

*Dorota Pruska**

ZASTOSOWANIE GŁĘBI LOKACYJNEJ I REGRESYJNEJ DO ANALIZY POZIOMU ZANIECZYSZCZENIA ŚRODOWISKA W POLSCE

Streszczenie. W pracy przeprowadzono porównanie poziomu zanieczyszczenia środowiska w województwach przed wstąpieniem i po wstąpieniu Polski do Unii Europejskiej. W analizie wykorzystano miary głębi lokacyjnej i regresyjnej. Zaproponowano zastosowanie mediany wielowymiarowej przy klasyfikacji województw ze względu na poziom zanieczyszczenia środowiska i skuteczność działań zapobiegających temu zjawisku oraz przy ocenie zmian strukturalnych zjawisk w czasie. Współczynniki regresji estymowano metodą najgłębszej regresji w celu wyeliminowania wpływu obserwacji nietypowych.

Słowa kluczowe: głębia lokacyjna, głębia regresyjna, metoda najgłębszej regresji, obserwacje nietypowe.

I. WSTĘP

W badaniach ekonomicznych analiza prawidłowości cechujących rozpatrywane zjawiska oraz określenie kierunku zachodzących w nich zmian wymaga często korzystania ze zbiorów danych zawierających obserwacje nietypowe, odstające od pozostałych. Do opracowania materiału badawczego wskazane jest wówczas zastosowanie metod i miar odpornych na występowanie obserwacji nietypowych.

Celem niniejszej pracy jest wskazanie kierunku i ocena zmian, jakie zaszły w poziomie zanieczyszczenia środowiska w Polsce przed i po wstąpieniu naszego kraju do Unii Europejskiej. Analizie na poziomie województw został podany ogólny stan środowiska w latach 2002–2006 oraz gospodarka odpadami w latach 2001–2006. Głębnię lokacyjną wykorzystano do porównania zmian ogólnego poziomu zanieczyszczenia środowiska ze względu na poszczególne jego elementy (powietrze, wodę i glebę) oraz do analizy natężenia działań mających na celu redukcję negatywnego wpływu działalności człowieka na środowisko. Za pomocą metody najgłębszej regresji oszacowano parametry dwóch modeli dotyczących ilości wytworzonych odpadów oraz nierekultywowanej powierzchni składowania odpadów w danym województwie. W pracy zaproponowano metodę porównania oszacowań parametrów regresji uzyskanych dla tej samej postaci modelu dla różnych lat opartą na głębi lokacyjnej.

* Dr, Katedra Metod Statystycznych, Uniwersytet Łódzki.

II. GŁĘBIA LOKACYJNA

Pojęcie głąbi lokacyjnej związane jest z określeniem położenia badanej jednostki w zbiorze danych i jej podobieństwa do innych elementów tego zbioru. Umożliwia to podział zbioru danych na mniejsze podzbiory charakteryzujące się podobnym poziomem badanej cechy oraz wyodrębnienie jednostek nietypowych, odstających.

Rozpatrujemy zbiór $Z_n = \{x_i = (x_{i1}, \dots, x_{ip}); i = 1, \dots, n\} \subset R^p$.

Głębią lokacyjną (*ldepth*) punktu $\theta \in R^p$ w zbiorze Z_n (por. Rousseeuw, Ruts [1998], Hubert, Rousseeuw, Van Aelst [1999]) nazywamy najmniejszą z liczb określających ilość punktów ze zbioru Z_n należących do poszczególnych domkniętych półprzestrzeni, których brzeg zawiera punkt θ , spośród wszystkich możliwych domkniętych półprzestrzeni, czyli

$$ldepth(\theta; Z_n) = \min_{\|u\|=1} \#\{i; u^T x_i \geq u^T \theta\}. \quad (1)$$

W przypadku gdy $p = 2$, głębię lokacyjną punktu $\theta \in R^2$ w zbiorze Z_n opisuje wzór:

$$ldepth(\theta, Z_n) = \min_H \#\{i; x_i \in H\}, \quad (2)$$

gdzie H to dowolna domknięta półpłaszczyzna ograniczona prostą przechodzącą przez θ .

Głębię lokacyjną danej obserwacji można zdefiniować w sposób umożliwiający jednoznaczne określenie, czy obserwacja jest nietypowa (odstająca). Przyjmuje się, że głębią lokacyjną obserwacji $\theta \in R^p$ w zbiorze $Z_n \subset R^p$ nazywamy najmniejszą liczbę obserwacji, które należy usunąć z tego zbioru, aby θ stała się obserwacją odstającą, przy czym obserwacja $\theta \in R^p$ jest odstająca w zbiorze $Z_n \subset R^p$, wtedy i tylko wtedy gdy leży ona na zewnątrz powłoki wypukłej zbioru Z_n .

Punktowi leżącemu na zewnątrz wypukłej powłoki zbioru Z_n będzie odpowiadała głębia lokacyjna równa zero.

Zbiór D_k punktów θ o głębi lokacyjnej większej lub równej k , czyli gdy $ldepth(\theta, Z_n) \geq k$, nazywamy obszarem głębi dla danego poziomu głębi k .

Brzeg obszaru głębi dla danego poziomu głębi k nazywamy konturem głębi lokacyjnej dla danego poziomu głębi k .

Wielkość $ldepth(\theta, Z_n)$ dla skończonego zbioru Z_n , traktowana jako funkcja wektora θ , przyjmuje wartości od 0 do n (por. Rousseeuw, Ruts [1998], Hubert, Rousseeuw, Van Aelst [1999]).

Głębia lokacyjna może służyć do wykrywania obserwacji nietypowych występujących w zbiorze danych, a także do podziału tego zbioru na mniejsze podzbiory na przykład ze względu na odpowiadające obserwacjom wartości głębi lokacyjnej.

III. GŁĘBIA REGRESYJNA I METODA NAJGŁĘBSZEJ REGRESJI

Głębia regresyjna pozwala na określenie jakości oszacowań parametrów modelu regresji ze względu na dopasowanie modelu do danych empirycznych. Metoda najgłębszej regresji została opracowana w celu wyznaczenia parametrów modelu regresji na podstawie danych zawierających obserwacje nietypowe.

