

STUDIA METODOLOGICZNE

Jadwiga BOŻEK

Klasyfikacja podregionów pod względem podobieństwa struktury agrarnej

Klasyfikacja obiektów jest jedną z podstawowych metod badania otaczającej nas rzeczywistości. Umożliwia ona porządkowanie zgromadzonych informacji, a uzyskane wyniki mogą stanowić podstawę do dalszych analiz (typologii czy prognozowania). Grupowanie jednostek terytorialnych (krajów, województw, podregionów) o podobnej strukturze agrarnej daje obraz jej zróżnicowania przestrzennego, a średnie wartości wskaźników w grupach wyznaczają typy badanej struktury (J. Bożek, 2010; J. Bożek, B. Bożek 2011). Metod grupowania jest bardzo wiele, ale na obecnym etapie badań nie ma metody obiektywnej, której wynik byłby niezależny od autora. Różne metody stosowane do tego samego materiału statystycznego dają rozmaite wyniki pod względem liczby grup, a także ich składu (Wysocki, Wagner, 1989; Putek-Szeląg, Aranowski, 2005; Bogocz i in., 2010). Wyniki grupowania nie są więc jednoznaczne i zależą w dużej mierze od wyboru przez autora metody klasyfikacji. Celem artykułu jest porównanie przestrzennego zróżnicowania struktury agrarnej Polski w podregionach, otrzymanego po zastosowaniu dwóch metod klasyfikacji — eliminacji wektorów i zbiorów rozmytych.

Obliczenia przeprowadzono na podstawie danych ostatniego Powszechnego Spisu Rolnego (PSR 2010) — liczby gospodarstw rolnych według grup obsza-

rowych w podregionach. Uwzględniono następujące grupy obszarowe gospodarstw: do 1 ha włącznie, powyżej 1 do mniej niż 5 ha, od 5 do mniej niż 10 ha, od 10 do mniej niż 15 ha, 15 ha i więcej.

METODA BADAWCZA

Metoda eliminacji wektorów

Strukturę agrarną w podregionach można przedstawić w postaci macierzy $\mathbf{A}$, której wiersze odpowiadają poszczególnym podregionom (obiektom), zaś kolumny — grupom obszarowym gospodarstw (składnikom struktury). Macierz $\mathbf{A}$ ma zatem postać:

$$\mathbf{A} = [a_{ik}] \quad (i = 1, 2, \dots, n; \quad k = 1, 2, \dots, r)$$

gdzie:

a_{ik} — udział liczby gospodarstw z k -tej grupy obszarowej w liczbie wszystkich gospodarstw w i -tym podregionie,

n — liczba podregionów,

r — liczba grup obszarowych gospodarstw.

Przy czym spełnione są następujące warunki:

$$0 \leq a_{ik} \leq 1 \quad (i = 1, \dots, n; \quad k = 1, \dots, r)$$

$$\sum_{k=1}^r a_{ik} = 1 \quad (i = 1, \dots, n)$$

Dla każdej pary podregionów i, j została obliczona miara zróżnicowania ε_{ij} według wzoru (Kukuła, 1989):

$$\varepsilon_{ij} = \frac{\sum_{k=1}^r |a_{ik} - a_{jk}|}{2} \quad (i, j = 1, \dots, n) \quad (1)$$

Elementy ε_{ij} tworzą macierz $\mathbf{E}_0$, która jest macierzą symetryczną z zerami na przekątnej. Następnie obliczono wartość średniego zróżnicowania między obiektami $\bar{\varepsilon}$:

$$\bar{\varepsilon} = \frac{\sum_{i=1}^n \sum_{j=1}^n \varepsilon_{ij}}{n(n-1)} \quad (2)$$

W kolejnym kroku, korzystając z wartości elementów macierzy E_0 i współczynnika $\beta = \bar{\varepsilon}$, tworzy się zero-jedynkową macierz E , której elementy e_{ij} są zdefiniowane następująco:

$$e_{ij} = \begin{cases} 0, & \text{gdy } \varepsilon_{ij} < \beta \\ 1, & \text{gdy } \varepsilon_{ij} \geq \beta \end{cases} \quad (i, j = 1, \dots, n) \quad (3)$$

Do utworzonej macierzy zastosowano metodę eliminacji wektorów (Chomętowski, Sokołowski, 1978), która sprowadza się do następującego algorytmu:

- 1) obliczamy sumę elementów w każdym wierszu macierzy E ;
- 2) w otrzymanym w ten sposób wektorze p wyszukujemy składową p_m przyjmującą największą wartość. Składowa ta wskazuje obiekt, który jest najbardziej niepodobny do pozostałych na poziomie β . Jeżeli składowa p_m jest większa od zera, to należy przejść do punktu 3. Jeżeli składowa p_m jest równa zeru, to należy przerywać pierwszy krok iteracji. Oznacza to, że jednostki odpowiadające wierszom, które pozostały w macierzy E (dla których wszystkie składowe są równe 0) tworzą pierwszą grupę obiektów jednorodnych. Z pozostałych jednostek tworzy się znowu macierz E i przechodzimy do punktu 1;
- 3) następnie w macierzy E wykreślamy m -ty wiersz i m -tą kolumnę;
- 4) przechodzimy do punktu 1.

