Colloquium Biometricum 39
2009, 33–44

EVALUATION OF REGIONAL VARIABILITY AND SPATIAL CORRELATIONS OF YIELDS FOR MAIN CROPS AND RATES OF FERTILIZATION

Zbigniew Laudański[1], Małgorzata Flaszka [2], Dariusz Gozdowski[3]

[1]Department of Seed Science and Technology, Plant Breeding and Acclimatization Institute
Radzików, 05-870 Błonie
[2]Department of Biometry, Warsaw University of Life Sciences,
Nowoursynowska 159, 02-776 Warszawa
[3]Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warszawa
e-mails: zbigniew_laudanski@sggw.pl; malgorzata_flaszka@sggw.pl;
dariusz_gozdowski@sggw.pl

Summary

This work presents evaluation of spatial correlations between yields of main crops and rates of fertilization in the sixteen voivodships of Poland. The analyses were conducted using data from the years 2000 and 2007. Spatial relationshps were evaluated on the base of coefficients of autocorrelation between quantities of crop yields. Voivodships significantly different from their neighbours were pointed out on the basis of the values of local spatial autocorrelation coefficients. Strong positive autocorrelation between crop yields of winter triticale and potatoes was proved, but the lack of such autocorrelation for yields of rapeseed and spring wheat was observed.

Key words and phrases: spatial autocorrelation, crop yields, fertilization, regional variability

Classification AMS 2000: 62H11

1. Introductiom

Quantity of crop yields can be investigated on a small scale e.g. experimental microplots or individual fields and on large scale e.g. regions or countries. Variability of quantity of crop yields have been different in different spatial scales, this phenomena can be interesting for farmers or economists because it is important for agricultural policy, insurances etc. (Doré 1997, Williams et al. 2008). Usually quantity of crop yield is similar for places which are situated next to each other and differences between yields are higher for distant areas (Ping et al 2004, Wood et al. 2004). Very often it is caused by environmental conditions (soil quality, climate), but it can be also caused by factors controlled by farmers (e.g. fertilization). The so-called first law of geography which was formulated by Tobler (1970) refers to this phenomena. This rule says that objects which are neighbours are more similar than distant objects.

We can measure similarity of objects in geographical space using e.g. geostatistical methods, which take into consideration position of objects (e.g. regions) and the values of variables for these objects. The spatial autocorrelation coefficients are one group of measures of similarity in geographical space. One of them is *Moran's I* coefficient which afford to evaluate if the values of variables are distributed randomly or spatial autocorrelation exists (Moran, 1950). Other related measures of spatial autocorrelation are *Geary's C* and *Ripley's K* (Geary, 1954; Ripley, 1981).

Positive autocorrelation is very common, it proves that similar objects have similar values of variables. Infrequently the negative autocorrelation exists, which is evidence of bigger differences between values for neighbouring objects (e.g. regions).

The level of autocorrelation affords to prove spatial correlations or lack of them. Furthermore, in case of positive autocorrelation we can point out objects significantly different from their neighbours. On the basis of local autocorrelation coefficients (e.g. local *Moran's I*) we can select outliers.

The aim of this work is evaluation of spatial autocorrelations between quantities of crop yields for main crops in voivodships in Poland and autocorrelation for rates of fertilization in these voivodships.

2. Materials and methods

For statistical analyses data for 16 voivodships from Central Statistical Office from years 2000 and 2007 were used (GUS 2008). Variables which characterize agricultural production for the voivodships are presented in Table 1.

Table 1. Variables which were used for analyses.