Niech dany będzie zbiór $Z_n = \{(x_{i1}, \dots, x_{ip-1}, y_i); i = 1, \dots, n\} \subset R^p$. Konstruujemy model postaci:

$$y = \theta_1 x_1 + \dots + \theta_{p-1} x_{p-1} + \theta_p + \varepsilon \quad (3)$$

opisujący zależność między zmiennymi Y i X , których realizacjami są wartości odpowiednio y_i oraz $(x_{i1}, \dots, x_{ip-1})$, gdzie $i=1, \dots, n$. Symbol ε oznacza składnik losowy. Niech θ oznacza wektor parametrów tego modelu, tzn. $\theta = [\theta_1, \dots, \theta_p]^T$. Wartości reszt modelu dla zbioru Z_n określone są wzorem:

$$r_i(\theta) = y_i - (\theta_1 x_{i1} + \dots + \theta_{p-1} x_{ip-1} + \theta_p), i=1, \dots, n. \quad (4)$$

Dla $p=2$ dany wektor θ nazywamy niedopasowanym do danych ze zbioru Z_n (por. Rousseeuw, Hubert [1999]), jeżeli istnieje taka liczba $v_\theta = v$ różna od wszystkich wartości zmiennej X , która spełnia warunki:

$$\begin{aligned} & r_i(\theta) < 0 \text{ dla każdego } x_i < v \text{ i } r_i(\theta) > 0 \text{ dla każdego } x_i > v \\ \text{lub} & r_i(\theta) > 0 \text{ dla każdego } x_i < v \text{ i } r_i(\theta) < 0 \text{ dla każdego } x_i > v, \end{aligned} \quad (5)$$

gdzie $x_i, i = 1, \dots, n$, to wartości zmiennej X .

Dla $p > 2$ dany wektor θ nazywamy niedopasowanym do danych ze zbioru Z_n (por. Rousseeuw, Hubert [1999]), wtedy i tylko wtedy gdy istnieje afiniczna hiperpłaszczyzna V w przestrzeni x -ów, taka że żaden $x_i = (x_{i1}, \dots, x_{ip-1}), i=1, \dots, n$ nie należy do V oraz

$$1) r_i(\theta) < 0 \text{ dla każdego } x_i \text{ należącego do jednej z otwartych półprzestrzeni wyznaczonych przez hiperpłaszczyznę } V \quad (6)$$

$$\text{i} \\ 2) r_i(\theta) > 0 \text{ dla każdego } x_i \text{ należącego do drugiej z otwartych półprzestrzeni wyznaczonych przez hiperpłaszczyznę } V,$$

gdzie $x_i, i r_i$ dla $i = 1, \dots, n$, to odpowiednio wartości zmiennej X i reszty z modelu o parametrach θ .

Głębią regresyjną wektora θ parametrów modelu określonego na zbiorze Z_n nazywamy najmniejszą liczbę obserwacji, które powinny zostać usunięte z tego zbioru, aby wektor θ stał się niedopasowany do powstałego zbioru.

Głębina regresji wektora θ w zbiorze Z_n wyraża się wzorem (por. Van Aelst, Rousseeuw, Hubert, Struyf [2000]):

$$rdepth(\theta, Z_n) = \min_{u, v} \{ \#(r_i(\theta) \geq 0 \text{ i } x_i^T u < v) + \#(r_i(\theta) \leq 0 \text{ i } x_i^T u > v) \}$$

gdzie $u = (u_1, \dots, u_{p-1})^T$ to wektory jednostkowe w przestrzeni R^{p-1} oraz v to liczby rzeczywiste takie, że $x_i^T u \neq v$ i $(x_i^T, y_i) \in Z_n$.

Głębina regresji może być wykorzystana przy szacowaniu parametrów modelu regresji.

Metoda najgłębszej regresji polega na wyznaczeniu parametrów regresji w taki sposób, aby uzyskanemu modelowi odpowiadała największa głębina regresyjna. Metoda ta została zaproponowana przez Rousseeuw'a i Hubert [1999].

Wartości parametru θ modelu (3) szacowane są na podstawie danych zawartych w zbiorze Z_n za pomocą wzoru:

$$T_r^*(Z_n) = \arg \max_{\theta} rdepth(\theta, Z_n). \quad (8)$$

W artykule Van Aelst, Rousseeuw, Hubert, Struyf [2000] przedstawiony jest algorytm szacowania parametrów modelu regresji wielorakiej metodą najgłębszej regresji.

IV. ANALIZA WOJEWÓDZTW ZE WZGLĘDU NA OGÓLNY STAN ZANIECZYSZCZENIA ŚRODOWISKA

Kompleksowa analiza stanu środowiska wymaga uwzględnienia różnych czynników wpływających na zanieczyszczenie powietrza, wód oraz gleby. Rozwój społeczno- gospodarczy może mieć negatywny wpływ na środowisko, dlatego konieczne jest podejmowanie działań zmierzających do zneutralizowania tego wpływu.

Pierwszy etap badania zmian, jakie zaszły w poziomie zanieczyszczania środowiska przed i po wstąpieniu Polski do Unii Europejskiej, stanowi dokonanie porównania sytuacji w 2006 r. w stosunku do 2003 r. (tuż przed wstąpieniem do UE) na podstawie danych ze strony internetowej www.stat.gov.pl, w oparciu o cztery cechy:

- emisja zanieczyszczeń pyłowych z zakładów szczególnie uciążliwych (t),
 - emisja zanieczyszczeń gazowych z zakładów szczególnie uciążliwych (t),
 - ścieki komunalne i wytworzone przez zakłady przemysłowe (dam³),
 - odpady wytworzone z wyłączeniem odpadów komunalnych (tys. t).
- Zmienne zostały oznaczone odpowiednio: emisp, emisg, ścieki, odpady.

