially dangerous objects increases the probability of accidents and natural, social and man-made character disasters. For safety analysis and risk assessment, it is proposed to use multidimensional statistics methods, which make it possible to divide the territory into homogeneous groups with similar characteristics. Based on the breakdown results, a quantitative assessment of the risk within each individual group was determined.

## Keywords 1

territorial risk, multivariate statistics, hierarchical cluster analysis

## 1. Introduction

The technique and technology development leads to the emergence of a huge range of threats, a risk-based approach is introduced for their analysis, assessment and forecasting. Today, there are good practical, social and economic reasons for using it. First of all, these reasons include:

1. determination of acceptable risks levels and development of alternative proposals for their reduction;
2. increased consistency to allow comparisons between different geographic areas.

A common risk understanding implies the "uncertainty" of the outcome of any random events. Uncertainty has two possible realizations: the event is "beneficial", or - will cause "harm". The essence of the risk-based approach is the classification of objects according to the level of danger (threat of harm to life and health of citizens, damage to the environment, cultural heritage, threats of natural, anthropogenic and social emergencies) [1].

The global approach to managing natural and man-made disaster risks is based on the UN Sendai Framework for Disaster Risk Reduction (2015–2030), adopted at the UN World Conference on March 14–18, 2015. Main areas of activity: concept of disaster risk; development and improvement of legal documents on risk management; funding for disaster risk reduction and resilience building; ensuring effective management and response.

---

VI International Conference Information Technologies and High-Performance Computing (ITHPC-2021), September 14–16, 2021, Khabarovsk, Russia

EMAIL: taseiko@gmail.com (A. 1); ulyana-ivanova@inbox.ru (A. 2); margaritageorgiev@gmail.com (A. 3); mhrane@gmail.com (A. 4)

ORCID: 0000-0002-0314-4881 (A. 1); 0000-0002-1535-3576 (A. 2); 0000-0003-3165-1992 (A. 3)

© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

Thus, the sustainable development of the territory is directly related to the analysis, assessment and minimization of territorial risks. The problem of ensuring the safety of industrial agglomerations, which is characterized by the presence of a large number of industrial facilities and a developed infrastructure, is especially acute. Territorial risk assessment and analysis are important tools for improving regional policies, strategies and tactics that will minimize impacts in a given area.

Theoretical and practical methods of risk analysis are presented in a number of Russian [2-9] and foreign publications [10-15]. Currently, various methods of forecasting and assessing risks have been developed, which models are classified depending on the source of occurrence, object of influence and purpose. Risk assessment models and methods can be divided into two groups: for the analysis of the safety of industrial facilities and for the analysis of a territorial entity. The aim of the work is to develop a method for analyzing territorial risks to improve sustainable development and effective management of industrial agglomerations.

## 2. The safety analysis of the industrial area

The need to assess territorial risks is determined by the following provisions [16]:

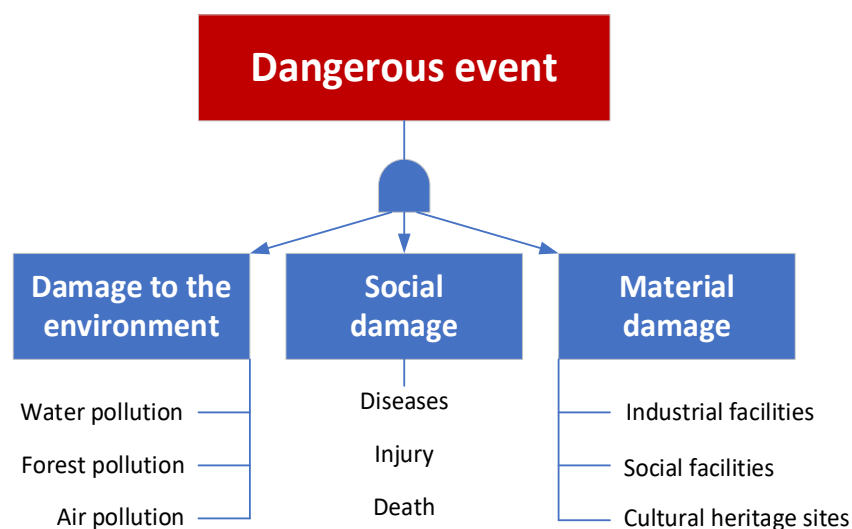
1 Control and regulation of social, natural and technogenic safety is one of the main factors in stabilizing crisis phenomena in the economy, ensuring the safety and functioning of fixed assets.

