



POLSKIE TOWARZYSTWO GEOGRAFICZNE · POLISH GEOGRAPHICAL SOCIETY

czasopismo  
geograficzne

geographical  
journal

# METODY STATYSTYCZNE W GEOGRAFII SPOŁECZNO- EKONOMICZNEJ: SZKIC HISTORYCZNY ORAZ OGRANICZENIA I KORZYŚCI STOSOWANIA W DOBIE CYFRYZACJI

## STATISTICAL METHODS IN SOCIO-ECONOMIC GEOGRAPHY: A HISTORICAL OUTLINE AS WELL AS LIMITATIONS AND ADVANTAGES OF THEIR USE IN THE DIGITAL AGE

Jan Hauke

Uniwersytet im. Adama Mickiewicza w Poznaniu, Wydział Geografii Społeczno-Ekonomicznej i Gospodarki  
Przestrzennej, Zakład Analiz Regionalnych i Lokalnych, Poznań, Polska

 <https://orcid.org/0000-0001-7933-123X>, e-mail: [jhauke@amu.edu.pl](mailto:jhauke@amu.edu.pl)

### Streszczenie

Metody statystyczne od wielu już lat stanowią pomocnicze narzędzie we wnioskowaniu w prawie wszystkich dyscyplinach naukowych. W pracy przedstawiono w krótkiej, zwartej formie wybrane korzyści, zagrożenia oraz nowe trendy odnoszące się do stosowania metod statystycznych w badaniach związanych z geografią społeczno-ekonomiczną w dobie powszechnej cyfryzacji. Konstrukcja pracy oparta jest na: a) szkicu historycznym dotyczącym początku używania tych metod, rozwoju i momentów przełomowych ze szczególnym uwzględnieniem wpływu polskich badaczy, b) szkicu historycznych zmian w edukacji, w efekcie prowadzących do zmian w używaniu metod statystycznych przez badaczy oraz c) przedstawieniu wybranych współczesnych problemów w stosowaniu w badaniach, korzyści oraz nowych trendów. Zarówno problemy, jak i korzyści wynikające ze stosowania tych metod nie są nowe, ale występują w nowej odsłonie, w sytuacji gdy mamy do dyspozycji wysoko wydajne komputery oraz zaawansowane oprogramowanie,

Hauke J. (2021). Metody statystyczne w geografii społeczno-ekonomicznej: szkic historyczny oraz ograniczenia i korzyści stosowania w dobie cyfryzacji. *Czasopismo Geograficzne*, 92(1): 73–93. <https://doi.org/10.12657/czageo-92-04>



Otrzymano/Received: 21.08.2021  
Zaakceptowano/Accepted: 04.10.2021

w tym (co jest szczególnie istotne dla geografów) Systemy Informacji Geograficznej (nazywane w skrócie zgodnie z angielską nazwą – GIS).

**Słowa kluczowe:** geografia społeczno-ekonomiczna, metody statystyczne, GIS, losowość, populacja, próba.

---

### **Abstract**

For many years, statistical methods have been an auxiliary tool in drawing conclusions in almost all scientific disciplines. The paper presents in a short, compact form selected benefits, threats and new trends in the use of statistical methods in research related to socio-economic geography in the era of widespread digitization. The structure of the work is based on: a) a historical sketch concerning the beginning of the use of these methods, development and breakthrough moments, with particular emphasis on the influence of Polish researchers, b) a sketch of historical changes in education, resulting in changes in the use of statistical methods by researchers, and c) the presentation of selected contemporary problems with their application in research, benefits and new trends. Both the problems and benefits resulting from the use of these methods are not new, but they appear in a new version in a situation where we have high-performance computers and advanced software, including (which is particularly important for geographers) Geographic Information Systems.

**Keywords:** human geography, statistical methods, GIS, randomness, population, sample.

---

## **WSTĘP**

Analiza treści i metod używanych w artykułach naukowych pozwala stwierdzić, że metody statystyczne są podstawą wnioskowania w prawie wszystkich dyscyplinach naukowych. Dane statystyczne wykorzystywane do obliczeń mogą mieć bardzo różny charakter. Mogą one dotyczyć cech ciągłych czy też skokowych, ale również takich, które inicjalnie nie były określane liczbowo i dopiero w trakcie obliczeń zostały liczbowo skategoryzowane. W naukach przyrodniczych, w których podstawę wnioskowania stanowią wartości uzyskane najczęściej poprzez pomiar określonym urządzeniem, obliczenia na tak uzyskanych danych wydają się naturalne. W naukach pozaprzyrodniczych, w szczególności społecznych, w których matematyczne wyrażanie cech jakościowych, takich jak np. skalowanie, jest dość subiektywne, stosowane we wnioskowaniu metody statystyczne nie są już tak naturalne. W geografii społeczno-ekonomicznej dodatkowym dość często spotykanym problemem jest brak losowości badanej cechy (nawet gdy jest ona wyrażona liczbowo) jako podstawowego założenia metod statystycznych. Tak się zdarza np. w badaniach regionalnych, w których dostępne dane reprezentują właściwie populację, a nie próbę. Uwagi odnoszące się zarówno do korzyści, jak i ograniczeń w używaniu metod statystycznych, czy też właściwego wyko-

rzystywania uzyskanych w wyniku obliczeń rezultatów, pojawiały się w publikacjach już wcześniej (patrz np. Barnes, 1994), w tym w literaturze polskojęzycznej, (patrz np. Ratajczak, Parysek, 2002). Jednakże wobec powszechnego dostępu do komputerów wyposażonych w bardzo zaawansowane pakiety statystyczne oraz bardzo dynamicznie rozwijających się GIS-ów zagrożenie niepoprawnego użycia tych metod drastycznie wzrosło. Właśnie ten aspekt jest przyczyną zainteresowania się podnoszonymi powyżej zagrożeniami w niniejszym artykule. Z drugiej strony, możliwość kompleksowych analiz na ogromnych zbiorach danych (big data) z użyciem zaawansowanych zaprogramowanych procedur stanowi niepodważalną zaletę. Obliczenia mogą być powtarzane na wielokrotnie korygowanych danych z użyciem narzędzi graficznych pozwalających w odpowiedni sposób przedstawić zarówno same dane, jak i uzyskane w obliczeniach wyniki, czyniąc to narzędzie dla wnikliwego badacza niezwykle użytecznym. Biorąc pod uwagę nakreślone powyżej dylematy, należy stwierdzić, że odpowiedź na pytanie, na ile używane metody statystyczne to bardzo cenne narzędzie pomocnicze, a na ile narzędzie sprzyjające manipulacji (niekoniecznie intencjonalnej), nie może być jednoznaczna. Ta niejednoznaczność w dobie powszechnej cyfryzacji stwarzającej możliwości przeprowadzenia kompleksowej analizy matematyczno-statystycznej na ogromnych zbiorach danych w sposób indywidualny na własnym komputerze odnosi się obecnie do szerszego kontekstu i prowokuje do podejmowania dalszych dyskusji. Należałoby też dodać, że w pracy nie rozważa się właściwości samych metod matematycznych, w tym sposobów kwantyfikacji wybranych cech i matematycznych operacji wykonywanych na danych według zaakceptowanej przez wszystkich procedury, np. w powszechnie używanych modelach grawitacji i potencjału. Zagadnienia te są bardzo ważne, ale bez wątpienia mogłyby wypełnić treść następnego oddzielnego artykułu. Przymiotnik „statystyczny” oznacza tu odwoływanie się w metodzie do losowości, a zatem rachunku prawdopodobieństwa.

W pracy przedstawiono krótki szkic historyczny na temat używania metod matematyczno-statystycznych (z naciskiem na jego początki i polski wkład), komentarze związane z edukacją studentów kierunków geograficznych w zakresie wykorzystania metod statystycznych oraz rozważania (oparte na literaturze) dotyczące nowych trendów, korzyści i zagrożeń wynikających ze stosowania aparatu statystycznego we wnioskowaniu w badaniach z dziedziny geografii społeczno-ekonomicznej.

