

Multidimensional Data Analysis for Evaluating the Natural and Anthropogenic Safety (in the Case of Krasnoyarsk Territory)

Tatiana Penkova

Institute of Computational Modelling of the Siberian Branch
of the Russian Academy of Sciences,
Siberian Federal University,
Krasnoyarsk, Russia
penkova_t@icm.krasn.ru

Abstract.

This paper presents an approach to evaluating the natural and technogenic safety of the one of the largest regions in Siberia through the comprehensive analysis of territorial indicators. In order to explore geographical variations and patterns in occurrence of emergencies the multidimensional data analysis technique is applied to data of the Territory Safety Passports. For data modeling, principal components are selected and interpreted taking account of the contribution of the data attributes to the principal components. Data distribution on the principal components is analyzed at different levels of the territory detail: municipal areas and settlements. The results of this analysis have allowed to identify the high-risk areas and rank the territories according to danger degree of occurrence of the natural and technogenic emergencies. It gives the basis for decision making and makes it possible for authorities to allocate the forces and means for territory protection more efficiently and develop a system of measures to prevent and mitigate the consequences of emergencies in the large region.

Keywords. Multidimensional data analysis, Principal component analysis, Evaluating the natural and anthropogenic safety, Prevention of emergencies, Territorial management.

1 Introduction

Prevention of natural and technogenic emergencies is a one of the major tasks of the territory management. Analytical support of decision-making processes based on modern technologies and efficient methods of data analysis is a necessary condition for improving the territorial safety system and management quality.

The Krasnoyarsk territory is the second largest federal subject of Russia and the third largest subnational governing body by area in the world. The Krasnoyarsk region lies in the middle of Siberia and occupies an area of 2,339,700 square kilometers, which is 13% of the country's total territory. This territory is characterised

by heightened level of natural and technogenic emergencies which is determined by social-economic aspects, large resource potential, geographical location and climatic conditions. In the territory there are many accident prone technosphere objects including radiation-related objects, chemically-dangerous objects, fire-hazardous and dangerously explosive objects; hydraulic facilities; critically important objects; a lot of survival objects including boiler plants, power plants, pipelines and networks. Moreover, the territory is located in seven climatic zones. A number of large-scale natural emergencies, such as flood, forest fire, gale-strength wind and anomalously low temperature are recorded each year [1]. In order to improve the population and territory safety, a lot of monitoring systems and control tools for on-line observation are being actively introduced within the region [2-4]. The Ministry of Emergency has enacted the structure and order of conducting the Territory Safety Passport, which defines a system of indicators to assess the state of territory safety, the risk of emergencies and possible damages to create efficient prevention and mitigation actions [5]. At present, there are massive data collections about the state of controlled objects, occurred events and sources of emergencies. However, we have to admit that the processing stored data, aimed at obtaining the new and useful knowledge, is insufficient. The local databases remain unused, while the reasonable decisions, comprehensive analysis and emergencies prediction are sorely needed. Thus, identification of risk factors of emergencies based on monitoring data and investigation of their impact on key indicators of human safety are topical and important tasks in territorial management.

Data mining techniques provide the effective tool for discovering previously unknown, nontrivial, practically useful and interpreted knowledge needed to make decisions [6]. This paper presents the results of comprehensive multidimensional analysis of natural and technogenic safety indicators of the Krasnoyarsk territory in order to explore geographical variations and patterns in occurrence of emergencies by applying the data mining technique – principal component analysis – to data of the Territory Safety Passports.

The outline of this paper is as follows: Section 1 contains introduction. Section 2 describes the initial data. Section 3 presents results of principal component analysis: identification and interpretation of principal components; analysis of data distribution on the principal components at different levels of the territory detail. Section 4 draws the conclusion.

2 Data Description

Evaluating the natural and technogenic safety indicators is based on data of the Territory Safety Passports of the Krasnoyarsk territory collected in Center of Emergency Monitoring and Prediction (CEMP). Original dataset contains 1,690 objects, essentially discrete settlements-level geographical entities of the Krasnoyarsk territory, each with 12 measured attributes. Data attributes are listed in Table 1. One part of attributes characterizes the sensitivity of the territory to the risk factors effects (e.g. population density, the presence of industrial and engineering facilities) that is

determined by the number of objects located on the territory (i.e. number of potential sources of emergencies), it is so-called "object attributes". The other part of attributes characterizes the presence of potential factor that can damage the health of people, can cause irreversible damage to the environment that is determined by the statistic of events occurred in the territory (i.e. number of emergencies), it is so-called "event attributes". In addition, some reference characteristics are used for data interpretation and map visualization. The preliminary correlation analysis of original data has shown a fairly strong relationship between "object" and "event" attributes, therefore for further analysis we will consider the attributes that characterize population and events. The correlation coefficients are presented in Table 2.

