

POZNAN 1997

CZĘŚĆ II

19

MELIORACJE I INŻYNIERIA ŚRODOWISKA



CCXCIV

ROZCZNIKI
AKADEMII ROLNICZEJ
W POZNANIU

JERZY BYKOWSKI, ANTONI MILER

ZASTOSOWANIE METOD GRUPOWAŃ HIERARCHICZNYCH DO CHARAKTERYSTYKI OBIEKTÓW DRENARSKICH W WIELKOPOLSCE

*Z Katedry Melioracji i Kształtowania Środowiska
Akademii Rolniczej w Poznaniu*

ABSTRACT. In the work there has been described characteristic of 584 drainage objects located in 7 provinces of Wielkopolska Region. Each of drainage objects was described by 13 coded parameters. Applied statistics method – cluster analysis (average method) showed that main group contain almost all analysed drainage objects.

Key words: drainage, exploitation and functioning of drainage objects

Wstęp

Jednym z celów działalności poznawczej jest charakterystyka, grupowanie oraz klasyfikacja zjawisk, przedmiotów lub zachowań. Podział jakiegoś zbioru przedmiotów, obiektów lub urządzeń na podzbiory oddaje szczególne usługi w pracach normalizacyjnych. W wyniku obiektywnej analizy struktury badanego zbioru jest możliwa właściwa korekta procesów produkcji, eksploatacji, modernizacji itp. Istotny jest jednak dobór odpowiednich metod grupowań lub klasyfikacji. Problem ten dotyczy także praktyki melioracyjnej. W niniejszej pracy podjęto próbę charakterystyki inwestycji drenarskich, stosując metody grupowań hierarchicznych – metody grupowań obiektów wielocechowych. Metody te, wykorzystywane w naukach ekonomicznych, dzięki przyjętym założeniom wydają się być przydatne do grupowań obiektów drenarskich – opisywanych wieloma charakterystykami niekoniecznie ilościowymi.

Zakres i metodyka badań

W pracy przedstawiono charakterystykę 584 obiektów drenarskich oddanych do eksploatacji w latach 1972-1985 na terenach województw: gorzowskiego, kaliskiego, konińskiego, leszczyńskiego, pilskiego, poznańskiego i zielonogórskiego. Rok 1972 przyjęto jako graniczny, ponieważ w tym roku wykonano technologią bezrowkową pierwszy obiekt na obszarze objętym badaniami. Dane archiwalne do analiz zebrano w latach 1986-1989 w ramach CPBR 10.8.3.1.A.17 (Kosturkiewicz i Bykowski 1989).

Do charakterystyki obiektów drenarskich, dla celów niniejszej pracy, przyjęto następujące cechy: powierzchnia drenowania, układ sieci, technologia drenowania, ukształtowanie terenu, warunki glebowe, stosowane zabezpieczenia rurociągów drenarskich oraz skuteczność funkcjonowania urządzeń. Część cech nie ma charakteru ilościowego, zatem postanowiono cechy opisać w skali kodowej 0, 1, 2, 3, 4 i 5.

Przyjęto następujący system kodowania:

powierzchnia drenowania została podzielona na przedziały, którym przypisano stosowne kody: do 50 ha – „1”, od 51 do 150 ha – „2”, od 151 ha do 400 ha – „3”, od 401 ha do 600 ha – „4”, powyżej 600 ha – „5”,

układ sieci drenarskiej scharakteryzowano przez procentowy udział w powierzchni drenowania sieci systematycznej i niesystematycznej: 0% – „0”, od 1 do 10% – „1”, od 11 do 25% – „2”, od 26 do 50% – „3”, od 51 do 75% – „4”, od 76 do 100% – „5”,

technologię drenowania opisano przez procentowy udział w powierzchni drenowania sączków wykonanych technologią rowkową i bezrowkową: kodowanie wykonano według przedziałów procentowych j.w.,

ukształtowanie terenu oceniono przez procentowy udział w terenie powierzchni o średnich spadkach do 20‰: kodowanie wykonano według przedziałów procentowych j.w.,

