

Alfred Stach, Patrycja Wysocka

ZASTOSOWANIE METODY KRIGINGU POISSONA W BADANIACH ROZKŁADU PRZESTRZENNEGO PROBLEMÓW SPOŁECZNYCH NA PRZYKŁADZIE POZNANIA

Analiza przestrzenna danych społecznych wymaga niejednokrotnie odfiltrowania wpływu nierealnych, odstających danych. Celem pracy jest omówienie podstaw teoretycznych bardzo efektywnej, a mało znanej metody do tego służącej – krigingu Poissona. Ilustrację praktyczną jej zalet przedstawiono na przykładzie identyfikacji obszarów występowania różnych kategorii problemów społecznych na obszarze Poznania.

Słowa kluczowe: *problemy społeczne, struktura społeczno-przestrzenna, Poznań, kriging Poissona*

1. Wstęp

Znajomość rozkładu przestrzennego występowania charakterystyk społecznych, takich jak poziom bezrobocia, przestępczości, alkoholizmu, ma duże znaczenie nie tylko dla rozwoju nauk społecznych, ale przede wszystkim dla służb publicznych oraz administracji. Pozwala wnioskować na temat ich przyczyn i na tej podstawie planować skuteczną politykę ich ograniczania. Wiedza na ten temat, często dobra na poziomie lokalnym, jest niewystarczająca, a czasami mylna, kiedy analizowany obszar staje się większy, bardziej zróżnicowany pod względem gęstości zaludnienia i struktury populacji. Efektywność działań prewencyjnych jest uzależniona od prawidłowej delimitacji stref/obszarów występowania problemów społecznych. Ryzyko ich pojawiania się jest oceniane na podstawie danych częstości zanotowanych w statystykach policyjnych czy służb społecznych i może być bardzo niepewne kiedy obliczane jest dla jednostek przestrzennych o niskim zaludnieniu lub odnoszone do małej części całej populacji. Dodatkowa niepewność wiąże się z koniecznością ochrony prywatności uniemożliwiającą identyfikację poszczególnych osób i związaną z tym agregacją danych punktowych do jednostek przestrzennych.

Problem ten ma znaczenie wykraczające poza nauki społeczne, ponieważ dane przestrzenne o charakterze zliczeń odnoszących się do określonego obszaru czy odcinka czasu (a więc charakteryzowane z pomocą rozkładu Poissona), są rutynowo zbierane w obrębie wielu dziedzin z zakresu nauk o Ziemi, takich jak:

epidemiologia, ekologia, geografia fizyczna. Wśród najczęstszych celów analizy tego rodzaju danych jest określenie wpływu na rejestrowaną liczebność zróżnicowanych przestrzennie różnych czynników sprzyjających (problem wyjaśniania) oraz problemy szacowania tej liczebności dla obszarów czy punktów, z których nie ma danych empirycznych (problem przewidywania).

Zazwyczaj, przy analizie przestrzennej danych społecznych, konieczna jest korekta nierealnych częstości empirycznych wynikających z tzw. problemu małych liczb (Buescher 1997; Waller, Gotway 2004). Występuje on na obszarach, gdzie jednostki przestrzenne charakteryzują się dużym zróżnicowaniem liczby mieszkańców. Dla części z nich mających bardzo małe zaludnienie, niewielkie, losowe zróżnicowanie liczby przypadków analizowanej cechy powoduje, że obliczenie wskaźników standaryzowanych w stosunku do populacji (np. liczba przypadków/1 000 mieszkańców) jest obciążone dużym i nieznanym błędem. Problem ten nie rozwiązuje stosowanie popularnych deterministycznych metod wygładzania danych przestrzennych, takich jak: średnie lokalne ważone populacją, regresja nieparametryczna (jądrowy estymator gęstości) lub różne warianty empirycznego wygładzania Bayesa (Kafadar 1994; Talbot i in. 2000; Pfeiffer i in. 2008). Nie uwzględniają one bowiem występującej prawie zawsze autokorelacji przestrzennej i prowadzą często do nadmiernego wygładzenia rozkładu i zamazania realnie istniejących skupień niskich i wysokich wartości analizowanej zmiennej. Poza tym standardowo metody te nie dają możliwości oceny błędu (niepewności) wygładzonych częstości empirycznych. Dlatego też nowe propozycje metodyczne, umożliwiające rozwiązanie lub choć zmniejszenie roli wyżej wymienionych problemów, są bardzo ważne i warte upowszechnienia. Przykładem z tego właśnie zakresu jest kriging Poissona (dalej określanego w skrócie jako PK), mający swoje korzenie w geostatystyce – metoda analizy i prognozowania rozkładów przestrzennych danych społecznych zaproponowana m.in. przez M.A. Oliver i in. (1998), a praktycznie zaimplementowana i opisana przez P. Goovaerts (2005). Obliczenia wykonane przez wspomnianego autora oraz P. Goovaerts i S. Gebreaba (2008) na danych symulowanych dla różnych scenariuszy częstości rejestrowanych zdarzeń, wielkości populacji i przestrzennego rozkładu ryzyka wykazały, że PK daje bardziej precyzyjne prognozy niż inne metody wygładzania częstości empirycznych. Jedyne wyjątek stanowi, rzadko spotykany przypadek, kiedy rozkład ryzyka w przestrzeni jest całkowicie losowy.

Metodyka PK wraz z kodem komputerowym realizującym jej obliczenia została, jak wspomniano wyżej, opisana przez P. Goovaerts w 2005 r. w nowo wówczas założonym czasopiśmie specjalizującym się w zagadnieniach geografii medycznej. Mało znane czasopismo, a także stosunkowo trudna od strony matematycznej procedura PK spowodowała, że mimo wielu zalet nie zdobyła ona w ciągu kolejnych lat większej popularności. Opublikowanych zostało zaledwie kilka prac wykorzystujących PK i to w dalszym ciągu głównie w zakresie analizy przestrzennej danych zachorowalności (Ali i in. 2006;

Goovaerts 2006, 2008; Shao i in. 2009; Ebenezer i in. 2013; Naeimeh i in. 2013;). Inne znane zastosowania obejmowały analizę przestrzenną zliczeń artefaktów archeologicznych (Bevan, Conolly 2009), przypadków różnych rodzajów przestępstw (Kerry i in. 2010; Bumpus 2012) oraz rozmieszczenia gatunków zwierząt (Kober i in. 2010; Kerry i in. 2012). Przegląd rozwoju metodyki krugingu Poissona został ostatnio omówiony od strony teoretycznej przez V. De Oliveirę (2013, 2014). Autorom niniejszej pracy nie jest znana żadna publikacja, w której PK byłby użyty przez polskiego autora. Na wzrost popularności tej metodyki w najbliższych latach może wpłynąć fakt zaimplementowania jej ostatnio w module Geostatistical Analyst powszechnie używanego komercyjnego pakietu GIS – ArcGIS wersja 10.1 firmy ESRI (Krivoruchko i in. 2011).

Celem niniejszego opracowania było omówienie podstaw teoretycznych PK, a także ilustracja praktyczna zalet tej metodyki analizy przestrzennej danych społecznych na przykładzie miasta Poznania. Zadaniem, które sobie postawiono, była identyfikacja obszarów, w obrębie granic administracyjnych miasta, występowania różnych kategorii problemów społecznych na podstawie informacji o miejscu zamieszkania osób pobierających zasiłki za pośrednictwem Miejskiego Ośrodka Pomocy Społecznej.

