ECOLOGICAL REGIONALIZATION METHODS OF OIL PRODUCING AREAS

Inna Pivovarova[1], Aleksei Makhovikov[1]

[1] Department of Informatics and Computer Technology, Saint-Petersburg Mining University, St. Petersburg, Russia, e-mail: I_Pivovarova@mail.ru

ABSTRACT

The paper analyses territory zoning methods with varying degrees of anthropogenic pollution risk. The summarized results of spatial analysis of oil pollution of surface water in the most developed oil-producing region of Russia. An example of GIS-zoning according to the degree of environmental hazard is presented. All possible algorithms of cluster analysis are considered for isolation of homogeneous data structures. The conclusion is made on the benefits of using combined methods of analysis for assessing the homogeneity of specific environmental characteristics in selected territories.

Keywords: zoning, ecosystem, cluster analysis, GIS, uniformity assessment

INTRODUCTION

Siberian water resources qualitative depletion as a result of oil-producing companies' activities currently becomes the problem of not only regional but also on a national scale. The territory of the West Siberian oil and gas province in Russia is one of the largest oil producing regions of the world. More than 55 percent or Russian oil is produced here; it is more than 6 percent of globally produced oil. Since the beginning of the development of oil fields in 1964, the cumulative oil production only in the territory of Khanty-Mansiysk Autonomous Okrug – Yugra exceeded 7,800 mln tons. According to environmental monitoring data, surface waters in the region are characterized by elevated concentration of ammonia phenols, zinc, manganese and iron compounds, but a particular hazard to rivers and lakes of the Ob-Irtysh basin is posed by oil hydrocarbons which account for about 80 percent of the overall pollution [Moiseenko et al. 2012].

Production activities are complicated by a wide spread of permafrost, considerable bogginess and low biological production potential reduce the stability of natural systems against techno-genes and complicate production activity. Experience in the construction and operation of oil and gas facilities in the north of Western Siberia clearly shows that this area is characterized by a high risk of environmental degradation [Panicheva et al. 2012]. Crude oil ingress to surface water as a result of leaks and accidents at oil-producing facilities what leads to the formation of a thin film on water surface, while heavy hydrocarbon fractions accumulate on the bottom of water bodies and then seeps into the groundwater and soil. All this leads to the destruction of the natural ecosystem and results in shrinking the territories of traditional natural resources use by indigenous minorities which are engaged in fishing, commercial hunting and reindeer breeding. According to the Center for Disease Control and Prevention (CDC), toxic components of crude oil may adversely affect the endocrine and cardiovascular system of man and damage to the DNA structure can lead to various cancers and congenital defects in future generations [EPA 2011]. Therefore, surface water quality assessment, identification of the most and least contaminated areas allows estimating the scale of man-induced impact on the natural ecosystem. The results of such ecological regionalization will make it possible to develop engineering and technical measures to preserve favorable and mitigate unfavorable changes in the conditions of development of oil and gas sector in this region.

For a long time environmental scientists of different countries focus their attention on the concept of ecological regionalization. In 1967, Crowley first presented the concept of ecoregion, which refers to the land and water areas with similar ecosystem or being supposed to play similar functions [Crowly 1967]. Basing on this concept, the purpose of ecological regionalization is to provide suitable spatial units for studying, evaluating, restoring and managing the ecosystem [Omernik 1997]. The concept of aquatic ecoregion originated from America. It refers to the freshwater ecosystem or living organism and the interrelated land units [Omernik 1987]. The aquatic ecosystems zoning is one of the most important fields of ecological regionalization, and it is also the field most successfully studied [Isaac 1999]. However, this paper focuses mainly on the development of methods for continuous monitoring and spatial zoning of areas with different degrees of anthropogenic pollution risk. In the optimized economic system each risk indicator threshold is managed in a way that ensures maximum profitability [Ugarov 2005]. Being aware of the extent of the ecological hazards, government bodies will be able to make reasoned management and investment decisions at local level. It is well-known that costs of forecasting of and preparation to natural disasters are on average 15 times less than the avoided damage [Osipov 2009].