Numer of variable	Description
Variables characterizing crop yields in dt per 1 ha	
X_1	Winter wheat
X_2	Spring wheat
X_3	Rye
X_4	Winter barley
X_5	Spring barley
X_6	Oat
X_7	Winter triticale
X_8	Spring triticale
X_9	Maize
X_{10}	Potato
X_{11}	Sugar beet
X_{12}	Rapeseed
Variables which characterize fertilization (kg per 1 ha)	
X_{13}	NPK
X_{14}	N – nitrogen
X_{15}	P – phosphorus
X_{16}	K – potassium
X_{17}	CaO – lime

For each variable the *Moran's I* coefficient of spatial autocorrelation was calculated according to formula (2.1) (Mitchell 2005).

$$I = \frac{N\sum_i \sum_j w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{(\sum_i \sum_j w_{ij})\sum_i (x_i - \overline{x})^2} \tag{2.1}$$

where: N – number of observations, x_i – value of variable for i-th location, x_j – value of variable for j-th location, $\overline{x}$ – average value of variable, w_{ij} – weight between locations i and j. These weights for pairs of objects (voivodships) are reciprocals of Euclidean distances between centroids (geometrical centres) of these objects. For large number of observation range of values of *Moran's I* is similar to Pearson coefficient of correlation. Values greater than 0 mean positive autocorrelation, while values less than 0 mean negative autocorrelation. Value equal to 0 or similar to 0 means random spatial distribution.

Testing of significance of autocorrelation was based on empirical value of the Z statistic which follows normal distribution Z~N(0, 1) (Mitchell, 2005).

$$Z(I) = \frac{I - E(I)}{S_{E(I)}} \tag{2.2}$$

where $E(I)$ is expected value of *Moran's I* coefficient of autocorrelation (Moran, 1950) and $S_{E(I)}$ is standard deviation of autocorrelation coefficient.

$$E(I) = \frac{-1}{N-1} \tag{2.3}$$

Simplified example of calculation of *Moran's I* coefficient is presented below. The variable which was used for the calculation was grain yield of winter wheat from year 2000 for 4 voivodships: (x_1 – Lubuskie, x_2 – Pomorskie, x_3 – Wielkopolskie, x_4- Zachodniopomorskie).

Table 2 presents Euclidean distances between geometrical centres (centroids) of pairs of voivodships, subsequent tables (Tab. 3 and Tab. 4) present further stages of I coefficient calculations using formula (2.1).

Table 2. Distances (in km) between centroids of voivodships (below main diagonal) and theirs reciprocals (above main diagonal).

	x_1	x_2	x_3	x_4
x_1	×	0.003569	0.007637	0.006442
x_2	280	×	0.004791	0.005791
x_3	131	209	×	0.005523
x_4	155	173	181	×

The sum of the weight is equal to $\sum_i \sum_j w_{ij}$=0,033754 on the base of data presented in Table 2

Table 3. Values of grain yield of winter wheat for 4 voivodships , deviations form the mean value and squares of these values.

	Grain yield (t per ha)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
x_1	30.0	6.6	42.9
x_2	41.1	-4.6	20.7
x_3	37.4	-0.9	0.7
x_4	37.7	-1.2	1.3
mean	36.6		$\sum_i (x_i - \bar{x})^2$ =65.7

Calculated value of denominator of the formula (2.1) is equal to $(\sum_i \sum_j w_{ij}) \sum_i (x_i - \bar{x})^2$ = 2.215956

Table 4. Products of deviations from the mean for pairs of voivodships (*i*, *j*) and their weighed values

i	*j*	$(x_i - \bar{x})(x_j - \bar{x})$	$w_{i,j}(x_i - \bar{x})(x_j - \bar{x})$
1	2	–29.80	–0.10637
1	3	–5.57	–0.04252
1	4	–7.53	–0.04853
2	3	3.87	0.01853
2	4	5.23	0.03030
3	4	0.98	0.00540
		$\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})$ = –0.14318	

Value of numerator of the formula (2.1) is equal to $N \sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})$ = –0.57271, and the value of autocorrelation coefficient is equal to *I* = –0,258.
It is important to notice that in this simplified example number of observations is very small (*N*=4), because of this expected value of *I* is equal to –0.33.
The principal component analysis (PCA) on the all 17 variables was used (Filipiak i Wilkos 1998) for multivariate evaluation of autocorrelation and *Moran's I* was calculated using values of the first principal component (PC1). Moreover the values of the first principal component (PC1) were used for calculation of local *Moran's I* coefficient (Anselin 1995).

