

**Agnieszka Sompolska-Rzechuła**

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

---

## JAKOŚĆ KLASYFIKACJI WOJEWÓDZTW POLSKI POD WZGLĘDEM STANU ZANIECZYSZCZENIA ŚRODOWISKA NATURALNEGO

---

**Streszczenie:** Celem opracowania jest ocena jakości klasyfikacji województw Polski pod względem stanu zanieczyszczenia środowiska naturalnego oraz wyłonienie najlepszego podziału analizowanych obiektów. Jakość dokonanego podziału została oceniona za pomocą mierników określających stopień heterogeniczności poszczególnych grup i punktów izolowanych w danej klasyfikacji. W klasyfikacji wykorzystano metodę mediany i Warda, na różnych poziomach dendrogramu oceniono jakość klasyfikacji oraz dokonano wyboru podziału najlepszego. Badanie przeprowadzono w przekroju województw Polski, a informacje statystyczne dotyczą roku 2005. Lista zmiennych diagnostycznych obejmowała główne obszary podlegające zanieczyszczeniom, tj. powietrze, wodę i glebę. Obliczenia wykonano w programach *Statistica* i *Taksonomia numeryczna*.

**Słowa kluczowe:** jakość klasyfikacji, stan środowiska naturalnego, dendrogram, mierniki heterogeniczności grup.

### 1. Wstęp

Jakość środowiska jest ściśle powiązana z poziomem rozwoju ekonomicznego oraz jakością życia ludności. Dynamiczny rozwój gospodarczy powoduje nadmierne wykorzystanie i obciążenie środowiska naturalnego. Stan środowiska jest wypadkową jakości trzech jego komponentów: powierzchni ziemi, zasobów wodnych i powietrza atmosferycznego. Ich zanieczyszczenie negatywnie oddziałuje na zasoby przyrody żywej, a także na warunki życia i zdrowie ludności. Zanieczyszczenie środowiska to wprowadzenie do powietrza, wody i ziemi substancji stałych, ciekłych lub gazowych albo energii w takich ilościach lub w takim składzie, że może to ujemnie wpłynąć na zdrowie człowieka, klimat, przyrodę żywą, glebę, wodę lub może spowodować inne zmiany w środowisku, w tym również kulturowym.

Rzetelna wiedza o stanie środowiska przyrodniczego i zmianach w nim zachodzących jest niezbędna do podejmowania optymalnych decyzji, które mogą dotyczyć np. przeznaczenia gruntów, lokalizacji wszelkiego rodzaju zakładów czy też – w szerszym zakresie – restrukturyzacji przemysłu w regionie (województwie, gminie).

Celem opracowania jest ocena jakości klasyfikacji województw Polski pod względem stanu zanieczyszczenia środowiska naturalnego oraz wyłonienie najlepszego podziału analizowanych obiektów.