2. Obszar badań, dane i metodyka analiz

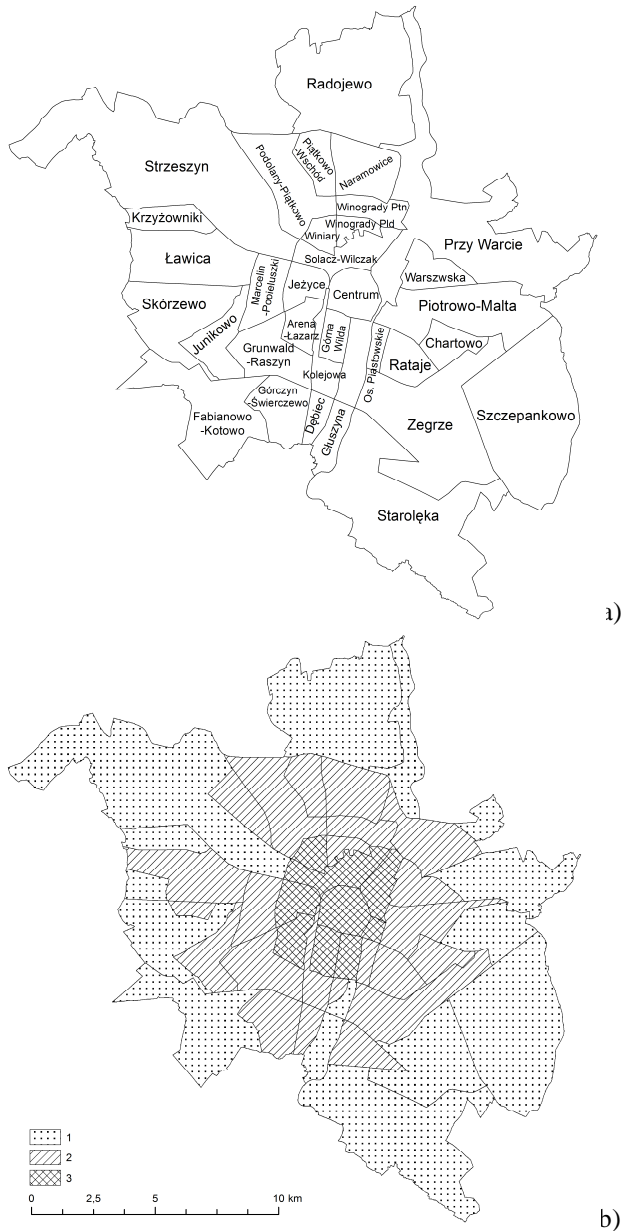
2.1. Obszar badawczy

Obszarem objętym badaniami było miasto Poznań. W obrębie miasta można wyróżnić trzy sektory: strefę centralną, pierścień zewnętrzny miasta i część peryferyjną (Parysek, Mierzejewska 2006) (rys. 1b).

Strefa centralna jest najgęściej zaludniona (do 13 000 os./km² zamieszkujących obszar jednego osiedla¹), chociaż pełni głównie funkcję usługowo-handlową.

Jednocześnie jest to obszar mocno zdegradowany, charakteryzujący się starym budownictwem (kamienice czynszowe z drugiej połowy XIX i początku XX w.) i zwartą zabudową, typową m.in. dla takich osiedli, jak: Jeżyce, Łazarz, czy Górna Wilda. Pierścień zewnętrzny spełnia funkcję mieszkaniowo-przemysłową. W jego zabudowie dominują zarówno nowe (np. Piątkowo czy Wschód), jak i stare osiedla mieszkaniowe (Warszawska, os. Piastowskie, większa część Rataj). Jest to tzw. zabudowa blokowa, która gdzieś tam przecinana jest przez gęste osiedla domków jednorodzinnych. Gęstość zaludnienia w tym obszarze osiąga wartość do 5 500 os./km². Ostatnia, peryferyjna strefa spełnia głównie funkcję rekreacyjną i rolniczą, choć zaczynają się rozwijać na niej coraz inten-

¹ Podstawowe jednostki administracyjne.



Rys. 1a. 33 jednostki administracyjne, dla których znana była struktura demograficzna
 Źródło: Prognoza Centrum Statystyki Regionalnej, Uniwersytet Ekonomiczny

Rys. 1b. Podział miasta Poznania ze względu na strukturę funkcjonalno-przestrzenną:
 1) część peryferyjna, 2) pierśień zewnętrzny miasta, 3) strefa centralna
 Źródło: opracowanie na podstawie: J.J. Parysek, L. Mierzejewska (2006)

sywniej nowe osiedla mieszkaniowe np. Szczepankowo. Dominuje luźna zabudowa jednorodzinna. Obszar ten zamieszkuje średnio 200–400 os./km² (rys. 1a).

Przestrzenny rozwój Poznania charakteryzują podobne zjawiska jak inne duże miasta (Parysek, Mierzejewska 2006), w tym: urbanizacja strefy podmiejskiej, postępująca przestrzenna segregacja oraz wzrastający problem nierówności społecznej. Pośród poważnych problemów społecznych, z którymi borykają się lokalne władze, należy wymienić bezdomność, biedę oraz alkoholizm. Opracowanie szczegółowego planu przeciwdziałania wymienionym powyżej problemom wymaga na początku ich dokładnego rozpoznania, a zasięg przestrzenny jest jego zasadniczym elementem.

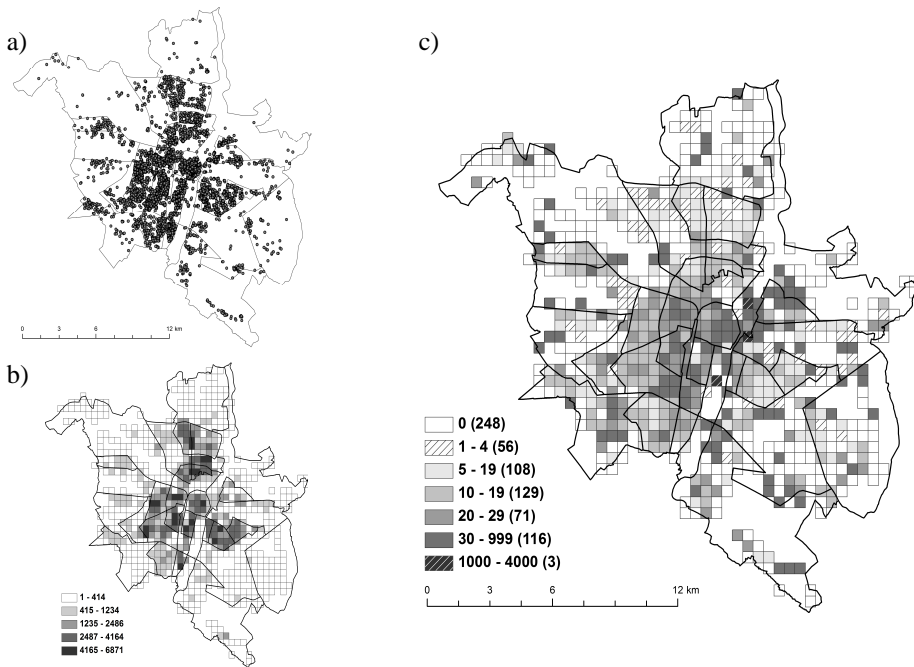
2.2. Dane

Podstawowymi danymi wykorzystanymi w analizie są informacje pochodzące z Miejskiego Ośrodka Pomocy Rodzinie w Poznaniu, dotyczące osób korzystających z pomocy społecznej w 2008 r. – miejsc ich zamieszkania, jak również wieku osoby, która pomoc uzyskała, liczebności jej rodziny oraz powodu przyznania pomocy.

MOPR przyznaje pomoc społeczną w 13 kategoriach: ubóstwo (UBS), bezdomność (BDOM), bezrobocie (BROB), bezradność (BRAD), alkoholizm (ALK), długotrwała lub ciężka choroba (CHOR), niepełnosprawność (NIEP), posiadanie rodziny wielodzietnej (WIELDZIET), potrzeba ochrony macierzyństwa (MACIERZ), sieroctwo (SIER), przemoc (PRZEM), narkomania (NAROM) oraz kategoria „inne” (INNE). W analizowanym okresie MOPR udzielił wsparcia finansowego 9 473 osobom (rys. 2a), co uwzględniając całe rodziny (18 264 osób) stanowiło ok. 3,3% populacji miasta (545 tys. w 2008 r.).

