Colloquium Biometricum 42
2012, 111–119

THE APPLICATION OF THE TOOLS OF SPATIAL STATISTICS TO STUDY REGIONAL DIFFERENTIATION OF POLISH ENVIRONMENT

Agnieszka Kamińska

Department of Applied Mathematics and Computer Science
University of Life Science in Lublin
Akademicka 13, 20-950 Lublin
e-mail: agnieszka.kaminska@up.lublin.pl

Summary

The article presents the application of the spatial autocorrelation analysis in evaluation of regional differentiation of Polish environment. The study is based on the chosen data for the sixteen provinces from the year 2005 and 2010. In order to estimate the level of environmental degradation and protection of the environment WAP methods were applied. On the basis of the synthetic measure, developed during the study, a ranking of regions was constructed. Additionally, the analyses were broadened by the use of spatial autocorrelation statistics which enabled to reveal the existing spatial relations.

Keywords and phrases: spatial autocorrelation, TOPSIS method, state of environment

Classification AMS 2010: 62H11

1. Introduction

The natural environment encompasses all living and non-living things occurring naturally on Earth. The concept of the natural environment can be distinguished by complete ecological units that function as natural systems without massive human intervention, including all vegetation, microorganisms, soil, rocks, atmosphere and universal natural resources and physical phenomena, such as air, water, climate.

The aim of this paper is to present the Polish environment. The order system has been shown to provinces according to environmental degradation and measures to protect it. In order to estimate regional differentiation TAPSIS method was applied (Hwang, Yoon 1981). In the paper, on the basis of the synthetic variable, developed during the study, the regions were classified and grouped into clusters, according to their levels of development. Additionally, the analyses were broadened by the use of spatial autocorrelation statistics which enabled to reveal the existing spatial relations.

2. Materials and methods

The study used selected data of the Central Statistical Office (Regional Data Bank) for the sixteen provinces from the years 2005 and 2010. An analytical framework of the synthetic indicator comprises of three main stages, such as: selection of diagnostic features, standardization of chosen features and calculation of indicators value (Wysocki and Lira 2005). Based on factual data a set of variables was proposed.

Variables describing the environmental degradation:

Z_1 – total emission of air pollutants in t per 1km^2,
Z_2 – total emission of air pollutants-gases in t per 1km^2,
Z_3 –industrial and municipal wastewater requiring treatment discharged through sewerage system in dam^3,
Z_4 – devastated and degraded land in percentage of agricultural land,
Z_5– acumulated industrial waste deposited in percentage of generated during the year.

Variables describing protection of the environment:

Y_1 – outlays on fixed assets expenditures on environmental protection on water management in zł. per capita,
Y_2 – devastated and degraded land, reclaimed and brought into cultivation in percentage of agricultural land,
Y_3 – pollutants retained in pollutant reduction systems in percentage of air pollutants generated during the year,
Y_4 – pollutants retained In pollutant reduction systems in percentage of air pollutants-gases generated during the year,
Y_5 – population connected to waste water treatment plant as percentage of total population,
Y_6 – capacity of waste-water treatment plants in percentage of discharged during the year.

All variables describing the environmental degradation were considered destimulants and all variables describing protection of the environment stimulants. In order to normalize the features the unitarization procedure was used based on the following formulas:

for stimulants:

$$c_{ik} = \frac{x_{ik} - \min_i \{x_{ik}\}}{\max_i \{x_{ik}\} - \min_i \{x_{ik}\}} \quad i = 1,2...,n; k = 1,2,...,m \tag{2.1}$$

for destimulants:

$$c_{ik} = \frac{\max_i \{x_{ik}\} - x_{ik}}{\max_i \{x_{ik}\} - \min_i \{x_{ik}\}} \quad i = 1,2...,n; k = 1,2,...,m \tag{2.2}$$

where

$\max_i \{x_{ik}\}$-maximum value of the k-th characteristic,

$\min_i \{x_{ik}\}$- minimum value of the k-th characteristic.