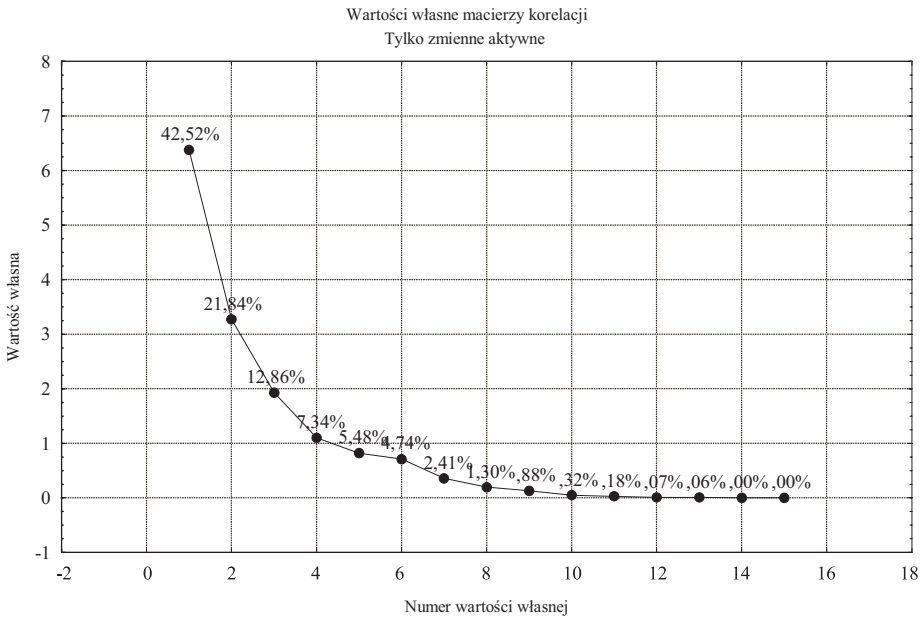
## 2. Materiał badawczy i metoda

Badanie dotyczy województw Polski w roku 2006. Dane statystyczne zostały zaczerpnięte z wydawnictwa Głównego Urzędu Statystycznego *Ochrona Środowiska 2006 [Ochrona środowiska 2005...]*. Wstępna lista zmiennych diagnostycznych obejmowała główne obszary, które podlegają zanieczyszczeniom, tj. powietrze, wodę i glebę.

Szczegółowa lista potencjalnych zmiennych diagnostycznych jest następująca:

- $X_1$  – ludność miast korzystająca z sieci wodociągowej w % ludności miast ogółem,
- $X_2$  – zużycie wody w gospodarstwach domowych na 1 mieszkańca miast w  $m^3$  na rok,
- $X_3$  – udział ścieków odprowadzanych bezpośrednio z zakładów w ogólnej ilości ścieków przemysłowych i komunalnych (w %),
- $X_4$  – udział ścieków odprowadzanych siecią kanalizacyjną w ogólnej ilości ścieków przemysłowych i komunalnych (w %),
- $X_5$  – liczba oczyszczalni ścieków na 100  $km^2$  powierzchni województwa,
- $X_6$  – ludność korzystająca z oczyszczalni ścieków w % ogólnej liczby ludności województwa,
- $X_7$  – ludność miast korzystająca z oczyszczalni ścieków w % ogólnej liczby ludności województwa,
- $X_8$  – ludność wsi korzystająca z oczyszczalni ścieków w % ogólnej liczby ludności,
- $X_9$  – emisja zanieczyszczeń pyłowych z zakładów szczególnie uciążliwych w tonach na  $km^2$  powierzchni województwa,
- $X_{10}$  – emisja dwutlenku siarki z zakładów szczególnie uciążliwych w tonach na  $km^2$  powierzchni województwa,
- $X_{11}$  – emisja dwutlenku węgla z zakładów szczególnie uciążliwych w tonach na  $km^2$  powierzchni województwa,
- $X_{12}$  – liczba zakładów szczególnie uciążliwie emitujących zanieczyszczenia powietrza na 100  $km^2$  powierzchni województwa,
- $X_{13}$  – odpady wytworzone w tys. ton na  $km^2$  powierzchni województwa,
- $X_{14}$  – odpady niebezpieczne wytworzone w tonach na  $km^2$  powierzchni województwa,
- $X_{15}$  – odpady komunalne zebrane w kg na 1 mieszkańca,
- $X_{16}$  – powierzchnia składowisk (wysypisk) w ogólnej powierzchni województwa (w %),
- $X_{17}$  – liczba dzikich wysypisk odpadów na 100  $km^2$  powierzchni województwa.

Cechy  $X_1$  i  $X_2$  charakteryzują się niską zmiennością (poniżej 10%) i zostały wykluczone z badania. Pozostałe zmienne wykazują duże zróżnicowanie (od 11% do ponad 195%), mogą zatem być podstawą do dokonania dyskryminacji obiektów pod względem analizowanego zjawiska. W celu zredukowania wstępnej listy zmiennych diagnostycznych i wyeliminowania cech silnie ze sobą skorelowanych zastosowano metodę głównych składowych (PCA). Metoda głównych składowych polega na ortogonalnym przekształceniu  $k$ -wymiarowego układu zmiennych opisujących obserwacje wielowymiarowe na nowy układ zmiennych nieskorelowanych [Morrison 1990; Ostasiewicz 1998]. Przekształcenia tego dokonuje się tak, że wariancje kolejnych składowych są coraz mniejsze, przy czym całkowita wariancja wszystkich zmiennych wyjściowych jest równa sumie wariancji wszystkich głównych składowych. Otrzymano wektory wartości własnych oraz wartości mówiące o tym, jaki procent wariancji zmiennych przyjętych w badaniu wyjaśniają poszczególne główne składowe. Cztery pierwsze składowe wyjaśniają 84,6% całkowitej zmienności analizowanych zmiennych, okazały się one istotne zgodnie z kryterium Kaisera. Ilustracją istotności głównych składowych jest wykres osypiska przedstawiony na rys. 1.



