Research Journal of Applied Sciences 11 (6): 321-326, 2016
ISSN: 1815-932X

Statistical Methods of Ecological Zoning

Inna Pivovarova and Aleksei Makhovikov
Department of Informatics and Computer Technology,
National Mineral Resources University "Mining", 199106 St. Petersburg, Russia

Abstract: The aim of our study was to develop specific methods that make possible to identify and visualize focuses of ecosystem pollution for the determinationof the permissible anthropogenic impact boundaries and prediction of their development. The two-level spatial zoning method is considered through the example of an environmentally unfavorable region of the Russian Federation. The set of analytic methods is proposed, which was applied to determine the preliminary boundaries of areas using the clustering procedure, while the algorithms of correlation and regression analyzes allow to assess the uniformityof specific environmental characteristics in selected areas.

Key words: Zoning, ecosystem, cluster analysis, GIS, correlation, algorithms, uniformity

INTRODUCTION

Every environmental study begins and ends with zoning. On the one hand, it is a part of the initial stage of the study of ecosystem's natural conditions and it contributes to the creation of an operational hypothesis on the ecosystem's chorologic structure (pre-model), which is verified and refined in the course of field studies, thus ensuring their purposeful nature. On the other hand, zoningis a body of knowledge about the nature of the ecosystem, similarities and differences between its areas. In the most general sense, zoningmeans a process of multifactorial division of the territory into a set of non-overlapping integral areas, which are compact crowding of some original cells (points) both in a three-dimensional physical space and in the multidimensional attribute space. The goal of zoning is to study the causes and factors of formation and differentiation of separate areas of the ecosystem and to identify the nature of interconnections between them.An equally important stage is the revealing of changes in areas under the influence of economic activity, definition of the boundaries between areas, development of a hierarchical system of zoning taxa, mapping of zoning schemeandpreparation of legends to relevant maps. Therefore, various methods of classification and zoning of ecosystems are available today (Bailey *et al*., 1985, Klijn *et al*., 1995). In our opinion, the most universal approach is the one where areas can be separated based both on one attribute (single attribute zoning) and on several attributes (multiple attribute zoning). If we base zoningon any single attribute, then, as a rule, there is only one possible choice when drawing the boundary. If there are more than one zoning factors, then there may be several variants of boundaries. In this case multivariate analysis methods are used. The "quality of zoning" i.e., the selected area's correspondence to the stated objectives largely depends on the choice of study method.The most widely used methods of regression analysis and cluster analysis (Acreman and Sinclair, 1986; Jongman *et al*., 1987) often exhibit a high degree of subjectivity. In fact, the use of different datasets, too careful consideration or on the contrary, lack of attention to the impact of constantly changing anthropogenic factors can lead to different zoning schemes. Therefore, in this study, we first propose cluster analysisfor multiple attribute zoning and then check the results by means of methods of regression and correlation analysis for a single attribute.

MATERIALS AND METHODS

Cluster procedures of multivariate analysis: Different methods of association and similarity measures are available for cluster procedures; selection of the most appropriate option is a problem faced by many researchers. Unfortunately, various clustering procedure applied to the same data setquite often produce different versions of clusters. The choice of clustering procedure is often implicitly specified depending on the data structure. Furthermore, if the individual objects actually do not form a cluster, then in any case such procedures will find non-existent clusters. If they really do exist, then different clustering procedures must produce similar

Corresponding Author: Inna Pivovarova, Department of Informatics and Computer Technology,
National Mineral Resources University "Mining", 199106 St. Petersburg, Russia

results (Gubareva 2012). In this case, the agglomerative hierarchical groupingwas chosen, which is recommended when there is no clear a priori grounds to assign objects to one group or another (Aivazyan *et al*., 1974).