In order to estimate the level of diversity of the environment synthetic measure was used, based on TOPSIS method (Technique for Order Performance by Similarity to an Ideal Solution) developed by Hwang and Yoon (Hwang, Yoon 1981). TOPSIS simultaneously considers the distances to the ideal solution and negative ideal solution regarding each alternative and selects the most relative closeness to the ideal solution as the best alternative. That is, the best alternative is the nearest one to the ideal solution and the farthest one from the negative ideal solution. A relative advantage of TOPSIS is the ability to identify the best alternative quickly. The procedure of TOPSIS is summarized as follows:

- Determination of the positive ideal and negative ideal solution:

$$A^{+} = \left(\max_i (c_{i1}), \max_i (c_{i2}),..., \max_i (c_{iK})\right) = \left(c_1^{+}, c_2^{+},..., c_K^{+}\right) \tag{2.3}$$

$$A^{-} = \left(\min_i (c_{i1}), \min_i (c_{i2}),..., \min_i (c_{iK})\right) = \left(c_1^{-}, c_2^{-},..., c_K^{-}\right) \tag{2.4}$$

- Calculation of Euclidean distance of each alternative from the positive ideal (d^{+}) and negative ideal (d_{-}):

$$d_i^+ = \sqrt{\sum_{k=1}^{K} \left(c_{ik} - c_k^+\right)^2} \quad i=1,\ldots,n \tag{2.5}$$

$$d_i^- = \sqrt{\sum_{k=1}^{K} \left(c_{ik} - c_k^-\right)^2} \quad i=1,\ldots,n \tag{2.6}$$

- Calculation of the relative closeness to the ideal solution:

$$z_i = \frac{d_i^-}{d_i^+ + d_i^-} \quad 0 \le z_i \le 1 \tag{2.7}$$

Ranking the preference order: the closer is z_i to 1 the higher priority of the *i*-th alternative.

On the basis of synthetic measure, administrative units were divided into four typological classes representing different level of the research issue:

– I class: $z_i \ge \bar{z} + s_z$

– II class: $\bar{z} \le z_i < \bar{z} + s_z$

– III class: $\bar{z} - s_z \le z_i < \bar{z}$

– IV class: $z_i < \bar{z} - s_z$

where: $\bar{z}$ - mean, s_z - standard deviation.

Spatial relationships were evaluated on the basis of global and local Moran's *I* coefficients (Anselin 1995). For the synthetic indicator *z* the global Moran's *I* coefficient was calculated according to the formula:

$$I = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \cdot \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2} \tag{2.8}$$

where n – number of observations,

z_i – value of variable for *i*-th location,

z_j – value of variable for *j*-th location,

$\bar{z}$ – average value of variable,

w_{ij} – weight between locations *i* and *j* based on contiguity where the definition of neighbor was based on sharing a common boundary. Contiguity spatial relationships were assigned a value of 1 to neighboring locations and 0 to all other ones. Spatial weights are standardized by row. Each weight is divided by its row sum.

Significant value of I greater than 0 indicates positive autocorrelation. It means that objects have similar values of variables (high values of the variable in location i tend to be clustered with high values of the same variable in locations that are neighbors of i, and vice versa). Value of I less than 0 denotes negative autocorrelation. Negative autocorrelation means, that exist big differences between values for neighboring objects (high values in a variable in location i tend to be co-located with lower values in the neighboring locations).Value of I equal to 0 or close to 0 means random spatial distribution. Significance test for global Moran's I statistics was presented in Cliff and Ord (1981).

These specific relationships can be identified from a Moran scatterplot. This graph depicts a standardized variable in the x-axis versus the spatial lag of that standardized variable. The spatial lag is a summary of the effects of the neighboring spatial units. In essence, Moran scatterplot presents the relation of the variable in the location i with respect the values of that variable in the neighboring locations. By construction, the slope of the line in the scatter plot is equivalent to the Moran's I coefficient. The four quadrants in the scatterplot box thus represent different types of association between the values at a given location and its spatial lag. The upper right and lower left quadrants represent positive spatial association, in the sense that a location is surrounded by similar valued locations. For the upper right this is association between high values , while for the lower left quadrant this is association between low values. The upper left and lower right quadrants correspond to negative association, low values are surrounded by high values (upper left) and high values are surrounded by low values (lower right). The relative densities of these quadrants indicate which of these patterns of negative spatial association (in the traditional sense) dominate (Anselin 1993).