Postępowanie kończy się, gdy wszystkie jednostki zostaną pogrupowane.

W metodzie tej wynik podziału zależy od wartości β — im mniejsza wartość β , tym (generalnie) większa liczba grup typologicznych. Większa liczba grup daje wprawdzie gwarancję większej jednorodności w grupach, ale jednocześnie zbyt duża liczba grup zaciera różnice między nimi, co w konsekwencji zaciemnia obraz sytuacji. Powstaje więc pytanie, przy jakiej wartości β klasyfikacja jest optymalna.

Jedną z metod wyboru wartości β jest porównanie wariancji wewnątrzgrupowych z wariancją całkowitą dla poszczególnych składników struktury (Wysocki, Wagner, 1989), którymi w tym przypadku są grupy obszarowe gospodarstw. Wariancje te obliczamy według podziałów (grupowania) otrzymanych przy różnych wartościach β dla każdego składnika struktury z osobna i na tej podstawie przy pomocy odpowiednio skonstruowanej funkcji F wyznaczana jest najlepsza wartość β , zwana wartością progową. Metoda ta wymaga zatem wielokrotnego przeprowadzania podziału przy różnych wartościach β_l .

Wartości te wybieramy z przedziału liczbowego $[a, b]$, gdzie: $a = \bar{\varepsilon} - s_{\varepsilon}$, $b = \bar{\varepsilon}$, przy czym $\bar{\varepsilon}$ — średnia arytmetyczna z ε_{ij} , ($i, j = 1, 2, \dots, n$), s_{ε} — odchylenie standardowe z ε_{ij} , ($i, j = 1, 2, \dots, n$), $\beta_l = a + (l-1)h$, gdzie $l = 1, 2, \dots, L$, $\beta_L = b$, h — krok.

Dla każdego grupowania oblicza się wartość funkcji jakości klasyfikacji F :

$$F^{(l)} = \sum_{k=1}^r F_{lk} \quad l=1, \dots, L \quad (4)$$

gdzie:

$F^{(l)}$ — wartość funkcji jakości klasyfikacji dla l -tego grupowania,

r — liczba składników struktury,

F_{lk} — wskaźnik jakości pogrupowania k -tego składnika struktury w l -tym grupowaniu:

$$F_{lk} = \frac{s_{k(o)}^2 / (n-1)}{s_{k(w)}^2 / (n-m-1)} \quad (5)$$

gdzie:

m — liczba wydzielonych grup przy danym β_l ,

$s_{k(o)}^2$ — wariancja ogólna k -tego składnika struktury,

$s_{k(w)}^2$ — wariancja wewnątrzgrupowa k -tego składnika struktury.

Optymalnym podziałem jest podział l_0 , przy którym funkcja $F^{(l)}$ przyjmuje największą wartość $F^{(l_0)} = \max \{F^{(1)}, \dots, F^{(L)}\}$. Odpowiadająca temu podziałowi wartość $\beta = \beta_{l_0}$ jest szukaną wartością progową.

Klasyfikacja rozmyta

Drugą metodą zastosowaną do grupowania podregionów o podobnej strukturze agrarnej jest klasyfikacja rozmyta, przekształcona w klasyfikację klasyczną. W odróżnieniu od klasyfikacji klasycznej, w której przynależność obiektów do danej klasy opisywana jest za pomocą zmiennej zero-jedynkowej, w klasyfikacji rozmytej przynależność obiektu do klasy opisywana jest za pomocą zmiennej ciągłej. Są to tzw. funkcje przynależności, które przyjmują wartości z przedziału $[0, 1]$.