2 Assessment of social, natural and man-made risks level makes it possible to develop economic mechanisms for regulating safety, including insurance of potentially dangerous objects and the population living in areas of possible damage in an emergency, which leads to a decrease in the volume of compensation payments from the local government's budgets.

3 Reducing emergencies a risk ensures more stable functioning of the economic potential and increases the competitive (investment) advantages of the region.

The level of territories safety is determined by a number of indicators: population size, area, number of hazardous objects and factors, statistics of emergencies and incidents, the number of people killed and injured in accidents and disasters of natural and man-made nature, the composition of special services and special equipment for eliminating the consequences of accidents and disasters.

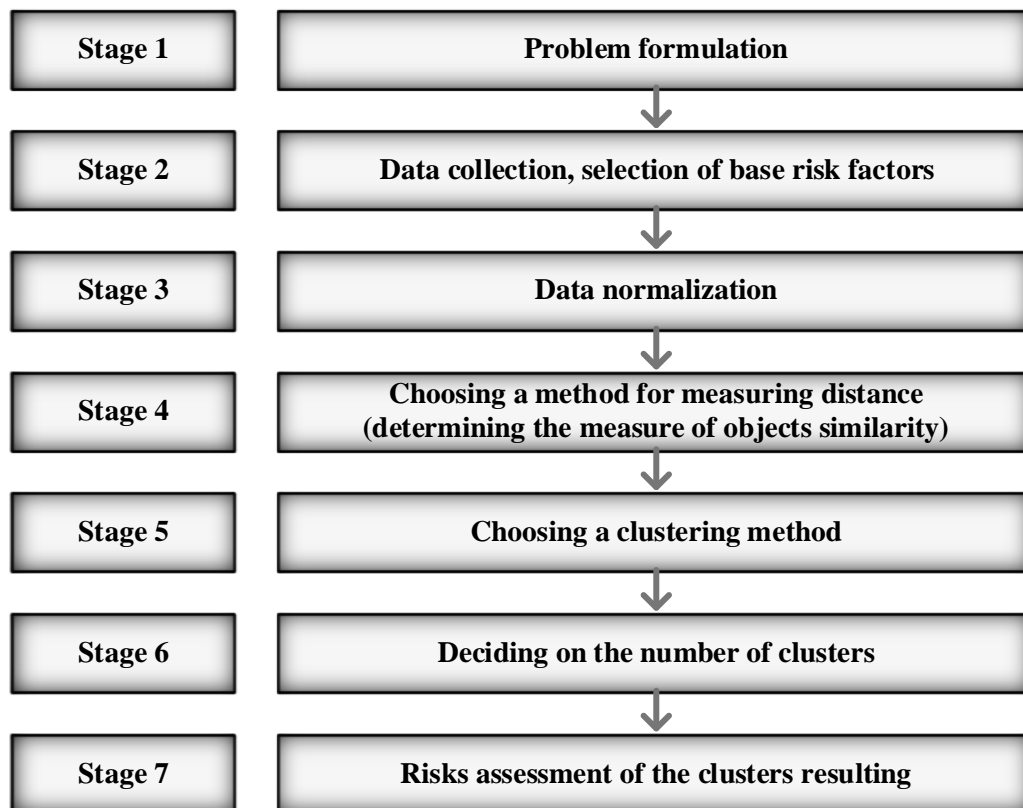
Any hazardous event can be realized according to different scenarios, but as a rule it leads to one or a combination of consequences, presented in Figure 1.