Rys. 1. Wykres wartości własnych – test osypiska

Źródło: opracowanie własne.

Na podstawie macierzy unormowanych wektorów własnych stwierdzono, że pierwsza główna składowa wyjaśnia 42,52% zmienności i jest najbardziej skorelo-

wana z sześcioma zmiennymi, które uznano za determinanty stanu zanieczyszczenia środowiska. Należą do nich:  $X_3$ ,  $X_7$ ,  $X_9$ ,  $X_{10}$ ,  $X_{11}$ ,  $X_{15}$ .

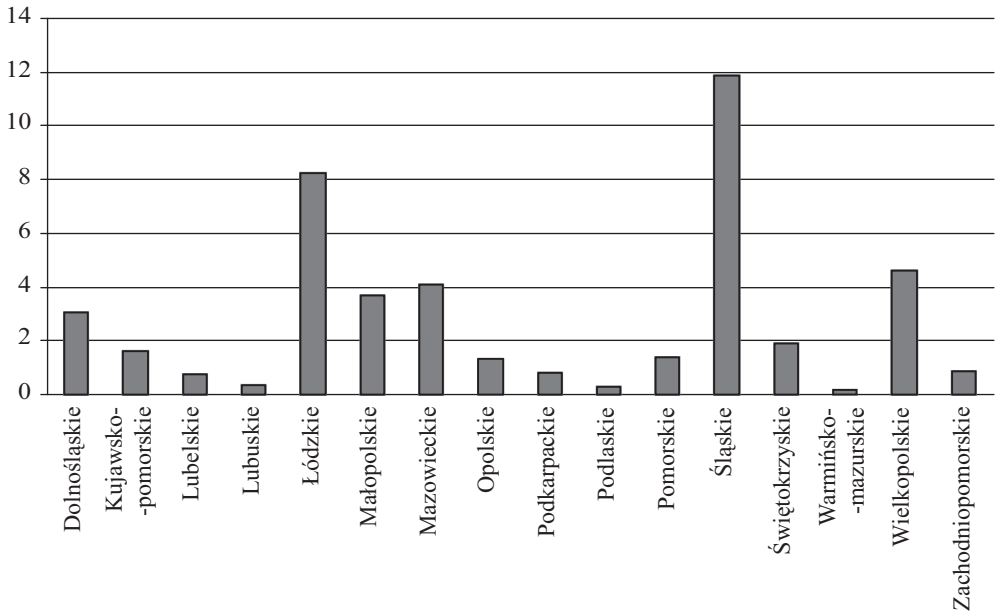
Podstawowe charakterystyki opisowe wyłonionych zmiennych diagnostycznych zostały przedstawione w tab. 1.

**Tabela 1.** Podstawowe charakterystyki opisowe zmiennych diagnostycznych

Zmienna	$X_3$	$X_7$	$X_9$	$X_{10}$	$X_{11}$	$X_{15}$
Średnia $\bar{x}$	60,27	87,93	0,40	2,82	778,40	236,38
Odchylenie standardowe $s$	28,97	9,36	0,40	3,22	845,46	51,51
Współczynnik zmienności $V_s$	48,08	10,64	100,44	114,20	108,62	21,79

Źródło: obliczenia własne.

Największą zmiennością charakteryzuje się zmienna  $X_{10}$  – emisja dwutlenku siarki z zakładów szczególnie uciążliwych w tonach na km<sup>2</sup> powierzchni województwa. Na rysunku 2 przedstawiono procentowy udział emisji dwutlenku siarki z zakładów szczególnie uciążliwych w tonach na km<sup>2</sup> powierzchni województwa.



**Rys. 2.** Emisja dwutlenku siarki z zakładów szczególnie uciążliwych w tonach na km<sup>2</sup> powierzchni województwa (w %)

Źródło: opracowanie własne.

Na podstawie wyłonionych determinant zanieczyszczenia środowiska przyrodniczego dokonano klasyfikacji województw Polski wybranymi metodami hierarchicznymi i zbadano jakość otrzymanych podziałów.

Zastosowanie metod hierarchicznych polega na tym, że tworzy się ciąg, czyli tzw. hierarchię. W zależności od sposobu otrzymywania ciągu klasyfikacji wyróżnia się dwie grupy metod hierarchicznych: hierarchiczne metody grupowania (zwane procedurami aglomeracyjnymi) oraz hierarchiczne metody podziału.

