

## Klasyfikacja pokrycia terenu z wykorzystaniem obrazów Sentinel-2A przetworzonych za pomocą metody głównych składowych (PCA)

*Land cover classification using Sentinel-2A images processed by the principal components method (PCA)*

Urszula KAŁUŻNA

Wyższa Szkoła Gospodarki Krajowej w Kutnie, Wydział Nauk Technicznych  
ul. Lelewela 7  
99-300 Kutno  
[u.kaluzna@powiat.plock.pl](mailto:u.kaluzna@powiat.plock.pl)

 Krzysztof BĘDKOWSKI

Uniwersytet Łódzki, Wydział Nauk Geograficznych  
ul. Kopcińskiego 31  
90-142 Łódź  
[krzysztof.bedkowski@geo.uni.lodz.pl](mailto:krzysztof.bedkowski@geo.uni.lodz.pl)

### Streszczenie

Celem badań jest ocena możliwości realizacji klasyfikacji nadzorowanej z wykorzystaniem obrazów (komponentów) uzyskiwanych w wyniku przetworzenia oryginalnych obrazów Sentinel-2A za pomocą metody głównych składowych (PCA). Klasyfikację wykonano w ośmiu wariantach, z wykorzystaniem algorytmów najmniejszej odległości (MD, *Minimum Distance*) oraz największego prawdopodobieństwa (ML, *Maximum Likelihood*), przy czym zastosowano oryginalne kanały 2, 3, 4, 8 Sentinel-2A oraz różną liczbę komponentów. Wyniki klasyfikacji oceniono poprzez porównanie z danymi o pokryciu terenu według Ewidencji Gruntów i Budynków (EGiB). Przeprowadzenie klasyfikacji na ograniczonej do dwóch liczbie komponentów uzyskanych w procedurze PCA tylko nieznacznie zmieniło wyniki w porównaniu do klasyfikacji na oryginalnych, nieprzetworzonych kanałach Sentinel-2A. Najbardziej zbliżone do danych EGiB rezultaty uzyskano stosując klasyfikację ML kanałów oryginalnych, nieprzetworzonych lub używając wszystkich komponentów PCA. Podjęta próba porównania pokrycia terenu ustalonego za pomocą klasyfikacji obrazów satelitarnych z klasami pokrycia, które zostały wyodrębnione z mapy EGiB wykazała, że przetworzenie mapy z postaci wektorowej na rastrową wpływa istotnie na uzyskiwane wyniki.

### Abstract

The aim of the research is to assess the feasibility of supervised classification using images (components) obtained through processing the original Sentinel-2A images by means of the principal component method (PCA). The classification was performed in eight variants, using the algorithms of the minimum distance (MD) and the maximum likelihood (ML), with the original channels 2, 3, 4, 8 of Sentinel-2A and a various number of components. The results of the classification were assessed by comparing them to the land coverage data of Land and Buildings Register (*Ewidencja Gruntów i Budynków* – EGiB). Performing the classification on a number of PCA components

limited to two only slightly altered the results compared to the classification on the original, raw Sentinel-2A channels. The results most similar to the EGiB data were obtained using the ML classification of the original channels, i.e. raw channels or using all PCA components. The attempt to compare the land coverage established by the classification of satellite images to the coverage classes that were extracted from the EGiB map revealed that processing the map from vector to raster form significantly influences the obtained results.

**Keywords:** remote sensing, land cover, Land and Buildings Register, Sentinel-2A, PCA, supervised image classification.

**Słowa kluczowe:** teledetekcja, pokrycie terenu, EGiB, Sentinel-2A, PCA, nadzorowana klasyfikacja obrazu.

## Wprowadzenie

Systemy satelitarne umożliwiają obecnie prowadzenie ciągłego monitoringu Ziemi. Na podstawie zobrażeń o różnej rozdzielczości (przestrzennej, spektralnej, radiometrycznej i czasowej) wykonywane są obserwacje komponentów środowiska naturalnego, w tym głównie pokrycia terenu. Interpretacja wyników pozwala także na wnioskowanie o użytkowaniu gruntów. Informacje te pozwalają zdobyć wiedzę na temat naturalnych procesów zachodzących w środowisku, jak również zmian będących wynikiem bezpośredniej działalności człowieka. Liczba dostępnych danych satelitarnych stale rośnie, przy czym dzięki udoskonaleniom parametrów technicznych systemów, dostarczane informacje obrazowe są coraz bardziej precyzyjne.

Jedną z głównych metod teledetekcji służących rozpoznawaniu pokrycia terenu jest klasyfikacja nadzorowana. Istnieje wiele algorytmów – oprócz klasycznych, takich jak metoda równoległościaków (*Paralelliped*), m. najmniejszej odległości (*Minimum Distance*), czy m. największego prawdopodobieństwa (*Maximum Likelihood*), coraz częściej stosowane są np. maszyna wektorów nośnych (*Support Vector Machine*, SVM), czy losowy las decyzyjny (*Random Forest*, RF). Sięga się także do narzędzi uczenia maszynowego, tj. sztucznej inteligencji (Babenko, Marmanis, McIlwraith 2017).

W klasyfikacji istotne znaczenie ma również dobór i ewentualne wstępne przetworzenie kanałów spektralnych użytych obrazów satelitarnych. Zastosowanie większej liczby kanałów wprawdzie na ogół zwiększa skuteczność klasyfikacji (choć z drugiej strony nadmiar kanałów skorelowanych może zwiększyć szum informacyjny), jednak znacznie wydłuża czas przetwarzania danych, co może mieć znaczenie przy wykonywaniu opracowań dla dużego obszaru (np. kraju lub kontynentu) albo z wykorzystaniem obrazów zarejestrowanych w wielu terminach. Znanym od dawna w teledetekcji sposobem redukcji liczby kanałów jest metoda głównych składowych (PCA, *Principal Components Analysis*), za której pomocą można zapisać istotną dla procesu klasyfikacji informację w nowych obrazach, tzw. komponentach (Adamczyk, Będkowski 2007). Komponenty są nieskorelowane ze sobą oraz „uporządkowane” według zawartości informacyjnej. Najwięcej przydatnych w klasyfikacji informacji zawierają pierwsze komponenty. Komponenty dalsze dosyć często zawierają szumy, tj. uszkodzenia obrazu spowodowane błędami systemu rejestrującego lub powstające podczas transmisji danych. Mogą być także związane z pewny-

mi obiektami, które są dosyć specyficzne pod względem ich charakterystyk spektralnych, lecz w obrazie jest ich niewiele (np. niezbyt liczna zabudowa z różnymi pokryciami dachów w krajobrazie rolno-leśnym). W odwrotnej transformacji PCA można „odtworzyć” kanały obrazu satelitarnego, korzystając wyłącznie z pierwszych komponentów. Otrzymane obrazy są w ten sposób oczyszczone z szumów. Zarówno komponenty PCA, jak i odtworzone kanały mogą być używane w dalszych procedurach przetwarzania danych, w tym m.in. do interpretacji kompozycji barwnych, ewentualnie nadzorowanej lub nienadzorowanej klasyfikacji treści obrazów (Bhattacharya i in. 2007).

## Cel i zakres badań

Celem badań jest ocena możliwości realizacji klasyfikacji nadzorowanej pokrycia terenu obszaru wielkości jednego powiatu, z wykorzystaniem różnej liczby komponentów uzyskiwanych w wyniku przetworzenia oryginalnych obrazów Sentinel-2A za pomocą metody głównych składowych (PCA). Klasyfikację wykonano w ośmiu wariantach, z wykorzystaniem algorytmów najmniejszej odległości (MD, *Minimum Distance*) oraz największego prawdopodobieństwa (ML, *Maximum Likelihood*). Uzyskane wyniki oceniono poprzez porównanie z wynikami klasyfikacji oryginalnych, nieprzetworzonych kanałów Sentinel-2A (2, 3, 4, 8) oraz z danymi zawartymi w bazie danych Ewidencji Gruntów i Budynków (EGiB).

## Obiekt badań, materiały i metody

### Obiekt badań

Badania dotyczą miasta Płock na prawach powiatu oraz powiatu płockiego<sup>1</sup>, które zlokalizowane są w środkowej części Polski, w województwie mazowieckim. Powiat płocki otacza ze wszystkich stron miasto Płock, które jest nie tylko ważnym ośrodkiem gospodarczym, turystycznym i kulturowym Mazowsza Zachodniego, ale też historyczną stolicą regionu<sup>2</sup>. Zasięg obszaru badań o łącznej powierzchni 1883,75 km<sup>2</sup> okre-

<sup>1</sup> W niniejszej pracy podajemy konsekwentnie dane odnoszące się łącznie do obszaru składającego się z powiatu płockiego oraz miasta Płock. Używając określenia „powiat” mamy na myśli obydwie jednostki administracyjne.

<sup>2</sup> <http://powiat-plock.geoportal2.pl/>

śląją współrzędne geograficzne skrajnych punktów: na północy N: 52° 46' 55", E: 19° 57' 43", na południu N: 52° 19' 38", E: 19° 56' 1", na zachodzie N: 52° 32' 5", E: 19° 19' 10", na wschodzie N: 52° 23' 58", E: 20° 15' 6".

Dane liczbowe przytoczone dalej podajemy za (GUS 2014) oraz (Uchwała 2016, Uchwała 2017). Obiekty pochodzenia antropogenicznego związane z infrastrukturą miejską i działalnością przemysłową położone są głównie na obszarach miast i miejscowości gminnych. Obszerne tereny przemysłowe należą do PKN Orlen w Płocku, Bazy Surowców PERN „Przyjaźń” S.A. w Plebance, zakładów przetwórstwa mięsnego, oczyszczalni ścieków. Tereny zabudowane skupiają się głównie w miastach Płock, Wyszogród, Drobin, Gąbin oraz w 407 wsiach należących do powiatu. Kompleksy osiedlowe zajmują w powiecie płockim ziemskim 10601 ha (14 114 ha po uwzględnieniu zabudowy miasta Płocka).

