



Kierunek studiów: Geoinformacja

Nr albumu: 444861

Błażej Kościański

**Porównanie miar niepodobieństwa struktury
przestrzennej w kontekście określania zmian
kategorii pokrycia terenu**

*Comparison of spatial structure dissimilarity measures
in the context of determining land cover changes*

Praca magisterska napisana
w Instytucie Geoekologii i Geoinformacji
pod kierunkiem
dr. hab. Jakuba Nowosada

Poznań, 2024

Streszczenie

Abstrakt

Analiza zmian w pokryciu terenu w różnych okresach czasowych lub między różnymi zestawami danych obrazowych stanowi istotne zagadnienie w badaniach geoprzestrzennych. Istnieje szereg technik służących do analizy różnic w danych rastrowych, w tym metody oparte na analizie struktur przestrzennych. W tych metodach, stopień zróżnicowania dwóch obszarów może być oszacowany za pomocą miar odległości, podobieństwa, niepodobieństwa i innych wskaźników. Współcześnie jednak nie określono, która z tych miar najlepiej odzwierciedla ludzkie postrzeganie różnic w pokryciu terenu. Celem tej pracy było porównanie zgodności miar umożliwiających określenie niepodobieństwa struktury przestrzennej z postrzeganiem zmian kategorii pokrycia terenu przez ludzi. W tym celu przeprowadzona została ankieta w której zadaniem respondentów była ocena podobieństwa między parami rastrów. Zbiory danych rastrowych zostały przygotowane poprzez symulację parametrów kompozycji i konfiguracji przestrzennej. W analizie uwzględniono 20 miar niepodobieństwa oraz dwie miary z dziedziny teorii informacji: entropię brzegową i względną informację wzajemną. Przeprowadzono także analizę różnic między wynikami czterech wybranych miar niepodobieństwa zastosowanych do analizy zmian pokrycia terenu na podstawie danych CORINE Land Cover. W efekcie przeprowadzonych badań udało się opisać relacje między sposobem postrzegania przez ludzi różnic w pokryciu terenu a wynikami miar niepodobieństwa oraz wartościami miar opisujących kompozycję i konfigurację przestrzenną rastrów.

Słowa kluczowe: zmiany pokrycia terenu, struktura przestrzenna, miary odległości, miary niepodobieństwa, ekologia krajobrazu

Abstract

The analysis of changes in land cover over different time periods or between different image datasets is an important issue in geospatial research. There are a number of techniques for analysing differences in raster data, including pattern-based change assessment methods. In these methods, the difference between two areas can be estimated using measures of distance, similarity, dissimilarity and other indicators. However, at present, it has not been determined which of these measures best reflects human perception of differences in land cover. The aim of this work was to compare the agreement of measures that allow the determination of spatial structure dissimilarity with people's perceptions of changes in land cover categories. For this purpose, a survey was conducted in which respondents were asked to assess the similarity between pairs of rasters. The raster datasets were prepared by simulating composition and spatial configuration parameters. The analysis included 20 measures of dissimilarity and two information theory measures: marginal entropy and relative mutual information. An analysis of differences between the results of four selected dissimilarity measures used to analyze land cover changes based on CORINE Land Cover data was also conducted. As a result of the study, it was possible to describe the relationships between people's perceptions of differences in land cover and the results of dissimilarity measures and the values of measures describing the spatial composition and configuration of raster images.

Keywords: land cover changes, spatial pattern, distance measures, dissimilarity measures, landscape ecology

Spis treści

Streszczenie	3
1 Wprowadzenie	7
2 Metody	11
2.1 Struktury przestrzenne	11
2.2 Metryki krajobrazowe	12
2.3 Sygnatury przestrzenne	13
2.4 Miary wywodzące się z teorii informacji	15
2.5 Metody analiz różnic pokrycia terenu	16
2.6 Miary odległości i niepodobieństwa	17
2.7 Obliczenie niepodobieństwa rastrów	18
3 Materiały	21
3.1 Symulowanie rastrów o określonej kompozycji i konfiguracji przestrzennej .	21
3.2 Przygotowanie ankiety	24
3.3 Dane CORINE Land Cover	25
4 Wyniki ankiety	27
4.1 Podsumowanie ankiety	27
4.2 Poziom zgodności odpowiedzi ankietowanych	30
4.3 Miary niepodobieństwa	32
4.4 Odpowiedzi i miary niepodobieństwa	32
4.5 Odpowiedzi a różnice między miarami z teorii informacji	40
5 Analiza zróżnicowania wybranych miar niepodobieństwa na podstawie danych CORINE Land Cover	43
5.1 Porównanie wybranych miar niepodobieństwa w kontekście określania zmian kategorii pokrycia terenu	44
5.2 Porównanie wyników miar niepodobieństwa	45
5.3 Miary niepodobieństwa a kompozycja i konfiguracja przestrzenna	50
6 Podsumowanie	53
Bibliografia	57

Rozdział 1

Wprowadzenie

Informacje geograficzne są wynikiem selekcji i przetwarzania danych związanych z otaczającą nas przestrzenią geograficzną. Pozwalają na bardziej zrozumiałe i efektywne analizowanie, modelowanie oraz interpretowanie złożonych zjawisk i procesów zachodzących w naszym otoczeniu. Informacje geograficzne i ich aspekty nie stanowią niepodważalnych faktów, lecz często powstają w wyniku działań jednostek, jak i wspólnych wysiłków grup ekspertów, którzy zajmują się wyborem, analizą i klasyfikacją danych geograficznych (Comber i in., 2005). W procesie tworzenia informacji geograficznych istnieje zatem pewien stopień subiektywności, który może wpłynąć na ostateczną postać tych informacji, ich interpretację, jak i na ich użyteczność w kontekście innych zastosowań. Przykładem informacji geograficznej, której ostateczna postać zależy od założeń przyjętych w trakcie tworzenia danych przestrzennych, jest pokrycie terenu.

Przyjmuje się, że termin pokrycie terenu obejmuje zbiór wszelkich elementów obecnych na powierzchni Ziemi (Fisher i in., 2005). W elementy pokrycia terenu włączają się obiekty związane z działalnością człowieka, skutkami sił przyrody oraz wszelkie inne istniejące obiekty, które mogą znaleźć się w przestrzeni geograficznej (Zwolinski i in., 2018). Tworzenie dokładnych i wiarygodnych danych dotyczących pokrycia terenu jest niezbędne w kontekście wielu zastosowań, takich jak planowanie przestrzenne (Bibby i in., 1999), ochrona środowiska (Mallinis i in., 2011), czy analiza zmian klimatycznych (Drašković i in., 2020). Ostateczna forma tych danych jest jednak w dużej mierze

determinowana przez wybory i założenia dokonywane w procesie ich tworzenia. W tym kontekście, analiza pokrycia terenu staje się istotnym polem badań, które skupia się na zarówno na technicznych aspektach zbierania danych, jak i na ich semantycznej interpretacji.

Dane oraz wynikowe mapy pokrycia terenu są rezultatem złożonego procesu przetwarzania i analizy informacji przestrzennych najczęściej w postaci obrazów satelitarnych (Lu i in., 2004; Jasiewicz i in., 2015). Na początku tego procesu, satelity wyposażone w sensory rejestrują obrazy Ziemi z różnych zakresów widmowych. Dane pochodzące z teledetekcji często mają postać danych rastrowych, gdzie informacje przestrzenne są zapisane w formie regularnej siatki komórek lub punktów (Głazewski, 2006). W modelu danych rastrowych każda z tych komórek lub punktów przechowuje jedną wartość, która stanowi reprezentację jakiejś charakterystyki danego fragmentu powierzchni Ziemi. Obrazy uzyskane w procesie teledetekcji mogą być interpretowane manualnie przez grupy specjalistów. Pozwala to na uzyskanie map pokrycia terenu o wysokiej dokładności, kosztem długiego procesu ich tworzenia (Cunningham, 2006). Dużo mniej czasochłonną metodą jest przetwarzanie danych teledetekcyjnych przy użyciu algorytmów. Umożliwiają one względnie szybką, półautomatyczną identyfikację i klasyfikację różnych typów powierzchni kosztem mniejszej dokładności mapy wynikowej (Cunningham, 2006). Ostatecznie, dane przekształcone w mapy pokrycia terenu mogą posłużyć do analiz zmian pokrycia terenu (Feranec i in., 2007; Sleeter i in., 2013; Mierzwiak i in., 2019).

Celem analiz zmian pokrycia terenu jest przede wszystkim monitorowanie i pogłębienie aktualnej wiedzy na temat ewolucji otaczającego nas krajobrazu (Lu i in., 2004). Jest to istotne w kontekście ochrony przyrody (Anderson i in., 2020), planowania przestrzennego (Enoguanbhor i in., 2019), oceny wpływu inwestycji i infrastruktury na środowisko (Pavlovic i in., 2022), a także w badaniach dotyczących zmian klimatycznych (Kepski, 2018), bioróżnorodności (Falcucci i in., 2006) oraz innych procesów ekologicznych (Lu i in., 2004). Dzięki analizie zmian pokrycia terenu można identyfikować obszary zagrożone degradacją, monitorować skutki urbanizacji, deforestacji czy erozji, co umożliwia podejmowanie odpowiednich działań w celu zrównoważonego zarządzania środowiskiem i zachowaniem jego integralności.

W badaniach nad zmianami pokrycia terenu wykorzystuje się różnorodne metody analityczne. Niemniej jednak, wiele z tych technik koncentruje się na analizie zmian na poziomie indywidualnych komórek w siatce rastra (Lu i in., 2004). Choć podejście to może dostarczać użytecznych informacji dotyczących trendów zmian pokrycia terenu, charakteryzuje się ono istotnymi ograniczeniami w kontekście interpretacji wyników. Szczególnie w przypadku badań obejmujących rozległe terytoria, takie jak kraje czy nawet kontynenty, bardziej efektywne staje się zastosowanie metod opartych na analizie struktur przestrzennych (Netzel i in., 2015). Głównym założeniem tych metod jest przekształcenie danych z postaci pojedynczych wartości komórek rastra w sygnatury przestrzenne, a następnie porównanie ich ze sobą za pomocą miar odległości i niepodobieństwa (Nowosad, 2021).

Sygnatury przestrzenne stanowią statystyczny opis struktur przestrzennych kategorii pokrycia terenu na mniejszych, wydzielonych obszarach w obrębie całego zbioru danych (Jasiewicz i in., 2015). W celu porównania ze sobą dwóch sygnatur przestrzennych, wykorzystywane są miary niepodobieństwa. Umożliwiają one określenie w jakim stopniu dwa analizowane obszary się od siebie różnią pod względem kompozycji oraz konfiguracji przestrzennej. Opracowane zostało wiele różnych miar niepodobieństwa, takich jak odległość euklidesowa, odległość Canberra, metryka Wave Hedgesa, współczynnik podobieństwa Jaccarda, dywergencja Jensena-Shannona czy dywergencja Pearsona (Cha, 2007). Współcześnie jednak nie określono, która z tych miar jest najbardziej zgodna zarówno z postrzeganiem przez człowieka, jak i wpływem zmian na procesy środowiskowe.

Celem tej pracy było porównanie zgodności miar umożliwiających określenie niepodobieństwa struktury przestrzennej z postrzeganiem zmian kategorii pokrycia terenu przez ludzi. Pierwszym etapem pracy było stworzenie zbioru rastrów poprzez proces symulacji. Założeniem przy tworzeniu rastrów było uwzględnienie szerokiego spektrum wartości kompozycji oraz konfiguracji przestrzennej. Na podstawie tych danych przeprowadzono ankietę, w której zadaniem respondentów było określenie stopnia podobieństwa między parami rastrów. Badanie przeprowadzono na rastrach składających się wyłącznie z dwóch

lub trzech kategorii. Na podstawie wyników ankiety określono sposób agregacji odpowiedzi, będący najbardziej wiarygodną reprezentacją postrzegania zmian przestrzennych przez ludzi. W oparciu o najlepszy sposób agregacji odpowiedzi, wyniki ankiety zestawiono z wartościami 20 miar niepodobieństwa oraz dwóch miar z teorii informacji: entropii i względnej informacji wzajemnej. Kolejnym etapem badania było wskazanie różnic w wynikach wybranych miar niepodobieństwa na podstawie rzeczywistych danych o pokryciu terenu. W tym celu z danych o pokryciu terenu CORINE Land Cover (CLC) dla lat 1990 i 2018 wydzielono regularną siatkę obszarów o wymiarach 10 na 10 km. Dla tych obszarów obliczono wartości niepodobieństwa dwóch miar charakteryzujących się największą zgodnością z ludzką percepcją zmian przestrzennych oraz dwóch najczęściej stosowanych: odległości euklidesowej i dywergencji Jensena-Shannona. Następnie wyselekcjonowano i opisano obszary, dla których miary wykazywały się największym zróżnicowaniem, co pozwoliło na scharakteryzowanie kluczowych różnic wynikających z zastosowania każdej z miar.

Rozdział 2

Metody

W tym rozdziale opisane są kolejno wszystkie zagadnienia kluczowe do zrozumienia tematyki określania zróżnicowania struktur przestrzennych w kontekście zmian pokrycia terenu. Pierwszym omawianym zagadnieniem są struktury przestrzenne oraz koncepcje kompozycji oraz konfiguracji przestrzennej. Następnie omawiane są wskaźniki umożliwiające określanie charakterystyk struktur przestrzennych, czyli metryki krajobrazowe oraz sygnatury przestrzenne. Opisana jest także idea reprezentacji rastrów w postaci macierzy i wektorów współwystępowania. W następnej kolejności przedstawione są dwie metody analiz zmian pokrycia terenu. Pierwsza opiera się na analizie różnic na poziomie indywidualnych komórek w siatce rastra, a druga na analizie struktur przestrzennych występujących wewnątrz rastra. Wyjaśnione są także zagadnienia miar odległości i niepodobieństwa oraz ich wykorzystania w analizach przestrzennych. Przedstawiany jest także proces obliczenia niepodobieństwa między parami rastrów.

2.1 Struktury przestrzenne

McGarigal i in. (2009) określa, że znaczna część dziedziny ekologii krajobrazu opiera się na paradygmacie płatów. Według tej idei, każdy krajobraz zbudowany jest z jednostek, nazywanych płatami. Płaty definiuje się jako wyodrębnione obszary, wyróżniające się od sąsiadujących elementów na podstawie różnych cech, takich jak wielkość, kształt, charakter granic, różnorodność czy kategoria pokrycia terenu (Forman, 1995; Solon,

2002; German, 2004). Każdy krajobraz, składający się z wielu płatów, natomiast cechuje się pewną strukturą przestrzenną, której najbardziej podstawowymi charakterystykami są kompozycja i konfiguracja przestrzenna (Gustafson, 1998).

Kompozycja rastra opisuje zróżnicowanie i liczbę płatów poszczególnych kategorii pokrycia terenu bez uwzględniania informacji o ich lokalizacji w przestrzeni (Gustafson, 1998; Solon, 2002; Kozak i in., 2014). Konfiguracja (ułożenie) natomiast opisuje sąsiedownictwo ze sobą poszczególnych płatów (Gustafson, 1998; Solon, 2002; Kozak i in., 2014). Kompozycja, konfiguracja, jak i inne różne cechy mogą następnie być opisywane na przykład za pomocą metryk krajobrazowych lub sygnatur przestrzennych.

2.2 Metryki krajobrazowe

Metryki krajobrazowe to ilościowe charakterystyki właściwości przestrzennych krajobrazów (McGarigal i in., 2023). Pozwalają na skwantyfikowanie różnorodności, kształtu, rozmieszczenia i innych cech na różnych poziomach hierarchii przestrzennej. Umożliwiają one także analizę struktury przestrzennej krajobrazu na podstawie danych o pokryciu terenu (Pukowiec-Kurda i in., 2016). Wyróżnia się trzy główne typy metryk krajobrazowych, z których każdy koncentruje się na innym szczeblu organizacji przestrzennej krajobrazu: na poziomie płatów, klas lub całych krajobrazów (McGarigal i in., 2023).

Metryki na poziomie płatów koncentrują się na indywidualnych obszarach krajobrazu, określanych jako płaty. Pozwalają na dokładniejsze zrozumienie charakterystyk poszczególnych obszarów w krajobrazie. Przykładem metryk na tym poziomie są powierzchnia płatu (ang. patch area) i długość granicy płatu (ang. patch perimeter).