We wszystkich metodach hierarchicznych istnieje możliwość graficznego przedstawienia podziału w postaci tzw. dendrogramu (drzewka połączeń), który ilustruje kolejne połączenia skupień coraz wyższego rzędu. Uzyskana hierarchia pozwala na określenie wzajemnego położenia skupień i obiektów w nich zawartych. Istotną wadą procedur tej grupy jest brak oczywistego kryterium ustalenia liczby skupień względnie jednorodnych klas oraz w niektórych przypadkach skłonność do tworzenia skupień w postaci łańcucha, a więc do powstawania skupień obiektów dość odległych od siebie.

Do oceny skonstruowanych podziałów wykorzystuje się wiele różnych mierników, które określają zwartość grupy lub jej odrębność. Jeżeli obiekty w danej grupie leżą blisko siebie, to grupa jest zwarta, natomiast gdy grupa jest bardziej oddalona od pozostałych klas, wówczas jest odrębna. Zwartość grupy nazywa się homogenicznością, a odrębność – heterogenicznością. W pracy [Grabiński, Wydymus, Zeliaś 1989, s. 136] zaproponowano przeprowadzanie oceny jakości metod taksonomicznych za pomocą analizy teoretycznej oraz ilościowej. Wśród miar określających poprawność grupowania można wyróżnić miary: homogeniczności skupienia, heterogeniczności skupień, poprawności skupień, zmienności wewnątrz- i międzygrupowej oraz równomierności rozkładu wielkości grup.

W analizie jakości klasyfikacji województw Polski pod względem zanieczyszczenia środowiska przyrodniczego wykorzystano miary zaproponowane w pracy [Kolenda 2006, s. 110-121]. Uwzględniono mierniki służące do oceny heterogeniczności grupy  $h$  oraz oceny całej klasyfikacji  $H$ , miernik  $SC$  (*silhouette coefficient*).

Do oceny heterogeniczności grupy wykorzystano trzy mierniki:  $h^{**}$ ,  $h^*$  oraz  $h$ . Mierniki te są uporządkowane i mają określoną hierarchię; największa heterogeniczność oznaczona jest symbolem  $h^{**}$ , a najniższa – symbolem  $h$ . Jeżeli grupa ma heterogeniczność typu  $h^{**}$ , to ma także heterogeniczność typu  $h^*$ ,  $h$ .

Określenia mierników  $h^{**}$ ,  $h^*$  oraz  $h$  są następujące [Kolenda 2006, s. 118]:

- Grupa  $c$  ma heterogeniczność typu  $h^{**}$  – bardzo dobrą heterogeniczność, jeśli:  $I5_{(c)} > \max(D_l)$ , gdzie:  $D_l$  – średnica taksonometryczna  $l$ -tej grupy, czyli maksymalna odległość obiektów w grupie,  $D_l = I2 = \max = \max\{d_{i,j}\}$  ( $i, j = 1, \dots, n_l$ );  $I5_{(c)}$  – odrębność grupy  $c$  wyznaczona za pomocą miernika  $I5$ , gdzie  $I5 = d_{\min}^{\{p\}\{j\}} = \min_{\{p\}} \min_{j \in \{p\}} \{d(p^{(l)}, j)\}$  oznacza minimalną odległość obiektu  $l$ -tej grupy do obiektu spoza grupy, a  $\{p\}$  to zbiór obiektów  $l$ -tej grupy. Grupa  $c$  ma bar-

dzo dobrą heterogeniczność, jeżeli odrębność grupy  $c$  jest większa od maksymalnej średnicy taksonomicznej grup. Miernik ten można wykorzystać także do określania odrębności punktu izolowanego.

- Grupa  $c$  ma heterogeniczność typu  $h^*$  – dobrą heterogeniczność, jeśli:  $IS_{(c)} > D_c$ . Grupa  $c$  ma dobrą heterogeniczność, jeżeli odrębność grupy  $c$  jest większa od średnicy tej grupy. Miernik ten można wykorzystać także do określania odrębności punktu izolowanego.
- Grupa  $c$  ma heterogeniczność typu  $h$  – słabą, lecz dopuszczalną heterogeniczność, jeżeli:  $IS_{(c)} > \max_i \left( d_{i,R} \right)$ . Grupa  $c$  ma słabą, lecz dopuszczalną heterogeniczność, jeżeli odrębność grupy  $c$  jest większa od maksymalnej odległości obiektów od reprezentanta ( $R$ ) w grupie  $c$  (większa od promienia grupy). Według tego miernika nie można określać odrębności punktu izolowanego.