W strefie badań sieć dróg publicznych tworzą: drogi krajowe, wojewódzkie, powiatowe i gminne zajmujące powierzchnię 4 528 ha (GUS 2014). Trzy przeprawy mostowe, dwie w Płocku i jedna w Wyszogrodzie zapewniają połączenie terenów położonych po prawej i lewej stronie Wisły.

Użytki rolne zajmują 139 497 ha co stanowi 74,05% powierzchni powiatu płockiego i wyraźnie wskazuje na jego rolniczy charakter (Uchwała 2016, Uchwała 2017). Obszar badań jest bardzo bogaty w gleby wysokich klas, a gruntów chronionych jest około 47% powierzchni użytków rolnych. Grunty chronione to użytki o wysokiej bonitacji, klasy I, II i III, które powstały z gleb pochodzenia mineralnego. Występują one na terenie gmin Bielsk, Drobin, Radzanowo, Słupno i Starożyby. Na obszarze gmin Gąbin, Łąck, Słubice oraz na pozostałych terenach w przyrzeczu Wisły znajdują się użytki rolne klas IV–VI wytworzone z gleb pochodzenia organicznego.

Na sieć hydrograficzną powiatu składają się śródładowe wody powierzchniowe, których powierzchnia wynosi 7 300 ha. Wody płynące to liczne rzeki i kanały. Wisła jest największą z rzek, której długość w granicach powiatu mierzy 69,7 km. Dorzecze Wisły rozwinięte jest asymetrycznie, przeważają prawe dopływy, z których największą powierzchnię zlewni posiada Skrwa Prawa. Duże znaczenie hydrograficzne, gospodarcze i krajobrazowe mają położone na analizowanym obszarze zbiorniki naturalne. Na terenie makroregionu Pradoliny Toruńsko-Eberswaldzkiej znajdują się jeziora: Ciechomickie (37,10 ha), Górskie (45 ha), Łąckie Duże (59,40 ha), Łąckie Małe (35,60 ha), Sendeń (14,2 ha) oraz Zdwońskie (352,50 ha) – największe jezioro w województwie mazowieckim. W wyniku przegradzania dolin rzecznych zaporami wodnymi w celu ochrony przeciwpowodziowej, zapewnienia rezerw przeciwpowodziowych i zwiększenia możliwości retencyjnych zlewni, utworzone zostały sztuczne zbiorniki wodne. Zbiornik Włocławek – inaczej jezioro Włocławskie – oddany do użytkowania w 1970 roku, zlokalizowany jest w północno-zachodniej części powiatu płockiego. Jego całkowita powierzchnia przy maksymalnym poziomie piętrzenia to 7 500 ha. Inne ważne zbiorniki to: Soczewka, Stanowo, Troszyn i Białobrzegi.

Powiat płocki jest jednym z najmniej zalesionych powiatów w województwie mazowieckim. Lasy zajmują 17,0% ogólnej powierzchni obszaru, co stanowi około 32 049 ha (Uchwała 2016, Uchwała 2017). W strukturze własnościowej dominują lasy publiczne Skarbu Państwa o powierzchni 23 104 ha (72,09% ogółu lasów). Zarząd nad lasami Skarbu Państwa sprawują trzy nadleśnictwa: Płock – 7 099,75 ha, Łąck – 9 759,15 ha i Gostynin – 6 245,21 ha. Pozostałe lasy powiatu należą głównie do osób fizycznych, wspólnot gruntowych, gmin, spółdzielni, kościołów i związków wyznaniowych. Ich łączna powierzchnia to 8 426 ha. Gminy leżące na południe od rzeki Wisły, które mogą poszczycić się największą lesistością w powiecie to Nowy Duninów i Łąck, w których lasy stanowią odpowiednio 72% i 49% powierzchni gminy. Ze względu na atrakcyjne warunki przyrodnicze i ukształtowanie terenu, obie gminy mają charakter turystyczno-rekreacyjny. Lesistość w pozostałych gminach jest na poziomie od 2,5% (Radzanowo) do 20,7% powierzchni (miasto i gmina Gąbin). Gminy te mają charakter typowo rolniczy.

Na obszarach leśnych przeważają gleby bielcowe i rdzawe, na których powstały siedliska borów świeżych i mieszanych w ujęciu typologii leśnej (Uchwała 2017). Warunki glebowe i klimatyczne sprawiają, że przeważającymi typami siedliskowymi są bory (60% powierzchni leśnej powiatu), głównie bór świeży (Bśw) i bór mieszany świeży (BMśw). Pozostałe typy siedliskowe lasów to bór mieszany świeży (BMśw), bór mieszany wilgotny (BMw), bór wilgotny (Bw). Tereny dolinne i grunty podmokłe z reguły zajmują lasy liściaste i mieszane: ols (Ol), las mieszany świeży (LMśw), las świeży (Lśw). Duży areał gruntów niskich klas bonitacyjnych powoduje, że głównym gatunkiem lasotwórczym jest sosna zwyczajna, występująca na około 80% powierzchni leśnej. Innymi gatunkami lasotwórczymi są: brzoza, dęby (szypułkowy i bezszypułkowy) oraz olśsza czarna. W domieszce występują: klon jawor, klon zwyczajny, modrzew, grab, lipa, jarzab pospolity, wiąz, jesion. Obcymi gatunkami są dąb czerwony, daglezia, sosna wejmutka czy sosna Banksa. Średni wiek drzewostanów wynosi 55 lat (Uchwała 2017).

### **Materiały teledetekcyjne i ich przetwarzanie**

W niniejszej pracy wykorzystano obrazy satelity Sentinel-2A Europejskiej Agencji Kosmicznej (ESA). Konstelacja Sentinel związana jest z projektem Copernicus – Europejskim Systemem Monitoringu Środowiska. Pierwszy europejski satelita z serii Sentinel został wystrzelony w 2014 roku. Obraz powierzchni Ziemi rejestrowany jest przez satelity Sentinel w zakresie optycznym, termalnym i radarowym ze zróżnicowaną rozdzielczością przestrzenną i wysoką częstotliwością pozyskiwania obrazów. Atuty misji Sentinel-2 to krótszy czas rewizyty, zwiększona rozdzielczość przestrzenna oraz szeroki pas obrazowania w porównaniu do amerykańskiego satelity Landsat 8 (Ryc. 1). Satelita Sentinel-2A został wystrzelony na kołową orbitę heliosferyczną 23 czerwca 2015 roku i osiągnął wysokość 799 km. Wielkość piksela wynosi 10 m w zakresie widzialnym, 20 m w bliskiej podczerwieni, natomiast

60 m w krótkofalowej podczerwieni. Sensor MSI pozyskuje zobrazenia dla pasa o szerokości 290 km, a rejestracja tego samego fragmentu powierzchni Ziemi

odbywa się co 10 dni. Czas rewizyty jest skrócony do 5 dni, jeśli uwzględnić dane pochodzące z dwóch satelitów (Tab. 1).

Tabela 1. Dane techniczne instrumentu MSI satelity Sentinel-2A.

Table 1. Specifications of the Sentinel 2A satellite MSI instrument

Kanał <i>Band</i>	Opis <i>Description</i>	Wielkość piksela <i>Ground sample distance</i> [m]	Długość fali dla środka kanału <i>Central wavelength</i> [nm]
B1	Ultra blue	60	443
B2	Blue	10	490
B3	Green	10	560
B4	Red	10	665
B5	Visible and Near Infrared (VNIR)	20	705
B6	VNIR	20	740
B7	VNIR	20	783
B8	VNIR	10	842
B8a	VNIR	20	865
B9	Short Wave Infrared (SWIR)	60	940
B10	SWIR	60	1375
B11	SWIR	20	1610
B12	SWIR	20	2190

Oznaczenia zakresów promieniowania: *ultra blue* – krótkie niebieskie, *blue* – niebieskie, *green* – zielone, *red* – czerwone, VNIR – widzialne i bliska podczerwień, SWIR – krótkofalowe termalne.

Źródło: [https://sentinel.esa.int/documents/247904/349490/S2\\_SP-1322\\_2.pdf](https://sentinel.esa.int/documents/247904/349490/S2_SP-1322_2.pdf)

Zobrazowania satelitarne pobrano ze strony internetowej Europejskiej Agencji Kosmicznej (ESA)<sup>3</sup>. Wykorzystano produkty Sentinel2Look, którymi są dwie sceny L1C o wymiarach 100 × 100 km z projekcją UTM/WGS84 (*Universal Transverse Mercator / World Geodetic System 1984*) w formacie JPEG2000, który jest skompresowanym plikiem z osadzonymi informacjami georeferencyjnymi. Pobrano sceny posiadające identyfikatory:

- L1C\_T34UDD\_A019953\_20190418T095032,
- L1C\_T34UCD\_A019953\_20190418T095032.

Wybrane obrazy, które zostały zarejestrowane dnia 18 kwietnia 2019 roku, były wolne od pokrywy chmur (Ryc. 1).

W badaniach wykorzystano następujące oprogramowanie:

- **Ewmapa** wersja 12.29 – do selekcji mapy ewidencyjnej, obliczenia powierzchni oraz transformacji mapy do układu UTM/WGS84.
- **PCI Geomatica** – do mozaikowania, przycinania, wyliczania wskaźnika wegetacyjnego NDVI, usunięcia korelacji obrazów, klasyfikacji nadzorowanej, analizy dokładności klasyfikacji oraz fuzji obrazu cyfrowego z danymi mapy ewidencyjnej. Korzystano z instalacji będącej w posiadaniu WSGK w Kutnie, a także z wersji krótkoterminowej 2018 SP2 Full Software (wersja 15 dniowa)<sup>4</sup>.