Punktowe dane miejsc zamieszkania osób pobierających pomoc społeczną zagregowano do większych obszarów, dla których posiadano informacje pochodzące z Zarządu Geodezji i Katastru Miejskiego miasta Poznania (GEOPOZ; <http://www.geopoz.pl>), a dotyczące liczby ich mieszkańców. Była to siatka złożona z 731 kwadratów (rys. 2b) o wymiarach 0,5 na 0,5 km (0,25 km²). Wiązało się to głównie z koniecznością standaryzacji ich zagęszczenia w odniesieniu do gęstości zaludnienia (liczba klientów MOPR/1 000 mieszkańców miasta). Rozmiar zastosowanej siatki jest wystarczający do dobrego rozpoznania zróżnicowania przestrzennego charakterystyk społecznych w obrębie Poznania, który ma powierzchnię 262 km².

Jak jednak widać na załączonej rycinie (rys. 2c) uzyskany w ten sposób obraz jest oczywiście w dalszym ciągu niewystarczająco precyzyjny, by możliwym była identyfikacja obszarów najbardziej obciążonych problemami społecznymi na poziomie lokalnym – sektora ulic. Należy również zwrócić uwagę, że niektóre z otrzymanych w ten sposób wartości przekraczają liczbę 1 000 klientów MOPR na 1 000 mieszkańców. Ten problem zostanie szerzej omówiony w kolejnym podrozdziale.



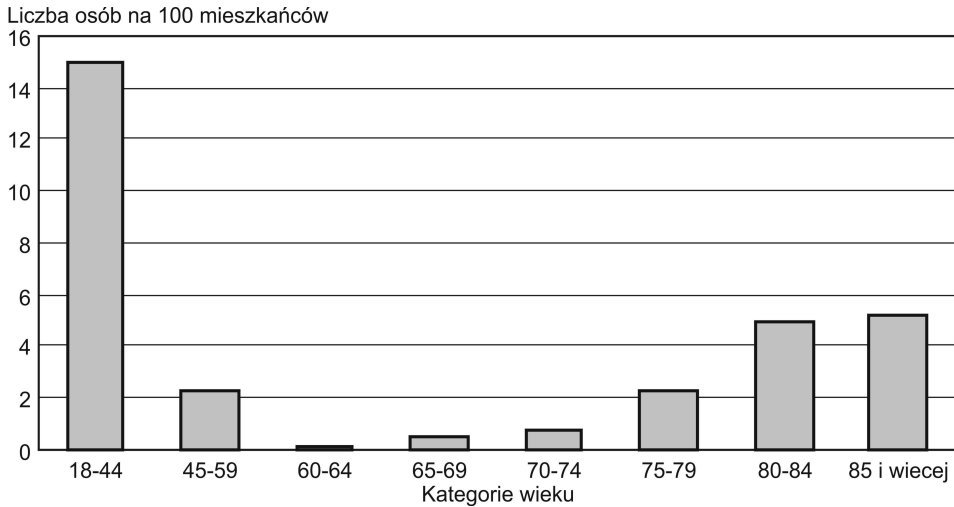
Rys. 2a. Rozmieszczenie przestrzenne danych dotyczących klientów MOPR na obszarze Poznania

Rys. 2b. Rozmieszczenie przestrzenne danych dotyczących gęstości zaludnienia w mieście (osób/0,25 km²) – klasyfikacji liczby mieszkańców w poligonach dokonano w oparciu o naturalne granice – metoda Jenksa (Jenks 1967)

Rys. 2c. Rozmieszczenie przestrzenne danych dotyczących liczby klientów MOPR w przeliczeniu na 1 000 mieszkańców – podział na klasy arbitralny (w nawiasie podano ilość przypadków na klasę)

Źródło: dane pochodzące z Urzędu Miasta Poznania

Ponadto w trakcie wstępnej analizy danych okazało się, że struktura wiekowa klientów MOPR różni się znacząco od całej populacji miasta. Różnice w liczbie osób reprezentujących daną grupę wiekową sięgają do 15 na 100 mieszkańców (rys. 3). Dotyczą tych grup mieszkańców, które są najbardziej narażone na występowanie takich problemów, jak długotrwała choroba (osoby starsze), czy bezrobocie (osoby młode, dopiero rozpoczynające dorosłe życie). Dane należało zatem poddać standaryzacji, nie tylko ze względu na gęstość zaludnienia, ale także rozkład wieku.



Rys. 3. Różnice udziału mieszkańców w poszczególnych klasach wieku pomiędzy całością mieszkańców Poznania a zbiorem klientów MOPR

Źródło: dane dotyczące rozkładu wieku mieszkańców pochodzący z Prognozy CSR dla 2010 r.

2.3. Standaryzacja danych społecznych z wykorzystaniem narzędzi GIS

Standaryzacja danych dotyczących klientów MOPR uwzględniająca ich strukturę wiekową była bardzo czasochłonnym zadaniem ze względu na liczbę porównywanych jednostek przestrzennych (731 poligonów). Do obliczeń tych opracowano zatem skrypt, działający w ArcGIS, który pozwolił zautomatyzować procedurę standaryzacji (statystyka danych punktowych w poligonach). Algorytm obliczeniowy został oparty na metodzie stosowanej w badaniach nad śmiertelnością (Anderson, Rosenberg 1998). W ten sposób dla każdego z 13 problemów społecznych występujących wśród klientów MOPR dokonano osobnej standaryzacji w obrębie każdego z 731 oczek siatki, zgodnie ze wzorem [1]:

$$R_j = \sum_{j=1}^m \sum_{i=1}^n w_i \frac{n_{ij} \times 1000}{d_{ij}} \quad [1]$$

R_j – ilość klientów MOPR na 1 000 mieszkańców zamieszkujących obszar j po uwzględnieniu poprawki ze względu na strukturę wiekową osób pobierających pomoc społeczną;

w_i – waga, wyliczona na podstawie struktury wiekowej mieszkańców Poznania;

n_{ij} – liczba klientów MOPR, będąca w grupie wiekowej i , zarejestrowana na obszarze j ;

D_{ij} – liczba mieszkańców będąca w grupie wiekowej i , zarejestrowana na obszarze j ;
 i – jedna z n grup wiekowych (tu 8 grup);
 j – dany obszar (jedno z m oczek siatki zawierającej informacje odnośnie ilości mieszkańców – tu 731 obszary);

Uzyskana w ten sposób mapa rozmieszczenia klientów MOPR w przeliczeniu na jednego mieszkańca nie różniła się w sposób znaczący od mapy nie uwzględniającej poprawki ze względu na wiek. Niemniej, otrzymana została dzięki temu dodatkowa informacja o rozkładzie przestrzennym analizowanych problemów społecznych. Wiadomo, że większe zagęszczenie klientów MOPR w pewnych obszarach miasta nie wynikało bezpośrednio ze specyficznej struktury wiekowej jego mieszkańców, a zatem musiały istnieć inne zależności przestrzenne wpływające na to zjawisko. Wpływ ten uwzględniono poprzez zastosowanie metod geostatystycznych.

W wyniku standaryzacji w odniesieniu do struktury wiekowej zmieniły się jednakże wyraźnie niektóre z podstawowych statystyk. Po wszystkich przeliczeniach średnia różnica w liczbie klientów MOPR na 1 000 mieszkańców w stosunku do danych surowych w każdym z poligonów wyniosła 3,25 osoby, a ich odchylenie standardowe 22,7 osoby. Największa różnica osiągnęła wartość 514 klientów MOPR/1000 mieszkańców. W kilku komórkach siatki (poligonach) otrzymane zagęszczenie klientów MOPR było większe od 1 000 osób na 1 000 mieszkańców (ryc. 2c). Wynika to z faktu, iż rzeczywisty adres zamieszkania podany przez niektórych klientów MOPR znajdował się na obszarach, odnośnie których miasto posiada niepełną informację o rzeczywistym ich zaludnieniu. Są to głównie tereny ogródków działkowych.

2.4. Analizy przestrzenne – kriging Poissona

Podstawą geostatystyki jest teoria funkcji losowych będąca częścią rachunku prawdopodobieństwa (Isaaks, Srivastava 1989). Fundamentalnym założeniem jest przyjęcie, że zmierzona/zaobserwowana w przestrzeni wartość cechy jest jedną z możliwych realizacji przestrzennie skorelowanych procesów losowych (tzw. zmiennych zregionalizowanych).

