

Tomasz Kossowski

Uniwersytet im. Adama Mickiewicza w Poznaniu

Wydział Geografii Społeczno-Ekonomicznej i Gospodarki Przestrzennej

tkoss@amu.edu.pl,  <https://orcid.org/0000-0002-9976-4398>

Zastosowanie metody przestrzenno-czasowych składowych głównych (STPCA) do identyfikacji przestrzennego zróżnicowania społeczno-gospodarczego powiatów

Zarys treści: Celem artykułu jest identyfikacja przestrzennego zróżnicowania społeczno-gospodarczego powiatów w latach 2005–2024. Wykorzystano zbiór 30 zmiennych uwzględniających sferę demograficzną, gospodarczą, stan infrastruktury oraz sferę społeczno-środowiskową. Do redukcji wymiaru obserwacji zastosowano przestrzenno-czasową metodę składowych głównych dla danych funkcjonalnych. Wyodrębnienia typów powiatów dokonano z kolei metodą analizy skupień, przeprowadzoną osobno dla dodatnich i ujemnych składowych głównych. W jej wyniku otrzymano dwa podziały na 3 typy powiatów oraz jeden na 9 typów. Uzyskana typologia potwierdza ciągle występujące zróżnicowanie powiatów ze względu na ich położenie na obszarze dawnych państw za-borczych, dychotomię pomiędzy powiatami grodzkimi a ziemskimi, a także lokalizację względem obszarów metropolitalnych.

Słowa kluczowe: przestrzenno-czasowe składowe główne, autokorelacja przestrzenna, zróżnicowanie przestrzenne, typologia powiatów, procesy aglomeracyjne

Wprowadzenie

Badania zróżnicowania społeczno-gospodarczego jednostek terytorialnych różnego poziomu podziału administracyjnego oraz rozkładu przestrzennego tego zróżnicowania mają długą tradycję geograficzną. Zazwyczaj powiązane są z problematyką rozwoju społeczno-gospodarczego jednostek terytorialnych, który charakteryzuje się zróżnicowaniem regionalnym (Churski 2008). Przedmiotem zainteresowania badaczy są również zazwyczaj czynniki wpływające na ten rozwój. Jak zauważają Józefowicz i Smolińska (2019), powodują one zmiany o charakterze ilościowym (np. w poziomie bezrobocia, w poziomie produkcji przemysłowej) oraz jakościowym (zachowania, zmiany hierarchii).

Opracowania dotyczące rozwoju lub zróżnicowania społeczno-gospodarczego jednostek terytorialnych mają głównie charakter ilościowy. Stosowane są zarówno ujęcia ograniczone do pojedynczej obserwacji w jednostce przestrzennej, jak i dynamiczne, choć z reguły ograniczone w czasie do dwóch-trzech obserwacji. Podział jednostek na grupy pod względem poziomu rozwoju dokonywany jest zazwyczaj za pomocą metody Hellwiga (1968), wskaźnika Perkala, a rzadziej z wykorzystaniem innego miernika syntetycznego. Z przykładowych podejść statycznych należy tutaj wymienić pracę Piszczka (2013) dotyczącą zróżnicowania przestrzennego poziomu rozwoju obszarów wiejskich w województwie kujawsko-pomorskim. Autor wykorzystał 10 zmiennych charakteryzujących obszary gmin wiejskich i metodę Hellwiga do ustalenia poziomu ich rozwoju. Dla gmin wiejskich województwa wielkopolskiego poziom wielofunkcyjnego poziomu ich rozwoju badał Wojtyra (2017). Wykorzystał 10 zmiennych dla roku 2016, a następnie przy wykorzystaniu metody Hellwiga wyodrębnił 4 klasy rozwojowe. Większego zbioru zmiennych, liczącego 38 cech podzielonych na 6 grup, użyła Bartkowiak-Bakun (2015) dla gmin województwa wielkopolskiego. Wyodrębniła ona za pomocą syntetycznego wskaźnika rozwoju 4 klasy gmin.

Wśród ujęć dynamicznych interesująca jest praca Dziekańskiego (2018), który na podstawie 25 zmiennych buduje syntetyczne miary sytuacji finansowej i rozwoju powiatów Polski Wschodniej, a następnie porównuje ich wartości dla lat 2010, 2012, 2016. Natomiast Siudek i in. (2017) posłużyli się 9 zmiennymi opisującymi gospodarkę powiatów i podregionów województwa mazowieckiego dla lat 2005 i 2014. Następnie porównali wartości wskaźnika rozwoju uzyskanego metodą Hellwiga i przeanalizowali jego zmiany w badanych jednostkach terytorialnych. Józefowicz i Smolińska (2018) oceniały poziom rozwoju społeczno-gospodarczego powiatów województwa wielkopolskiego. Wykorzystały 24 zmienne, które uśredniały dla lat 2005–2007 oraz 2015–2017, i wyznaczyły syntetyczny miernik rozwoju Hellwiga, a następnie dokonały analizy porównawczej powiatów.

Przytoczone powyżej prace nie uwzględniały podejścia przestrzennego biorącego pod uwagę możliwość wystąpienia zależności przestrzennych pomiędzy obserwacjami, a ponadto ustalały poziom rozwoju dla jednego lub kilku wybranych lat. Wyjątkiem jest tutaj praca Bartkowiak-Bakun (2015), która wykorzystała współczynnik *I* Morana (1950) do identyfikacji autokorelacji przestrzennej zróżnicowania rozwoju społeczno-gospodarczego, a następnie posługując się lokalną statystyką Morana (Anselin 1995), wyodrębniła klastry zależności przestrzennej. Z kolei Pietrzak (2014) wykorzystał przestrzenny taksonomiczny miernik rozwoju (pTMR) do analizy poziomu rozwoju gospodarczego 66 podregionów Polski w 2011 roku. Miernik ten wykorzystuje metodę Hellwiga, ale zmienne wykazujące autokorelację przestrzenną są w wyniku przekształcenia pozbawiane efektu przestrzennej zależności.

Celem pracy jest identyfikacja zróżnicowania typów rozwoju społeczno-gospodarczego powiatów w latach 2005–2024. Do realizacji tego celu wykorzystano metodę przestrzenno-czasowych składowych głównych dla danych funkcjonalnych, zaproponowaną przez Krzyśkę i in. (2024). Stanowi ona rozszerzenie

klasycznej metody składowych głównych Hotellinga (1933), uwzględniając aspekt przestrzenny (autokorelację przestrzenną zmiennych) oraz dynamikę (czas).

Punktem wyjścia jest wielokrotnie wykorzystywana w geografii (Parysek, Ratajczak 1978, Maćkiewicz, Ratajczak 1993) klasyczna metoda składowych głównych. Wymiar przestrzenny został wprowadzony przez Jombarta i in. (2008) i jest znany jako metoda przestrzennych składowych głównych. Analizę składowych głównych dla wielowymiarowych danych funkcjonalnych rozwijali Górecki i in. (2012, 2014, 2018) oraz Krzyśko i in. (2022). Metoda przestrzenno-czasowych składowych głównych łączy ze sobą te trzy podejścia. Zawiera także wymiar czasowy, przy czym nie dokonuje porównania pomiędzy obserwacjami początkowymi i końcowymi, ale niejako uwzględnia zmiany zachodzące w całym analizowanym okresie. Podobnie jak klasyczny wariant, służy do redukcji wymiaru przestrzeni obserwacji i jest szczególnie użyteczna, gdy wykorzystywany jest duży zbiór zmiennych. Należy jednak zauważyć, że jej stosowanie jest sensowne, gdy wariancja objaśniana przez ustaloną liczbę czasowo-przestrzennych składowych głównych jest większa od wariancji objaśnianej przez tę samą liczbę składowych wyodrębnionych tylko z uwzględnieniem wymiaru czasowego (Krzyśko i in. 2024).