- **TerrSet** – TerrSet Geospatial Monitoring and Modeling Software, dostępne na Wydziale Nauk Geograficznych Uniwersytetu Łódzkiego w Łodzi (zastosowanie jak wyżej).

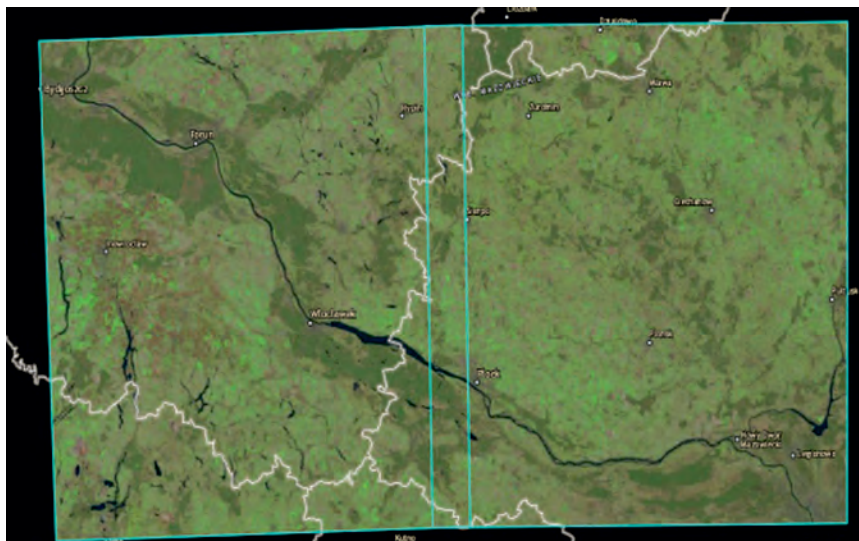
Wyboru pól treningowych dokonano na podstawie (Ryc. 2): kompozycji barwnych – w barwach naturalnych (RGB), tj. z wykorzystaniem kanałów widzialnych (niebieski, zielony, czerwony), kompozycji barwnej z podczerwienią (CIR, *Color Infra Red*), czyli złożonej z dwóch kanałów widzialnych (zielony i czerwony) i bliskiej podczerwieni (NIR, *Near Infra Red*) oraz indeksu spektralnego NDVI (*Normalized Difference Vegetation Index*). Na podstawie wizualnej analizy tych obrazów stwierdzono, że możliwe jest wyróżnienie pięciu klas pokrycia terenu, które w niniejszej pracy nazywamy klasami teledetekcyjnymi (T):

- T1. Lasy i zadrzewienia,
- T2. Wody,
- T3. Obszary zabudowane (drogi oraz zabudowa wiejska, miejska i przemysłowa),
- T4. Grunty rolne bez pokrywy roślinnej,
- T5. Grunty rolne z pokrywą roślinną.

Przy wstępnej wizualnej ocenie pokrycia terenu kierowano się zróżnicowaniem barwnym obrazu oraz jego strukturą i teksturą. Wyodrębnienie gruntów rolnych bez pokrywy roślinnej było konieczne ze względu na stan wielu upraw rolnych (obrazy są z początku sezonu wegetacyjnego i wiele roślin jeszcze nie weszło lub dopiero zaczyna swój rozwój) oraz na stosunkowo duży ich udział w powierzchni powiatu. Gruntami rolnymi

<sup>3</sup> <https://scihub.copernicus.eu>

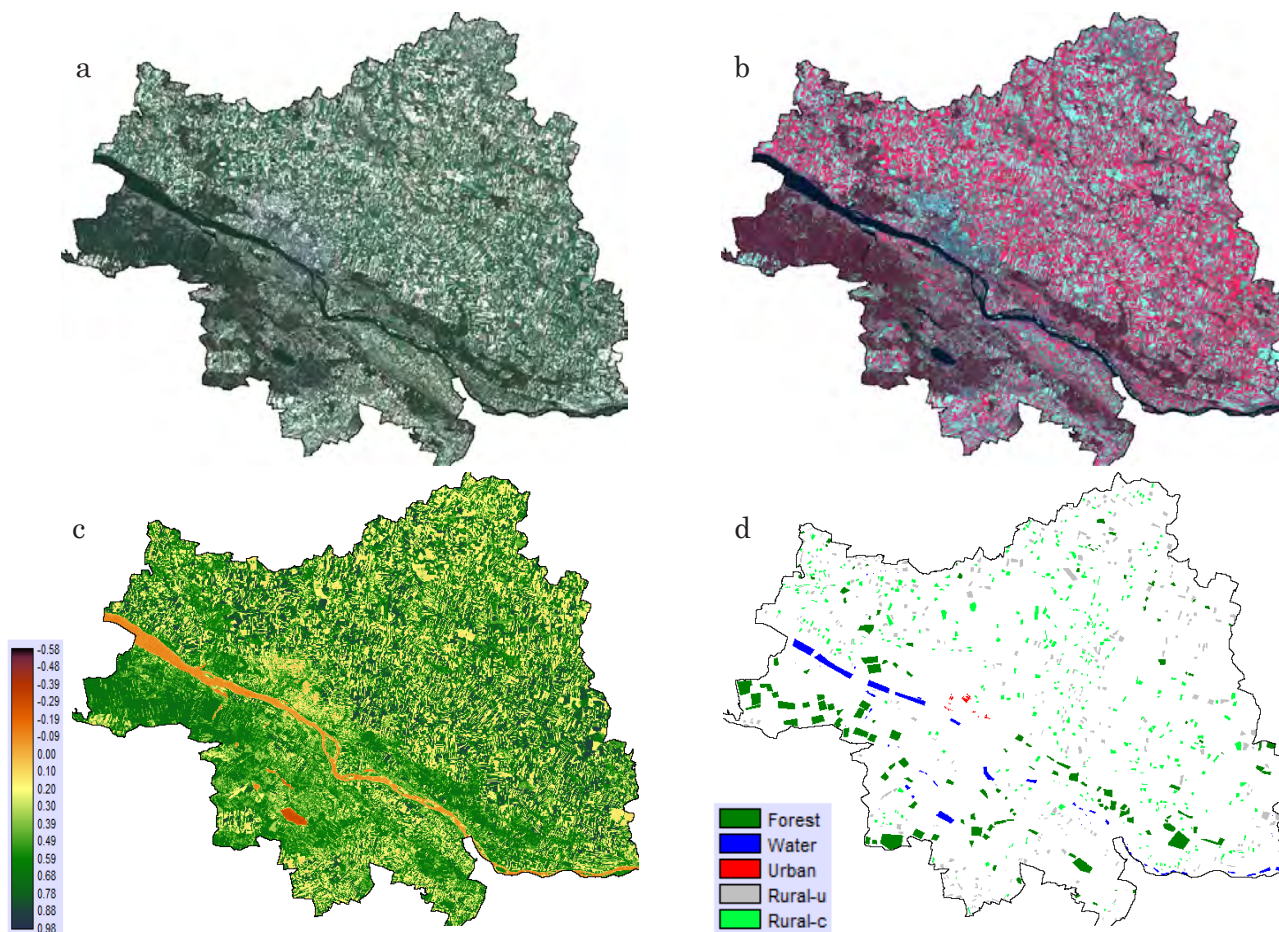
<sup>4</sup> <https://www.pci-geomatics.com>



Ryc. 1. Zasięg obrazów Sentinel-2A zastosowanych w badaniach (<https://scihub.copernicus.eu>)  
 Fig. 1. Coverage of Sentinel-2A images used in the study (<https://scihub.copernicus.eu>)

z pokrywą roślinną są sady, łąki i pastwiska oraz oziminy i inne rozwinięte w tym momencie uprawy roślin. Nie sprawiło większych trudności wyodrębnienie lasów i zadrzewień oraz wód (Wisła i mniejsze rzeki, jeziora, stawy) i obszarów zabudowanych.

Na potrzeby klasyfikacji nadzorowanej dokonano wyboru pól treningowych, które starano się zlokalizować w miarę równomiernie na terenie całego powiatu. Ogółem wyznaczono 961 pól treningowych, które zajmują łącznie 7,37% powierzchni powiatu (Ryc. 2, Tab. 2).