Metody geostatystyczne są rzadko wykorzystywane do analizy przestrzennej danych społeczno-ekonomicznych. Pierwsze próby ich wykorzystania w badaniach przestrzennego zróżnicowania statusu społecznego i materialnego mieszkańców miasta miały miejsce dopiero pod koniec ubiegłego wieku (Wong 1999). Z biegiem czasu zaczęto stosować je coraz szerzej i to nie tylko w badaniach dotyczących segregacji miejskiej (Reardon, Sullivan 2004; Jargowsky, Kim 2005), ale również w innych dziedzinach będących przedmiotem zainteresowania socjologii miasta (David i in. 2003; Wang 2004; Goodchild, Janelle 2004; Parker, Asencio 2008).

Aby metody geostatystyczne mogłyby zostać szeroko wykorzystane do ich przetwarzania trzeba najpierw było uwzględnić w algorytmach obliczeniowych ich specyfikę. Efektem tych prac jest opracowanie krigingu Poissona (dalej określanego w skrócie jako PK) wykorzystującego, jako model autokorelacji przestrzennej, nie tradycyjny estymator semiwariogramu Matherona, ale semiwariogramu ryzyka. Metodyka ta, wraz z opisem wykonującego obliczenia kodu komputerowego, została opisana przez P. Goovaerts (2005).

2.5. Teoria

Wskaźnik klientów MOPR w poligonie (jednostce przestrzennej) v_α to ich skorygowana, ze względu na strukturę wiekową, liczba $d(v_\alpha)$ w owym poligonie podzielona przez liczbę zamieszkujących go ludzi $n(v_\alpha)$ i pomnożona przez 1 000. Bardzo często, kiedy $n(v_\alpha)$ jest małe, to wskaźnik $z(v_\alpha) = d(v_\alpha)/n(v_\alpha)$ uzyskuje nierealnie małą albo dużą wartość. Ta cecha nazywana jest problemem „małych liczb”.

Klasyczny wzór miary niepodobieństwa – semiwariancji – musi być w tego typu analizie zmodyfikowany tak, aby uwzględniając wagę populacji można było zredukować znaczenie problemu „małych liczb”. Semiwariancja ryzyka $\hat{\gamma}_{Rv}(h)$ dla par jednostek przestrzennych oddalonych od siebie o wektor \mathbf{h} jest następująca [2]:

$$\hat{\gamma}_{Rv}(h) = \frac{1}{2 \sum_{\alpha, \beta}^{N(h)} \frac{n(v_\alpha)n(v_\beta)}{n(v_\alpha) + n(v_\beta)}} \sum_{\alpha, \beta}^{N(h)} \left\{ \frac{n(v_\alpha)n(v_\beta)}{n(v_\alpha) + n(v_\beta)} [z(v_\alpha) - z(v_\beta)]^2 - m^* \right\} \quad [2]$$

gdzie $N(\mathbf{h})$ oznacza liczbę par jednostek przestrzennych (v_α, v_β), których ważone liczbą mieszkańców centroidy oddalone są o wektor \mathbf{h} , a m^* jest średnią częstości zdarzeń dla N obszarów, ważoną poprzez liczbę mieszkańców. Występujące w tradycyjnej postaci semiwariancji podniesione do kwadratu różnice $[z(v_\alpha) - z(v_\beta)]^2$ są tutaj modyfikowane (ważone) funkcją ich liczby mieszkańców pod postacią wyrażenia $n(v_\alpha)n(v_\beta)/[n(v_\alpha) + n(v_\beta)]$, które jest odwrotnie proporcjonalne do ich błędów standardowych. Dzięki temu do par danych mających większą wiarygodność przypisywane jest większe znaczenie. Znaczenie to bowiem zależy od liczby mieszkańców według zasady: im jest większa, tym niższy błąd standardowy. Wprowadzone w powyższym wzorze modyfikacje klasycznego estymatora semiwariancji mają tym większe znaczenie im większe jest zróżnicowanie liczby mieszkańców w zbiorze analizowanych jednostek przestrzennych.

Aby tradycyjny algorytm estymacji geostatystycznej – zwykły kriging (ang. *Ordinary Kriging*, Goovaerts 1997) można było wykorzystać do wskaźników

społecznych, takich jak zróżnicowanie przestrzenne liczby klientów MOPR, trzeba go było również zmodyfikować. Liczbę klientów MOPR $d(v_a)$ traktuje się jako jedną realizację zmiennej losowej $D(v_a)$, której rozkład jest zgodny z modelem Poissona, a parametr tego rozkładu (oczekiwana liczba klientów MOPR) jest iloczynem liczby mieszkańców $n(v_a)$ i lokalnego ryzyka zdarzenia (zostania klientem MOPR) $R(v_a)$. $R(v_a)$ może być traktowane jako liczba klientów MOPR w jednostce v_a pozbawiona losowego szumu (składnika losowego). Wartość tą szacuje się używając wariantu krigingu zwanego krigingiem Poissona (Goovaerts 2005; Monestiez i in. 2005).

Ryzyko zdarzenia (bycie klientem MOPR, $\hat{r}(X)$) i związana z nim wariancja estymacji (wariancja krigingowa $\sigma^2(x)$) w jednostce przestrzennej X jest obliczane według następujących wzorów [3] i [4]:

$$\hat{r}(x) = \sum_{i=1}^K \lambda_i z(v_i) \quad [3]$$

$$\sigma^2(x) = \bar{C}_R(X, X) - \sum_{i=1}^K \lambda_i \bar{C}_R(v_i, X) - \mu(X_0) \quad [4]$$

gdzie X reprezentuje powierzchnię jednostki v_a . Wagi krigingowe λ_i oraz parametr Lagrange'a $\mu(X)$ są obliczane poprzez rozwiązanie przedstawionego poniżej [5] układu równań krigingu Poissona:

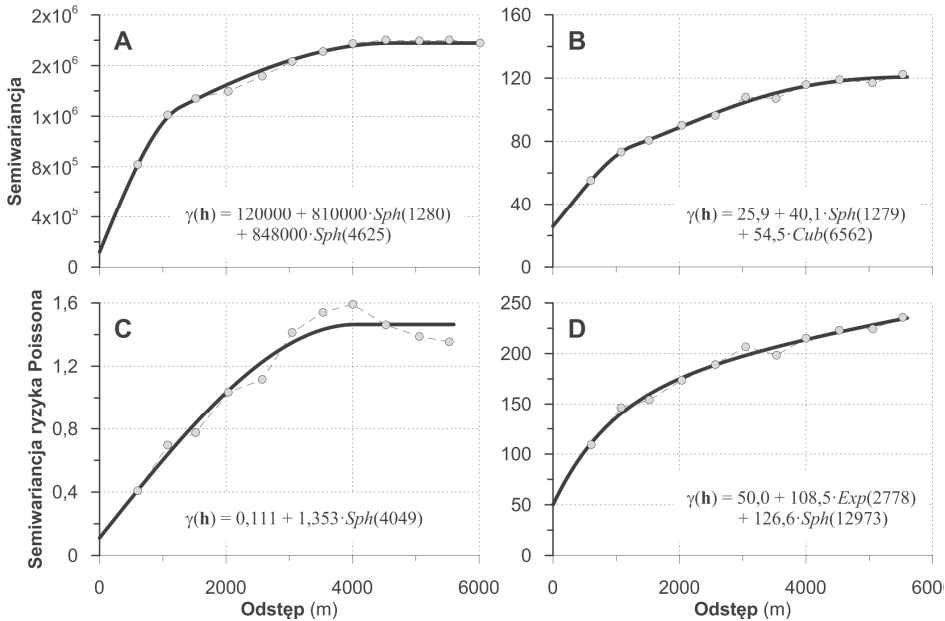
$$\left. \begin{array}{l} \left\{ \sum_{j=1}^K \lambda_j \left[C_R(v_i, v_j) + \sigma_{ij} \frac{m^*}{n(v_i)} \right] + \mu(X) = \bar{C}_R(v_i, X), \quad i = 1, \dots, \right. \\ \left. \sum_{j=1}^K \lambda_j = 1 \right\} \quad [5]$$

gdzie $\delta_{ij} = 1$, jeśli $i = j$, a 0 w pozostałych przypadkach. Wyrażenie wariancji błędu $m^*/n(v_i)$ daje mniejsze wagi dla częstości zarejestrowanych w jednostkach z małą liczbą mieszkańców. Kowariancje $\bar{C}_R(v_i, v_j)$ są szacowane jako średnie ważone populacją punktowych kowariancji $C_R(\mathbf{h})$ obliczanych pomiędzy dwoma lokalizacjami określającymi jednostki v_j i v_i , czyli ich centroidami. Punktowe kowariancje $C_R(\mathbf{h})$ definiowane są zaś jako różnica wariancji próby $C_R(\mathbf{0})$ i semiwariancji ryzyka dla odstępów \mathbf{h} $\gamma_R(\mathbf{h})$.