Zmiana stanu środowiska w latach 2002–2006 spowodowana łącznym wpływem wszystkich czterech wielkości na stan środowiska w województwach została zbadana za pomocą głębi lokacyjnej. Dla każdego z rozpatrywanych lat oddzielnie określone zostały wartości współrzędnych punktów o największej głębi lokacyjnej $Me_{w_sr}^{(4)}$. Rezultaty zostały zaprezentowane w tabelicy 4.1. Na podstawie porównania współrzędnych mediany wielowymiarowej można zauważyć, że w analizowanym okresie wielkość emisji zanieczyszczeń pyłowych z zakładów szczególnie uciążliwych stopniowo malała od ponad 5,8 tys. t w 2002 r. do około 3,8 tys. t w 2006 r. Emisja zanieczyszczeń gazowych z zakładów szczególnie uciążliwych w latach 2002–2003 r. wynosiła ponad 6,6 mln t, jednak w roku 2004 nastąpił jej wzrost i w 2005 r. osiągnęła wartość 8,3 mln t, następnie nastąpił spadek i w 2006 r. odnotowano najniższą wartość emisji gazowej w całym analizowanym okresie niecałe 6,5 mln t. Ilość ścieków komunalnych i wytworzonych przez zakłady przemysłowe wykazywała tendencję malejącą i w analizowanym okresie spadła z 272 tys. dam^3 do 230 tys. dam^3 . Ilość wytworzonych odpadów (z wyłączeniem odpadów komunalnych) w latach 2002 – 2006 na przemian rosła i malała, wahając się od niecałych 2,4 mln t do ponad 3,3 mln t.

Uzyskane wyniki wskazują, że stan środowiska w 2006 r. uległ poprawie względem 2003 r., czyli tuż przed wstąpieniem Polski do Unii Europejskiej, gdyż wszystkie współrzędne punktu, będącego medianą wielowymiarową, przyjmowały niższe wartości dla 2006 r. w porównaniu z 2003 r. Poszczególne województwa charakteryzował jednak różny poziom badanych zmiennych. Kontury głębi dla ścieków komunalnych i wytworzonych przez zakłady przemysłowe oraz odpadów wytworzonych (z wyłączeniem odpadów komunalnych) dla lat 2003 i 2006 przedstawione są na rysunkach 4.1. i 4.2. Literami A, B, C, D, E zaznaczono na nich punkty reprezentujące województwa odpowiednio: dolnośląskie, śląskie, mazowieckie, wielkopolskie, zachodniopomorskie.

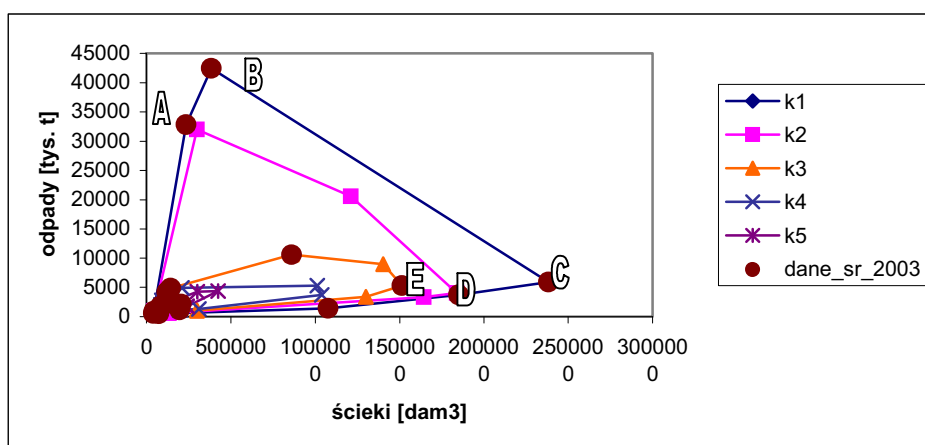
W badaniu zostało uwzględnione również inne podejście do przestrzennego zróżnicowania poziomu zanieczyszczenia środowiska. W celu doprowadzenia wartości poszczególnych zmiennych do porównywalności przeprowadzono unitaryzację danych. Następnie dla uzyskanego zbioru danych ponownie wyznaczono medianę wielowymiarową $Me_{w_sr_u}^{(4)}$ i obliczono odległości euklidesowe poszczególnych województw od tego punktu. Wyniki przedstawiono w tabelicy 4.2. Najbardziej oddalone od $Me_{w_sr_u}^{(4)}$ były województwa śląskie i mazowieckie. Najbliższe mediany wielowymiarowej było województwo pomorskie. Na rysunkach 4.3. i 4.4. zostały zaprezentowane odległości poszczególnych województw od median wielowymiarowych w latach 2003 i 2006. Wyodrębniono również województwa charakteryzujące się wartościami wszystkich badanych zmiennych większymi od odpowiadających im współrzędnych dla $Me_{w_sr}^{(4)}$

dla poszczególnych lat. W 2003 r. były to województwa śląskie, mazowieckie, małopolskie i wielkopolskie, natomiast w 2006 r. dołączyło do nich dolnośląskie. W obydwu analizowanych latach wszystkimi wartościami zmiennymi mniejszymi od odpowiadających im dla $Me_{w_sr}^{(4)}$ charakteryzowały się województwa pomorskie, warmińsko-mazurskie, podlaskie, podkarpackie i lubuskie.

Tablica 4.1. Wartości współrzędnych punktu $Me_{w_sr}^{(4)}$ dla województw dla lat 2002–2006

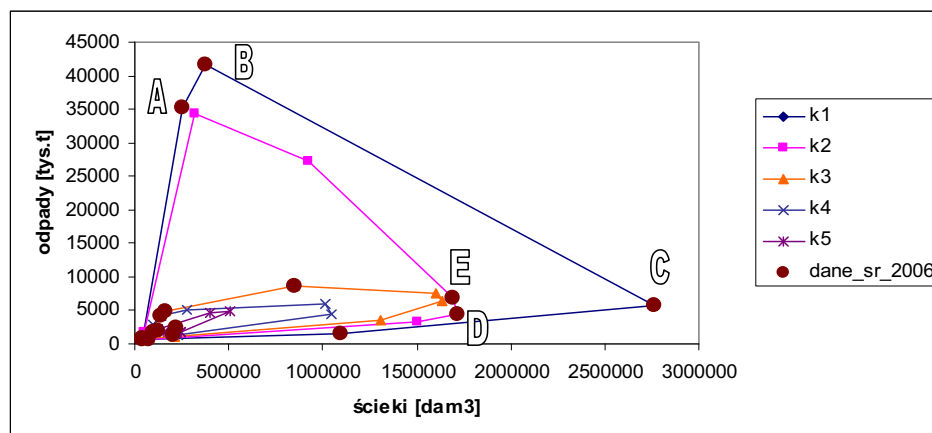
Rok	emisp [t]	emisg [t]	ścieki [dam ³]	odpady [tys.t]
2002	5839	6698620	272135	2395
2003	5687	6639591	245885	3317
2004	4798	7101778	237903	2712
2005	4666	8322843	232844	3223
2006	3772	6469745	230231	2872

Źródło: Obliczenia własne.