MATERIALS AND METHODS

The algorithm integrated assessment and zoning of the area of interest according to the degree of environmental hazard included two phases:
1) Development of an updatable relational database. Interpolation and visualization of processed parameters using capabilities of GIS-applications.
2) Clusterization and identification of environmental areas. Building the hierarchical system of taxa for zoning.

The input data used are materials of environmental and geochemical monitoring of surface waters over the allocated subsoil reserve and information about the level of anthropogenic impact for one calendar year (January – December 2012) for 16 largest oil fields Khanty-Mansiysk Autonomous Okrug – Yugra) [Moskovchenko et al. 2014]. To evaluate the intensity of oil pollution we used such indicators as the average oil content in water bodies within the license site, the median value and the frequency of detection of exceeding maximum permissible concentrations (MPCs) as percent of the total number of observations (Table 1).

First stage – GIS-zoning of the territory in terms of environmental hazard level. In general, this module is a Geographic Information System (GIS) for data ranking by pollution of water bodies with oil products in the territory, including:

Table 1. Initial data

No.	Fields	Longitude (DD. DD)	Latitude (DD. DD)	MPC (%)	Mean (mg/dm³)	Median (mg/dm³)
1	Kamynskoe	71.600	62.150	25.000	0.071	0.042
2	Kamennoe	67.250	61.667	11.800	0.052	0.020
3	Vakhskoe	79.000	60.917	39.800	0.064	0.041
4	Druzhnoye	74.817	62.417	0.000	0.043	0.045
5	Ombinskoe	73.073	60.907	15.800	0.037	0.033
6	Prirazlomnoe	70.750	60.883	12.300	0.054	0.032
7	Pravdinskoye	71.833	60.917	64.900	0.084	0.069
8	Priobskoe	70.167	61.167	25.000	0.038	0.034
9	Samotlorskoe	76.633	61.167	40.500	0.155	0.050
10	Salymskoe	71.283	60.800	26.200	0.043	0.038
11	Mid-Ugutskoye	73.917	60.500	48.600	0.052	0.048
12	Srednebalykskoe	72.533	60.467	29.400	0.036	0.023
13	Ugutskoye	74.150	60.517	36.000	0.066	0.044
14	South-Yagunskoe	74.250	62.717	0.000	0.047	0.050
15	South-Surgutskoe	73.000	61.133	36.700	0.058	0.045
16	Petelenskoe	71.980	60.610	12.100	0.045	0.034

- databases (formalized data tables with five indicators),
- software for referencing and interpolation of spatially distributed data (Mapinfo-GIS),
- visualization and electronic mapping software (Mapinfo-GIS and Global Mapper GIS).

A bitmap-vector image was obtained where the field of characteristic under study (frequency of cases of MPCs exceeding) is classified according to the degree of increase in the form of Grid-theme. Algorithms built into GIS interpolate the displayed indicator over the grid nodes using various mathematical methods (solving systems of linear equations, triangles converging algorithms, inverse distance method, etc.), the above gives visual representation of the factor as probability surfaces. Layer-by-layer imposition of objects to be mapped (in this case it is point coordinate-referenced symbols of producing oil fields) over the layer of element under study allowed obtaining an easy-to-see distribution of the indicator on the territory over the area. It should be noted that MapInfo-GIS has a variety of methods for data interpolation to build theme surfaces. The main ones are inverse distance weighting (IDW) method (weighted average values in adjacent points for on a given number of neighbors or within a given radius); Kriging (multi-stage fitting of a mathematical function for on a given number of points or for points within a given radius to propagate dependencies to all points); Natural Neighbor (finds the closest subset of input samples to a query point and applies weights to them based on proportionate areas in order to interpolate a value); Bilinear (bilinear interpolation in which the value of interpolated point in the new image is calculated by linear interpolation between the values of the four closest points); TIN (method in which all reference points are connected by triangles resulting in formation of an irregular triangulated network) [Pivovarova 2015]. In this study we used the inverse distance weighting method that allowed us to carry out zoning of the territory according to pollution degree. Seven license sites have been classified as heavy pollution zones (maximum – Pravdinskoye field). Six sites fell into medium pollution zones; in the area of two fields oil in the nearby water bodies was practically absent (Figure 1).

