Applying cluster analysis to ranking the vulnerabilities of railway influence territory

Viktor Katin¹, Vladimir Kosygin²*, Midkhat Akhtiamov¹ and Andrey Lutsenko¹

¹Far Eastern State Transport University, Khabarovsk, Russia
²Computer Center of Far-Eastern Branch of Russian Academy of Sciences, Khabarovsk, Russia

Abstract. The article considers the proprietary method of ranking the territory of the railway influence, based on the use of vulnerability matrices and application of cluster analysis, which allows optimizing the decision-making process for responding to emergencies associated with rail transport. The "point" ranking of vulnerabilities in the zone of railway influence was used for more accurate information on the events with applying the identity matrices of vulnerabilities. The partition cell is taken as a single "point". The method is based on the ranking of ecological, economic, environmental, bioresource and cultural significance of the territory by means of defining identical matrices of vulnerabilities that describe each point of the event. The partition cells were grouped into three clusters: moderately vulnerable territory, highly vulnerable territory and extremely vulnerable territory. The result of using the method was the compiling of vulnerability ranking maps.

1 Introduction

The issues of ensuring security in railway transport are among the top-priority issues in the development of Russian transport sector. "Strategy for the Development of Russian Railways up to 2030" [1] implies the elaboration of measures aimed at improving the security system on railway transport. Particularly, this document defines the aim to modernize and maintain databases for assessing the vulnerability of categorized objects. The uniqueness of the specific territory passed by the railway and the risk of irreversible consequences require special attention to the problem of ensuring the transportation safety, including issues of environmental protection.

Since the statistic data indicate a high level and a tendency to increase the number of emergencies associated with the transport of dangerous goods by rail, it is important to develop and apply new methods and tools to minimize the corresponding risks. In the conditions of limited material resources, measures to prevent emergencies and to organize an effective set of response measures to eliminate its consequences are particularly important.

One of the directions of solving this problem is the ranking of the railway influence territory by its perceptivity to the effect of hazards, or in other words by the vulnerability of

* Corresponding author: kosyginv@inbox.ru

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).
the territory under consideration. The issues of zoning the railway in terms of emergency risk are considered in the works [2, 3, 4].

In this context, the authors of the present article developed a method for ranking the railway influence territory using vulnerability matrices and applying the cluster analysis. This method enables optimizing the decision-making taking into account the rational disposing of emergency response facilities associated with rail transportation.

2 Vulnerability ranking method

The works [3, 4] are dedicated to the issues of ranking the railway by hazard categories. They propose categorizing the railway line according to the hazard ranks and using a comprehensive risk map that summarizes natural hazards without differentiation by the types of vulnerabilities and of linking to the event points.

As opposite to the methods mentioned above, for a more accurate specification of the event site, a "point" ranking of the vulnerabilities of the railroad influence zone is proposed using identity matrices of vulnerabilities. The complex of elementary cells obtained by dividing the zone of railway influence by a grid are understood by identity matrices of vulnerabilities. [5]. Each cell is assigned with a weight characteristic, which depends on the location of the vulnerability object based on available information resources.

The considered method is based on the ranking of the territory according to ecological, economic, environmental, bioresource and cultural value by defining identity matrices of vulnerabilities that describe each point of the event [5].

The ranking of the territory, which is necessary to optimize the response in the event of an accident or emergency with dangerous goods carried by the railway, is suggested by the authors on the basis of a cluster analysis of identity matrices of vulnerabilities.

Cluster analysis is used to divide the set of considered objects or characteristics into groups or clusters, combined according to certain criteria [6]. The objects in each group have some common features: proximity by a certain extent (distance) [6]. In our case, the main advantage of applying cluster analysis to the ranking of vulnerabilities is the ability to divide the objects into groups (clusters) by a set of their various parameters.

The split of the set of investigated objects that satisfy a certain criterion of optimality – the target function – is the solution of the cluster analysis problem. As a target function, the intragroup sum of the squared deviations of the object parameter values from its average value is usually assumed.

The questions of cluster analysis are comprehensively described in monographs [6,7,8].

As a starting point for further cluster analysis, identity matrices of vulnerabilities are used.

To implement the procedure for the typology of the cells under analysis, it is necessary to define a set of parameters that involves selecting indicators, selecting weight coefficients, selecting a method for processing initial data, and selecting the determining sequence of visualization.

When selecting the indicators, the weight characteristics of the vulnerability in the form of the coefficients of potential damage due to emergencies are displayed. The susceptibility of the territory is taken into account as well as and the ability to respond to the incident with damage (loss of initial quality) caused by air pollution; pollution of a water body; pollution, degradation, clogging or cluttering of soil; damage to the fauna and flora; damage from long-term consequences of an emergency situation, damage to population and infrastructure.

Identity matrices of vulnerabilities, compiled on the basis of territorial indicators, according to the methodology requirements [9], are a factual foundation for constructing the respective maps.
When clustering the set of unit cells of an identity matrice of vulnerabilities, Hierarchical Clustering Explorer 3.5 (HCE) was applied, which uses a hierarchical clustering algorithm that is a procedure for creating a sequence of inset partitions from the perspective of the proximity matrix data [10]. The hierarchical algorithm allows defining a natural grouping with interactive feedback visualization (dendogram and color mosaic) and dynamic request management [5]. As a result, the analyzed weight characteristics of single cells are displayed in color mosaic, uniting into clusters (Fig. 1).

![Vulnerability clustering map](image)

**Fig. 1.** Vulnerability clustering map.

The arrangement of lines and columns on the display and the color of the mosaic are changing according to the result of the clustering. A graphic drawing on the basis of the data shows the color of each tile on the basis of a numerical value corresponding to the tile. The coordinates of the cell correspond to the coordinates of the cell of the original matrices of the vulnerabilities, tied to geographic coordinates. The parameters for displaying the color and, hence, the range of the cluster, are set by the controller using a histogram of numerical values.