Rys. 4.1. Ilość ścieków komunalnych i wytworzonych przez zakłady przemysłowe oraz wielkość odpadów wytworzonych (z wyłączeniem odpadów komunalnych) w województwach dla 2003 r. (dane_sr_2003) oraz kontury głębokości (k1 – k5)

Źródło: Opracowanie własne.

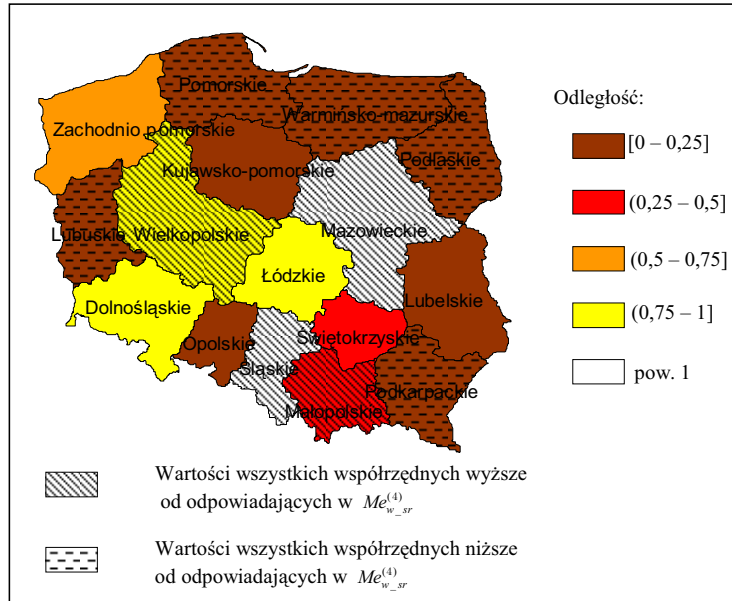


Rys. 4.2. Ilość ścieków komunalnych i wytworzonych przez zakłady przemysłowe oraz wielkość odpadów wytworzonych (z wyłączeniem odpadów komunalnych) w województwach dla 2006 r. (dane_sr_2006) oraz kontury głębi (k1 – k5)
 Źródło: Opracowanie własne.

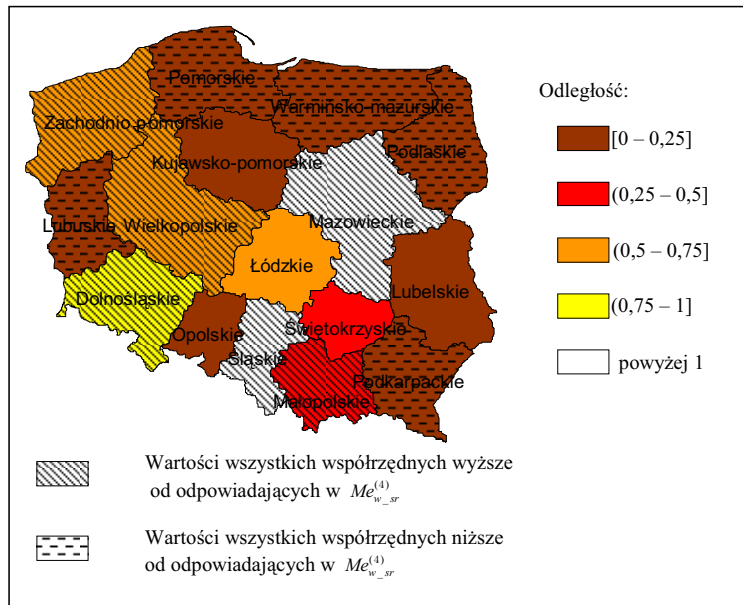
Tablica 4.2. Odległości od $Me_{w_sr_u}^{(4)}$ dla województw w latach 2002–2006 oraz odpowiadające im rangi

Województwo	2002		2003		2004		2005		2006	
	Odległość od $Me_{w_sr_u}^{(4)}$	Ranga	Odległość od $Me_{w_sr_u}^{(4)}$	Ranga	Odległość od $Me_{w_sr_u}^{(4)}$	Ranga	Odległość od $Me_{w_sr_u}^{(4)}$	Ranga	Odległość od $Me_{w_sr_u}^{(4)}$	Ranga
Dolnośląskie	0,76	5	0,77	4	0,79	3	0,83	3	0,87	3
Kujawsko-pomorskie	0,19	11	0,13	13	0,13	13	0,15	13	0,12	14
Lubelskie	0,08	15	0,08	15	0,10	15	0,16	12	0,13	13
Lubuskie	0,16	12	0,17	11	0,18	11	0,20	11	0,14	12
Łódzkie	0,88	3	0,90	3	0,73	4	0,76	4	0,69	5
Małopolskie	0,42	7	0,40	7	0,45	7	0,41	7	0,46	7
Mazowieckie	1,01	2	1,04	2	1,07	2	1,08	2	1,15	2
Opolskie	0,13	13	0,16	12	0,17	12	0,15	14	0,19	9
Podkarpackie	0,13	14	0,11	14	0,11	14	0,13	15	0,07	15
Podlaskie	0,20	10	0,19	10	0,20	10	0,22	9	0,16	11
Pomorskie	0,05	16	0,04	16	0,04	16	0,07	16	0,01	16
Śląskie	1,44	1	1,45	1	1,55	1	1,54	1	1,59	1
Świętokrzyskie	0,38	8	0,37	8	0,33	8	0,22	10	0,34	8
Warmińsko-mazurskie	0,21	9	0,20	9	0,21	9	0,23	8	0,17	10
Wielkopolskie	0,77	4	0,77	5	0,70	5	0,67	5	0,71	4
Zachodniopomorskie	0,60	6	0,54	6	0,51	6	0,47	6	0,56	6

Źródło: Obliczenia własne.

Rys. 4.3. Odległości od $Me_{w_sr_u}^{(4)}$ dla województw w 2003 r.

Źródło: opracowanie własne.

Rys. 4.4. Odległości od $Me_{w_sr_u}^{(4)}$ dla województw w 2006 r.

Źródło: opracowanie własne.