Table 1. List of the data attributes of Territory Safety Passports

No	Attributes	Description
1	Pop	Population
2	Soc_object	Number of important social facilities (e.g. educational, health, social, cultural and sports facilities)
3	Water_object	Number of dangerous water bodies
4	Indust_object	Number of potentially dangerous industrial objects (e.g. plants, factories, mines)
5	Oil_line	Number of pipeline sectors in 5 km. radius from borders of settlement
6	Munic_object	Number of municipal facilities (e.g. power supply, water supply and heating facilities)
7	Flood_event	Number of floods
8	NFire_event	Number of natural fires
9	TFire_event	Number of technogenic fires
10	Munic_event	Number of accidents at municipal facilities
11	Nat_event	Number of natural events (excluding natural fires and floods)
12	Tech_event	Number of technogenic events (excluding technogenic fires and accidents at municipal facilities)

Table 2. Correlation coefficients between data attributes

No	2	3	4	5	6	7	8	9	10	11	12
1	0.97	0.39	0.96	0.04	0.28	0.29	0.08	0.96	0.95	0.08	0.60
2		0.36	0.96	0.01	0.25	0.25	0.05	0.91	0.94	0.06	0.59
3			0.39	-0.01	0.32	0.60	0.12	0.39	0.36	0.17	0.30
4				0.01	0.24	0.29	0.05	0.91	0.91	0.07	0.56
5					0.08	-0.02	0.06	0.07	0.02	0.05	0.14
6						0.29	0.08	0.31	0.43	0.13	0.48
7							0.06	0.33	0.30	0.13	0.28
8								0.10	0.06	-0.02	0.05
9									0.93	0.11	0.63
10										0.08	0.58
11											0.13

Within this research, the analysis and visualisation of multidimensional data are conducted using the ViDaExpert [7]. Data visualization on geographical maps is performed by applying the mapping tools «ArcGIS» [8].

3 Principal Component Analysis

Principal Component Analysis (PCA) is one of the most common techniques used to describe patterns of variation within a multi-dimensional dataset, and is one of the simplest and robust ways of doing dimensionality reduction. PCA is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components [9]. The number of principal components is always less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance and each subsequent component, respectively, has the highest variance possible under the constraint that it is orthogonal to the preceding components.

3.1 Contribution of the Data Attributes to the Principal Components

One of the greatest challenges in providing a meaningful interpretation of multi-dimensional data using PCA is determining the number of principal components. In general, the method allows to identify k components based on k initial attributes. Table 3 shows the results of calculating the eigenvectors of the covariance matrix arranged in order of descending eigenvalues.

Table 3. Results of principal components calculation

Components	1	2	3	4	5	6	7
Eigenvalues	0.404	0.249	0.141	0.116	0.075	0.010	0.005
Accumulated dispersion	0.504	0.652	0.793	0.909	0.985	0.995	1
Pop	0.509	0.109	0.111	0.113	0.227	0.182	0.787
TFire_event	0.513	0.083	0.061	0.088	0.171	0.616	-0.557
NFire_event	0.060	0.439	-0.876	0.186	-0.022	-0.033	0.012
Munic_event	0.503	0.096	0.120	0.084	0.251	-0.764	-0.263
Flood_event	0.235	-0.314	-0.325	-0.853	0.109	-0.004	0.029
Nat_event	0.086	-0.822	-0.311	0.458	0.103	-0.015	0.010
Tech_event	0.397	-0.072	0.019	0.013	-0.913	-0.051	0.024

Based on combination of Kaiser's rule and the Broken-stick model [10], two principal components for data attributes were identified (PC1 and PC2) with 65% accumulated dispersion. Figure 1(a) illustrates the eigenvalues of components. As can be seen from Figure 1(a), Kaiser's rule determines two principal components – eigenvalues of first two components are significantly greater than the average value and the Broken-stick model gives also two principal components – the line of Broken-stick model also cuts the eigenvalues of first two components. The contribution of the data attributes to principal components is presented in Figure 1(b).