W przypadku klasyfikacji rozmytej zakłada się, że dany jest zbiór Ω , liczący n obiektów (w tym przypadku podregionów) — $P_1, P_2, \dots, P_n$. Obiekty te opisane są przez wartości r zmiennych — $X_1, X_2, \dots, X_r$ (w artykule X_l oznacza udział liczby gospodarstw z l -tej grupy obszarowej w ogólnej liczbie gospodarstw

w danym podregionie). Na zbiorze Ω należy tak określić rodzinę klas rozmytych $S_1, S_2, \dots, S_K$ ($1 < K < n$), aby spełnione były warunki:

1) $0 \leq f_{S_j}(P_i) \leq 1$ ($i = 1, \dots, n$; $j = 1, \dots, K$), gdzie: $f_{S_j}(P_i)$ oznacza stopień przynależności obiektu P_i do klasy S_j ;

2) $\sum_{j=1}^K f_{S_j}(P_i) = 1$ ($i = 1, \dots, n$);

3) obiekty, dla których stopnie przynależności do tej samej klasy są duże — są bardzo podobne, natomiast obiekty, dla których stopnie przynależności do różnych klas są duże — są mało podobne.

Utworzenie klasyfikacji rozmytej polega zatem na wyznaczeniu dla każdego obiektu $P_i \in \Omega$ takiego wektora $f(P_i) = (f_{S_1}(P_i), f_{S_2}(P_i), \dots, f_{S_K}(P_i))$, że spełnione są warunki 1—3.

Istnieje kilka metod tworzenia klasyfikacji rozmytej (Jajuga, 1984). W opracowaniu zastosowano metodę iteracyjną, wykorzystującą pojęcie rozmytego środka ciężkości. W metodzie tej w kolejnych iteracjach dokonuje się zmiany wartości stopni przynależności obiektów do poszczególnych klas. Procedura ta kontynuowana jest aż do momentu, gdy te wartości przestaną się zmieniać w stopniu znaczącym.

Tak więc, wychodząc od rozmytej klasyfikacji początkowej $f_{S_j}^0(P_i)$ ($i = 1, \dots, n$; $j = 1, \dots, K$) w ν -tym kroku iteracyjnym, wyznacza się:

a) dla każdej klasy rozmytej środki ciężkości poszczególnych zmiennych X_l według wzoru:

$$g_{jl}^{\nu} = \frac{\sum_{i=1}^n (f_{S_j}^{\nu-1}(P_i))^2 x_{il}}{\sum_{i=1}^n (f_{S_j}^{\nu-1}(P_i))^2} \quad (j = 1, \dots, K; \quad l = 1, \dots, r) \quad (6)$$

gdzie:

g_{jl}^{ν} — wartość l -tej zmiennej dla rozmytego środka ciężkości j -tej klasy, wyznaczona w ν -tej iteracji,

$x_{il} = X_l(P_i)$ — wartość l -tej zmiennej w i -tym obiekcie;

b) dla każdego obiektu odległości od rozmytych środków ciężkości klas według wzoru:

$$d_{ij}^{\nu} = \sum_{l=1}^r (x_{il} - g_{jl}^{\nu})^2 \quad (i = 1, \dots, n; \quad j = 1, \dots, K) \quad (7)$$

gdzie d_{ij}^v — odległość i -tego obiektu od środka ciężkości j -tej klasy, wyznaczona w v -tej iteracji;

c) nowe stopnie przynależności według wzoru dla każdego obiektu P_i :

gdy $d_{ij}^v \neq 0$ ($j = 1, \dots, K$), wtedy:

$$f_{S_j}^v(P_i) = \frac{(d_{ij}^v)^{-1}}{\sum_{l=1}^K (d_{il}^v)^{-1}} \quad (i = 1, \dots, n; \quad j = 1, \dots, K) \quad (8)$$

gdy $d_{ij}^v = 0$ dla pewnego t , wtedy:

$$f_{S_j}^v(P_i) = \begin{cases} 1, & \text{gdy } j = t \\ 0, & \text{gdy } j \neq t \end{cases} \quad (9)$$

Procedurę opisaną w punktach a), b), c) kontynuujemy, aż spełniony zostanie warunek $\max_{i,j} |f_{S_j}^v(P_i) - f_{S_j}^{v-1}(P_i)| < \varepsilon$, gdzie ε jest liczbą dodatnią bliską 0.

Przyjmuje się wówczas, że ostateczną klasyfikacją rozmytą jest klasyfikacja otrzymana w iteracji v -tej.