## **METODY STATYSTYCZNE W NAUKACH POZAPRZYRODNICZYCH. SZKIC HISTORYCZNY**

Rozważań na temat używania metod statystycznych we współczesnych badaniach dotyczących szeroko rozumianej geografii społeczno-ekonomicznej nie

sposób oderwać nie tylko od początków ich stosowania w naukach pozaprzyrodniczych, ale także od początków powstawania statystyki jako nauki. Statystyka (rozumiana jako część matematyki) to nauka, która ma niezbyt długą historię. Jej intensywny rozwój zaczął się w drugiej połowie XIX w., choć słowo „statystyka” w formie rzeczownika zostało użyte znacznie wcześniej. Wprowadził je w 1749 r. urodzony w Elblągu G. Achenwall. Jak wyjaśnia Ostasiewicz (2014), przyjmował on, że statystyka to wiedza o państwie i jego konstytucji. Jest to bliższe temu, co dziś związane jest z kolekcją danych statystycznych. Przedstawione tu rozważania obejmują jednak metody statystyczne jako szczególne metody matematyczne i ich zastosowania.

Lista badaczy, którzy byli pionierami lub istotnie wpłynęli na rozwój tej aplikacyjności, jest długa, ale ze względu na cel pracy warto się odnieść do wybranych trzech bardzo istotnych postaci: A. Quételeta, F. Galtona oraz B. Janowskiego. Dodajmy też, że trzej ich wielcy poprzednicy: P. de Laplace, C. Gauss i S. Poisson jako znakomici XVIII-wieczni matematycy zajmujący się też rachunkiem prawdopodobieństwa i statystyką matematyczną (wg dzisiejszego rozumienia tej nauki), mimo swej szerokiej działalności naukowej w innych dziedzinach, nie wychodzili z zastosowaniami metod statystycznych poza nauki przyrodnicze. Inny rówieśnik tych uczonych Jean Antoine Nicolas Caritat markiz de Condorcet (1743–1794) – francuski filozof racjonalista, matematyk, ekonomista i polityk, podjął próbę wejścia z matematyką w nauki pozaprzyrodnicze. Crépel i Rieucou (2005) w artykule pod znamiennym tytułem *Condorcet's Social Mathematic, a few tables* (Matematyka społeczna Condorceta, kilka tabel) piszą o nim, że stworzył załączki planu zastosowania rachunku różniczkowego w naukach politycznych i moralnych już na początku lat 70. XVIII w., umieszczając „naukę o stosunkach moralnych lub politycznych” w ramach gałęzi „matematyki fizycznej”. Był w tym względzie pionierem, ale jego idee, generalnie rzecz ujmując, nie były później kontynuowane (po części odnosił się do nich A. Quételet). Metody matematyczne do nauk społecznych (politologii) próbował wprowadzać także Józef Hoene-Wroński (1776–1883), polski polihistor: matematyk, fizyk, filozof, ekonomista i prawnik. Zaproponował on matematyczny wzór na właściwe sprawowanie rządów, za pomocą którego wyznacza się stopień anarchii i despotyzmu (Pragacz, 2007). Nie ma chyba potrzeby szerzej wyjaśniać, że takie zastosowania nie znalazły jednak uznania.

Adolphe Quételet (1796–1874) to belgijski uczony, który zajmował się astronomią, matematyką, statystyką, meteorologią, socjologią, kryminologią. Uważał on, że matematyka to centrum, do którego powinny zbiegać (poprzez jej stosowanie) inne nauki w celu osiągnięcia swoich doskonałości, a teoria prawdopodobieństwa powinna stanowić podstawę badań wszystkich nauk, zwłaszcza zaś nauk, w których korzysta się z obserwacji. Wykorzystywał badania statystyczne do zagadnień z zakresu antropologii i socjologii. Zajmował się też statystyką przestępczości,

a prawidłowości w życiu społecznym zidentyfikowane za pomocą statystyki uważał za nieodmienne prawa przyrody, których człowiek nie może zmienić.

Stworzył fizykę społeczną, aby policzyć wszystkie fizyczne, moralne i intelektualne cechy ludzkości (Donnelly, 2015). Jest twórcą koncepcji *hommemoien* („przeciętny człowiek”) uznającej, że wartości każdej z ludzkich cech są skoncentrowane wokół wartości centralnej tej cechy zgodnie z rozkładem normalnym (co dziś określilibyśmy, że wartości tej cechy podlegają rozkładowi normalnemu). Wraz z pracą Quetéleta z 1835 r., jak zauważa Freudenthal (1975), rozpoczęła się nowa era w statystyce. Przedstawił nową technikę obliczeń statystycznych, a raczej pierwszą taką technikę w ogóle. Materiał, na którym pracował (dane), został przez niego starannie uporządkowany zgodnie z pewnymi wcześniej ustalonymi zasadami. W pracy Quetéleta nie ma zbyt wielu danych statystycznych, ale każda podana liczba ma sens. Dla każdej liczby Quetélet próbował znaleźć to, co miało decydujący wpływ na jej wartość, jej naturalne przyczyny i perturbacje wywołane przez człowieka. Praca ta zawierała opis przeciętnego człowieka jako zjawiska zarówno statycznego, jak i dynamicznego.

Francis Galton (1822–1911) to polimat (*universalman*), zajmował się statystyką, socjologią, psychologią, antropologią, badaniem tropików, geografą, meteorologią, protogenetyką i psychometrią. Był także wynalazcą oraz (co dziś nie przynosi mu sławy) zwolennikiem darwinizmu społecznego, eugeniki i rasizmu naukowego. Maier (2011) określił Galtona następująco: „He strove always to take the ineffable, the abstract, the emotional, the human and to scale and calibrate it” („Zawsze dążył do ustalenia skali i skalibrowania tego co niewypowiedziane, abstrakcyjne, emocjonalne i ludzkie”, tłum. własne). W roku 1883 opublikował książkę (Galton, 1883), w której zajął się analizą statystyczną (dziś nazwalibyśmy to statystyką opisową) różnych cech człowieka. W latach 1884–1885 Galton założył laboratorium do pomiaru statystyki ludzkiej (mierzącej cechy ludzi). Zebrał dane (ilościowo bardzo duże, jak na ówczesne czasy), takie jak wzrost, waga i siła dużej liczby osób, konstruując samodzielnie aparat do pomiarów. W doświadczeniach biologicznych (z nasionami groszku) zaobserwował, że mediana średnicy potomstwa dużych nasion była mniejsza niż ich rodziców, podczas gdy mediana średnicy potomstwa małych nasion była większa niż ich rodziców, co doprowadziło go do stwierdzenia, że potomstwo ma tendencję do powrotu do średniej wielkości. Początkowo nazwał to zjawisko „rewersją”, ale później zmienił nazwę na „**regresja**”. Pojęcie to stało się potem synonimem zależności w statystyce. W kolejnych badaniach dotyczących zależności wzrostu rodziców i ich dorosłych dzieci utworzył wykresy tych zależności w taki sposób, że współczynnik regresji stał się nachyleniem linii regresji. Z kolei w badaniach porównawczych wielkości dwóch różnych narządów tej samej osoby opracowane przez siebie metody do badania stopnia powiązania rozmiarów zdefiniował i zastosował współczynnik korelacji jako miarę stopnia, w jakim były one powiązane. Współczynnik ten bywa na-

zywany współczynnikiem korelacji Pearsona (C. Pearson był uczniem Galtona). Analiza regresji i korelacji była rozwijana później zarówno pod względem teoretycznym, jak i praktycznym przez wielu jego następców. Ale to Galton wprowadził badanie statystycznej zależności na właściwe tory i nie sposób przecenić użyteczność aplikacyjną jego propozycji w różnych dziedzinach nauki, w tym również w badaniach w geografii społeczno-ekonomicznej.

Benon Janowski (1873–1945? – rok śmierci nie jest pewny), który uzyskał doktorat na Wydziale Filozofii Uniwersytetu Lwowskiego w roku 1903, nie będąc zatrudniony na uczelni (pracował jako nauczyciel i dyrektor szkół), prowadził badania w zakresie geografii, stosując metody matematyczno-statystyczne w swoich publikacjach (Ratajczak, 2016). Perełką w zakresie tych zastosowań jest praca traktująca o znaczeniu odległości w antropogeografii (Janowski, 1908). Arthur Getis, jeden z najważniejszych współczesnych geografów na świecie zajmujących się statystyką przestrzenną, o pracy tej, po wydrukowaniu w 2013 r. jej przetłumaczonej na angielski wersji, wyraził się z niezwykłym uznaniem w sposób następujący: „I want to thank you again for the translation of Benon Janowski's book. It is fascinating and **wellahead of his time**. If the translation were available in 1908, perhaps geography and regional science would be quite different today” („Jeszcze raz dziękuję za przetłumaczenie książki Benona Janowskiego. Jest fascynująca i **wyprzedza jego czasy**. Gdyby tłumaczenie było dostępne w 1908 r., być może geografia i badania regionalne byłyby dziś zupełnie inne”, tłum. własne) (Ratajczak, 2016). O innych pracach publikowanych przez Janowskiego stosujących metody zarówno matematyczne, jak i statystyczno-matematyczne pisze Jędrzejczyk (1992, 2000).

