

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS

of Wrocław University of Economics

Nr 427

Taksonomia 27

**Klasyfikacja i analiza danych –
teoria i zastosowania**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronach internetowych
www.pracnaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

Publikacja udostępniona na licencji Creative Commons
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2016

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

e-ISSN 2392-0041

ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
ul. Komandorska 118/120, 53-345 Wrocław
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp	9
Beata Bal-Domańska: Propozycja procedury oceny zrównoważonego rozwoju w układzie <i>presja – stan – reakcja</i> w ujęciu przestrzennym / Proposal of the assessment of poviats sustainable development in the pressure – state – response system in spatial terms.....	11
Tomasz Bartłomowicz: Pomiar preferencji konsumentów z wykorzystaniem metody <i>Analytic Hierarchy Process</i> / Analytic Hierarchy Process as a method of measurement of consumers’ preferences.....	20
Maciej Beręsewicz, Marcin Szymkowiak: Analiza skupień wybranych lokalnych rynków nieruchomości w Polsce z wykorzystaniem internetowych źródeł danych / Cluster analysis of selected local real estate markets in Poland based on Internet data sources.....	30
Beata Bieszk-Stolorz: Wybrane modele przeciętnego efektu oddziaływania w analizie procesu wychodzenia z bezrobocia / Chosen average treatment effect models in the analysis of unemployment exit process.....	40
Justyna Brzezińska: Modele IRT i modele Rascha w badaniach testowych / IRT and Rasch models in test measurement.....	49
Mariola Chrzanowska, Nina Drejerska: Geograficznie ważona regresja jako narzędzie analizy poziomu rozwoju społeczno-gospodarczego na przykładzie regionów Unii Europejskiej / Geographically weighted regression as a tool of analysis of socio-economic development level of regions in the European Union.....	58
Sabina Denkowska: Zastosowanie analizy wrażliwości do oceny wpływu nieobserwowanej zmiennej w <i>Propensity Score Matching</i> / The application of sensitivity analysis in assessing the impact of an unobserved confounder in Propensity Score Matching.....	66
Adam Depta: Zastosowanie analizy czynnikowej do wyodrębnienia aspektów zdrowia wpływających na jakość życia osób jękających się / The application of factor analysis to the identification of the health aspects affecting the quality of life of stuttering people.....	76
Mariusz Doszyń, Sebastian Gnat: Taksonomiczno-ekonometryczna procedura wyceny nieruchomości dla różnych miar porządkowania / Taxonomic and econometric method of real estate valuation for various classification measures.....	84

Marta Dziechciarz-Duda, Anna Król: Segmentacja konsumentów smartfonów na podstawie preferencji wyrażonych / Segmentation of smartphones' consumers on the basis of stated preferences	94
Ewa Genge: Zmienne towarzyszące w ukrytym modelu Markowa – analiza oszczędności polskich gospodarstw domowych / Latent Markov model with covariates – Polish households' saving behaviour	103
Joanna Górna, Karolina Górna: Modelowanie wzrostu gospodarczego z wykorzystaniem narzędzi ekonometrii przestrzennej / Economic growth modelling with the application of spatial econometrics tools	112
Alicja Grześkowiak: Wielowymiarowa analiza kompetencji zawodowych według grup wieku ludności / Multivariate analysis of professional competencies with respect to the age groups of the population	122
Agnieszka Kozera, Feliks Wysocki: Problem ustalania współrzędnych obiektów modelowych w metodach porządkowania liniowego obiektów / The problem of determining the coordinates of model objects in object linear ordering methods	131
Mariusz Kubus: Lokalna ocena mocy dyskryminacyjnej zmiennych / Local evaluation of a discrimination power of the variables.....	143
Paweł Lula, Katarzyna Wójcik, Janusz Tuchowski: Analiza wydzźwięku polskojęzycznych opinii konsumenckich ukierunkowanych na cechy produktu / Feature-based sentiment analysis of opinions in Polish.....	153
Aleksandra Łuczak, Agnieszka Kozera, Feliks Wysocki: Ocena sytuacji finansowej jednostek samorządu terytorialnego z wykorzystaniem rozmytych metod klasyfikacji i programu R / Assessment of financial condition of local government units with the use of fuzzy classification methods and program R	165
Dorota Rozmus: Badanie stabilności taksonomicznej czynnikowej metody odległości probabilistycznej / Stability of the factor probability distance clustering method	176
Adam Sagan, Aneta Rybicka, Justyna Brzezińska: <i>Conjoint analysis</i> oparta na modelach IRT w zagadnieniu optymalizacji produktów bankowych / An IRT-approach for conjoint analysis for banking products preferences.....	184
Michał Stachura: O szacowaniu centrum populacji określonego obszaru na przykładzie Polski / On estimating centre of population of a given territory. Poland's case	195
Michał Stachura, Barbara Wodecka: Wybrane aspekty i zastosowania modeli zdarzeń ekstremalnych / Selected facets and application of models of extremal events	205
Iwona Staniec, Jan Żółtowski: Wykorzystanie analizy log-liniowej do wyboru czynników determinujących współpracę w przedsiębiorczości	