3. Results

Both essential and statistic reasons decided about diagnostic features selection. Hence complex variables were chosen. Moreover mutually strong correlated features were eliminated to dispose of duplicate information. In order to eliminate excessively correlated variables, an inverse matrix was established for correlation coefficients between considered variables. On the basis of the analysis of diagonal elements of the matrix Z2 was eliminated from further investigations. Coefficients of variation were also included in statistical analysis- and, in results – the quasi-constant variable Y3 was rejected.

The results of the classification made on the basis synthetic indicator and assigning to groups due to degradation and protection of the environment for individual provinces are presented in Tables 1-4 and on the maps in Figure 1.

There was a clustering trend in Poland's provincial level environmental degradation (represented by indicator z). Śląskie and Łódzkie had the highest values of synthetic measure z. It means, that the most polluted environment occurs in that areas. Low values of z were observed in the northern and eastern part of Poland. Efforts to protect the environment were carried out in along the line from Pomorskie to Podlaskie. The situation was quite stable in both analyzed years, only a few provinces changed groups.

Table 1. Divide into groups due to environmental degradation for individual provinces in the year 2005

Class number	no	Measure values	provinces
I	3	<0.89; 1>	lubelskie, podlaskie, podkarpackie
II	6	<0.73; 0,89)	lubuskie, warmińsko-mazurskie, mazowieckie, pomorskie, kujawsko-pomorskie, małopolskie
III	6	<0.58; 0.76)	świętokrzyskie, opolskie, zachodniopomorskie, wielkopolskie, dolnośląskie,
IV	1	<0; 0.58)	łódzkie , śląskie

Table 2. Divide into groups due to environmental degradation for individual provinces in the year 2010

Class number	no	Measure values	provinces
I	2	<0.93; 1>	podlaskie, podkarpackie
II	8	<0.77; 0.93)	lubuskie, lubelskie, mazowieckie, małopolskie, pomorskie, kujawsko-pomorskie, warmińsko-mazurskie, zachodniopomorskie
III	3	<0.61; 0.77)	wielkopolskie, dolnośląskie, świętokrzyskie
IV	3	<0; 0.61)	opolskie, łódzkie, śląskie

Table 3. Divide into groups due to protection of environment for individual provinces in the year 2005

Class number	no	Measure values	provinces
I	3	<0.57; 1>	śląskie, łódzkie, dolnośląskie
II	6	<0.47; 0.57)	pomorskie, warmińsko-mazurskie, kujawsko-pomorskie, lubuskie, opolskie, podlaskie
III	3	<0.36; 0.47)	lubelskie, wielkopolskie, świętokrzyskie
IV	4	<0; 0.36)	zachodniopomorskie, mazowieckie, małopolskie, podkarpackie

Table 4. Divide into groups due to protection of environment for individual provinces in the year 2010

Class number	no	Measure values	provinces
I	3	<0.59; 1>	śląskie, dolnośląskie, łódzkie
II	6	<0.49; 0.59)	pomorskie, podlaskie, kujawsko-pomorskie, lubuskie, opolskie, warmińsko-mazurskie
III	4	<0.39; 0.49)	wielkopolskie, zachodniopomorskie, małopolskie, świętokrzyskie
IV	3	<0; 0.39)	lubelskie, podkarpackie, mazowieckie