W końcówce lat 20. XX w. nastąpił przełom w rozwoju statystyki matematycznej, w którym istotny udział miał polski matematyk i statystyk J. Sława-Neyman, publikując prace (razem z E. Pearsonem) zawierające fundamentalne rezultaty dotyczące testowania hipotez statystycznych. To właśnie ten element statystyki matematycznej jest drugim (obok regresji i korelacji) najbardziej użytecznym w stosowanych metodach statystycznych w różnych dziedzinach nauki, w tym również w geografii.

Z kolei przełom w zastosowaniach statystyki w geografii związany był z tzw. rewolucją ilościową (*quantitative revolution*), która miała miejsce w latach 50. i 60. XX w. i oznaczała gwałtowną zmianę w metodach badań geograficznych, w tym przejście z geografii regionalnej na naukę przestrzenną. Była ona odpowiedzią na postępujący w latach 40. i jeszcze na początku lat 50. XX w. spadek pozycji geografii, kiedy to nie była traktowana jako poważna nauka jako mająca tylko charakter opisowy. W tym czasie następowało też rozdzielanie geografii na dwie nauki: geografie fizyczną i społeczną (ang. *human*), dziś określaną w języku polskim jako społeczno-ekonomiczna. Pojęcie „geografia ilościowa” i używanie metod matematycznych miało już miejsce znacznie wcześniej, np. w pracy Wallisa (1912) czy też

Huntingtona (1927), ale w bardzo ograniczonej skali. Sama rewolucja ilościowa miała na celu opracowanie bardziej rygorystycznej i systematycznej metodologii dla geografii jako dyscypliny nauki. Główną zasługą rewolucji ilościowej jest to, że doprowadziła ona do przejścia od geografii opisowej (idiograficznej) do empirycznej geografii prawotwórczej (nomotetycznej). W konsekwencji w końcówce lat 50. i dalej w latach 60. XX w. intensywnie wprowadzano do geografii zarówno tej fizycznej, jak i społeczno-ekonomicznej, takie bardziej zaawansowane metody (w tym również metody wielowymiarowe), jak analiza wariancji, analiza składowych głównych, analiza czynnikowa, metody klasyfikacji wielowymiarowej, rozbudowane metody nieparametryczne, regresja nieliniowa czy analizy oparte na entropii. O kolejnych etapach rozwoju tak rozumianej geografii ilościowej pisali m.in.: Garrison (1956), Mackay (1959), Burton (1963), Curry (1967). Marchand (1974) zastanawiał się, na ile to była rewolucja, a na ile kontrewolucja (czyli siła niszcząca rozwój geografii) ze względu na trwający w tamtym czasie już od kilku lat kryzys spowodowany nie tylko niewłaściwym użyciem w badaniach matematyki (aby tylko była!!!), ale przede wszystkim matematyki pozbawionej racjonalnej formalizacji na gruncie geografii społeczno-ekonomicznej. Wymieniając tylko niektórych spośród wielu uznanych później geografów tworzących podwaliny wykorzystywania metod matematyczno-statystycznych przede wszystkim trzeba wspomnieć takich badaczy, jak: R. Chorley, A. Wilson czy P. Haggett, który w artykule podsumowującym rozwój geografii ilościowej (Haggett, 2008) opisał tę rewolucję w badaniach na wiodącym wówczas Uniwersytecie w Cambridge, a w pracy z 2019 r. zwrócił uwagę na niektóre aspekty współczesnych zmian w geografii ekonomicznej, w tym na „trwanie” rewolucji ilościowej i analizy systemów jako oczywistą prawdę (Haggett, 2019).

Wśród polskich badaczy wykorzystujących nowatorsko te metody (od lat 70. XX w.) należałoby wymienić: T. Czyż, Z. Chojnickiego i W. Ratajczaka z UAM w Poznaniu, którzy nie tylko byli pionierami zastosowań zaawansowanych metod statystycznych we wnioskowaniu, ale również zajmowali się analizą wartości i użyteczności tych metod pod względem teoretycznym (Czyż, Chojnicki, 2019). Znaczące osiągnięcia w tym zakresie (dokładniej istotny wkład w metody taksonomiczne) mieli też J. Perkal i B. Kostrubiec z Uniwersytetu we Wrocławiu.

Do wspomnianego powyżej przełomu skutkującego wprowadzeniem nowych metod statystycznych w badaniach w geografii przyczynili się też P.A.P. Moran i R.C. Geary, którzy odpowiednio w roku 1950 i 1954 zaproponowali metodę pomiaru w sposób matematyczny wpływu sąsiedztwa w przestrzeni poprzez wprowadzenie współczynnika autokorelacji przestrzennej. W konsekwencji intensywnie zaczął się rozwijać zarówno pod względem teoretycznym, jak i praktycznym interdyscyplinarny dział nazywany dzisiaj statystyką przestrzenną. Myśląc o promotorach tych działań, należy wymienić Cliffa i Orda, którzy przyczynili się nie tylko w istotny sposób do rozwoju podstaw teoretycznych statystyki przestrzen-

nej, ale też stosowali je do zagadnień praktycznych. W szczególności ważne w tym zakresie są: artykuł Clifffa i Orda (1969), monografie Clifffa i Orda (1973, 1981) oraz monografia Clifffa i in. (1975). Monografie te, zawierając istotne elementy – swoiste kamienie milowe statystyki przestrzennej – wywarły ogromny wpływ na analityków danych geograficznych i używane przez nich metody.

W latach 80. XX w. zaczęto produkować i wykorzystywać do obliczeń komputery, co stworzyło możliwość wykonywania bardziej skomplikowanych matematycznie analiz oraz obliczenia na większych zbiorach danych. Także w rozwoju statystyki przestrzennej i ekonometrii przestrzennej nastąpiło przejście do następnej fazy ze znaczącym udziałem takich badaczy – geografów społeczno-ekonomicznych – jak L. Anselin (w 2010 r. opisał w monografii 30-letnią historię ekonometrii przestrzennej), D.A. Griffith (patrz np. Griffith, 1987), S. Openshaw, D. Wartenberg, M. Fischer czy P. Nijkamp. Należy też wspomnieć o dwóch kluczowych badaczach, którzy przyczynili się do rozwoju ekonometrii przestrzennej, mianowicie o J.H.P. Paelnicku i L. Klaassenie. To właśnie J.H.P. Paelnick w 1974 r. wprowadził pojęcie *spatial econometrics* (ekonometria przestrzenna) i przez kilkadziesiąt kolejnych lat intensywnie tę dyscyplinę naukową rozwijał oraz stosował jej metody w badaniach empirycznych. Jest on współautorem (m.in. z L. Klaassenem) wydanej w roku 1979 monografii, o której znaczeniu świadczy bardzo wysoka jej cytowalność (Paelinck i in., 1979). Lata 90. XX w. to początek ery komputerów – pecetów, wyposażanych w coraz większą pamięć wewnętrzną (zarówno RAM, jak i ROM) oraz dyski twarde. Za ich pomocą GIS (używany wcześniej do sporządzania map i ich stosowania w analizach) został przestawiony na tory przyspieszonego rozwoju. Dotyczy to zarówno aspektów teoretycznych, jak i aplikacyjnych, w tym zautomatyzowania metod graficznych, w szczególności sporządzania coraz lepszych jakościowo, „bogatszych” map. W tworzeniu oprogramowania związanego z GIS w ramach pakietu R (oraz jego poprzednika RCran) duży wkład wniósł R. Bivand. Rozpoczął on karierę naukową na UAM w Poznaniu, później trwale związał się z Norwegian School of Economics (Bergen), ale przez cały czas współpracował i współpracuje do dziś z UAM i innymi uczelniami w Polsce (w monografii Bivand i in., 2008 opisano metody analizy danych za pomocą pakietu R). Równoległe w ramach pakietu Matlab tworzeniem oprogramowania dla metod ekonometrii przestrzennej zajmował się LeSage (1998).