Kolejnym etapem badania poziomu zanieczyszczenia środowiska było porównanie województw pod względem skuteczności działań mających na celu redukcję negatywnego wpływu oddziaływania człowieka na środowisko w latach 2002–2006. Do analizy wykorzystano następujące zmienne:

- udział zanieczyszczeń pyłowych zatrzymanych lub zneutralizowanych w urządzeniach do redukcji w całości wytworzonych zanieczyszczeń pyłowych w zakładach szczególnie uciążliwych (%),
- udział zanieczyszczeń gazowych zatrzymanych lub zneutralizowanych w urządzeniach do redukcji w całości wytworzonych zanieczyszczeń gazowych w zakładach szczególnie uciążliwych (%),
- udział ścieków przemysłowych i komunalnych oczyszczanych w ściekach wymagających oczyszczenia (%),
- udział odpadów poddanych odzyskowi w ilości odpadów wytworzonych w ciągu roku z wyłączeniem odpadów komunalnych (%).

Zmienne zostały oznaczone odpowiednio: z_emisp , z_emisg , $o_ścieki$, o_odpady .

Analiza efektywności zapobiegania zanieczyszczeniu środowiska przeprowadzona została analogicznie do poprzedniej za pomocą głębi lokacyjnej. Dla każdego z rozpatrywanych lat wyznaczono punkt o największej głębi lokacyjnej $Me_{w_rz}^{(4)}$. Współrzędne median wielowymiarowych przedstawiono w tablicy 4.3.

Na podstawie wartości tych współrzędnych można zauważyć, że w latach 2002 – 2006 zanieczyszczenia pyłowe były zatrzymane lub zneutralizowane prawie całkowicie (w około 99%), natomiast gazowe jedynie w około 0,5% w latach 2004–2006, a w 2002 r. i 2003 r. w jeszcze mniejszym stopniu. Udział ścieków oczyszczanych w wymagających oczyszczenia wynosił ponad 92%, natomiast udział odzyskanych odpadów w ogólnej ich ilości (z wyłączeniem odpadów komunalnych) wahał się od około 72% do ponad 76%. Wartości wszystkich współrzędnych dla 2006 r. były wyższe niż dla 2003 r. Wartości zmiennych w poszczególnych województwach były zróżnicowane. Kontury głębi dla udziału ścieków przemysłowych i komunalnych oczyszczanych w ściekach wymagających oczyszczenia oraz udziału odpadów poddanych odzyskowi w ilości odpadów wytworzonych w ciągu roku z wyłączeniem odpadów komunalnych dla województw w latach 2003 i 2006 przedstawione zostały na rysunkach 4.5. i 4.6. Literami A, B, C zaznaczono na nich punkty reprezentujące województwa odpowiednio: śląskie, mazowieckie, zachodniopomorskie.

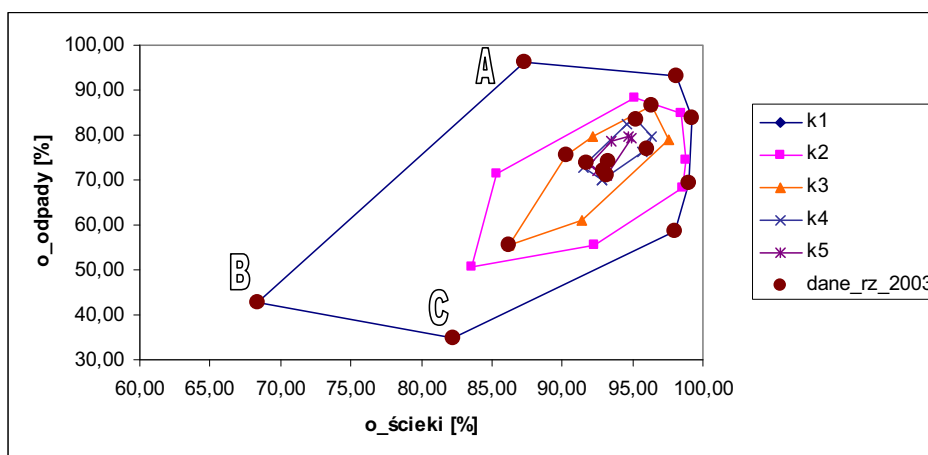
Efektywność zapobiegania zanieczyszczeniu środowiska została również poddana innego typu analizie przestrzennej. Wartości zmiennych doprowadzono do porównywalności poprzez unitaryzację. Następnie wyznaczono dla otrzymanego zbioru danych punkt o największej głębi lokacyjnej $Me_{w_rz_u}^{(4)}$ i obliczono odległości euklidesowe poszczególnych województw od tego punktu. Wyniki zaprezentowano w tablicy 4.4. Najbardziej odległe od $Me_{w_rz_u}^{(4)}$ były woje-

wództwa mazowieckie i dolnośląskie, a najbliższe medianie wielowymiarowej pomorskie i podkarpackie. Na rysunkach 4.7. i 4.8. zostały zaprezentowane odległości poszczególnych województw od median wielowymiarowych w latach 2003 i 2006. Wyodrębniono również województwa charakteryzujące się wartościami wszystkich badanych zmiennych większymi od odpowiadających im w $Me_{w_rz}^{(4)}$ dla poszczególnych lat. W 2003 r. były to łódzkie, opolskie i dolnośląskie, w 2006 r. już tylko opolskie. Wszystkimi wartościami zmiennych mniejszymi od $Me_{w_rz}^{(4)}$ w 2003 r. nie charakteryzowało się żadne województwo, a w 2006 r. – zachodniopomorskie i lubuskie.

Tablica 4.3. Wartości współrzędnych punktu $Me_{w_rz}^{(4)}$ dla województw dla lat 2002–2006

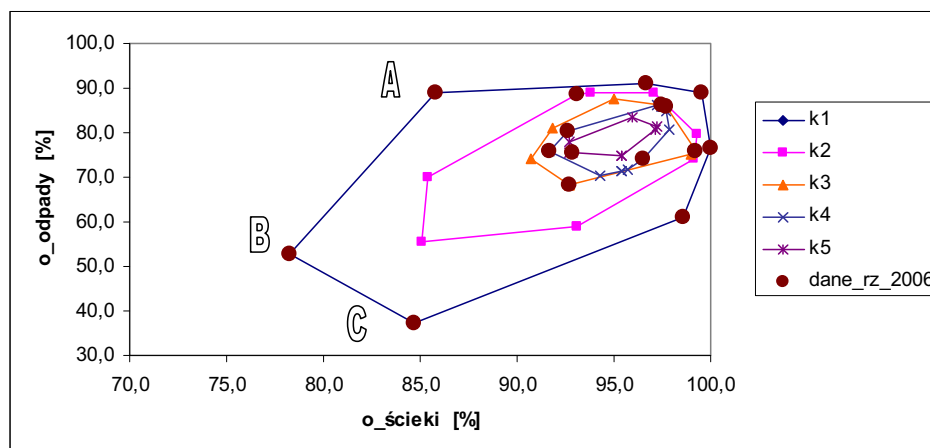
Rok	z_emisp [%]	z_emisg [%]	o_ścieki [%]	o_odpady [%]
2002	98,61	0,45	93,84	74,58
2003	98,67	0,35	92,09	73,93
2004	98,65	0,49	92,39	71,71
2005	98,94	0,49	92,88	76,19
2006	99,02	0,51	93,75	76,40

Źródło: Obliczenia własne.