3. Wyniki

Korekty danych surowych uzyskane w wyniku zastosowania procedury PK są stosunkowo nieduże. Dla 22–30% poligonów były one mniejsze od 0,1 osoby na 1 000 mieszkańców, dla 50–59% poligonów – mniejsze od 1 osoby na 1 000 mieszkańców, a 85–91% poligonów – mniejsze od 10 osób na 1 000 mieszkań-

ców. Wygładzanie przekraczające wartość 100 osób na 1 000 mieszkańców dotyczyło tylko od 5 do 15 poligonów (0,70–2,05% przypadków). Rozmiar korekt jest silnie uwarunkowany liczbą mieszkańców w poligonie (rys. 4). Zależność ta ma charakter potęgowej, a przykładowe współczynniki determinacji wynoszą: dla wszystkich kategorii klientów MOPR = 72,2%, dla pobierających pomoc ze względu na BROB = 59,0%, ze względu na CHOR = 59,3%, ze względu na NIEP = 61,3% i ze względu na UBS = 68,4%.



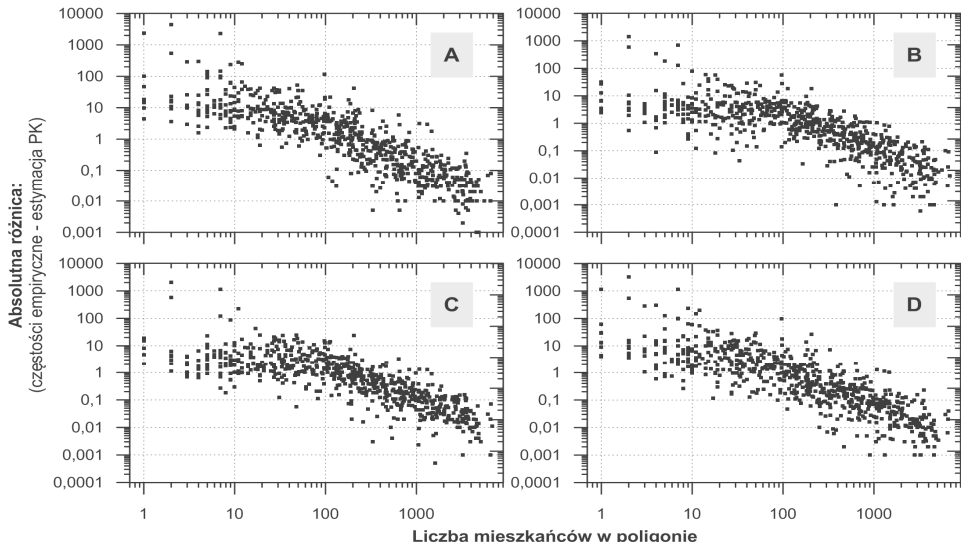
Rys. 4. Semiwariogramy empiryczne i ich modele dla wybranych zmiennych:

A – gęstość zaludnienia, B – BROB, C – WIELDZIET i D – NIEP

Na poszczególnych wykresach umieszczono uproszczone wzory modeli
– szczegóły w tekście i w tab. 1

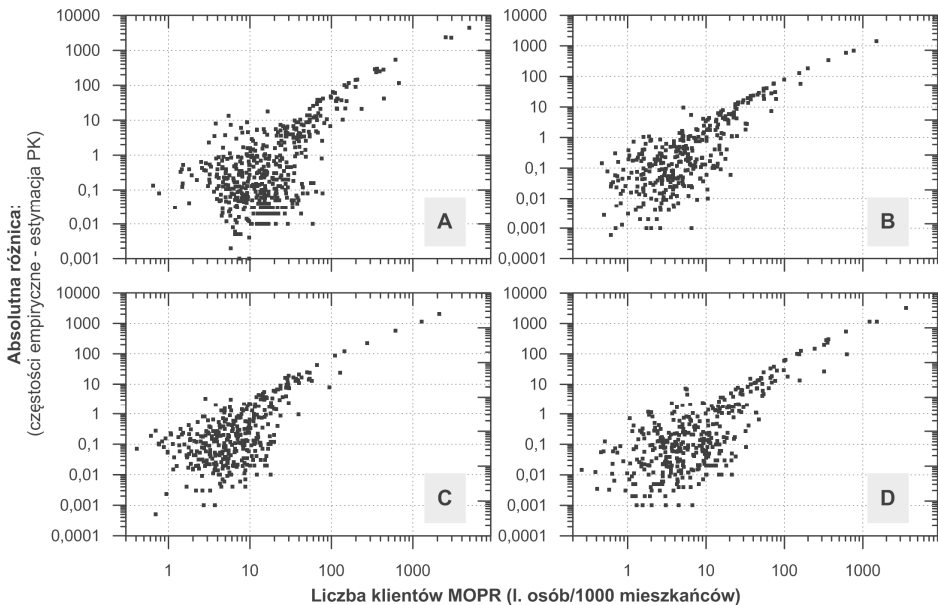
Wielkość wygładzania jest również w pewnym stopniu skorelowana dodatnio z liczbą klientów MOPR w danej komórce siatki (rys. 6). Zależność ta jest wyrażna jednak tylko wtedy, kiedy liczba klientów MOPR jest wyższa od 10 osób na 1 000 mieszkańców. Ma ona również charakter potęgowej, a współczynniki determinacji dla poszczególnych kategorii wahają się od ok. 35 do 60%.

Zarówno gęstość zaludnienia, jak i wszystkie, oprócz „potrzeby ochrony macierzyństwa”, kategorie klientów MOPR, zagregowane w siatce kwadratów 0,5 na 0,5 km wykazują na terenie miasta Poznania silną autokorelację przestrzenną (tab. 1; rys. 4).



Rys. 5. Rozmiar wygładzenia danych empirycznych przez algorytm krigingu Poissona w relacji do gęstości zaludnienia

Objaśnienia: A – wszystkie kategorie klientów MOPR, B – otrzymujący pomoc społeczną ze względu na brak pracy (BROB), C – NIEP, D – UBS



Rys. 6. Rozmiar wygładzenia danych empirycznych przez algorytm krigingu Poissona w relacji do liczby klientów MOPR

Objaśnienia: A – wszystkie kategorie klientów MOPR, B – otrzymujący pomoc społeczną ze względu na brak pracy (BROB), C – NIEP, D – UBS

Tabela 1

Zestawienie modeli semiwariogramów rozkładu przestrzennego gęstości zaludnienia w Poznaniu i poszczególnych kategorii klientów MOPR