Od początku XXI w. dostępność wysoko wydajnych komputerów, poszerzających swoje moce obliczeniowe z każdym rokiem, jest już powszechna a używanie zaawansowanych pakietów obliczeniowych (SPSS, Statistica, pakiet R, ArcGis, Open GeoDa, Matlab) pozwala na przeprowadzanie dogłębnych i kompleksowych analiz statystycznych na ogromnych zbiorach danych, będąc niezwykle silnym aparatem narzędziowym wspomagającym modelowanie i wnioskowanie wszystkich badaczy, a zwłaszcza geografów społeczno-ekonomicznych.



W powyższym szkicu historycznym wspomniane są nazwiska tylko wybranych badaczy, co nie umniejsza istotnych osiągnięć innych naukowców, w tym polskich, którzy przyczynili się do rozwoju aplikacji metod statystycznych w geografii społeczno-ekonomicznej. Bardziej kompletny opis zająłby całą monografię.

## **METODY STATYSTYCZNE W TREŚCIACH KSZTAŁCENIA GEOGRAFÓW**

Naukowcy w każdej dyscyplinie w swoich badaniach (poza powtarzaniem metod użytych w sposób twórczy przez innych wprowadzających nowe metody) opierają się głównie na metodologii, której uczyli się w czasie studiów. Pierwsze podręczniki ze statystyki przeznaczone dla studentów geografii w Polsce to podręczniki Kostrubca i Jokiel (1976) oraz przetłumaczone z angielskiego podręczniki Gregory'ego (1970) i Norcliffe'a (1977). Dwa ostatnie w wersji angielskiej doczekały się wielu wydań, co potwierdza ich wartość dydaktyczną. W kolejnych latach wydawane były skrypty innych autorów. Podręczniki te zawierały elementy statystyki opisowej wraz z prezentacją graficzną danych, podstawy testowania hipotez parametrycznych i nieparametrycznych oraz analizy korelacji i regresji. Zawartością merytoryczną nie różniły się one generalnie od podręczników do statystyki dla studentów innych kierunków, ale zgodnie z tytułem dedykowane były studentom geografii i zawierały geograficzne przykłady. W latach 70. XX w. nauczanie studentów realizowane było jeszcze poprzez manualne wykonywanie obliczeń (czasami już z użyciem kalkulatorów). Jednakże wraz z powstawaniem coraz doskonalszych komputerów (zarówno pod względem szybkości procesora, jak i wielkości pamięci RAM) w drugiej połowie lat 80. XX w. (z niewielkim opóźnieniem względem państw Europy Zachodniej) także w Polsce zaczęto do obliczeń stosować komputery. Lata 90. XX w. to intensywna informatyzacja, możliwość bezpośredniego dostępu do komputerów i zaprogramowanych metod statystycznych, a przede wszystkim rozwój GIS. Należy tu podkreślić, że angielski skrót GIS odnosi się do dwóch pojęć: Geographic Information Science oraz Geographic Information System. Pierwsze z nich oznacza dyscyplinę badawczą (niewyróżnioną w polskiej klasyfikacji dyscyplin), która zajmuje się technikami uzyskiwania, przedstawiania, przetwarzania i analizowania informacji geograficznych, podczas gdy drugie odnosi się do narzędzi programowych. W języku polskim skrót ten odnosi się w większym stopniu do drugiego pojęcia (w tłumaczeniu: „system informacji geograficznej”), w szczególności w edukacji, ale z rozszerzeniem jego znaczenia w badaniach naukowych na to co związane jest z pierwszym pojęciem. GIS od początku powstania wprowadzany był kolejno jako obowiązkowy przedmiot nauczania studentów różnych kierunków studiów, ale w szczególności na wszystkich kierunkach studiów na wydziałach geograficznych uczelni w Polsce. Zatem w drugiej połowie lat 90. XX w. studenci kierunków geo-

graficznych dysponowali już narzędziem i programami pozwalającymi na wykonywanie samemu obliczeń statystycznych, można by powiedzieć według zautomatyzowanej procedury: wrzucić dane do komputera, dokonaj obliczeń i wydrukuj wynik. Czy założenia teoretyczne są spełnione i czy interpretacja uzyskanych wyników jest prawidłowa, może okazać się już nie tak ważne. Sprzyja temu również ograniczanie w nauczaniu podstaw matematycznych (co łatwo można sprawdzić, porównując obniżający się poziom trudności rozwiązywanych zadań w kolejnych wydaniach podręczników do matematyki, np. wydawnictwa Nowa Era). Działo się tak w ciągu ostatnich dwudziestu lat w szkołach podstawowych i średnich. Co prawda w 2010 r. przywrócono (po ponad 25 latach) obowiązek zdawania matury z matematyki, ale aby zdać tę maturę, wystarczy uzyskać 30% punktów (z coraz łatwiejszych zadań maturalnych). W konsekwencji spowodowało to obniżenie poziomu nauczania matematycznego na studiach pozamatematycznych (nie tylko w Polsce), w tym zwłaszcza na studiach geograficznych. Przedmioty wcześniej mające w nazwie „elementy matematyki i statystyki” (sama nazwa działała odstraszała) obecnie często zamienia się na przedmioty mające w nazwie „statystyczna analiza danych”, z prawie całkowitym wyeliminowaniem nauczania elementów matematyki wyższej. Powoduje to, że absolwenci posiadają uboższą wiedzę w zakresie matematyki i podstaw teoretycznych statystyki. Aktualne jest więc pytanie, czy student, przyszły badacz w dziedzinie geografii, jest świadom czekającej go pułapki wynikającej z niepełnej wiedzy. Jest to szczególnie istotne w dobie powszechnej cyfryzacji. Zgodnie z przyjmowaną logiką myślenia: wynik uzyskany za pomocą komputera nie powinien być kwestionowany. Można by oczywiście dyskutować o rozszerzeniu edukacji studentów o zaawansowane metody ilościowe, jak sugerują (w odniesieniu do brytyjskiej edukacji) Harris i in. (2013). Celem artykułu nie jest krytyka zmian w edukacji (można na nie spojrzeć jak na swoiste *signum temporis*) i sugestia nowej rewolucji w tym zakresie, ale uwypuklenie ewentualnych niepożądanych konsekwencji wynikających ze zmian już dokonanych.

## **OGRANICZENIA, KORZYŚCI ORAZ NOWE TRENDY W UŻYCIU METOD STATYSTYCZNYCH WSPÓŁCZEŚNIE W GEOGRAFII SPOŁECZNO-EKONOMICZNEJ**

### **Ograniczenia**

Metody statystyczne, jak zaznaczono we wstępie, są stosowane jako pomocne narzędzie we wnioskowaniu w prawie wszystkich dziedzinach nauk. Nie można jednak zapominać o tym, że nawet najlepsze narzędzie podlega pewnym ograniczeniom (jak każde narzędzie) oraz że jego użytkownik powinien mieć pełną tego świadomość, w tym również – do czego ono służy. Świadomość ograniczeń

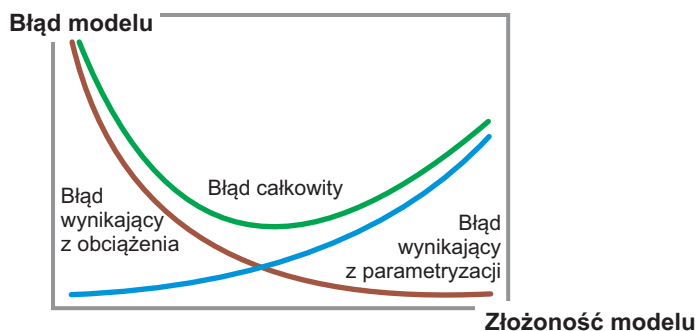
w korzystaniu z metod matematyczno-statystycznych w geografii społeczno-ekonomicznej mieli już badacze, którzy pioniersko ich używali, zarówno w USA i krajach Europy Zachodniej, jak i w Polsce (np. Czyż, Ratajczak, 1986; Parysek, Ratajczak, 2002).