On the base of local *Moran's I* (formula 2.4) the voivodships which are significantly different than their neigbours were detected (standardized *Z* value of *I* was used for this purpose).

$$I_i = \frac{(x_i - \bar{x})\sum_{i=1}^{n} w_{ij}(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2 \frac{1}{N}} \tag{2.4}$$

The calculations of spatial autocorrelation coefficients were conducted using ArcGIS 9.3 (Mitchell, 2005), while principal component analysis was conducted using Statistica 7.1 (Dobosz, 2004).

3. Results

The values of autocorrelation coefficients prove significant spatial correlations (Tab. 5).

In both years (2000 and 2007) positive significant autocorrelations were obtained for yield of winter triticale (respectively in 2000 and 2007 they were equal to 0.375 and 0.463), yields of potato (respectively 0.597 and 0.493) and nitrogen fertilization (0.416 and 0.530). Voivodships with the highest yields and with the highest N fertilization were along the line from Pomorskie voivodship to Opolskie voivodship (GUS 2008). Decrease of yields and decrease of N fertilization were observed in the northern-east direction (Lubuskie and Zachodniopomorskie voivodships) and in the eastern direction from this line. Quite strong positive autocorrelations were observed for yields of maize and barley, they were statistically significant in the year 2007 and very near of significance level (α=0.05) in 2000.

The lowest values of autocorrelations were observed for yields of spring wheat and rapeseed in both years. Spatial distribution of voivodships with high yields of these crops was random, moreover voivodshiops with high yields of these two crops were different in 2000 comparing with 2007.

Values of spatial autocorrelation coefficients were lower in 2000 year than in 2007. The reason of this phenomena could be drought in 2000, especially it was visible in voivodships Wielkopolskie and Kujawsko-pomorskie which are located along the line of voivodships with usually high yields.

Table 5. Values of *Moran's I* coefficients, *Z* statistics and *p*-values for year 2000 and 2007

	2000			2007		
	Moran's I	*Z*	*p*-value	*Moran's I*	*Z*	*p*-value
X_1	0.243	1.501	0.133	0.329	1.943	0.052
X_2	0.105	0.872	0.383	0.080	0.734	0.463
X_3	0.327	1.959	0.050	0.226	1.520	0.129
X_4	0.297	1.814	0.070	0.415	2.338	0.019
X_5	0.277	1.706	0.088	0.048	0.058	0.561
X_6	0.274	1.690	0.091	0.136	1.051	0.293
X_7	0.375	2.117	0.034	0.463	2.572	0.010
X_8	0.164	1.191	0.234	–0.168	–0.495	0.621
X_9	0.312	1.843	0.065	0.364	2.087	0.037
X_{10}	0.597	3.225	0.001	0.493	2.703	0.007
X_{11}	0.354	2.064	0.039	0.161	1.126	0.260
X_{12}	–0.014	0.253	0.800	–0.034	0.155	0.877
X_{13}	0.269	1.628	0.103	0.540	2.981	0.003
X_{14}	0.416	2.311	0.021	0.530	2.917	0.004
X_{15}	–0.190	–0.721	0.471	0.432	2.434	0.015
X_{16}	–0.102	–0.182	0.855	0.556	3.042	0.002
X_{17}	0.168	1.132	0.257	0.509	2.997	0.003
PC1	0.286	1.741	0.082	0.449	2.540	0.011
PC2	0.359	2.097	0.036	0.308	1.827	0.068

The principal component analysis was conducted for multivariate evaluation of autocorrelation and *Moran's I* were calculated on the base of the values of first principal component (PC1). Such a way of evaluation of spatial autocorrelation seems to be right approach because of share of the first principal component in total variability equal to c.a. 60%. In the year 2000 this share was equal to 58.0% and in 2007 it was equal to 59.6%. This large share of the first principal component was caused by strong linear correlations between values of different crop yields and their relationships with N, P and K fertilization. Strong correlation with the first principal component (PC1) means positive correlation with crop yields and with rates of fertilization.