Rys. 4.5. Udział ścieków przemysłowych i komunalnych oczyszczanych w ściekach wymagających oczyszczenia oraz udział odpadów poddanych odzyskowi w ilości odpadów wytworzonych w ciągu roku z wyłączeniem odpadów komunalnych w województwach dla 2003 r. (dane_rz_2003) oraz kontury głębokości (k1 – k5)

Źródło: Opracowanie własne.



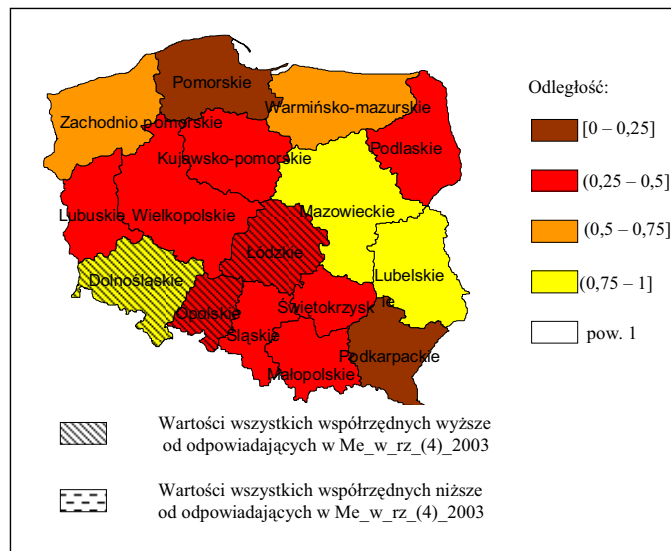
Rys. 4.6. Udział ścieków przemysłowych i komunalnych oczyszczanych w ściekach wymagających oczyszczenia oraz udział odpadów poddanych odzyskowi w ilości odpadów wytworzonych w ciągu roku z wyłączeniem odpadów komunalnych w województwach dla 2006 r. (dane_rz_2006) oraz kontury głębi (k1 – k5)

Źródło: Opracowanie własne.

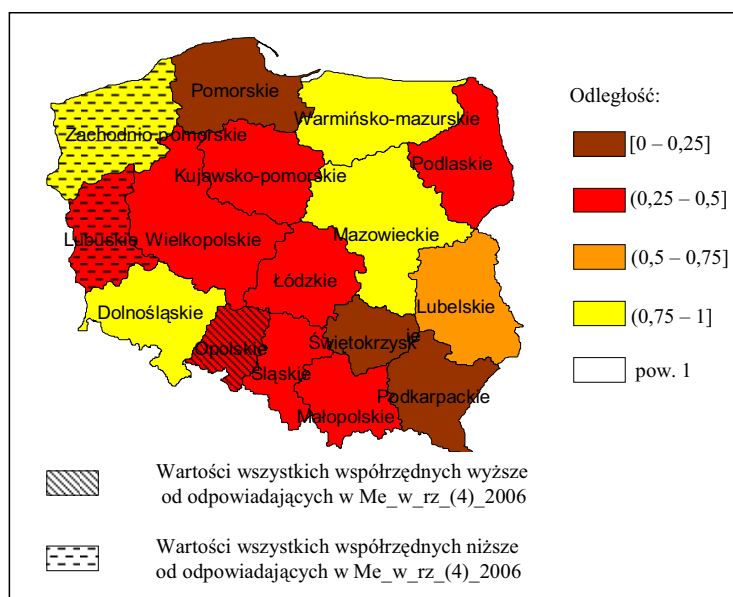
Tablica 4.4. Odległości od $Me_{w_rz_u}^{(4)}$ dla województw w latach 2002–2006 oraz odpowiadające im rangi

Województwo	2002		2003		2004		2005		2006	
	Odległość od $Me_{w_rz_u}^{(4)}$	Ranga	Odległość od $Me_{w_rz_u}^{(4)}$	Ranga	Odległość od $Me_{w_rz_u}^{(4)}$	Ranga	Odległość od $Me_{w_rz_u}^{(4)}$	Ranga	Odległość od $Me_{w_rz_u}^{(4)}$	Ranga
Dolnośląskie	0,94	2	0,96	1	0,94	1	0,93	1	0,94	1
Kujawsko-pomorskie	0,38	7	0,26	14	0,23	15	0,25	14	0,30	11
Lubelskie	0,50	6	0,78	3	0,75	3	0,90	3	0,75	5
Lubuskie	0,60	5	0,33	9	0,28	13	0,35	8	0,43	7
Łódzkie	0,30	10	0,28	12	0,35	9	0,30	11	0,28	13
Małopolskie	0,21	14	0,26	13	0,31	12	0,32	10	0,28	12
Mazowieckie	0,95	1	0,93	2	0,82	2	0,92	2	0,84	2
Opolskie	0,28	11	0,34	8	0,40	7	0,36	7	0,40	9
Podkarpackie	0,13	15	0,06	16	0,25	14	0,24	15	0,25	14
Podlaskie	0,21	13	0,29	11	0,34	10	0,34	9	0,40	8
Pomorskie	0,12	16	0,12	15	0,04	16	0,04	16	0,10	16
Śląskie	0,36	9	0,45	6	0,42	6	0,39	6	0,46	6
Świętokrzyskie	0,36	8	0,41	7	0,37	8	0,27	12	0,24	15
Warmińsko-mazurskie	0,73	4	0,75	4	0,75	5	0,77	5	0,79	4
Wielkopolskie	0,24	12	0,30	10	0,31	11	0,26	13	0,37	10
Zachodniopomorskie	0,80	3	0,70	5	0,75	4	0,78	4	0,84	3

Źródło: Obliczenia własne.