technologicznej / Use of log-linear analysis for the selection determinants of cooperation in technological entrepreneurship.....	215
Marcin Szymkowiak, Wojciech Roszka: Potencjał gospodarczy gmin aglomeracji poznańskiej w ujęciu taksonomicznym / The economic potential of municipalities of the Poznań agglomeration in the light of taxonomy analysis.....	224
Lucyna Wojcieszka: Zastosowanie modeli klas ukrytych w badaniu opinii respondentów na temat roli państwa w gospodarce / Implementation of latent class models in the respondents' survey on the role of the country in economy.....	234

Wstęp

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania”, zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego.

W trakcie dwóch sesji plenarnych oraz 13 sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 27. Teksty 25 recenzowanych artykułów naukowych znajdują się w Taksonomii nr 26.

Krzysztof Jajuga, Marek Walesiak

Michał Stachura

Uniwersytet Jana Kochanowskiego w Kielcach
e-mail: michal.stachura@ujk.edu.pl

**O SZACOWANIU CENTRUM POPULACJI
OKREŚLONEGO OBSZARU
NA PRZYKŁADZIE POLSKI**

**ON ESTIMATING CENTRE OF POPULATION
OF A GIVEN TERRITORY. POLAND'S CASE**

DOI: 10.15611/pn.2016.427.20

Streszczenie: Metody centrograficzne nie tylko dostarczają narzędzi do prowadzenia analiz demograficznych i rozpoznawania typu sieci osadniczej pewnego terytorium, lecz także mogą wspomagać decyzje planistyczne. W pracy rozważane jest centrum populacji, czyli taki punkt danego terytorium, z którego łączna odległość mierzona do wszystkich mieszkańców jest najmniejsza. Głównym celem pracy jest zaproponowanie szacunkowej i skutecznej metody wyznaczania centrum populacji. Metoda ta bazuje na symulacyjnym i wielokrotnie powtarzanym wyborze punktu startowego, od którego rozpoczyna się procedura optymalizacyjna. Metodę tę zastosowano dla przypadku Polski z odwołaniem do danych z podziałem na gminy. Tak oszacowane centrum populacji Polski zlokalizowane jest nieopodal Pabianic. Pobocznym celem pracy jest analiza układów k centrów populacji danego terytorium, rozumianych jako k punktów takich, że zminimalizowana jest suma odległości od każdego mieszkańca do najbliższego spośród k centrów.

Słowa kluczowe: metody centrograficzne, centrum populacji, optymalizacja.

Summary: Centographic methods not only serve to carry out demographic analyses or to recognise types of settlement networks, but also they provide such tools that may assist in planning and decision making. The paper deals with a centre of population that is the point to which the whole population of a given territory has the smallest possible sum of distances. The main goal of the paper is to propose an approximate and effective numerical method that leads to determine the centre of population of a given territory. This method relies on a multiple simulation of a start point of the carried optimisation problem. According to this idea, the centre of population of Poland has been determined with reference to the division of Poland into communes or municipalities. An auxiliary goal of the paper is to analyse systems of k centres of population, which is understood as such k points of a territory that the sum of distances, measured from every citizen to his or her closest centre, is minimal.

Keywords: centographic technics, centre of population, optimisation.

1. Wstęp

Ogólnie rzecz ujmując, metody centrograficzne mają na celu wyznaczenie, stosownie do potrzeb definiowanych, miar tendencji centralnej badanej cechy, która rozłożona jest (zwykle nierównomiernie) na *a priori* określonym terytorium. Z perspektywy aspektów geograficznych miary te służą przede wszystkim rozpoznawaniu typu sieci osadniczej danego terytorium oraz szeroko rozumianym analizom demograficznym – w szczególności, gdy badanie prowadzone jest dodatkowo w ujęciu dynamicznym. Z perspektywy aspektów ekonomicznych metody centrograficzne dostarczają pewnych wskazówek¹, jak optymalizować lokalizację scentralizowanych usług (*sensu largo*).

