

Olsztyn, 17 grudnia 2024 r.

Dr hab. inż. Elżbieta Lewandowicz
Katedra Geoinformacji i Kartografii
Instytut Geodezji i Budownictwa
Wydział Geoinżynierii
Uniwersytet Warmińsko Mazurski w Olsztynie

Recenzja rozprawy doktorskiej
mgr inż. Sylwii Borkowskiej
pt. Metodyka wielocechowej oceny porównawczej przydatności
jakościowych danych przestrzennych

1. Podstawa formalno-prawna

Podstawą sporządzenia recenzji jest pismo Przewodniczącego Rady Dyscypliny Naukowej Inżynieria Lądowa, Geodezja i Transport Wojskowej Akademii Technicznej z dnia 22 października 2024 roku i umowa zawarta pomiędzy Wojskową Akademią Techniczną a recenzentem. Zgodnie z umową, recenzja została wykonana w oparciu o wymagania określone w art. 187 ust. 1, 2, 3 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2023 r., poz. 724 z późn. zm.). Celem recenzji jest ocena dorobku, w tym oryginalnego rozwiązania problemu naukowego możliwego do zastosowania w sferze gospodarczej, prezentowanego w rozprawie doktorskiej mgr inż. Sylwii Borkowskiej.

2. Wstęp - ocena problematyki badawczej

Doktorantka podjęła tematykę badawczą związaną z jakością danych przestrzennych. Międzynarodowy standard, w normie ISO 19157-1:2023, określa mierniki jakości danych, które powinny być opisane w metadanych urzędowych zbiorów. Jednak obok zbiorów urzędowych dostępne są i powszechnie używane zbiory tworzone przez wolontariuszy, np. Open Street Map (OSM). Są one także elementem urzędowych systemów informacji przestrzennej, wykorzystywane są do licznych wizualizacji, analiz i służą do podejmowania decyzji. Jakość zbiorów danych OSM, wykonywanych przez społeczników o różnych umiejętnościach, jest widoczna przy szczegółowym przeglądaniu danych geometrycznych i atrybutowych pojedynczych obiektów. Jest ona systematycznie poprawiana przez organizację OSM, szczególnie po zgłoszeniach błędów. Jakość danych OSM jest obiektem licznych prac naukowych. Tematyka doktoratu przyłącza się do tych prowadzonych badań.

3. Ocena merytoryczna przeprowadzonych badań

3.1. Ogólna charakterystyka rozprawy

Rozprawę stanowi cykl 4 publikacji wieloautorskich, w których doktorantka jest pierwszym autorem. Dwie prace są opublikowane w wydawnictwie Multidisciplinary Digital Publishing Institute (MDPI), w czasopiśmie *Sustainability i Applied Science*, o wskaźnikach Impact Factor (IF), odpowiednio powyżej 3 i 2, z przypisanymi 100 punktami na liście ministerialnej z 2024r. Dwie kolejne, opublikowane zostały w polskich wydawnictwach - Państwowej Akademii Nauk i Akademii Górniczo-Hutniczej, z przypisanymi 70 punktami na liście ministerialnej.

- [1] Borkowska, S.; Pokonieczny, K. Analysis of OpenStreetMap Data Quality for Selected Counties in Poland in Terms of Sustainable Development. *Sustainability* **2022**, *14*, 3728. <https://doi.org/10.3390/su14073728>
- [2] Borkowska, S.; Pokonieczny, K.; Bielecka, E. OpenStreetMap – building data completeness visualization in terms of “Fitness for purpose”. *Advances in Geodesy and Geoinformation*, Vol.72 no 1, article n. e35, 2023 <https://doi.org/10.24425/agg.2022.141922> ;
- [3] Borkowska, S.; Bielecka, E.; Pokonieczny, K. Comparison of Land Cover Categorical Data Stored in OSM and Authoritative Topographic Data. *Applied Sciences*. 2023, *13*, 7525. <https://doi.org/10.3390/app13137525>
- [4] Borkowska, S.; Pokonieczny, K.; Bielecka, E. Weight Impact on Comparative Evaluation of Topographic Data. *Geomatics and Environmental Engineering*. • Volume 18, Number 4, 20240 <https://doi.org/10.7494/geom.2024.18.4.97>

3.2. Cele rozprawy, tezy badawcze

W autoreferacie rozprawy doktorskiej przedstawiono cel główny rozprawy. Jest nim opracowaniem uwarunkowań kompleksowej oceny porównawczej jakości danych przestrzennych. Wiąże się on ze wskazaniem metod analitycznych możliwych do przeprowadzenia przy podejmowaniu decyzji, związanej z wyborem bazy danych przestrzennych do wykonania konkretnego zadania. Sformułowano tezę badawczą, że metoda wielo cechowej analizy porównawczej i wizualizacji jakości danych oraz autorskie wskaźniki oceny przydatności danych przestrzennych stanowią podstawy do spójnej ich oceny przez użytkownika.