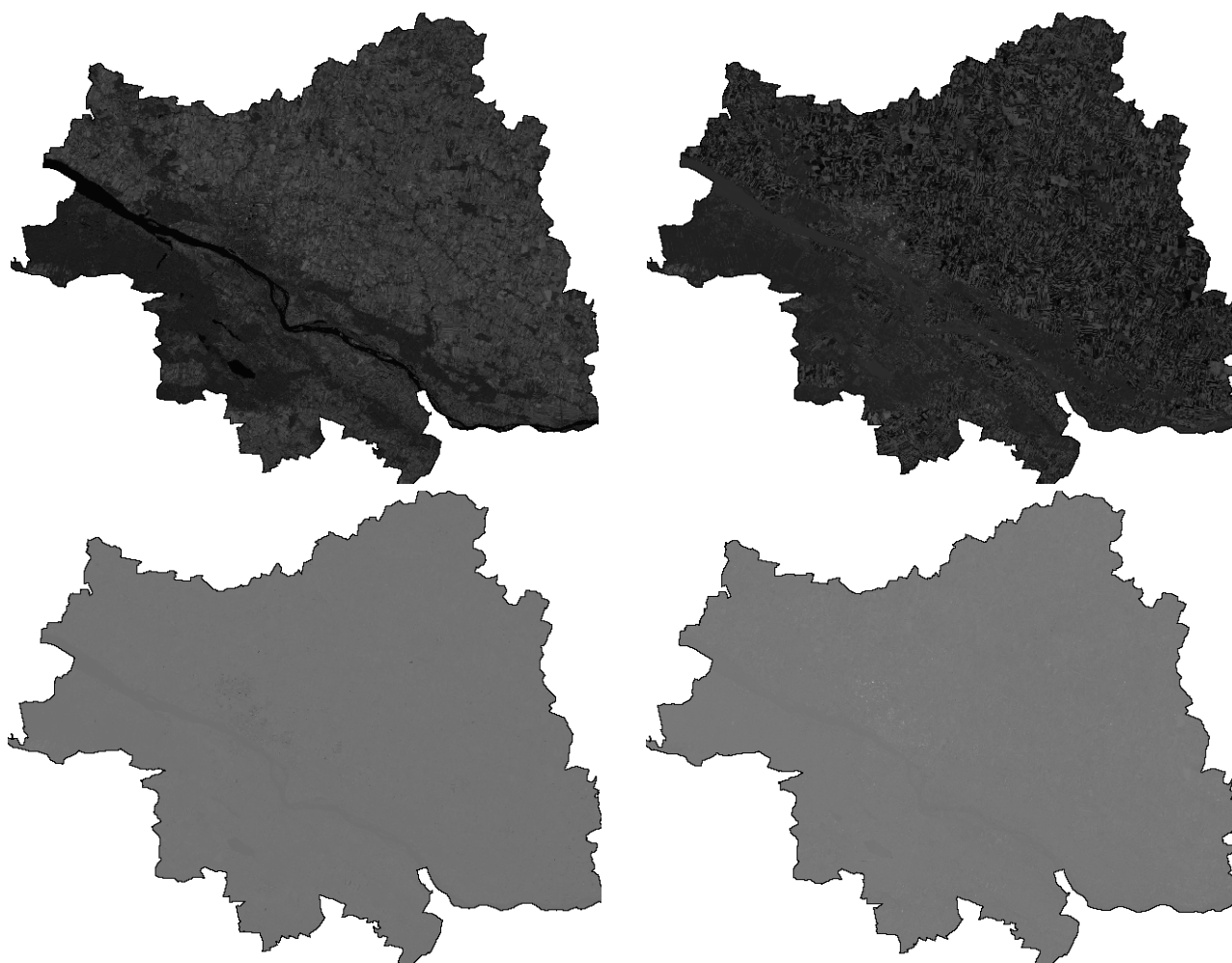


Ryc. 2. Kompozycje oryginalnych kanałów Sentinel-2A wykorzystane w procesie wyboru pól treningowych: (a) w barwach naturalnych – RGB, (b) barwna z podczerwienią – CIR, (c) indeks spektralny NDVI oraz (d) rozmieszczenie pól treningowych  
 Fig. 2. Color composites of original Sentinel-2A bands used to locate training fields: (a) Natural Color RGB composite, (b) Color Infra Red – CIR composite, (c) Normalized Difference Vegetation Index NDVI and (d) distribution of training fields over area of the study

Tabela 2. Charakterystyka pól treningowych

Table 2. Training fields

Teledetekcyjna klasa pokrycia terenu <i>Remote Sensing Land cover class</i>			Pola treningowe <i>Training fields</i>		
Nr <i>No</i>	Nazwa <i>Name</i>	Nazwa skrócona <i>Short name</i>	Liczba <i>Number</i>	Liczba pikseli obrazu <i>No of pixels</i>	Udział w powierzchni powiatu [%] <i>Share in the area [%]</i>
T1	Lasy i zadrzewienia <i>Forests and woodlots</i>	<i>Forest</i>	133	565 710	3,00
T2	Wody <i>Water</i>	<i>Water</i>	67	136 751	0,73
T3	Obszary zabudowane <i>Built-up areas</i>	<i>Urban</i>	13	8 295	0,04
T4	Grunty rolne bez pokrywy roślinnej <i>Arable areas not covered by vegetation</i>	<i>Rural-u</i>	334	335 690	1,78
T5	Grunty rolne z pokrywą roślinną <i>Arable areas covered by vegetation</i>	<i>Rural-c</i>	414	342 780	1,82
Razem / <i>Total</i>			961	1 389 226	7,37



Ryc. 3. Komponenty uzyskane w wyniku przetworzenia czterech kanałów (2, 3, 4, 8) Sentinel-2A za pomocą metody głównych składowych PCA

Fig. 3. Components achieved with PCA procedure from original bands 2, 3, 4 and 8 of Sentinel-2A

Kanały 2, 3, 4, 8 obrazu Sentinel-2A poddano przetworzeniu za pomocą metody głównych składowych (PCA, *Principal Components Analysis*). Uzyskano cztery tzw. komponenty (Ryc. 3). Z raportu zastosowanej procedury wynika, że początkowe komponenty skupiły znaczną część istotnej dla potrzeb klasyfikacji informacji obrazowej, kolejno odpowiednio: 89,370%, 10,533%, 0,072% i 0,025% ogólnej wariancji (zmienności). Taki rezultat sugeruje, że efektywną klasyfikację pokrycia terenu można przeprowadzić z wykorzystaniem kilku komponentów PCA.

Klasyfikację treści obrazów wykonano w kilku wariantach (Tab. 3) z wykorzystaniem dwóch algorytmów: najmniejszej odległości (MD, *Minimum Distance*) oraz największego prawdopodobieństwa (ML, *Maximum Likelihood*). Warianty pierwszy i drugi dotyczą klasyfikacji z wykorzystaniem czterech oryginalnych, nieprzetworzonych kanałów Sentinel-2A. Uzyskane wyniki służą do porównania z dalszymi wariantami, w których zastosowano różną liczbę komponentów uzyskanych w metodzie PCA (odpowiednio 4, 3 lub tylko 2 kolejne komponenty).

Tabela 3. Warianty klasyfikacji z wykorzystaniem algorytmów MD – najmniejszej odległości oraz ML – największego prawdopodobieństwa

Table 3. Classification variants using Minimum Distance (MD) and Maximum Likelihood (ML) algorithms

Wariant klasyfikacji <i>Classification scheme</i>	Wykorzystane obrazy <i>Image data applied</i>	Algorytm <i>Algorithm</i>
1	Oryginalne nieprzetworzone kanały Sentinel-2A	MD
2	<i>Original unprocessed Sentinel-2 bands</i> (2, 3, 4, 8)	ML
3	Komponenty PCA	MD
4	<i>PCA components</i> (1, 2, 3, 4)	ML
5	Komponenty PCA	MD
6	<i>PCA components</i> (1, 2, 3)	ML
7	Komponenty PCA	MD
8	<i>PCA components</i> (1, 2)	ML

### Referencyjne dane Ewidencji Gruntów i Budynków

Właściwą ocenę dokładności wyników klasyfikacji danych teledetekcyjnych można przeprowadzić korzystając z odpowiednich danych referencyjnych – zebranych w ramach samodzielnie przeprowadzonych badań w terenie lub pochodzących z rzetelnych materiałów kartograficznych, baz danych lub danych statystycznych GUS. W niniejszych badaniach jako dane referencyjne wykorzystano numeryczną mapę ewidencji gruntów i budynków (EGiB), z której uzyskano kontury użytków gruntowych. Mapa ewidencji gruntów i budynków powiatu plockiego została pozyskana w dniu 29 lipca 2019 roku z Powiatowego Ośrodka Dokumentacji Geodezyjnej i Kartograficznej w Płocku oraz w dniu 2 lipca 2019 roku, dla części dotyczącej miasta Płocka, z Miejskiego Ośrodka Dokumentacji Geodezyjnej i Kartograficznej w Płocku. Prowadzona jest w geodezyjnym układzie współrzędnych prostokątnych

kątnych płaskich 2000 zwanym „układem 2000”, który oparty jest na odwzorowaniu Gaussa-Krügera dla elipsoidy GRS 80. Mapa ma szczegółowość właściwą dla skali 1:5000, w formacie pliku GML i EMU.

Ujęte na mapie ewidencyjnej użytki (Tab. 4) zgeneralizowano do sześciu klas pokrycia terenu (plus klasa terenów niesklasyfikowanych), które dalej nazywamy klasami pokrycia według EGiB, w skrócie E (Tab. 5, Ryc. 4): E1 – lasy i zadrzewienia, E2 – wody, E3 – tereny zabudowane, E4 – grunty rolne (bez Ł, Ps, S) – z pokrywą roślinną lub bez pokrywy roślinnej, E5 – grunty rolne z pokrywą roślinną (Ł, Ps, S), E6 – nieużytki oraz E0 – tereny niesklasyfikowane.

Wykonano transformację wyodrębnionych użytków z układu 2000/7 na układ UTM/WGS84, a następnie przetransformowano je z postaci wektorowej do rastrowej, odpowiadającej pikselom 10×10 m Sentinel-2A. Transformacja do formatu rastrowego niesie znaczne konsekwencje zarówno w graficznej reprezentacji użytków, jak i w bilansach pól powierzchni (Tab. 6).