Second stage – the use of cluster analysis procedures for zoning of territory by pollution level. Cluster analysis accompanies a variety of methods to detect structures inherent to a complex set of data. The data basis is most often the sample of objects each of which is described by a set of individual variables. The problem is to combine variables or elements in a group in clusters in such a way so that elements within a cluster would have a high degree of "natural affinity" to each other, while the clusters themselves would be "quite different" [Dulepov et al. 2004].

In concept, cluster analysis as sums that nothing or little is known about the structure inherent to the data set. All that we have is the set of data. The purpose of analysis in this case is the discovery of certain "categorical" structure, which would be consistent with observations and would allow to highlight uniform environmental areas in the territory of interest.
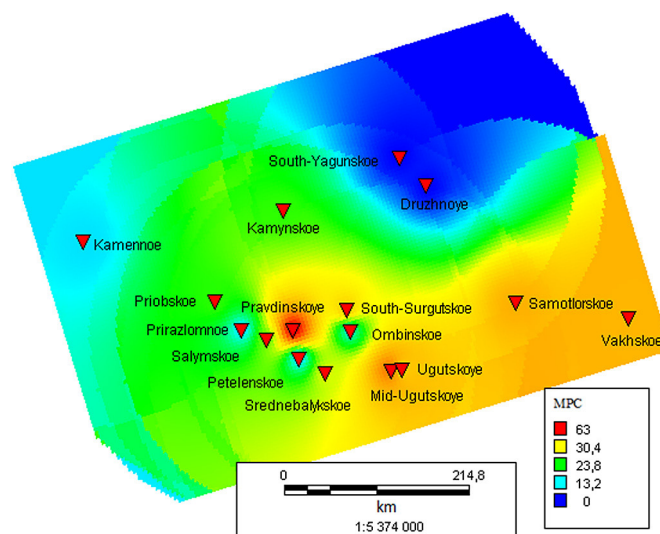

Figure 1. GIS-zoning in MapInfo

In general, main stages of cluster analysis represented the following:

- selection of objects comparable to each other;
- selection of a set of attributes which will be used to compare and describe the objects;
- calculation of similarity measure for objects (or inequality measure for objects) in accordance with chosen metric;
- grouping objects into clusters using one or another grouping procedure;
- verification of applicability of the obtained cluster solution.

In this study we performed cluster analysis of variables using two methods: hierarchical and non-hierarchical, namely, by constructing dendrograms and by *k*-means method.

Hierarchical cluster analysis system

The hierarchical method was used to process data in all possible ways: Average Linkage (Between Groups), Average Linkage (Within Groups), Single Linkage, Complete Linkage, Centroid Linkage, Median Linkage, Ward Linkage. Preliminary assessment of the cluster procedure was carried out by the dendrogram on which the distances of similarity measure were displayed between the individual values in the observation points and groups of the same characteristics. The squared Euclidean distance was used as the main similarity measure. Data were pre-normalized. In the process of clustering method of analysis, the distance formula and the number of clusters in the reference algorithm were determined. As a result, the cluster procedure showed decomposition into three groups: cluster 7; clusters 4 and 14; the last group – all the rest of the 16 studied (Figure 2).

Clustering using k-means method

Although hierarchical clustering methods are accurate, they are laborious: at each step you need to build a distance matrix for all current clusters. The time needed for computations increases in proportion to the cubed number of observations; in the presence of large amounts of data it can result in errors in the calculation even in such powerful software as SPSS Statistics (this is a computer application we used in our study to perform cluster analysis). The k-means algorithm has such benefits as speed and ease of implementation. In our case shortcomings of the method, namely, the uncertainty of the choice of initial cluster centers, and that the number of clusters must be set in advance, was easily offset by the availability of a priori information obtained at the previous stages. In general, the computation algorithm is an iterative procedure with the following steps:

1) A certain number of clusters *k* is selected.
2) From the original set of data *k*-records are randomly selected, which will serve as initial cluster centers.
3) For each record in the original sample a cluster center closest to it is determined. Records "attracted" by centers form initial clusters.
4) Centroids are computed – cluster's centers of gravity. Each centroid is a vector whose elements are the means of attributes calculated over all the records of the cluster. The cluster center is then moved to its centroid.