Ze względu jednak na specyfikę metod centrograficznych typowe wielowymiarowe miary tendencji centralnej definiowane w przestrzeni wektorowej \mathbb{R}^n z naturalną metryką euklidesową wymagają uogólnienia. Dzieje się tak, gdyż analizowane terytorium nie powinno być traktowane ani jako obszar w \mathbb{R}^2 z odległością mierzoną na płaszczyźnie, ani też jako podzbiór \mathbb{R}^3 z odległością mierzoną w przestrzeni. Zatem za model terytorium należy przyjąć pewną dwuwymiarową powierzchnię – opisaną parametrycznie² – która, choć zawarta w przestrzeni \mathbb{R}^3 , to wyposażona jest w swoistą metrykę, pozwalającą liczyć odległości po tzw. geodezyjnych³.

Rozważaną w opracowaniu cechą jest liczba ludności, a wyznaczaną dla niej miarą tendencji centralnej jest **centrum populacji**⁴, rozumiane jako ten punkt danego terytorium, z którego łączna odległość mierzona do wszystkich mieszkańców jest najmniejsza (zob. [Kostrubiec 1972]). Jak widać, miara ta wprost nawiązuje do definicji mediany przestrzennej, lecz różni się od niej odwołaniem do innej metryki. Centrum populacji jest wyznaczone w nielicznych krajach świata, a zwykle czynią to instytucje państwowe (np. w Stanach Zjednoczonych jest to United States Census Bureau [<http://www.census.gov/2010census/data/center-of-population.php>]).

Zasadniczym celem opracowania jest zaproponowanie szacunkowej i skutecznej **metody wyznaczenia centrum populacji** danego terytorium⁵, opartej na pew-

¹ Wskazówek lub jedynie sugestii – na pewno nie są to bowiem jednoznaczne przesłanki, ponieważ omawiana metoda nie uwzględnia np. naturalnych barier terenowych.

² Najlepiej, gdy nadmieniona powierzchnia modelowana jest jako tzw. różniczkowa różniczkowa (różniczkowa riemannowska), która uwzględnia nie tylko fakt spłaszczenia kuli ziemskiej (elipsoida, geoida), lecz wręcz ukształtowanie powierzchni danego terytorium. Zob. np. [Austin 2015; Aboufadel, Austin 2006; Barmore 1993].

³ Geodezyjna jest taką linią powierzchni, która zawiera najkrótszą drogę łączącą parę punktów.

⁴ Centrum populacji (*centre of population*), zwane też środkowym centrum lub środkiem zbitejności, było rozważane już w 1872 r. (zob. [Hilgard 1872; Sviatlovsky, Eells 1937]).

⁵ Szacunkowej – w rozumieniu przybliżonej, zaś skutecznej – w sensie prostej, dającej się łatwo stosować i prowadzącej do uzyskania wyników w krótkim czasie.

nych upraszczających założeniach. Wśród celów dodatkowych znajdują się zastosowanie metody dla przypadku Polski oraz propozycja uogólnienia pojęcia centrum populacji – wraz z metodą jego szacowania i zastosowaniem ponownie dla przypadku Polski. Motywacją zaś do podjęcia takiej tematyki była chęć wypełnienia luki w wyznaczaniu centrum populacji Polski – i w literaturze, i w zasobach GUS stwierdzono brak takich opracowań.

2. Metodologia

Pierwsze przyjęte uproszczenie sprowadza się do założenia, że rozważane terytorium (fragment kuli ziemskiej) jest odpowiednim fragmentem sfery⁶ o arbitralnie przyjętym promieniu $r = 6371$ km. Nie są więc siłą rzeczy poruszane i uwzględniane takie kwestie, jak modelowanie globu za pomocą geoidy, a danego terytorium przez pryzmat ukształtowania jego terenu. Drugie natomiast uproszczenie dotyczy zarówno współrzędnych geograficznych, które notowane są w formacie stopni dziesiętnych, jak i geodezyjnego systemu odniesienia, który po prostu nie jest uwzględniany.