Tabela 4. Podział użytków gruntowych w Ewidencji Gruntów i Budynków (Ustawa 1989)

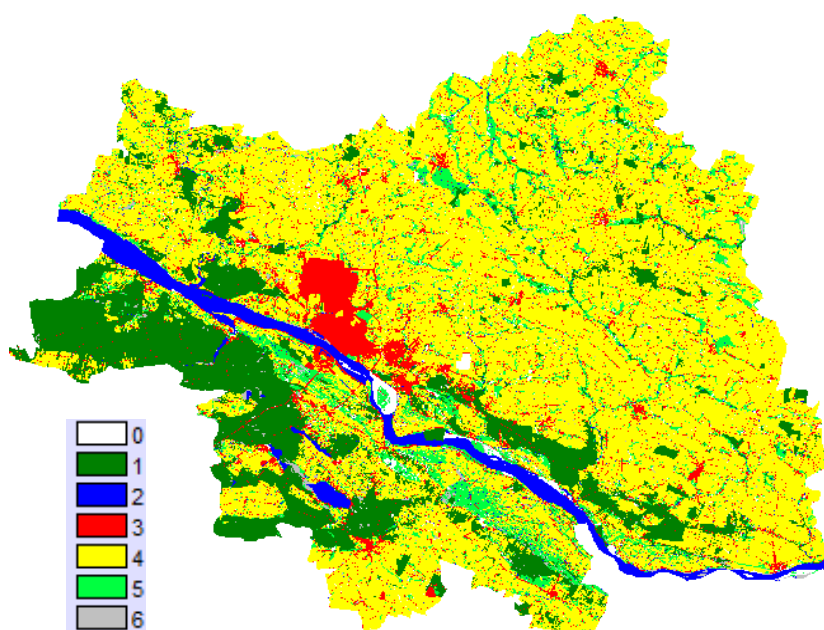
Table 4. Land and Buildings Register (EGiB) land use classes (Ustawa 1989)

Grupa użytku gruntowego <i>Land use class</i>	Podgrupa użytku gruntowego <i>Land use subclass</i>	Rodzaj użytku gruntowego <i>Type of land use</i>	Oznaczenie użytku gruntowego <i>Symbol of land use type</i>	
Grunty rolne	użytki rolne	grunty orne	R	
		sady	S	
		łąki trwałe	Ł	
		pastwiska trwałe	Ps	
		grunty rolne zabudowane	Br	
		grunty pod stawami	Wsr	
		grunty pod rowami	W	
		grunty zadrzewione i zakrzewione na użytkach rolnych	Lzr	
Grunty leśne	nieużytki	–	N	
	lasy	–	Ls	
	grunty zadrzewione i zakrzewione	–	Lz	
Grunty zabudowane i zurbanizowane	tereny mieszkaniowe	–	B	
	tereny przemysłowe	–	Ba	
	inne tereny zabudowane	–	Bi	
	zurbanizowane tereny niezabudowane lub w trakcie zabudowy	–	Bp	
	tereny rekreacyjno-wypoczynkowe	–	Bz	
	użytki kopalne	–	K	
	tereny komunikacyjne	drogi		dr
		tereny kolejowe		Tk
		inne tereny komunikacyjne		Ti
		grunty przeznaczone pod budowę dróg publicznych lub linii kolejowych		Tp
Użytki ekologiczne			E	
Grunty pod wodami	grunty pod morskimi wodami wewnętrznymi	–	Wm	
	grunty pod wodami powierzchniowymi płynącymi	–	Wp	
	grunty pod wodami powierzchniowymi stojącymi	–	Ws	
Tereny różne			Tr	



Tabela 5. Dane referencyjne – użytki gruntowe powiatu plockiego według EGİB i sposób ich podzielenia na 6 klas pokrycia terenu  
 Table 5. Reference data – Land use classes in accordance to Land and Buildings Register (EGİB) database being aggregated into 6 Land cover classes

Oznaczenie użytku gruntowego według EGİB <i>Land use class in accordance to EGİB database</i>	Oznaczenie konturu użytku gruntowego (plik SHP numerycznej mapy ewidencyjnej) <i>Signature on EGİB map</i>	Klasa pokrycia terenu EGİB wynikająca z aglomeracji użytków gruntowych <i>EGİB agglomerated Land cover class</i>
Grunty niesklasyfikowane lub sklasyfikowane błędnie <i>Not clasified or classified incorrectly</i>	–	E0. Tereny niesklasyfikowane <i>Not clasified</i>
B, Bi, Br, Ba, Bz, Bp	B	E3. Tereny zabudowane <i>Built-up</i>
dr	dr	E3. Tereny zabudowane <i>Built-up</i>
K	K	E3. Tereny zabudowane <i>Built-up</i>
Ls	Ls	E1. Lasy i zadrzewienia <i>Forests and woodlots</i>
Lz, Lzr	Lz	E1. Lasy i zadrzewienia <i>Forests and woodlots</i>
Ł_I – VI	Ł	E5. Grunty rolne z pokrywą roślinną <i>Arable land covered by vegetation</i>
N	N	E6. Nieużytki <i>Waste ground</i>
Ps_I – VI	Ps	E5. Grunty rolne z pokrywą roślinną <i>Arable land covered by vegetation</i>
R_I – VI	R	E4. Grunty rolne (bez Ł, Ps, S) – z pokrywą roślinną lub bez pokrywy roślinnej <i>Arable land (excepting Ł-meadows, Ps-pastures, S-orchards) – covered or not covered by vegetation</i>
S	S	E5. Grunty rolne z pokrywą roślinną <i>Arable land covered by vegetation</i>
Tp, Ti, Tk	T	E3. Tereny zabudowane <i>Built-up</i>
W, Wsr, Wp, Ws	W	E2. Wody <i>Water</i>



Ryc. 4. Pokrycie terenu powiatu plockiego uzyskane poprzez połączenie użytków gruntowych wykazanych na mapie ewidencyjnej EGİB, zgodnie z Tabelą 5. Mapa stanowi warstwę referencyjną służącą do porównania i oceny wyników wykonanej klasyfikacji obrazów. Objaśnienia: 0 – tereny niesklasyfikowane, 1 – lasy i zadrzewienia, 2 – wody, 3 – tereny zabudowane, 4 – grunty rolne z pokrywą roślinną lub bez pokrywy roślinnej, 5 – grunty rolne z pokrywą roślinną, 6 – nieużytki

Fig. 4. EGİB Land cover classes of Plock district achieved through aggregation of land use classes from EGİB database, in acc. to Table 5. This is reference layer used to compare and assess results achieved with diverse classification methods. Explanations: 0 – not clasified, 1 – forests and woodlots, 2 – water, 3 – built-up, 4 – arable land covered or not covered by vegetation, 5 – arable land covered by vegetation, 6 – waste ground

Tabela 6. Powierzchnia klas pokrycia terenu powiatu plockiego ustalona na podstawie mapy ewidencyjnej EGİB i danych GUS (GUS 2014)

Table 6. Area covered by EGİB Land cover classes and (GUS 2014) data

Klasa pokrycia terenu utworzona przez aglomerację użytków gruntowych EGİB <i>Land cover class achieved through aggregation of land use classes from EGİB database</i>	Pole powierzchni <i>Area</i>					
	według warstw wektorowych EGİB po ich przetworzeniu do postaci rastrowej odpowiadającej pikselom 10 × 10 m Sentinel-2A <i>Area calculated based on Land and Buildings Register vector data being transformed into raster data corresponding with 10 × 10 m Sentinel-2A pixels</i>		według warstw wektorowych EGİB <i>Area calculated based on Land and Buildings Register vector data</i>		na podstawie danych GUS (GUS 2014) <i>Area calculated based on (GUS 2014) data</i>	
	[km <sup>2</sup> ]	[%]	[km <sup>2</sup> ]	[%]	[km <sup>2</sup> ]	[%]
E0. Tereny niesklasyfikowane <i>Not classified</i>	27,0314	1,4	4,5651	0,2	0	0
E1. Lasy i zadrzewienia <i>Forests and woodlots</i>	312,2726	16,6	329,0631	17,5	326,1900	17,3
E2. Wody <i>Water</i>	74,2479	3,9	77,4153	4,1	82,4000	4,4
E3. Tereny zabudowane <i>Built-up</i>	147,6517	7,8	150,5114	8,0	141,1100	7,5
E4. Grunty rolne (bez L, Ps, S) – z pokrywą roślinną lub bez pokrywy roślinnej <i>Arable land (excepting: L-meadows, Ps-pastures, S-orchards) – covered or not covered by vegetation</i>	1 143,4709	60,7	1 129,5296	59,9	1 145,6900	60,8
E5. Grunty rolne z pokrywą roślinną (L, Ps, S) <i>Arable land covered by vegetation (i.e L-meadows, Ps-pastures and S-orchards)</i>	157,8301	8,4	171,9861	9,1	167,3500	8,9
E6. Nieużytki <i>Waste ground</i>	21,2495	1,1	21,7139	1,2	22,0300	1,1
Razem <i>Total</i>	1 883,7541	100,0	1 884,7845	100,0	1 884,7700	100,0

## Wyniki i ich omówienie

### Wyniki różnych wariantów klasyfikacji

Uzyskano różniące się wizualnie rezultaty klasyfikacji (Ryc. 5–8), co jest także wyraźnie widoczne w zestawieniach pól powierzchni klas pokrycia terenu (Tab. 7). Znacząco różnią się wyniki uzyskane za po-