Rys. 4.7. Odległości od $Me_{w_rz}^{(4)}$ ($Me_{w_rz}(4)_{2003}$) dla województw w 2003 r.
 Źródło: opracowanie własne



Rys. 4.8. Odległości od $Me_{w_sr}^{(4)}$ ($Me_{w_rz}(4)_{2006}$) dla województw w 2006 r.
 Źródło: opracowanie własne.

V. ANALIZA WOJEWÓDZTW ZE WZGLĘDU NA GOSPODARKE ODPADAMI

Wytwarzanie odpadów jest jednym z głównych przejawów ingerencji człowieka w środowisko naturalne. Odpady, które nie zostały poddawane odzyskowi, wymagają składowania na odpowiednio dużych powierzchniach przeznaczonych na ten cel.

W analizie gospodarki odpadami (z wyłączeniem odpadów komunalnych) w Polsce przed i po wstąpieniu naszego kraju do Unii Europejskiej wykorzystane zostały dwa modele pozwalające na określenie ilości wytwarzanych odpadów oraz niezrekultywowanej powierzchni składowisk, na której mogą być one gromadzone.

Pierwszy z modeli opisuje zależność ilości wytworzonych odpadów (z wyłączeniem odpadów komunalnych) od liczby ludności i liczby dużych przedsiębiorstw w poszczególnych województwach i ma postać:

$$\text{odpady}_i = a_1 * \text{lud}_i + a_2 * \text{przedd}_i + a_3 + \varepsilon_i, \quad (9)$$

gdzie odpady_i – odpady wytworzone (tys. t) w i -tym województwie (z wyłączeniem odpadów komunalnych), lud_i – ludność zameldowana na stałe na dzień 31 XII w i -tym województwie, przedd_i – liczba dużych przedsiębiorstw w i -tym województwie, a_1, a_2, a_3 , – parametry modelu, ε_i – składnik losowy.

Analiza została przeprowadzona na poziomie województw dla lat 2003 – 2006, oddzielnie dla każdego roku, na podstawie danych ze strony internetowej www.stat.gov.pl.

W zbiorach danych wystąpiły obserwacje odstające, a więc do oszacowania parametrów modelu (9) zastosowano metodę najgłębszej regresji, czyli metodę odporną na obserwacje nietypowe.

Model opisujący ilości wytworzonych odpadów ma postać:

$$\text{dla 2003 r.: } \text{odpady}_i = 0,00262 * \text{lud}_i + 7,17 * \text{przedd}_i - 3020, \quad (10)$$

$$\text{dla 2004 r.: } \text{odpady}_i = 0,00242 * \text{lud}_i + 11,6 * \text{przedd}_i - 3230, \quad (11)$$

$$\text{dla 2005 r.: } \text{odpady}_i = 0,00283 * \text{lud}_i + 11,1 * \text{przedd}_i - 3610, \quad (12)$$

$$\text{dla 2006 r.: } \text{odpady}_i = 0,00277 * \text{lud}_i + 7,25 * \text{przedd}_i - 2910, \quad (13)$$

Do analizy podobieństwa struktury zjawiska w czasie zaproponowano wykorzystanie mediany wielowymiarowej. Na podstawie parametrów modeli (10) – (13) utworzony został zbiór punktów o współrzędnych (a_{1k}, a_{2k}, a_{3k}) odpowiadających wartościom parametrów, gdzie $k = 2003, \dots, 2006$. W celu doprowadzenia do porównywalności współrzędne punktów zostały poddane unitaryzacji. Na

podstawie uzyskanego zbioru wyznaczono współrzędne punktu o największej głębi lokacyjnej $Me_{w_o}^{par(3)}$ i obliczono odległość od mediany wielowymiarowej dla każdego z punktów (por. tablica 5.1.). Za odległość przyjęto dla danego roku największą co do modułu różnicę między odpowiadającymi sobie współrzędnymi danego punktu i $Me_{w_o}^{par(3)}$. Analiza tych odległości pozwala zauważyć, że spośród analizowanych lat najbardziej podobne pod względem struktury badanego zjawiska były lata 2003 i 2006, których odległości od mediany były najmniejsze. Wyraźna zmiana w strukturze nastąpiła w latach 2004 i 2005, dla których odległości od $Me_{w_o}^{par(3)}$ były znacznie większe od pozostałych.

Tablica 5.1. Parametry modeli (10)-(13) oraz odległości od $Me_{w_o}^{par(3)}$ dla odpowiadających im wartości zunitaryzowanych

Rok k	Parametr			Odległość od $Me_{w_o}^{par(3)}$
	a_{1k}	a_{2k}	a_{3k}	
2003	0,00262	7,17	-3020	0,108
2004	0,00242	11,6	-3230	0,892
2005	0,00283	11,1	-3610	0,826
2006	0,00277	7,25	-2910	0,303

Źródło: Obliczenia własne.

Drugi z modeli opisujących gospodarkę odpadami w województwach odzwierciedla zależność między niezrekultywowaną powierzchnią składowisk a liczbą zakładów wytwarzających odpady i ilością wytworzonych odpadów. Ma on postać:

$$\text{powsk}_i = b_1 * \text{zakl}_i + b_2 * \text{odpady_sk}_i + b_3 + \varepsilon_i \quad (14)$$

powsk_i – niezrekultywowana powierzchnia składowania odpadów (z wyłączeniem odpadów komunalnych) w i -tym województwie, zakl_i – zakłady wytwarzające odpady w i -tym województwie, odpady_sk_i – odpady wytworzone w ciągu roku składowane na składowiskach własnych i innych w i -tym województwie (z wyłączeniem odpadów komunalnych), b_1 , b_2 , b_3 – parametry modelu, ε_i – składnik losowy.

Analiza została przeprowadzona na poziomie województw dla lat 2001–2005, oddzielnie dla każdego roku, na podstawie danych ze strony internetowej www.stat.gov.pl.

Ze względu na występowanie w zbiorze danych obserwacji nietypowych oszacowania parametrów modelu dokonano metodą najgłębszej regresji.

