Cluster analysis and individual anthropogenic risk

Vladimir V. Moskvichev^{1,2}, Ulyana S. Postnikova^{1,2} and Olga V. Taseiko^{1,3}

¹Krasnoyarsk Branch of the Federal Research Center for Information and Computational Technologies, Krasnoyarsk, Russia

²Siberian Federal University, Krasnoyarsk, Russia

³Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russia

Abstract

Models and assessment methods of anthropogenic risk are analyzed at this article, general basis of mathematical approach for risk analysis is disclosed. Based on multivariate statistic methods, algorithm of analysis for Siberian territories safety is formulated, it allows to define acceptable level of risk for each territorial group (cities with population density more than 70 000, towns with population less than 70 000, and municipals areas).

Keywords

Territorial risk, hierarchical clustered analysis, k-means method, acceptable level of risk.

1. Introduction

Technological exploration of territories and industrial development have negative influence on ecological and social safety and form some problems, which can reflect on country development. Key problems, which require careful consideration from government bodies and public authorities:

- high concentration of potential risk on limited territories (spent nuclear fuel, uranium waste, nuclear and chemical weapon, defense manufacture pipelines, gas-holder, waterpower plants, chemical industry, aviation etc.);
- increased risk of accidents due to high degree of equipment wear;
- human factor, connected to low safety culture.

For effective realization of Presidential decree "Strategy on the Russian Federation national safety" 02.07.2021, a range of measures is required, which are referred to population and territory protection from natural to anthropogenic accidents.

Risk analysis forms main mechanism in safety control, which involves playing into kill effect. Key reasons of risk formation are human and his living, natural and industrial processes. At present different methods of prediction of the risk and risk assessment are developed, which associated with natural and man-made emergencies. Risk assessment models and methods can be divided into two groups:

SDM-2021: All-Russian conference, August 24-27, 2021, Novosibirsk, Russia

[☆] krasn@ict.nsc.ru (V. V. Moskvichev); ulyana-ivanova@inbox.ru (U. S. Postnikova); taseiko@gmail.com (O. V. Taseiko)

^{© 02021} Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

- for the analysis of the safety of industrial facilities;
- for the analysis of territorial entity.

Minimization of technogenic risks which allows to reduce cost on remediation is necessary for sustainable economic development.

2. Multivariate statistic method in problems solving of acceptable level risk

These days acceptable level risk is considered $1 \cdot 10^{-5}$, which can be interpret as one death from 100 000 people [1], another method proposes to measure acceptable risk as aggregate figure in some years [2]. However, official methods of risk assessment don't take into account the fact that the number of populations in territorial entity differs from 5 000 to one million people and more. Such changes in population base influence the final risk level and lead to uncertainty. For example, while having the equal number of accidents and survivors in territories with different population density the risk also will be different.

At this work the method of hierarchical clustered analysis is proposed to use, which allows to divide the Siberian Federal District's (SFD) territory into clustered groups, to choose a reference group, in which comparisons with determination of acceptable level of risk are held [3]. The given method is widely used in various fields of science [4, 5].

Figure 1 shows the algorithm of the method for analyzing hazardous anthropogenic events in the territory under consideration.



Figure 1: Algorithm of the method for analysis of hazardous events.

At the 1st stage, the task to analyze anthropogenic safety of SFD is set. At the 2nd stage, quantitative indicator is selected on the basis of which the analysis is carried out. Instead of data normalization, value of territory vulnerability from different kinds of accidents is used, where data of automated information and control system for the prevention and liquidation of emergency situations period 1999–2017 years is taken. Territory vulnerability is a complex indicator, which has a probable origin. It includes probability of hazardous event formation, probability of emergency and death in different anthropogenic events:

$$\vartheta = \{p_a; p_f; p_e\},\$$

where ϑ – territory vulnerability; p_a – probability of hazardous event formation; p_f – probability of death during hazardous event; p_e – probability of emergency.

Next stage is connected with determination the distance between objects. More common methods to determine the distance between two points, formed by coordinate axes x and y are Euclidean distance, Chebyshev, city-block distance (Manhattan). To be accurate in cluster dividing different distances should be used, at adequate distribution hierarchical tree will have the similar form.

At the 5thstage, cluster method should be chosen (calculation way od distance between clusters). Ward's method is the best suited for objects having "blurred" structure with vague concentration. As a result, small in size and very compact clusters are formed. This method differs from the others in that it uses analysis of variance methods to estimate the distance between clusters [3].

Next stage is connected with finding out the number of clusters. There are various methods to calculate the number of clusters in a hierarchical tree. However, nowadays the universal method doesn't exist [6]. At the work k-means method is used to determine the number of clusters. It allows to set the number of clusters and gradually, beginning with number two, to cheek the adequacy of the division of the hierarchical tree [3].

The 7th stage is related to a quantitative assessment of the investigated individual anthropogenic risk. The acceptable level of risk is taken in the last stage.

3. Assessment of individual anthropogenic risk for Siberian Federal District's territories

Analysis of SFD territories is necessary to be done separately in different kinds of administrativeterritorial units cities (with population more than 70 000 people), towns (population less than 70 000 people) and territorial entities. Anthropogenic load is different in the cities and districts.