Następne zaś uproszczenia są bezpośrednio związane z analizowanymi cechą i miarą tendencji centralnej. Po trzecie więc, przyjmuje się, że rozmieszczenie całej populacji danego terytorium reprezentowane jest przez L punktów – zwanych dalej węzłami – spośród których każdy charakteryzowany jest współrzędnymi geograficznymi φ , λ (szerokość i długość) oraz przypisaną sobie liczbą ludności n .⁷

W świetle poczynionych założeń węzłowi w_i o współrzędnych geograficznych (φ_i, λ_i) , w którym umiejscowiono n_i osób, przypisywany jest punkt $p_i = (x_i, y_i, z_i) = r \cdot (\cos \frac{\pi\lambda}{180^\circ} \cdot \cos \frac{\pi\varphi}{180^\circ}, \cos \frac{\pi\lambda}{180^\circ} \cdot \sin \frac{\pi\varphi}{180^\circ}, \sin \frac{\pi\lambda}{180^\circ})$ na sferze o promieniu r w przestrzeni \mathbb{R}^3 (dla $i \in \{1, 2, \dots, L\}$).⁸ Aby wyznaczyć odległość ρ_{ij} na sferze między parą punktów $p_i = (x_i, y_i, z_i)$, $p_j = (x_j, y_j, z_j)$ tej sfery wystarczy obliczyć ich odległość euklidesową d_{ij} w \mathbb{R}^3 , a następnie przekształcić ją uwzględniając krzywiznę

⁶ Jest to daleko idące uproszczenie. Niemniej jednak, po pierwsze, błąd obliczeń tym powodowany nie zdaje się dominujący – przynajmniej dla przypadku Polski – wśród innych popełnianych błędów (zob. przyp. 7), a po drugie istota rozważanego modelu bez względu na przyjęty typ powierzchni pozostaje ta sama i może zostać odtworzona w innych przypadkach. Choć pewnie o wiele trudniej będzie procedurę wdrożyć w przypadku bardzo wiernego opisu ukształtowania terenu.

⁷ Oczywiście jest, że ze względu na – powodowaną wielorakiej natury czynnikami – permanentną mobilność przypisanie każdej osoby do określonego punktu terytorium, a następnie grupowanie osób w określonych węzłach jest daleko idącym uproszczeniem. Co więcej, właśnie chyba w tym miejscu należy upatrywać czynnika generującego zasadniczy, a przy tym absolutnie nieunikniony, błąd prowadzonych szacunków centrum populacji.

⁸ W przypadku współrzędnych geograficznych przekraczających skrajne wartości ich zakresów, wystarczy dokonać ich rozsądnej zamiany, np. ciąg współrzędnych: 179 °E, 180 °E, 179 °W zastąpić ciągiem: 179 °E, 180 °E, 181 °E lub 181 °W, 180 °W, 179 °W. W sensie teorii geometrii różniczkowej wybiera się po prostu inną, *nomen omen*, mapę.

sfery: $\rho_{ij} = 2r \cdot \arcsin \frac{d_{ij}}{2r}$ (dla $i, j \in \{1, 2, \dots, L\}$). Z kolei jako oszacowanie centrum populacji przyjmuje się punkt w_0 (niekoniecznie będący węzłem), który realizuje minimum wyrażenia: $m(w_0) = \sum_{i=1}^L n_i \rho_{0i}$. Wartość $n_i \rho_{0i}$ naturalnie interpretuje się jako łączną odległość z punktu w_0 do wszystkich osób w węźle w_i .

W tym miejscu widoczne staje się użycie parametryzacji. Otóż, pewien fragment sfery o promieniu r parametryzowany jest fragmentem płaszczyzny \mathbb{R}^2 poprzez ciągłą i gładką funkcję: $(\varphi_i, \lambda_i) \rightarrow (x_i, y_i, z_i)$. Użycie parametryzacji gwarantuje, że wyznaczone centrum populacji (jak również i inne ewentualnie rozważane miary tendencji centralnej) znajduje się na sferze (a w pełni ogólności na stosownej podrozmaitości).

Osobną kwestią jest sposób wyznaczenia argumentu w_0 realizującego minimum wyrażenia $m(w_0)$. W opracowaniu przyjęto następujący, prosty algorytm⁹.

1. Losowany jest punkt (para współrzędnych geograficznych) spomiędzy najmniejszych i największych współrzędnych geograficznych terytorium.
2. Punkt ten jest punktem startu procedury optymalizacyjnej. (Użyto funkcji `optim` z jej domyślnymi ustawieniami.)
3. Wyniki procedury (argument w_0 i wartość $m(w_0)$) są zapisywane.
4. Kroki 1.–3. powtórzone są J razy i wybrany jest argument, któremu odpowiada najmniejsza wartość.