We can assume that the higher value of PC1 is favourable. The values of the first and the second principal components are presented in Table 6. Moreover, the values of the first principal component are presented on the maps in Figure 1A. In both years the highest values of the first principal component were observed for voivodships in the western part of Poland with the exception of Zachodniopomorskie and Lubuskie voivodships .

In the year 2007 *Moran's I* coefficient points out significant spatial autocorrelation (p=0.011), while in the year 2000 p-value for coefficient of autocorrelation was near to significance level (p=0.082).

Table 6. The values of correlation coefficients between all examined variables and the values of the first and the second principal component (PC1 and PC2).

	2000		2007	
	PC1	PC2	PC1	PC2
X_1	0.89	0.32	0.93	0.09
X_2	0.90	–0.23	0.83	0.30
X_3	0.90	–0.29	0.75	0.55
X_4	0.98	–0.12	0.92	0.19
X_5	0.83	–0.43	0.83	0.39
X_6	0.88	–0.38	0.80	0.49
X_7	0.86	0.22	0.92	–0.03
X_8	0.81	–0.38	0.78	0.29
X_9	0.74	–0.33	0.51	–0.03
X_{10}	0.67	0.35	0.81	–0.29
X_{11}	0.76	–0.15	0.42	0.54
X_{12}	0.58	0.00	0.40	–0.02
X_{13}	0.55	0.80	0.79	–0.59
X_{14}	0.41	0.73	0.77	–0.57
X_{15}	0.38	0.60	0.78	–0.54
X_{16}	0.64	0.64	0.77	–0.60
X_{17}	0.83	–0.15	0.85	–0.02

The local *Moran's I* coefficients based on the value of the first principal component (PC1) were calculated to detect the voivodships , which differ significantly in terms of most of the variables (yields and rates of fertilizers) from neighbouring voivodships . The values of these coefficients below zero mean negative autocorrelation. The *Z* standardized values for local autocorrelation coefficients were calculated in order to easily point out outliers which voivodships differ significantly form their neigbours,. These values are presented on the map at the figure 1B. the voivodships marked by dark colour have lower values of *Z* statistic. Lubuskie voivodship has the lowest value of Z in year 2000 and in 2007. It is because of much lower values of the first principal component than neigbouring voivodships.

Table 7. The values of the first (PC1) and the second (PC2) principal component and *Z* statistic for local values of *Moran's I*.

	PC1 (58,0%)	PC2 (17,7%)	*Z*	PC1 (59,6%)	PC2 (15,4%)	*Z*
Dolnośląskie	4.424	–1.716	2.497	2.58	0.16	1.846
Kujawsko-pomorskie	–0.058	1.302	0.084	3.78	–1.41	1.461
Lubelskie	–1.183	–0.712	0.293	–2.62	–1.03	1.041
Lubuskie	–2.660	0.808	–1.309	–1.55	–2.21	–0.592
Łódzkie	–2.325	–0.702	–0.378	1.21	–1.47	0.240
Małopolskie	0.699	–1.878	0.044	–2.52	3.34	1.066
Mazowieckie	–2.015	–0.001	0.575	–2.84	–1.86	–0.288
Opolskie	6.932	0.921	2.916	7.31	2.27	1.964
Podkarpackie	–1.748	–2.562	0.569	–3.54	2.43	2.025
Podlaskie	–5.232	1.629	0.878	–3.75	0.13	0.359
Pomorskie	1.408	3.882	0.062	0.93	–0.14	0.439
Śląskie	2.859	–2.192	1.043	0.50	0.92	0.300
Świętokrzyskie	–3.258	–0.926	0.222	–3.14	–0.12	1.005
Warmińsko-mazurskie	–1.434	–0.485	0.878	–0.74	–0.27	0.359
Wielkopolskie	2.842	0.682	0.347	3.70	–1.41	1.301
Zachodniopomorskie	0.748	1.951	–0.143	0.70	0.66	–0.043