In SFD with population more than 70 000 people 31 cities are located, three of them are cities with one million people and more. Omsk, Novosibirsk and Krasnoyarsk are the most vulnerable cities, because of significant population density, developed industry and infrastructure. 46 towns are situated in Siberian territory with population less than 70 000 and there are 268 different municipal districts. For these territories figures of vulnerability have been got as well. The most event frequency is defined in Irkutsk, Novosibirsk and Emelyanova districts.

On basis of this method, with the use of software STATISTICA the sources and factors of technogenic accident are reviewed with the example of SFD cities to assess the anthropogenic



Figure 2: Tree diagrams of clustering Siberian cities by Ward's method using Euclidean (a), Manhattan (b) and Chebyshev (c) distances.

hazardous level. 31 cities (population more than 70 000 people) are chosen to analyze. Figure 2 shows the tree diagrams which are regimented by using different distances (Euclidean, Chebyshev, Manhattan). The tree diagram has been received with one-type distribution. Thus, we can make a conclusion about the adequate distribution of clusters. To calculate the number of clusters k-means method is used, which allows to distinguish five equable groups.

For each cluster different kinds of hazardous anthropogenic events are analyzed (for big cities 17 various kinds of accidents are followed), there are appertained mean values in Table 1. In every cluster group there are its dominating negative factors. The biggest number of anthropogenic accidents is seen in V cluster, where high population density and developed industry are.

Therefore, the values of individual risk and interval values for every city have been received. The 3rdgroup has the minimum (the least value). This cluster is considered to be the reference and risk interval value is an acceptable level. Thus, there are three zones to analyze risk for big cities SFD (Figure 3).

- acceptable level $R \leq 1.07 \cdot 10^{-5}$;
- higher level $R \in (1.07 \cdot 10^{-5}; 1.85 \cdot 10^{-5}];$
- high-lying level $R > 2.85 \cdot 10^{-5}$.

Table 1

Mean values of hazardous events in clusters groups.

No.	Factors	Clusters groups				
		I	11	III	IV	V
1	Accidents in life support system	7	10	2	39	77
2	Accidents in airline service	2	3	0	14	28
3	Highly toxic chemicals accidents	2	2	0	8	11
4	Explosion in industrial objects	1	2	0	2	4
5	Household explosion	1	3	0	5	8
6	Major traffic accident	3	5	1	7	41
7	Crane falling	0	0	0	2	5
8	Mass presence of people fires	5	18	1	28	82
9	Fires in industrial objects	5	7	1	23	45
10	Household fires	37	50	17	145	264
11	Chemical substances findings and radioactive materials	1	1	0	7	12
12	Construction collapsing	2	4	1	6	12
13	Main pipelines accidents	0	0	0	0	2
14	Industrial object accidents	2	7	3	1	2
15	Railway accidents	0	1	0	1	2
16	Water transport accidents	1	0	0	2	4
17	Radioactive substances emission	0	0	0	1	5



Figure 3: The curve of changes in individual risk in the SFD cities with a population of more than 70 000 people.

In acceptable risk level zone there are 11 cities, which are in III cluster. There are 13 cities (I and II cluster) in a higher level, and there are 6 cities in a high-lying level from VI and V clusters and one city from I and II clusters (Mezhdurechinsk and Biisk).

By a similar way individual anthropogenic risk values have been got for towns with population less 70 000 people:

- acceptable level $R \leq 3.48 \cdot 10^{-6}$;
- higher level $R \in (3.48 \cdot 10^{-6}; 4.3 \cdot 10^{-6}];$
- high-lying level $R > 4.3 \cdot 10^{-6}$.

In municipal districts this method allows you to divide 268 ones into five uniform groups, analyze each group individually and find out a reference cluster and risk level. Risk values for municipal districts change from 0 to $6.3 \cdot 10^{-6}$. The following risk level values are done:

- acceptable level $R \leq 3.85 \cdot 10^{-6}$;
- higher level $R \in (3.85 \cdot 10^{-6}; 6.3 \cdot 10^{-6}];$
- high-lying level $R > 6.3 \cdot 10^{-6}$.

After analyzing the technogenic safety of Siberian it was determined that there are 74 territorial units in a high risk zone, where it is necessary to hold event to privent and minimize risk.

4. Conclusion

The sustainable development depends on the analysis, assessment and minimization of anthropogenic risk. Assessment and analysis of territorial technogenic risk are the important tools for improving the regional policy, strategy and tactics which allow to minimize the consequences on the targeted territory. With man-made burden moving higher, which threatens restoration of natural recourses, with rising risks for life and heath of population, high-quality management based on using the assessment approaches of anthropogenic risk necessary.

Acknowledgments

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.

References

- [1] MR 2-4-71-40 methodical recommendations "The Order of development, verification, evaluation and correction of electronic passports of Territories (objects): UTV". Ministry of the Russian Federation on Civil Defense, emergency situations and liquidation of consequences of natural disasters from 15.07.2016. Available at: http://docs.cntd.ru/document/456080084.
- [2] Standard 22.10.02-2016 Safety in emergency situations. Emergency risk management. Permissible risk of emergency situations. Approved by the order of Rosstandart No. 724art. of 29.06.2016.
- [3] 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-August (3.1). P. 777–784.
- [4] 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. P. 1–21.

- [5] Shan M., Li S.F.Y., Yu S., Qian Y., Guo S., Zhang L., Ding A. 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. P. 41–48.
- [6] Yatskiv I., Gusarova L. Methods for determining the number of clusters by classifying without training // Transport and Telecommunication 2003. Vol. 4. No. 1. P. 23–28.