warunki glebowe obiektów scharakteryzowano według *Wytucznych drenowania gruntów ornych* (Śniadowski 1978), dzieląc gleby na 4 grupy: gleby zwięzłe (A), gleby zwięzłe na lekkich (B), gleby lekkie na zwięzłych (C) i gleby lekkie (D). Dla każdej grupy określono procentowy udział w drenowanej powierzchni: kodowanie wykonano według przedziałów procentowych j.w.,

zabezpieczenia sieci sączków opisano przez procentowy udział w powierzchni drenowania sączków poddanych słomowaniu i zwirowaniu: kodowanie wykonano według przedziałów procentowych j.w.,

skuteczność funkcjonowania urządzeń (rowy, budowle na rowach, zbieracze, sączki, budowle drenarskie) oceniono przez odnotowane reklamacje dotyczące ich działania: brak reklamacji – „0”, reklamacja – „1”. Ostatecznie po zsumowaniu tych 5 ocen otrzymano wartość kodową (0-5).

Każdy zatem z 584 obiektów drenarskich został w sumie opisany przez 13 parametrów – kodów określających jego wybrane 13 cech. Dla właściwego pogrupowania tak opisanych obiektów drenarskich zastosowano metodę średnich połączeń z grupy metod skupiania SAHN.

Metody analizy skupień. Metody SAHN

Grupowanie obiektów jest w istocie jedną z metod analiz wielowymiarowych, która umożliwi dokonanie abstrakcji, a w konsekwencji odkrycie praw i zbudowanie teorii. Istnieją dwa zasadnicze typy grupowania: metody klasyfikacji i metody skupiania (Mezzich i Solomon 1980). W przypadku klasyfikacji dysponuje się określoną liczbą grup obiektów, zadanie polega na przydzieleniu, zgodnie z określonymi zewnętrznymi kryteriami, określonych obiektów do odpowiadających im grup (np. metoda analizy dyskryminacyjnej Fishera (Marek 1989)). Metody skupiania bazują na wewnętrznym kryterium podziału, jakim jest matematycznie zdefiniowane podobieństwo między obiektami. Odpowiednie procedury skupiania tworzą grupy (skupienia) obiektów. Liczba tych grup nie jest z reguły zadawana z góry. Funkcję podobieństwa s określa się następująco (Marek 1989):

$$s : G \times G \rightarrow R, \quad S_{ij} = s(O_i, O_j) \quad (1)$$

gdzie: G – zbiór obiektów,
 (O_i, O_j) – para obiektów,
 $S_{ij} \in R$ – liczba rzeczywista (przeważnie z przedziału $[0, 1]$).

Metryczna funkcja podobieństwa musi spełniać trzy podstawowe aksjomaty:

$$S_{ij} \leq S_{max} \text{ (ograniczoność), } S_{ij} = S_{ji} \text{ (symetria) i}$$

$$S_{ii} = S_{max} \text{ – wartość max. (zwrotność)} \quad (2)$$

oraz dwa warunki:

$$S_{ij} = S_{max}, \quad \text{gdy } O_i = O_j \text{ (gdy obiekty są identyczne)}$$

$$\bigwedge |(S_{ij} + S_{jk}) \cdot S_{ik}| \leq |S_{ij} \cdot S_{jk}| \text{ („nierówność trójkąta")} \quad (3)$$

$$O_i, O_j, O_k \in G.$$

Odwrotnością funkcji podobieństwa jest funkcja odległości d (funkcja niepodobieństwa). Funkcję d definiuje się analogicznie jak funkcję s , czyli:

$$d: G \times G \rightarrow R_+, \quad D_{ij} = d(O_i, O_j) \quad (4)$$

Metryczna funkcja odległości (np. norma lub odległość euklidesowa), podobnie jak funkcja podobieństwa, musi spełniać trzy aksjomaty:

$$D_{ij} \geq 0, \quad D_{ij} = D_{ji}, \quad D_{ii} = 0 \quad (5)$$

oraz dwa warunki:

$$D_{ij} = 0 \Rightarrow O_i = O_j, \quad D_{ij} \leq D_{ij} + D_{jk} \text{ (oznaczenia jak wyżej)} \quad (6)$$