Wśród znanych od lat problemów (prowadzących do ograniczeń w użyciu metod matematyczno-statystycznych w geografii społeczno-ekonomicznej) wymienić należy przede wszystkim:

1. wątpliwości, czy mamy do czynienia ze zmienną losową,
2. trudności w rozróżnieniu, czy mamy do czynienia z próbą, a nie populacją,
3. ograniczenia we wnioskowaniu dla cech jakościowych.

Nie można podać jednej kompletnej recepty, jak na etapie wstępnym uniknąć błędów we wnioskowaniu związanych z tymi problemami. Jedna ogólna sugestia (dość przydatna zawsze, gdy mamy do czynienia z „zakłóceniem” losowości) brzmi: w razie wątpliwości ograniczamy się do stosowania tylko metod matematycznych. Oczywiście oznacza to niestety, że nie możemy posiłkować się pojęciem „statystycznego poziomu istotności”.

Kolejny ważny problem (także znany od dawna) związany jest z doborem właściwej metody statystycznej czy też modelu w analizie danych reprezentujących określone zagadnienie. Jeśli model jest niepoprawny, np. ze względu na niespełnienie określonych założeń statystycznych, to prowadzi on do mało wiarogodnych wniosków. W tym przypadku pomocne są: ciągły rozwój statystyki pod względem teoretycznym (propozycje modeli z uproszczonymi założeniami) oraz komputery o wielkiej mocy obliczeniowej pozwalające na uzyskiwanie skomplikowanych estymatorów. Bardzo często są to estymatory asymptotyczne lub bayesowskie, obliczane w sposób iteracyjny, w tym z użyciem metody flip-flop. Istnieje możliwość implementacji zaawansowanych procedur numerycznych, w których kontrolowana jest zbieżność i punkt stopu. Trzeba tu dodać, że konieczność wykonania nawet bardzo dużej liczby iteracji nie stanowi już dziś tak poważnej przeszkody jak dawniej. Nie ma ponadto tak dużego znaczenia jak – jeszcze kilkanaście lat wcześniej – koszt dłuższych obliczeń. Ale i na tym polu w dalszym ciągu możemy spodziewać się przeszkód trudnych do pokonania. Saltelli (2017) w pracy pod znamienym tytułem *Should statistics rescue mathematical modeling?* („Czy statystyka powinna ratować modelowanie matematyczne?”) stwierdza, że statystyka znajduje się obecnie w trudnym okresie spowodowanym zauważalnym nadużywaniem lub błędnym używaniem jej metod w kontekście tzw. kryzysu odtwarzalności. Metody i sposoby kwantyfikacji stosowane w modelowaniu matematycznym rzadko podnoszone są jako główny temat dyskusji naukowych, chociaż praktycy zajmujący się modelowaniem piszący w naukowych czasopismach dyscyplinarnych wskazują na wiele problemów w tej dziedzinie. Jednym z zagadnień poruszanych przez Saltellego, na które warto zwrócić uwagę, jest kwestia błędu w modelowaniu wynikającego ze stosowania coraz bardziej skomplikowanych



**Ryc. 1.** Zależność wielkości błędu modelu od stopnia złożoności modelu przy uwzględnieniu w analizie: błędu wynikającego z obciążenia oraz błędu wynikającego z parametryzacji

Źródło: Saltelli (2017, s. 8).

(złożonych) modeli. Rycina 1 przedstawia istotę tego problemu (jest to konceptualna ilustracja analogii między statystycznym a matematycznym modelowaniem zaprezentowana już w 1973 r. przez R.W. O’Neilla). Całkowity błąd modelu jest wypadkową dwóch odwrotnie proporcjonalnych błędów. Minimalizując błąd wynikający z parametryzacji (większa liczba parametrów), poprzez użycie bardziej złożonego modelu zwiększamy jednocześnie błąd wynikający z obciążenia (błąd systematyczny), a złotym środkiem jest przyjęcie takiego poziomu złożoności, który na rycinie odpowiada minimum paraboli (kolor zielony) przedstawiającej błąd całkowity – swoiste equilibrium między obciążeniem a złożonością modelu.

O trudnościach wynikających ze stosowania metod matematyczno-statystycznych w geografii biorąc pod uwagę różnorodne aspekty, pisali też: Ferreira i Vale (2020), Johnston i in. (2020), Johnston i Jones (2020) oraz Derbyshire (2020).

Najważniejszy dziś problem ze stosowaniem metod matematyczno-statystycznych w geografii bez wątplenia związany jest z GIS (traktowanym jako zestaw narzędzi obliczeniowych). GIS to ogromny magazyn z narzędziami, przy użyciu których (jak sądzi niemała grupa badaczy geografów) można obliczyć wszystko i (w konsekwencji) wnioskować o wszystkim, jeśli tylko dysponujemy dużym zbiorem danych. Pułapka tkwi w tym, że nawet w bardzo dużym zbiorze danych niekoniecznie zakodowana jest prawidłowość, którą (o dziwo!) można „wykryć”, stosując GIS.

W rozwoju GIS od początku dostrzegano pewne niebezpieczeństwo. Stan Openshaw w artykule z 1991 r. zwrócił uwagę na fakt, że geografia nie przypomina fizyki czy chemii, w których to dyscyplinach mamy do czynienia z silnym stopniem uniwersalnego uporządkowania leżącym u podstaw złożonego zjawiska. Problemy geografii są pod pewnymi względami znacznie trudniejsze, zwłaszcza z perspektywy normalnego naukowego punktu widzenia. Jednak niektórzy

geografowie zareagowali na nie, tworząc coś, co jest niczym więcej niż prawdopodobnym dziełem fikcji, a następnie przypisują mu wiarygodność akademicką poprzez rozwijanie podstaw filozoficznych, które sprawiają, że normalny dowód naukowy nie jest konieczny. W ten sposób empiryczną weryfikację przewidywalności procesów uważa się za nieadekwatną, niewystarczającą, a nawet wadliwą. Dodaje on jednak, że spostrzeżenie to nie uprawnia do pełnego krytycyzmu GIS. Swoje rozważania Openshaw kontynuuje w artykule z 1998 r. zauważając, że to nie geografia ilościowa zawiodła, wskazując na mocne i słabe aspekty stosowania GIS. Do mocnych zalicza m.in. fakt, że zapewniają użyteczną technologię analizy danych empirycznych a w stosownych przypadkach można korzystać z analizy numerycznej, modelowania i przewidywania. Do słabych (poza wskazaną powyżej uwagę ogólną) zalicza możliwość zaistnienia sytuacji, w której obiektywizm może być iluzją, ponieważ metody i dane nie muszą być neutralne. Dziesięcioletnią krytyczną dyskusję o GISach podsumowuje Schuurman (2000). Jego podstawowy zarzut zawarty jest w stwierdzeniu, że w geografii najbardziej strzeżonym terytorium nie jest ziemia, ale metody dyscyplinarne i, co za tym idzie, soczewki, przez które patrzy się na Ziemię. Wprowadzenie GIS nie przyczyniło się do złagodzenia tych podziałów. Ale to dopiero prawdziwy początek w dyskusjach na powyższy temat. Harvey (2014) zadaje przekorne pytanie o to, czy możliwy jest świat bez GIS, wskazując jednocześnie na ograniczenia, ale też możliwe kierunki rozwoju tego narzędzia. Zauważa on, że w dobie panowania (dominacji) technologii komputacyjnych następuje coraz większe zbliżenie świata fizycznego i cyfrowego, co doprowadzi do istotnej zmiany czy też zastąpienia obecnych technologii lokalizacyjnych czymś nowym, jeszcze dziś nieznanym.

Problemy ze stosowaniem GIS nie zostały rozwiązane przez kolejne lata ich stosowania, o czym piszą Czyż i Chojnicki (2019), stwierdzając: „Umiejętność uruchomienia i realizacja funkcji oprogramowania GIS nie oznacza jednak rozumienia przez badacza określonej metody statystycznej i matematycznej oraz jej właściwego zastosowania”. Jednocześnie dodają, że „metody matematyczne stały się skutecznym narzędziem uściślenia wiedzy geograficznej”, co w konsekwencji prowadzi do rozwoju „twardego” jej składnika.