Zmienna	C_0	F_1	A_1 (m)	C_1	F_2	A_2 (m)	C_2	$\%C_0$	$\%C_1$	$\%C_2$	Typ
Gęstość zaludnienia (*)	120 000	<i>Sph</i>	1 280	810 000	<i>Sph</i>	4 625	848 000	6,7	45,6	47,7	.
Wszyscy klienci MOPR	346,8	<i>Sph</i>	1 261	484,4	<i>Cub</i>	6 627	650,2	23,4	32,7	43,9	1
ALK	10,1	<i>Sph</i>	1 164	24,4	<i>Sph</i>	6 135	28,1	16,1	39,0	44,9	1
BDOM	102,3	.	.	.	<i>Sph</i>	4 700	154,5	39,8	0,0	60,2	2
BRAD	2,7	<i>Exp</i>	2 331	13,6	<i>Cub</i>	6 296	9,1	10,5	53,5	35,9	1
BROB (*)	25,9	<i>Sph</i>	1 279	40,5	<i>Cub</i>	6 562	54,5	21,4	33,5	45,1	1
CHOR	3,1	<i>Cub</i>	1 110	377,8	<i>Sph</i>	10 345	421,9	0,4	47,1	52,6	3
MACIERZ	0,001	<i>Cub</i>	1 465	0,007	<i>Cub</i>	24 884	0,045	1,9	13,0	85,1	3
NARKOM	0,059	<i>Exp</i>	1 262	1,747	<i>Exp</i>	4 428	0,912	2,2	64,3	33,6	2
NIEP (*)	50,0	<i>Exp</i>	2 778	108,5	<i>Sph</i>	12 973	126,6	17,5	38,1	44,4	3
PRZEM	0,298	<i>Cub</i>	949	0,378	<i>Sph</i>	24 885	1,787	12,1	15,3	72,6	3
SIER	42,6	<i>Cub</i>	1 056	89,8	<i>Sph</i>	24 885	285,2	10,2	21,5	68,3	3
UBS	191,6	<i>Sph</i>	1 508	157,5	<i>Cub</i>	6 464	275,9	30,7	25,2	44,1	1
WIELDZIET (*)	0,111	.	.	.	<i>Sph</i>	4 049	1,353	7,6	0,0	92,4	2
Inne	4,3	<i>Cub</i>	1 052	27,3	<i>Sph</i>	24 885	64,2	4,5	28,5	67,0	3

Objaśnienia: C_0 – efekt samorodka (semiwariancja nuggetowa), F_x – typ funkcji dla składowej x (*Sph* – sferyczna, *Exp* – wykładnicza, *Cub* – sześcienna), A_x – zasięg składowej x w metrach, C_x – semiwariancja progowa (sill) składowej x , $\%C_x$ – udział procentowy składowej x w sumie semiwariancji, Typ – typ rozkładu przestrzennego – objaśnienia w tekście. Modele semiwariogramów dla zmiennych zaznaczonych (*) zostały przedstawione na rys. 4.

Kryterium optymalizacyjnym jakości dopasowania była ważona suma kwadratów odchyłeń z wartością wagi równą „liczba par danych podniesiona do potęgi $\frac{1}{2}$ semiwariancji empirycznej”.

Nie stwierdzono, aby wzrost niepodobieństwa analizowanych zmiennych, wyrażony poprzez semiwariancje empiryczne (Goovaerts 1997), wykazywał istotną zależność od kierunku (anizotropię). Dlatego też uwzględniono jedynie rolę odległości obliczając semiwariancje bezkierunkowe (izotropowe) dla 11 odstępów o szerokości 0,5 km.

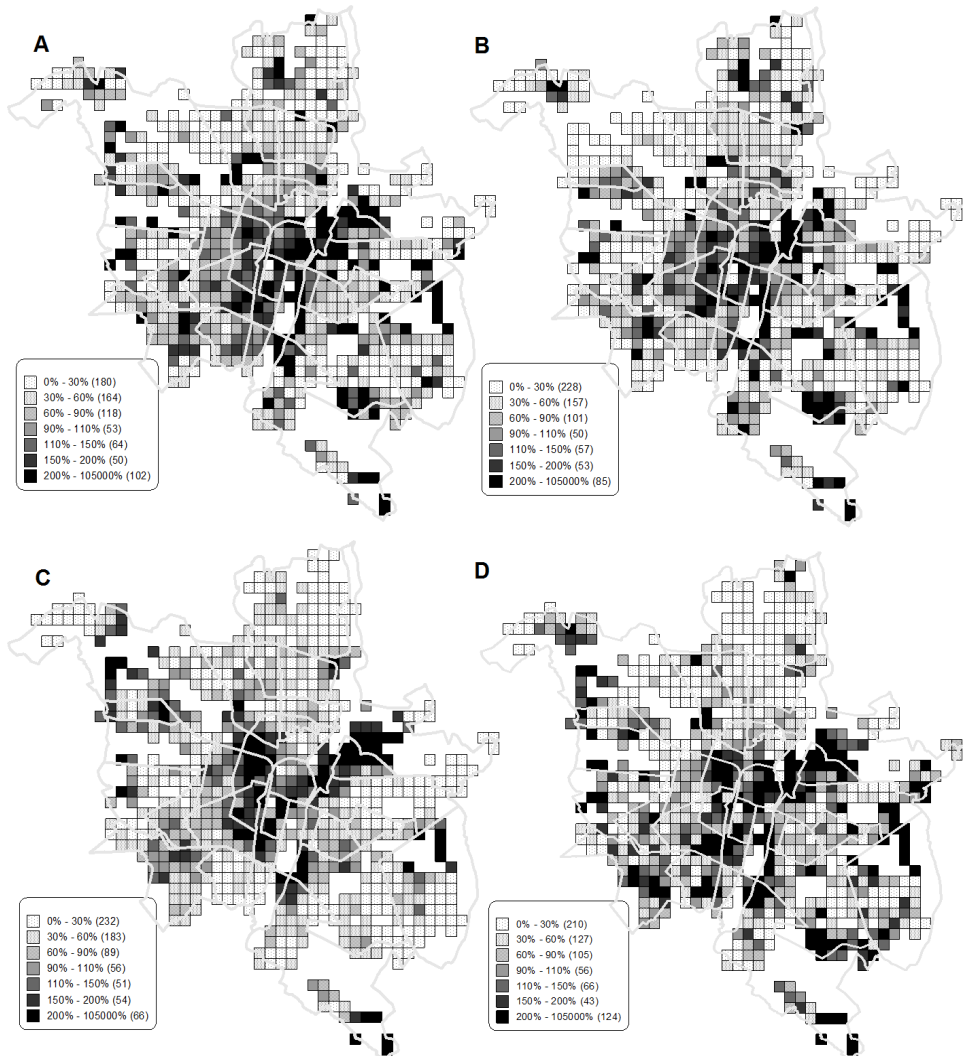
Do semiwariancji empirycznych dopasowano metodą automatyczną (Pardo-Igúzquiza 1999) modele będące kombinacją czterech dopuszczalnych funkcji matematycznych: efektu samorodka (nuggetowej), sferycznej (*Sph*), wykładniczej (*Exp*) i sześcienną (*Cub*).

Semiwariogram gęstości zaludnienia (rys. 4A, tab. 1) wykazuje istnienie na terenie miasta Poznania dwóch struktur o zasięgach odpowiednio ok. 1 250 i 4 600 m. Ich znaczenie jest bardzo zbliżone, bo każda stanowi około 45% całkowitej zmienności rozkładu przestrzennego ludności. Zmienność krótkodystansowa, wewnątrz pojedynczej komórki 0,5 na 0,5 km, wyrażona poprzez udział semiwariancji nuggetowej wynosi około 7%. Taka charakterystyka rozkładu mieszkańców jest efektem struktury przestrzennej miasta związanej z jego historycznym rozwojem. Odzwierciedla bowiem przeciętną wielkość osiedli, które z kolei składają się na dzielnice.