Dane i metody

Wykorzystane w pracy dane pochodziły z Banku Danych Lokalnych Głównego Urzędu Statystycznego (Statistics Poland). Z bazy danych pobrano 51 zmiennych w układzie powiatowym dla lat 2005–2024, które następnie poddano analizie pod kątem braków w przekrojach czasowych i przestrzennych, oraz wyeliminowano zmienne mogące powielać informacje zawarte również w innych zmiennych. Ostateczny zbiór danych w układzie 380 powiatów dla lat 2005–2024 zawierał następujące zmienne: X_1 – gęstość zaludnienia w osobach na 1 km², X_2 – zmiana liczby ludności na 1000 mieszkańców, X_3 – ludność w wieku nieprodukcyjnym na 100 osób w wieku produkcyjnym, X_4 – ludność w wieku poprodukcyjnym na 100 osób w wieku produkcyjnym, X_5 – współczynnik obciążenia demograficznego osobami starszymi, X_6 – przyrost naturalny na 1000 ludności, X_7 – współczynnik feminizacji, X_8 – dochody budżetów powiatów na 1 mieszkańca w złotych, X_9 – wydatki budżetów powiatów na 1 mieszkańca w złotych, X_{10} – podmioty wpisane do rejestru REGON na 1000 ludności, X_{11} – produkcja sprzedana przemysłu (w przedsiębiorstwach powyżej 9 pracujących) na 1 mieszkańca w złotych, X_{12} – stopa bezrobocia rejestrowanego w procentach, X_{13} – przeciętne miesięczne wynagrodzenia brutto w złotych, X_{14} – przeciętna powierzchnia użytkowa mieszkania na 1 osobę w m², X_{15} – odsetek ludności korzystającej z wodociągu, X_{16} – odsetek ludności korzystającej z kanalizacji, X_{17} – odsetek ludności korzystającej z sieci gazowej, X_{18} – gęstość sieci wodociągowej w kilometrach na 100 km², X_{19} – gęstość sieci kanalizacyjnej w kilometrach na 100 km², X_{20} – gęstość sieci gazowej w kilometrach na 100 km², X_{21} – gęstość sieci dróg gminnych i powiatowych o twardej nawierzchni w kilometrach na 100 km², X_{22} – ludność na 1 placówkę biblioteczną (łącznie z punktami bibliotecznymi ujętymi zgodnie z siedzibą

jednostki macierzystej), X_{23} – księgozbiór bibliotek na 1000 ludności, X_{24} – łóżka w szpitalach ogólnych na 10 000 mieszkańców, X_{25} – mieszkania oddane do użytkowania na 1000 ludności, X_{26} – przeciętna powierzchnia użytkowa 1 mieszkania oddanego do użytkowania w m^2 , X_{27} – dzieci w przedszkolach i innych formach wychowania przedszkolnego na 1000 dzieci w wieku 3–5 lat, X_{28} – zużycie wody na 1 mieszkańca w m^3 , X_{29} – udział procentowy ścieków przemysłowych i komunalnych oczyszczanych w ściekach wymagających oczyszczania, X_{30} – udział procentowy ludności korzystającej z oczyszczalni ścieków w ogólnej liczbie ludności.

Wyżej wymieniony zbiór zmiennych podzielono na cztery grupy o podobnej liczebności: 1) X_1 – X_7 – zmienne charakteryzujące potencjał demograficzny powiatów, 2) X_8 – X_{13} – zmienne opisujące gospodarkę (w tym samorządową) i rynek pracy powiatów, 3) X_{14} – X_{21} – zmienne infrastrukturalne i 4) X_{22} – X_{30} – zmienne dotyczące społeczeństwa i środowiska. Dla potrzeb dalszych analiz zbiorowi zmiennych nadano formę $T = 20$ macierzy o wymiarach $n \times p$, gdzie $n = 380$ powiatów, $p = 30$ zmiennych. Ze względu na duże zróżnicowanie wartości zmiennych, dokonano ich przekształcenia do przedziału $[1, 7]$ (nawiązującego do skali Lickerta).

W celu redukcji wymiaru zbioru danych wykorzystano metodę przestrzenno-czasowych składowych głównych. Metoda ta została zaproponowana przez Krzyśkę i in. (2024) i łączy podejście do przestrzennych składowych głównych (SPCA) wprowadzone przez Jombarta i in. (2008) oraz wielozmiennych funkcjonalnych składowych głównych (Górecki i in. 2014, 2018, Krzyśko i in. 2022). Jak piszą Krzyśko i in. (2024), innowacyjność metody polega na tym, że uwzględniając momenty czasowe obserwacji zmiennych w kontekście danych funkcjonalnych, jednocześnie bierze pod uwagę lokalizację obserwacji jak SPCA (Jombart i in. 2008) oraz podejście do zarządzania wariancją w obserwowanych wartościach zmiennych jak w klasycznej analizie składowych głównych (Hotelling 1933). Precyzyjny matematyczny opis tej metody zawarty jest w pracy Krzyśki i in. (2024) i dlatego poniżej przedstawiony będzie tylko jej ogólny zarys.

W pierwszym kroku oryginalne zmienne zostają przekształcone do ciągłych funkcji. Niech x_{kj} będzie obserwacją zmiennej X_k w chwili t_j ($k = 1, 2, \dots, p$, $j = 1, 2, \dots, T$). Dane te następnie wygładza się przy wykorzystaniu ciągłych funkcji

$x_k(t) = \sum_{b=0}^{B_k} \alpha_{kb} \psi_b(t)$, gdzie $\alpha_{k0}, \alpha_{k1}, \dots, \alpha_{kB_k}$ są nieznanymi losowymi współczynnikami, a ψ_b , $b = 0, 1, \dots, B_k$ są ortonormalnymi funkcjami bazy, $t \in [0, T]$, przy czym $t_j \in [0, T]$, $j = 1, 2, \dots, T$. Współczynniki α_{kb} szacuje się metodą najmniejszych kwadratów (Krzyśko i in. 2024, równanie 3). Górecki i Krzyśko (2012) sugerują użycie bazy Fouriera, określonej następująco:

$$\{\psi_0(t) = 1/\sqrt{T}, \psi_{2b-1}(t) = \sqrt{2/T} \sin(2\pi bt/T), \psi_{2b}(t) = \sqrt{2/T} \cos(2\pi bt/T)\}.$$

Przyjmując

$$\mathbf{x}(t) = (x_1(t), \dots, x_p(t))^T, \boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_p^T)^T, \boldsymbol{\Psi}(t) = \text{diag}(\boldsymbol{\psi}_1^T(t), \dots, \boldsymbol{\psi}_p^T(t)),$$

$$\boldsymbol{\Psi}_k^T(t) = (\psi_{k0}(t), \psi_{k1}(t), \dots, \psi_{kB_k}(t)),$$

można zapisać wielozmienne dane funkcjonalne w postaci $\mathbf{x}(t) = \boldsymbol{\Psi}(t)\boldsymbol{\alpha}$. Oryginalne szeregi czasowe są teraz reprezentowane przez dane funkcjonalne w postaci macierzy obserwacji $A = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)^T$ o wymiarach $n \times (B + p)$, $B = B_1 + B_2 + \dots + B_p$, gdzie $\mathbf{a}_k = \widehat{\mathbf{a}}_k$ jest estymatorem uzyskanym metodą najmniejszych kwadratów.

W drugim kroku wyznaczane są przestrzenno-czasowe składowe główne. W tym celu szukamy wektora $\boldsymbol{\gamma}$, spełniającego warunek $\boldsymbol{\gamma}^T \boldsymbol{\gamma} = \mathbf{1}$, który maksymalizuje kryterium $f(\boldsymbol{\gamma}) = \widehat{\text{Var}}(\mathbf{U})I(\mathbf{z})$, gdzie $\mathbf{U} = \boldsymbol{\gamma}^T \mathbf{A}$, $\widehat{\text{Var}}(\mathbf{U}) = (1/n)\boldsymbol{\gamma}^T \mathbf{A} \mathbf{A}^T \boldsymbol{\gamma}$,

a $I(\mathbf{z}) = \frac{n}{s_0} \frac{\mathbf{z}^T \mathbf{W} \mathbf{z}}{\mathbf{z}^T \mathbf{z}}$ jest współczynnikiem autokorelacji przestrzennej I (Moran 1950,

Cliff, Ord 1973)¹. Elementy wektora $\mathbf{z}$ są równe $z_i = x_i - \bar{x}$, $i = 1, 2, \dots, n$. Macierz wag przestrzennych $\mathbf{W}$ jest symetryczna o nieujemnych elementach, spełniająca

warunek $w_{ii} = 0$, $i = 1, 2, \dots, n$ i o sumie elementów $S_0 = \sum_{i,j=1}^n w_{ij}$ (Cliff, Ord 1973). Po przekształceniach, kryterium to można zapisać w postaci $f(\boldsymbol{\gamma}) = \boldsymbol{\gamma}^T \mathbf{C} \boldsymbol{\gamma}$, gdzie $\mathbf{C} = (1 / S_0) \mathbf{A}^T \mathbf{W} \mathbf{A}$ jest macierzą korelacji przestrzennych w przestrzeni danych funkcjonalnych. Przestrzenno-czasowe składowe główne są wyznaczane w oparciu o ortonormalne wektory własne macierzy $\mathbf{C}$, która w przypadku braku symetrii może być zastąpiona macierzą symetryczną $(1 / 2)(\mathbf{C} + \mathbf{C}^T)$.