Następnie otrzymaną klasyfikację rozmytą przekształcono w klasyfikację klasyczną, przyjmując że obiekt P_i należy do klasy (grupy typologicznej) S_j , gdy:

$$f_{S_j}(P_i) = \max_l f_{S_l}(P_i) \quad \text{ i } \quad f_{S_j}(P_i) \geq 0,4 \quad (10)$$

WYNIKI BADAŃ

Według przedstawionej metody eliminacji wektorów przeprowadzono grupowanie podregionów Polski wykorzystując dane PSR 2010. Średnie zróżnicowanie strukturalne obliczone według wzoru (1) wynosi 0,2527. Grupowanie przeprowadzono według różnej wartości progowej. W tabl. 1 podano wyniki otrzymane dla następujących wartości: $\beta_1 = 0,2527$, $\beta_2 = 0,24$, $\beta_3 = 0,22$, $\beta_4 = 0,20$. W przypadku każdego grupowania obliczono wartość funkcji jakości klasyfikacji F według wzoru (4).

TABL. 1. GRUPY PODREGIONÓW O PODOBNEJ STRUKTURZE AGRARNEJ W 2010 R. WYODRĘBNIŁONE METODĄ ELIMINACJI WEKTORÓW I METODĄ KLASYFIKACJI ROZMYTEJ. WARTOŚCI FUNKCJI JAKOŚCI KLASYFIKACJI *F*

Grupy Wartość funkcji <i>F</i>	Metoda eliminacji wektorów według wartości progowej zróżnicowania β				Metoda zbiorów rozmytych — klasyfikacja	
	0,2527	0,24	0,22	0,20	5-klasowa	6-klasowa
	skład grup (nr podregionów) ^a					
I	1, 2, 3, 4, 6, 9, 13, 14, 31, 32, 37, 40, 41, 42, 45, 54, 55, 56, 57, 58, 59, 60, 61, 63, 64, 66	2, 4, 6, 7, 8, 9, 31, 32, 37, 40, 41, 42, 54, 55, 56, 57, 58, 59, 60, 61, 63, 64, 66	6, 7, 8, 31, 37, 38, 40, 41, 42, 54, 55, 56, 58, 59, 60, 61, 63, 64, 66	6, 7, 8, 31, 37, 40, 41, 42, 54, 56, 58, 59, 60, 61, 63, 64, 66	21, 43, 44, 47, 48, 49, 50, 51, 62, 65	21, 43, 44, 47, 48, 49, 50, 51, 65
II	5, 20, 21, 23, 33, 34, 43, 44, 47, 49, 50, 51, 62, 65	1, 3, 5, 14, 20, 22, 23, 24, 33, 34, 35, 36, 45, 46, 50, 51, 52	1, 2, 3, 4, 5, 11, 14, 20, 32, 34, 36, 45, 46, 50, 52	1, 2, 3, 5, 13, 14, 20, 34, 43, 50, 51, 62	10, 12, 15, 17, 18, 27, 28, 29, 30, 53	10, 12, 15, 17, 18, 27, 28, 29, 30, 41, 53
III	10, 11, 12, 16, 17, 27, 28, 29, 30, 36, 46, 52, 53	10, 11, 12, 15, 16, 17, 27, 28, 29, 30, 53	9, 10, 12, 15, 17, 18, 19, 27, 29, 30, 53, 57	9, 10, 11, 12, 15, 17, 18, 29, 30, 53, 57	16, 20, 22, 23, 24, 33, 34, 35, 36, 46, 52	16, 20, 22, 23, 24, 33, 34, 35, 36, 46, 52
IV	7, 8, 15, 18, 19, 25, 26, 38	21, 43, 44, 47, 48, 49, 62, 65	21, 33, 43, 44, 47, 48, 49, 51, 62, 65	22, 24, 33, 35, 36, 45, 46, 52	7, 8, 9, 19, 25, 37, 38, 39, 40, 41, 42, 55, 56, 57, 58, 59, 60, 61, 64	7, 8, 9, 19, 25, 26, 37, 38, 39, 40, 57, 58, 59
V	22, 24, 35	19, 25, 26, 38, 39	16, 22, 23, 24, 35	21, 44, 47, 48, 49, 65	1, 2, 3, 4, 6, 13, 14, 31, 32, 63, 66	1, 2, 3, 4, 6, 13, 14, 31, 32, 62
VI	48	13	25, 39	19, 25, 26, 38	5	42, 54, 55, 56, 60, 61, 63, 64, 66
VII	39	x	28	16, 27, 28	11	5
VIII	x	x	26	39, 55	26	11
IX	x	x	13	4, 32	45	45
X	x	x	x	23	54	x
Wartość funkcji <i>F</i>	17,81	20,99	20,74	24,74	26,38	30,67