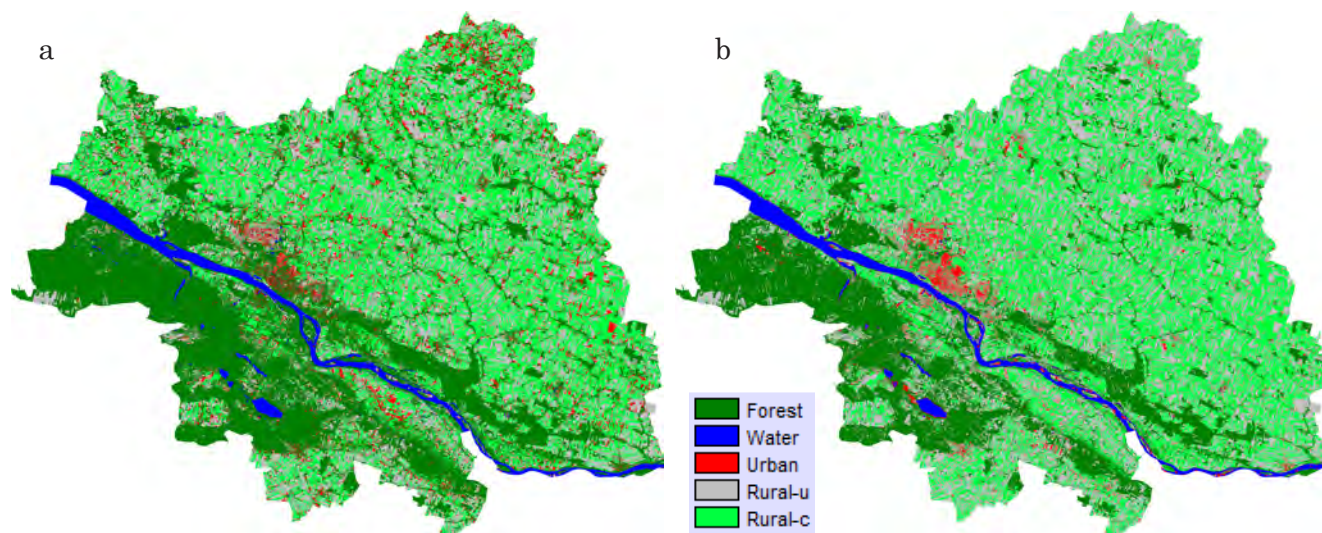
mocą metody MD od wyników metody ML. Natomiast mimo zastosowania różnej liczby komponentów PCA, uzyskane wyniki klasyfikacji zmieniają się nieznacznie.

Z analizy zawartości Tabeli 8 i 9 (klasyfikacja kanałów oryginalnych) wynika, że z klasy E1 (lasy i zadrzewienia) najwięcej pikseli zostało zaliczonych do klasy T1 pokrycia terenu, tj. lasów i zadrzewień. Wody

(E2) w większości zostały zaklasyfikowane do wód (T2), lecz znaczna ich część trafiła do klasy T1, tj. lasów i zadrzewień. Zabudowa (E3) w dużej części została sklasyfikowana do T1, czyli lasów (metoda MD, Tab. 8) lub do T4 – gruntów bez pokrywy roślinnej (Tab. 9, metoda ML, kl. 4). Grunty rolne (bez Ł, Ps, S), tj. klasa E4 według EGİB, zawierająca grunty z pokrywą roślinną lub bez pokrywy roślinnej, zostały przyporządkowane za pomocą obydwu metod głównie do klasy T4 (grunty bez pokrywy roślinnej) oraz T5 (grunty z pokrywą roślinną). Klasa E5 (grunty rolne z pokrywą, tj. łąki,

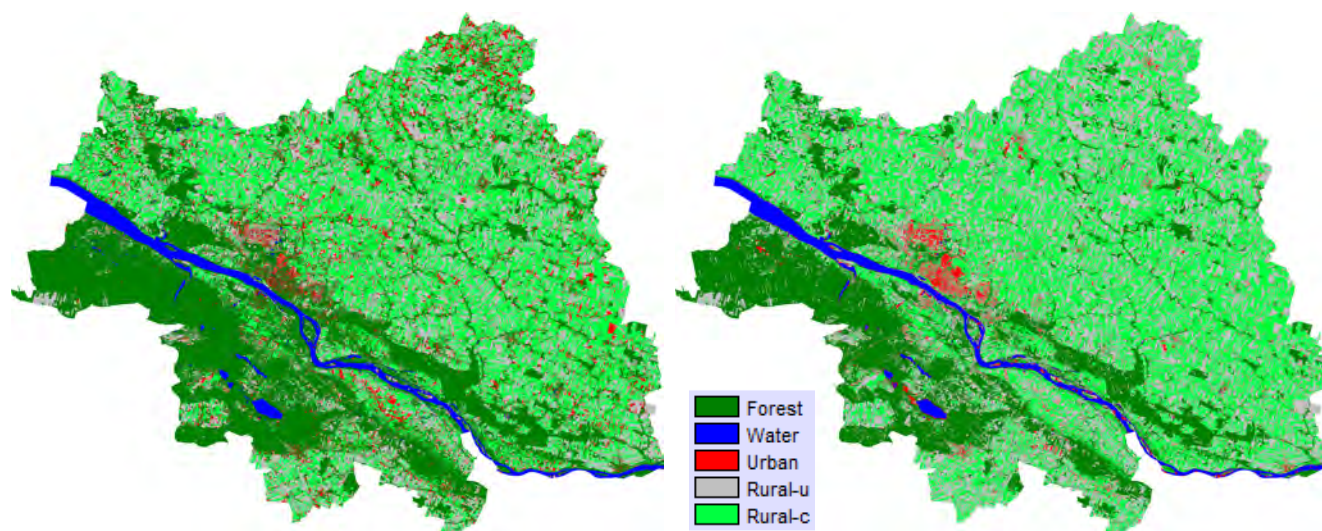
pastwiska i sady) została wliczona głównie do klasy T1 (lasy i zadrzewienia) oraz T5 (grunty rolne z pokrywą roślinną), natomiast klasa E6 (nieużytki) w znacznym stopniu została zakwalifikowana jako lasy i zadrzewienia, co świadczy o tym, że nieużytki są w znacznym stopniu pokryte roślinnością.

Przedstawione prawidłowości powtórzyły się w pozostałych wariantach klasyfikacji, tj. realizowanych za pomocą metod MD oraz ML, z wykorzystaniem zredukowanej liczby komponentów otrzymanych w metodzie PCA (Tab. 10–15).



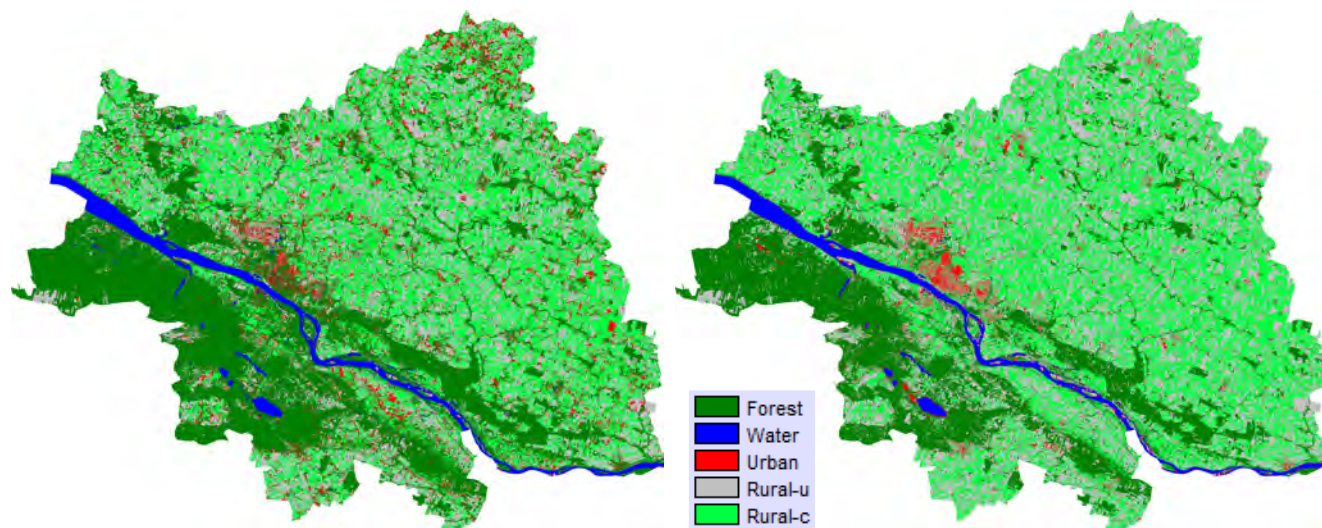
Ryc. 5. Wyniki klasyfikacji oryginalnych kanałów (2, 3, 4, 8) obrazu Sentinel-2A za pomocą metod: (a) najmniejszej odległości, (b) największego prawdopodobieństwa

Fig. 5. Classification results using original bands 2, 3, 4 and 8 of Sentinel-2A and (a) Minimum Distance, (b) Maximum Likelihood algorithms



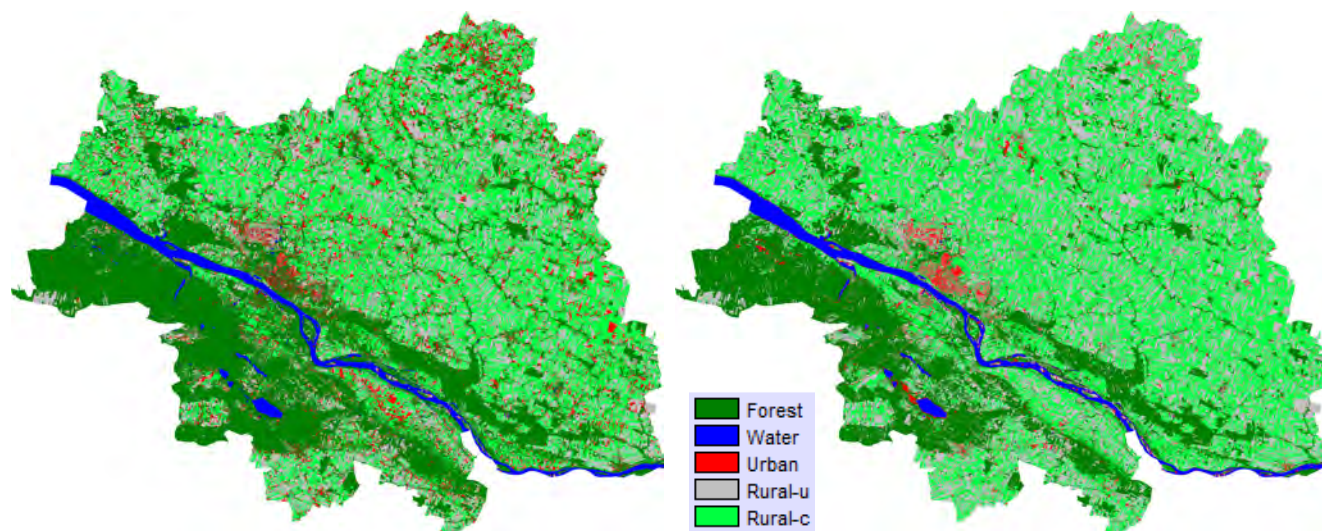
Ryc. 6. Wyniki klasyfikacji z wykorzystaniem czterech pierwszych (1, 2, 3, 4) komponentów uzyskanych za pomocą metody PCA z przetworzenia kanałów (2, 3, 4, 8) obrazu Sentinel-2A