W odróżnieniu od klasycznej analizy składowych głównych, macierz $\mathbf{C}$ nie musi być nieujemnie określona, a zatem jej wartości własne mogą być ujemne. Stąd też można ustalić dwa zbiory przestrzenno-czasowych składowych głównych („dodatnich” i „ujemnych”), tj. U^+ , U^- , których elementy są równe $U_i^+ = (\boldsymbol{\gamma}_i^+)^T \boldsymbol{\alpha}$, $U_j^- = (\boldsymbol{\gamma}_j^-)^T \boldsymbol{\alpha}$, natomiast $\boldsymbol{\gamma}_i^+$, $\boldsymbol{\gamma}_j^-$ są wektorami odpowiadającymi i -tej dodatniej oraz j -tej ujemnej wartości własnej. Rzut wektora $\boldsymbol{\alpha}$ na i -tą przestrzenno-czasową składową główną jest równy $U_{ij} = \boldsymbol{\gamma}_j^T \boldsymbol{\alpha}$, przy czym U_i może być dodatnią lub ujemną składową główną w zależności od znaku wartości własnej λ_i , której odpowiada wektor $\boldsymbol{\gamma}_i$.

Krzyśko i in. (2024) wprowadzają osobne miary jakości M^+ , M^- dla dodatnich i ujemnych wyodrębnionych składowych głównych. Stosowany w klasycznej analizie składowych głównych iloraz wartości własnej i sumy wartości własnych jest nieinterpretowalny w przypadku przestrzenno-czasowych składowych głównych, ponieważ wartości własne mogą przyjmować wartości ujemne. Oznaczając przez λ^+ , λ^- zbiory odpowiednio dodatnich i ujemnych wartości własnych oraz porządkując je malejąco $\lambda_1^+ > \lambda_2^+ > \dots > \lambda_r^+ | \lambda_1^- | > | \lambda_2^- | > \dots > | \lambda_s^- |$, de-

finiuje się miary jakości następująco: $M_i^+ = (\lambda_1^+ + \lambda_2^+ + \dots + \lambda_i^+) / \sum_{j=1}^r \lambda_j^+$, oraz

$M_i^- = (|\lambda_1^-| + |\lambda_2^-| + \dots + |\lambda_i^-|) / \sum_{j=1}^s |\lambda_j^-|$.

¹ Możliwe jest też wykorzystanie współczynnika autokorelacji przestrzennej Geary’ego (1954), por. Krzyśko i in. (2023).

Otrzymane przestrzenno-czasowe składowe główne interpretuje się na podstawie udziału poszczególnych zmiennych w ich konstrukcji. Wektory γ_i mają $B + p$ składowych i można je podzielić na p podwektorów $\gamma_i(k)$ o $B_k + 1$ składowych, $k = 1, 2, \dots, p$, $\gamma_i = (\gamma_i(1), \gamma_i(2), \dots, \gamma_i(p))$. Udział k -tej zmiennej w i -tej składowej głównej zależy od długości podwektora $\gamma_i(k)$ (im dłuższy podwektor, tym większy udział) i wyliczany jest według formuły $\gamma_i(k)\gamma_i^T(k) \cdot 100\%$.

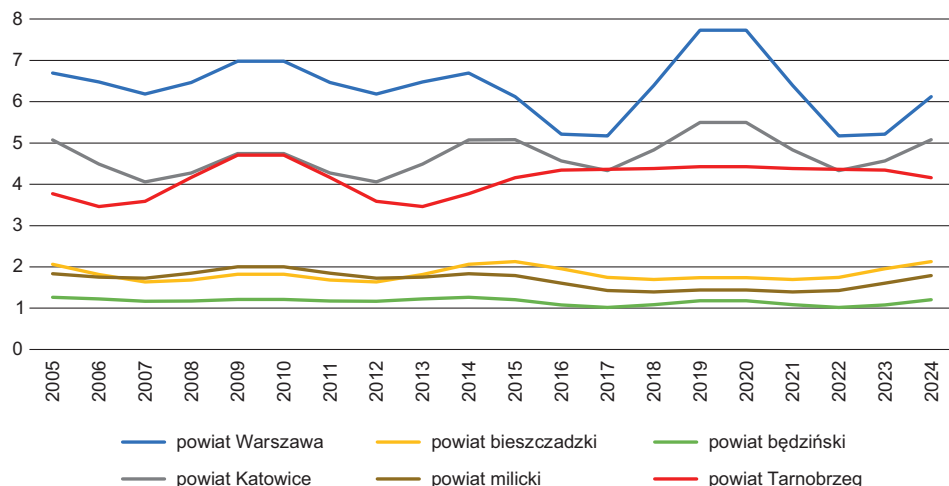
Do przekształcenia oryginalnych danych do danych funkcjonalnych, wyodrębnienia składowych głównych oraz ich interpretacji wykorzystano kod w języku R (Wołyński 2024) dołączony do artykułu Krzyński i in. (2024), który dostosowano do potrzeb niniejszego artykułu.

Typologię regionów przeprowadzono w odmienny sposób niż w pracach Krzyński i in. (2023, 2024). Proponują oni wykorzystać do tego celu dwie pierwsze składowe (osobno dodatnie i ujemne) i następnie wydzielić cztery typy na podstawie kombinacji znaków ich wartości. W artykule zastosowano analizę skupień, osobno dla dodatnich i ujemnych składowych głównych. Wykorzystano metodę grupowania hierarchicznego Warda (1963) oraz odległość euklidesową. Optymalną liczbę grup ustalono za pomocą analitycznych wskaźników grupowania, takich jak np. silhouette (Rousseeuw 1987, Kaufman, Rousseeuw 1990). Obliczenia wykonano w R (R Core Team, 2025) z użyciem pakietów factoextra (Kassambara, Mundt 2020) oraz NbClust (Charrad i in. 2014).

Wyniki

Oryginalny zbiór zmiennych pierwotnych w pierwszym kroku przekształcono do wielozmiennych danych funkcjonalnych, wykorzystując bazę Fouriera i przyjmując $B_1 = B_2 = \dots = B_{30} = 4$, $B = 120$. Wybrano bazę Fouriera, gdyż według Góreckiego i Krzyński (2012) wymaga minimalnej liczby wyrażeń dla rozwinięcia funkcji w szereg (w porównaniu np. z bazą Legendre'a). Natomiast liczba funkcji bazowych $B_k = 4$ jest wystarczająca dla wymaganej dokładności reprezentacji danych. Liczbę funkcji bazowych można też ustalić na podstawie kryterium BIC (Krzyński i in. 2024, wzór 4), ale ma ono tendencję do jej przeszacowania (powodującego długie rozwinięcia w szereg). Przykładowe dane w postaci funkcjonalnej przedstawia rycina 1. Otrzymano w ten sposób macierz $\mathbf{A}$ o wymiarach 380×150 reprezentującą zbiór obserwacji w przestrzeni współczynników wielozmiennych danych funkcjonalnych. W następnym kroku zbudowano binarną i symetryczną macierz wag przestrzennych $\mathbf{W}$ na podstawie kryterium odległości pomiędzy centroidami powiatów. Jako kryterium maksymalnej odległości przyjęto dystans 50 km, zapewniając spójność grafu reprezentującego sąsiedztwo jednostek (minimalna odległość zachowująca spójność wynosi 40,886 km). Macierz wag przestrzennych jest egzogeniczna i może zostać określona na różne sposoby. Krzyński i in. (2024) w studiach empirycznych nad charakterystyką społeczno-gospodarczą województw w Polsce zastosowali 9 definicji najczęściej wykorzystywanych w analizach przestrzennych macierzy wag przestrzennych. Ustalili oni, że dla macierzy określonej jak w niniejszym artykule, wariancja objaśniana przez dwie

pierwsze dodatnie składowe główne jest największa (dla ustalonej arbitralnie maksymalnej odległości pomiędzy centroidami województw na poziomie średniej odległości powiększonej o ich odchylenie standardowe).



Ryc. 1. Dane w postaci funkcjonalnej dla wybranych powiatów

Źródło: opracowanie własne.

Macierze **W** oraz **A** zostały wykorzystane do konstrukcji macierzy korelacji przestrzennej **C** o wymiarach 150×150 , a na jej podstawie wyodrębniono 4 pierwsze składowe odpowiadające największym dodatnim wartościom własnym oraz 4 ostatnie składowe dla ujemnych wartości własnych o największych wartościach bezwzględnych. Liczbę składowych głównych dobrano w taki sposób, aby przynajmniej w jednej z grup skumulowany udział zmienności przekroczył 75%. Na dodatnie składowe główne przypadało łącznie 87,23% skumulowanego udziału w zmienności przypadającej na składowe odpowiadające dodatnim wartościom własnym, a na ujemne składowe przypadało 72,13% skumulowanego udziału

Tabela 1. Charakterystyka wyodrębnionych przestrzenno-czasowych składowych głównych

Składowe główne odpowiadające							
dodatnim wartościom własnym				ujemnym wartościom własnym			
STPC	Wartość własna λ_i^+	Udział w zmienności %	Skumulowany udział M_i^+ %	STPC	Wartość własna λ_i^-	Udział w zmienności %	Skumulowany udział M_i^- %
1	94,11244	54,38701	54,38701	150	-4,98854	51,92716	51,92716
2	24,29829	14,04183	68,42884	149	-0,71444	7,43682	59,36397
3	18,02356	10,41571	78,84455	148	-0,67730	7,05024	66,41421
4	14,51753	8,38959	87,23414	147	-0,54928	5,71764	72,13185

Źródło: opracowanie własne.

w zmienności odpowiadającej ujemnym wartościom własnym. Wyniki te uzasadniają zastosowanie podejścia przestrzenno-czasowego, gdyż 4 pierwsze składowe główne wyodrębnione z uwzględnieniem tylko wymiaru czasowego objaśniały łącznie 73,41% zmienności. Zarówno pierwsza dodatnia składowa (STPC1) odpowiadająca największej dodatniej wartości własnej, jak i pierwsza ujemna składowa (STPC150) odpowiadająca ujemnej wartości własnej o największej wartości bezwzględnej miały ponad 50% udziału w zmienności dodatnich lub ujemnych składowych głównych (tab. 1).