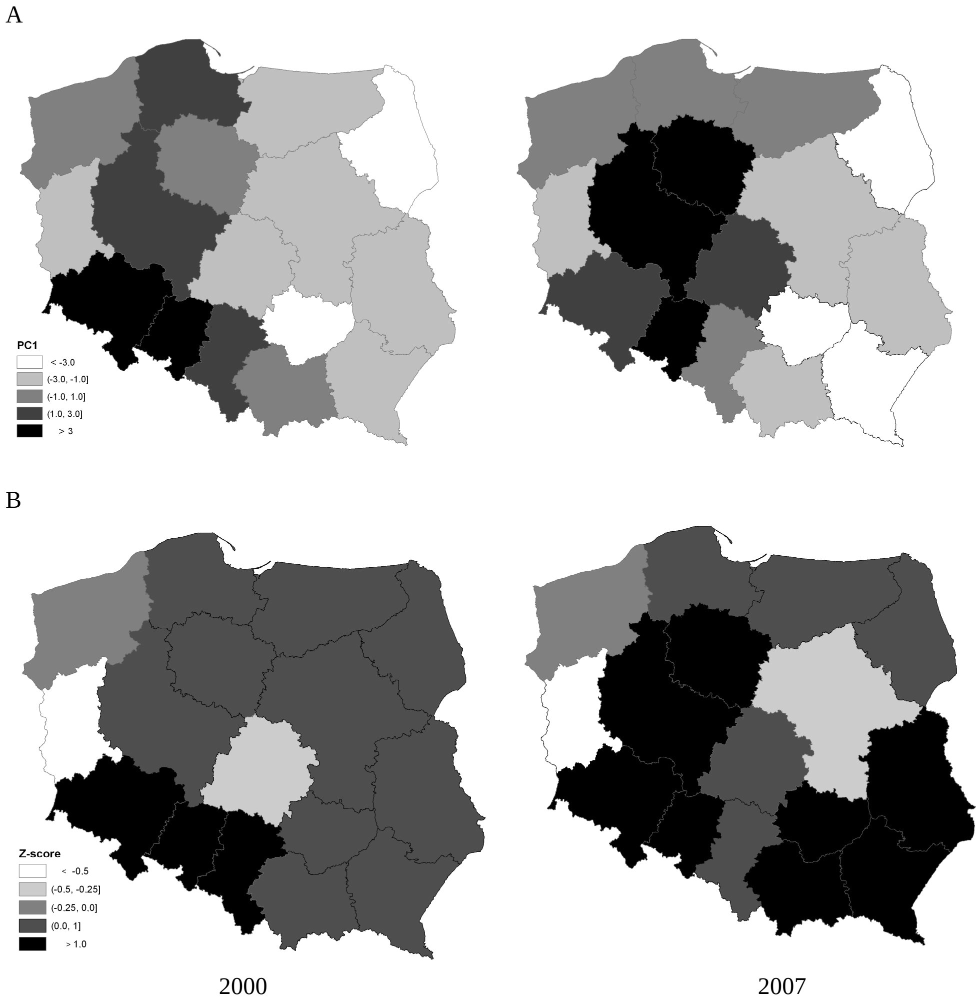

Fig. 1. The values of the first principal component (A) and values of *Z* statistic for local *Moran's I* coefficient (B) calculated on the values of PC1.