Według tego miernika nie można określać odrębności punktu izolowanego.

- Grupa  $c$  jest heterogeniczna według miernika  $h'$ , jeśli wartość  $IS_{(c)}$  jest większa od średniej wartości średnicy grup bez punktów izolowanych:  $IS_{(c)} > \frac{1}{k-p} \sum_{i=1}^{k-p} D_i$ , gdzie:  $k$  – liczba grup,  $p$  – liczba obiektów izolowanych. Miernik ten ocenia słabszą heterogeniczność punktu izolowanego. Grupa nie jest heterogeniczna (nie ma odrębności), jeżeli nie ma heterogeniczności przynajmniej typu  $h$ , a dla obiektu izolowanego przynajmniej typu  $h'$ .

Do oceny jakości klasyfikacji wykorzystano mierniki typu  $H$  zaproponowane w [Kolenda 2006, s. 120]. Według tej propozycji ocena klasyfikacji zostaje przeprowadzona na podstawie określenia stopnia heterogeniczności wszystkich poszczególnych grup i punktów izolowanych w danej klasyfikacji. O jakości całej klasyfikacji decyduje najniższa heterogeniczność grupy lub punktu izolowanego w tej klasyfikacji. Wyróżniono następujące mierniki charakteryzujące poszczególne rodzaje klasyfikacji:

- Miernik  $H^{**}$  – oznacza bardzo dobrą klasyfikację – wszystkie grupy w tej klasyfikacji mają heterogeniczność typu  $h^{**}$ .
- Miernik  $H^*$  – oznacza dobrą klasyfikację – wszystkie grupy w tej klasyfikacji mają heterogeniczność typu  $h^*$ , a punkty izolowane – przynajmniej typu  $h'$  i istnieje chociaż jedna grupa, która ma heterogeniczność typu  $h^*$  lub punkt izolowany typu  $h'$ .
- Miernik  $H$  – oznacza słabą, lecz dopuszczalną klasyfikację – wszystkie grupy w tej klasyfikacji mają heterogeniczność przynajmniej typu  $h$ , a punkt izolowany przynajmniej typu  $h'$  i istnieje chociaż jedna grupa o heterogeniczności typu  $h$  lub punkt izolowany o heterogeniczności typu  $h'$ .
- Klasyfikacje uznaje się za złą, jeśli przynajmniej jedna z grup w danej klasyfikacji lub punkt izolowany nie będzie mieć heterogeniczności.

Miernik  $SC$  wyznacza się ze wzoru [Kolenda 2006, s. 120]:

$$SC = \max_k(\bar{s}_k),$$

gdzie:  $\bar{s}_k = SK = \frac{1}{n_k} \sum_{i=1}^{n_k} s_i$  – średnia wartość dla wszystkich punktów  $i$  należących do grupy  $k$ ,

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)},$$

$a_i$  – średnia odległość obiektu  $i$  względem wszystkich obiektów w swojej grupie,

$b_i$  – średnia odległość obiektu  $i$  względem wszystkich obiektów w grupie sąsiedniej (grupa sąsiednia dla obiektu  $i$  jest grupą, w której znajduje się obiekt  $j$  (obiekt  $j$  jest najbliższym sąsiadem dla obiektu  $i$  spośród obiektów spoza swojej grupy)),

$s_i = 0$  – dla grupy zawierającej jeden punkt,

$k$  – numer grupy,

$n_k$  – liczebność grupy,

$i$  – obiekt należący do grupy  $k$ .