^a Podregiony podano zgodnie z numeracją przyjętą przez GUS: 1 — jeleniogórski, 2 — legnicko-głogowski, 3 — waltburski, 4 — wrocławski, 5 — m. Wrocław, 6 — bydgosko-toruński, 7 — grudziądzki, 8 — włocławski, 9 — białski, 10 — chełmsko-zamojski, 11 — Lubelski, 12 — puławski, 13 — gorzowski, 14 — zielonogórski, 15 — łódzki, 16 — m. Łódź, 17 — piotrkowski, 18 — sieradzki, 19 — skierniewicki, 20 — krakowski, 21 — m. Kraków, 22 — nowosądecki, 23 — oświęcimski, 24 — tarnowski, 25 — ciechanowsko-płocki, 26 — ostrołęcko-siedlecki, 27 — radomski, 28 — m. Warszawa, 29 — warszawski wschodni, 30 — warszawski zachodni, 31 — nyski, 32 — opolski, 33 — krosiński, 34 — przemyski, 35 — rzeszowski, 36 — tamobrzeski, 37 — białostocki, 38 — łomżyński, 39 — suwalski, 40 — gdański, 41 — słupski, 42 — starogardzki, 43 — trójmiejski, 44 — bielecki, 45 — bytomski, 46 — częstochowski, 47 — głiwicki, 48 — katowicki, 49 — rybnicki, 50 — sosnowiecki, 51 — tyski, 52 — kielecki, 53 — sandomiersko-jędrzejowski, 54 — ełbski, 55 — ełcki, 56 — olsztyński, 57 — kaliski, 58 — koniński, 59 — leszczyński, 60 — piski, 61 — poznański, 62 — m. Poznań, 63 — koszaliński, 64 — stargardzki, 65 — m. Szczecin, 66 — szczeciński.

Źródło: obliczenia własne na podstawie danych PSR 2010.

Wyniki przeprowadzonego grupowania (tabl. 1) wskazują, że niewielka zmiana wartości progowej zróżnicowania β powoduje duże zmiany w grupowaniu. Zmienia się liczba grup, ich skład, a także rozmieszczenie. Otrzymujemy zatem różne obrazy zróżnicowania przestrzennego badanej struktury w zależności od przyjętej wartości β . W celu porównania przedstawiono graficzny obraz wyodrębnionych grup podregionów według $\beta = 0,2527$ (wykr. 1) i $\beta = 0,20$ (wykr. 2). Średnie wskaźniki struktury w poszczególnych grupach prezentuje tabl. 2. Grupy jednoelementowe, jako elementy „niepasujące” do wyodrębnionych skupisk, podano w tablicy kursywą.

W obydwu grupowaniach uwidoczniło się największe rozdrobnienie gospodarstw w Polsce południowo-wschodniej, a najmniejsze w Polsce północno-zachodniej i zachodniej. Jednakże obydwie grupowania różnią się pod względem liczby grup, ich składu, średnich wskaźników struktury i rozmieszczenia terytorialnego. Wraz ze spadkiem wartości zróżnicowania progowego β rośnie wartość funkcji F . Oznacza to, że otrzymane grupy są bardziej jednorodne, jednak równocześnie zwiększa się liczba grup, przez co otrzymany obraz staje się mniej przejrzysty.

W następnej kolejności przeprowadzono grupowanie podregionów z zastosowaniem metody zbiorów rozmytych. Obliczenia wykonano według klasyfikacji 5- i 6-klasowej¹. Wartości początkowe stopni przynależności do klas rozmytych były ustalane losowo, co nie miało wpływu na klasyfikację końcową. Obliczenia przerywano, gdy maksimum modułu różnicy wartości stopni przynależności w dwóch kolejnych iteracjach było mniejsze od 0,000001. Otrzymałą w ten sposób klasyfikację rozmytą przekształcono w klasyfikację klasyczną według warunku (10) i obliczono wartość funkcji jakości klasyfikacji F według wzoru (4). Wyniki przedstawia tabl. 1. Obiekty (podregiony), dla których stopień przynależności do otrzymanych skupisk był mniejszy od 0,4, wyszczególniono jako grupy jednoelementowe.