Rozkład przestrzenny gęstości poszczególnych kategorii klientów MOPR wyraźnie nawiązuje do rozkładu zaludnienia (rys. 4B–D; tab. 1). Na podstawie zasięgów poszczególnych struktur składających się na model autokorelacji przestrzennej wyróżniono 3 typy. Do pierwszego z nich należy 5 kategorii klientów MOPR. Mają one dwie struktury o zasięgach 1–2 km i ok. 6,5 km (ekstrapolacja). Na rys. 4B jako przykład tego typu zaprezentowano semiwariogram empiryczny i jego model dla rozkładu przestrzennego osób pobierających pomoc społeczną ze względu na bezrobocie. W drugiej grupie, na którą składają się 3 kategorie klientów MOPR wyraźnie zaznacza się jedynie jedna struktura o zasięgu około 4,5 km. Do tego typu należy m.in. model semiwariogramu „wielodzietności” (rys. 4C). Do ostatniej, trzeciej, grupy zaklasyfikowano sześć kategorii klientów MOPR. Modele ich semiwariogramów, oprócz struktury „krótkiej” (1–1,5 km), wykazują istnienie ważnej struktury długodystansowej o zasięgu mogącym obejmować nawet całe miasto (ekstrapolacja). Na rys. 4D jako przykład tej grupy zaprezentowano semiwariogram empiryczny i jego model dla rozkładu przestrzennego osób pobierających pomoc społeczną ze względu na niepełnosprawność. Rozkład przestrzenny poszczególnych kategorii klientów MOPR różni się też znacznie ze względu na udział poszczególnych struktur (tab. 1). Najmniejszym udziałem zmienności krótkodystansowej (losowej), bo mniejszym od 10%, charakteryzują się kategorie: „długotrwała lub ciężka choroba”, „potrzeba ochrony macierzyństwa”, „narkomania”, „wielodzietność” i „inne”, największym zaś, przekraczającym 30% – „bezdомność” i „ubóstwo”. Duże jest także zróżnicowanie udziału struktury „krótkodystansowej” (osiedlowej) i „długodystansowej” (dzielnicowej/miejskiej).

Większość analizowanych problemów społecznych koncentruje się w centralnej części miasta (rys. 7). Jednak konkretne obszary, w których występuje największe zagęszczenie poszczególnych kategorii klientów MOPR, różnią się między sobą. Rozmieszczenie przestrzenne osób pobierających pomoc ze względu na długotrwałą i ciężką chorobę jest najbardziej zbliżone do rozmieszczenia wszystkich klientów MOPR (rys. 7A–B). Wynika to z faktu, iż jest to jedna z największych grup osób korzystających z pomocy społecznej. W przypadku problemu bezradności widzimy wyraźnie dwa centra tego problemu – jedno na osiedlu Łazarz, a drugie w okolicach osiedla Warszawskiego (rys. 7C). Obszary te dotyka także problem bezrobocia, jednak nie w tak dużym stopniu. Jest on

również silnie związany ze ścisłym centrum. Jednak to, co wyróżnia to zjawisko od pozostałych, to dużo większa częstotliwość jego występowania na obrzeżach miasta (rys. 7D). Wiarygodność prezentowanych wyników potwierdziły wyniki wariacji krigingu, która dla większości analizowanych problemów była stosunkowo niewielka nie tylko w strefie centralnej, ale również w obszarze całego zewnętrznego pierścienia miasta.



Rys. 7. Rozkład przestrzenny zagęszczenia klientów MOPR na terenie miasta Poznania skorygowany metodą krigingu Poissona – wyniki przedstawione w skali relatywnej (ryzyko wystąpienia danego problemu w poligonie w stosunku do średniej gęstości dla całego miasta [%]); w nawiasach podano liczbę poligonów; wyniki dla A – wszystkich klientów MOPR, B – CHOR, C – BRAD oraz D – BROB

4. Wnioski

W wyniku przeprowadzonych analiz rozmieszczenia klientów MOPR możliwe było zlokalizowanie na terenie Poznania tych obszarów, które były w 2008 r. (i prawdopodobnie w dalszym ciągu są) potencjalnie najbardziej narażone na występowanie różnych problemów społecznych. Zgodnie z oczekiwaniami (Wilson 1987, 2012) – ściśle centrum Poznania zdaje się być najbardziej narażone na występowanie takich zjawisk, jak bezrobocie czy bezradność; jak również dzielnice położone wokół centrum, w tym: Jeżyce, Górna Wilda, os. Warszawskie i obszar „Przy Warcie”. Większość klientów MOPR zamieszkuje najstarsze dzielnice Poznania, co nie musi jednak oznaczać, iż dzielnice te sprzyjają rozwojowi negatywnych zjawisk społecznych. Wiąże się to raczej z ucieczką mieszkańców, którzy mogą sobie na to pozwolić, do innych części miasta o wyższym standardzie zabudowy mieszkaniowej. Jednak w przypadku niektórych problemów społecznych (w tym bezrobocia) można również zauważyć pojedyncze osiedla zlokalizowane na obrzeżach miasta, w których proporcja klientów MOPR w całej populacji jest relatywnie bardzo wysoka.

Rozkład przestrzenny częstości występowania poszczególnych kategorii klientów MOPR nawiązuje do rozkładu gęstości zaludnienia tylko w ogólnym zarysie – w wielu przypadkach widoczne są znaczące odchylenia. Większość analizowanych rozkładów klientów MOPR wykazywała autokorelację w dwóch skalach przestrzennych: pojedynczego osiedla (ok. 1–1,2 km) oraz dzielnicy (ok. 6–6,5 km) lub nawet całego miasta (ok. 25 km).

Statystyki przestrzenne wykorzystane w prezentowanej pracy pozwoliły na identyfikację konkretnych dzielnic, w których istnieje wysokie ryzyko wystąpienia poszczególnych problemów społecznych. Niektóre z nich są narażone nawet kilka razy bardziej niż średnia dla całego miasta. Dzięki krigingowi Poissona możliwe było określenie wiarygodności otrzymanych wyników na podstawie wariancji krigingowej.

Interpretując powyższe wyniki należy pamiętać, iż rozkład przestrzenny klientów MOPR jest tylko pewnym przybliżeniem rzeczywistej struktury społeczno-przestrzennej miasta. Jest to dlatego obraz niepełny, gdyż MOPR nie posiada wiedzy o wszystkich rodzinach faktycznie borykających się z różnymi problemami społecznymi. Dlatego ważnym byłoby poszerzenie powyższych badań o analizę innych źródeł wiedzy dotyczących omawianych zjawisk, takich jak wywiad środowiskowy w poszczególnych dzielnicach, informacje pochodzące z komisariatów czy też kuratorów sądowych. Takie badania mogłyby też być dodatkową weryfikacją dokładności szacunków uzyskanych przy pomocy krigingu Poissona.

LITERATURA

- Ali M. i in., 2006, *Application of Poisson kriging to the mapping of cholera and dysentery incidence in an endemic area of Bangladesh*, „International Journal of Health Geographics”, 5, s. 45, <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1617092&tool=pmcentrez&rendertype=abstract>.
- Anderson R.N., Rosenberg H.M., 1998, *Age Standardization of Death Rates: Implementation of the Year 2000 Standard*, „National Vital Statistics Reports”, 47(3).
- Bevan A., Conolly J., 2009, *Modelling Spatial Heterogeneity and Nonstationarity in Artifact-Rich Landscapes*, „Journal of Archaeological Science”, 36(4), s. 956–964.
- Buescher P.A., 1997, *Problems with rates based on small numbers*, [w:] *Statistical Primer*, North Carolina State Center for Health Statistics, s. 1–7, <http://www.schs.state.nc.us/SCHS/pdf/primer12.pdf>.
- Bumpus S., 2012, *Analysing and visualising areal crime data. A case study of residential burglary in San Francisco, USA*, Master of Science Thesis in Geospatial Technologies, Universidade Nova de Lisboa, <http://hdl.handle.net/10362/8316>.
- David K., Higgs G., White S., 2003, *Socio-Economic Applications in Geographical Information Science*, Taylor & Francis, London.
- De Oliveira V., 2013, *Hierarchical Poisson models for spatial count data*, „J. Multivariate Anal.”, 122, s. 393–408.
- De Oliveira V., 2014, *Poisson kriging: A closer investigation*, „Spatial Statistics”, 7, s. 1–20.
- Ebenezer B., Ebenezer O.S., Linda O., 2013, *Application of Area to Point Kriging to Buruli Ulcer Incidence in Ashanti and Brong Ahafo Regions of Ghana*, Geoinfor. Geostat: An Overview, 1:1.
- Goodchild M.F., Janelle D.G., 2004, *Spatially integrated social science*, Oxford University Press, Inc., New York.
- Goovaerts P., 1997, *Geostatistics for Natural Resources Evaluation.*, Oxford University Press, Inc., New York.
- Goovaerts P., 2005, *Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging*, „International Journal of Health Geographics”, 4(1), s. 31, <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1360096&tool=pmcentrez&rendertype=abstract>.
- Goovaerts P., 2006, *Geostatistical Analysis of Disease Data: Accounting for Spatial Support and Population Density in the Isopleth Mapping of Cancer Mortality Risk Using Area to Point Poisson Kriging*, „International Journal of Health Geographics”, 5(52).
- Goovaerts P., 2008, *Accounting for rate instability and spatial patterns in the boundary analysis of cancer mortality maps*, „Environmental and Ecological Statistics”, 15, s. 421–446.
- Goovaerts P., Gebreab S., 2008, *How does Poisson kriging compare to the popular BYM model for mapping disease risks?*, „International Journal of Health Geographics”, 7(1), s. 6, <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2276482&tool=pmcentrez&rendertype=abstract>.
- Isaaks E.H., Srivastava R.M., 1989, *Applied Geostatistics*, Oxford University Press, Inc., New York.