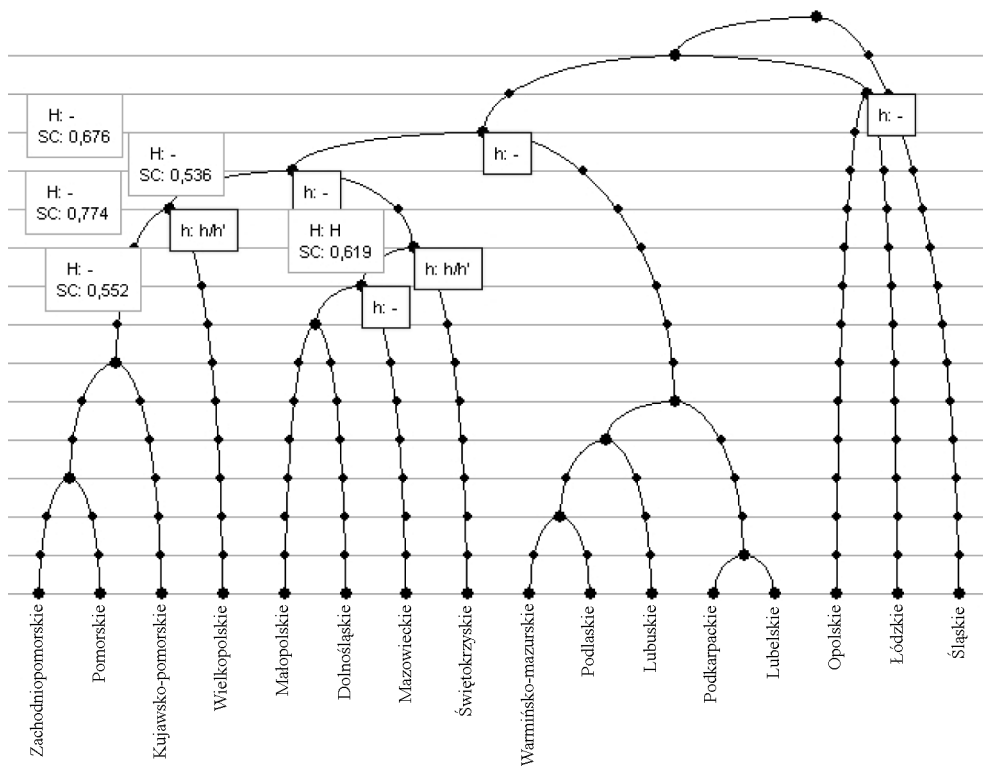
Interpretacja miernika  $SC$  jest następująca:

- wartość miernika z przedziału 0,71-1,00 – podział na grupy w danej klasyfikacji należy uznać za bardzo dobry,
- wartość miernika z przedziału 0,51-0,70 – podział jest dobry,
- wartość miernika z przedziału 0,26-0,50 – podział jest słaby, zaleca się stosowanie innej klasyfikacji,
- wartość miernika z przedziału 0,25-0,00 – podział jest zły, wyodrębnione grupy nie tworzą żadnych grup obiektów podobnych.

### 3. Wyniki badania

Wyłoniony zbiór sześciu zmiennych diagnostycznych został wykorzystany do budowy klasyfikacji województw Polski pod względem stanu zanieczyszczenia środowiska przyrodniczego. Zastosowano metody hierarchiczne, następnie za pomocą wybranych mierników określających jakość otrzymanych podziałów podjęto próbę wskazania najlepszej klasyfikacji. Stosując metody hierarchiczne, otrzymuje się całą strukturę różnych podziałów. Za najlepszy proponuje się uznać taki podział, w którym jest mała liczba grup  $i$  w którym wszystkie grupy są heterogeniczne, najlepiej typu  $H^{**}$  lub  $H^*$ , może też być typ  $H$ . W opracowaniu sprawdzano kolejno rzędy dendrogramu, począwszy od góry, za pomocą mierników  $H$ , jednocześnie biorąc pod uwagę liczebności poszczególnych grup. W większości analizowanych klasyfikacji, z wykorzystaniem różnych metod hierarchicznych, otrzymywano podziały charakte-

ryzujące się brakiem heterogeniczności lub niekiedy słabą heterogenicznością grup. Taka sytuacja została zaprezentowana na rys. 3, na którym pokazano dendrogram otrzymany metodą mediany. Na dendrogramie tylko w jednym przypadku można zaobserwować podział obiektów, dla którego istnieje słaba heterogeniczność grup, a wartość miernika  $SC$  wynosi 0,619 (według tego miernika podział uznaje się za dobry). W wyniku takiego podziału otrzymuje się trzy grupy oraz trzy punkty izolowane. Do pierwszej grupy należą województwa: zachodniopomorskie, pomorskie, kujawsko-pomorskie oraz wielkopolskie, druga klasa to także cztery obiekty: małopolskie, dolnośląskie, mazowieckie i świętokrzyskie, trzecia grupa to: warmińsko-mazurskie, podlaskie, lubuskie, podkarpackie i lubelskie. Odrębne jednoelementowe skupienia tworzą województwa: opolskie, łódzkie i śląskie. Województwo śląskie charakteryzuje się najwyższymi wartościami cech określających emisję: zanieczyszczeń pyłowych, dwutlenku siarki i dwutlenku węgla z zakładów szczególnie uciążliwych w tonach na  $\text{km}^2$  powierzchni województwa. Pozostałe dwa województwa izolowane także mają wyższe od średnich wartości zmiennych diagnostycznych. Skupienie trzecie (pięć obiektów) charakteryzuje się najkorzystniejszą sytuacją pod