Fig. 6. Classification results using four first components (1, 2, 3, 4) achieved with PCA procedure from original bands 2, 3, 4 and 8 of Sentinel-2A



Ryc. 7. Wyniki klasyfikacji z wykorzystaniem trzech pierwszych (1, 2, 3) komponentów uzyskanych za pomocą metody PCA z przetworzenia kanałów (2, 3, 4, 8) obrazu Sentinel-2A

*Fig. 7. Classification results using three first components (1, 2, 3) achieved with PCA procedure from original bands 2, 3, 4 and 8 of Sentinel-2A*



Ryc. 8. Wyniki klasyfikacji z wykorzystaniem dwóch pierwszych (1, 2) komponentów uzyskanych za pomocą metody PCA z przetworzenia kanałów (2, 3, 4, 8) obrazu Sentinel-2A

*Fig. 8. Classification results using two first components (1, 2) achieved with PCA procedure from original bands 2, 3, 4 and 8 of Sentinel-2A*

### Porównanie wyników klasyfikacji z danymi referencyjnymi EGİB

Porównanie wyników klasyfikacji i danych referencyjnych polega na sprawdzeniu, do jakich teledetekcyjnych klas pokrycia terenu (ujętych w Tab. 2) zostały zaliczone grunty należące do klas pokrycia terenu EGİB utwo-

rzonych poprzez połączenie użytków gruntowych EGİB wskazanych w Tab. 5 oraz na Ryc. 4. Wyniki zestawiono według zastosowanych metod klasyfikacji (MD, ML) oraz użytych obrazów (obrazy oryginalne lub różna liczba komponentów uzyskanych w procedurze PCA) (Tab. 8–15).

Tabela 7. Wyniki klasyfikacji pokrycia terenu dla powiatu plockiego (objaśnienia wariantów podano w Tabeli 3)  
 Table 7. Classification result of Płock district land cover (for classification schemas refer to Table 3)

Wariant klasyfikacji wg Tabeli 3 <i>Classification scheme acc. to Table 3</i>	Teledetekcyjne klasy pokrycia terenu <i>Remote Sensing Land Cover Classes</i>									
	T1. Lasy i zadrzewienia <i>Forests and Woodlots</i>		T2. Wody <i>Water (rivers and reservoirs)</i>		T3. Tereny zabudo- wane <i>Built-up areas</i>		T4. Grunty rolne bez pokrywy roślinnej <i>Arable areas not covered by vegetation</i>		T5. Grunty rolne z pokrywą roślinną <i>Arable areas covered by vegetation</i>	
	[km <sup>2</sup> ]	[%]	[km <sup>2</sup> ]	[%]	[km <sup>2</sup> ]	[%]	[km <sup>2</sup> ]	[%]	[km <sup>2</sup> ]	[%]
1	653,55	34,7	58,16	3,1	85,10	4,5	495,71	26,3	591,24	31,4
2	446,90	23,7	50,78	2,7	43,83	2,3	631,01	33,5	711,25	37,8
3	653,54	34,7	58,16	3,1	85,10	4,5	495,70	26,3	591,25	31,4
4	446,86	23,7	50,78	2,7	43,66	2,3	631,24	33,5	711,22	37,8
5	653,22	34,7	58,17	3,1	86,69	4,6	494,40	26,2	591,26	31,4
6	443,23	23,5	50,83	2,7	43,25	2,3	621,53	33,0	724,92	38,5
7	649,87	34,5	58,23	3,1	103,13	5,5	481,30	25,6	591,23	31,4
8	455,22	24,2	50,80	2,7	44,45	2,4	593,33	31,5	739,96	39,3

Tabela 8. Porównanie klas pokrycia terenu EGiB (dane referencyjne) z klasami utworzonymi podczas klasyfikacji oryginalnych kanałów (2, 3, 4, 8) Sentinel-2A, z wykorzystaniem algorytmu MD

Table 8. Comparison between EGiB reference data and classification results of original unprocessed Sentinel-2A bands (2, 3, 4, 8) achieved using MD algorithm

Teledetekcyjna klasa pokrycia terenu, Ryc. 5a <i>Remote Sensing Land cover class, Fig. 5a</i>	Jedn. miary <i>Unit</i>	Dane referencyjne (klasy pokrycia terenu według EGiB, Ryc. 4) <i>Reference data (Land cover classes from EGiB database, Fig. 4)</i>						
		E1	E2	E3	E4	E5	E6	Razem <i>Total</i>
T1	Pixels	3050854	174881	730963	1499185	739821	155697	6351401
	%	16.431	0.942	3.937	8.074	3.985	0.839	34.208
T2	Pixels	16587	495359	11562	4640	7477	20554	556179
	%	0.089	2.668	0.062	0.025	0.040	0.111	2.995
T3	Pixels	11193	8829	189294	560975	71137	3333	844761
	%	0.060	0.048	1.020	3.021	0.383	0.018	4.550
T4	Pixels	33944	35425	424635	4187660	224703	19484	4925851
	%	0.183	0.191	2.287	22.554	1.210	0.105	26.530
T5	Pixels	10148	27985	120063	5182249	535163	13427	5889035
	%	0.055	0.151	0.647	27.911	2.882	0.072	31.717
Razem <i>Total</i>	Pixels	3122726	742479	1476517	11434709	1578301	212495	18567227
	%	16.818	3.999	7.952	61.585	8.500	1.144	100.000

Tabela 9. Porównanie klas pokrycia terenu powstałych z połączenia użytków gruntowych EGiB (dane referencyjne) z klasami utworzonymi podczas klasyfikacji oryginalnych kanałów (2, 3, 4, 8) Sentinel-2A, z wykorzystaniem algorytmu ML

Table 9. Comparison between EGiB reference data and classification results of original unprocessed Sentinel-2A bands (2, 3, 4, 8) achieved using ML algorithm

Teledetekcyjna klasa pokrycia terenu, Ryc. 5b <i>Remote Sensing Land cover class, Fig. 5b</i>	Jedn. miary <i>Unit</i>	Dane referencyjne (klasy pokrycia terenu według EGiB, Ryc. 4) <i>Reference data (Land cover classes from EGiB database, Fig. 4)</i>						Razem <i>Total</i>
		E1	E2	E3	E4	E5	E6	
T1	Pixels	2 884 411	84 371	226 940	670 842	366 538	108 399	4 341 501
	%	15.535	0.454	1.222	3.613	1.974	0.584	23.383
T2	Pixels	6 491	469 054	1 983	1 108	2 312	6 143	487 091
	%	0.035	2.526	0.011	0.006	0.012	0.033	2.623
T3	Pixels	16 916	51 322	226 543	69 846	33 995	26 852	425 474
	%	0.091	0.276	1.220	0.376	0.183	0.145	2.292
T4	Pixels	163 087	63 139	723 322	4 877 718	390 752	39 480	6 257 498
	%	0.878	0.340	3.896	26.271	2.105	0.213	33.702
T5	Pixels	51 821	74 593	297 729	5 815 195	784 704	31 621	7 055 663
	%	0.279	0.402	1.604	31.320	4.226	0.170	38.001
Razem <i>Total</i>	Pixels	3 122 726	742 479	1 476 517	11 434 709	1 578 301	212 495	18 567 227
	%	16.818	3.999	7.952	61.585	8.500	1.144	100.000

Tabela 10. Porównanie klas pokrycia terenu powstałych z połączenia użytków gruntowych EGiB (dane referencyjne) z klasami utworzonymi podczas klasyfikacji, za pomocą algorytmu MD, komponentów (1, 2, 3, 4) uzyskanych w metodzie głównych składowych PCA

Table 10. Comparison between EGiB reference data and classification of PCA components (1, 2, 3, 4) using MD algorithm

Teledetekcyjna klasa pokrycia terenu, Ryc. 6a <i>Remote Sensing Land cover class, Fig. 6a</i>	Jedn. miary <i>Unit</i>	Dane referencyjne (klasy pokrycia terenu według EGiB, Ryc. 4) <i>Reference data (Land cover classes from EGiB database, Fig. 4)</i>						Razem <i>Total</i>
		E1	E2	E3	E4	E5	E6	
T1	Pixels	3 050 855	174 879	730 949	1 499 143	739 797	155 698	6 351 321
	%	16.431	0.942	3.937	8.074	3.984	0.839	34.207
T2	Pixels	16 585	495 358	11 561	4 640	7 477	20 552	556 173
	%	0.089	2.668	0.062	0.025	0.040	0.111	2.995
T3	Pixels	11 194	8 829	189 301	561 011	71 140	3 334	844 809
	%	0.060	0.048	1.020	3.022	0.383	0.018	4.550
T4	Pixels	33 943	35 427	424 633	4 187 590	224 699	19 484	4 925 776
	%	0.183	0.191	2.287	22.554	1.210	0.105	26.529
T5	Pixels	10 149	27 986	120 073	5 182 325	535 188	13 427	5 889 148
	%	0.055	0.151	0.647	27.911	2.882	0.072	31.718
Razem <i>Total</i>	Pixels	3 122 726	742 479	1 476 517	11 434 709	1 578 301	212 495	18 567 227
	%	16.818	3.999	7.952	61.585	8.500	1.144	100.000