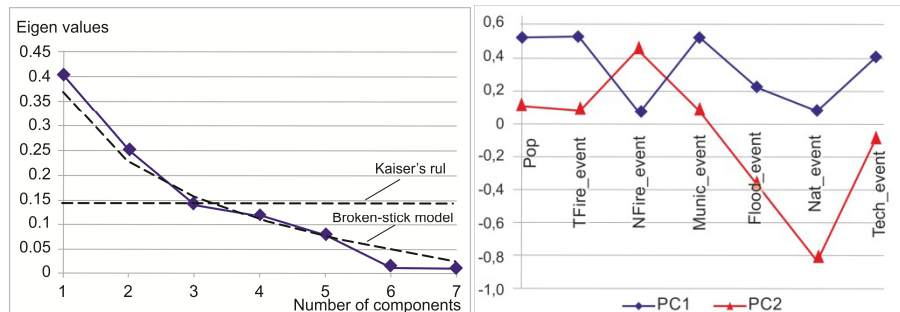


Fig. 1. (a) Eigenvalues of components. (b) Contribution of the data attributes to the first (PC1) and second (PC2) principal components

From Figure 1 (b) we can see that the first principal component (PC1) is characterised by the following attributes: a high level of population, high proportions of technogenic fires, accidents at municipal facilities and other technogenic events, a low percentage of natural events including natural fires and floods. In combination, these characteristics present the big settlements (e.g. cities) with high levels of technogenic hazards. The second principal component (PC2) is characterised by the following attributes: a low level of population, high proportion of natural fires, strong negative correlation with the percentage of natural events including floods and technogenic events including fires and accidents at municipal facilities. In combination, these characteristics present relatively small settlements (e.g. villages) with high levels of natural fires. This means that in comparison with other types of emergencies the technogenic and natural fires are the greatest threat for the Krasnoyarsk territory.

3.2 Data Distribution on the Principal Components

The data can be divided into groups according to where the settlements are located in terms of Territory Classifier. There are three levels of the territory detail: settlements, municipal areas and groups of municipal areas that give 1,690 objects, 65 objects and 8 objects respectively for the Krasnoyarsk territory. Figure 2 shows the visualisation of territorial groups (groups of municipal areas) on the geographic coordinates and the PCA plot, where: group 1 (green) – Angarsk Group; group 2 (rose) – Eastern Group; group 3 (purple) – Yeniseisk Group; group 4 (light blue) – Western Group; group 5 (yellow) – Central Group; group 6 (red) – Southern Group; group 7 (blue) – Taymyr Autonomous Okrug; group 8 (brown) – Evenk Autonomous Okrug. On a data map, the points in the form of triangles are settlements, and the color of these points corresponds to the color of the territorial group. Objects in the form of circles represent centroids of clusters of territorial groups.

As can be seen from Figure 2, along the first principal component (PC1) the territorial groups are concentrated quite densely, it means that technogenic fires are general characteristic for all territorial groups of region, but along the second principal component (PC2) the territorial groups are distributed significantly and we can see that the natural fires are indicative of northern territorial groups.

The lowest values of projections on the first principal component (Figure 3, blue points) are observed for such settlements as: Ust-Kamo, Shigashet, Kasovo, Verhnekemskoe, Komorowskiy, Angutiha, Lebed. It can be explained by the fact that these settlements are very small villages and, at present, in these settlements there are no any socially significant objects and residents. The complete absence of the economic activity in these settlements leads to the lowest level (or absence) of technogenic fires. The highest values of the projections on the first principal component (Figure 3, red points) are observed for such large settlements as Krasnoyarsk, Norilsk, Achinsk, Kansk, Minusinsk Lesosibirsk. These settlements present the big cities of the Krasnoyarsk territory where the population and number of socially significant and industrial facilities are above average level in region.