Grupowania otrzymane w wyniku zastosowania metody zbiorów rozmytych charakteryzują wyższe wartości funkcji F , a więc większa jednorodność grup, niż grupowania uzyskane poprzednią metodą. Wartość funkcji F jest wyższa dla klasyfikacji 6-klasowej i wynosi 30,67, podczas gdy dla klasyfikacji 5-klasowej wynosi 26,38. Jednakże stosując obie klasyfikacje 5- i 6-klasową otrzymujemy bardzo podobne wyniki — grupa III w klasyfikacji 5-klasowej ma identyczny skład, jak w klasyfikacji 6-klasowej, natomiast grupy I, II i V w obu klasyfikacjach różnią się tylko jednym elementem. Istotna różnica dotyczy grupy IV z klasyfikacji 5-klasowej. Podregiony z tej grupy (w klasyfikacji 5-klasowej) podzielono na dwie grupy w klasyfikacji 6-klasowej — IV i VI. Dlatego też typy struktury wyznaczone przez średnie wskaźniki grup (tabl. 3) w klasyfikacji 5-klasowej pozostają prawie takie same w klasyfikacji 6-klasowej, z wyjątkiem typu generowanego przez IV grupę. Jednak podregiony grupy IV (w klasyfikacji 5-klasowej) charakteryzuje wystarczająco wysokie podobieństwo (współczynnik zmienności w przypadku każdej z grup obszarowych nie przekracza 25%), a obraz podziału jest bardziej czytelny przy mniejszej liczbie klas, dlatego klasyfikację 5-klasową można uznać za najlepszą.

¹ Do tego celu zastosowano autorski program komputerowy.

TABL. 2. ŚREDNIE WSKAŹNIKI STRUKTURY WEDŁUG POSZCZEGÓLNYCH GRUP PODREGIONÓW PRZY RÓŻNYCH WARTOŚCIACH ZRÓŻNICOWANIA PROGOWEGO β W %

β	Grupy	Liczebność grupy	Gospodarstwa o powierzchni użytków rolnych w ha				
			0—1	1—5	5—10	10—15	≥ 15
0,2527	I	26	32,2	27,9	14,4	9,0	16,5
	II	14	55,6	32,7	6,4	2,0	3,2
	III	13	25,9	47,5	17,9	4,9	3,8
	IV	8	17,2	31,1	23,4	12,5	15,7
	V	3	40,5	50,7	7,1	1,0	0,7
	VI	1	71,4	20,4	5,1	1,0	2,1
	VII	1	15,3	19,2	18,5	17,7	29,3
0,24	I	23	29,0	26,9	15,5	10,1	18,4
	II	17	44,3	41,1	8,5	2,5	3,5
	III	12	21,8	45,8	21,1	6,4	4,9
	IV	8	62,6	25,8	5,9	1,9	3,8
	V	5	14,3	28,0	23,7	14,9	19,1
	VI	1	44,1	27,3	10,7	5,7	12,2
0,22	I	19	27,7	25,8	15,5	10,8	20,2
	II	15	41,2	37,6	11,1	4,0	6,1
	III	12	20,4	41,3	23,1	8,4	6,8
	IV	10	60,9	27,7	5,9	2,0	3,5
	V	5	40,5	49,3	7,8	1,4	1,0
	VI	2	13,9	23,9	21,3	16,0	24,9
	VII	1	14,9	59,8	14,8	3,9	6,6
	VIII	1	10,4	33,5	26,6	14,7	14,9
	IX	1	44,1	27,3	10,7	5,7	12,2
0,20	I	17	28,2	26,0	15,3	10,5	19,9
	II	12	47,7	31,9	9,6	4,0	6,8
	III	11	22,2	41,3	22,2	7,9	6,5
	IV	8	40,3	47,5	8,6	1,8	1,9
	V	6	65,2	25,9	4,8	1,4	2,7
	VI	4	14,1	30,2	25,0	14,2	16,5
	VII	3	20,3	53,1	17,7	4,7	4,2
	VIII	2	22,7	20,4	14,8	14,1	28,0
	IX	2	37,2	30,9	13,5	6,5	12,0
	X	1	51,0	44,0	3,7	0,6	0,7

Źródło: obliczenia własne.

**TABL 3. ŚREDNIE WSKAŹNIKI STRUKTURY WEDŁUG POSZCZEGÓLNYCH
GRUP PODREGIONÓW WYODRĘBNIONYCH METODĄ ZBIORÓW ROZMYTYCH W %**

Grupa	Liczebność grupy	Gospodarstwa o powierzchni użytków rolnych				
		0—1 ha	1—5	5—10	10—15	15 ha i więcej

Klasyfikacja:

5-klasowa

I	10	61,0	27,1	6,2	2,1	3,6
II	10	20,0	45,9	22,2	6,8	5,2
III	11	41,5	47,4	8,4	1,5	1,3
IV	19	23,7	26,3	17,6	12,1	20,3
V	11	39,1	28,7	12,4	6,5	13,3
VI	I	49,7	35,9	6,1	2,9	5,3
VII	I	32,1	40,0	17,5	5,5	4,8
VIII	I	10,4	33,5	26,6	14,7	14,9
IX	I	34,8	39,8	13,1	4,9	7,4
X	I	32,2	20,9	12,2	11,0	23,7

6-klasowa

I	9	62,3	26,9	5,8	1,9	3,2
II	11	20,1	45,9	22,2	6,8	5,2
III	11	41,5	47,4	8,4	1,5	1,3
IV	13	20,3	30,0	21,7	12,6	15,4
V	10	41,8	29,7	11,8	5,8	10,9
VI	9	29,1	24,3	13,2	10,5	23,0
VII	I	49,7	35,9	6,1	2,9	5,3
VIII	I	32,1	40,0	17,5	5,5	4,8
IX	I	34,8	39,8	13,1	4,9	7,4

Źródło: obliczenia własne.