J repetycji zostało wprowadzonych jedynie po to, by zwiększyć prawdopodobieństwo trafienia – w sensie osiągnięcia minimum. Jeżeli bowiem prawdopodobieństwo trafienia w jednej iteracji jest równe p , to prawdopodobieństwo p' choć jednego trafienia w J iteracjach wynosi $p' = 1 - (1 - p)^J$. Dla $J = 1000$ (tak właśnie przyjęto w dalszym ciągu) oraz $p \geq 0,01$ mamy i tak $p' \geq 0,999956$.

3. Centrum populacji terytorium Polski

W przykładzie ilustrującym metodę szacowania centrum populacji rozważono przypadek terytorium Polski. Jako węzły przyjęto siedziby gmin (stąd $L = 2479$). Posłużono się bazą współrzędnych geograficznych siedzib gmin z portalu bazamiejscowosci.pl [<http://www.bazamiejscowosci.pl/>] oraz danymi dotyczącymi liczby ludności z podziałem na gminy za lata 2011–2014 publikowanymi przez Główny Urząd Statystyczny [<http://stat.gov.pl/>], w tym z 2011 r. z Narodowego Spisu Powszechnego.

Zastosowanie opisanego algorytmu poskutkowało wyznaczeniem centrum populacji Polski w poszczególnych latach. Jego współrzędne geograficzne zamieszczone zostały w tab. 1., w której dodatkowo podano także węzły najbliższe temu centrum (kolumna: gmina i powiat) oraz liczby tych spośród J iteracji, w przypad-

⁹ Wszystkie obliczenia oraz wykresy wykonane zostały z użyciem autorskich kodów napisanych w programie R ([R Development Core Team 2015]).

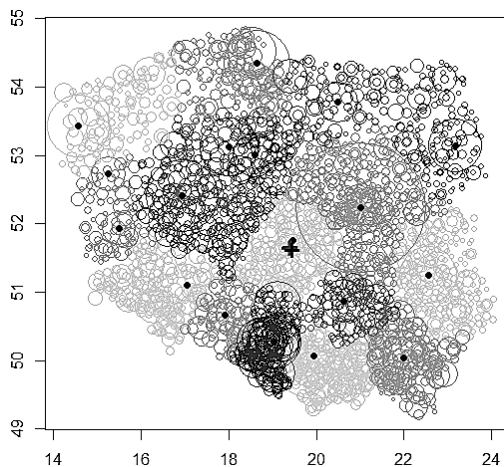
ku których procedura optymalizacyjna wskazywała na ten sam węzeł, co w iteracji najlepszej (kolumna: zgodność). Stwierdzone wartości równe 1000 pozwalają szacować już prawdopodobieństwo p jako równe wręcz 1. Tym bardziej więc uznać można prawdopodobieństwo p' za równe 1.

Tabela 1. Lokalizacja centrum populacji w latach 2011–2014

Rok	Centrum populacji		Najbliższy węzeł		Zgodność
	szerokość	długość	gmina	powiat	
2011	51,6591	19,4198	Ksawerów	pabianicki	1000
2012	51,6609	19,4002	Ksawerów	pabianicki	1000
2013	51,6618	19,4209	Ksawerów	pabianicki	1000
2014	51,6282	19,4682	Rzgów	łódzki wschodni	1000

Źródło: opracowanie własne na podstawie obliczeń w programie R.

Wizualizację położenia centrum populacji w poszczególnych latach stanowi schematyczna mapa Polski zamieszczona na rys. 1. Widoczne na niej okręgi symbolizują gminy, przy czym powierzchnie zamknięte przez te okręgi są proporcjonalne do liczby mieszkańców gmin. Z kolei czarnymi wypełnionymi punktami zaznaczono miasta wojewódzkie.



Rys. 1. Lokalizacja centrum populacji w latach 2011, 2012, 2013 i 2014 (oznaczone odpowiednio: +, +, +, +)

Źródło: opracowanie własne na podstawie danych z GUS i portalu bazamiejscowości.pl.

Centrum populacji w latach 2011–2014 na tyle nieznacznie zmienia swoje położenie, że jego znaczniki na rys. 1 wręcz się pokrywają. Ma to także swoje odbicie w tab. 2 prezentującej odległości między lokalizacjami centrum w poszczególnych latach (odległości mierzone konsekwentnie na modelowej sferze).

Tabela 2. Odległości między lokalizacjami centrum populacji w latach 2011–2014

	2011	2012	2013
2012	1,367 km		
2013	0,309 km	1,431 km	
2014	4,792 km	5,935 km	4,962 km

Źródło: opracowanie własne na podstawie obliczeń w programie R.