Utworzenie i przeanalizowanie wszystkich możliwych podziałów zbioru $G = \{O_1, O_2, \dots, O_n\}$ obiektów jest bardzo trudne już przy stosunkowo niedużej liczbie elementów n . Liczba L wszystkich możliwych zupełnych podziałów n -elementowego zbioru obiektów na h rozłącznych skupień jest dana formułą (liczba Stirlinga II rodzaju):

$$L = \left(\frac{1}{h!}\right) \cdot \sum_{i=0}^h (-1)^i \cdot \binom{h}{i} \cdot (h-i)^n \quad (7)$$

Na przykład dla $n = 10$ i $h = 3$ liczba wszystkich możliwych zupełnych podziałów wynosi $L = 9330$ (Marek 1989).

Powstała zatem potrzeba opracowania takich metod, jakie mogłyby w efektywny sposób skupiać obiekty. Wielu autorów twierdzi, iż sekwencyjne, aglomeracyjne, hierarchiczne i niepokrywające metody SAHN zajmują eksponowane miejsce wśród metod analizy skupień. W zasadzie wszystkie metody skupiania SAHN opierają się na jednym ogólnym schemacie postępowania:

– wyszukanie w macierzy odległości lub podobieństwa D najbliższej sobie pary skupień (s_i, s_j) ,

– połączenie skupienia s_i oraz s_j w jedno, nowe zaglomerowane skupienie s_a ,

– uaktualnienie macierzy D , uwzględniając fakt powstania nowego skupienia.

Powtarzana sekwencyjnie procedura kończy się, gdy w macierzy D występuje tylko jedna wartość. Poszczególne metody SAHN różnią się w zasadzie jedynie trzecim krokiem powyższego algorytmu. Uogólniona formuła rekurencyjna opisująca trzeci krok algorytmu ma postać:

$$d(s_a, s_k) = \alpha_i \cdot d(s_i, s_k) + \alpha_j \cdot d(s_j, s_k) + \beta \cdot d(s_i, s_j) + \gamma \cdot |d(s_i, s_k) - d(s_j, s_k)| \quad (8)$$

gdzie: $d(\dots)$ – odległość (podobieństwo) pomiędzy dwoma skupieniami,

s_i, s_j, s_k – skupienia i, j oraz k ,

s_a – zaglomerowane (nowe) skupienie, powstałe z połączenia s_i oraz s_j ,

$\alpha_i, \alpha_j, \beta, \gamma$ – parametry charakteryzujące metody skupiania.

Tabela 1

Wartości średnie (wg klasyfikacji kodowej 0-5) dla 10 wydzielonych grup
Average values (according to code classification 0-5) of 10 selected groups

A	B	C	D	E	F	G	H	I	J	K	L	M	N
1/3	1,66	0	5,0	5,0	0	5,0	0	0	5,0	0,33	1,0	1,66	0,66
2/559	2,93	4,97	0,17	3,77	1,76	4,39	1,14	0,29	3,49	1,90	1,18	0,40	0,86
3/5	2,8	5,0	0	2,2	4,8	4,8	5,0	0	1,6	0,4	0,4	1,0	1,4
4/2	2,5	5,0	0	5,0	0	5,0	1,0	0	0,5	5,0	0,5	3,5	0
5/2	2,0	0	5,0	5,0	0	5,0	1,5	0	0	4,5	0	0	0
6/3	3,67	3,0	3,67	5,0	0	4,0	0,67	0	5,0	0,33	0,33	0,33	0,67
7/3	2,0	0	5,0	5,0	0	5,0	0	0	0	5,0	4,0	0	0
8/1	3,0	0	5,0	5,0	0	5,0	5,0	0	0	0	2,0	0	3,0
9/1	1,0	5,0	0	5,0	0	5,0	0	5,0	2,0	0	0	4,0	0
10/5	3,8	5,0	0	3,8	3,2	5,0	0	3,2	3,4	0	2,0	0,6	0,2