**Figure 1:** Consequences from a dangerous event

### 3. The method of multivariate statistics in solving problems of the risk assessment

To analyze the safety of industrial agglomerations, it is proposed to use the method of hierarchical cluster analysis, which allows dividing territories into groups of clusters with similar characteristics [17]. This method is widely used in various fields of science [18, 19]. Figure 2 shows the algorithm of the method for analyzing hazardous technogenic events in the territory under consideration.



**Figure 2:** Algorithm of the method for analysis of hazardous events

At the first stage, it is necessary to formulate a task for which it is necessary to classify territories into groups (for example, when analyzing technogenic safety or when assessing the risks of forest, aquatic ecosystems, etc.). At the second stage, quantitative indicators are selected, on the basis of which the analysis is carried out. Statistical data for the analysis should be normalized (depending on the nature of the source data, different methods of normalization can be used), this is necessary to reduce the scatter between data that are presented in different temporal and spatial scales.

The next step is related to determining the distance between objects. The most common ways to determine the distance between two points formed by the x and y coordinate axes are: Euclidean distance, the square of the Euclidean distance; distance of city blocks (Manhattan), Chebyshev distance.

Euclidean distance is the most general type, its mathematical meaning is the smallest distance (straight line) between points x and y, which is calculated by the formula:

$$dist = \left( \sum_{i=1}^m (x_i - y_i)^2 \right)^{1/2}, \quad (1)$$

The Manhattan distance is defined as the sum of the absolute differences of the pairs of values. In most cases, this method leads to the same results as the Euclidean distance. However, the influence of individual large differences (outliers) is reduced due to the lack of squaring. Looks like this:

$$dist = \sum_{i=1}^m |x_i - y_i|. \quad (2)$$

The Chebyshev distance is used when it is necessary to define objects as "different" and is calculated by the formula:

$$dist = \sum_{i=1}^m |x_i - y_i|. \quad (3)$$

For the accuracy of dividing into clusters, different distances should be used; with an adequate distribution, hierarchical trees will have a similar appearance.

At the 5th stage, it is necessary to choose a clustering method (a method for calculating the distances between clusters). The main methods are:

- Between-groups linkage - average of all distances between all possible pairs of points;
- Within-groups linkage - the distance calculated on the basis of all possible pairs of observations belonging to both clusters, taking into account the pairs of observations formed within the clusters;
- Nearest neighbor - the distance between a pair of observations located closest to each other, and each observation is taken from its own cluster;
- Furthest neighbor is defined as the distance between the farthest values of observations, with each observation taken from its own cluster;
- Centroid clustering the distance between two clusters is defined as the distance between two averaged observations;
- Median clustering - determined as in the previous case, but the center of the combined cluster is calculated as the average of all objects;
- Ward's method - the average values of individual variables for all available observations are determined, then the squared Euclidean distances from individual observations are calculated. The distances are summed up. One group of clusters combines those clusters that give the smallest increase in the total amount of distances.

For objects with a "blurred" structure with indistinctly pronounced "condensations", the Ward's method is best suited. As a result of this method, small and very compact clusters are formed. This method differs from the others in that it uses analysis of variance methods to estimate the distances between clusters [17].

The next step is related to determining the number of clusters. There are various approaches to determine the number of clusters in a hierarchical tree, but there is currently no universal method [20]. In this work, the k-means method was used to determine the number of clusters. It allows you to set the number of clusters and sequentially, starting with the second, to check the adequacy of the division of the hierarchical tree [17].

The last stage is associated with a quantitative assessment of the investigated technogenic risks.

Dividing the territory into homogeneous groups of clusters with similar characteristics allows you to determine your own risk levels, as well as to identify a reference group, the risk indicators in which can be used to determine acceptable levels.

## 4. Acknowledgements

This work was carried out with the financial support of the Krasnoyarsk regional fund of Science and Technology support within the framework of the project No. 2020061506473.