Steps 3 and 4 are iteratively repeated. It is obvious that after each iteration clusters' boundaries are changed and their centers are shifted. As a result, the distance between elements within clusters is minimized. The algorithm is stopped
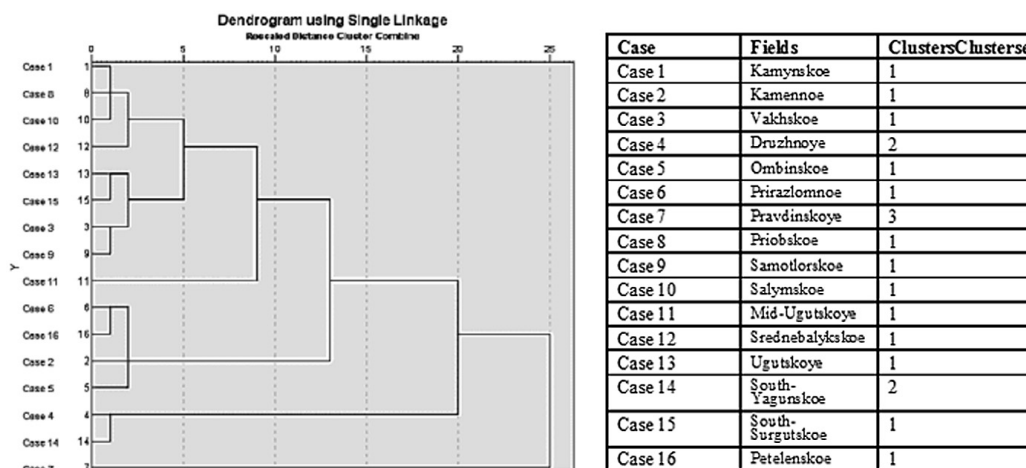

Case	Fields	ClustersClusterse
Case 1	Kamynskoe	1
Case 2	Kamennoe	1
Case 3	Vakhskoe	1
Case 4	Druzhnoye	2
Case 5	Ombinskoe	1
Case 6	Prirazlomnoe	1
Case 7	Pravdinskoye	3
Case 8	Priobskoe	1
Case 9	Samotlorskoe	1
Case 10	Salymskoe	1
Case 11	Mid-Ugutskoye	1
Case 12	Srednebalykskoe	1
Case 13	Ugutskoye	1
Case 14	South-Yagunskoe	2
Case 15	South-Surgutskoe	1
Case 16	Petelenskoe	1

Figure 2. Hierarchical method

when clusters' boundaries and centroid locations stop changing from iteration to iteration, that is the same set of records remains at each iteration, in each cluster. In our case, the algorithm found a set of stable clusters during a few tens of iterations. As a result, two areas were identified: the first area (clusters 3, 7, 9, 11, 13 and 15) – most polluted zone, clusters 1, 2, 4, 5, 6, 8, 10, 12, 14 and 16 – less polluted zone (Figure 3).

RESULTS AND DISCUSSION

Thus, in this paper we evaluated methods of spatial analysis and presented summary results in the form of thematic map with allocation of ecological risk zones (Figur4). The level of oil pollution of a watershed was evaluated using three statistical indicators: arithmetic mean, median and the proportion of samples exceeding the MPC. Notably, some discrepancy between the rank values of oil pollution indicators was identified. For example, the maximum value of the arithme-

tic mean was obtained for Samotlor license site, while in terms of the median it is inferior compared to Pravdinskoye and in terms of percentage of samples exceeding MPC it is ranked third. This is due to the presence in the sample of outliers where MPC is exceeded by several ten-folds, indicating that pollutants are from point sources located in close proximity to the watershed. Cases when oil flows directly into water bodies are rare. According to the reported data on accidents, direct flow of oil into the rivers and lakes in the year under review was observed only at four licensed sites. The weight of oil coming into water bodies was less than 0.1% the total weight of all pollutants. This is consistent with the opinion of the prevalence of diffuse sources of oil pollution (local spills) over point sources [Kalinin 2001], [Kalinin, 2010].