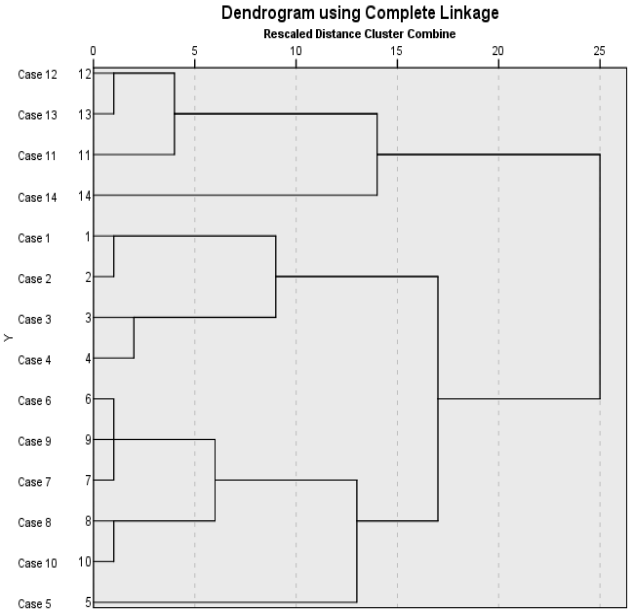
The principle of agglomerative hierarchical algorithms is based on successive merging of groups of elements, i.e., the creation of a hierarchical class structure. The object of our study was the stretch of Klyazma River with length of 163 km with environmental monitoring data collected at 14 monitoring stations located between the city of Moscow and the city of Vladimirin Central Russia. For the selected study area data from ecological reports showed significant river water pollution by heavy metals.

The most prevailing elements are copper, zinc, iron and manganese; nickel and lead are found more rarely. Cluster analysis on a set of six attributes was performed (levels of pollutants in the river system averaged over 10 years), (Table 1). Preliminary assessment of the cluster procedure was carried out using a dendrogram,which displays the distance of similarity measure between the individual values in the monitoring points and groups of the same characteristics. The main similarity measure used was the Euclidean distance squared. Data were normalized before hand. In the course of clustering, the following were determined: method of analysis, formula for the distance andthe number of clusters in the reference algorithm. The following methods were tried: Average Linkage (Between Groups), Average Linkage (Within Groups), Single Linkage, Complete Linkage, Centroid Linkage, Median Linkage and Ward Linkage. Three methods (Average Linkage (Between Groups), Complete Linkage and Ward Linkage) gave matching cluster cores with one exception: the area was divided into 2 final groups, within the first group 9 cores coincided; within the second group 3 cores.

Clusters 1-10, except the 5th cluster and 12th, 13th and 14th are grouped into two different groups for all three clustering methods and different versions of data normalization (Fig.1). Cluster 5 is far from the second group geo graphically, so grouping would look unrea sonable, from the view point of the ecosystem formation conditions. The variant of grouping cluster 11 June 30, 2016in one group with clusters 12 13 and 14 needs additional verification but, using other methods of analysis.

Table 1: Levels of pollutants in the river system averaged over 10 years (1995-2004)

Monitoring stations	lat	dol	Fe	Zn	Cu	Mn
Tarasovka	55,95	37,82	160	27	4	44
Schelkovo1	55,92	37,98	250	37	8	40
Sverdlovskiy	55,88	38,12	480	46	7	44
Losino-Petrovsky	55,87	38,2	375	64	13	36
Noginsk1	55,85	38,37	800	25	9	100
Noginsk2	55,87	38,47	500	20	15	60
Noginsk3	55,85	38,4	345	25	21	59
A/m bridge	55,85	38,55	400	35	13	30
Gavrino	55,8	38,65	590	27	20	60
Or.Zuevo1	55,8	38,97	410	25	8	41
abover.Kirzhach	55,87	39,08	510	10	8	14
Petushki	55,92	39,45	690	10	7	30
Sobinka	55,98	40,02	690	7	5	29
Vladimir	56,13	40,42	740	32	5	38



Cluster Membership

Case	2 Clusters
1:Case 1	1
2:Case 2	1
3:Case 3	1
4:Case 4	1
5:Case 5	1
6:Case 6	1
7:Case 7	1
8:Case 8	1
9:Case 9	1
10:Case 10	1
11:Case 11	2
12:Case 12	2
13:Case 13	2
14:Case 14	2