Metryki na poziomie klas opisują charakterystyki płatów o wspólnych właściwościach. Pozwalają zrozumieć rozmieszczenie, zróżnicowanie i relacje między obszarami o podobnych cechach. Do przykładowych metryk na tym poziomie należą: liczba płatów (ang. number of patches) oraz średnia odległość do najbliższego płatu tej samej klasy (ang. mean nearest neighbor).

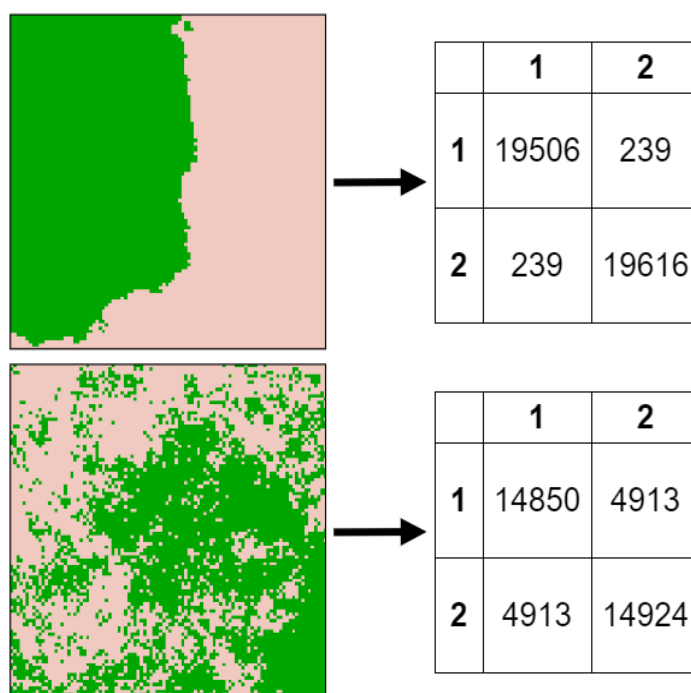
Metryki na poziomie krajobrazów skupiają się na aspektach przestrzennych i strukturalnych całego krajobrazu. Uwzględniają wszystkie płyty znajdujące się w obrębie analizowanego obszaru. Opracowano wiele różnych metryk na poziomie krajobrazu, których celem jest opis różnych aspektów struktury przestrzennej rastrów kategoryzowanych (McGarigal i in., 2023). Można je podzielić na dwie podstawowe grupy: wskaźniki kompozycji, które opisują zróżnicowanie i liczbę płatów danego typu, oraz wskaźniki konfiguracji przestrzennej, które opisują sposób rozmieszczenia i sąsiedowności płatów (Solon, 2002; Kozak i in., 2014). Do pierwszej grupy należą między innymi indeks różnorodności Shannona (ang. Shannon's diversity index) oraz liczba typów pokrycia terenu (ang. patch richness), natomiast do drugiej grupy przykładami są wskaźnik agregacji płatów (ang. aggregation index) i wskaźnik przestrzennej spójności płatów (ang. patch cohesion index).

Szczególną zaletą metryk krajobrazowych jest to, że mogą one być obliczone dla różnorodnych jednostek przestrzennych, które mogą być zdefiniowane na podstawie aspektów administracyjnych, geograficznych, biogeograficznych lub umownych (Pukowiec-Kurda i in., 2016). Jednostki te mogą obejmować zarówno granice gmin, zlewni, ekoregionów czy nawet abstrakcyjne obszary na mapie, co z kolei umożliwia wszechstronne stosowanie tych metryk w badaniach związanych z analizą krajobrazu.

2.3 Sygnatury przestrzenne

Często w celu lepszej reprezentacji analizowanych rastrów możemy wykorzystywać także bardziej skomplikowane sygnatury przestrzenne. Sygnatura przestrzenna to statystyczny opis pewnych struktur przestrzennych występujących wewnątrz rastra (Jasiewicz i in., 2015; Nowosad, 2021). Mogą one być dwuwymiarową reprezentacją kompozycji i konfiguracji przestrzennej rastra, czyli jego najbardziej podstawowych charakterystyk.

Przykładem sygnatury łączącej zarówno kompozycję, jak i konfigurację przestrzenną jest macierz współwystępowania (ang. co-occurrence matrix). Jest to macierz o wymiarach k na k , gdzie k reprezentuje liczbę kategorii pokrycia terenu obecnych w analizowanym rastrze (Haralick i in., 1973; Jasiewicz i in., 2015). Macierz tę możemy skonstruować poprzez zliczanie kolejno wszystkich par sąsiadujących ze sobą komórek w rastrze.



Rycina 2.1: Przykład dwóch macierzy współwystępowania dla rastrów o zbliżonej kompozycji i różnej konfiguracji przestrzennej

Wewnątrz tej macierzy informacje o kompozycji rastra otrzymujemy poprzez zliczenie wartości w kolumnach lub wierszach (proporcje każdej kategorii), natomiast relacja przekątnej do pozostałych wartości w macierzy informuje nas o konfiguracji przestrzennej. Przykład dwóch macierzy współwystępowania dla rastrów o zbliżonej kompozycji, ale różnej konfiguracji przestrzennej przedstawia Rycina 2.1.

W celu porównywania ze sobą sygnatur dwóch rastrów w postaci dwuwymiarowej macierzy należy je sprowadzić do postaci jednowymiarowych wektorów (histogramów), a następnie przeprowadzić ich normalizację, tak aby wszystkie wartości sumowały się do 1. Taka postać umożliwia obliczanie miar odległości lub podobieństwa, pozwalających na porównywanie histogramów wartości (Cha, 2007). Miary te następnie pozwalają określić stopień odmienności dwóch rastrów. Podejście to może być także wykorzystane w innych analizach przestrzennych, jak wyszukiwanie obszarów o podobnej strukturze przestrzennej, grupowanie tych obszarów oraz wykrywanie ich zmian (Jasiewicz i in., 2015; Nowosad, 2021).

2.4 Miary wywodzące się z teorii informacji

Nowosad i in. (2019) określa, że podstawowymi jednostkami analizy dla miar wywodzących się z teorii informacji nie są poszczególne komórki rastra, lecz pary sąsiadujących komórek. Parą natomiast określa się dwuwartościową zmienną losową (x, y) , dla której x reprezentuje klasę analizowanej komórki, a y jest klasą sąsiedniej komórki. Zmienna (x, y) przyjmuje wartości (c_i, c_j) , które mieszczą się w zakresie od 1 do K , gdzie K reprezentuje liczbę klas obecnych na rastrze. Na tej podstawie, macierz współwystępowania może być traktowana jako dwuwymiarowy histogram par komórek, w którym każdy przedział wskazuje liczbę par (c_i, c_j) .

W tej pracy wykorzystane zostały dwie miary z teorii informacji. Pierwszą jest entropia brzegowa (ang. marginal entropy). Opisuje ona zróżnicowanie kompozycji rastra, czyli udziałów każdej z kategorii w rastrze. Im wyższa wartość tej miary, tym bardziej zróżnicowane są poszczególne klasy. Entropia brzegowa może być obliczona zgodnie ze wzorem:

$$H(y) = - \sum_{j=1}^K p(y = c_j) \log_2 p(y = c_j)$$

- K - liczba kategorii pokrycia terenu
- $p(y = c_j)$ - prawdopodobieństwo, że sąsiadująca komórka y ma przypisaną klasę c_j

Drugą miarą, która została wykorzystana w tej pracy jest względna informacja wzajemna (ang. relative mutual information). Opisuje ona stopień sąsiadowania ze sobą kategorii w rastrze, czyli jego konfigurację przestrzenną. Względna informacja wzajemna przyjmuje wartości w zakresie od 0 do 1. Im niższa wartość tej miary, tym bardziej pofragmentowany jest cały krajobraz. Względną informację wzajemną można obliczyć ze wzoru:

$$U = I(y, x) / H(y)$$

gdzie $I(y, x)$ oznacza informację wzajemną, liczoną ze wzoru:

$$I(y, x) = H(y) - H(y|x)$$

natomiast $H(y|x)$ reprezentuje entropię warunkową, obliczaną zgodnie ze wzorem:

$$H(y|x) = \sum_{i=1}^K \sum_{j=1}^K p(x = c_i, y = c_j) \log_2 p(y = c_j | x = c_i)$$

- $p(x = c_i, y = c_j)$ - prawdopodobieństwo, że analizowana komórka x jest klasy c_i oraz sąsiadująca komórka y jest klasy c_j
- $p(y = c_j | x = c_i)$ - prawdopodobieństwo, że komórka y jest klasy c_j pod warunkiem, że analizowana komórka jest klasy c_i

2.5 Metody analiz różnic pokrycia terenu

Zmiany pokrycia terenu w czasie lub różnice w pokryciu terenu pomiędzy obrazami można analizować przy użyciu wielu metod. Wiele z nich koncentruje się na analizie różnic na poziomie indywidualnych komórek w siatce rastra. Najbardziej podstawowym przykładem takiego podejścia jest analiza ilościowa różnic w pokryciu terenu. Zaletą tego podejścia jest jego prostota: wystarczy zliczyć wszystkie komórki należące do poszczególnych kategorii dla wybranych rastrów, a następnie porównać ze sobą te wartości, aby otrzymać wynik informujący nas o ilościowych różnicach między analizowanymi rastrami. Analiza ilościowa najczęściej wykorzystywana jest w celu wskazania ogólnych trendów zmian pokrycia terenu dla określonego obszaru badań jak na przykład zmniejszanie się obszarów leśnych lub wzrost terenów zurbanizowanych.

Wszelkie metody analiz zmian pokrycia terenu opierające się na analizie poszczególnych komórek w siatce rastra są użyteczne na obszarach, gdzie zmiany między indywidualnymi komórkami dostarczają istotnych informacji. Ich przydatność jednak maleje, gdy informacja na poziomie pojedynczej komórki przestaje być tak istotna, na przykład dla rastrów o wysokiej rozdzielczości lub znacznym zasięgu przestrzennym (Jasiewicz i in., 2015). W takiej sytuacji bardziej efektywne staje się zastosowanie metod opartych na

analizie struktur przestrzennych (ang. pattern-based change assessment) (Netzel i in., 2015).

Pozwalają one przede wszystkim na opis oraz obliczenie podobieństwa struktur przestrzennych. Głównym zamysłem tych metod jest przekształcenie danych z postaci dużych rastrów zbudowanych z wielu indywidualnych komórek zawierających pojedyncze informacje w metryki krajobrazowe oraz sygnatury przestrzenne, a następnie porównanie ich za pomocą miar odległości lub niepodobieństwa. Zastosowanie metryk krajobrazowych w kontekście kompleksowych analiz przestrzennych ma jednak istotną wadę. Jako że pojedyncza metryka krajobrazowa reprezentuje wyłącznie jedną, konkretną cechę analizowanego obszaru, to nie jest w stanie opisać całej charakterystyki struktury przestrzennej danego rastra. W tym celu korzystniejsze może okazać się zastosowanie sygnatur przestrzennych. Dzięki temu, że są dwuwymiarową reprezentacją struktury przestrzennej rastrów, można je porównywać ze sobą przy użyciu szerokiej gamy istniejących miar odległości i niepodobieństwa.

2.6 Miary odległości i niepodobieństwa

Odległość oraz rozbieżność (inaczej niepodobieństwo, dywergencja) stanowią pewien policzalny stopień różnorodności pary obiektów. Największą różnicą między nimi jest to, że odległości są symetryczne, podczas gdy rozbieżności są niesymetryczne. Oznacza to, że wyłącznie dla miar odległości otrzymujemy identyczny wynik przy porównywaniu par obiektów A i B, jak i par B i A.

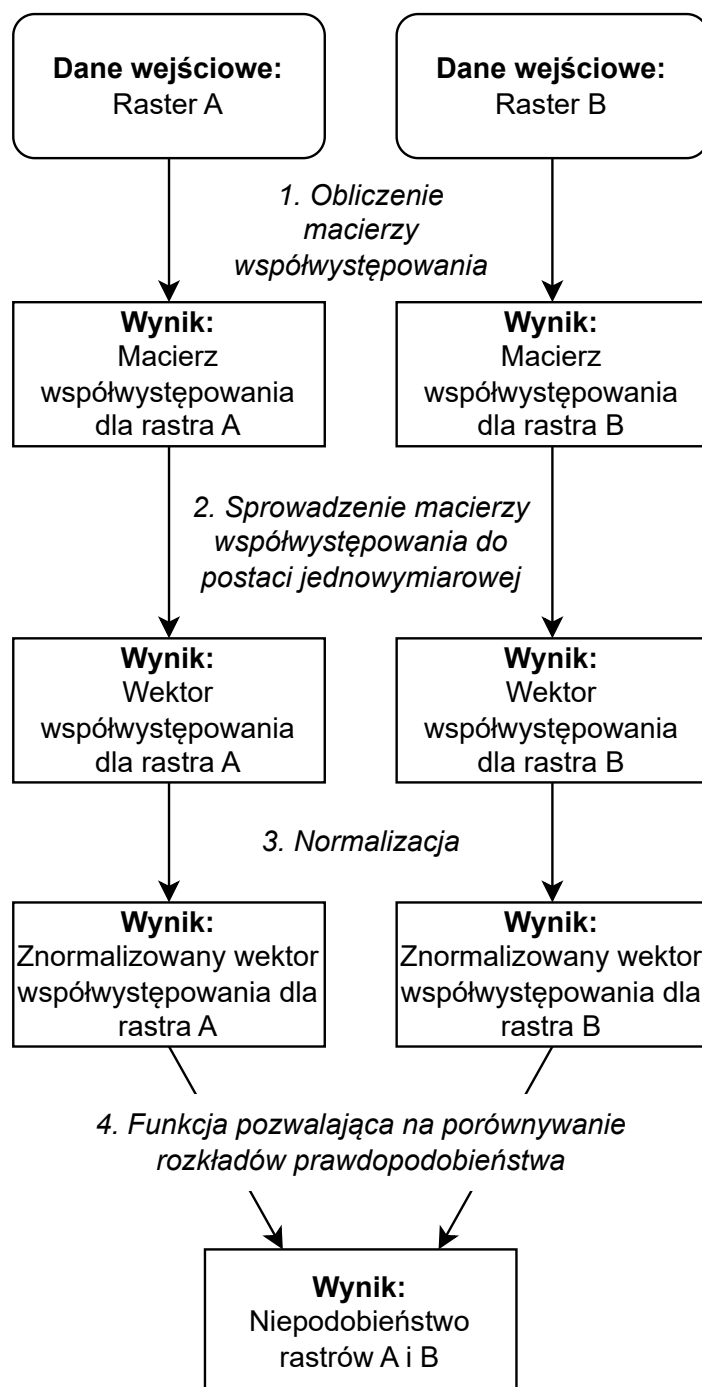
Rozbieżność jest przeciwieństwem podobieństwa. Ponadto, miary podobieństwa można łatwo przekształcić w miary niepodobieństwa (Niesterowicz i in., 2016a). W związku z tym, w celu uproszczenia terminologii, wszystkie miary odległości, podobieństwa oraz te wywodzące się z dziedziny teorii informacji, które zostały wykorzystane w tej pracy, będą dalej zbiorowo nazywane miarami niepodobieństwa.

Na podstawie podobieństw syntaktycznych, wyróżnia się kilka grup rodzin miar niepodobieństwa (Cha, 2007): rodzina Minkowski (odległość euklidesowa, odległość Minkowskiego, odległość Manhattan), rodzina L_1 (odległość Canberra, Sorensen, Kulczyński),

rodzina Intersection (Intersection, Wave Hedges, Ruzicka), rodzina Inner Product (Jaccard, średnia harmoniczna), rodzina Squared-chord (Fidelity, Matusita), rodzina χ^2 (Clark, Pearson χ^2 , Neyman χ^2), rodzina Shannon's Entropy (Jensen-Shannon, Kullback-Leibler), a także miary będące połączeniem innych miar (Taneja, Kumar-Johnson) oraz miary wywodzące się z teorii informacji (informacja wzajemna, entropia Shannona). Wybór odpowiedniej miary niepodobieństwa zależy między innymi od rodzaju pomiaru lub sposobu reprezentacji obiektów (Cha, 2007).