- Jargowsky P.A., Kim J., 2005, *A Measure of Spatial Segregation: The Generalized Neighborhood Sorting Index A Measure of Spatial Segregation*, National Poverty Center Working Paper Series, http://www.nationalpovertycenter.net/publications/working_papers/.
- Jenks G.F., 1967, *The Data Model Concept in Statistical Mapping*, „International Yearbook of Cartography”, 7, s. 186–190.
- Kafadar K., 1994, *Choosing among two-dimensional smoothers in practice*, „Computational Statistics and Data Analysis”, 18, s. 419–439.
- Kerry R. i in., 2010, *Applying Geostatistical Analysis to Crime Data: Car-Related Thefts in the Baltic States*, „Geographical Analysis”, 42(1), s. 53–77, <http://doi.wiley.com/10.1111/j.1538-4632.2010.00782.x>
- Kerry R. i in., 2012, *A comparison of multiple indicator kriging and area-to-point Poisson kriging for mapping patterns of herbivore species abundance in Kruger National Park, South Africa*, „International Journal of Geographical Information Science”.
- Kober K. i in., 2010, *An analysis of the numbers and distribution of seabirds within the British Fishery Limit aimed at identifying areas that qualify as possible marine SPAs*.
- Krivoruchko K., Gribov A., Krause E., 2011, *Multivariate areal interpolation for continuous and count data*, „Procedia Environ. Sci.”, 3, s. 14–19.
- Monestiez P., Dubroca L., Bonnin E., Durbec J.P., Guinet C., 2005, *Comparison of Model Based Geostatistical Methods in Ecology: Application to Fin Whale Spatial Distribution in Northwestern Mediterranean Sea*, [w:] Leuanghton O., Deutsch C.V. (red.), „Geostatistics Banff 2004”, 2, s. 777–786.
- Naeimeh S.A. i in., 2013, *Area-to-Area Poisson Kriging Analysis of Mapping of County-Level Esophageal Cancer Incidence Rates in Iran*, „Asian Pacific Journal of Cancer Prevention”, 14, s. 11–13.
- Oliver M.A. i in., 1998, *Binomial cokriging for estimating and mapping the risk of childhood cancer*, „IMA Journal of Mathematics Applied in Medicine and Biology”, 15, s. 79–297.
- Pardo-Igúzquiza E., 1999, *VARFIT: a Fortran-77 program for fitting variogram models by weighted least squares*, „Computers and Geosciences”, 25, s. 251–261.
- Parker R.N., Asencio E.K., 2008, *GIS and Spatial Analysis for the Social Sciences Coding, Mapping and Modeling*, Taylor & Francis, New York.
- Parysek J.J., Mierzejewska L., 2006, *City Profile Poznań*, „Cities”, 23(4), s. 291–305, <http://linkinghub.elsevier.com/retrieve/pii/S0264275106000291>.
- Pfeiffer D.U. i in., 2008, *Spatial Analysis in Epidemiology*, Oxford University Press, Oxford, UK.
- Reardon S.F., Sullivan D.O., 2004, *Measure of spatial segregation*, „Sociological Methodology”, 34(1), s. 121–162.
- Shao C.Y., Mueller U., Cross J., 2009, *Area-to-point Poisson kriging analysis for lung cancer incidence in Perth areas*, [w:] 18th World IMACS/MODSIM Congress 13–17 July 2009, Cairns, Australia, s. 1–7.
- Talbot T.O. i in., 2000, *Evaluation of spatial filters to create smoothed maps of health data*, „Statistics in Medicine”, 19, s. 2399–2408.
- Waller L.A., Gotway C.A., 2004, *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons, Inc., Hoboken, NJ, USA.

- Wang F., 2004, *Geographic Information Systems and Crime Analysis*, [w:] Wang F. (red.), *IGI Global*, <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-59140-453-8>.
- Wilson W.J., 1987, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, Bridge G., Watson S. (red.), University of Chicago Press, <http://www.amazon.com/Truly-Disadvantaged-Underclass-Public-Policy/dp/0226901319>.
- Wilson W.J., 2012, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, Second Edition II, University of Chicago Press, USA.
- Wong D.W.S., 1999, *Geostatistics as measures of spatial segregation*, „Urban Geography”, 20(7), s. 635–647, <http://bellwether.metapress.com/openurl.asp?genre=article&id=doi:10.2747/0272-3638.20.7.635>.

POISSON KRIGING AS A TOOL FOR SOCIAL PROBLEMS ANALYSIS – POZNAŃ CASE STUDY

Planning of social policy is a complicated and multidimensional issue, especially in complex urban structures characteristic for big cities. However econometric indicators of spatial dependence provide us some information about spatial autocorrelation, they do not show the differences in local variability. Geostatistics is an answer for this challenge. This method is not only helpful in more accurate determination of the most important problems but it also enables identification of their location, scale and possible reasons.

This paper presents possibilities given by Poisson Kriging for analysis of social problems in urban space. It was applied in Poznań for identification of neighbourhoods or local communities (related to the basic administrative units called „osiedla”) in which concentration of people needing social help is bigger than population distribution might suggest.

The data used in the analysis was taken from urban centre helping families in difficult social situation (MOPR). They concern people who received financial support in 2008. The basic information taken into account was their place of residence and the reason for which they received financial aid. MOPR distinguishes 13 categories of social problems needing support, including poverty, chronic diseases, alcoholism and domestic violence. 9 473 persons received financial aid in the analysed period of time. Taking into account their families it gives us at least 18 264 people struggling with social problems – 3,3% of the city population (545 000 inhabitants).

In order to receive a comparable measure of issues analysed in urban space, the number of people needing social support must be compared with the population distribution. Thus, information about place of residence of people who receive financial aid was aggregated to bigger areas – 731 regular polygons for which the number of city inhabitants was known. Side length of single polygon was 500 meters. In each polygon data needed also to be age-adjusted. It is a very sophisticated task, therefore a special script dedicated for ArcGIS was created. The age-adjusted data aggregated in the polygons were bases for main spatial analysis.

Application of Poisson Kriging resulted in more precise identification of areas affected by the major social problems in Poznań. Presence of autocorrelation was noticeable in case of majority of analysed social problems. The most common ranges of autocorrelation were 1–1,2 km (which is similar to the spatial range of single local

communities) and 6–6,5 km (the range of single neighbourhoods). Analysis showed that there are some neighbourhoods in Poznań where occurrence of social problems is significantly higher than mean occurrence for the whole city. Presented method enabled smoothing of unreliable, extremely high relative risks values but without loss of the local variability.

Key words: *social problems, Poznań, socio-spatial structure, Poisson kriging method*

Prof. UAM, dr hab. Alfred Stach

Mgr inż. Patrycja Wysocka

Zakład Geoinformacji, Instytut Geoekologii i Geoinformacji

Wydział Nauk Geograficznych i Geologicznych, Uniwersytet im. Adama Mickiewicza
w Poznaniu