Fig. 1: Continue

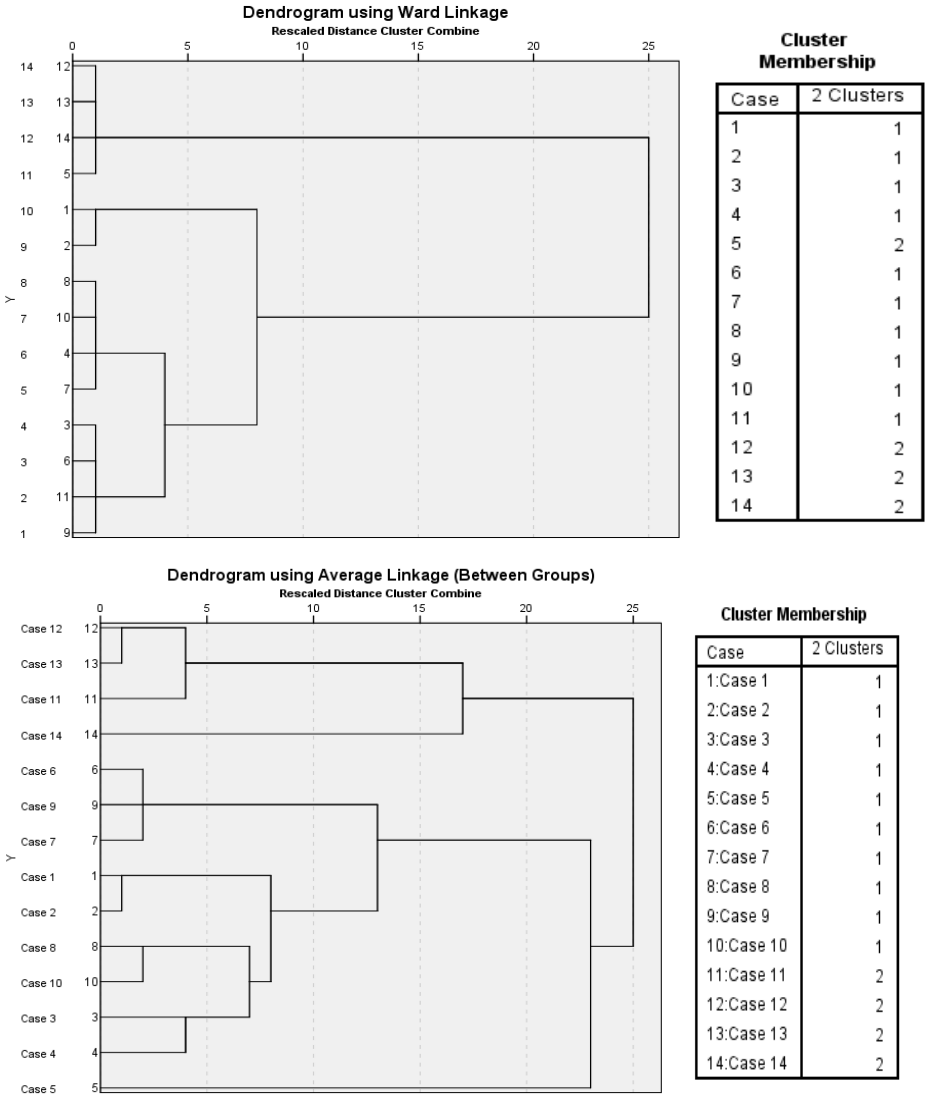


Fig. 1: Grouping by cluster analysis; a) Complete Linkage; b) Ward Linkage; c) Average linkage (between groups)