2.7 Obliczenie niepodobieństwa rastrów

Pierwszym krokiem, jaki należy podjąć w celu obliczenia podobieństwa struktur przestrzennych danych rastrowych z wykorzystaniem metod opartych o sygnatury przestrzenne jest sprowadzenie rastrów wejściowych do postaci macierzy współwystępowania. Proces utworzenia macierzy współwystępowania polega na zliczeniu wartości każdej indywidualnej komórki rastra, a także przylegających do niej komórek (najczęściej czterech lub ośmiu). Przykładowe macierze współwystępowania widoczne są na Rycinie 2.1. Następnie, dwuwymiarową macierz należy sprowadzić do postaci jednowymiarowej, czyli wektora współwystępowania. Kolejnym etapem analizy jest normalizacja wektora współwystępowania. Po dodaniu do siebie wszystkich wartości tego wektora powinniśmy otrzymać wynik równy 1. Po wykonaniu powyższych czynności otrzymujemy reprezentację rastrów wejściowych, która umożliwia porównanie ich ze sobą przy użyciu miar niepodobieństwa. Proces obliczenia niepodobieństwa dwóch rastrów w postaci schematu przedstawia Rycina 2.2.



Rycina 2.2: Schemat procesu obliczenia niepodobieństwa dwóch rastrów

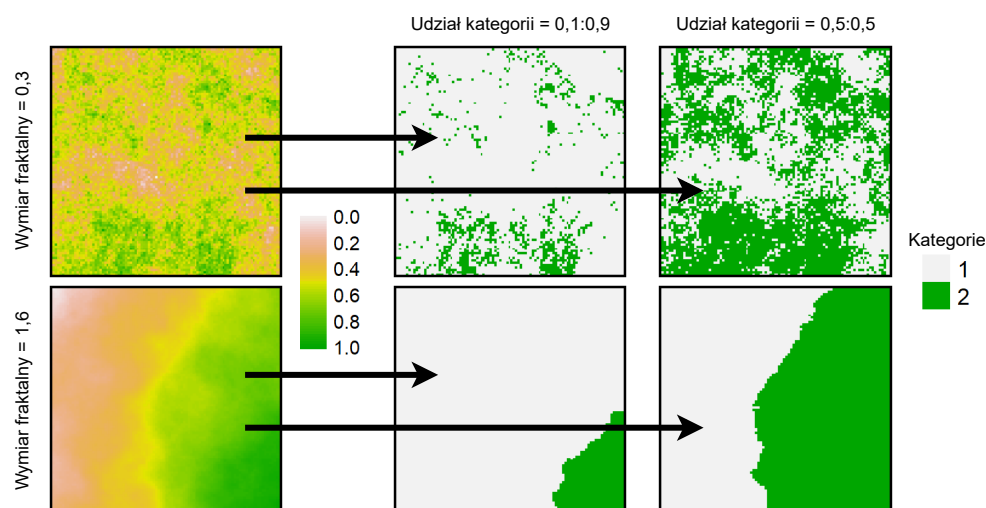
Rozdział 3

Materiały

W pierwszej kolejności w tym rozdziale są omawiane aspekty związane z formą przeprowadzonej ankiety. Przedstawiony jest proces symulacji, który został wykorzystany do stworzenia zbioru danych o określonych parametrach kompozycji i konfiguracji przestrzennej. Następnie opisany jest sposób doboru pytań do ankiety. W końcowej części rozdziału są omówione dane przestrzenne o pokryciu terenu CORINE Land Cover, które zostały wykorzystane do porównania stosowania wybranych miar niepodobieństwa w kontekście analiz zmian pokrycia terenu.

3.1 Symulowanie rastrów o określonej kompozycji i konfiguracji przestrzennej

Najważniejszym założeniem przy tworzeniu zbioru rastrów do ankiety było przygotowanie ich w sposób umożliwiający uzyskanie pełnej reprezentacji wszystkich możliwych wartości kompozycji, jak i konfiguracji przestrzennej. Zbiór rastrów został przygotowany w języku programowania R (R Core Team, 2023), w oparciu o wykorzystanie funkcji *nlm_fbm* z pakietu NLMR (Sciaini i in., 2018). Powyższa funkcja umożliwia uzyskanie danych rastrowych wypełnionych wartościami zmiennoprzecinkowymi mieszczącymi się w zakresie od 0 do 1 oraz określonej konfiguracji przestrzennej. Funkcja ta pozwala na symulację rastrów przy użyciu ułamkowych ruchów Browna, będących uproszczeniem ruchów Browna (Schlather i in., 2015). W tej funkcji poziom autokorelacji między kolejnymi



Rycina 3.1: Symulowanie rastrów o określonej kompozycji i konfiguracji przestrzennej

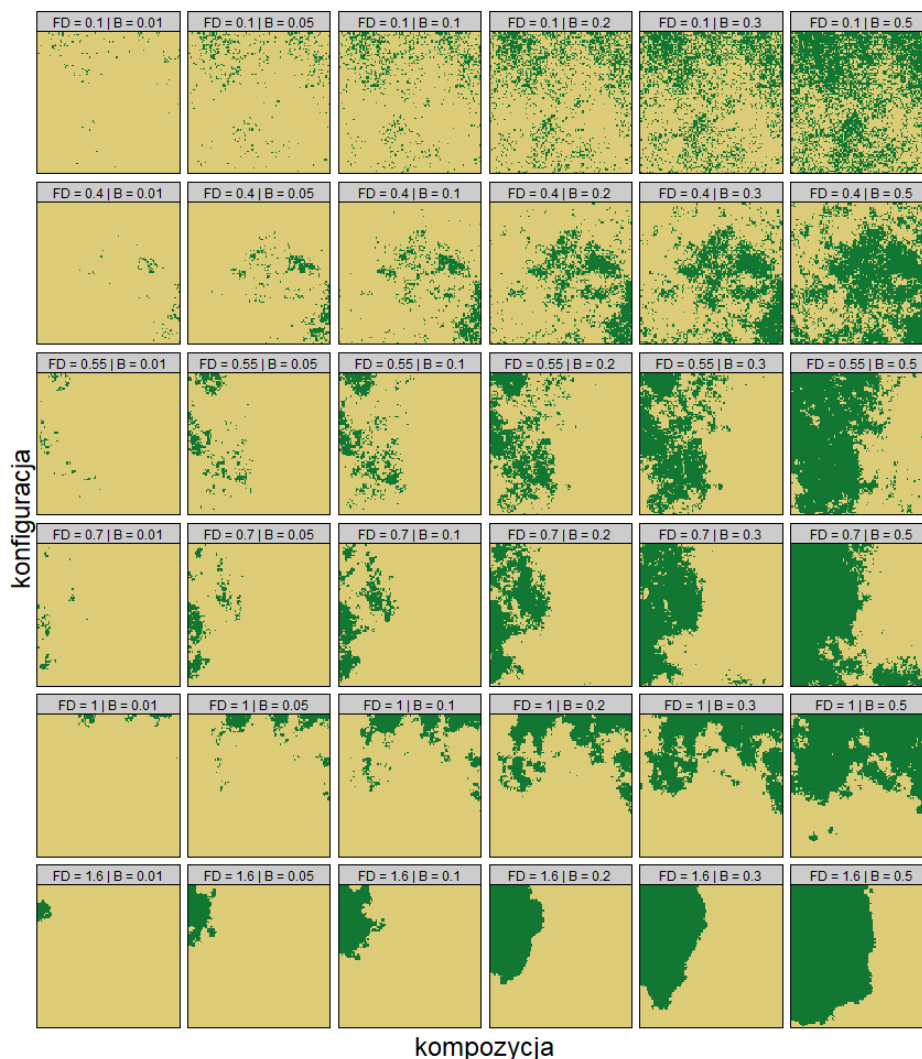
symulacjami jest kontrolowany za pomocą parametru wymiaru fraktalnego ("fract_dim"). W kontekście tego badania, parametr ten reguluje konfigurację przestrzenną. Oznacza to, że w przypadku, gdy "fract_dim" przyjmuje niską wartość, zbliżoną do 0, wartości w generowanym rastrze rozmieszczone są w sposób losowy, zbliżony do szumu. Natomiast w przypadku wysokiej wartości "fract_dim", zbliżonej do 2, na wynikowym rastrze tworzą się skupiska najwyższych i najniższych wartości, a przejścia pomiędzy nimi mają płynny, wygładzony charakter.

Następnie, aby otrzymać zbiór rastrów uwzględniający także pełen przekrój kompozycji należy przeprowadzić proces rekasyfikacji rastrów stworzonych w poprzednim kroku. Procedura ta polega na podziale każdego dotychczas utworzonego rastra na kategorie w różnych proporcjach, na przykład 90:10, 70:30 oraz 50:50, w przypadku rastrów zawierających wyłącznie dwie kategorie. Wykonanie tego procesu ułatwia funkcja *util_binarize* z pakietu *landscapetools* (Sciaini i in., 2018). W tej funkcji proporcje kategorii kontrolowane są za pomocą parametru "breaks". Przykładowo, ustawienie parametru "breaks" na poziomie 0.2 skutkuje otrzymaniem rastra o kategoriach w proporcjach 20 do 80. Oznacza to, że jedna z kategorii będzie pokrywała 20% komórek rastra, podczas gdy druga kategoria wypełni pozostałe 80% komórek. Na Rycinie 3.1 przedstawiony został przykład przygotowania rastrów podzielonych na dwie kategorie w sposób opisany powyżej.

Ostatecznie wygenerowane zostały zbiory rastrów składających się wyłącznie z dwóch lub trzech kategorii pokrycia terenu. Przykład jednego ze zbiorów rastrów przedstawia Rycina 3.2. Rastry zawierające trzy kategorie zostały uwzględnione w badaniu w celu próby wskazania czy liczba kategorii na rastrach ma wpływ na odpowiedzi udzielane przez ankietowanych.

Po utworzeniu zbioru danych rastrowych bardzo istotne było potwierdzenie, że obejmuje on pełen zakres rozkładu kompozycji i konfiguracji. Na potwierdzenie tego założenia pozwoliło obliczenie wybranych miar opisujących struktury przestrzenne: entropii oraz względnej informacji wzajemnej.

W rezultacie uzyskano dwa zestawy rastrów, które cechowały się znacznym zróżnicowaniem w zakresie kompozycji i konfiguracji przestrzennej. Pierwszy zestaw składał się z 36 rastrów, z których każdy zawierał dwie klasy (Rycina 3.2). Minimalna wartość entropii w tym zestawie wyniosła 0,077, podczas gdy maksymalna osiągnęła wartość 1. Minimalna wartość względnej informacji wzajemnej wyniosła 0,015, natomiast maksymalna wyniosła 0,912. Drugi zbiór obejmował 54 rastry zawierające trzy klasy. Minimalna wartość entropii w tym zbiorze wyniosła 0,157, a maksymalna osiągnęła wartość 1,585. Minimalna wartość względnej informacji wzajemnej w tym zbiorze wyniosła 0,014, z kolei maksymalna wyniosła 0,904.



Rycina 3.2: Przykład zbioru wygenerowanych rastrów (2 kategorie). *FD* - parametr wymiaru fraktalnego "fract_dim", *B* - parametr "breaks" kontrolujący proporcje kategorii

3.2 Przygotowanie ankiety

Ostatnim krokiem przygotowania danych do ankiety było wybranie par rastrów tworzących poszczególne pytania, tak aby uwzględniały one wszelkie możliwe różnice struktury przestrzennej rastrów. W tym celu, pytania w ankietach podzielone zostały na dwie grupy, wewnątrz których znalazły się po trzy podgrupy pytań.

W pierwszej kolejności respondenci zetknęli się z 24 pytaniami dotyczącymi rastrów uwzględniających wyłącznie dwie kategorie pokrycia terenu, a następnie z 24 pytaniami uwzględniającymi trzy kategorie pokrycia terenu. Pierwsza podgrupa pytań (6 par

rastrów) składała się z par rastrów różniących się między sobą wyłącznie entropią. Podgrupa druga (6 par rastrów) zawierała wyłącznie rastry różniące się względną informacją wzajemną. Ostatnia podgrupa (12 par rastrów) składała się z pytań zróżnicowanych zarówno pod względem entropii, jak i względnej informacji wzajemnej.

Taki sposób doboru pytań pozwolił na zredukowanie liczby odpowiedzi wymaganych od respondentów, jak i ograniczenie wpływu błędu selekcji, który powstałby w wyniku niewłaściwego doboru pytań. Respondenci celowo nie zostali poinformowani o występujących różnicach pomiędzy kolejnymi pytaniami, ponieważ mogłoby mieć to wpływ na udzielane przez nich odpowiedzi, co z kolei mogłoby wpłynąć na ostateczne wyniki badania.

3.3 Dane CORINE Land Cover

Jednym z dalszych etapów pracy było zastosowanie uzyskanych wcześniej wyników do analizy zmian pokrycia terenu na podstawie danych CORINE Land Cover (CLC). Dane CLC reprezentują szczegółowe informacje o pokryciu terenu, klasyfikując obszary według różnych kategorii, takich jak lasy, tereny rolnicze, obszary zurbanizowane czy zbiorniki wodne. Stanowią one istotny instrument w analizie i monitorowaniu zmian środowiska, a także służą jako narzędzie wspierające procesy decyzyjne na poziomie europejskim.

Zbiory danych przestrzennych CORINE Land Cover stanowią integralną część programu CORINE (Coordination of Information on Environment), wprowadzonego przez Komisję Wspólnot Europejskich w 1985 roku. Program ten został stworzony w celu skoordynowania przedsięwzięć związanych z gromadzeniem i przetwarzaniem informacji na temat stanu środowiska geograficznego w krajach należących do Wspólnoty Europejskiej oraz standaryzację tych danych w celu ułatwienia wymiany informacji między państwami członkowskimi (Bielecka i in., [2004](#)).

Wyniki programu CORINE są udostępniane w formatach wektorowych ESRI i SQLite geodatabase oraz formacie rastrowym GeoTiff o rozdzielczości przestrzennej 100 metrów, co oznacza, że jedna komórka rastra obejmuje 1 hektar powierzchni. Do celów tej pracy, wykorzystane zostały dane CLC dla lat 1990 i 2018 udostępnione do pobrania w formacie

GeoTiff na witrynie Copernicus Land Monitoring Service (<https://land.copernicus.eu/pan-european/corine-land-cover>). Dane te wykorzystują układ współrzędnych ETRS-LAEA (EPSG:3035). Zasięg obszaru badań został ograniczony do terytorium Polski, włącznie z morskimi wodami wewnętrznymi i morzem terytorialnym.

Dane CLC są zorganizowane na trzech poziomach szczegółowości. Na najwyższym poziomie tej hierarchii wyodrębniono pięć głównych typów pokrycia terenu: tereny antropogeniczne, tereny rolne, lasy i ekosystemy seminaturalne, obszary podmokłe oraz obszary wodne. W celu uproszczenia analizy wyników, dane zostały poddane reklasyfikacji. W wyniku tego procesu zredukowano liczbę kategorii pokrycia terenu do wyłącznie dwóch: lasy i pozostałe obszary.

Rozdział 4

Wyniki ankiety

Niniejszy rozdział przedstawia podsumowanie oraz analizę wyników przeprowadzonej ankiety. W pierwszej kolejności omówione są ogólne informacje dotyczące samej ankiety, takie jak termin jej przeprowadzenia, sposób realizacji badania oraz profil respondentów. Przedstawione jest podsumowanie wyników ankiety, obejmujące między innymi liczbę udzielonych odpowiedzi poszczególnych typów oraz wskazanie poziomów zgodności podgrup pytań. Następnie opisana jest analiza relacji między miarami niepodobieństwa a różnymi sposobami podsumowania odpowiedzi. Dalsza analiza koncentruje się na badaniu związku między średnią odpowiedzią a miarami niepodobieństwa, w podziale na grupy pytań uwzględniające dwie oraz trzy kategorie pokrycia terenu. Końcowa część rozdziału skupia się na analizie relacji między odpowiedziami a różnicami wartości miar z teorii informacji.