4. Conclusions

The results obtained demonstrate presence of spatial autocorrelation for yields of most crops. Moreover the analyses proved that in case of some of crops such as rapeseed and spring wheat autocorrelation does not exist. The results were slightly different in each year; it may indicate influence of variable weather conditions on different yield spatial variability (Igras and Lipiński, 2006). Strong positive spatial autocorrelation for some of crops (in our case potato and winter triticale) can be a factor which afford e.g. prediction of yield quantity of these crops (Priya i Shibasaki 2001). The analyses should be extend to find out the reasons why the outliers (selected voivodships) are not similar to their neighbours in terms of yield quantity. In our survey the most visible outlier was Lubuskie voivodship, in which the yields of most crops were much lower than in neigbouring voivodships .

References

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

Dobosz M. (2004). *Wspomagana komputerowo statystyczna analiza wyników badań*. Wyd. II. Wydawnictwo Exit.

Doré T., Sebillotte M., Meynard J. M. (1997). A diagnostic method for assessing regional variations in crop yield. *Agricultural Systems* 54, 169–188.

Filipiak K., Wilkos S. (1998). *Wybrane metody analizy wielozmiennej i ich zastosowanie w badaniach przestrzennych*. IUNG Puławy, R 349.

Geary R. (1954). The contiguity ratio and statistical mapping. *Incorporated Statistician* 5, 115–145.

GUS (2008). http://www.stat.gov.pl

Igras J., Lipiński W. (2006). Ocena wybranych elementów stanu żyzności gleby i jakości płytkich wód gruntowych na tle intensywności produkcji roślinnej w ujęciu regionalnym. *Pamiętnik Puławski* 142, 147–162.

Mitchell A. (2005). *The ESRI guide to GIS analysis. Vol.2 Spatial measurements and statistics.* ESRI Press, Redlands, California.

Moran P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika* 37, 17–23.

Ping J.L., Green C.J., Zartman R.E., Bronson K.F. (2004). Exploring spatial dependence of cotton yield using global and local autocorrelation statistics. *Field Crops Research* 89, 219–236.

Priya S., Shibasaki R. (2001). National spatial crop yield simulation using GIS-based crop production model. *Ecological Modelling* 135, 113–129.

Ripley B.D. (1981). *Spatial Statistics*. Wiley-Interscience, pp. 252.

Tobler W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, 234–240.

Williams C.L., Liebman M., Edwards J.W., James D.E., Singer J.W., Arritt R., Herzmann D. (2008). Patterns of regional yield stability in association with regional environmental characteristics. *Crop Science*, 48, 1545–1559.

Wood S., You L., Zhang X. (2004). Spatial patterns of crop yields in Latin America and the Caribbean. *Cuadernos De Economía*, 41, 361–381.

OCENA REGIONALNEJ ZMIENNOŚCI I ZALEŻNOŚCI PRZESTRZENNYCH PLONÓW PODSTAWOWYCH ROŚLIN UPRAWNYCH I WYSOKOŚCI NAWOŻENIA

Streszczenie

W pracy przedstawiono ocenę zależności przestrzennych pomiędzy plonami podstawowych gatunków roślin uprawnych oraz nawożeniem w poszczególnych regionach Polski. Analizy zostały wykonane z wykorzystaniem danych dla województw z roku 2000 i 2007. Na podstawie wartości globalnych i lokalnych współczynników autokorelacji przestrzennej określono siłę związków przestrzennych pomiędzy wielkością plonów oraz określono województwa odbiegające pod względem badanych zmiennych od sąsiadów. Stwierdzono silną dodatnią autokorelację przestrzenną pomiędzy plonami pszenżyta ozimego oraz ziemniaków, natomiast brak tej autokorelacji dla rzepaku i pszenicy jarej.

Słowa kluczowe: autokorelacja przestrzenna, plony roślin, nawożenie, zmienność regionalna

Klasyfikacja AMS 2000: 62H11