A high value and in our case the highest vulnerability of the territory has a bright red color by default, when a low value and accordingly a small vulnerability has a bright green color. The average value is black.

As a result of the cluster analysis, the detection of clusters and their interpretation into ranks become possible. At the final stage, based on the results obtained, clusters are interpreted, including the estimation of the size and composition of clusters and the analysis of their content, which is necessary for determination of the response measures [11]. The algorithm for ranking vulnerabilities using the cluster analysis method is shown in figure 2.
Following the results of clustering based on the numerical and color characteristics, the ranking of vulnerabilities of the railway influence territory (moderately vulnerable, highly vulnerable, extremely vulnerable) is executed, followed by the elaboration of an appropriate response procedure [5].

Based on the visualization of vulnerability clustering, a vulnerability map of the railway influence zone is drawn up, which allows planning measures for preventing and eliminating emergencies, optimizing the function of “response” to the incident.

3 The results of using the vulnerability ranking method

Let us consider the issue of ranking the vulnerabilities of the railway influence territory on the example of the Khabarovsk-Korfovsky section.

The site is characterized by the settlements of various categories, essential structures, crossing with water bodies, interlinking with specially protected natural areas.

At the first stage, the territory of railway influence was determined. For this purpose, the territory adjacent to the railway line was divided on both sides into unit cells oriented to the cardinal points and bound to the reference network.

To implement the procedure for the typology of the analyzed cell parameters, a set of indicators were determined based on the provisions of the methodology [9].

For the selected site, the following vulnerability indicators were considered: soil, hydrological, population, object and biocenosis.

Based on available information sources, each cell was assigned with weight characteristics depending on the particular location of its vulnerability object. At the next stage, the numerical weight coefficients were distributed over identity matrices.

Let us consider the soil identity matrix as an example. The coefficients of this matrix were taken in accordance with the recommended regulatory documents:

- for lands of specially protected areas, the coefficient is 1;
- for moss-lichen deer and meadow-herbage mountain pastures in the composition of lands of all categories - 0.9;
- for water protection zones in the composition of lands of all categories - 0.9;
- for agricultural land in the composition of agricultural land - 0.7;
- for forested areas in the land of all categories - 0.5;
- for lands of settlements (except land plots referred to the territorial zones of industrial, special purpose, engineering or transport infrastructures, military facilities) - 0.3;
- for other categories and types of special purposes - 0.1.

The identity matrix of vulnerabilities compiled for the soil indicator is presented on figure 3.

|   | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
| A | 0 | E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| B | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| D | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| E | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| H | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| I | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| J | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| L | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| M | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| N | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| O | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| P | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Q | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| R | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| S | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| T | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| U | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| V | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| W | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| X | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Y | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Z | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Fig. 3. Identity matrix of soil vulnerability.
The matrices for the remaining four vulnerability indexes were compiled the same way.

In the next stage, comparing the five identity matrices obtained for each vulnerability index to each other, a complex vulnerability matrix was constructed, where each unit cell was characterized by five vulnerability indicators. The resulting complex vulnerability matrix was used in the subsequent cluster analysis, which became the basis for a comprehensive vulnerability ranking map was obtained as a result of visualization (Fig. 4).

![Fig. 4. Map of the ranking of vulnerabilities in the zone of railway influence in the Khabarovsk-Korovsky section.](image)

Based on the cluster visualization, the vulnerability categorization was performed: green cells indicate moderately vulnerable territory of the railway influence, black cells – highly vulnerable territory, red cells – extremely vulnerable territory.

According to the analysis of the obtained categorized map of vulnerability of the railway influence territory, the most dangerous areas were identified for the purpose of determination of the procedure for dealing with the consequences of the emergency.
In our case, these are the places of connection of the railway with water bodies, essential structures, areas of increased population density and specially protected natural areas.

4 Conclusion

Providing security for dangerous transportation of goods and organizing prompt response to incidents and emergencies require the use of modern means of implementation.

Electronic map of ranked vulnerabilities (vulnerable, particularly vulnerable, extremely vulnerable) is one of the solutions of this problem. Using cluster analysis, it is possible to solve the problem of classifying and categorizing the initial set of data on territorial vulnerabilities. Information on the characteristics of the matrix cells is required for that purpose, which enables judging their similarity or difference on the basis of cartographic and statistical data.

The analysis results are the maps of ranking the vulnerabilities of railway influence territory allowing to organize the "response" function effectively in case of incidents and emergencies related to the transportation of dangerous goods.

The use of a vulnerability ranking map and the elaboration of appropriate regulations for minimizing the consequences of emergencies according to ranks of the territory allows instantly identifying the location of the incident, preventing the developing of unfavorable scenario of the event, identifying possible damage more accurately and mobilizing the necessary emergency response forces and resources.

References

3. I.O. Teslenko, Sovershenstvovaniye usloviy perevozki opasnykh gruzov na zheleznodorozhnom transporte: dis. kand. tekh. nauk (Novosibirsk, 2001)
5. V.D. Katin, A.N. Lutsenko, Yevraziyskiy soyuz uchenykh (YESU) 4, 109-111 (2014) ISSN 2575-7999
7. B. Dyuran, P. Odell, Klasternyy analiz (Statistika, Moscow, 1977)
8. I.D. Mandel’, Klasternyy analiz (Finansy i statistika, Moscow, 1988)
10. P. Anand, Profitability of Reduction of Ecological Risk from the Car of the Tank of the Railway Transportation of Dangerous Materials, the thesis for a doctor's degree (University of Illinois in Urbana-Plain, 2006)