RESULTS AND DISCUSSION

Methods of correlation and regression analyses of the environmental data uniformity: To analyze the relationships between variables, correlation and regression analyses are used. The correlation analysis describes the closeness of correlations between variables. The regression analysis is used to obtain the most suitable equations, approximating those correlations (Pivovarova, 2015). As a mathe matical algorithm for problem solution we propose to use Fischer's Z-distribution, when the statistical estimate obtained for compared groups of characteristics is compared with the theoretical value at a given confidence level (Fisher, 1928). In brief, the calculation method included construction a space-correlation function and evaluation of the difference between the actual correlation coefficient and the expected coefficient in the total population (Alekseev, 1971). Mathematical algorithm calculations were as follows: according to Fischer, the values of empirical $(a)_{jk}$ and theoretical r (a_{jk}) correlation functions are used to determineinter mediate quantities:

$$z_{jk} = \frac{1}{2}\ln\frac{1+r(a_{jk})}{1-r(a_{jk})}$$

$$\tilde{z}_{jk} = \tilde{z}(\alpha_{jk}) = \frac{1}{2}\ln\frac{1+\tilde{r}(\alpha_{jk})}{1-\tilde{r}(\alpha_{jk})} + \frac{r(\alpha_{jk})}{2(N_{jk}-1)}$$

Table 2: Evalutation of homogeneity of correlation function (a fragment over of calculation is brought on 13 paris from 91)

Pairs	Distance between monitoring points	Pairwise correlation coefficient	Value of empirical correlation functions	Value of theoretical correlation functions	Z_{JK}	$\tilde{Z}_{JK}$	Rejection (Δ)	Mean square error (δ)	Doubled mean square error (2δ) exceeds	Case,when a rejection exceeds a MSE (Δ>δ)	Case, when a rejection a doubled MSE (Δ>2δ)
1-14	41,5	0,199	0,658	0,817	1,149	0,835	0,314	0,33	0,67		
2-14	31,3	0,132	0,695	0,862	1,302	0,905	0,397	0,33	0,67	1	
3-14	21,9	0,383	0,728	0,904	1,492	0,976	0,516	0,33	0,67	1	
4-14	16,6	0,390	0,748	0,927	1,636	1,019	0,617	0,33	0,67	1	
5-14	6,7	0,482	0,783	0,971	2,101	1,107	0,994	0,33	0,67	1	1
6-14	5	0,299	0,789	0,978	2,249	1,124	1,125	0,33	0,67	1	1
7-14	5,4	0,397	0,788	0,976	2,210	1,120	1,090	0,33	0,67	1	1
8-14	13,9	0,437	0,757	0,939	1,728	1,042	0,686	0,33	0,67	1	1
9-14	32,38	0,584	0,691	0,858	1,284	0,897	0,387	0,33	0,67	1	
10-14	38,26	0,343	0,670	0,832	1,193	0,856	0,337	0,33	0,67	1	
11-14	62,1	0,273	0,584	0,727	0,922	0,708	0,213	0,33	0,67		
12-14	97,7	0,279	0,456	0,570	0,648	0,523	0,124	0,33	0,67		
13-14	125,5	0,515	0,355	0,447	0,481	0,396	0,085	0,33	0,67		

and deviations (differences) are calculated, $z_{jk} - \tilde{z}(\alpha_{jk})$ for all $c_1^2 = \frac{l(l-1)}{2}$ pairwise distances between monitoring points. Standard deviations σ_{zjk} of intermediate quantities from their conditional mean values $\tilde{z}(\alpha_{jk})$ are then determined using the formula:

$$\sigma_{zjk} = \frac{1}{\sqrt{N_{jk} - 1}}$$

Based on normal distribution of the normalized deviations from the mean value, the confidence limits:

$$\tilde{z}(\alpha_{jk}) - t_{\sigma jk} < z_{jk} < \tilde{z}(\alpha_{jk}) + t_{\sigma zjk}$$

shall contain P(1) = 0.683 = 68.3% of all empirical values with t = 1, or P(2) = 0.954 = 95.4% with t = 2. Therefore, a necessary and almost sufficient condition of the correlation functionuniformity within the area of interest is that the inequalities:

$$\left| z_{jk} - \tilde{z}(\alpha_{jk}) \geq \sigma_{zjk} \right| \text{or} \geq 2\sigma_{zjk}$$

are valid in approximately 31.7% or 4.6% of the total number of cases $c_1^2 = \frac{l(l-1)}{2}$ of empirical values . In other words, at t = 1 and t = 2 the total empiric number of excesses:

$$K_e(1) = K_e\left[z_{jk} - \tilde{z}_{jk} \right] > \sigma_{zjk}$$
$$K_e(1) = K_e\left[z_{jk} - \tilde{z}_{jk} \right] > 2\sigma_{zjk}$$