Dodać należy, że średni czas wszystkich obliczeń wykonywanych w pojedynczej iteracji wynosił ok. 8,5 sekundy¹⁰.

4. Układ k centrów populacji

Jako uogólnienie pojęcia centrum populacji, zaproponowano **układ k centrów populacji**, czyli zbiór takich k punktów danego terytorium, że zminimalizowana jest suma odległości od każdego mieszkańca do najbliższego temu mieszkańcowi jednego spośród k wyróżnionych punktów.

Wprowadzenie pojęcia układu k centrów $(w_{\vartheta_1}, \dots, w_{\vartheta_k})$ skutkuje stosownymi zmianami w rozważanych uprzednio i metodzie szacowania, i algorytmie. Otóż, zamiast losować jeden punkt losuje się ich k , a optymalizowana jest k razy więcej wymiarowa funkcja celu: $m(w_{\vartheta_1}, \dots, w_{\vartheta_k}) = \sum_{i=1}^L n_i \min(\rho_{\vartheta_1 i}, \dots, \rho_{\vartheta_k i})$, która tym razem jest nieróżniczkowalna. Nie nastęrcza to jednak większych trudności obliczeniowych. Okazuje się ponadto, że średni czas obliczeń w pojedynczej iteracji wzrasta niewspółmiernie wolniej, niż wzrasta liczba rozważanych w układzie centrów populacji, gdyż jest to czas odpowiednio ok. 10 i 12 sekund dla układów 2 i 3 centrów populacji – a właśnie przy $k = 2$ i $k = 3$ zastosowano uogólnioną procedurę i wyznaczono układy centrów dla Polski.

Tabela 3. Lokalizacja układu dwu centrów populacji w latach 2011–2014

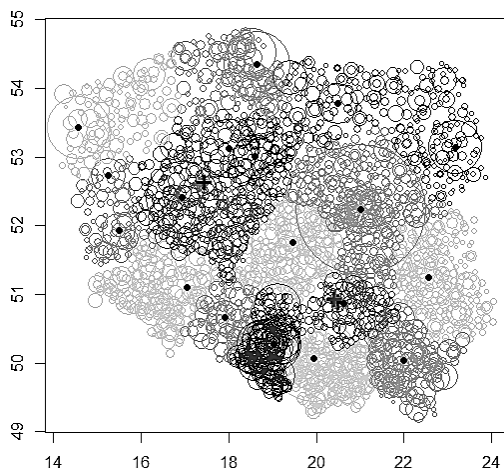
Rok	2 centra populacji				Najbliższe węzły		Zgodność
	punkt 1		punkt 2		punkt 1	punkt 2	
	szer.	Dług.	Szer.	Dług.	Gmina/powiat	gmina/powiat	
2011	52,6618	17,4533	50,9295	20,4182	Mieleszyn/ gnieźniński	Strawczyn/ kielecki	581
2012	52,6125	17,3500	50,9507	20,4366	Klecko/ gnieźniński	Strawczyn/ kielecki	622
2013	52,6571	17,4436	50,9343	20,4254	Klecko/ gnieźniński	Strawczyn/ kielecki	474
2014	52,6455	17,4518	50,8973	20,5162	Klecko/ gnieźniński	Piekoszów/ kielecki	509

Źródło: opracowanie własne na podstawie obliczeń w programie R.

¹⁰ Czas (ten i kolejne) obliczeń na „domowym” laptopie.

Współrzędne geograficzne układu dwu centrów populacji zamieszczone zostały w tab. 3, w której dodatkowo podano także analogiczne jak w tab. 1 informacje, tzn. węzły najbliższe tym centrom oraz liczby „dobrych”¹¹ spośród J iteracji. Tym razem nadal prawdopodobieństwo p' można uznać za bardzo bliskie 1, jeśli przyjąć za oszacowanie prawdopodobieństwa p odsetki „dobrych” spośród wszystkich J iteracji.

Z kolei wizualizację położenia układu dwu centrów populacji w poszczególnych latach stanowi schematyczna mapa Polski zamieszczona na rys. 2.



Rys. 2. Układ dwu centrów populacji w latach 2011, 2012, 2013 i 2014 (oznaczone odpowiednio: +, +, +, +)

Źródło: opracowanie własne na podstawie danych z GUS i portalu bazamiejscowości.pl.

Maksymalne przemieszczenia punktów w układzie dwu centrów populacji wynoszą w latach 2011–2014 po 8,862 km i 8,145 km odpowiednio w okolicy Gniezna i Kielc. Są to znów niezbyt znaczące przemieszczenia i dlatego też znaczniki centrów na rys. 2 pokrywają się w znacznym stopniu.