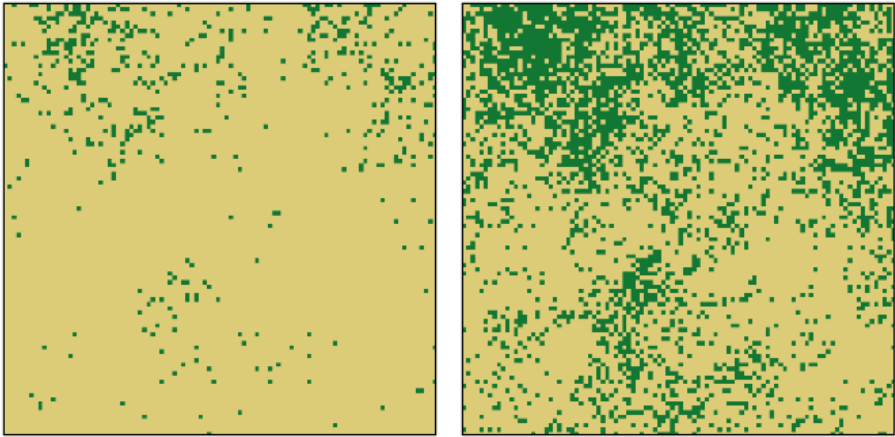
4.1 Podsumowanie ankiety

W celu zidentyfikowania potencjalnych związków pomiędzy percepcją zmian w pokryciu terenu przez ludzi a miarami niepodobieństwa, które te zmiany kwantyfikują, przeprowadzona została ankieta. Głównym celem ankiety było uzyskanie wstępnych informacji na temat tego w jaki sposób różnice w entropii oraz względnej informacji wzajemnej między analizowanymi rastrami wpływają na subiektywne postrzeganie zmian w pokryciu terenu przez ludzi.

Z uwagi na znaczną liczbę rastrów, które miałyby być przedstawione ankietowanym w trakcie badania, proces zbierania odpowiedzi respondentów przyjął formę ankiety internetowej. Pozwoliło to respondentom na wygodny udział w badaniu przy użyciu komputera lub urządzenia mobilnego. Kwestionariusz stworzony został w formie aplikacji internetowej za pomocą języka programowania R, na podstawie pakietów *shiny* (2022) oraz *shinysurveys* (2021). Sama aplikacja umieszczona została na platformie *shinyapps.io* (<https://www.shinyapps.io/>). Przeprowadzenie ankiety w tej formie umożliwiło systematyczne gromadzenie oraz przechowywanie odpowiedzi w formie tabelarycznej, ułatwiając tym samym dalszą analizę i interpretację danych.

Badanie było realizowane w terminie od 21 do 24 listopada 2022 roku. Respondenci stanowili grupę 50 studentów drugiego i trzeciego roku studiów inżynierskich na kierunku Geoinformacja na Wydziale Nauk Geograficznych i Geologicznych Uniwersytetu im. Adama Mickiewicza. Wybór tej grupy respondentów oznacza, że byli oni już zaznajomieni z tematyką tworzenia i analiz map w formie rastrowej oraz pojęciem zmian pokrycia terenu.

Każdy z ankietowanych otrzymał do wypełnienia jeden z dwóch wcześniej przygotowanych zbiorów pytań. Każdy ze zbiorów składał się z 48 pytań, przy czym część pytań między zbiorami się pokrywała. Oznacza to, że łącznie uzyskano odpowiedzi na 93 unikatowe pytania. W każdym z pytań zadaniem respondentów było określenie podobieństwa na podstawie dwóch załączonych rastrów. W ramach badania respondenci mieli możliwość wyrażania swoich odpowiedzi za pomocą pięciostopniowej skali Likerta (Likert, 1932), która obejmowała poziomy podobieństwa od “Brak” przez “Bardzo małe”, “Umiarkowane”, “Bardzo duże” aż po “Pełne”. Wykorzystanie skali Likerta o nieparzystej liczbie przedziałów, pozwoliło na zastosowanie przedziału środkowego, którego celem było reprezentowanie odpowiedzi neutralnych lub trudnych do określenia. Początkowo, zamiast skali Likerta planowano wykorzystać skalę liczbową, w zakresie mieszczącym się od 1 do 100, jednakże zrezygnowano z tego pomysłu, jako że znaczenie wartości na skali liczbowej może być interpretowane inaczej przez każdego respondenta oraz skala ta nie pozwala na jednoznaczne uwzględnienie wspomnianej wcześniej odpowiedzi neutralnej. Przykład pytania przedstawionego respondentom ilustruje Rycina 4.1.



Pytanie 2/49: Określ podobieństwo (liczba kolorów i ich ułożenie) między dwoma obrazami

Brak Umiarkowane Pełne

Brak Bardzo małe Umiarkowane Bardzo duże Pełne

Rycina 4.1: Przykładowe pytanie z ankiety

Łącznie uzyskano 2400 odpowiedzi. Ich podsumowanie przedstawione zostało w Tabeli 4.1. Według ankietowanych prawie 36% par rastrów charakteryzowała się brakiem podobieństwa, 32,6% uzyskanych odpowiedzi wskazywało na bardzo małe podobieństwo, 18,1% na umiarkowane, 11,6% bardzo duże, natomiast mniej niż 2% wskazywało na pełne podobieństwo. Warto tutaj także zwrócić uwagę, że zestawienia wszystkich odpowiedzi w zależności od liczby kategorii widocznych na rastrach nie wskazują na znaczące różnice w liczbie odpowiedzi dla danej kategorii. Największą różnicę stanowi w tym przypadku kategoria “Bardzo duże”, dla której liczba odpowiedzi dla rastrów z dwoma i trzema kategoriami pokrycia terenu różni się zaledwie o 2,7%. Najmniejszą różnicą charakteryzuje się kategoria “Pełne”, gdzie liczba odpowiedzi pomiędzy zestawami różni się o jedyne 0,9%.

Tabela 4.1: Podsumowanie odpowiedzi z ankiety

Typ odpowiedzi	Łącznie	Dwie klasy	Trzy klasy
Brak	862	420	442
Bardzo małe	783	398	385
Umiarkowane	434	211	223
Bardzo duże	278	155	123
Pełne	43	16	27
Suma	2400	1200	1200

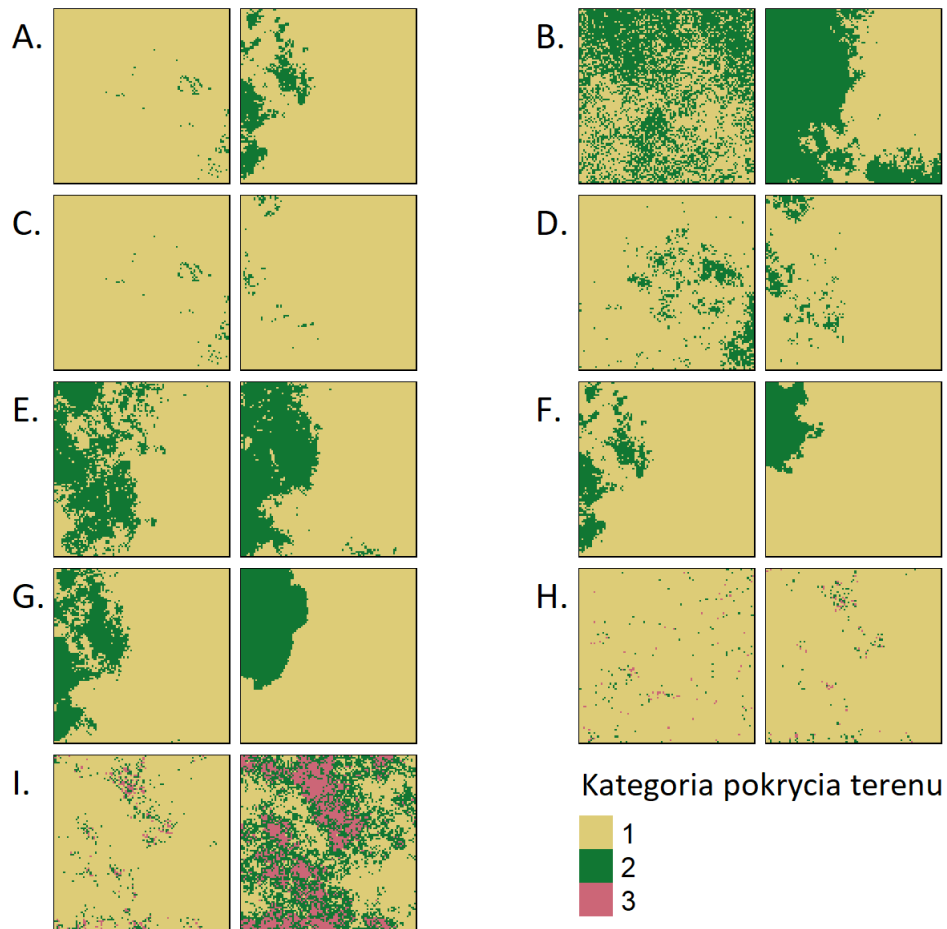
4.2 Poziom zgodności odpowiedzi ankietowanych

Aby określić jakie odpowiedzi należały do grupy najczęstszych odpowiedzi dla poszczególnych pytań, obliczona została miara określana dalej "poziomem zgodności". Poziom zgodności każdego pytania został obliczony jako stosunek najczęściej udzielonej odpowiedzi względem całkowitej liczby odpowiedzi udzielonych na to pytanie.

Poziomy zgodności ankietowanych w podziale na rodzaje pytań znajdują się w Tabeli 4.2. Całkowity poziom zgodności ankietowanych został oszacowany na 55%. Oznacza to, że 1321 z 2400 udzielonych odpowiedzi znalazło się w grupie najczęściej udzielonych odpowiedzi na pytania. Pytania zawierające rastry różniące się zarówno entropią, jak i względną informacją wzajemną cechowały się najwyższym poziomem zgodności odpowiedzi wynoszącym 61%, podczas gdy pytania różniące się wyłącznie entropią uzyskały wynik 53%, a pytania różniące się wyłącznie względną informacją wzajemną uzyskały wynik 52%.

Tabela 4.2: Poziom zgodności ankietowanych w podziale na rodzaje pytań

Podgrupa pytań	Liczba najczęściej udzielonych odpowiedzi	Liczba odpowiedzi	Poziom zgodności [%]
Różna entropia, różna RMI	429	704	61
Różna entropia, identyczna RMI	425	796	53
Identyczna entropia, identyczna RMI	467	900	52
Łącznie	1321	2400	55
RMI - względna informacja wzajemna			



Rycina 4.2: Pytania o najwyższym (A) oraz najniższym (B-I) poziomie zgodności odpowiedzi

Najwyższy poziom zgodności odpowiedzi osób ankietowanych wyniósł 92% i dotyczył pytania różniącego się zarówno entropią, jak i względną informacją wzajemną. Rastry, których dotyczyło to pytanie przedstawia Rycina 4.2.A. Aż 24 z 26 osób wskazało na brak podobieństw między rastrami uwzględnionymi w tym pytaniu. Jest to interesujący wynik, ponieważ rastry w tym pytaniu wcale nie należały do par rastrów najbardziej zróżnicowanych pod względem entropii i względnej informacji wzajemnej. Pierwszy z rastrów charakteryzował się entropią na poziomie 0,08 oraz względną informacją wzajemną wynoszącą 0,09. W przypadku drugiego rastra, wspomniane wartości wynosiły odpowiednio 0,47 i 0,48. Oznacza to, że w przypadku obu miar różnica między parami entropii oraz wynikami względnej informacji wzajemnej wyniosły 0,39.

Najniższy poziom zgodności, czyli zaledwie 38%, osiągnęło aż 8 pytań (Rycina 4.2, pary B-I). Wyłącznie dwa z nich dotyczyły pytań zawierających trzy kategorie pokrycia terenu

(pary rastrów H oraz I). 75% z tych pytań uwzględniało rastry różniące się pod względem względnej informacji wzajemnej (różna konfiguracja) przy zachowaniu tej samej entropii (identyczna kompozycja) (pary B, C oraz E-H).

4.3 Miary niepodobieństwa

Do analizy relacji między miarami niepodobieństwa a wynikami ankiety wybrano 20 miar niepodobieństwa, za pomocą których zostało obliczone niepodobieństwo każdej pary rastrów uwzględnionych w ankiecie. Uwzględnione zostały wyłącznie miary, których wyniki dla zbiorów rastrów nie wykazały bardzo silnego podobieństwa z wynikami innych miar. Decyzja ta pozwoliła skoncentrować się na bardziej zróżnicowanym zestawie miar niepodobieństwa.

W analizie uwzględniono miary: Addytywne χ^2 , odległość Canberra, Czebyszewa, Clarka, Hellingera, iloczyn skalarny, Jaccarda, Jensena-Shannona, Kullbacka-Leiblera, Kumara-Hassebrooka, Kumara-Johnsona, odległość Manhattan, Neymana, odległość euklidesową, Pearsona, podobieństwo cosinusowe (Cosinus), Ruzicki, Tanejy, średnią harmoniczną oraz Wave Hedgesa.

4.4 Odpowiedzi i miary niepodobieństwa

Analizę relacji między odpowiedziami a miarami niepodobieństwa rozpoczęto od obliczenia korelacji między różnymi sposobami podsumowania odpowiedzi a miarami niepodobieństwa. W tym celu, oryginalna skala odpowiedzi z ankiety, która obejmowała wartości od "Brak" do "Pełne", została przekształcona na skalę liczbową. W tym procesie odpowiedzi "Brak" została przypisana wartość 5, a każdej kolejnej odpowiedzi została przyporządkowana niższa cyfra, aż do "Pełne", której przypisano wartość 1.

Pod uwagę wzięto cztery sposoby podsumowania odpowiedzi: średnią z odpowiedzi, medianę odpowiedzi, najczęstszą odpowiedź (modę) oraz odchylenie standardowe (SD). Następnie wartości współczynnika korelacji Spearmana podniesiono do kwadratu, co poskutkowało przesunięciem ich zakresu z pierwotnie wynoszącego od -1 do 1, na nowy

Tabela 4.3: Zestawienie wskaźników korelacji Spearmana wybranych metod podsumowania odpowiedzi dla 2 klas

Sposób podsumowania odpowiedzi	Średnia	Mediana	Moda	SD
Średnia	1.00	0.93	0.86	0.37
Mediana	0.93	1.00	0.90	0.23
Moda	0.86	0.90	1.00	0.06
SD	0.37	0.23	0.06	1.00

Tabela 4.4: Zestawienie wskaźników korelacji Spearmana wybranych metod podsumowania odpowiedzi dla 3 klas

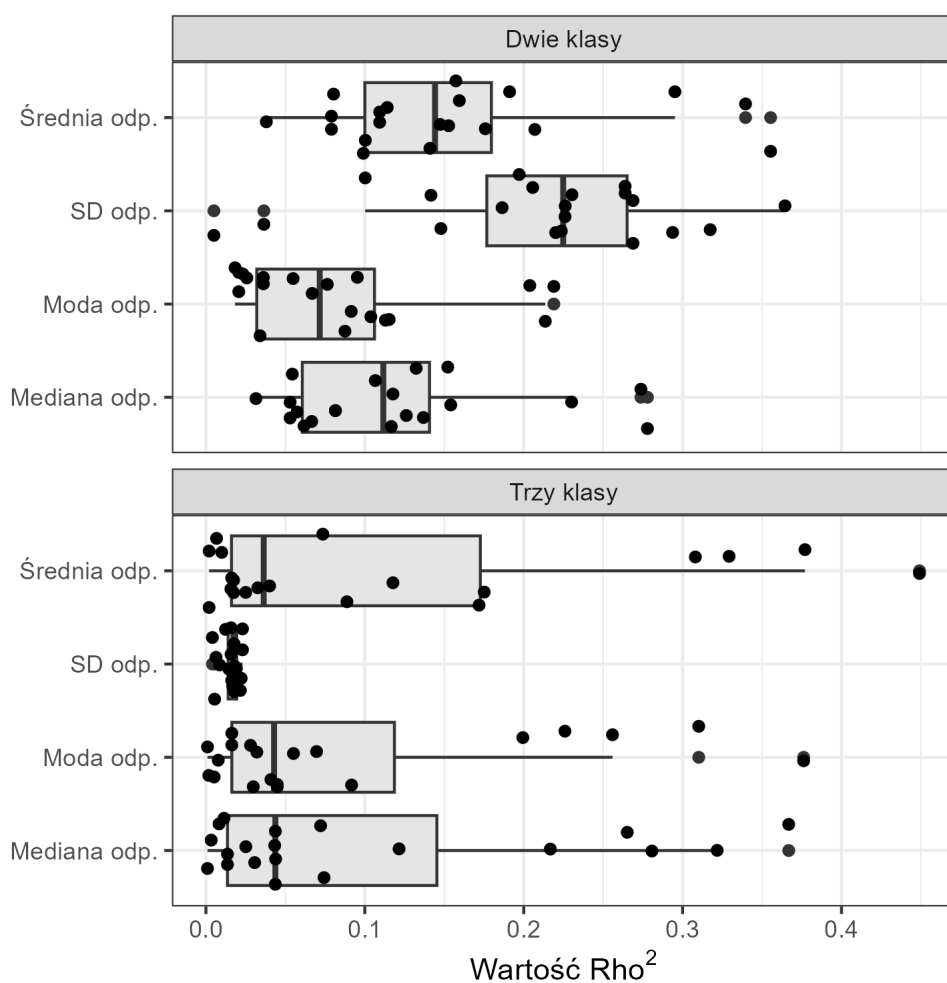
Sposób podsumowania odpowiedzi	Średnia	Mediana	Moda	SD
Średnia	1.00	0.93	0.89	0.32
Mediana	0.93	1.00	0.91	0.17
Moda	0.89	0.91	1.00	0.11
SD	0.32	0.17	0.11	1.00

zakres od 0 do 1. Relacja wyników danej miary z odpowiedziami ankietowanych jest tym silniejsza, im bliższa 1 jest uzyskana wartość współczynnika Rho^2 .