Z przeprowadzonych badań wynika, że metoda eliminacji wektorów jest mało stabilna — niewielka zmiana β powoduje duże zmiany w wynikach grupowania. Jest to duże utrudnienie w analizie przestrzenno-czasowej badanej struktury, dlatego w badaniach takich autorka stosowała metodę zbiorów rozmytych.

Spośród 6 przedstawionych grupowań najlepsze okazało się grupowanie otrzymane metodą zbiorów rozmytych przy 5 klasach, gdyż spełnia warunek o wysokim stopniu jednorodności grup i istotnym zróżnicowaniu międzygrupowym. Zatem w Polsce w 2010 r. można wyróżnić 5 wieloelementowych grup podregionów podobnych pod względem struktury agrarnej (wykr. 3). Grupy te skupiają 61 podregionów. Pozostałe 5 podregionów odbiega strukturą od wyodrębnionych skupisk, tworząc grupy jednoelementowe.

Największe rozdrobnienie gospodarstw wystąpiło w I grupie obejmującej podregiony: m. Kraków, trójmiejski, bielski, gliwicki, katowicki, rybnicki, sosnowiecki, tyski, m. Poznań, m. Szczecin, w których prawie 90% gospodarstw nie przekraczało 5 ha (średnio 61% stanowiły gospodarstwa o powierzchni użytków rolnych do 1 ha, a ok. 27,1% — gospodarstwa 1—5 ha). Udziały gospodarstw 5—10 ha wynosiły przeciętnie 6,2%, 10—15 ha — 2,1%, a powyżej 15 ha — 3,6%. Równie duże rozdrobnienie cechowało podregiony grupy III (krakowski, nowosądecki, oświęcimski, tarnowski, krośnieński, przemyski, rzeszowski, tarnobrzewski, częstochowski, kielecki, m. Łódź). Tam także prawie 90% gospodarstw nie przekraczało 5 ha (średnio 41,5% stanowiły gospodarstwa do 1 ha, a ok. 47,4% — gospodarstwa 1—5 ha). Udziały kolejnych grup obszarowych wynosiły średnio: 8,4%, 1,5%, 1,3%.

Najmniej rozdrobniona struktura charakteryzowała grupę IV (podregiony: grudziądzki, włocławski, białski, skierniewicki, ciechanowsko-płocki, białostocki, łomżyński, suwalski, gdański, słupski, starogardzki, ełcki, olsztyński, kaliski, koniński, leszczyński, pilski, poznański, stargardzki), gdzie wskaźniki struktury kształtowały się następująco: 23,7%, 26,3%, 17,6%, 12,1%, 20,3%. Występuje tam największy odsetek gospodarstw 10—15 ha (12,1%) i gospodarstw powyżej 15 ha (20,3%). Gospodarstwa mające do 5 ha stanowiły w sumie połowę ogółu wszystkich gospodarstw i był to najniższy odsetek spośród wszystkich grup typologicznych.

W grupie II podregionów (chełmsko-zamojski, puławski, łódzki, piotrkowski, sieradzki, radomski, m. Warszawa, warszawski wschodni, warszawski zachodni, sandomiersko-jędrzejowski) gospodarstwa do 1 ha stanowiły 20%, a gospodarstwa 1—5 ha — 45,9%. Gospodarstw mających 5—10 ha było 22,2% i jest to najwyższy odsetek tych gospodarstw w porównaniu z innymi grupami. Udziały zaś gospodarstw 10—15 ha i powyżej 15 ha wynosiły odpowiednio 6,8% i 5,2%.

W podregionach grupy V (jeleniogórski, legnicko-głogowski, wałbrzyski, wrocławski, bydgosko-toruński, gorzowski, zielonogórski, nyski, opolski, koszański, szczeciński) wskaźniki struktury kształtowały się następująco: 39,1%, 28,7%, 12,4%, 6,5%, 13,3%.