Wyodrębnione 4 dodatnie składowe główne charakteryzowały się również najwyższymi spośród wszystkich składowych wartościami statystyki I Morana, które wynosiły odpowiednio $I(\text{STPC1}) = 0,3858$, $I(\text{STPC2}) = 0,3935$, $I(\text{STPC3}) = 0,4267$ oraz $I(\text{STPC4}) = 0,4746$ i były istotne na poziomie 0,001. W przypadku ujemnych składowych głównych wartości statystyki Morana były również ujemne: $I(\text{STPC150}) = -0,0682$ ($p = 0,004$), $I(\text{STPC149}) = -0,0869$ ($p = 0,002$), $I(\text{STPC148}) = -0,0676$ ($p = 0,002$) oraz $I(\text{STPC147}) = -0,0981$ ($p = 0,002$).

Interpretacja przestrzenno-czasowych składowych głównych opiera się na długości podwektorów odpowiadających zmiennym w długości wektora $\mathbf{y}_t$. Długości tych podwektorów przedstawia tabela 2. W interpretacji przyjęto, że wektory dłuższe niż 5% mają istotny wpływ na konstrukcję danej składowej. Kryterium to wynika również z: 1) stosunkowo dużej liczby zmiennych pierwotnych (30), która przekłada się na mniejsze udziały poszczególnych zmiennych w składowych, 2) faktu, że przy ustalonym progu 5% składowe mogą być szersze i szczegółowiej empirycznie interpretowane przez 27 zmiennych pierwotnych, a przy progu 10% (zastosowanym w pracy Krzyżski i in. 2024 dla 12 zmiennych pierwotnych) tylko przez 17 zmiennych.

W pierwszej dodatniej składowej głównej (STPC1) największe udziały (ponad 13%) mają zmienne X_{17} (odsetek ludności korzystającej z sieci gazowej), X_{20} (gęstość sieci gazowej w kilometrach na 100 km²) oraz X_{18} (gęstość sieci wodociągowej w kilometrach na 100 km²). Długość powyżej 5% przekraczały również podwektory zmiennych X_{21} (gęstość sieci dróg gminnych i powiatowych o twardej nawierzchni w kilometrach na 100 km²), X_{19} (gęstość sieci kanalizacyjnej w kilometrach na 100 km²), oraz nieznacznie zmienna X_1 (gęstość zaludnienia w osobach na 1 km²). Pierwsza dodatnia składowa zdominowana jest przez zmienne infrastrukturalne X_{17} – X_{21} , których łączna długość wynosi 61,9%.

Druga dodatnia składowa główna (STPC2) ma dwa podwektory przekraczające długość 14%: pierwsza z nich to odsetek ludności korzystającej z kanalizacji (X_{16}), a kolejną jest odsetek ludności korzystającej z oczyszczalni ścieków (X_{30}). Znaczący udział (29,8%) mają zmienne charakteryzujące potencjał demograficzny powiatów takie, jak: ludność w wieku nieprodukcyjnym na 100 osób w wieku produkcyjnym (X_3), ludność w wieku poprodukcyjnym na 100 osób w wieku produkcyjnym (X_4), współczynnik obciążenia demograficznego osobami starszymi (X_5) i przyrost naturalny na 1000 ludności (X_6). Udział przekraczający 5% ma jeszcze zmienna X_{17} . Ze względu na taką strukturę udziałów, składową tę można

interpretować jako demograficzną, choć ma ona też spory udział (przekraczający 20%) zmiennych infrastrukturalnych.

Składowa STPC3 ma 8 podwektorów o długościach przekraczających 5%. Jest to najbardziej zróżnicowana składowa pod względem struktury zmiennych. Każda grupa jest reprezentowana przez dwie zmienne pierwotne. Wśród nich zmienne charakteryzujące potencjał demograficzny (X_2 – zmiana liczby ludności na 1000 mieszkańców oraz X_6) mają łączny udział 20,2%. O zaledwie 0,4% mniejszy

Tabela 2. Udziały zmiennych w przestrzenno-czasowych składowych głównych

Zmienna	Składowa główna							
	dodatnia				ujemna			
	STPC1	STPC2	STPC3	STPC4	STPC150	STPC149	STPC148	STPC147
Długość podwektora %								
X_1	5,220	0,507	0,471	0,230	0,363	0,122	1,537	15,209
X_2	0,053	1,118	8,753	0,027	1,491	0,175	3,088	1,733
X_3	1,484	7,476	3,353	4,972	5,000	1,486	8,045	6,716
X_4	0,170	7,831	2,280	1,519	1,810	0,183	2,323	1,103
X_5	0,178	8,182	1,505	0,922	1,698	0,123	2,609	0,764
X_6	0,090	6,297	11,422	3,786	0,860	0,613	0,367	1,424
X_7	0,969	0,919	0,117	2,954	4,328	0,135	0,527	2,301
X_8	2,435	0,061	0,797	0,011	8,161	14,433	1,466	3,354
X_9	2,374	0,079	0,816	0,004	8,492	14,229	1,289	3,363
X_{10}	0,141	3,657	7,174	7,620	1,339	0,567	0,480	3,196
X_{11}	1,250	0,316	0,412	1,187	0,235	4,260	7,506	2,231
X_{12}	4,630	2,537	12,605	3,601	0,234	0,887	6,583	0,610
X_{13}	1,163	0,014	0,753	1,860	0,390	1,427	1,457	7,555
X_{14}	0,074	4,249	11,042	8,214	1,969	0,127	8,287	0,792
X_{15}	1,462	1,342	6,252	23,300	1,863	0,819	2,230	0,297
X_{16}	4,496	14,981	2,442	2,738	0,567	1,537	1,210	0,040
X_{17}	15,852	8,962	2,244	18,095	0,932	0,612	1,733	1,433
X_{18}	13,054	4,645	0,174	0,043	0,709	0,331	0,425	1,845
X_{19}	8,616	0,360	0,211	1,501	0,573	1,618	1,193	9,091
X_{20}	13,347	0,655	0,352	5,185	0,065	0,158	2,381	3,568
X_{21}	10,991	4,016	0,048	0,300	0,423	0,019	2,966	0,609
X_{22}	0,953	0,161	0,073	0,796	0,030	0,707	0,149	1,927
X_{23}	0,139	0,354	1,678	1,797	1,181	20,767	11,303	0,619
X_{24}	0,054	0,463	0,554	0,898	21,525	19,365	17,819	10,918
X_{25}	0,185	1,123	14,256	1,054	0,380	2,322	7,527	0,688
X_{26}	1,030	4,656	0,145	1,570	16,561	3,031	1,885	11,832
X_{27}	3,158	0,537	5,160	3,607	7,475	1,845	0,627	1,257
X_{28}	0,035	0,059	0,001	0,062	0,656	0,287	0,450	1,581
X_{29}	2,830	0,368	1,768	0,235	9,278	3,741	1,499	3,615
X_{30}	3,569	14,073	3,142	1,912	1,411	4,073	1,040	0,328

Udziały zmiennych powyżej 5% wytłuszczono.

Źródło: opracowanie własne.

jest udział zmiennych opisujących gospodarkę powiatów: podmioty wpisane do rejestru REGON na 1000 ludności (X_{10}) oraz stopa bezrobocia rejestrowanego w procentach (X_{12}). Udział tych zmiennych w składowej o drugiej co do wielkości wartości miernika autokorelacji przestrzennej Morana nie jest przypadkowy. Zmienne te również mają jedno z najwyższych wartości I Morana w różnych układach terytorialnych (Hauke, Kossowski 2012). Zmienne społeczno-środowiskowe mają z kolei udział wynoszący 19,4% i składają się na niego długość podwektora zmiennej liczby mieszkań oddanych do użytkowania na 1000 ludności (X_{25}) i udział dzieci w przedszkolach i innych formach wychowania przedszkolnego na 1000 dzieci w wieku 3–5 lat (X_{27}). Długość podwektora zmiennej X_{25} jako jedyna przekraczała 14%. Najmniejszy udział w konstrukcji trzeciej składowej miały zmienne infrastrukturalne: przeciętna powierzchnia użytkowa mieszkania na 1 osobę w m^2 (X_{14}) oraz odsetek ludności korzystającej z wodociągu (X_{15}) – i wynosił on łącznie 17,3%. Ze względu na niewielkie różnice udziałów trzech pierwszych wymienionych grup, składowa ta ma niejednoznaczny charakter.