## 5. References

- [1] Makhutov N.A. Comprehensive analysis of safety and risks of promising life support systems and life activity // In the collection: Safety and monitoring of natural and technogenic systems. materials and reports. 2020. Pp. 7-11.
- [2] Akimov, V.A., Novikov V.D., Radaev N.N. Natural and technogenic emergency situations: dangers, threats, risks. M.: Business Express, 2001. 344 p.

- [3] Akimov, V. A., Lesnykh, V. V. and Radaev, N. N. Risks in nature, technosphere, society and economy. M.: Business Express, 2004. 352 p.
- [4] Lesnykh V.V. Risk analysis and mechanisms of compensation for damage from accidents at energy facilities. Novosibirsk "Science", 1999, 251 p.
- [5] Makhutov N.A., Gadenin M.M., Yudina O.N. Scientific risks analysis in the life support of a person, society and the state // Problems of risk analysis. Vol. 16.№ 2. 2019.Pp. 70-86.
- [6] Moskvichev V.V., Shokin Yu.I. Anthropogenic and natural risks in Siberia // Herald of the Russian Academy of Sciences. № 2. 2012. Pp. 131-140.
- [7] Shokin Yu.I., Moskvichev V.V., Nicheporchuk V.V. Methodology for assessing anthropogenic risks of territories and constructing risk cartograms using geographic information systems // Computational technologies. 2010. № 1. Pp. 120-131.
- [8] Safety of Russia. Legal, socio-economic and scientific and technical aspects. Safety and sustainable development of large cities. - M.: MGF "Knowledge", 1998. - 492 p.
- [9] Safety of Russia. Legal, socio-economic and scientific and technical aspects. Analysis of risk and security problems. Part IV. Scientific and methodological base of risk and safety analysis - M.: MGF "Knowledge", 2007. - 857 p.
- [10] Risk Management Standard AS/NZS ISO 31000:2009. 24 p.
- [11] Ales Bernatik, Pavel Senovsky, Michail Senovsky, David Rehak Territorial Risk Analysis and Mapping // Chemical Engineering Transactions. 2013. Vol 31(1). P. 69-74
- [12] Maestria en dirección y gestión pública local curso de experto en dirección y gestión pública local. Modulo 3. Prof. Carlos M. Rodrigues Otero. Documentacion De Apoyo. 2009. 70 p
- [13] Yafei Zhou\* and Mao Liu Risk Assessment of Major Hazards and its Application in Urban Planning: A Case Study // Risk Analysis, Vol. 32, No. 3, 2012
- [14] National Estimates Methodology for Building Fires and Losses. U.S. Fire Administration // National Fire Data Center. 2012. 18 p.
- [15] Katarina Holla Complex model for risk assessment of industrial processes // Journal of integrated disaster risk management. 2014. 4(2). P. 93-102
- [16] Moskvichev V.V., Bychkov I.V., Potapov V.P., Taseiko O.V., Тасейко О.В., Shokin Yu.I. Information system of territorial management of development risks and security // Herald of the Russian Academy of Sciences. Vol. 87, № 8. 2017. Pp. 696-705.
- [17] Taseiko, O., Ivanova, U., Rihter, E., Pitt, A. Using multivariate statistics to solve risk assessment problems for forest ecosystems // International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM, 2020, Pp. 777–784
- [18] Tromelin A., Chabanet C., Audouze K., Koensgen F. Multivariate statistical analysis of a large odorants database aimed at revealing similarities and links between odorants and odors // Flavour Fragr J. 2017. Pp. 1-21.
- [19] Mingqiu Shan, Sam Frong Yau Li, Sheng Yu, Yan Qian, Shuchen Guo, Li Zhang and Anwei Ding Chemical fingerprint and quantitative analysis for the quality evaluation of platyclade cacumen by ultra-performance liquid chromatography coupled with hierarchical cluster analysis // Journal of Chromatographic Science. 2018. Vol. 56. No 1. Pp. 41-48.
- [20] Yatskiv I., Gusarova L. Methods for determining the number of clusters by classifying without training. Transport and Telecommunication. Vol. 4, № 1, 2003. Pp. 23-28.