Dyskusja na temat problemów z używaniem GIS w dalszym ciągu trwa (patrz np. Yuan, 2021a, b, Wilson, 2021) i trwać będzie bez wątpienia przez następne lata, co wynika z ciągłej rozbudowy tego narzędzia.

Problemy powyższe dotyczą oczywiście również innych narzędzi analizy danych. Programy statystyczne często traktowane są jak czarne skrzynki, które po wprowadzeniu danych zwracają liczne statystyki i sugerują wnioski, które bez dokładnej analizy i znajomości założeń nie powinny być wyciągane. Dodatkowym mankamentem niektórych zaawansowanych metod statystycznych jest kwestia oczekiwanej odtwarzalności (ang. *reproducibility*) – powtórzone obliczenia na tych samych danych powinny prowadzić do tych samych wyników – oraz kwe-

stia oczekiwanej powtarzalności (ang. *replicability*) – podobne wyniki powinno się uzyskać przy użyciu różnych próbek danych reprezentujących tę samą populację, różnego oprogramowania, które implementuje te same metody, lub obydwu tych rzeczy jednocześnie. Zwracają na to uwagę m.in. Goodchild i in. (2021) oraz Waters (2021).

Trzeba też dodać, że zdarzają się przypadki (ciągłe nowe), iż badacze używają metod matematyczno-statystycznych po to, by „naciągać” wyniki obliczeń w celu potwierdzenia przyjętych przez nich wcześniej hipotez, które nie do końca są prawdziwe, piszą o tym np. Wang (1992) i Barker (2021).

## **Korzyści**

Dla badań w geografii powstanie GIS i jego intensywny rozwój w latach 90. XX w. to szczególnie ważny moment. To tak jak wyposażenie manualnego laboratorium w sprzęt do badań automatycznych wielokrotnie zwiększających szybkość uzyskiwania wyników. Dzisiejszy aparat w postaci wysoko wydajnych (jeśli chodzi o szybkość obliczeń) komputerów pozwala wykonać analizę na ogromnych zbiorach danych (np. dane panelowe, czyli uwzględniające czas i przestrzeń), umożliwiając jednocześnie przeprowadzenie wnioskowania na prawie każdym poziomie szczegółowości. Powszechna cyfryzacja stworzyła możliwości kompleksowej kolekcji danych w skali europejskiej. Faktyczne zbieranie danych na wszystkich poziomach, takich jak: Unia Europejska, państwa, regiony, subregiony, województwa, powiaty, gminy, miasta, jednostki municypalne, instytucje, wybrane grupy zawodowe czy gospodarstwa domowe i dostęp do nich (czasami niestety płatny), to szansa, o jakiej mogliby tylko pomarzyć badacze w niedalekiej przeszłości. Należy też dodać, że to, co z jednej strony jest wadą – automatyzacja obliczeń oraz uzyskiwanie niemalże natychmiast „gotowych” (niekoniecznie uzyskanych właściwie pod względem statystycznym) wyników – może być z drugiej strony zaletą. Stwarza bowiem możliwość natychmiastowej weryfikacji różnych skomplikowanych obliczeń oraz szansę porównywalności wyników uzyskiwanych z różnych pakietów. Sprzyja też „rozprzestrzenianiu się” wiedzy i umiejętności poprzez korzystanie z tych pakietów, a nawet „zmusza” tych bardziej dociekliwych do poszerzania swoich horyzontów badawczych.

## **Nowe trendy**

Mankamenty obciążające wnioskowanie w używanych metodach statystycznych, ujawniające się w trakcie ich empirycznego stosowania, są punktem wyjścia do korekty starych metod i konstrukcji nowych, wyznaczając nowe trendy w stosowaniu metod statystycznych. Dzieje się tak również w szczególności w przypadku geografii społeczno-ekonomicznej, w ramach której rozważania teoretyczne od-

grywają istotną rolę na gruncie statystyki i ekonometrii przestrzennej. Można by powiedzieć, że nowatorstwo oznacza tu też rozwiązywanie starych problemów nowymi metodami lub wcześniej już używanymi, ale zmodyfikowanymi. Śledząc współczesną literaturę, można wskazać (odwołując się do wybranych publikacji) m.in. następujące nowe trendy badań:

1. modelowanie czasowo-przestrzenne (panelowe) oraz analizy big data – Elhorst (2003, 2014), Ferreira, Vale (2020);
2. nowe podejście do zmiennych jakościowych w modelowaniu, uwzględnianie danych symbolicznych – Bivand i in. (2017), Anselin, Li (2019);
3. nowe spojrzenie na modele regresyjne – przestrzenna regresja kwantylowa, modele regresji lokalnej – Rachdi i in. (2021), przestrzenne modele wielopoziomowe – Łaskiewicz (2016), semiparametryczne modele przestrzenne – Goodwin i in. (2021);
4. nieaprioryczne podejście do macierzy wag – Herrera i in. (2019);
5. przestrzenność w statystyce małych obszarów – Mehrabi i in. (2020).

Można by powiedzieć, że nowatorstwo oznacza również rozwiązywanie starych problemów nowymi metodami lub wcześniej już używanymi, ale zmodyfikowanymi. W tej grupie nowych trendów warto zwrócić uwagę na cztery następujące nowe podejścia, które poza wskazaniem literaturowym zostaną pokrótce scharakteryzowane:

1. Nowe metody estymacji.

Ze względu na problem tzw. endogeniczności składnika opóźnienia przestrzennego w modelach regresji, w ogólności parametry modelu nie mogą być estymowane bezpośrednio metodą najmniejszych kwadratów. Wynika stąd potrzeba poszukiwania specjalistycznych metod estymacji. Za metodę podstawową można uznać podejście oparte na metodzie największej wiarygodności (MNV), zaprezentowane już w latach 80. Niemniej jednak, jak zauważają Olejnik i Olejnik (2020a), konieczność numerycznej optymalizacji funkcji, zawierającej trudny obliczeniowo wyznacznik macierzy dużych rozmiarów, spowodował, że dopiero wzrost mocy obliczeniowej komputerów oraz wypracowanie wachlarza rozwiązań algorytmicznych pozwoliły na rozkwit tej metodologii już w pierwszym dziesięcioleciu nowego milenium. Metodami alternatywnymi dla MNV stały się algorytmy estymacyjne oparte na metodzie zmiennych instrumentalnych oraz uogólnionej metodzie momentów (UMM), w tym, dla szczególnych specyfikacji, procedury wieloetapowe (zapoczątkowane przez Kelejian i Prucha, 1998). W pewnym sensie problematyczne technicznie pozostają specyfikacje tzw. wyższego rzędu, w których opóźnienie przestrzenne ma charakter wielowymiarowy i reprezentowane jest przez więcej niż jedną macierz wag. W takiej sytuacji mamy do czynienia z zauważalnym, choć niekrytycznym spadkiem wydajności metod największej wiarygodności. Stosowane jest wówczas podejście UMM (Lee, Liu,

2010) lub metody bayesowskie, oparte na iteratywnym próbkowaniu Monte Carlo i łańcuchach Markowa (patrz LeSage, 2020).

## 2. Modele asymptotyczne (Olejnik, Olejnik, 2020a, b).

Zwiększające się możliwości obliczeniowe komputerów dostępnych dla badaczy pozwalają na przeprowadzanie wnioskowania opartego na coraz większych zbiorach danych. W związku z tym faktem pojawia się potrzeba formułowania teorii matematycznej zachowania asymptotycznego stosowanych w praktyce estymatorów. I tak np. teoria dużej próby dla oszacowań MNW dla popularnych w geografii społeczno-ekonomicznej ekonometrycznych modeli przestrzennych, choć zainicjowana znacznie wcześniej, została sformalizowana dopiero w pracy Lee (2004). Nastąpił wtedy gwałtowny rozwój tej metodologii, a twierdzenia graniczne dotyczące estymatorów zarówno MNW, jak i nieco ogólniejszego podejścia opartego na UMM były rozszerzane w kierunku coraz to bardziej rozbudowanych specyfikacji modelowych. Niedawno zaś w publikacjach Olejnik i Olejnik (2020a, b) jej matematyczne podstawy zostały zrewidowane, a w efekcie uzyskano teorię pozwalającą na wnioskowanie wielkiej próby dla modeli z macierzami wag reprezentującymi większą niż w teorii klasycznej ilością interakcji przestrzennych. Jak dotąd teorie asymptotyczne estymatorów dla modeli przestrzennych wymagały założenia o tzw. sumowalności wag (wierszowej i kolumnowej) występujących w specyfikacji macierzy. Jak pokazali Olejnik i Olejnik (2020a, b), takie założenie można złagodzić. W szczególności opracowanie to dostarczyło formalne uzasadnienie dla używania popularnych w badaniach geograficznych macierzy wag opartych na odwrotności odległości (ang. *Inverse Distance Weighting*) w „newtonowskiej” drugiej potędze.