Zarówno w przypadku rastrów z dwoma (Tabela 4.3) i trzema klasami (Tabela 4.4) średnia, mediana i moda odpowiedzi są ze sobą silnie skorelowane oraz wszystkie mają względnie niewielką, ale widoczną korelację z odchyleniem standardowym odpowiedzi. Najwyższą wartość współczynnika korelacji z odchyleniem standardowym ma średnia odpowiedzi (0,37).

Rycina 4.3 pokazuje relacje między sposobami podsumowania odpowiedzi a miarami niepodobieństwa dla dwóch i trzech klas. Każda kropka na wykresie reprezentuje wartość współczynnika Rho^2 pojedynczej miary. Na podstawie wykresu można zauważyć, że każdy ze sposobów podsumowania odpowiedzi charakteryzuje się dużym rozrzutem wartości współczynnika.

Najlepszym sposobem agregacji odpowiedzi dla rastrów uwzględniających dwie klasy jest obliczenie średniej. Metoda ta charakteryzuje się zarówno jedną z najwyższych zgodności z miarami niepodobieństwa (miara Wave Hedgesa = 0,36), jak i mediana tej formy podsumowania odpowiedzi jest najwyższa. W przypadku rastrów uwzględniających trzy klasy, prawie wszystkie metody podsumowania odpowiedzi dają zbliżone wyniki. Wyjątek



Rycina 4.3: Rozkład wartości współczynnika Rho^2 pomiędzy odpowiedziami a miarami niepodobieństwa dla dwóch i trzech klas

stanowi tutaj odchylenie standardowe, które charakteryzuje się niskimi wartościami współczynnika Rho^2 dla wszystkich miar. W przypadku pozostałych sposobów agregacji najwyższą zgodność z miarami niepodobieństwa ma średnia odpowiedzi (miara Clarka = 0,45), natomiast mediana współczynników jest najwyższa dla metody grupowania odpowiedzi za pomocą mediany.

Interesującą relację z miarami niepodobieństwa ukazuje odchylenie standardowe odpowiedzi. W przypadku rastrów uwzględniających dwie klasy niektóre miary niepodobieństwa, takie jak odległość euklidesowa i miara Ruzicka, przedstawiają dosyć wysoką zgodność z tą formą agregacji, natomiast zgodność miar z odchyleniem standardowym odpowiedzi dla rastrów z trzema klasami jest bardzo niska. To oznacza, że zróżnicowanie

Tabela 4.5: Zestawienie współczynników Rho^2 miar niepodobieństwa ze średnią odpowiedzi dla dwóch i trzech klas wraz z oznaczeniami istotności statystycznej tych relacji

Miara niepodobieństwa	Rho^2 dla dwóch klas	Rho^2 dla trzech klas
Wave Hedges	0.36***	0.31***
Odl. Canberra	0.34***	0.38***
Clark	0.3***	0.45***
Kumar-Johnson	0.21**	0.17**
Addytywne χ^2	0.19**	0.12*
Taneja	0.18**	0.09*
Hellinger	0.16**	0.03
Kullback-Leibler	0.16**	0.07*
Jensen-Shannon	0.15**	0.03
Pearson	0.15**	0.33***
Śr. harmoniczna	0.14*	0.02
Neyman	0.11*	0.02
Odl. Manhattan	0.11*	0
Ruzicka	0.11*	0
Odl. euklidesowa	0.1*	0.01
Czebyszew	0.1*	0.01
Cosinus	0.08*	0.04
Kumar-Hassebrook	0.08*	0.02
Jaccard	0.08*	0.02
Iloczyn skalarny	0.04	0.18**

Przedziały istotności statystycznej: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

odpowiedzi respondentów w większości przypadków jest bardziej zgodne z wynikami miar niepodobieństwa w przypadku rastrów uwzględniających dwie klasy, niż trzy.

Współczynniki Rho^2 dla relacji 20 miar niepodobieństwa ze średnią odpowiedzi dla dwóch i trzech klas przedstawia Tabela 4.5. Średnia odpowiedzi ankietowanych z reguły charakteryzuje się silniejszą relacją z miarami niepodobieństwa obliczonymi dla pytań dotyczących rastrów uwzględniających wyłącznie dwie kategorie pokrycia terenu, niż uwzględniających trzy kategorie pokrycia terenu. Wyjątek stanowią miary Clarka oraz Canberra, które także osiągnęły najwyższe wyniki współczynnika dla trzech klas.

Spośród wszystkich analizowanych metod obliczania niepodobieństwa między parami rastrów, najsilniejsze relacje z odpowiedziami ankietowanych wykazują trzy metody: Wave Hedgesa, Canberra oraz Clarka. Dla tych miar wartości współczynnika przekroczyły poziom 0,3 w przypadku obu grup pytań.

Najwyższą wartość współczynnika w przypadku zbioru rastrów z dwoma klasami osiągnęła miara Wave Hedgesa, z Rho^2 na poziomie 0,36. Do pozostałych miar, które osiągnęły dosyć wysokie wyniki współczynnika należą miary Canberra (0,34) oraz Clarka (0,3). Wyniki trzech wymienionych powyżej miar należą też do najbardziej istotnych statystycznie, dla których wartość p wyniosła poniżej 0.001.

Cztery miary osiągnęły wartość współczynnika Rho^2 dla dwóch klas poniżej 0,1: podobieństwo cosinusowe (0,08), Hassebrookka (0,08), Jaccarda (0,08) oraz iloczyn skalarny (0,04). Wyniki te w połączeniu z niską istotnością statystyczną tych relacji na poziomie powyżej 0,05, czy nawet tak jak w przypadku iloczynu skalarnego powyżej 0,1, wskazują na bardzo niewielki lub brak związku tych miar niepodobieństwa ze średnią odpowiedzi udzielanych przez respondentów.

W przypadku średniej odpowiedzi dla rastrów uwzględniających trzy kategorie pokrycia terenu najwyższą wartość współczynnika Rho^2 na poziomie 0,45 osiągnęła miara Clarka. Jest to jednocześnie najlepszy wynik spośród wszystkich miar dla obu zbiorów rastrów. Dobre wyniki uzyskały ponownie miary Canberra oraz Wave Hedgesa, dla których wartości współczynnika wyniosły kolejno 0,38 i 0,31. Wyjątkowo silną relację wykazuje tutaj także miara Pearsona (0,33).

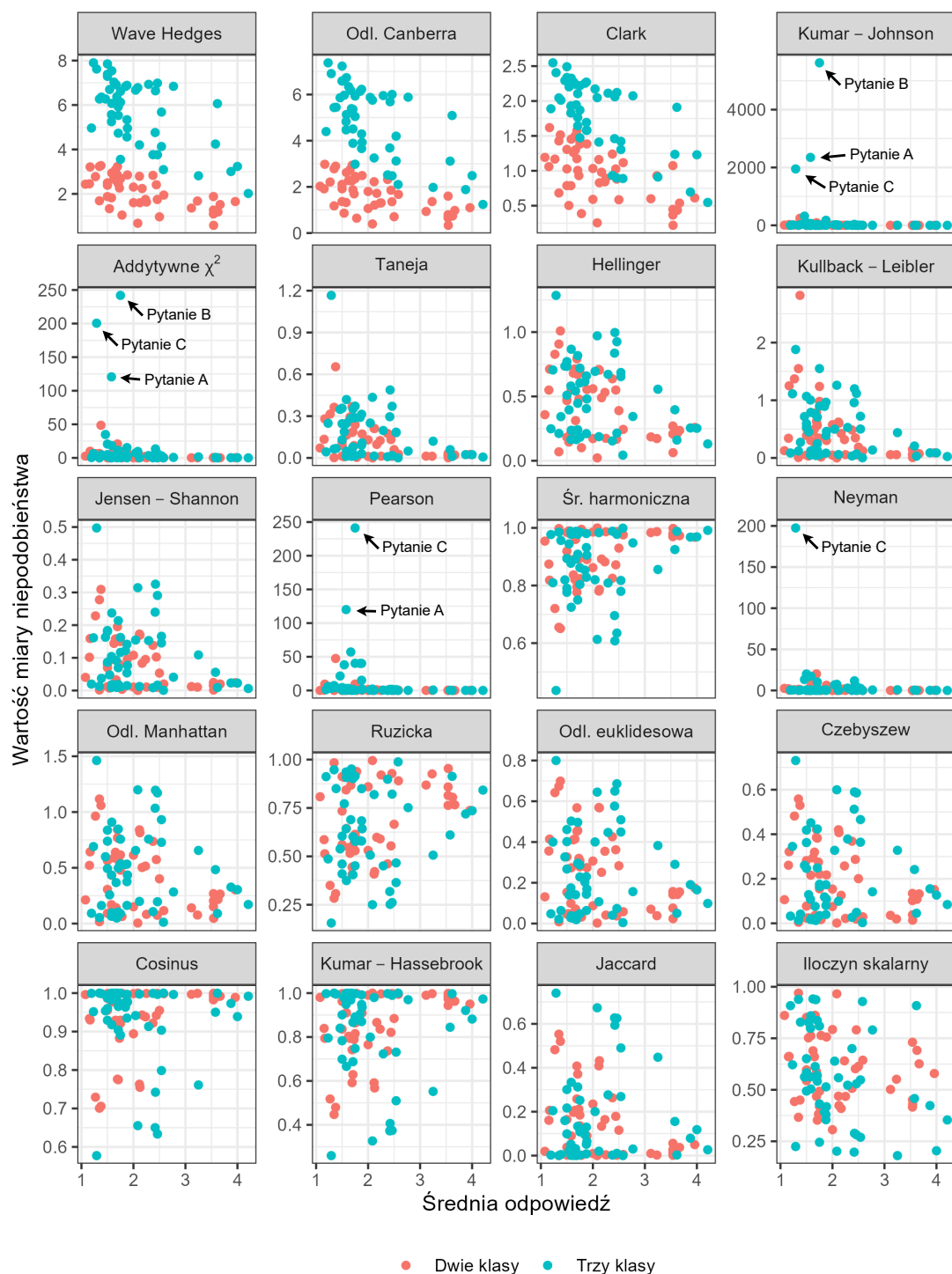
Należy także zauważyć, że w przypadku odpowiedzi na pytania dotyczące rastrów z trzema klasami, spora liczba miar charakteryzuje się bardzo słabą relacją ze średnią odpowiedzią przy jednoczesnej bardzo niskiej istotności statystycznej. Współczynnik Rho^2 aż 11 miar wyniósł poniżej 0,05. Jednocześnie dla każdej miary z tej grupy wartość p wyniosła powyżej 0,1. Najniższe wyniki w tej grupie osiągnęły miary Ruzicki i odległość Manhattan ($<0,01$) oraz Czebyszewa i odległość euklidesowa (0,01).

Rycina 4.4 przedstawia relacje między średnią odpowiedzią a miarami niepodobieństwa dla dwóch i trzech klas. Analizując wykresy można zauważyć, że w przypadku części miar zakres wartości w którym mieszczą się ich wyniki zależy od liczby klas obecnych na rastrach. Do tej grupy należą miary Wave Hedgesa, Canberra oraz Clarka, natomiast zakres wartości pozostałych miar nie jest zależny od liczby klas.

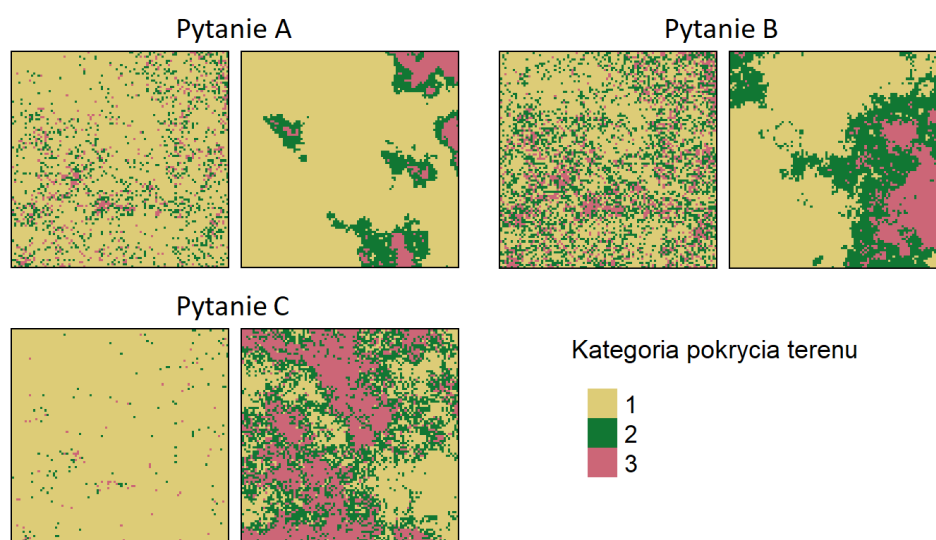
Dla miar Kumara-Johnsona, Pearsona, Neymana oraz Addytywnego χ^2 widoczne są także pojedyncze punkty znacznie przewyższające wszystkie pozostałe uzyskaną wartością niepodobieństwa. Wyniki tak znacznie odstające mają duży wpływ na ostateczny wynik współczynnika Rho^2 tych miar. Dla wszystkich czterech miar odstające punkty odpowiadają grupie tych samych trzech pytań dotyczących wyłącznie par rastrów z trzema klasami. Pary rastrów związane z odstającymi wynikami współczynnika przedstawia Rycina 4.5.

Dla czterech wcześniej wspomnianych miar, dwa najbardziej odstające wyniki odpowiadają pytaniom, które należały do podgrupy pytań o identycznej kompozycji i różnej konfiguracji przestrzennej. Rastry uwzględnione w tych pytaniach zostały oznaczone na Rycinie 4.5 jako "Pytanie A" oraz "Pytanie B". Różnica we względnej informacji wzajemnej dla tych par rastrów wynosi 0,58 i 0,63. Oznacza to, że rastry różnią się pod względem konfiguracji w dużym stopniu. W odniesieniu do pierwszej pary, 62% udzielonych przez respondentów odpowiedzi wskazywało na brak podobieństwa, natomiast dla drugiej pary rastrów 50% odpowiedzi wskazywało na bardzo niskie podobieństwo.

Jeden z odstających wyników, oznaczony na Rycinie 4.5 jako "Pytanie C", dotyczy pytania z podgrupy o różnej kompozycji i konfiguracji. Dla tej pary rastrów różnica entropii wynosi 1,42, natomiast różnica względnej informacji wzajemnej wynosi 0,22. Oznacza to, że rastry są bardzo mocno zróżnicowane pod względem kompozycji, natomiast w małym stopniu różnią się konfiguracją przestrzenną. W tym przypadku 75% odpowiedzi ankietowanych wskazywało na brak podobieństwa.