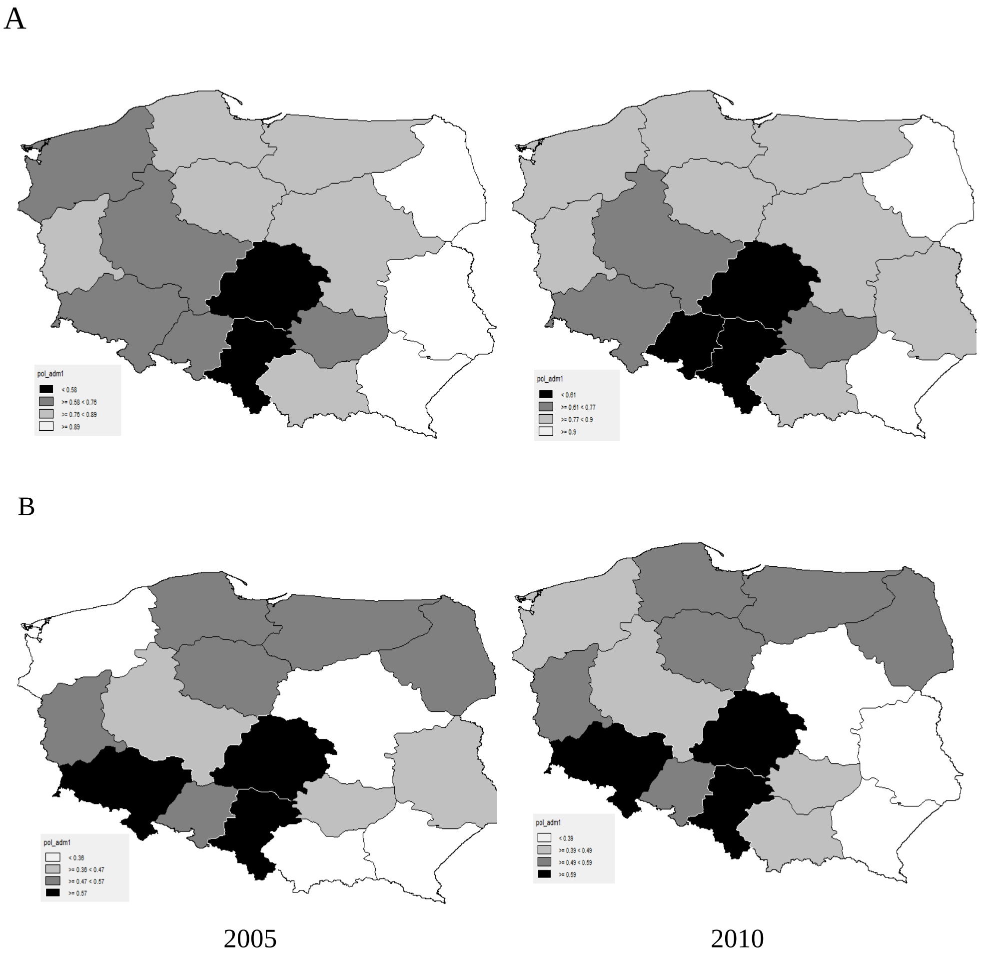
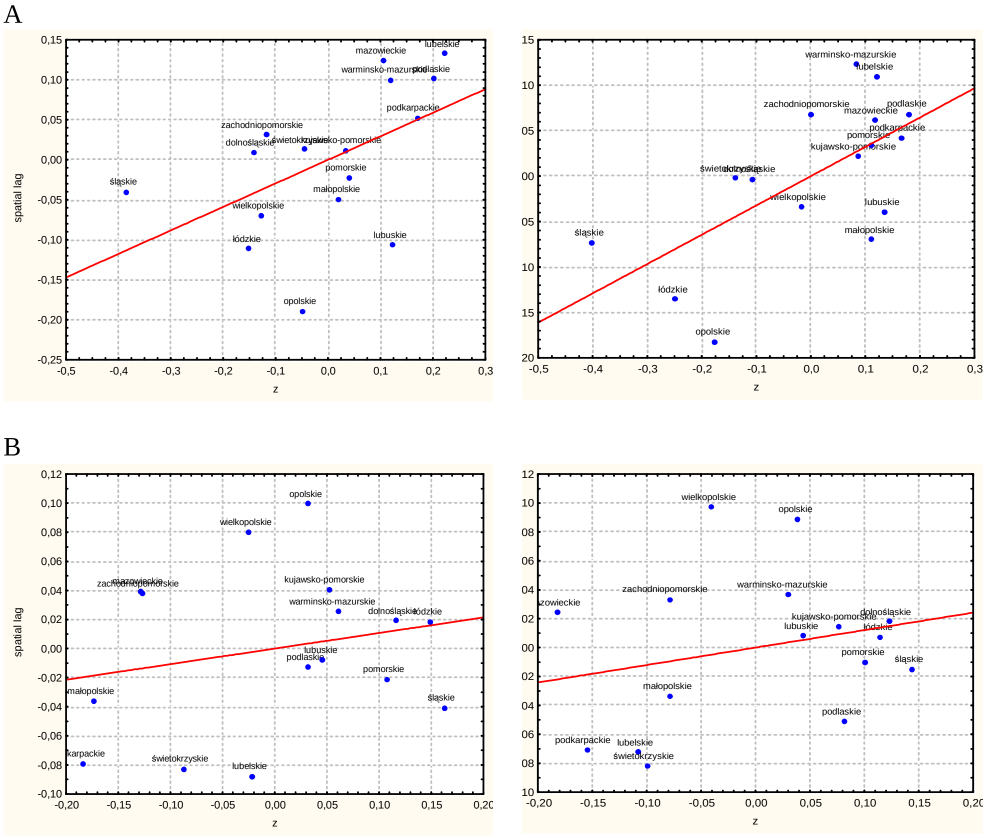