Konstrukcja ostatniej wyodrębnionej dodatknej składowej głównej (STPC4) ma tylko 3 zmienne o długości podwektora przekraczającej 5%. Były to zmienne: X_{10} – podmioty wpisane do rejestru REGON na 1000 ludności, X_{15} – odsetek ludności korzystającej z wodociągu oraz X_{17} – odsetek ludności korzystającej z sieci gazowej, przy czym długość podwektora zmiennej przekraczała 23%, a podwektora zmiennej X_{15} wynosiła ponad 17%. W składowej tej dominują zmienne infrastrukturalne, ale należy zauważyć, że jej udział w zmienności jest ponad czterokrotnie mniejszy niż pierwszej składowej.

Analogicznie interpretuje się ujemne przestrzenno-czasowe składowe główne. Pierwsza z nich (STPC150) ma najdłuższe podwektory związane ze zmiennymi społeczno-środowiskowymi (łącznie 54,84%): liczba łóżek w szpitalach ogólnych na 10000 mieszkańców (X_{24} , długość podwektora 21,5%) i przeciętna powierzchnia użytkowa 1 mieszkania oddanego do użytkowania w m^2 (X_{26} , długość podwektora 16,6%), odsetek ścieków przemysłowych i komunalnych oczyszczanych w ściekach wymagających oczyszczania (X_{29} , długość podwektora 9,28%) oraz zmienna X_{27} (długość podwektora 7,48%). Ważne są jeszcze zmienne związane z dochodami i wydatkami powiatów (X_8 , X_9) o długościach przekraczających 7%. Składowa ta ma charakter społeczno-środowiskowy i finansów publicznych.

Druga ujemna składowa (STPC149) strukturalnie podobna jest do pierwszej, ale jest bardziej zrównoważona. Zmienne społeczne (X_{23} , czyli księgozbiór bibliotek na 1000 ludności oraz X_{24}) miały podwektory o długości łącznie przekraczającej 40%, natomiast zmienne dotyczące finansów powiatów (X_8 , X_9) miały łączną długość ponad 28%. Z kolei składowa STPC148 miała 7 zmiennych o długościach podwektorów przekraczających 5%. Prawie 37% łącznej długości podwektorów przypadało na charakterystyki społeczne powiatów opisywane zmiennymi X_{23} , X_{24} , X_{25} , a ich poszczególne długości wahały się od ponad 7% do ponad 17%. Zmienne takie, jak X_{14} (przeciętna powierzchnia użytkowa mieszkania na 1 osobę w m^2) oraz X_3 , należące do grupy wskaźników infrastrukturalnych i demograficznych, miały długości ponad 8%. Zmienne X_{11} (produkcja sprzedana przemysłu w przedsiębiorstwach powyżej 9 pracujących na 1 mieszkańca w złotych) oraz

X_{12} opisujące gospodarkę powiatów miały łączny udział przekraczający nieco 14%. Taki udział poszczególnych zmiennych pozwala traktować tę składową jako społeczno-gospodarczą.

Ostatnia wyodrębniona ujemna składowa (STPCA147) ma dwie grupy zmiennych o podobnej łącznej długości podwektorów: społeczne (X_{24} , X_{26} , dające 22,8% długości) i demograficzne (X_1 , X_3 , razem 21,93% długości). Dwie pozostałe zmienne o podwektorach przekraczających 5% długości to X_{19} (9,09%) i X_{13} (przeciętne miesięczne wynagrodzenia brutto w złotych, 7,56%).

Wśród analizowanych zmiennych, 26 ma w co najmniej jednej wyodrębnionej składowej podwektor o długości powyżej 5%. Cztery zmienne nie osiągają takiego udziału i są to następujące wskaźniki: X_4 (ludność w wieku poprodukcyjnym na 100 osób w wieku produkcyjnym), X_7 (współczynnik feminizacji), X_{22} (ludność na 1 placówkę biblioteczną łącznie z punktami bibliotecznymi ujętymi zgodnie z siedzibą jednostki macierzystej) i X_{28} (zużycie wody na 1 mieszkańca w m^3).

Dodatknie przestrzenno-czasowe składowe główne reprezentują procesy, które skutkują występowaniem wzorca dodatniej autokorelacji przestrzennej. Analogicznie ujemne składowe związane są z procesami generującymi wzorzec ujemnej autokorelacji przestrzennej. W tej procedurze wyodrębniania składowych do oceny autokorelacji przestrzennej wykorzystywana jest miara I Morana, której konstrukcja umożliwia identyfikację występowania efektów aglomeracyjnych. W konsekwencji dodatnie przestrzenno-czasowe składowe możemy interpretować jako procesy aglomeracyjne, a ujemne składowe jako reprezentantów procesów różnicujących jednostki.

Analiza struktury składowych głównych pozwala na ustalenie, jakie własności jednostek przestrzennych sprzyjają procesom aglomeracyjnym, a jakie wręcz przeciwnie. W tym przypadku efekty aglomeracyjne wywoływane są w największym stopniu przez infrastrukturę wodną, gazową i kanalizacyjną, a także przez podobne wzorce przyrostu naturalnego i silnie skorelowane przestrzennie rozmieszczenie podmiotów gospodarczych zarejestrowanych w REGON. Efekty różnicujące widoczne są zwłaszcza w przypadku zmiennych społecznych, ze szczególnym uwzględnieniem dostępu do opieki medycznej na poziomie szpitali. Jest to najbardziej różnicujący przestrzennie powiaty czynnik. Również finanse publiczne różnicują powiaty, gdyż odmienne są konstrukcje budżetów powiatów ziemskich i grodzkich.

W odniesieniu do części zmiennych trudności nastręcza ustalenie, czy związane są z efektami aglomeracyjnymi, czy też wręcz przeciwnie. Występują bowiem w strukturze zarówno dodatnich, jak i ujemnych składowych głównych i należą do nich: X_1 , X_3 , X_{12} , X_{14} , X_{25} , X_{27} . Biorąc pod uwagę fakt, że autokorelacja wyodrębnionych dodatnich składowych jest silniejsza niż ujemnych składowych, można przyjąć, że przynajmniej w przypadku zmiennych X_{12} (bezrobocie) i X_{14} (warunki mieszkaniowe) przeważają efekty aglomeracyjne (z możliwością wystąpienia również obserwacji odstających).

W kolejnym kroku postępowania badawczego przeprowadzono analizę skupień. Wybrano metodę grupowania hierarchicznego Warda na podstawie odległości euklidesowej pomiędzy obserwacjami. Analizę skupień wykonano osobno na