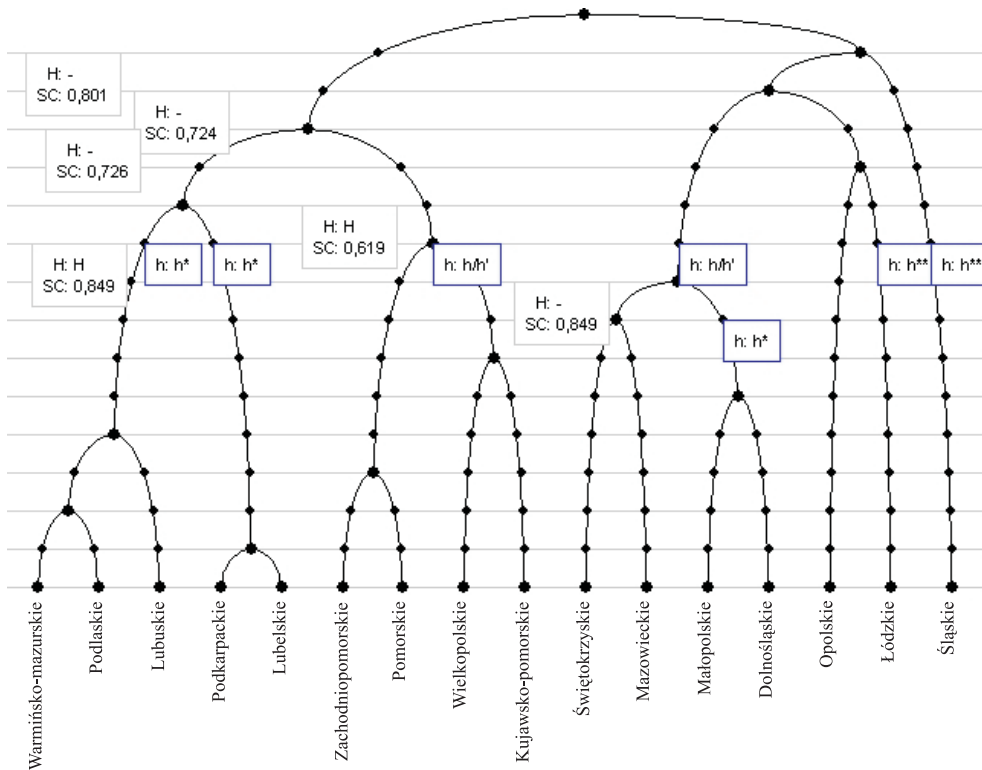
Rys. 3. Dendrogram otrzymany metodą mediany

Źródło: opracowanie własne.



względem stanu zanieczyszczenia środowiska przyrodniczego, uwzględniając przyjęte zmienne diagnostyczne (np. niskie procentowe wartości udziału ścieków odprowadzanych bezpośrednio z zakładów w ogólnej ilości ścieków przemysłowych i komunalnych czy emisji).

Klasyfikując Polskę w ujęciu województw pod względem stanu zanieczyszczenia środowiska przyrodniczego, wykorzystano także metodę Warda. Dendrogram otrzymany za pomocą tej metody przedstawiono na rys. 4. Pokazano na nim także wartości mierników określających jakość otrzymanych podziałów. Klasyfikacja otrzymana metodą Warda daje polepszenie klasyfikacji obiektów w rzędzie niżej o jeden w porównaniu z metodą mediany. Jest to jednak tylko polepszenie wartości miernika  $SC$  z 0,619 na 0,849, natomiast miernik  $H$  pozostaje bez zmian. Uwzględniając wartość miernika  $SC$  0,849, otrzymuje się cztery skupienia kilkuelementowe oraz w dalszym ciągu trzy punkty izolowane. Różnica między podziałami polega na tym, że klasa składająca się z pięciu elementów uzyskanych metodą mediany została podzielona w metodzie Warda na dwie grupy.



Rys. 4. Dendrogram otrzymany metodą Warda

Źródło: opracowanie własne.