The role of the coordinate referencing of data should be noted particularly because without it the spatial analysis does not make sense. Sources of environmental hazard geographically correlate with polluted areas. The proximity of the risk el-
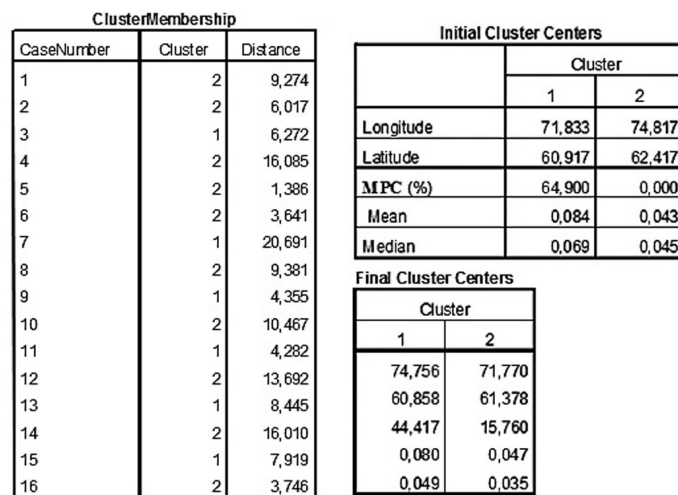
ClusterMembership

CaseNumber	Cluster	Distance
1	2	9,274
2	2	6,017
3	1	6,272
4	2	16,085
5	2	1,386
6	2	3,641
7	1	20,691
8	2	9,381
9	1	4,355
10	2	10,467
11	1	4,282
12	2	13,692
13	1	8,445
14	2	16,010
15	1	7,919
16	2	3,746

Initial Cluster Centers

	Cluster	
	1	2
Longitude	71,833	74,817
Latitude	60,917	62,417
MPC (%)	64,900	0,000
Mean	0,084	0,043
Median	0,069	0,045

Final Cluster Centers

Cluster	
1	2
74,756	71,770
60,858	61,378
44,417	15,760
0,080	0,047
0,049	0,035

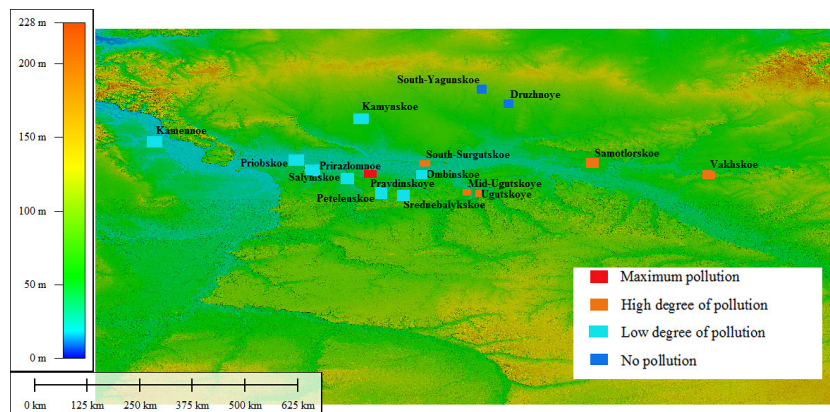**Figure 3.** K-means method

Figure 4. Ecological risk zones

ement to the pollution focus determines the intensity of hazardous impact and possible damage depend on, and the frequency of dangerous occurrences determines the risk. Thus, when identifying areas of adverse impact, the use geographic coordinate space is necessary to assess the area and intensity of environmental damage. That is why geographic coordinates, along with the other three attributes were used in all types of cluster analysis conducted. Speaking about the proximity of pollution source to the watershed, an important role of landscape should be noted: it can weaken or strengthen the adverse impact of pollution source. Typing and landscape zoning are very important in the identification of environmental anomalies. That is why a schematic representation of areas was made in the geographic information system Global Mapper GIS. We used data on the topography of the region studied as the topographic base for further application of thematic layers. Digital elevations of the underlying surface provided by Consortium for Spatial Information (CGIAR-CSI) [CGIAR-CSI 2015], also available on the official NASA website [NASA 2015], have a high degree of resolution (30-90 meters). The absolute error of elevation data for Eurasia is 6.2 m; the relative error of elevation is 8.7, all errors within the confidence interval 90%. The topography with such a degree of detail and accuracy serves very well as the basis for any GIS project and allows more clearly determine the lines of the watercourse [Makhovikov and Pivovarova 2015].