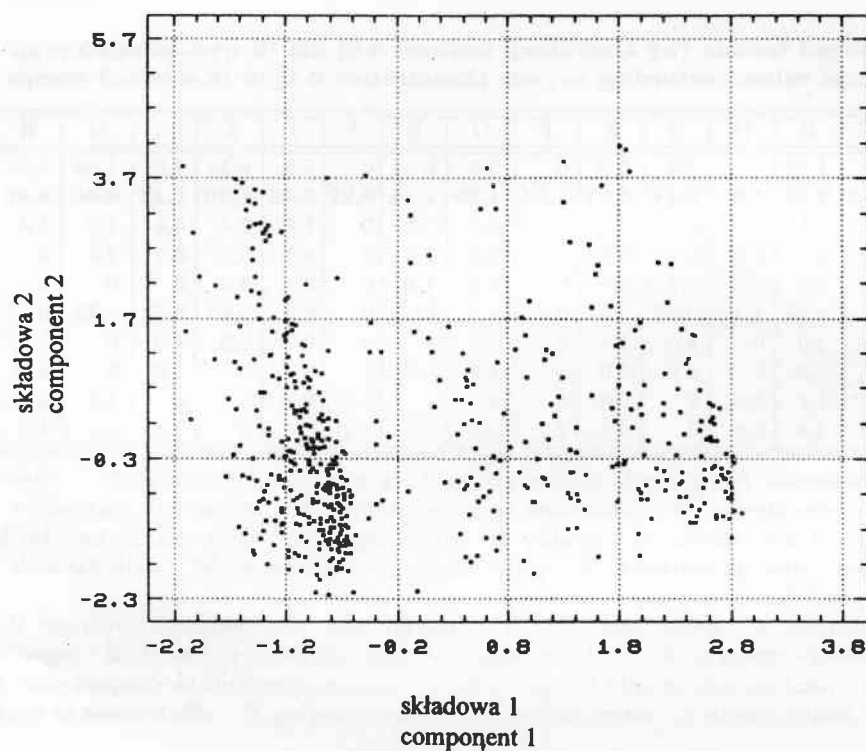
Oznaczenia: A – grupa/liczebność grupy, B – powierzchnia drenowania, C – drenowanie systematyczne, D – drenowanie niesystematyczne, E – technologia rowkowa, F – technologia bezrowkowa, G – spadki, H – gleby zwarte, I – gleby zwarte na lekkich, J – gleby lekkie na zwężłych, K – gleby lekkie, L – słomowanie, M – żwirowanie, N – skuteczność funkcjonowania.

Notations: A – group/group size, B – drainage area, C – systematic drainage, D – unsystematic drainage, E – ditch technique, F – mole technique, G – slope, H – compact soils, I – compact soils on light textured soils, J – light textured soils on compact soils, K – light textured soils, L – straw covering, M – gravel covering, N – effectiveness of work.

Dla celów niniejszej pracy przyjęto (często stosowaną) metodę średnich połączeń (average method), tzn. metodę SAHN o parametrach: $\alpha_i = 0,5$, $\alpha_j = 0,5$, $\beta = 0$ i $\gamma = 0$.

Wyniki i dyskusja

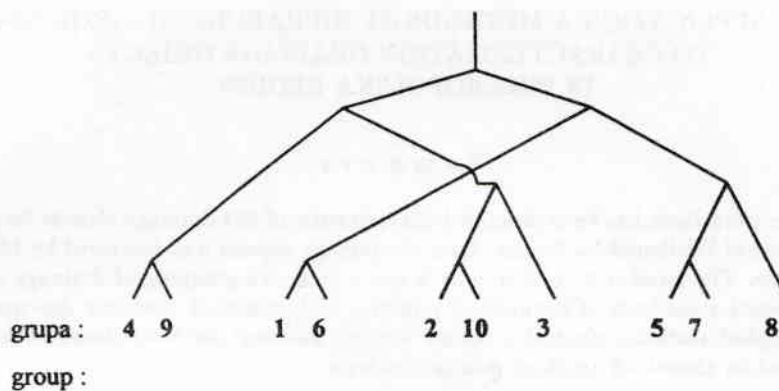
Ze względu na dużą liczbę analizowanych obiektów drenarskich (584) opisanych dużą liczbą cech (13) grupowanie prowadzono dwuetapowo. Postępowanie takie wynikało z możliwości obliczeniowych komputerów PC. W pierwszym etapie podzielono arbitralnie obiekty na 8 zbiorów i przeprowadzono niezależnie dla każdego zbioru rozdział na 10 grup. Uzyskano następujące rozkłady: w 6 zbiorach jedna z wydzielanych grup zdecydowanie dominowała (zawierała 71-79% obiektów), a w 2 zbiorach dominowały dwie z wydzielanych grup (zawierały po 82 i 84% obiektów). W konsekwencji po pierwszym etapie grupowania uzyskano 80 skupień, z 10 dominującymi. Drugi etap grupowania polegał na rozdziale tych 80 skupień na 10 nowych grup. Ostatecznie uzyskane rezultaty grupowania obiektów