Szczegółowe cele badawcze przedstawiono w 4 publikacjach. Wiązały się z określeniem dokładności danych przestrzennych i atrybutowych danych OSM względem danych Bazy Danych Obiektów Topograficznych, opracowanej dla skali 1:10000 (BDOT_10k.). Bazę urzędową BDOT_10k przyjęto jako bazę referencyjną. Cele cząstkowe, zdefiniowane w kolejnych publikacjach, wiązały się z:

- [1] oceną elementów jakości danych za pomocą dokładności lokalizacji, kompletności danych i zgodności atrybutów,
- [2] oceną liczby budynków w zbiorze OSM, w stosunku do BDOT_10k, określoną za pomocą wskaźników kompletności i dokładności, z uwzględnieniem wskaźników statystycznych i analizą wyników przedstawionych za pomocą kartogramów dwuwymiarowych,

[3] analizą porównawczą danych ilościowych z OSM i BDOT_10k, za pomocą autorskich wskaźników Compound Correspondence Indeks (CCI), z uwzględnieniem wskaźników lokalnych CCI_{Ln} i regionalnych, CCI_{Rn} , z wagowaniem, zależnym od rozpoznawalności obiektów na zdjęciach satelitarnych i lotniczych,

[4] kontynuacją badań z publikacji [3], związaną z oceną ilościową, za pomocą autorskich wskaźników CCI, z uwzględnieniem wag przypisanych obiektom topograficznym, zależnych od preferencji użytkowników.

3.3. Obiekt badań, metodyka

Do badań przyjęto zbiory danych przestrzennych i atrybutowych w wybranych 7 powiatach, z różnych regionów Polski, charakteryzujących się różnymi cechami środowiska geograficznego i historycznego. Realizując cel, z dwóch baz OSM i BDOT_k10, wybrano klasy obiektów o różnej geometrii:

- powierzchniowej: budynki, wody powierzchniowe, lasy,
- liniowej: osie dróg, linii kolejowych, sieć rzeczna.

Badania oparto na polach podstawowych, w formie regularnej siatki sześciokątów o powierzchni 1 km^2 , wyznaczonych w granicach każdego powiatu. W każdej komórce sześciokątów wyznaczano wartości porównujące obiekty OSM, względem obiektów BDOT_10k.

W kolejnych publikacjach [1- 4], na danych z jednego, pięciu lub siedmiu wybranych powiatów, stosowano różne metody badawcze.

W publikacjach [1-2] badania prowadzono w oparciu o wskaźniki ISO, w publikacjach [3-4], w oparciu o autorską propozycję:

[1] w oparciu o dane z 5 powiatów, przeprowadzono badania dokładności geometrycznej i kompletności obiektów powierzchniowych, oparte o współrzędne punktów homologicznych, a dokładność obiektów liniowych, oparto o strefy buforowe wokół obiektów bazy referencyjnej BDOT_10k. Dokładność geometryczną obiektów powierzchniowych określono za pomocą błędów RMSE, w oparciu o współrzędne punktów homologicznych. Dokładność obiektów liniowych, określono w procentach (%). Kompletność obiektów powierzchniowych (C), określono w %, za pomocą porównania powierzchni obiektów OSM i BDOT_10k w polach podstawowych. Przy nakładaniu obiektów powierzchniowych wyznaczono dodatkowe cztery wskaźniki kompletności, w trzech różnych formach: część wspólna (TP), część tylko w OSM (FP), część tylko w BDOT_k10 (FN). Wyniki prezentowano za pomocą kartogramów w polach podstawowych. Zgodność atrybutową oceniono w bazie OSM poprzez ocenę procentową ilości obiektów z przypisanymi informacjami, do całkowitej liczby obiektów. Badania uzupełnią podstawowe wskaźniki statystyczne, wiążące wskaźniki jakości danych, z dochodem na mieszkańca w budżecie powiatów.

[2] metodę badawczą oparto o dane z jednego powiatu Piaseczyńskiego, z dwóch baz OSM i BDOT10K, w klasie budynków. Zastosowano metodykę wyznaczenia kompletności zabudowy, podobną jak w publikacji [1]. Metodę badań rozbudowano o analizy statystyczne oceniające normalność zbiorów za pomocą testu Shapiro-Wilk. Do prezentacji wyników w polach podstawowych, zastosowano kartogramy dwuwymiarowe, które rozszerzają interpretacje wyników, poprzez łączenie dwóch zmiennych.