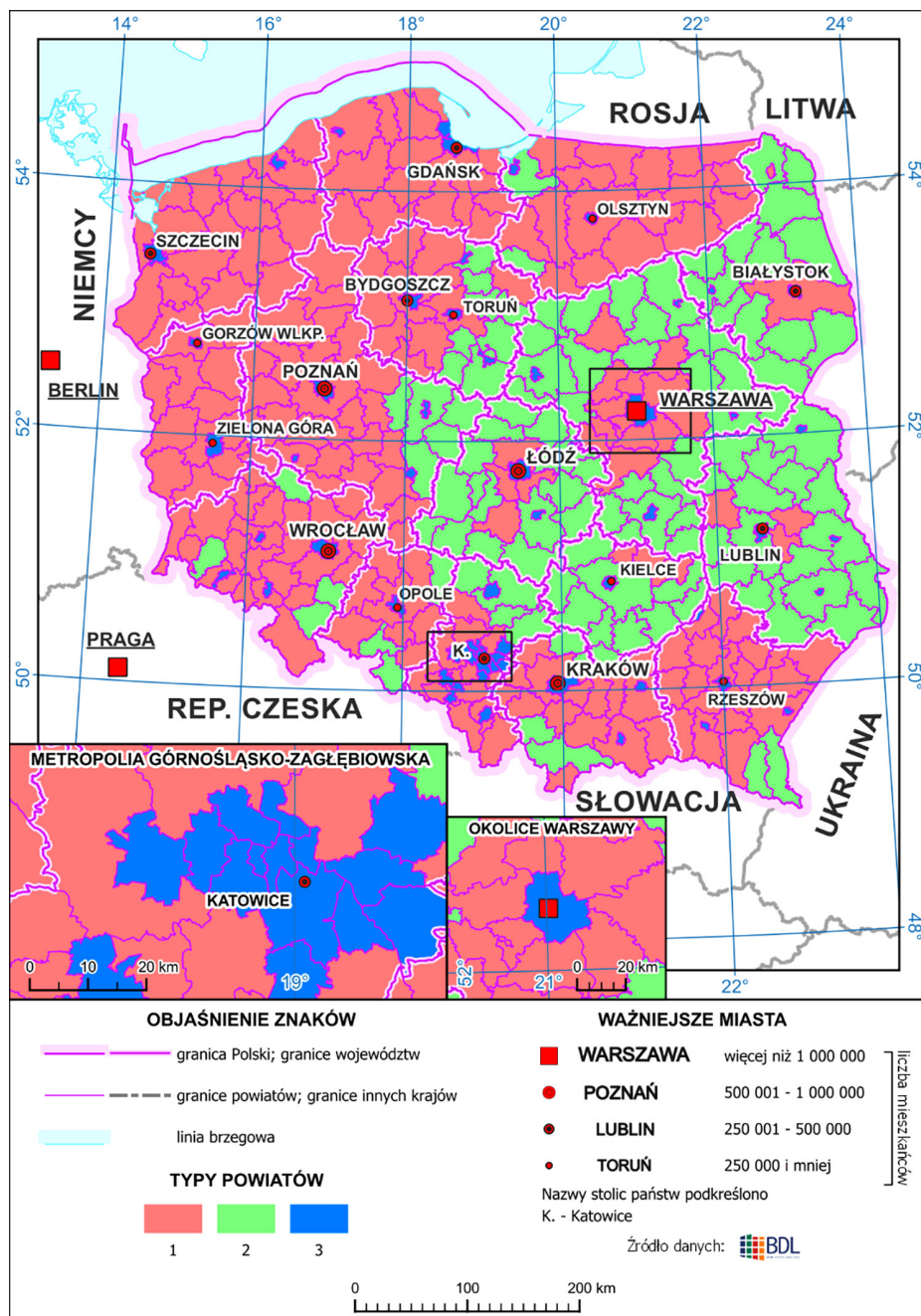
dodatnich i ujemnych składowych głównych. Liczbę skupień ustalono na podstawie miar jakości grupowania i w obu przypadkach wskazały one, że najbardziej optymalny jest podział na 3 grupy. Rozkłady przestrzenne skupień dla dodatnich i ujemnych przestrzenno-czasowych składowych głównych pokazane są odpowiednio na rycinach 2 i 3.

Rozkład przestrzenny typów rozwoju powiatów wyprowadzony na podstawie dodatnich składowych głównych pokazuje charakterystyczny dla wielu zjawisk społeczno-ekonomicznych w Polsce podział nawiązujący do granic państw zaborczych przed 1914 r. Typ 1 dominuje na obszarach dawnego zaboru pruskiego i austriackiego oraz na terenach, które weszły w skład Polski po 1945 r. Typ 2 obejmuje głównie powiaty na terenie dawnego zaboru rosyjskiego. Z wzorca tego wyłamują się okolice Warszawy, Łodzi i Białegostoku, których powiaty również zaliczają się do typu 1. Z kolei nieliczne powiaty na terenie dawnych zaborów pruskiego i austriackiego zaliczają się do typu 2 i są to głównie obszary peryferyjne przy granicy państwa lub granicy województw. Typ 3 to miasta na prawach powiatu (bez Suwałk) i niektóre powiaty ziemskie województwa śląskiego. Rozkład przestrzenny typów jest wynikiem działania istotnych efektów aglomeracyjnych skutkujących występowaniem wysokiej autokorelacji przestrzennej i w konsekwencji grupowania się powiatów o podobnych charakterystykach.

Rycina 3 przedstawia rozkład przestrzenny typów rozwoju powiatów w wyniku działania procesów różnicujących. Wyraźnie wyodrębniają się ze swojego otoczenia miasta na prawach powiatu (typ 3). Rozkład przestrzenny typów 1 i 2 nie ma zauważalnego wzorca przestrzennego. Wynika to z faktu, że autokorelacja przestrzenna dla wyodrębnionych ujemnych przestrzenno-czasowych składowych jest niewielka i waha się w przedziale $[-0,0981; -0,0682]$. Jest więc zbyt słaba i niewiele różniąc się od układu losowego, aby rozkład przestrzenny typów 1 i 2 przybrał charakterystyczny dla silnej ujemnej autokorelacji przestrzennej wzorzec szachownicowy. Można więc stwierdzić, że procesy różnicujące polskie powiaty są znacznie słabsze niż procesy aglomeracyjne, których z kolei skutkiem jest narastająca polaryzacja przestrzenna.

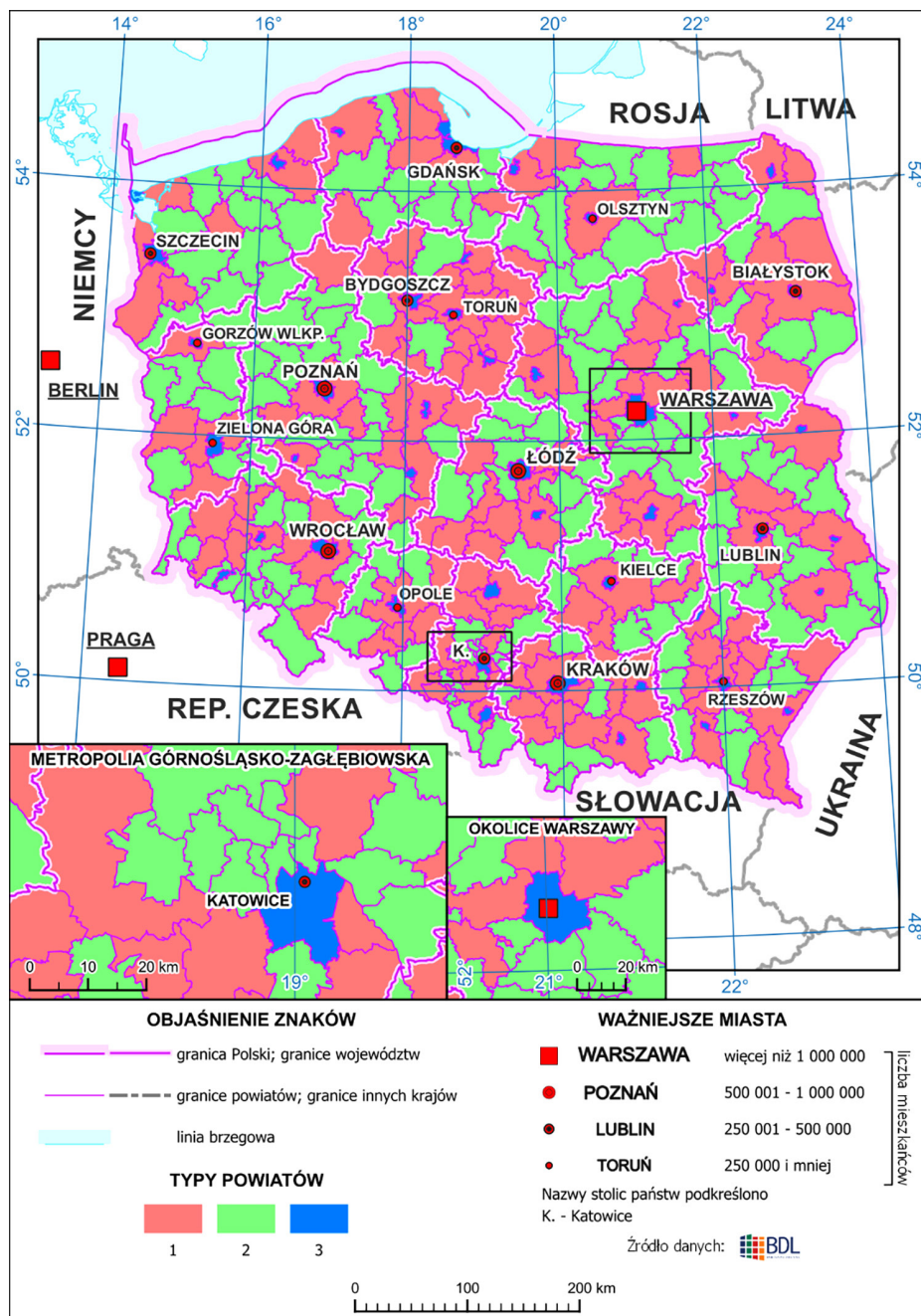
Przedstawione na poniższych rycinach podziały na 3 typy dają obraz bardzo zgeneralizowany. Dlatego uszczegółowiono podział dla dodatnich składowych głównych, wybierając kolejny sugerowany przez miary jakości układ 9 typów. Ze względu na słabą autokorelację ujemną, zrezygnowano z dokładniejszego podziału dla ujemnych składowych głównych.