Rycina 4.4: Relacje między średnią odpowiedzią a miarami niepodobieństwa dla dwóch i trzech klas. Miary uporządkowane według wartości współczynnika Rho^2 dla dwóch klas. Wartości 1-5 na osi X odpowiadają kolejno odpowiedziom "Pełne", "Bardzo duże", "Umiarkowane", "Bardzo małe" oraz "Brak"



Rycina 4.5: Pytania związane z odstającymi wynikami współczynnika Rho^2 w relacji ze średnią odpowiedzi

4.5 Odpowiedzi a różnice między miarami z teorii informacji

Niniejszy podrozdział koncentruje się na badaniu relacji pomiędzy odpowiedziami ankietowanych a różnicami wartości miar z teorii informacji. Analiza korelacji przeprowadzona została na podstawie przygotowanych do ankiety dwóch zbiorów rastrów z dwoma i trzema klasami oraz dwoma miarami z teorii informacji: entropii oraz względnej informacji wzajemnej. Te same miary zostały wcześniej wykorzystane w trakcie przygotowania rastrów do ankiety w celu potwierdzenia uzyskania pełnego zakresu kompozycji i konfiguracji przestrzennej (Podrozdział 3.1). Wyniki korelacji Spearmana sposobów agregacji odpowiedzi z różnicami entropii i względnej informacji wzajemnej w podziale na zbiory rastrów z dwoma i trzema klasami przedstawia Tabela 4.6.

W przypadku dwóch klas, średnia odpowiedź jest negatywnie skorelowana zarówno z różnicą entropii jak i różnicą względnej informacji wzajemnej. Inaczej mówiąc, im mniejsza jest różnica entropii lub względnej informacji wzajemnej pomiędzy parą rastrów, tym wyższa jest wartość średniej odpowiedzi. Oznacza to, że im mniejsza różnica pomiędzy właściwościami informacyjnymi pary rastrów, tym według respondentów są one do siebie bardziej podobne, co potwierdza Rycina 4.6.

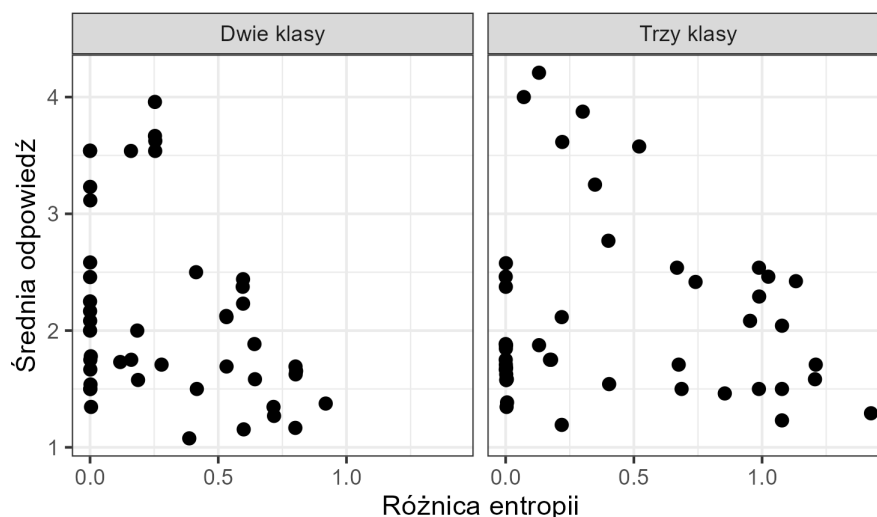
W przypadku rastrów z trzema klasami, analiza korelacji Spearmana nie wykazała istotnej relacji między średnią odpowiedzią a entropią. Wykazała natomiast negatywną relację

Tabela 4.6: Zestawienie wskaźników korelacji Spearmana sposobów agregacji odpowiedzi z różnicami entropii i względnej informacji wzajemnej

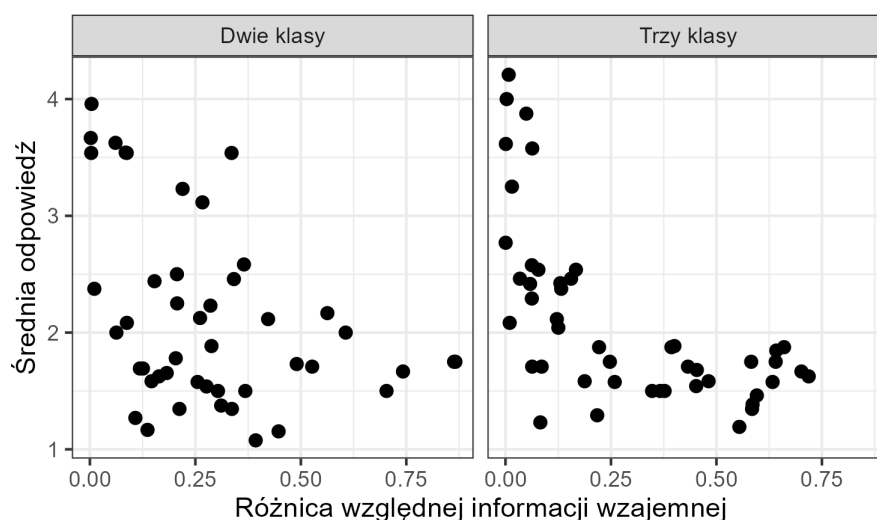
Nazwa miary	Dwie klasy		Trzy klasy	
	Różnica entropii	Różnica RMI	Różnica entropii	Różnica RMI
Średnia odp.	-0.42**	-0.37**	-0.07	-0.69
Mediana odp.	-0.31*	-0.42*	0.01	-0.7
Moda odp.	-0.21	-0.43	0	-0.66
SD odp.	-0.7***	0.24***	-0.14	0.07
Różnica entropii	1***	-0.29***	1***	-0.43***
Różnica RMI	-0.29*	1*	-0.43**	1**

RMI - względna informacja wzajemna

Przedziały istotności statystycznej: 0 '***' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1



Rycina 4.6: Relacja między średnią odpowiedzią a różnicą entropii dla dwóch i trzech klas. Wartości 1-5 na osi X odpowiadają kolejno odpowiedziom "Pełne", "Bardzo duże", "Umiarkowane", "Bardzo małe" oraz "Brak"



Rycina 4.7: Relacja między średnią odpowiedzią a różnicą względnej informacji wzajemnej dla dwóch i trzech klas. Wartości 1-5 na osi X odpowiadają kolejno odpowiedziom "Pełne", "Bardzo duże", "Umiarkowane", "Bardzo małe" oraz "Brak"

z względną informacją wzajemną na poziomie -0,69, choć istotność statystyczna tej relacji jest bardzo niska. Niemniej jednak, na Rycinie 4.7 można zauważyć, że dla różnic względnej informacji wzajemnej z zakresu od 0 do 0,25, średnia odpowiedź gwałtownie maleje. Wskazuje to na występowanie potencjalnej, nieliniowej relacji między tymi zmiennymi, co sugeruje potrzebę dalszych badań w celu jej potwierdzenia.

Odchylenie standardowe odpowiedzi dla dwóch klas jest wyraźnie negatywnie skorelowane z różnicą entropii. Im większa była różnica w entropii między parą rastrów, tym mniejsze było zróżnicowanie odpowiedzi. Oznacza to, że respondenci są tym bardziej ze sobą zgodni, im bardziej rastry różnią się od siebie pod względem entropii. Natomiast wynik wskaźnika korelacji odchylenia standardowego ze względną informacją wzajemną nie jest istotny statystycznie.

Odchylenie standardowe odpowiedzi dla trzech klas nie wykazuje istotnej statystycznie relacji z różnicami między miarami z teorii informacji. To sugeruje, że zmienność kompozycji i konfiguracji na rastrach z większą liczbą kategorii nie mają bezpośredniego wpływu na zróżnicowanie odpowiedzi ankietowanych.

Rozdział 5

Analiza zróżnicowania wybranych miar niepodobieństwa na podsta- wie danych CORINE Land Cover

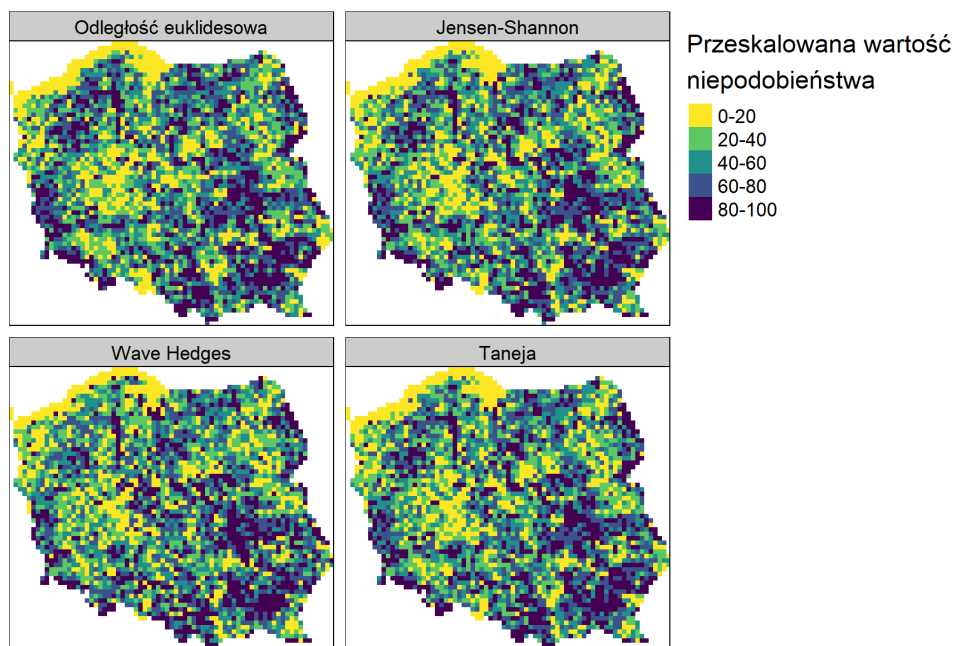
W tym rozdziale skupiono się na próbie wyjaśnienia różnic między wynikami czterech wybranych miar niepodobieństwa: odległości euklidesowej, miary Jensena-Shannona, Wave Hedgesa oraz Tanejy, które zostały obliczone dla obszarów wydzielonych z danych rastrowych o pokryciu terenu CORINE Land Cover. W pierwszej sekcji omówione jest zastosowane podejście do porównania tych miar oraz wykorzystane w tym celu dane przestrzenne CORINE Land Cover. Następnie, w drugiej sekcji opisano relacje między poszczególnymi miarami niepodobieństwa. Przedstawiony jest także sposób wyboru obszarów, dla których miary niepodobieństwa wykazały największe zróżnicowanie a także opis tych obszarów. W ostatniej sekcji tego rozdziału opisana jest relacja czterech miar niepodobieństwa ze zmianami kompozycji i konfiguracji przestrzennej.

5.1 Porównanie wybranych miar niepodobieństwa w kontekście określania zmian kategorii pokrycia terenu

Dalsza część porównania miar niepodobieństwa opierała się na zestawieniu wyników wybranych miar przy zastosowaniu w analizie zmian pokrycia terenu na podstawie danych rzeczywistych. W tym celu wykorzystano dane CORINE Land Cover dla obszaru Polski dla lat 1990 i 2018, które opisano w Podrozdziale 3.3. Dla uproszczenia analizy, dane CLC zostały zreklasyfikowane tak aby uwzględniały wyłącznie dwie kategorie pokrycia terenu: lasy i pozostałe obszary.

Do oceny zmian pokrycia terenu zastosowano metody oparte na analizie struktur przestrzennych, opisane dokładniej w Podrozdziale 2.5. Dane wejściowe podzielono na regularną siatkę kwadratów o wymiarach 100 na 100 komórek rastra (10 na 10 km). W ten sposób, obszar badań został podzielony na 3337 mniejszych jednostek. Następnie, dla tych obszarów obliczono zmiany ich struktur przestrzennych dla danych z 1990 i 2018 roku wykorzystując do tego cztery miary niepodobieństwa wybrane na podstawie różnych kryteriów:

- Odległość euklidesowa – najpopularniejsza miara niepodobieństwa, która nie jest zależna od liczby klas
- Jensen-Shannon – miara niepodobieństwa, która również nie jest zależna od liczby klas a jest wykorzystywana w pracach dotyczących podobieństwa struktur przestrzennych (Jasiewicz i in., [2014](#); Niesterowicz i in., [2016a,b](#); Briseño Montes i in., [2024](#))
- Wave Hedges – miara niepodobieństwa, która jest zależna od liczby klas, ale która dała najwyższą wartość współczynnika Rho^2 dla średniej odpowiedzi dla dwóch klas oraz jedną z najwyższych dla trzech klas
- Taneja – miara niepodobieństwa, która nie jest zależna od liczby klas, ale która dała relatywnie wysokie wartości współczynnika Rho^2 dla odchylenia standardowego odpowiedzi dla dwóch oraz trzech klas



Rycina 5.1: Mapy wyników czterech miar niepodobieństwa po przeskalowaniu dla wydzielonej siatki obszarów o wymiarach 10 na 10 km

Ze względu na to że oryginalnie wyniki tych miar mieszczą się w różnych przedziałach wartości, zostały one przeskalowane do zakresu od 0 do 100. Dzięki temu można łatwiej porównać wyniki tych miar. Należy jednak wziąć pod uwagę, że w takiej sytuacji wartość miar równa 0 niekoniecznie będzie oznaczać całkowity brak zmian w strukturze przestrzennej, a po prostu najmniejszą zmianę z całego zbioru danych. Wartość 100 natomiast reprezentować będzie obszary o największej zmianie struktury przestrzennej.

Uwzględnione miary niepodobieństwa wykazują między sobą na tyle silne relacje, że na pierwszy rzut oka trudno zauważyć istotne różnice pomiędzy nimi na mapach (Rycina 5.1). Dlatego też, aby dokładniej zrozumieć różnice w wynikach tych miar, konieczna jest analiza danych uwzględniająca bezpośrednie porównanie ze sobą wyników miar dla obszarów o największych różnicach wartości.

5.2 Porównanie wyników miar niepodobieństwa

W celu wskazania obszarów dla których miary niepodobieństwa dają najbardziej zróżnicowane wyniki, dla każdego obszaru obliczono odchylenie standardowe czterech wybranych miar niepodobieństwa. Następnie wybrano obszary, dla których odchylenie

standardowe miar znalazło się w 1% najwyższych wyników. Pozwoliło to na wskazanie 12 obszarów, dla których wyniki miar niepodobieństwa są najbardziej zróżnicowane z całego zbioru danych. Próg dla najwyższych 1% odchyleń standardowych wyniósł 31,03. Wyniki powyżej tego progu wskazują na bardzo duże zróżnicowanie wyników miar niepodobieństwa. Dodatkowo, wybrano cztery obszary dla których wartości miar Tanejy i Jensena-Shannona różniły się od siebie w znacznym stopniu.

Na podstawie wykresów na Rycinie 5.2 możemy zauważyć, że wszystkie cztery miary niepodobieństwa są ze sobą silnie skorelowane. W szczególności miary Jensena-Shannona oraz Tanejy charakteryzują się prawie pełną zgodnością. Wyjątek stanowią obszary oznaczone kolorem czerwonym, które reprezentują sytuacje o dużym zróżnicowaniu wyników miar niepodobieństwa. Obszary te widocznie odstają także w przypadku pozostałych relacji miar. Najsłabszą relację wykazują miara Wave Hedgesa z odległością euklidesową.

Obszary, dla których miary niepodobieństwa wykazały największe zróżnicowanie, zostały przedstawione na Rycinie 5.3. Wśród tych obszarów można wyróżnić dwie główne grupy: pierwsza grupa obejmuje obszary oznaczone jako 1, 5, 6, 8, 9, 10, 11 i 12, na których nastąpiły bardzo niewielkie zmiany, natomiast druga grupa to obszary 2, 3, 4, 7, 13, 14, 15 oraz 16, na których doszło do zmiany liczby klas pokrycia terenu. Każda z tych grup liczy po osiem obszarów.