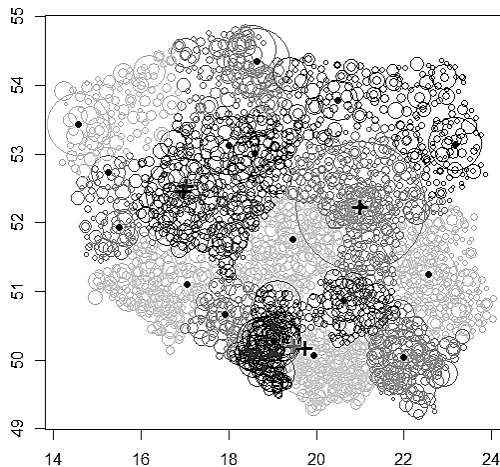
Analogiczne tabela i mapa (tab. 4 i rys. 3) prezentują dane i lokalizacje w odniesieniu do oszacowanego układu trzech centrów populacji. Tym razem maksymalne przemieszczenia punktów w układzie trzech centrów populacji wynoszą w latach 2011–2014 po 11,590 km, 21,019 km i 0,001 km odpowiednio w okolicy Poznania, Krakowa i Warszawy. Najbardziej wyraźne przemieszczenie następuje w okolicy Krakowa, gdzie centrum w 2014 r. najsilniej odstaje od wcześniejszych bardzo bliskich sobie (maksymalnie do 3,434 km) lokalizacji. Choć i tak trudno jest uznać to przemieszczenie za szczególnie znamienne.

¹¹ Dobrych, czyli tych iteracji, w przypadku których procedura optymalizacyjna wskazywała na te same k węzłów równocześnie, co w iteracji najlepszej.

Tabela 4. Lokalizacja układu trzech centrów populacji w latach 2011–2014

Rok	Punkt 1		Punkt 2		Punkt 3		Zgodność
	szer.	dług.	szer.	dług.	szer.	dług.	
2011	52,5497	17,0620	50,2225	19,4616	52,2297	21,0122	85
	Murowana Goślina/ poznański		Bukowno/ olkuski		Warszawa		
2012	52,4644	16,9636	50,2100	19,5057	52,2297	21,0122	50
	Czerwonak/ poznański		Trzebinia/ chrzanowski		Warszawa		
2013	52,5443	17,0541	50,2170	19,4747	52,2297	21,0122	104
	Murowana Goślina/ poznański		Bukowno/ olkuski		Warszawa		
2014	52,4696	16,9899	50,1843	19,7508	52,2297	21,0122	191
	Czerwonak/ poznański		Jerzmanowice-Przebinia/ krakowski		Warszawa		

Źródło: opracowanie własne na podstawie obliczeń w programie R.



Rys. 3. Układ trzech centrów populacji w latach 2011, 2012, 2013 i 2014 (oznaczone odpowiednio: +, +, +, +)

Źródło: opracowanie własne na podstawie danych z GUS i portalu bazamiejscowości.pl.

Odnotować jeszcze należy, że w przypadku zarówno układu dwu, jak i trzech centrów populacji jeden z punktów tworzących te układy jest wyraźnie „ciągnięty” w kierunku południowego wschodu (okolice Kielc), lub południa kraju (obszar między Katowicami a Krakowem).

5. Podsumowanie

Zamieszczony w opracowaniu opis prezentowanej metody, a także zgodnie z nią przeprowadzone szacowanie centrum populacji Polski uwidaczniają pożądane własności samej metody: prostotę oraz szybkość obliczeniową. Po części szybkość ta jest bezpośrednim skutkiem specyfiki sfery jako modelu badanego terytorium, gdyż odległość między każdą parą punktów sfery zależy jedynie od ich względnego (wobec siebie), a nie bezwzględnego położenia, co umożliwiło skorzystanie z wbudowanej w R funkcji `dist` wyznaczającej macierz odległości euklidesowych i następnie jej bezpośrednie przekształcenie w macierz odległości na sferze.

Wobec powyższego, z punktu widzenia analiz geograficzno-demograficznych, prezentowana metoda może dostarczać wartościowych informacji o rozmieszczeniu ludności na danym terytorium – również w ujęciu dynamicznym, tym bardziej, że barierą w tym względzie są raczej czas i częstość zbierania danych demograficznych, a nie czas wykonywania na podstawie tych danych niezbędnych obliczeń¹².