Ryc. 1. Rozmieszczenie 584 obiektów drenarskich w układzie dwóch pierwszych składowych głównych

Fig. 1. Distribution of 584 drainage objects in plot of first two principal components

tów drenarskich wskazują na istnienie jednej dominującej grupy (zawierającej aż 559 obiektów) (tab. 1). Dla graficznej prezentacji rozkładu badanych obiektów zastosowano metodę składowych głównych. W tym celu wyżej wymienione 13 cech pierwotnych opisujących 584 obiekty poddano transformacji w składowe główne (kombinacje liniowe cech pierwotnych). Transformacja taka zapewnia, z przyjętego algorytmu postępowania, iż pierwsza składowa główna wnosi największą część informacji zawartej w cechach pierwotnych. Przeprowadzone obliczenia wykazały, iż dwie pierwsze składowe główne wnoszą w sumie 55% informacji zawartej w cechach pierwotnych. Na rycinie 1 przedstawiono wynik obliczeń. Można ostrożnie skomentować (45% informacji nie ujmuje rycina) otrzymany wynik – obiekty rozkładają się na dwa ugrupowania główne. Metodę składowych głównych zastosowano również do wartości ujętych w tabeli 1. Pozwoliło to w konsekwencji na utworzenie poglądowego dendrogramu obrazującego związku pomiędzy wydzielonymi 10 grupami obiektów drenarskich (ryc. 2).



Ryc. 2. Związki pomiędzy 10 wydzielonymi grupami obiektów drenarskich
 Fig. 2. Relations between 10 selected groups of drainage objects

Podsumowanie

Proponowana metoda grupowań hierarchicznych wraz z systemem kodowania do charakterystyki obiektów drenarskich wydaje się zasadna, jako iż cechy opisujące te obiekty mają często charakter jakościowy. Uzyskany wynik grupowania – zdecydowana większość obiektów została przypisana do jednej grupy – wskazuje na duże podobieństwo obiektów drenarskich położonych w jednym regionie (Wielkopolsce). Istnieją duże trudności w jednoznacznej ocenie obiektów drenarskich – wskazuje na to wynik uzyskany metodą składowych głównych. Przyjęte w pracy rozwiązanie: wydzielone cechy obiektów drenarskich, system kodowania nie należy traktować jako ostateczny, obrazują one jednak zasadniczą koncepcję autorów.

Literatura

- Kosturkiewicz A., Bykowski J. (1989): Ocena skuteczności drenowania w regionie Wielkopolski. Sprawozdanie końcowe z badań prowadzonych w latach 1986-1989 w ramach CPBR 10.8.3.1.A.17. Maszyn. KMiKŚ AR, Poznań.
- Marek T. (1989): Analiza skupień w badaniach empirycznych. Metody SAHN. PWN, Warszawa.
- Mezzich J.E., Solomon H. (1980): Taxonomy and behavioural science. Academic Press, New York.
- Śniadowski Z. (1978): Wytyczne drenowania gruntów ornych. Wyd. IMUZ, Falenty.

AN APPLICATION A METHODS OF HIERARCHICAL GROUPING
TO CHARACTERIZATION DRAINAGE OBJECTS
IN WIELKOPOLSKA REGION

S u m m a r y

In the work there has been described characteristic of 584 drainage objects located in 7 provinces of Wielkopolska Region. Each of drainage objects was described by 13 coded parameters. The number 0, 1, 2, 3, 4 or 5 was a code. To grouping of drainage objects was proposed a methods of hierarchical grouping and presented synthetic description of them. Applied statistics method – cluster analysis (average method) showed that main group contain almost all analysed drainage objects.