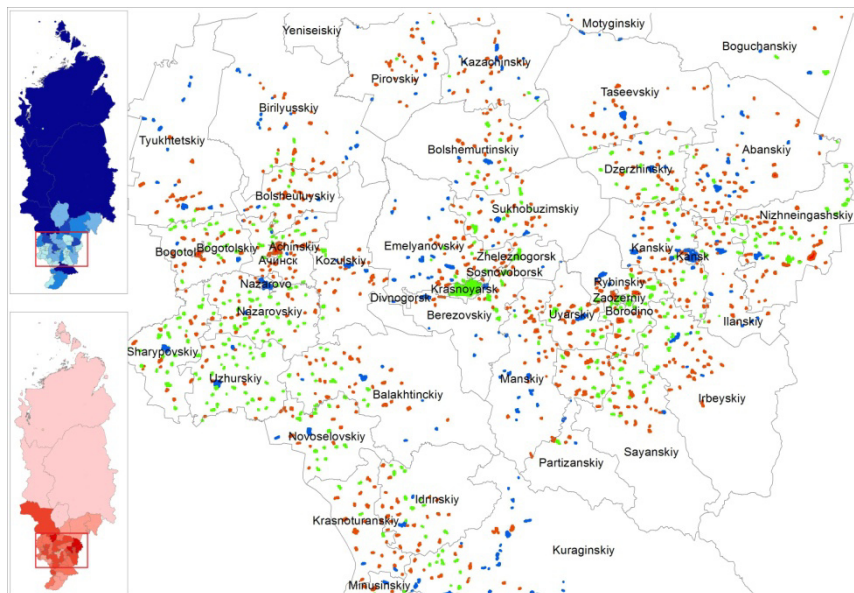


Fig. 4. Visualisation of the projections on the second principal component for municipal areas and settlements

The lowest values of projections for the second principal component (Figure 4, blue points) are observed for such settlements as: Turuhansk, Cheremshanka, Tanzybey, Emelyanovo, Ermakovskoe. Low levels of natural fires can be explained by the following facts: the absence of vegetation as a source of emergency in steppe areas (e.g. Western and Southern groups) and the absence of settlements in forest zone (e.g. Evenk Autonomous Okrug, Yeniseiysk and Turukhansky areas). The highest values of projections for the second principal component (Figure 4, red points) are observed for such settlements as: Startsevo, Tilichet, Kuray, Baikal, Glinniy. The high risk of natural fires is observed in the large settlements that are located close to the forest zones. In addition, there is probability of natural fires in the big cities where the forests constitute the part of their territories.

4 Conclusion

In this paper the evaluating of natural and technogenic safety of the Krasnoyarsk territory in the context of settlements is carried out first time by applying the multidimensional data analysis technique – principal component analysis – to data of the Territory Safety Passports. The data analysis results show that the technogenic and natural fires are the greatest threat for territory of the Krasnoyarsk region. The explored geographical variations and patterns allow to identify the high-risk municipal areas and particular settlements, rank the territories according to danger degree of occurrence of the natural and technogenic emergencies. The results of this research make it possible for specialists of CEMP to develop a system of measures to prevent and mitigate the consequences of emergencies in the Krasnoyarsk territory.

The techniques and tools used in this paper make it easy to change the initial dataset (e.g. territories or threats) for other tasks. The presented approach to comprehensive multidimensional analysis of the territories can be adopted for different control objects in various areas.

References

1. The State of Natural and Anthropogenic Emergencies Protection of Territory and Population in the Krasnoyarsk Region: Annual Report of Ministry of Emergency, Krasnoyarsk, 230 p. (2016) (in Russian)
2. Penkova T., Nicheporchuk V., Metus A.: Comprehensive operational control of the natural and anthropogenic territory safety based on analytical indicators, LNAI, Part I, Vol.10313, pp. 263-270 (2017)
3. Shaparev N.Y.: Environmental Monitoring of the Krasnoyarsk Region in Terms of Sustainable Environmental Management. Informational and Analytical Bulletin (Scientific and Technical Journal), Vol.18, No 12, pp. 110-113 (2009) (in Russian)
4. Bryukhanova E.A., Kobalinskiy M.V., Shishatskiy N.G., Sibgatulin V.G.: Improvement of Environmental Monitoring Information Maintenance as an Instrument for Sustainable Social and Economic Development (in the case of Krasnoyarsk Region). Informatization and Communication, No.1, pp. 43-47 (2014) (in Russian)
5. The Standard Territory Passport of Regions and Municipal Areas: The Regulation of Ministry of Emergency, No.484, 25/10/2004 (in Russian)
6. Williams G.J., Simoff S.J.: Data Mining: Theory, Methodology, Techniques, and Applications: Springer, 329 p (2006)
7. Gorban A., Pitenko A., Zinovyev A.: ViDaExpert: User-friendly Tool for Nonlinear Visualization and Analysis of Multidimensional Vectorial Data. Cornell University Library. <http://arxiv.org/abs/1406.5550>
8. Using ArcViewGIS: The Geographic Information System of Everyone: ESRI Press, 350 p (1996)
9. Abdi H., Williams L.: Principal Components Analysis. Wiley Interdisciplinary Reviews: Computational Statistics, Vol. 2(4), pp. 439-459 (2010)
10. Peres-Neto P., Jackson D., Somers K.: How Many Principal Components? Stopping Rules for Determining the Number of Non-trivial Axes Revisited, Computational Statistics & Data Analysis, N 49(4), pp. 974-997 (2005).