Fig. 1. The values of the synthetic indicator z due to (A) environmental degradation and (B) protection of the environment for individual provinces



2005 2010

Fig. 2. Moran scatterplot for indicator *z* due to (A) environmental degradation (B) protection of the environment and for individual provinces

The value of Spearman correlation coefficient was equal to 0.40 for year 2005 and 0.44 in the year 2010 and were not significant. It means that there was no correlation between levels of degradation and protection of the environment.

The values of global Moran's *I* for environmental degradation was equal to 0.29 for the year 2005 and 0.32 for the year 2010 and proved significant positive spatial autocorrelation (α=0.05). The values of global Moran's *I* for protection of the environment was equal to 0.11 for the year 2005 and 0.12 for the year 2010 and were insignificant (α=0.05).

In figure 2 the Moran scatterplot for this data is presented, with a linear smoother superimposed. Positive significant Moran's I was obtained for environmental degradation. Almost all of the associations fall in the lower left and upper right quadrants indicating presence of clusters. Points in the upper right quadrant corresponding to cluster of high values and its similar neighbors

(Pomorskie, Warmińsko-Mazurskie, Mazowieckie, Lubelskie, Podlaskie, Podkarpackie, Kujawsko-Pomorskie). These provinces had not polluted environment. Points in the lower left quadrant corresponding to Śląskie , Łódzkie i Opolskie . It is the cluster of the most polluted provinces. The value of the Moran's *I* was not significant for the level of protection of the environment. It means that points are evenly situated in all four quadrants.

4. Conclusions

The performed analyses showed regional differentiation of environment in Poland. The actual state of ecosystem is strongly connected with localization and concentration of industry in a given period. It is related to the ongoing processes of industrialization and urbanization. The northern and eastern regions of Poland are scarcely urbanized, not very industrialized and do not have big cities and that is why they relatively do not have degraded natural environment. Efforts aimed at protection of the environment were carried out in the most polluted provinces and along the line from Pomorskie to Podlaskie.

Additionally, the comparative analysis was broadened by the use of the tools of spatial autocorrelation statistics which enabled to reveal the existing spatial relations. Presence of positive spatial autocorrelation for environment degradation was shown. It means the trend of clustering among Poland's provinces due to level of environmental degradation. There was no spatial correlation for protection of environment.

References

Anselin L. (1993). The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. *Research Paper* 9330.

Anselin L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis* 27, 93–115.

Cliff A., Ord J. K. (1981). *Spatial Process: Models and Applications.* Pion, London.

Hwang C. L., Yoon K. (1981). *Multiple attribute decision making. Methods and applications*. Springer, Berlin.

Wysocki F. (2008). Zastosowanie metody TOPSIS do oceny regionalnego zróżnicowania poziomu rozwoju sektora mleczarskiego. *Wiadomości Statystyczne* 1, 38–49.

Wysocki F., Lira J. (2005). *Statystyka opisowa*. Wyd. AR, Poznań.