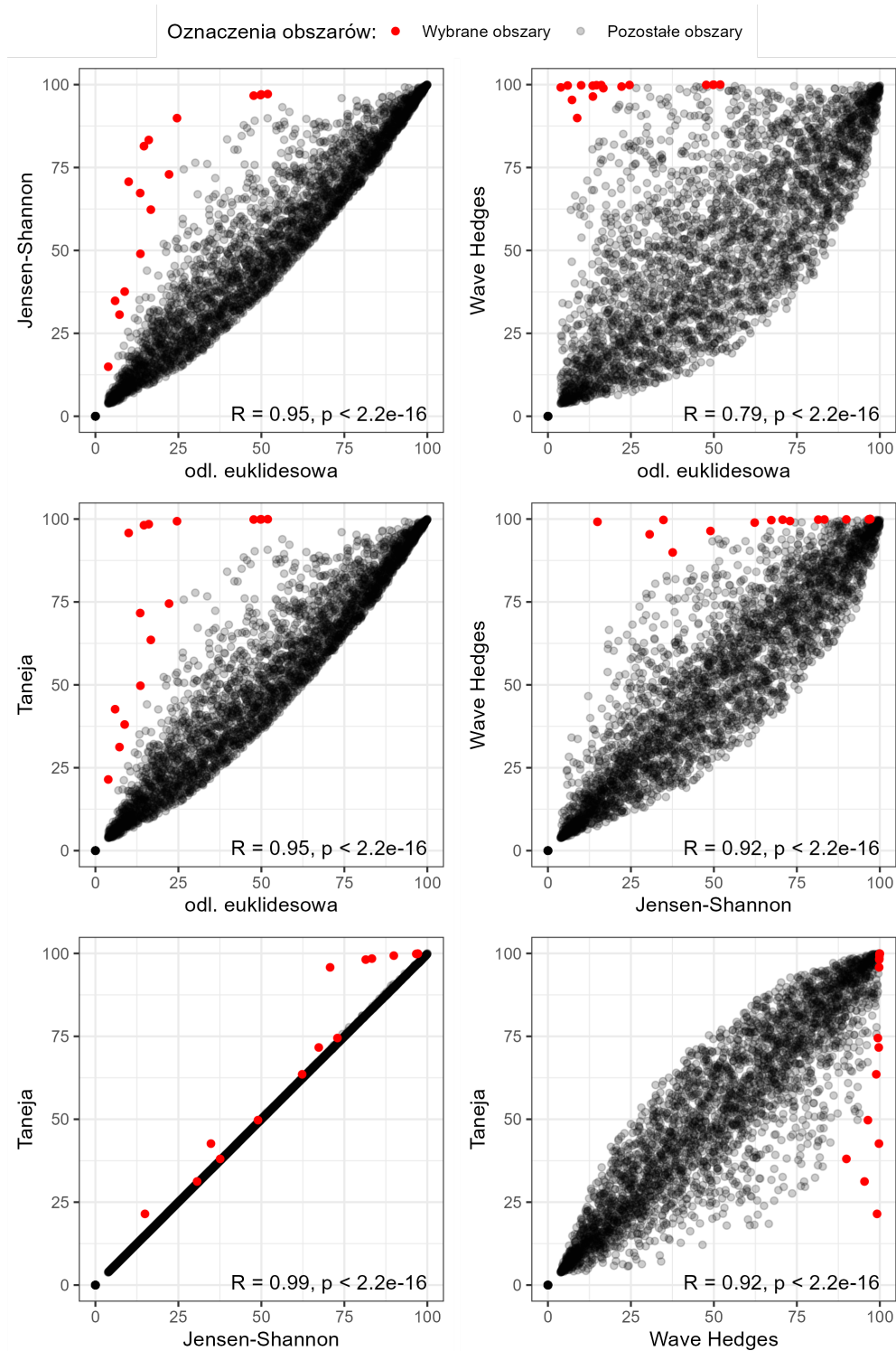
W przypadku drugiej grupy, rastry początkowe zawierają wyłącznie jedną kategorię pokrycia terenu, natomiast rastry przedstawiające stan po zmianach uwzględniają obie klasy. Na podstawie Tabeli 5.1 można zauważyć, że tym obszarom miary Tanejy oraz Jensena-Shannona przypisały dużo wyższe wartości niż pozostałym. Miara Tanejy przypisała wszystkim tym obszarom wartości powyżej 98, natomiast miara Jensena-Shannona wartości powyżej 70. Wynika z tego, że miary te są wyjątkowo wrażliwe na zmianę liczby klas między porównywanymi rastrami.

Wyniki odległości euklidesowej wykazują mniejszą zależność od zmian liczby klas uwzględnionych na rastrach w porównaniu do pozostałych miar. Jest to szczególnie widoczne w przypadku obszarów oznaczonych jako 2, 4 oraz 7 dla których pozostałe

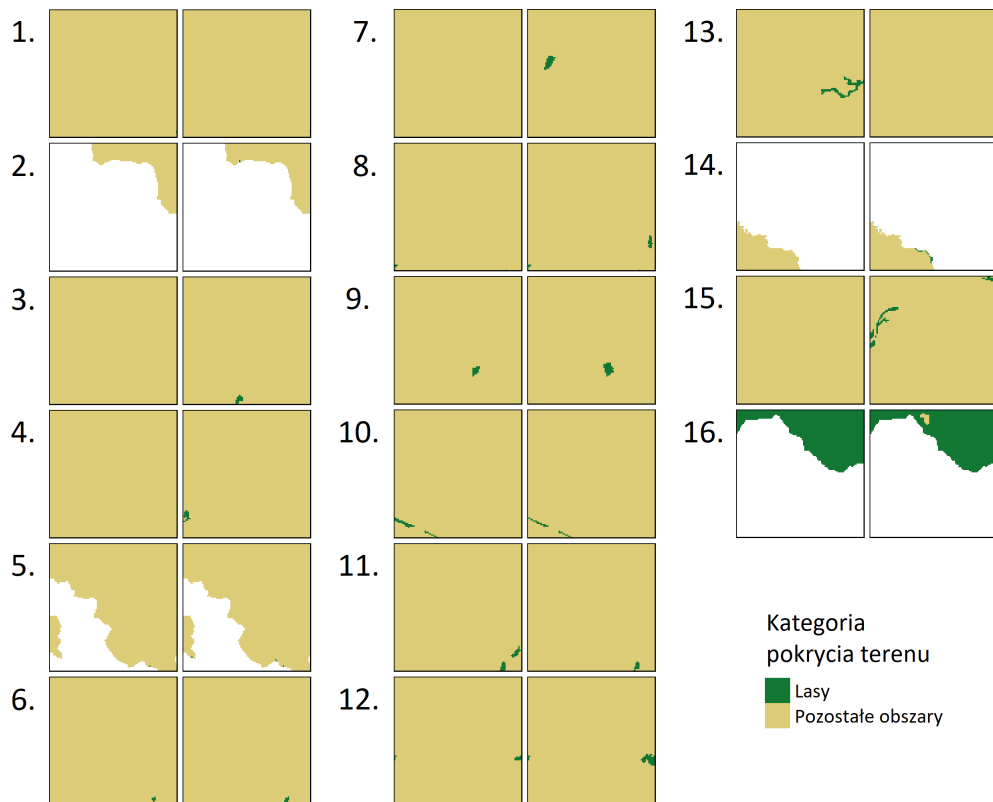
miary wskazują na bardzo duże zmiany (od 70 do 99), podczas gdy odległość euklidesowa na dosyć niewielkie (10, 16 i 24).

Warto także zauważyć, że miara Wave Hedgesa dla obszarów odstających wykazała wyłącznie bardzo wysokie wartości powyżej 89. Oznacza to, że w bardzo specyficznych sytuacjach ta miara wskazuje na bardzo duże zmiany pokrycia terenu w przypadku obszarów, na których zmiany w rzeczywistości były dosyć niewielkie.

W przypadku każdego z wybranych obszarów, miara Tanejy wskazuje na większe zmiany pokrycia terenu w porównaniu do miary Jensena-Shannona.



Rycina 5.2: Wykresy relacji między odległością euklidesową, miarą Jensena-Shannona, miarą Wave Hedgesa oraz miarą Taneja dla wydzielonych obszarów 10 na 10km



Rycina 5.3: Obszary dla których miary niepodobieństwa dają najbardziej zróżnicowane wyniki. Posortowane malejąco według odchylenia standardowego miar. Kolor biały reprezentuje komórki znajdujące się poza obszarem badań niewzeględzone podczas poniższej analizy

Tabela 5.1: Zestawienie wyników czterech miar niepodobieństwa obszarów odstających. Posortowane malejąco według odchylenia standardowego miar

Numer obszaru	Różnica entropii	Różnica RMI	odległość euklidesowa	Jensen-Shannon	Wave Hedges	Taneja
1	0.001	0.241	3.87	14.93	99.16	21.46
2	0.013	0.822	10.01	70.71	99.79	95.80
3	0.028	0.365	14.63	81.44	99.82	98.14
4	0.028	0.460	16.07	83.30	99.85	98.44
5	0.004	0.088	5.94	34.80	99.73	42.66
6	0.009	0.119	7.25	30.64	95.38	31.24
7	0.048	0.317	24.58	89.90	99.88	99.34
8	0.022	0.184	13.49	67.30	99.67	71.64
9	0.021	0.058	13.55	48.98	96.40	49.73
10	0.012	0.157	8.81	37.59	89.93	38.04
11	0.026	0.008	16.70	62.29	98.92	63.55
12	0.037	0.069	22.18	72.93	99.40	74.49
13	0.097	0.475	47.69	96.67	99.91	99.85
14	0.088	0.698	49.76	96.85	99.94	99.88
15	0.098	0.541	49.94	97.03	99.97	99.91
16	0.110	0.379	51.92	97.15	100.00	99.94

RMI - względna informacja wzajemna

5.3 Miary niepodobieństwa a kompozycja i konfiguracja przestrzenna

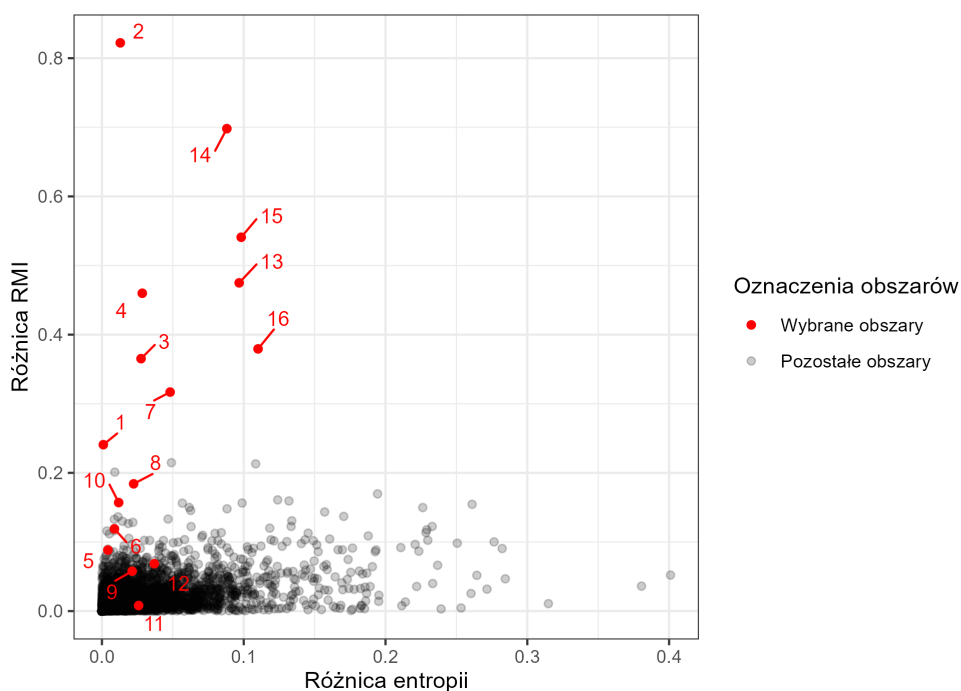
W celu oceny stopnia zależności wybranych miar niepodobieństwa od zmian w kompozycji i przestrzennej konfiguracji obszarów, przeprowadzono analizę ich korelacji Spearmana z dwiema miarami z teorii informacji: entropią oraz względną informacją wzajemną. Wyniki te umożliwiają oszacowanie wpływu zmian w kompozycji i konfiguracji przestrzennej obszarów na wyniki miar niepodobieństwa. Im wyższa wartość korelacji z różnicą entropii lub różnicą względnej informacji wzajemnej, tym teoretycznie większy wpływ mają zmiany kompozycji lub konfiguracji przestrzennej na wyniki danej miary.

Analiza wyników korelacji przedstawionych w Tabeli 5.2 ukazuje, że wszystkie miary wykazują silną relację z różnicą entropii na zbliżonym poziomie, wynoszącym między 0,71 a 0,8. Miary wykazują natomiast umiarkowaną relację z różnicą względnej informacji wzajemnej, których zakres wartości mieści się w przedziale od 0,42 do 0,67. Najsilniejszy związek z obiema miarami z teorii informacji wykazuje miara Wave Hedgesa. Miary Tanejy i Jensena-Shannona uzyskały identyczne wyniki w przypadku obu miar, natomiast najslabszą relację w obu przypadkach wykazuje odległość euklidesowa. Wyniki te oznaczają, że w przypadku tych czterech miar niepodobieństwa większy wpływ na ostateczny wynik mają zmiany udziałów kategorii w rastrze niż zmiany sąsiedowania ze sobą kategorii.

Na podstawie Ryciny 5.4, która przedstawia rozkład wartości różnicy entropii i różnicy względnej informacji wzajemnej, można zauważyć, że dla znacznej części obszarów wartości obu tych miar znajdują się w przedziale od 0 do 0,1. Oznacza to, że zmiany

Tabela 5.2: Zestawienie korelacji Spearmana pomiędzy czterema wybranymi miarami niepodobieństwa a różnicą entropii i różnicą względnej informacji wzajemnej (RMI)

Miara niepodobieństwa	Różnica entropii	Różnica RMI
odległość euklidesowa	0.71	0.42
miara Jensena-Shannona	0.78	0.55
miara Wave Hedgesa	0.80	0.67
miara Tanejy	0.78	0.55



Rycina 5.4: *Relacja różnicy entropii z różnicą względną informacji wzajemnej (RMI)*

kompozycji i przestrzennej konfiguracji na tych obszarach były stosunkowo niewielkie. Punkty na wykresie oznaczone kolorem czerwonym reprezentują 16 wcześniej wybranych obszarów charakteryzujących się wysokim odchyleniu standardowym miar niepodobieństwa. Wartości różnicy entropii obszarów dla których miary różnią się od siebie najbardziej mieszczą się w przedziale od 0,02 do 0,11. W przypadku różnic względnej informacji wzajemnej dla tych obszarów wartości są bardzo zróżnicowane i mieszczą się w przedziale od 0,008 do ponad 0,82, gdzie 9 z tych obszarów charakteryzuje się najwyższą różnicą spośród całego zbioru danych. Analizując rozmieszczenie tych 16 punktów można wywnioskować, że wyniki miar niepodobieństwa różnią się od siebie znacząco głównie w przypadku obszarów o niewielkiej różnicy entropii i dosyć dużej różnicy względnej informacji wzajemnej. To sugeruje, że istotny wpływ na zgodność wyników miar niepodobieństwa dla różnic między dwoma rastrami ma to, jak bardzo analizowane rastry różnią się od siebie pod względem sąsiadowania ze sobą poszczególnych kategorii.