Shall be approximately equal to the theoretically possible number of excesses based on the normal distribution law:

$$K_e(1) \approx 0.317 C_1^2 = 0.317 \frac{l(l-1)}{2}$$
$$K_e(2) \approx 0.317 C_1^2 = 0.046 \frac{l(l-1)}{2}$$

Thus, all the possible pairs were tested of the correlation ratios between the input data describing river water pollution with iron oxides. The correlation was plotted and the equations of empirical and theoretical correlation functions were obtained which are needed for further calculations. They were used to calculate intermediate quantities z_{jk} and $\tilde{z}(a_{jk})$. Standard deviations σ_{zjk} of intermediate quantities from their conditional mean values were determined. As a result, the conclusion was made that the space-correlation function of the area of interest is non-uniform because the total empirical number of excesses is more than the theoretically allowed number equal (Table 2). For the number of excessesis 43 as compared to 29 maximum expected according to the normal distribution law, for $2\sigma_{jk}$-14 and 4, accordingly. In order not to make the calculation process too long and cumbersome, as well as to exclude manual calculation of all intermediate values, it was decided to replace the graphical module for obtaining equationsfor empirical and theoretical correlation functions with the construction of approximating dependences and subsequent determination of the approximation parameters based on minimization of the total squared error (least square method). Thus, the parameters of an equation y = ax+b were obtained using the formulas:

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

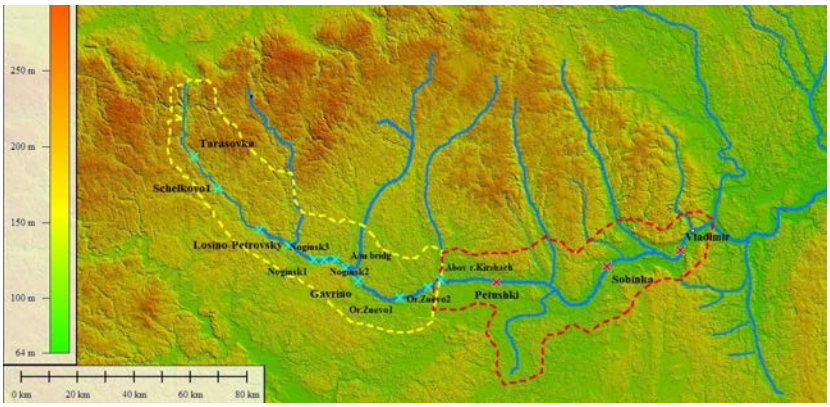


Fig. 2: Lower and higher contamination sub aears

$$b = \frac{\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i}{n}$$

wheren: The number of series terms, x_i = The distance between monitoring points, y_i = pairwise correlation coefficient. Further calculations followed the above algorithm. We then automated calculations, developed VBA application for MS Excel, so similar calculations to solve the ultimate problem of study area subdivision into uniform subareas (zones) was not time and labor intensive. The study area was divided into two subareas: Klyazma-1 and Klyazma-2 for each of which the whole procedure described above was repeated. The geographical boundary of these two zones determined from the results of the cluster analysis. We tested two versions of grouping. The best result was achieved with grouping where the lower contamination subarea is the one from Tarasovka settlement to Kirzhach River (Klyazma-1) and the higher contamination subarea is the one from Kirzhach River to the city of Vladimir (Klyazma-2) (Fig.2). Both subareas turn out to be uniform. The number of excesses for Klyazma-1 by $\sigma_{jk}$is eight and by is $2\sigma_{jk}$ zero, according to the normal distribution law the maximum is 29 and 2,accordingly; forKlyazma-2there were no excesses at all.