Rycina 4 prezentuje rozkład przestrzenny 9 typów powiatów dla dodatnich składowych głównych. Typ 1a obejmuje swoim zasięgiem powiaty-obwarzanki lub położone w strefie podmiejskiej największych polskich miast oraz w aglomeracji górnośląsko-zagłębiowskiej. Typ 1b to z kolei część województwa małopolskiego i prawie całe województwo podkarpackie. Kolejny typ, 1d, to powiaty w sąsiedztwie mniejszych miast na prawach powiatu lub w drugim pierścieniu powiatów wokół największych polskich miast (na skraju ich miejskich obszarów funkcjonalnych). Typowo rolnicze oraz słabiej rozwinięte powiaty Polski północnej i zachodniej należą do typów 1c i 1e. Dawny typ 2 został rozbity na typy 2a i 2b, przy czym pierwszy przeważa w Polsce Wschodniej, a drugi jest



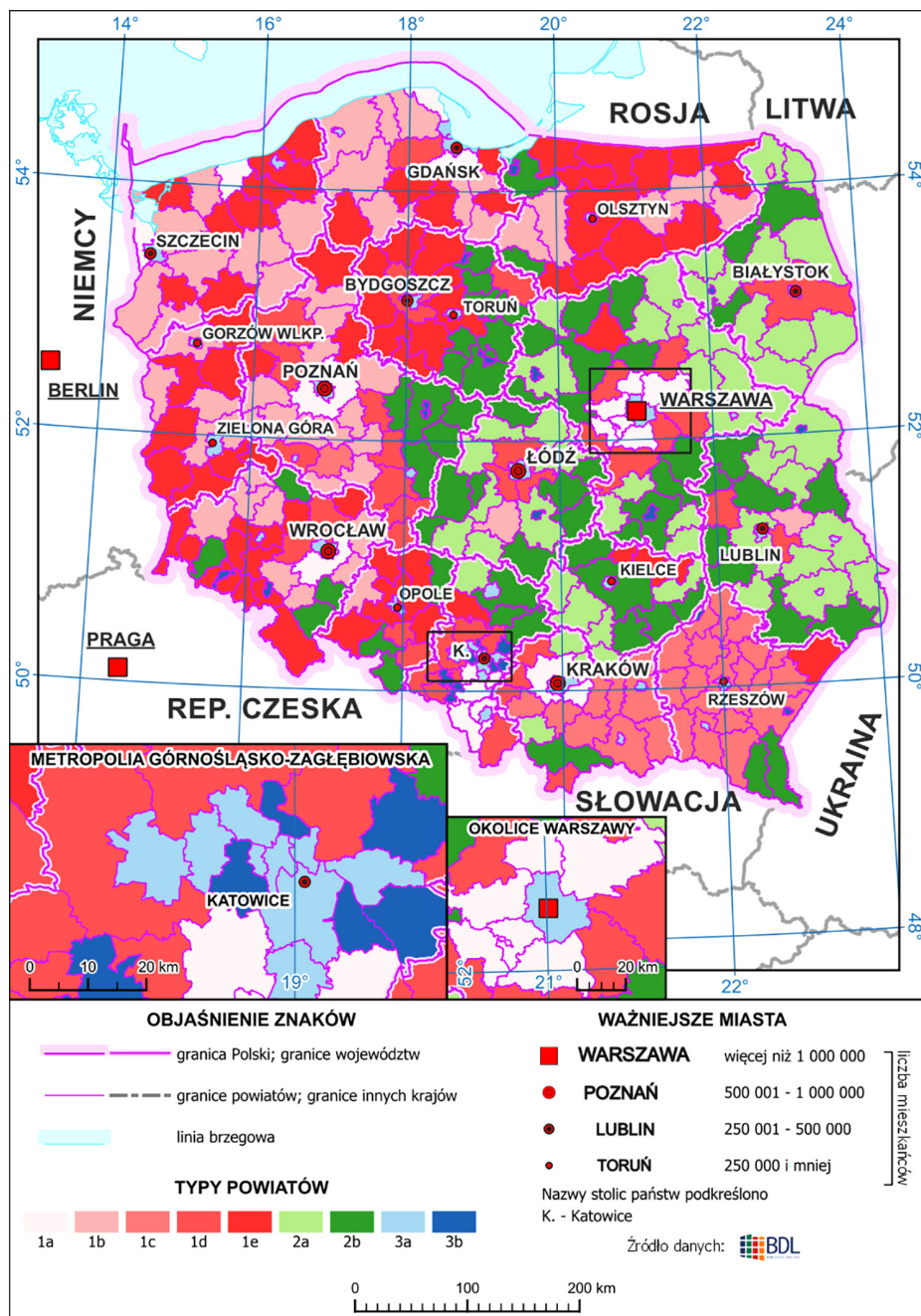
Ryc. 2. Rozkład przestrzenny typów rozwoju powiatów wyodrębnionych na podstawie 4 pierwszych dodatknych składowych głównych (efekt dodatniej autokorelacji przestrzennej)

Źródło: opracowanie własne.



Ryc. 3. Rozkład przestrzenny typów rozwoju powiatów wyodrębnionych na podstawie 4 pierwszych ujemnych składowych głównych (efekt ujemnej autokorelacji przestrzennej)

Źródło: opracowanie własne.



Ryc. 4. Rozkład przestrzenny typów rozwoju powiatów na podstawie 4 dodatknych składowych głównych (z uwzględnieniem dodatniej autokorelacji przestrzennej), podział na 9 typów

Źródło: opracowanie własne.

powszechniejszy w rejonie pogranicza dawnego zaboru pruskiego i rosyjskiego. Miasta na prawach powiatu reprezentują teraz jeden z dwóch typów – 3a lub 3b. Do pierwszego z nich należą wszystkie obecne miasta wojewódzkie oraz część byłych miast wojewódzkich sprzed reformy w 1999 r. W typie 3b wyróżnić można miasta, które doświadczały w ostatnim dwudziestolecu pewnych trudności rozwojowych, np. znacznego odpływu mieszkańców (Wałbrzych, Jelenia Góra) czy wygaszania eksploatacji surowców energetycznych (Konin).

Podsumowanie

W pracy tej przeprowadzono identyfikację przestrzennego zróżnicowania społeczno-gospodarczego powiatów dla lat 2005–2024. W tym celu zastosowano kilkuetapową procedurę badawczą, na którą złożyły się: 1) zebranie danych opisujących sytuację społeczno-gospodarczą powiatów i ich uporządkowanie, 2) przekształcenie zmiennych do postaci funkcjonalnej oraz wybór macierzy reprezentującej strukturę interakcji przestrzennych, 3) redukcja wymiaru przestrzeni obserwacji za pomocą metody przestrzenno-czasowych składowych głównych dla danych funkcjonalnych, 4) wydzielenie typów rozwoju powiatów oraz 5) interpretacja uzyskanych wyników. O ile pierwszy etap procedury nie nastręczył trudności, o tyle kolejne wymagają dodatkowego komentarza.

Przekształcenie zmiennych do postaci funkcjonalnej spowodowało redukcję wymiaru czasowego przestrzeni obserwacji z 20 wymiarów do pięciowymiarowej przestrzeni współczynników funkcji bazy Fouriera. Ruch ten pozwolił na zmniejszenie rozmiaru estymowanej macierzy kowariancji i liczby szacowanych elementów z $[pT \times (pT + 1)] / 2 = (600 \times 601) / 2 = 180300$ (jak w metodzie Hotelinga) do $[(B + p) \times (B + p + 1)] / 2 = (150 \times 151) / 2 = 11325$ elementów.

Struktura przestrzenna została określona za pomocą macierzy wag przestrzennych, w której uwzględniono sąsiadów w odległości do 50 km od centroidów powiatów. Macierz miała postać binarną. Tak zdefiniowana macierz charakteryzowała się dużą rzadkością i gwarantowała spójność grafu bezpośredniego sąsiedztwa. Ponadto lepsza jest macierz o mniejszej licznie sąsiadów niż zbyt dużej, gdyż nie występuje wtedy ryzyko przeszacowania wielkości interakcji przestrzennych.

Metoda składowych głównych zredukowała przestrzeń obserwacji do 4 dodatnich i 4 ujemnych składowych. Ich interpretacja pozwoliła na stwierdzenie, że w przypadku składowych dodatnich najważniejszy udział mają zmienne infrastrukturalne przed demograficznymi, gospodarczymi i społecznymi. Za zróżnicowanie zaś odpowiadają zmienne opisujące dostęp do opieki szpitalnej i finansów powiatów.

Metodą analizy skupień wydzielono po 3 typy rozwoju powiatów dla dodatnich i ujemnych składowych, przy czym optymalną liczbę skupień ustalono za pomocą kryteriów jakości dopasowania podziału. Ze względu na ogólny charakter tej typologii (mała liczba typów w porównaniu z liczbą jednostek) uszczegółowiono podział dla dodatnich składowych, wyodrębniając 9 typów. Uzyskany podział w układzie 3 typów potwierdza występujące zróżnicowanie powiatów ze względu

na ich położenie na obszarze dawnych państw zaborczych. Widoczna jest również dychotomia typu rozwoju pomiędzy powiatami grodzkimi a ziemskimi. Uszczegółowiona typologia dokonała podziału powiatów grodzkich na dwie grupy różniące się zmianami rozwoju w analizowanym dwudziestoleciu. Wyodrębnione zostały również typy powiatów-obwarzanków dużych miast, powiatów o charakterze rolniczym i szczególny typ powiatów Polski południowo-wschodniej.