Podsumowanie

1. Wybór metody grupowania obiektów ma znaczący wpływ na wyniki badań — w przypadku klasyfikacji podregionów Polski pod względem podobieństwa struktury agrarnej metodą klasyfikacji rozmytej otrzymujemy inne wyniki niż metodą eliminacji wektorów. Różnice dotyczą liczby grup, ich składu, średnich wskaźników struktury, rozmieszczenia terytorialnego i stopnia jednorodności grup.
2. Metoda eliminacji wektorów jest mało stabilna. Niewielka zmiana wartości progowej zróżnicowania powoduje zwykle istotne zmiany w wynikach gru-

- powania, co np. w przypadku badań przestrzenno-czasowych znacznie utrudnia interpretację. W tego typu badaniach bardziej wskazana jest metoda klasyfikacji rozmytej przekształcona w klasyfikację klasyczną niż metoda eliminacji wektorów.
3. Często w zastosowaniach metody eliminacji wektorów wartość progową ustala się na poziomie średniego zróżnicowania struktur badanych obiektów, otrzymując sensowne wyniki podziału. Jednakże wówczas zróżnicowanie wewnątrzgrupowe jest zwykle wysokie.
 4. W wyniku zastosowania metody zbiorów rozmytych otrzymano grupy o wyższym stopniu jednorodności niż w przypadku metody eliminacji wektorów. Spośród sześciu grupowań przedstawionych w artykule, najlepsze było grupowanie otrzymane metodą zbiorów rozmytych przy 5 klasach, gdyż spełnia warunek o wysokim stopniu jednorodności grup i istotnym zróżnicowaniu międzygrupowym. Oznacza to, że w Polsce w 2010 r. można było wyróżnić 5 grup podregionów o podobnej strukturze agrarnej. Jest to równoznaczne z wyznaczeniem 5, znacznie różniących się między sobą, typów badanej struktury podregionów.

dr Jadwiga Bożek — *Uniwersytet Rolniczy w Krakowie*

LITERATURA

- Bogocz D., Bożek J., Kukuła K., Strojny J. (2010), *Statystyczne studium struktury agrarnej w Polsce*, PWN, Warszawa
- Bożek J. (2010), *Typologia krajów Unii Europejskiej pod względem podobieństwa struktury agrarnej*, „Acta Scientiarum Polonorum, Oeconomia”, nr 9(3)
- Bożek J., Bożek B. (2011), *Typologia struktury agrarnej województw w ujęciu dynamicznym z zastosowaniem klasyfikacji rozmytej. Metody ilościowe w badaniach ekonomicznych*, tom XII, nr 2, Wydawnictwo SGGW
- Chomątowski S., Sokołowski A. (1978), *Taksonomia struktur*, „Przegląd Statystyczny”, nr 2
- Jajuga K. (1984), *Zbiory rozmyte w zagadnieniu klasyfikacji*, „Przegląd Statystyczny”, z. 3/4
- Kukuła K. (1989), *Statystyczna analiza strukturalna i jej zastosowanie w sferze usług produkcyjnych dla rolnictwa*, Zeszyty Naukowe AE w Krakowie, seria specjalna „Monografie”, nr 89, Kraków
- Putek-Szeląg E., Aranowski A. (2005), *Zastosowanie statystycznych metod klasyfikacji danych do ujednoludnienia zbiorowości nieruchomości rolnych na przykładzie powiatu pyrzyckiego*, „Zeszyty Naukowe Uniwersytetu Szczecińskiego”, Prace Katedry Ekonometrii i Statystyki, nr 16
- Wysocki F., Wagner W. (1989), *O ustalaniu wartości progowej zróżnicowania struktur z danych empirycznych*, „Wiadomości Statystyczne”, nr 9

SUMMARY

The purpose of this article is to compare the diversity of the agrarian structure in Polish sub-regions based on two classification methods: the elimination of vectors and the fuzzy set theory. The data used in the calculations by area groups of farms come from the Agricultural Census of 2010. Each method generated a different picture of the spatial diversity of the test structure. Differences related to the number of groups, their composition, average structure indicators, territorial distribution and degree of the group homogeneity.

РЕЗЮМЕ

Целью статьи является сравнение дифференциации аграрной структуры в субрегионах Польши разработанной на основе двух методов классификации: элиминации векторов и нечетких множеств. Для исчислений были использованы данные по группам размеров хозяйств из Всеобщей сельскохозяйственной переписи 2010 г. Каждый из методов создал иное изображение пространственной дифференциации обследуемой структуры. Различия касались числа групп, их состава, средних показателей структуры, территориального размещения и степени однородности групп.