## 3. Modelowanie przestrzennej niestacjonarności (Ghiringhelli i in., 2021).

Wśród podstawowych założeń w statystyce przestrzennej istotna jest stacjonarność drugiego stopnia. Implikuje ona homogeniczność oraz izotropię. Jednakże gdy bazuje się na przestrzennym polu losowym w modelowaniu społeczno-ekonomicznym czy też danych epidemiologicznych, zdarza się, że przyjęcie powyższego założenia jest nieuzasadnione. W pracy Ghiringhelli i in. (2021) zaproponowano w takich sytuacjach wprowadzanie do modeli utajonych (ukrytych) losowych pól Markowa, które mają zapewnić bardziej realistyczną realizację analizowanego procesu. Prowadzi to w autoregresji przestrzennej typu modelu liniowego Cliffa-Orda do uzyskiwania wartości współczynników poprzez identyfikację grup obserwacji o podobnym zachowaniu. Proponowane rozwiązanie zawiera zarówno uzasadnienie teoretyczne, jak i empiryczną analizę. Stanowi więc doskonały punkt wyjścia do dalszych uogólnień i bez wątpienia zajmować się tym będą inni badacze, rozwijając to podejście.



4. Przestrzenne analizy efektywności – SDEA (*spatial data envelopment analysis*) (Żółtaszek, Olejnik, 2021).

W pracy zaprezentowano nowe podejście do pomiaru efektywności w badaniach regionalnych poprzez konstrukcję SDEA. Autorzy rozszerzyli klasyczną metodę DEA poprzez jawne włączenie interakcji przestrzennych (przy użyciu macierzy wag przestrzennych) do modelu. Zbiór zmiennych rozszerzono o przestrzennie autokorelowane zmienne wejściowe i wyjściowe. W rezultacie dopuszczono, by nakłady w jednym regionie wpływały na wyniki w regionach sąsiednich, a efekty mogły być wykorzystywane jako nakłady lub efekty dla regionów sąsiednich. Dzięki temu analiza efektywności przeniesiona na grunt geografii społeczno-ekonomicznej jest w stanie uwzględniać zjawiska współpracy i konkurencji między geograficznie zbliżonymi/sąsiadującymi jednostkami przestrzennymi. Podobnie jak w wersji klasycznej, wykonywanie takich analiz uwarunkowane jest możliwością przeprowadzenia relatywnie pracochłonnych obliczeniowo optymalizacji z ograniczeniami. Zainteresowanie tymi metodami można niewątpliwie wiązać z ogólną dostępnością środowisk i pakietów programowania służących do obliczeń naukowych.

## PODSUMOWANIE

Geografia społeczno-ekonomiczna (wraz z gospodarką przestrzenną) dwa lata temu została wyodrębniona jako nowa dyscyplina w dziedzinie nauk społecznych w zreformowanej klasyfikacji dziedzin naukowych. Jest to niewątpliwym sukcesem pokazującym jej wysoką pozycję, w szczególności w świetle wątpliwości i zarzutów co do jakości całej geografii jako nauki podnoszonych ponad 70. lat temu (o czym wspomniano powyżej). Bez wątpienia do tego sukcesu w dużej mierze przyczyniło się jej metodologiczne oparcie na szkieletcie zbudowanym z metod statystycznych stosowanych we wnioskowaniu przez badaczy w tej dziedzinie. W rozbudowywaniu tego szkieletu biorą udział sami geografowie społeczno-ekonomiczni będący często jednocześnie specjalistami w zakresie statystyki i ekonometrii przestrzennej. Zatem jakikolwiek rozwój stosowanych tu metod jest wewnętrznie kontrolowany, w tym również zdarzające się merytoryczne nadużycia. Niezależnie od tego konieczne jest jednak ciągle przypomnianie o istniejących zagrożeniach, zwłaszcza wobec używania nowych lub ulepszanych metod oraz nowych trendów w dobie powszechnej (ale jeszcze nie zakończonej) cyfryzacji.

## LITERATURA

Anselin L. (2010). Thirty years of spatial econometrics *Papers in Regional Science*, 89(1): 3–25.

- Anselin L., Li X. (2019). Operational local join count statistics for cluster detection. *Journal of Geographical Systems*, 21(2): 189–210.
- Barker H. (2020). *Lying Numbers: How Maths and Statistics Are Twisted and Abused*. London: Little, Brown Book (polskie tłumaczenie: 2021, *Liczby, które nie kłamią*, PWN).
- Barnes T.J. (1994). Probable writing: Derrida, deconstruction, and the quantitative revolution in human geography. *Environment and Planning, A*, 26(7): 1021–1040.
- Birkin M., Clarke G., Clarke M., Wilson A. (2014). The achievements and future potential of applied quantitative geography: a case study. *Geographia Polonica*, 87(2): 179–202.
- Bivand R.S., Pebesma E.J., Gómez-Rubio V., Pebesma E.J. (2008). *Applied spatial data analysis with R* (vol. 747248717, 237–268), New York: Springer.
- Bivand R.S., Wilk J., Kossowski T. (2017). Spatial association of population pyramids across Europe: The application of symbolic data, cluster analysis and join-count tests. *Spatial Statistics*, 21: 339–361.
- Burton I. (1963). The quantitative revolution and theoretical geography. 1. *Canadian Geographer/Le Géographe canadien*, 7(4): 151–162.
- Cliff A.D., Haggett P., Ord J.K., Bassett K.A., Davies R., Bassett K.L. (1975). *Elements of Spatial Structure: A Quantative Approach*. Vol. 6. Cambridge University Press.
- Cliff A.D., Ord J.K. (1969). The Problem of Spatial Autocorrelation, W: A.J. Scott (red.), *Studies in Regional Science London*. Papers in Regional Science, s. 25–55.
- Cliff A.D., Ord J.K. (1973). *Spatial Autocorrelation*. London: Pion.
- Cliff A.D., Ord J.K. (1981). *Spatial processes: models & applications*. Taylor & Francis.
- Crépel P., Rieucan J.N. (2005). Condorcet's social mathematic, a few tables. *Social Choice and Welfare*, 25(2): 243–285.
- Curry L. (1967). Quantitative geography. *Canadian Geographer/Le Géographe canadien*, 11(4): 265–279.
- Czyż T., Chojnicki Z. (2019). Rola poznańskiego ośrodka geograficznego w implementacji metod i modeli matematycznych w geografii społeczno ekonomicznej. *Rozwój Regionalny i Polityka Regionalna*, 4: 9–21.
- Czyż T., Ratajczak W. (1986). Mathematical methods in economic geography. *Concepts and Methods in Geography*, 1: 99–126.
- Derbyshire J. (2020). Answers to questions on uncertainty in geography: Old lessons and new scenario tools. *Environment and Planning, A: Economy and Space*, 52(4): 710–727.
- Donnelly K. (2015). *Adolphe Quetelet, social physics and the average men of science, 1796–1874*. Routledge.
- Elhorst J.P. (2003). Specification and estimation of spatial panel data models. *International regional science review*, 26(3): 244–268.
- Elhorst J.P. (2014). *Spatial econometrics from cross-sectional data to spatial panels*. Springer.
- Ferreira D., Vale M. (2020). Geography in the big data age: an overview of the historical resonance of current debates. *Geographical Review*. <https://doi.org/10.1080/00167428.2020.1832424>
- Freudenthal H. (1975). Quetelet, Lambert-Adolphe-Jacques. *Dictionary of Scientific Biography*, 11: 236–238.
- Galton F. (1883). *Inquiries into human faculty and its development*. Macmillan.
- Garrison W. (1956). Applicability of statistical inference in geographical research. *Geographical Review*, 46: 427–429.