W publikacjach [3-4], badania prowadzono w oparciu o autorskie rozwiązania na podstawie danych z 7 powiatów:

[3] rozszerzenie badań poprzez zastosowanie autorskiego wskaźnika Commodity Channel Index (CCI) - wskaźnika różnic w powierzchniach budynków, lasów, wód i długościach dróg, torów kolejowych, sieci wodnej, w dwóch zbiorach danych przestrzennych. Rozszerzono badania poprzez przyjęcie zbiorów z 7 powiatów. Pozwoliło to na określenie wskaźników różnic CCI, w skali lokalnej (CCI_{Lni}), dla poszczególnych 7 powiatów, i regionalnej dla różnych grup powiatów CI_{Rn} . Zastosowanie różnych podzbiorów regionalnych, do wyznaczenia wskaźników CCI, powinno wskazać w wynikach zmienność regionalną. Przy badaniu CCI różnic w dwóch zbiorach danych, uwzględniono wagi przypisane podzbiорom tematycznym (budynom, lasom, wodom, ...), związane z rozpoznawalnością obiektów, na obrazowaniach satelitarnych i zdjęciach lotniczych. Zastosowanie wielokryterialnej analizy decyzyjnej (MCDA), z zastosowaniem metody TOPSIS, stwarza możliwości wyboru preferowanych podzbiorów wybranej bazy do badań,

[4] rozwinięto badania CCI, poszerzając je o zastosowanie metody TOPSIS, z nowym wariantem wagowania. Wagi ustalano na podstawie analizy statystycznej, uwzględniając przestrzenne zmiany względne, wyznaczone analizami statystycznymi (RC_{CCI}). Przy wyznaczeniu wag, zastosowano analizę HOT SPOT. Określono wrażliwości wskaźników CCI regionalnych i lokalnych, przy zmianie wagowania. Do ustalania preferencji, zastosowano wielokryterialną analizę decyzyjną (MCDA), wykorzystującą metodę TOPSIS.

4. Ocena badań i wyników

Biorąc pod uwagę, że większość wcześniejszych analiz jakości i użyteczności OSM dotyczyła dużych miast lub aglomeracji, wyniki badań przedstawione w dysertacji są nowatorskie, ponieważ obejmują obszary całych powiatów, o różnej charakterystyce geograficznej. Analiza jakości danych obejmuje drogi i budynki, powszechnie badane i uwzględnia klasy obiektów często pomijane w badaniach: lasy, wody, sieć kolejową, wodną.

Wyniki [1] badań wskazują na zbiory o różnej zgodności w dwóch bazach, w aspekcie dokładności geometrycznej i atrybutowej. Zgodność nie dotyczy całych baz, ale określona jest w wybranych klasach obiektów powierzchniowych i liniowych. Jest różna w poszczególnych klasach i regionach. W pierwszym badaniu autorka wykazała, że nie można mówić o dokładności OSM, a o regionalnej i tematycznej dokładności. Wykazała związek jakości danych OSM, z zamożnością powiatów.

W drugiej publikacji [2], rozwinięto badania kompletności budynków poprzez wizualizację wartości wskaźników, z wykorzystaniem diagramów dwuwymiarowych w sześciokątnych podzbiорach pól podstawowych 1 km^2 . Taka wizualizacja pozwoliła na lepszą interpretację wyników, w ocenie przydatności danych.

W trzecim i czwartym badaniu [3-4], wykorzystano własne wskaźniki oceny, oparte o CCI regionalne i lokalne. Pozwoliło to na ocenę danych w powiatach, w porównaniu ze wskaźnikami regionalnymi. Zastosowanie metody TOPSIS i różnych metod wagowania klas obiektów, rozszerzyło możliwości badań i oceny jakości, z uwzględnieniem jakości pozyskania danych, czy potrzeb użytkowników.

Przedstawiona metodyka badań, z analizą wskaźników, interpretacją wyników, powinna pozwolić na podejmowanie decyzji co do wyboru i przyjęcia określonej klasy danych z jednej z wybranych dwóch baz, do szczegółowych badań, spełniając preferencje potencjalnych użytkowników.

5. Dyskusja i uwagi

Praca jest interesująca. Doktorantka wykorzystuje znane wskaźniki ISO oceny jakości danych przestrzennych i atrybutowych i tworzy własne.

Nowością w opracowaniu jest kompleksowe podejście do oceny jakości danych OSM dla głównych klas pokrycia terenu na obszarze powiatu, regionu. Nowością jest również analiza jakości danych OSM w porównaniu z dochodem budżetów powiatów, w przeliczeniu na głowę (per capita) jednego mieszkańca. przy szukaniu wskaźników zrównoważonego rozwoju między powiatami.