In this study two areas have been identified based on the combined results of cluster analysis using *k*-means method and GIS-zoning. In each of these areas spots of maximum and minimum pollution have been identified (the result of a hierarchical clustering procedure). The conclusion was made about the promising outlook for application of algorithms in ecological regionalization and visualization of zones with a maximum risk of anthropogenic pollution.

CONCLUSION

Negative impact of oil production is due both to the direct pollution of water bodies, and the influence of oil components on the neighboring compartments; that is why oil transformation products are found in different biosphere objects and of course has a very negative impact on habitat in the region. It should also be noted that current high degree of chronic contamination of the oil-producing areas can promote the negative impact of potentially dangerous facilities in case of man-made and combined natural and man-made emergency situations. All this determines the need to develop environmental protection technologies in the Russian oil and gas industry. However, despite some steps taken by the leading Russian companies, the prospects for their mass adoption today seem unlikely [Nikolaichuk and Tsvetkov 2016]. Joint activities are needed of public and local authorities and oil industry companies in order to ensure an acceptable level of environmental safety in the territories of intensive oil production.

REFERENCES

1. Crowly J M., 1967. Canadian Geographer. Biogeography, 11(4), 312–326.
2. CGIAR-CSI official web site. URL: http://srtm.csi.cgiar.org (accessed 06.10.2016)
3. Dulepov V.I., Leskov A.A., Maiorov I.S. 2004. System ecology: textbook Vladivostok. 252
4. Inventory of U.S. Greenhouse gas emissions and sinks: 1990–2009. 2011. EPA 430-R-11-005. U.S. Environmental Protection Agency. Released for printing: April 15.
5. Isaac B. 1999. The conceptual development and use of ecoregion classifications. Master's Thesis, Corvallis, Oregon: Oregon State University.
6. Kalinin V.M. 2001. Petroleum products from diffuse sources in the river network (casestudy of the Middle Ob). Vestnik Tiumenskogo gosudarstvennogo universiteta —Tyumen State University Herald. 2, 11–21.
7. Kalinin V.M. 2010. Formation of technological background of petroleum hydrocarbons in natural waters of oil and gas fields. Vestnik Tiumenskogo gosudarstvennogo universiteta — Tyumen State University Herald. 7, 137–146.
8. Makhovikov A., Pivovarova I. 2015. Free Data Use for Designing GIS-Projects in the Ecology Students Training Course. Journal of Engineering and Applied Sciences, 10(8), 257–260.

9. Moiseenko T., Shalabodov A., Gashev S. 2012. The quality of Siberian waters. Science in Russia. 4, 13–19.

10. Moskovchenko D.V, Ubaydullaev A.A. 2014. Impact of oil spills on the surface waterS pollution IN Khanty-Mansi Autonomous AREA – Yugra. Vestnik Tiumenskogo gosudarstvennogo universiteta – Tyumen State University Herald. 4, 5–16.

11. NASA official web site. URL: http://www.nasa.gov (accessed 06.10.2016)

12. Nikolaichuk L.A., Tsvetkov P.S. 2016. Prospects of Ecological Technologies Development in the Russian Oil Industry. International Journal of Applied Engineering Research. 11, 5271–5276.

13. Omernik J. M., Bailey R. G. 1997. Distinguishing between watershed andecoregion. Journal of American Water Resources Association, 33(5), 935–949.

14. Omernik J M. 1987. Ecoregions of the Conterminous United States (MapSupplement). Annals of the Association of American Geographers, 77(1), 118–125

15. Panicheva L.P., Moiseenko T.I., Kremleva T.A., Volkova S.S. 2012. Biochemical transformation of petroleum hydrocarbons in the waters of Western Siberia. Vestnik Tiumenskogo gosudarstvennogo universiteta – Tyumen State University Herald..12, 38–48.

16. Pivovarova I. 2015. Systematic Approach in Ecological Zoning. Journal of Engineering and Applied Sciences. 10(1), 11–15

17. Osipov V.I. 2001. Natural disasters at the turn of the xxi century. Herald of the Russian academy of sciences. 71(4), 291–302

18. Ugarov A.N. 2005.Information conveyer. CADmaster. 3, 34–37.