- Geary R.C. (1954). The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5: 115–41.
- Ghiringhelli C., Bartolucci F., Mira A., Arbia G. (2021). Modelling Nonstationary Spatial Lag Models with Hidden Markov Random Fields. *Spatial Statistics*, 44:100522.
- Goodwin B.K., Holt M.T., Prestemon J.P. (2021). Semi-parametric models of spatial market integration. *Empirical Economics*, s. 1–27.
- Goodchild M.F., Fotheringham A.S., Kedron P., Li W. (2021). Introduction: Forum on reproducibility and replicability in geography. *Annals of the American Association of Geographers*, 111(5): 1271–1274.
- Gregory S. (1963). *Statistical methods and the geographer*. London: Longman (polskie tłumaczenie, 1970, *Metody statystyki w geografii*, Państwowe Wydawnictwo Naukowe).
- Griffith D.A. (1987). *Spatial autocorrelation. A Primer*. Washington, DC, Association of American Geographers.
- Haggett P. (2008). The local shape of revolution: reflections on quantitative geography at Cambridge in the 1950s and 1960s. *Geographical Analysis*, 40(3): 336–352.
- Haggett P. (2019). Changing concepts in economic geography. W: *Frontiers in Geographical Teaching*. Routledge, s. 101–117.
- Harris R., Fitzpatrick K., Souch C., Brunson C., Jarvis C., Keylock C., ... , Tate N. (2013). *Quantitative methods in geography: making the connections between schools, universities and employers* (<https://orca.cardiff.ac.uk/59957/1/report-130906041556-.pdf>).
- Harvey F. (2014). A world without GIS? Post-GIS futures for the new millennium. *Geographia Polonica*, 87(2): 241–250.
- Herrera M., Mur J., Ruiz M. (2019). A comparison study on criteria to select the most adequate weighting matrix. *Entropy*, 21(2): 160.
- Huntington E. (1927). The quantitative phases of human geography. *The Scientific Monthly*, 25(4): 289–305.
- Janowski B. (1908). *O odległościach jako czynniku rozwoju kultury*. Nakł. Towarzystwa dla Popierania Nauki Polskiej (angielskie tłumaczenie: 2013, *On distances as a factor of cultural development*. Poznań: Bogucki Wyd. Nauk.
- Jędrzejczyk D. (1992). Benon Janowski as the Pioneer of Application of Gravitation Models (Patterns) in the Geographico-Economic Research. *Miscellanea Geographica, Regional Studies on Development*, 5(1): 213–216.
- Jędrzejczyk D. (2000). Mathematical Method of Analysing Forms of the Rural Settlements of Benon Janowski. *Miscellanea Geographica, Regional Studies on Development*, 9(1): 159–164.
- Johnston R., Harris R., Jones K., Manley D., Wang W.W., Wolf L. (2020). Quantitative methods II: How we moved on – Decades of change in philosophy, focus and methods. *Progress in Human Geography*, 44(5): 959–971.
- Johnston R., Jones K. (2020). A classic that wasn't: Statistical Geography and paths only later taken. *Progress in Human Geography*, 44(2): 357–373.
- Kelejian H.H., Prucha I.R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), 99–121.
- Kostrubiec B., Jokiel B. (1976). *Statystyka z elementami matematyki dla geografów*. Uniwersytet Wrocławski im. Bolesława Bieruta.

- Lee L.F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6): 1899–1925.
- Lee L.F., Liu X. (2010). Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory*, 26(1): 187–230.
- LeSage J.P. (1998). *Applied econometrics using MATLAB*. University of Toledo.
- LeSage J.P. (2020). Fast MCMC estimation of multiple W-matrix spatial regression models and Metropolis–Hastings Monte Carlo log-marginal likelihoods. *Journal of Geographical Systems*, 22(1): 47–75.
- Łaszkiwicz E. (2016). *Ekonometria przestrzenna III: modele wielopoziomowe-teoria i zastosowania*. Wydawnictwo CH Beck.
- Mackay J.R. (1959). Regional geography: A quantitative approach. *Cahiers de Géogr. de Québec*, 3me ann. (Apr.-Sept. 1959): 57–63.
- Maier D. (2011). Francis Galton: Measuring the immeasurable. *Significance*, 8(3): 122–123.
- Marchand B. (1974). Quantitative geography: Revolution or counter revolution? *Geoforum*, 5(1): 15–23.
- Mehrabi Y., Kavousi A., Baghestani A.R., Soltani-Kermanshahi M. (2020). Spatiality in small area estimation: A new structure with a simulation study. *Geospatial Health*, 15(2).
- Moran P.A.P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1): 17–23.
- Norcliffe G.B. (1977). *Inferential statistics for geographers: an introduction*. London: Hutchinson (polskie tłumaczenie: 1981, *Statystyka dla geografów: wprowadzenie*. Państwowe Wydawnictwo Naukowe).
- Olejnik A., Olejnik J. (2020a). *Metody stochastyczne w ekonometrii przestrzennej – nowoczesna analiza asymptotyczna*. Wydawnictwo Uniwersytetu Łódzkiego.
- Olejnik A., Olejnik J. (2020b). QML estimation with non-summable weight matrices. *Journal of Geographical Systems*, 22(4): 469–495.
- O’Neill R.V. (1973). *Error analysis of ecological models*. W: *Radionuclides in ecosystems*. National Technical Information Service, Springfield, s. 898–908.
- Openshaw S. (1991). A view on the GIS crisis in geography, or, using GIS to put Humpy-Dumpty back together again. *Environment and Planning*, A, 23(5): 21–628.
- Openshaw S. (1998). Towards a more computationally minded scientific human geography. *Environment and Planning*, A, 30(2): 317–332.
- Paelinck J.H.P., Klaassen L.H., Ancot J.P., Verster A.C.P. (1979). *Spatiale conometrics*. Vol. 1. Saxon House.
- Parysek J., Ratajczak W. (2002). Analiza składowych głównych, jej korzyści i ograniczenia z punktu widzenia badań geograficznych. W: H. Rogacki (red.), *Możliwości i ograniczenia zastosowania metod badawczych w geografii i gospodarce przestrzennej*. Bogucki Wyd. Nauk., Poznań, s. 61–73.
- Pragacz P. (2007). Życie i dzieło Józefa Marii Hoene-Wrońskiego. *Roczniki Polskiego Towarzystwa Matematycznego*, seria 2: *Wiadomości Matematyczne*, 43: 67–86.
- Rachdi M., Laksaci A., Al-Awadhi F.A. (2021). Parametric and nonparametric conditional quantile regression modeling for dependent spatial functional data. *Spatial Statistics*, 43, 100498.

- Ratajczak W. (2016). *Great Minds in Regional Science: Social Physics in Benon Janowski's Spatial Economics Works*. Prezentacja podczas konferencji: URBAN EMPIRES: Cities as Global Rules in the New Urban World, Poznań, 15 August 2016 (<https://www.slideshare.net/regionalscienceacademy/great-minds-in-regional-science-benon-janowski>).
- Saltelli A. (2017). Discussion Paper: Should statistics rescue mathematical modelling? *arXiv preprint arXiv: 1712.06457*.
- Schuurman N. (2000). Trouble in the heartland: GIS and its critics in the 1990s. *Progress in Human Geography*, 24(4): 569–590.
- Yuan, M. (2021a). GIS research to address tensions in geography. *Singapore Journal of Tropical Geography*, 42(1), 13–30.
- Yuan M. (2021b). Cogitating on GIS. Response to David O'Sullivan and Matthew Wilson. *Singapore Journal of Tropical Geography*, 42(1): 40–44.
- Wallis B.C. (1912). The Use of Formulæ in Quantitative Geography. *Geographical Journal*, s. 175–178.
- Wang C. (1992). *Sense and Nonsense of Statistical Inference: Controversy: Misuse, and Subtlety*. Vol. 6. CRC Press.
- Waters N. (2021). Motivations and methods for replication in geography: Working with data streams. *Annals of the American Association of Geographers*, 11(5): 1291–1299.
- Wilson M.W. (2021). Corrective optics. A commentary on May Yuan's 'GIS research to address tensions in geography'. *Singapore Journal of Tropical Geography*, 42(1): 36–39.
- Żółtaszek A., Olejnik A. (2021). Regional effectiveness of innovation: leaders and followers of the EU NUTS 0 and NUTS 2 regions. *Innovation, The European Journal of Social Science Research*. <https://doi.org/10.1080/13511610.2020.1870440>.