Wysoka zgodność geometryczna, w klasie budynków, a niska w klasie lasów, zbadana w oparciu o punkty homologiczne, nie dziwi. Przy budynkach, wybór punktów jest bardziej jednoznaczny, niż przy określaniu granicy lasów. Czy informacja statystyczna o różnicy długości werteksów, w poligonach prezentujących obiekty, np. lasów i wód, nie wniosłaby istotnego nowego parametru w ocenie jakości geometrycznej obiektów w tych klasach, z dwóch baz? Zgodność geometryczna przy analizie sieci rzecznej wyszła najniższa. Jest to też zrozumiałe. Oś koryta rzeki, przy niejednoznacznych i zmiennych granicach jej koryta, trudno jednoznacznie określić w niezależnych edycjach. Wyniki przeprowadzonych badań wskazują na interesujące wartości tych różnic.

Wyniki analizy kompletności obiektów w dwóch zbiorach, wskazują na obszary silnie zurbanizowane, o najwyższym wskaźniku TP (nakładania się obszarów), w klasie budynków, a także o nadmiarowości budynków w OSM. Podobne wyniki uzyskano dla sieci drogowej. Sieć rzeczna charakteryzuje się niższymi wskaźnikami. Dokładność semantyczna w OSM jest niższa niż BDOT_10k i widocznie różna w poszczególnych powiatach. W publikacji [1], a rysunkach 8-11, na mapach, Indeks FP >100 występuje rzadko w pojedynczych polach, w dużym rozproszeniu. Czy to działanie przypadkowych wolontariuszy, edytujących miejsca swojego zamieszkania, tak jak to czynią moi studenci w ramach dodatkowych prac domowych, mapując wiaty, drewnutnie, polne drogi dojazdowe? Czy nie warto by było, przeprowadzić badań w tym kierunku?

Czy nadmiarowość budynków i dróg w OSM, względem BDOT_10k, nie wskazuje na to, że do OSM dodano dane z uwolnionych zbiorów BDOT_10k? Doktorantka twierdzi, że w analizowanych powiatach nie dokonano importu danych z BDOT_10k. Proszę o uzasadnienie!

Wagowanie (TOPSIS), zależne rozpoznawalności obiektów na zdjęciach satelitarnych i lotniczych, zależne jest od wyrazistości konturów. Czy reguła ta nie wiąże się z jakością edycji obiektów w dwóch badanych bazach OSM i BDOTk10? W publikacji [3] budynki i lasy otrzymały tą samą wagę 0.25. Natomiast w publikacji [1], przy analizie dokładności geometrycznej, budynki mają najwyższy wskaźnik dokładności geometrycznej, a lasy i wody uzyskały najniższą dokładność geometryczną. Stąd pytanie, czy w badaniu [3], lasy nie powinny otrzymać niższą wartość wagi, np. 0.10, taką jak wody i cieki wodne?

Przedstawiona metodyka badań, z analizą wskaźników, powinna pozwolić na podejmowanie decyzji co do wyboru i przyjęcia określonej klasy danych z jednej z wybranych dwóch baz, do szczegółowych badań, działań, wspierających podejmowanie decyzji, spełniając preferencje potencjalnych użytkowników.

Uważam, że przedstawiona praca doktorska zawiera oryginalne rozwiązanie problemu naukowego, możliwe do zastosowania przy ocenie jakości innych baz. Jestem przekonana, że proponowane rozwiązanie jest także możliwe do zastosowania przy ocenie zmian w przestrzeni geograficznej, na podstawie tożsamyh baz, z różnych okresów czasowych.

Założone cele badawcze, główny i szczegółowe, zostały zrealizowane. Hipotezy badawcze zostały zweryfikowane pomyślnie. Nie zgadzam się tylko z wyrażonym poglądem, że wskaźniki ISO i proponowane przez doktorantkę, są wystarczające do całościowej oceny jakości danych przestrzennych. Przyjęcie tej tezy, nie inspiruje do dalszych badań w tym temacie.

6. Wniosek końcowy

Stwierdzam, że rozprawa doktorska mgr inż. Sylwii Borkowskiej spełnia wymogi stawiane pracom klasyfikacyjnym na stopień doktora nauk inżynieryjno-technicznych w dyscyplinie inżynieria lądowa, geodezja i transport. Praca doktorska zawiera oryginalne rozwiązanie problemu naukowego, możliwego do zastosowania w wybranych dziedzinach gospodarki związanych z podejmowaniem decyzji przestrzennych, a mgr inż. Sylwia Borkowska wykazała się ogólną wiedzą teoretyczną oraz umiejętnością samodzielnego prowadzenia pracy naukowej. Warsztat badawczy oceniam wysoko.

Podsumowując, wnioskuję do Rady Dyscypliny Inżynieria Lądowa, Geodezja i Transport Wojskowej Akademii Technicznej o przyjęcie rozprawy doktorskiej i dopuszczenie mgr inż. Sylwii Borkowskiej do publicznej obrony.

Alibita Smalbin