Podsumowując, można stwierdzić, że zastosowana metoda przestrzenno-czasowych składowych głównych dla danych funkcjonalnych wykazała swoją przydatność dla budowy typologii powiatów. Wydzielone na podstawie składowych typy uwzględniają efekty aglomeracyjne (dodatnia autokorelacja przestrzenna) lub procesy różnicujące (ujemna autokorelacja przestrzenna). W przypadku tego badania empirycznego efekty aglomeracyjne mają większe znaczenie niż procesy różnicujące powiaty. Wydzielone typy poddają się prostej interpretacji i ich rozkład przestrzenny jest możliwy do wyjaśnienia.

Metoda przestrzenno-czasowych składowych głównych ma też przydatne właściwości, które mogą ją wyróżniać w porównaniu z metodami klasycznymi. Przede wszystkim integruje ona element czasowy (dynamikę) i element przestrzenny. Klasyczna metoda składowych głównych może zostać zastosowana do ujęć statycznych, a podejście dynamiczne można realizować przez powtarzanie procedury dla każdej serii obserwacji w czasie. Przekształcenie danych do postaci funkcjonalnej pozwala na redukcję wymiaru czasowego i wymaga tylko jednokrotnego zastosowania metody. Wprowadzenie do metody składowych informacji o strukturze interakcji przestrzennych pomiędzy jednostkami powoduje włączenie elementu przestrzennego, ważnego z punktu widzenia badań geograficznych.

Konflikt interesów

Autor deklaruje brak występowania konfliktu interesów. Oświadczam, że tekst artykułu jest w całości jego dziełem.

Literatura / References

- Anselin L. 1995. Local Indicators of Spatial Association – LISA. *Geographical Analysis*, 27: 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- Bartkowiak-Bakun N. 2015. Zróżnicowanie poziomu rozwoju społeczno-gospodarczego obszarów wiejskich województwa wielkopolskiego. *Roczniki Naukowe Ekonomii Rolnictwa i Rozwoju Obszarów Wiejskich*, 102(2): 7–13.
- Charrad M., Ghazzali N., Boiteau V., Niknafs A. 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw.*, 61(6): 1–36. (<http://www.jstatsoft.org/v61/i06/>).
- Churski P. 2008. Czynniki rozwoju regionalnego i polityka regionalna w Polsce w okresie integracji z Unią Europejską. Wydawnictwo Naukowe UAM, Poznań.
- Cliff A.D., Ord J.K. 1973. Spatial autocorrelation. Pion.
- Dziekański P. 2018. Sytuacja finansowa a rozwój zrównoważony powiatów Polski Wschodniej. *Optimum. Economic Studies*, 2: 25–37.
- Geary R.C. 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5: 115–145.

- Górecki T., Krzyśko M. 2012. Functional principal components analysis. [W:] J. Pocięcha, R. Decker (red.), *Data analysis methods and its applications*. C.H. Beck, s. 71–87.
- Górecki T., Krzyśko M., Waszak Ł., Wołyński W. 2014. Methods of reducing dimension for functional data. *Statistics in Transition – New Series*, 15(2): 231–242.
- Górecki T., Krzyśko M., Waszak Ł., Wołyński W. 2018. Selected statistical methods of data analysis for multivariate functional data. *Statistical Papers*, 59(1): 153–182. <https://doi.org/10.1007/s00362-016-0757-8>
- Hauke J., Kossowski T. 2012. An analysis of labour market in metropolitan areas. *Quaestiones Geographicae*, 31(2): 39–48.
- Hellwig Z. 1968. Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Przegląd Statystyczny*, 15, 4: 307–327.
- Hotelling H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6): 417–441. <https://doi.org/10.1037/h0071325>
- Jombart T., Devillard S., Dufour A.B., Pontier D. 2008. Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, 101(1): 92–103. <https://doi.org/10.1038/hdy.2008.34>
- Józefowicz K., Smolińska K. 2019. Poziom rozwoju społeczno-gospodarczego w powiatach województwa wielkopolskiego. *Turystyka i Rozwój Regionalny*, 11: 37–49.
- Kassambara A., Mundt F. 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7 (<https://CRAN.R-project.org/package=factoextra>).
- Kaufman L., Rousseeuw P.J. 1990. *Finding Groups in Data: an Introduction to Cluster Analysis*. Wiley Series in Probability and Statistics. <https://doi.org/10.1002/9780470316801>
- Krzyśko M., Nijkamp P., Ratajczak W., Wołyński W. 2022. Multidimensional economic indicators and multivariate functional principal component analysis (MFPCA) in a comparative study of countries' competitiveness. *Journal of Geographical Systems*, 24(1): 49–65. <https://doi.org/10.1007/s10109-021-00352-8>
- Krzyśko M., Nijkamp P., Ratajczak W., Wołyński W., Wenerska B. 2024. Spatio-temporal principal component analysis. *Spatial Economic Analysis*, 19(1): 8–29. <https://doi.org/10.1080/17421772.2023.2237532>
- Krzyśko M., Nijkamp P., Ratajczak W., Wołyński W., Wojtyła A., Wenerska B. 2023. A novel Spatio-temporal principal component analysis based on Geary's contiguity ratio. *Computers, Environment and Urban Systems*, 103, 101980. <https://doi.org/10.1016/j.compenvurbsys.2023.101980>
- Maćkiewicz A., Ratajczak W. 1993. Principal components analysis (PCA). *Computers & Geosciences*, 19(3): 303–342. [https://doi.org/10.1016/0098-3004\(93\)90090-R](https://doi.org/10.1016/0098-3004(93)90090-R)
- Parysek J., Ratajczak W. 1978. Analiza głównych składowych cech charakteryzujących własności społeczno-ekonomiczne i środowisko geograficzne Polski. [W:] Z. Chojnicki, T. Czyż, J. Parysek, W. Ratajczak (red.), *Badania przestrzennej struktury społeczno-ekonomicznej Polski metodami czynnikowymi*. PAN, Oddział w Poznaniu, Seria Geografia, t. II. PWN, Warszawa, s. 83–106.
- Pietrzak M.B. 2014. Taksonomiczny miernik rozwoju (TMR) z uwzględnieniem zależności przestrzennych. *Institute of Economic Research Working Papers*, 10. Institute of Economic Research (IER), Toruń.
- Piszczyk S. 2013. Zróżnicowanie przestrzenne poziomu rozwoju społeczno-gospodarczego obszarów wiejskich województwa kujawsko-pomorskiego. *Nierówności Społeczne a Wzrost Gospodarczy*, 31: 334–346.
- R Core Team 2025. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (<https://www.R-project.org/>).
- Rousseeuw P.J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20: 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Ward Jr. J.H. 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58: 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Wojtyra B. 2017. Poziom wielofunkcyjnego rozwoju obszarów wiejskich województwa wielkopolskiego. *Rozwój Regionalny i Polityka Regionalna*, 40: 149–161.
- Wołyński W. 2024 (<http://woylanski.home.amu.edu.pl/Software.html>).

Application of the spatio-temporal principal component analysis (STPCA) method for identifying the spatial differentiation of socio-economic conditions of counties

Abstract: The aim of the article is to identify the spatial socio-economic differentiation of counties in the years 2005–2024. A set of 30 variables was used, including demographic and economic variables, the state of infrastructure, and socio-environmental variables. A spatio-temporal principal component method for functional data was applied to reduce the dimensionality of the observations. First, the variables were transformed into a functional form, and a matrix representing the spatial interaction structure was determined. Then, 4 positive and 4 negative principal components were extracted. Their interpretation allowed them to state that, in the case of positive components, infrastructural variables make the most important contribution. County differentiation is determined by variables describing access to hospital care and county finances.

The classification of county types was carried out using the cluster analysis method, conducted separately for the positive and negative principal components. As a result, two divisions into 3 county types each and one into 9 types were obtained. The resulting typology confirms the persistent differentiation of counties based on their location within the territories of former partitioning states, the dichotomy between urban and rural counties, and their location relative to metropolitan areas. The obtained results confirm the usefulness of the applied method for constructing a typology of counties. The types distinguished based on the cluster analysis of the components take into account both agglomeration effects, which are more significant than processes that differentiate counties. Moreover, the county typology lends itself to straightforward interpretation, and its spatial distribution can be explained.

Keywords: spatio-temporal principal components, spatial autocorrelation, spatial heterogeneity, county typology, agglomeration processes

Data przekazania tekstu: 4.05.2026; **data zaakceptowania tekstu:** 20.07.2026.

Article submitted: 4.05.2026; **article accepted:** 20.07.2026.