CONCLUSION

This study applies the two-level spatial zoning through the example of environmentally problematic region. The proposed classification using cluster analysis methods showed very good results in the initial stage of our study to identify groups of cases, when the grouping was not previously known. Preliminary boundaries of subareas were determined; it made it possible to use the correlation analysis algorithm at the second stage of the study to assess spatial uniformity of environmental characteristics ofpreviously identified subareas. As a result, two subareas were identified for each of which the iron oxides pollution can be attributed to the same general population, where the differences between the characteristics are in the range of random fluctuations. Schematic representation of regions was made by means of a geographic information system Global Mapper GIS. The topographic base for further application of GIS layers was represented by topography data of the region. The digital data on the underlying surface elevation which are provided by the Consortium for Spatial Information (CGIAR-CSI) (http://srtm.csi.cgiar.org/) andare also available at NASA website. (http://www2.jpl. nasa.gov/srtm) are characterized by high resolution (30-90 meters). The absolute uncertainty of elevation data for Eurasiais 6.2 m, the relative uncertainty is 8.7 and all uncertainties are within the confidence interval 90%. The relief of such degree of detail and accuracy is a very good basis for any GIS project and allows more clearly identify the watercourses (Makhovikov and Pivovarova, 2015).

In general, the development and practical application of ecological zoningmethods is fully consistent with the UN recommendations for strategic approach to monitoring and assessment of rivers, lakes and groundwater: “Converting data into information implies their analysis and interpretation. In particular, a comprehensive data management shall be supported by simulation models and GIS. It is recommended to use the soft waread apted to the specific conditions” (International Hydrological Programme (IHP) Eighth Phase in 2011). Among the most significant contributors to the impact on the environment we should note power generation; in Russia it is based mostly on the direct burning of fossil fuels (Tsvetkov, and Strizhenok, 2016). The undeniable superiority in terms of volume of discharged waste water belongs to thermal power industry, drains of the oil industry andgas generator plants. Therefore, the development of methods

for identification and visualization of hot spots is needed to determine the boundaries of permissible anthropogenic impact and to predict their development.

REFERENCES

Acreman, M.C. and C.D. Sinclair, 1986. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. J. Hydrol., 84: 365-380.

Aivazyan, S.A., Z.I. Bezhaeva and O.V. Staroverov, 1974. Classification of Multidimensional Observations. Statistika, Moscow, Russia, Pages: 238.

Alekseev, G.A., 1971. Objective Methods for Alignment and Normalization of the Correlations Relationships-Leningrad. Gidrometeoizdat, Leningrad, Pages: 180.

Bailey, R.G., S.C. Zoltai and E.B. Wiken, 1985. Ecological regionalization in Canada and the United States. Geoforum, 16: 265-275.

Gubareva, T.S., 2012. Classification of river basins and hydrological zoning (an example of Japan). Geogr. Nat. Resour., 33: 74-82.

Jongman, R.H.G., C.J.F. Braaks and O.F.R.V. Tongeren, 1987. Data Analysis in Community and Landscape Ecology. Pudoc, Wageningen, The Netherlands.

Klijn, F., D.R.W. Waal and J.H.O. Voshaar, 1995. Ecoregions and ecodistricts: Ecological regionalizations for the Netherlands environmental policy. Environ. Manage., 19: 797-813.

Makhovikov, A. and I. Pivovarova, 2015. Free data use for designing GIS-projects in the ecology students training course. J. Eng. Appl. Sci., 10: 257-260.

Pivovarova, I., 2015. Systematic approach in ecological zoning. J. Eng. Appl. Sci., 10: 11-15.

Tcvetkov, P. and A. Strizhenok, 2016. Ecological and economic efficiency of peat fast pyrolysis projects as an alternative source of raw energy resources. J. Ecol. Eng., 17: 56-62.