Model opisujący ilości wytworzonych odpadów ma postać:

$$\text{dla 2001 r.: powsk}_i = 2,870 * \text{zakl}_i + 0,180 * \text{odpady_sk}_i + 5,72, \quad (15)$$

$$\text{dla 2002 r.: powsk}_i = 2,980 * \text{zakl}_i + 0,237 * \text{odpady_sk}_i + 23,2, \quad (16)$$

$$\text{dla 2003 r.: powsk}_i = 6,060 * \text{zakl}_i + 0,183 * \text{odpady_sk}_i - 189, \quad (17)$$

$$\text{dla 2004 r.: powsk}_i = 3,840 * \text{zakl}_i + 0,286 * \text{odpady_sk}_i - 195, \quad (18)$$

$$\text{dla 2005 r.: powsk}_i = 0,921 * \text{zakl}_i + 0,459 * \text{odpady_sk}_i - 36,8. \quad (19)$$

Analiza podobieństwa struktury zjawiska w czasie została przeprowadzona analogicznie jak w poprzednim modelu przy pomocy miary głębi lokacyjnej. Na podstawie parametrów modeli (15) – (19) utworzony został zbiór punktów o współrzędnych (b_{1l}, b_{2l}, b_{3l}) odpowiadających wartościom parametrów, gdzie $l = 2001, \dots, 2005$. Następnie współrzędne punktów zostały poddane unitaryzacji. Na podstawie uzyskanego zbioru wyznaczono współrzędne punktu o największej głębi lokacyjnej $Me_{w_sk}^{par(3)}$ i obliczono odległość od mediany wielowymiarowej dla każdego z punktów (por. tablica 5.2.). Za odległość przyjęto dla danego roku największą co do modułu różnicę między odpowiadającymi sobie współrzędnymi danego punktu i $Me_{w_o}^{par(3)}$. Analiza tych odległości pozwala zauważyć, że spośród analizowanych lat najbardziej podobne pod względem struktury badanego zjawiska były lata 2001 i 2002, których odległości od mediany były najmniejsze. Zmiana w strukturze zjawiska nastąpiła w latach 2003 i 2004, dla których odległości od $Me_{w_o}^{par(3)}$ były większe od poprzednich, natomiast w roku 2005 struktura zjawiska nie była podobna do tej z pozostałych lat.

Tablica 5.2. Parametry modeli (15)–(19) oraz odległości od $Me_{w_sk}^{par(3)}$ dla odpowiadających im wartości zunitaryzowanych

Rok l	Parametr			Odległość od $Me_{w_sk}^{par(3)}$
	b_{1l}	b_{2l}	b_{3l}	
2001	2,870	0,180	5,72	0,261
2002	2,980	0,237	23,20	0,341
2003	6,060	0,183	-189,00	0,631
2004	3,840	0,286	-195,00	0,659
2005	0,921	0,459	-36,80	0,772

Źródło: Obliczenia własne.

Na podstawie otrzymanych wyników można zauważyć, że gospodarka odpadami w Polsce w 2003 r., czyli tuż przed wstąpieniem do Unii Europejskiej pod względem zależności ilości wytworzonych odpadów (z wyłączeniem odpadów komunalnych) od liczby ludności i liczby dużych przedsiębiorstw w poszczególnych województwach była podobna pod względem struktury zjawiska do sytuacji w 2006 r. Natomiast zależność między niezrekultywowaną powierzchnią składowisk a liczbą zakładów wytwarzających odpady i ilością wytworzonych odpadów (z wyłączeniem odpadów komunalnych) w latach 2003 i 2005 miała inny charakter.

Wartości największej głębi lokacyjnej, granice konturów głębi lokacyjnej zaprezentowane w tej pracy zostały obliczone na podstawie programów odpowiednio *Deeplock* i *Isodepth* (por. Ruts, Rousseeuw (1996), Struyf, Rousseeuw (2000)), natomiast parametry modelu oszacowano za pomocą programu *Med-sweep*. Algorytmy programów zamieszczone są na stronie internetowej www.agoras.ua.ac.be.

VI. UWAGI KOŃCOWE

Analiza poziomu zanieczyszczenia środowiska w poszczególnych województwach przeprowadzona za pomocą głębi lokacyjnej i regresyjnej umożliwiła wskazanie kierunków i zmian strukturalnych, jakie zaszły na obszarze naszego kraju od roku 2001 do 2006. Zastosowanie do szacowania parametrów modelu metody najgłębszej regresji, odpornej na występowanie w zbiorze danych obserwacji nietypowych, pozwoliło uzyskać wyniki, które nie są obciążone błędami wynikającymi z nierównomiernego nasilenia przestrzennego analizowanych zjawisk. Zaproponowane w pracy wykorzystanie mediany wielowymiarowej do analizy zmian parametrów regresji uzyskanych dla tej samej postaci modelu dla różnych lat, może być stosowane w przypadkach różnych zjawisk ekonomicznych.

BIBLIOGRAFIA

- Hubert M., Rousseeuw P. J., Van Aelst S. (1999), *Similarities between location depth and regression depth*, strona internetowa www.agoras.ua.ac.be
- Rousseeuw P. J., Hubert M. (1999) *Regression Depth*, *JASA* 94, 388–402
- Rousseeuw P. J., Ruts I. [1998], *Constructing the bivariate Tuckey median*, *Statistica Sinica* 8, 827–839
- Ruts I., Rousseeuw P. J. (1996), *Computing depth contours of bivariate point clouds*, *Computational Statistics & Data Analysis* 23, 153–168
- Struyf A., Rousseeuw P. J. (2000), *High-dimensional computation of the deepest location*, *Computational Statistics & Data Analysis* 34, 415–426
- Van Aelst S., Rousseeuw P. J., Hubert M., Struyf A. (2002), *The Deepest Regression Method*, *Journal of Multivariate Analysis*, 81, 138–166.

Dorota Pruska

**APPLICATION OF LOCATION DEPTH AND REGRESSION DEPTH
IN ANALYSIS OF LEVEL OF ENVIRONMENT POLLUTION IN POLAND**

Abstract

In the paper the comparison of the level of environment pollution in voivodships was conducted for Poland before and after the accession to the European Union. The measures of location and regression depth were used in the analysis. The classification of voivodships was conducted using the multivariate median. The usage of the multivariate median was also proposed in analysis of structural changes of phenomenon in time. The regression coefficients were estimated by the deepest regression method in order to eliminate the influence of outliers.