Podział na większą liczbę skupień nie wnosi do klasyfikacji bardziej dokładnych informacji o analizowanych obiektach. Wartości średnie dla przyjętych zmiennych diagnostycznych w wyodrębnionej klasie są na poziomie średnich ogólnych. Wykluczenie dwóch województw z wcześniejszej grupy podkreśliło lepszą sytuację pozostałych trzech województw w porównaniu z innymi grupami pod względem stanu środowiska przyrodniczego.

#### 4. Podsumowanie

W opracowaniu wykorzystano wybrane mierniki do oceny klasyfikacji Polski w ujęciu województw pod względem stanu środowiska przyrodniczego. Do wyznaczenia determinant analizowanego kryterium wykorzystano metodę głównych składowych (PCA), która pozwoliła wyłonić te najistotniejsze spośród wielu czynników wpływających na zanieczyszczenie powietrza, gleby oraz wody. Zredukowana lista zmiennych diagnostycznych stanowiła podstawę klasyfikacji województw Polski.

Analiza zgromadzonego materiału pozwoliła na sformułowanie następujących wniosków:

- Za pomocą metody głównych składowych wyodrębniono cztery pierwsze składowe wyjaśniające 84,6% całkowitej zmienności analizowanych zmiennych. Pierwsza główna składowa wyjaśniała 42,5% ogólnej zmienności i była skorelowana z sześcioma zmiennymi, które uznano za determinanty stanu zanieczyszczenia środowiska w przyrodniczego województwach Polski.
- Metody hierarchiczne posłużyły do utworzenia klasyfikacji województw i pokazały strukturę otrzymanych podziałów.
- Mierniki oceny klasyfikacji dały możliwość wyboru najlepszego podziału, charakteryzującego się niedużą liczbą heterogenicznych klas. Ocena otrzymanych podziałów została potwierdzona zgodnymi wartościami wybranych mierników.
- Wybierając najlepszy podział obiektów, należy brać pod uwagę nie tylko wartości mierników, ale także liczbę klas, od której uzależniona jest wartość miar oceniających jakość klasyfikacji, oraz liczebności poszczególnych grup.
- Wykorzystanie metod wielowymiarowej analizy porównawczej jest bardzo przydatne w badaniu wielu zjawisk, stąd znajduje zastosowanie w tak wielu dziedzinach życia. Należy prowadzić takie analizy w czasie, z zastosowaniem różnych zmiennych diagnostycznych, nie tylko w przekroju województw, ale także krajów, regionów, powiatów czy gmin oraz obszarów chronionych czy przedsiębiorstw. Badanie dynamiczne umożliwia zaobserwowanie prawidłowości analizowanego zjawiska, uwzględnienie zmian i wahań w jego przebiegu, tendencji rozwojowych w czasie. Badania porównawcze powinny mieć charakter analizy przekrojowo-dynamicznej będącej najpełniejszą formą wielowymiarowej analizy porównawczej [Sej-Kolasa 2005, s. 224].

## Literatura

- Grabiński T., Wydymus S., Zeliaś A., *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, PWN, Warszawa 1989.
- Kolenda M., *Taksonomia numeryczna. Klasyfikacja, porządkowanie i analiza obiektów wielocechowych*, AE, Wrocław 2006.
- Morrison D.F., *Wielowymiarowa analiza porównawcza*, PWN, Warszawa 1990.
- Ochrona środowiska 2005*, GUS, Warszawa 2006.
- Ostasiewicz W., *Statystyczne metody analizy danych*, AE, Wrocław 1998.
- Sej-Kolasa M., *Metody taksonomiczne w badaniach środowiska naturalnego*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, *Ekonometria 15, Zastosowania metod ilościowych*, AE, Wrocław 2005.

### QUALITY OF THE CLASSIFICATION OF POLISH VOIVODESHIPS BY THE NATURAL ENVIRONMENT POLLUTION STATE

**Summary:** The purpose of the study is an appraisal of the quality of the classification of Polish voivodeships by the environment pollution and the identification of the best division of objects under study. The quality of the division is estimated by measures determining the homogeneity degree of certain groups and points isolated within the given classification.

While establishing the classification median and Ward methods were applied, the classification quality was rated on different dendrogram levels and the best division was chosen. The study covers all voivodeships and the statistic information refers to the year 2005. The list of diagnostic variables includes key issues subject to the pollution, i.e. soil, water resources and air. The *Statistica* and *Taksonomia numeryczna* software were used for computing.