Z kolei, z punktu widzenia nadmienionych we wstępie aspektów ekonomicznych – w tym optymalizacji lokalizacji scentralizowanych usług – możliwa jest przynajmniej dwójaka adaptacja prezentowanej metody, prowadząca do wiernego oddania specyfiki zarówno analizowanego zagadnienia, jak i odpowiadającego mu adekwatnego i rozsądnego pomiaru dystansu (w sensie np. odległości drogowej, kolejowej, czy też czasu przejazdu). Po pierwsze, możliwe jest dokonanie takiej deformacji modelowej powierzchni, by lepiej dało się odzwierciedlić pożądaną metrykę. Po drugie, rozważania z całej modelowej powierzchni można zawęzić jedynie do samych węzłów i zadać metrykę przez adekwatną macierz odległości między nimi¹³.

W odniesieniu do centrów populacji wyznaczonych dla terytorium Polski nadmienić jeszcze należy, że uzyskane oszacowania świetnie odzwierciedlają specyfikę rozmieszczenia gęstości ludności naszego kraju (wyraźnie większa od średniej gęstość występuje na obszarach województw śląskiego, małopolskiego i podkarpackiego). Pojedyncze centrum populacji leży bowiem na południe od geometrycznego środka Polski, który znajduje się w gminie Piątek na północ od Łodzi. Z kolei w przypadku podwójnego centrum populacji jeden z punktów (nieopodal Kielc) zlokalizowany jest dość blisko nadmienionego obszaru większej gęstości ludności, a dla potrójnego – punkt taki leży już wewnątrz tego obszaru.

¹² Przedstawiony problem opisu dynamiki położenia miar tendencji centralnej wpisuje się w szerszy nurt tego typu badań, uwzględniający ponadto pomiar stopnia niezrównoważenia przestrzennego cechy (por. np. [Zhao, Stough, Li 2003]). Warto w tym miejscu wspomnieć też analizy dotyczące rozmieszczenia innych cech – m.in: zasobu bogactwa (np. [Grasland, Madelin 2001]), wielkości produkcji (np. [Grether, Mathys 2014]), czy podaży energii i popytu na nią (np. [Zhang i in. 2012]).

¹³ W tym nurcie są m.in. badania dotyczące lokalizacji usług medycznych – por. [Bagheri, Benwell, Holt 2005].

Literatura

- Aboufadel E., Austin D., 2006, *A new method for computing the center of population of the United States*, *The Professional Geographer*, vol. 58, no. 1, s. 65–69.
- Austin D., *The Center of Population of the United States*, <http://www.ams.org/samplings/feature-column/fcarc-population-center> (17.10.2015).
- Bagheri N., Benwell G.L., Holt A., 2005, *Measuring spatial accessibility to primary health care*, *The 17th Annual Colloquium of the Spatial Information Research Centre*, s. 103–108.
- Barmore F.E., 1993, *Where are we? Comments on the Concept of Center of Population*, *The Wisconsin Geographer*, no. 9, s. 8–21.
- Grasland C., Madelin M., 2001, *The unequal distribution of population and wealth in the world*, *Populations & Sociétés*, no.368, s. 1–4.
- Grether J.-M., Mathys N., *Is the World's Economic Center of Gravity Already in Asia?*, <http://www.etsg.org/ETSG2007/papers/grether.pdf> (17.5.2014).
- Hilgard J.E., 1872, *The advance of population in the United States*, *Scribner's Monthly*, no. 4, s. 214–218.
- Kostrubiec B., 1972, *Analiza zjawisk koncentracji w sieci osadniczej*, Zakład Narodowy im. Ossolińskich, Wrocław.
- R Development Core Team, 2015, *R: A language and environment for statistical computing*, The R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>.
- Sviatlovsky E.E., Eells W.C., 1937, *The centrographical method and regional analysis*, *Geographical Review*, vol. 27, s. 240–254.
- Zhang Y., Zhang J., Yang Z., Li J., 2012, *Analysis of the distribution and evolution of energy supply and demand centers of gravity in China*, *Energy Policy*, vol. 49, s. 695–706.
- Zhao Z., Stough R.R., Li N., 2003, *Note on the measurement of spatial imbalance*, *Geographical Analysis*, vol. 35, no. 2, s. 170–176.

Źródła internetowe

<http://stat.gov.pl>.

<http://www.census.gov/2010census/data/center-of-population.php> (12.10.2015).

<http://www.bazamiejscowosci.pl/> (17.05.2014)