Rozdział 6

Podsumowanie

Zasadniczym celem niniejszej pracy było porównanie miar umożliwiających określenie niepodobieństwa struktury przestrzennej w kontekście określania zmian kategorii pokrycia terenu przez ludzi. Do realizacji tego celu stworzono dwa zbiory danych rastrowych uwzględniających dwie i trzy kategorie pokrycia terenu. Oba zbiory przygotowano za pomocą symulacji, co pozwoliło na uzyskanie pełnej reprezentacji wszystkich możliwych wartości kompozycji i konfiguracji przestrzennej. Rastry wykorzystano do przeprowadzenia internetowej ankiety na grupie 50 studentów Wydziału Nauk Geograficznych i Geologicznych Uniwersytetu im. Adama Mickiewicza w Poznaniu. W tej ankiecie, zadaniem respondentów było określenie stopnia podobieństwa między parami rastrów w skali która obejmowała poziomy od "Brak" przez "Bardzo małe", "Umiarkowane", "Bardzo duże" aż po "Pełne". Łącznie uzyskano 2400 odpowiedzi na 93 pytania. Poziom zgodności ankietowanych, czyli stosunek najczęściej udzielonych odpowiedzi względem całkowitej liczby pytań wyniósł 55%. Respondenci najczęściej udzielali zgodnych odpowiedzi w kwestii oceny podobieństwa w przypadku par rastrów różniących się od siebie pod względem kompozycji jak i konfiguracji przestrzennej - poziom zgodności dla tej grupy pytań wyniósł 61%. Najniższy wynik, na poziomie 52%, uzyskała grupa par rastrów różniących się wyłącznie względną informacją wzajemną.

W kolejnym etapie analizy dla wszystkich par rastrów uwzględnionych w ankiecie zostały obliczone wartości 20 miar niepodobieństwa. Pierwszym krokiem był wybór metody

agregacji odpowiedzi, która wykazywała największą zgodność z tymi miarami. Wzięto pod uwagę cztery sposoby grupowania odpowiedzi: średnią, medianę, modę i odchylenie standardowe. Analiza relacji między tymi metodami a miarami niepodobieństwa pozwoliła stwierdzić, że dla zbioru rastrów uwzględniających dwie klasy najlepszym sposobem agregacji odpowiedzi jest obliczenie średniej. W przypadku zbioru rastrów z trzema klasami wszystkie metody uzyskały podobne wyniki. Szczególnie interesującą relację z miarami niepodobieństwa ukazuje odchylenie standardowe odpowiedzi - część miar wykazała silną relację z tą metodą agregacji w przypadku zbioru rastrów uwzględniających dwie klasy, podczas gdy dla zbioru z trzema klasami relacje te były bardzo słabe.

Na podstawie relacji 20 miar niepodobieństwa ze średnią odpowiedzi wykazano, że ten sposób grupowania z reguły cechuje się silniejszą relacją z miarami niepodobieństwa obliczonymi dla pytań dotyczących rastrów uwzględniających wyłącznie dwie kategorie pokrycia terenu, niż uwzględniających trzy kategorie pokrycia terenu. Dla obu zbiorów rastrów najlepsze wyniki uzyskały trzy miary: miara Wave Hedgesa, odległość Canberra oraz miara Clarka. Jest to godne uwagi, ponieważ są to jedyne miary, których zakres wartości jest zauważalnie zależny od liczby klas obecnych na analizowanych rastrach. W związku z tym, przy wykorzystaniu tych miar do analiz pokrycia terenu należy zastanowić się nad ewentualnym zastosowaniem normalizacji wyników. Warto również rozważyć ograniczenie ich stosowania do przypadków, w których liczba klas pozostaje stała, aby uzyskać bardziej spójne i porównywalne wyniki analizy.

Dla wszystkich par rastrów, dodatkowo obliczono różnicę entropii oraz różnicę względnej informacji wzajemnej, które dostarczają informacji o zróżnicowaniu tych rastrów pod względem kompozycji i przestrzennej konfiguracji. W przypadku rastrów składających się z dwóch klas, średnia odpowiedź jest negatywnie skorelowana zarówno z różnicą entropii, jak i różnicą względnej informacji wzajemnej. Wynika z tego, że im mniejsza różnica między właściwościami informacyjnymi pary rastrów, tym są one przez respondentów postrzegane jako bardziej podobne. Ponadto, odchylenie standardowe odpowiedzi dla rastrów uwzględniających dwie klasy wykazuje wyraźną ujemną korelację z różnicą entropii, co oznacza, że im większa różnica między rastrami pod względem entropii tym

większa zgodność odpowiedzi ankietowanych. W przypadku rastrów z trzema klasami, analiza korelacji Spearmana nie wykazała istotnej zależności między średnią odpowiedzią a różnicą entropii oraz różnicą względnej informacji wzajemnej. Wskazano jednak potencjalne występowanie nieliniowej relacji między średnią odpowiedzią a entropią, co sugeruje konieczność przeprowadzenia dalszych badań w celu jej potwierdzenia.

W dalszej części pracy skupiono się na próbie wyjaśnienia różnic między wynikami czterech wybranych miar niepodobieństwa: odległości euklidesowej, miary Jensena-Shannona, Wave Hedgesa oraz Tanejy przy zastosowaniu w analizie zmian pokrycia terenu na podstawie danych rzeczywistych. Do tego celu wykorzystano dane rastrowe o pokryciu terenu CORINE Land Cover (CLC) dla obszaru Polski dla lat 1990 i 2018, które zostały zreklasyfikowane, aby obejmowały wyłącznie dwie kategorie pokrycia terenu: lasy i pozostałe obszary. Obszar badań podzielono na mniejsze jednostki, dla których obliczono wartości czterech miar niepodobieństwa oraz różnicę entropii i różnicę względnej informacji wzajemnej. Następnie, na podstawie odchylenia standardowego wybranych miar niepodobieństwa wskazano obszary, dla których te miary uzyskały najbardziej zróżnicowane wyniki. Wszystkie cztery miary charakteryzują się dosyć silną relacją, a w szczególności miary Jensena-Shannona oraz Tanejy, które wykazują prawie pełną zgodność. Analiza obszarów o największym zróżnicowaniu wyników miar niepodobieństwa pozwoliła określić, że miary Tanejy oraz Jensena-Shannona wskazują na duże zmiany w strukturze przestrzennej w przypadku obszarów, na których doszło do zmiany liczby kategorii pokrycia terenu, natomiast odległość euklidesowa wykazuje w tym przypadku mniejszą zależność. Miara Wave Hedgesa dla tych obszarów wykazała wyłącznie wysokie wartości. Oznacza to, że w specyficznych sytuacjach ta miara wskazuje na bardzo duże zmiany pokrycia terenu, nawet gdy w rzeczywistości były one dosyć niewielkie. Na podstawie relacji wybranych miar niepodobieństwa z różnicami miar z teorii informacji wskazano, że wyniki miar niepodobieństwa różnią się od siebie znacząco głównie w przypadku obszarów o niewielkiej różnicy entropii i dosyć dużej różnicy względnej informacji wzajemnej. To sugeruje, że istotny wpływ na zgodność wyników miar niepodobieństwa dla różnic między dwoma rastrami ma to, jak bardzo analizowane rastry różnią się od siebie pod względem sąsiedowania ze sobą poszczególnych kategorii.

W świetle przeprowadzonych badań, zalecane jest wykorzystanie miar niepodobieństwa takich jak odległość Canberra, miara Wave Hedgesa, czy miara Clarka do analiz zmian pokrycia terenu obszarów o stałej liczbie kategorii. Rekomendacja ta wynika z obserwowanej silnej relacji między wartościami tych miar a postrzeganiem różnic w strukturze przestrzennej przez ludzi. Dodatkowo, w przypadku analizy obszarów o różnej liczbie klas istnieje możliwość skalowania wartości miar, co mogłoby ujednolicić ich przedziały wartości. Alternatywnie zaleca się wykorzystanie miar Kullbacka-Leiblera oraz Tanejy, które mimo tego, że charakteryzują się słabszą relacją z oceną zmian przez ludzi, to ich zakresy wartości są w mniej zależne od liczby kategorii obecnych na rastrach.

W tej pracy ustalono, że istnieją miary niepodobieństwa, które wykazują istotne różnice w strukturze przestrzennej przy porównywaniu rastrów o różnej liczbie klas. Jednakże w ankiecie nie uwzględniono pytań, w których zadaniem respondentów byłoby określenie podobieństwa między takimi parami rastrów. W związku z tym, w kolejnych pracach zaleca się przeprowadzenie analizy wpływu zmiany liczby klas obecnych na rastrach na ocenę zmian pokrycia terenu przez ludzi. Wykazano także, że przedziały wartości niektórych miar też są zależne od liczby klas obecnych na badanych rastrach. Warto rozważyć w kolejnych badaniach analizę zakresów wartości przy porównywaniu par rastrów o zróżnicowanej liczbie klas.

Bibliografia

- Anderson, E i C Mammides (2020). Changes in land-cover within high nature value farmlands inside and outside Natura 2000 sites in Europe: A preliminary assessment. *AMBIO A Journal of the Human Environment*.
- Bibby, PR i J Shepherd (1999). Monitoring land cover and land use for urban and regional planning. *Geographical Information Systems* **2**, 953–965.
- Bielecka, E i A Ciołkosz (2004). Mapa pokrycia terenu w Polsce w skali 1:1 000 000 jako wynik wizualizacji bazy danych CLC-2000. *Polski Przegląd Kartograficzny* **T. 36, nr 4**, 274–287.
- Briseño Montes, A, J Salas, EA Villaseñor Garcia, RR Suarez i D Wood (2024). Assessing Human Settlement Sprawl in Mexico via Remote Sensing and Deep Learning. *IEEE Latin America Transactions* **22(3)**, 174–185.
- Cha, SH (2007). Comprehensive Survey on Distance/Similarity Measures Between Probability Density Functions. *Int. J. Math. Model. Meth. Appl. Sci.* **1**.
- Chang, W, J Cheng, J Allaire, C Sievert, B Schloerke, Y Xie, J Allen, J McPherson, A Dipert i B Borges (2022). *shiny: Web Application Framework for R*. R package version 1.7.4. <https://CRAN.R-project.org/package=shiny>.
- Comber, A, P Fisher i R Wadsworth (2005). What is Land Cover? *Environment and Planning B: Planning and Design* **32**, 199–209.
- Cunningham, MA (2006). Accuracy assessment of digitized and classified land cover data for wildlife habitat. *Landscape and Urban Planning* **78(3)**, 217–228.
- Dražković, B, B Miletić, M Gutalj i S Stjepanović (2020). Climate changes and fires in Bosnia and Herzegovina. W: *Proceedings of the XI International Scientific Agricultural Symposium "Agrosym"*, pp.694–702.

- Enoguanbhor, EC, F Gollnow, JO Nielsen, T Lakes i BB Walker (2019). Land Cover Change in the Abuja City-Region, Nigeria: Integrating GIS and Remotely Sensed Data to Support Land Use Planning. *Sustainability* **11**(5).
- Falcucci, A, L Maiorano i L Boitani (2006). Changes in land-use/land-cover patterns in Italy and their implications for biodiversity conservation. *Landscape Ecology* **22**, 617–631.
- Feranec, J, G Hazeu, S Christensen i G Jaffrain (2007). Corine Land Cover Change Detection in Europe (Case Studies of the Netherlands and Slovakia). *Land Use Policy* **24** (2006) **1** 24.
- Fisher, P, A Comber i R Wadsworth (2005). "Land use and land cover: contradiction or complement". W: s. 85–98.
- Forman, R (1995). *Land Mosaics: The Ecology of Landscapes and Regions*. Cambridge University Press. https://books.google.pl/books?id=sSRNU_5P5nwC.
- German, K (2004). Zastosowanie koncepcji płatów i korytarzy do analizy funkcjonalnej krajobrazu wyżynnego. *Problemy Ekologii Krajobrazu* (14).
- Głazewski, A (2006). Modele rzeczywistości geograficznej a modele danych przestrzennych. *Polski Przegląd Kartograficzny* **38**(3), 217–225.
- Gustafson, EJ (1998). Quantifying landscape spatial pattern: what is the state of the art? *Ecosystems* **1**(2), 143–156.
- Haralick, RM, K Shanmugam i I Dinstein (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-3**(6), 610–621.
- Jasiewicz, J, P Netzel i TF Stepinski (2014). Retrieval of pattern-based information from giga-cells categorical rasters — Concept and new software. W: *2014 IEEE Geoscience and Remote Sensing Symposium*, pp.1785–1788.
- Jasiewicz, J, P Netzel i TF Stepinski (2015). GeoPAT: A toolbox for pattern-based information retrieval from large geospatial databases. *Comput. Geosci.* **80**, 62–73.
- Kepski, D (2018). Wpływ rzeźby i pokrycia terenu na rozkład przestrzenny i dynamikę zmian pokrywy śnieżnej na tundrze w okolicy Polskiej Stacji Polarnej na Spitsbergenie.
- Kozak, J, M Luc, K Ostapowicz i E Ziolkowska (2014). "Pozyskiwanie i analiza danych o pokryciu terenu a badania struktury przestrzennej krajobrazu". W: s. 63–84.
- Likert, R (1932). A technique for the measurement of attitudes. *Archives of psychology*.

- Lu, D, P Mausel, E Brondizio i E Moran (2004). Change Detection Techniques. *International Journal of Remote Sensing* **25**.
- Mallinis, G, D Emmanoloudis, V Giannakopoulos, F Maris i N Koutsias (2011). Mapping and interpreting historical land cover/land use changes in a Natura 2000 site using earth observational data: The case of Nestos delta, Greece. *Applied Geography* **31**, 312–320.
- McGarigal, K, S Cushman, M Neel i E Ene (2023). McGarigal K, Cushman SA, Neel MC, Ene E. FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. www.umass.edu/landeco/research/fragstats/fragstats.html.
- McGarigal, K, S Tagil i S Cushman (2009). Surface metrics: an alternative to patch metrics for the quantification of landscape structure. *Landscape Ecology* **24**, 433–450.
- Mierzwiak, M i B Całka (2019). Zmiany pokrycia terenu w Polsce w latach 1990-2012 na podstawie danych CORINE Land Cover. *Roczniki Geomatyki* **T. 17, z. 2(85)**, 127–140.
- Netzel, P i TF Stepinski (2015). Pattern-Based Assessment of Land Cover Change on Continental Scale With Application to NLCD 2001–2006. *IEEE Transactions on Geoscience and Remote Sensing* **53(4)**, 1773–1781.
- Niesterowicz, J i T Stepinski (2016a). On using landscape metrics for landscape similarity search. *Ecological Indicators* **64**, 20–30.
- Niesterowicz, J, TF Stepinski i J Jasiewicz (2016b). Unsupervised regionalization of the United States into landscape pattern types. *International Journal of Geographical Information Science* **30(07)**, 1450–1468.
- Nowosad, J (2021). Motif: an open-source R tool for pattern-based spatial analysis. *Landscape Ecology* **36**, 29–43.
- Nowosad, J i T Stepinski (2019). Information theory as a consistent framework for quantification and classification of landscape patterns. *Landscape Ecology* **34**.
- Pavlovic, M, S Ilic, N Antonic i D Culibrk (2022). Monitoring the Impact of Large Transport Infrastructure on Land Use and Environment Using Deep Learning and Satellite Imagery. *Remote Sensing* **14(10)**.
- Pukowiec-Kurda, K i M Sobala (2016). Nowa metoda oceny stopnia antropogenicznego przekształcenia krajobrazu na podstawie metryk krajobrazowych. *Prace Komisji Krajobrazu Kulturowego* **nr 31**, 71–84.

- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Schlather, M, A Malinowski, P Menck, M Oesting i K Stokorb (2015). Analysis, Simulation and Prediction of Multivariate Random Fields with Package Random Fields. *Journal of statistical software* **63**, 1–25.
- Sciaini, M, M Fritsch, C Scherer i CE Simpkins (2018). NLMR and landscapetools: An integrated environment for simulating and modifying neutral landscape models in R. *Methods in Ecology and Evolution* **00**, 1–9.
- Sleeter, B, T Sohl, T Loveland, R Auch, W Acevedo, M Drummond, K Sayler i S Stehman (2013). Land-cover change in the conterminous United States from 1973 to 2000. *Global Environmental Change* **23**, 733–748.
- Solon, J (2002). *Ocena różnorodności krajobrazu na podstawie analizy struktury przestrzennej roślinności*. Prace Geograficzne - Polska Akademia Nauk. PAN IG i PZ. <https://books.google.pl/books?id=00dcAwAAQBAJ>.
- Trattner, J i L D'Agostino McGowan (2021). *shinysurveys: Create and Deploy Surveys in 'Shiny'*. R package version 0.2.0. <https://CRAN.R-project.org/package=shinysurveys>.
- Zwolinski, Z i J Gudowicz (2018). "Zmienność przestrzenna typów pokrycia terenu i użytkowania ziemi w zlewniach rzecznych i jeziornych ZMŚP". W: s. 491–524.