Tabela 11. Porównanie klas pokrycia terenu powstałych z połączenia użytków gruntowych EGiB (dane referencyjne) z klasami utworzonymi podczas klasyfikacji, za pomocą algorytmu ML, komponentów (1, 2, 3, 4) uzyskanych w metodzie głównych składowych PCA

Table 11. Comparison between EGiB reference data and classification of PCA components (1, 2, 3, 4) using ML algorithm

Teledetekcyjna klasa pokrycia terenu, Ryc. 6b <i>Remote Sensing Land cover class, Fig. 6b</i>	Jedn. miary <i>Unit</i>	Dane referencyjne (klasy pokrycia terenu według EGiB, Ryc. 4) <i>Reference data (Land cover classes from EGiB database, Fig. 4)</i>						Razem <i>Total</i>
		E1	E2	E3	E4	E5	E6	
T1	Pixels	2 884 303	84 354	226 944	670 763	366 384	108 402	4 341 150
	%	15.534	0.454	1.222	3.613	1.973	0.584	23.381
T2	Pixels	6 491	469 051	1 983	1 108	2 312	6 142	487 087
	%	0.035	2.526	0.011	0.006	0.012	0.033	2.623
T3	Pixels	16 883	51 247	225 700	69 225	33 887	26 834	423 776
	%	0.091	0.276	1.216	0.373	0.183	0.145	2.282
T4	Pixels	163 205	63 279	724 478	4 878 425	390 943	39 502	6 259 832
	%	0.879	0.341	3.902	26.274	2.106	0.213	33.714
T5	Pixels	51 844	74 548	297 412	5 815 188	784 775	31 615	7 055 382
	%	0.279	0.402	1.602	31.320	4.227	0.170	37.999
Razem <i>Total</i>	Pixels	3 122 726	742 479	1 476 517	11 434 709	1 578 301	212 495	18 567 227
	%	16.818	3.999	7.952	61.585	8.500	1.144	100.000

Tabela 12. Porównanie klas pokrycia terenu powstałych z połączenia użytków gruntowych EGiB (dane referencyjne) z klasami utworzonymi podczas klasyfikacji, za pomocą algorytmu MD, komponentów (1, 2, 3) uzyskanych w metodzie głównych składowych PCA

Table 12. Comparison between EGiB reference data and classification of PCA components (1, 2, 3) using MD algorithm

Teledetekcyjna klasa pokrycia terenu, Ryc. 7a <i>Remote Sensing Land cover class, Fig. 7a</i>	Jedn. miary <i>Unit</i>	Dane referencyjne (klasy pokrycia terenu według EGiB, Ryc. 4) <i>Reference data (Land cover classes from EGiB database, Fig. 4)</i>						Razem <i>Total</i>
		E1	E2	E3	E4	E5	E6	
T1	Pixels	3 050 713	174 778	730 454	1 497 240	739 372	155 637	6 348 194
	%	16.431	0.941	3.934	8.064	3.982	0.838	34.190
T2	Pixels	16 627	495 340	11 624	4 646	7 495	20 579	556 311
	%	0.090	2.668	0.063	0.025	0.040	0.111	2.996
T3	Pixels	11 431	9 143	190 782	573 227	72 499	3 443	860 525
	%	0.062	0.049	1.028	3.087	0.390	0.019	4.635
T4	Pixels	33 805	35 230	423 576	4 177 141	223 731	19 407	4 912 890
	%	0.182	0.190	2.281	22.497	1.205	0.105	26.460
T5	Pixels	10 150	27 988	120 081	5 182 455	535 204	13 429	5 889 307
	%	0.055	0.151	0.647	27.912	2.883	0.072	31.719
Razem <i>Total</i>	Pixels	3 122 726	742 479	1 476 517	11 434 709	1 578 301	212 495	18 567 227
	%	16.818	3.999	7.952	61.585	8.500	1.144	100.000

Tabela 13. Porównanie klas pokrycia terenu powstałych z połączenia użytków gruntowych EGiB (dane referencyjne) z klasami utworzonymi podczas klasyfikacji, za pomocą algorytmu ML, komponentów (1, 2, 3) uzyskanych w metodzie głównych składowych PCA

Table 13. Comparison between EGiB reference data and classification of PCA components (1, 2, 3) using ML algorithm

Teledetekcyjna klasa pokrycia terenu, Ryc. 7b <i>Remote Sensing Land cover class, Fig. 7b</i>	Jedn. miary <i>Unit</i>	Dane referencyjne (klasy pokrycia terenu według EGiB, Ryc. 4) <i>Reference data (Land cover classes from EGiB database, Fig. 4)</i>						Razem <i>Total</i>
		E1	E2	E3	E4	E5	E6	
T1	Pixels	2 881 473	87 149	222 156	651 261	353 683	107 515	4 303 237
	%	15.519	0.469	1.196	3.508	1.905	0.579	23.177
T2	Pixels	6 639	469 691	2 139	1 029	2 260	5 947	487 705
	%	0.036	2.530	0.012	0.006	0.012	0.032	2.627
T3	Pixels	19 526	44 847	209 760	76 157	41 876	28 460	420 626
	%	0.105	0.242	1.130	0.410	0.226	0.153	2.265
T4	Pixels	169 713	71 591	688 131	4 816 156	3 759 29	40 453	6 161 973
	%	0.914	0.386	3.706	25.939	2.025	0.218	33.187
T5	Pixels	45 375	69 201	354 331	5 890 106	804 553	30 120	7 193 686
	%	0.244	0.373	1.908	31.723	4.333	0.162	38.744
Razem <i>Total</i>	Pixels	3 122 726	742 479	1 476 517	11 434 709	1 578 301	212 495	18 567 227
	%	16.818	3.999	7.952	61.585	8.500	1.144	100.000

Tabela 14. Porównanie klas pokrycia terenu powstałych z połączenia użytków gruntowych EGiB (dane referencyjne) z klasami utworzonymi podczas klasyfikacji, za pomocą algorytmu MD, komponentów (1, 2) uzyskanych w metodzie głównych składowych PCA

Table 14. Comparison between EGiB reference data and classification of PCA components (1, 2) using MD algorithm

Teledetekcyjna klasa pokrycia terenu, Ryc. 8a <i>Remote Sensing Land cover class, Fig. 8a</i>	Jedn. miary <i>Unit</i>	Dane referencyjne (klasy pokrycia terenu wg EGiB, Ryc. 4) <i>Reference data (Land cover classes from EGiB database, Fig. 4)</i>						Razem <i>Total</i>
		E1	E2	E3	E4	E5	E6	
T1	Pixels	3 048 833	174 069	725 542	1 476 453	734 936	155 208	6 315 041
	%	16.421	0.938	3.908	7.952	3.958	0.836	34.012
T2	Pixels	16 874	495 413	11 558	4 691	7 571	20 744	556 851
	%	0.091	2.668	0.062	0.025	0.041	0.112	2.999
T3	Pixels	15 959	10 938	202 196	704 619	85 482	4 524	1 023 718
	%	0.086	0.059	1.089	3.795	0.460	0.024	5.514
T4	Pixels	30 910	34 078	417 179	4 066 786	215 147	18 591	4 782 691
	%	0.166	0.184	2.247	21.903	1.159	0.100	25.759
T5	Pixels	10 150	27 981	120 042	5 182 160	535 165	13 428	5 888 926
	%	0.055	0.151	0.647	27.910	2.882	0.072	31.717
Razem <i>Total</i>	Pixels	3 122 726	742 479	1 476 517	11 434 709	1 578 301	212 495	18 567 227
	%	16.818	3.999	7.952	61.585	8.500	1.144	100.000



Tabela 15. Porównanie klas pokrycia terenu powstałych z połączenia użytków gruntowych EGiB (dane referencyjne) z klasami utworzonymi podczas klasyfikacji, za pomocą algorytmu ML, komponentów (1, 2) uzyskanych w metodzie głównych składowych PCA

Table 15. Comparison between EGiB reference data and classification of PCA components (1, 2) using ML algorithm

Teledetekcyjna klasa pokrycia terenu, Ryc. 8b <i>Remote Sensing Land cover class, Fig. 8b</i>	Jedn. miary <i>Unit</i>	Dane referencyjne (klasy pokrycia terenu wg EGiB, Ryc. 4) <i>Reference data (Land cover classes from EGiB database, Fig. 4)</i>						
		E1	E2	E3	E4	E5	E6	Razem <i>Total</i>
T1	Pixels	2876874	99885	301854	672447	355553	106585	4413198
	%	15.494	0.538	1.626	3.622	1.915	0.574	23.769
T2	Pixels	6572	469344	2271	1167	2223	5935	487512
	%	0.035	2.528	0.012	0.006	0.012	0.032	2.626
T3	Pixels	36933	48290	163871	90366	56294	35890	431644
	%	0.199	0.260	0.883	0.487	0.303	0.193	2.325
T4	Pixels	148943	58970	669593	4658890	319091	33443	5888930
	%	0.802	0.318	3.606	25.092	1.719	0.180	31.717
T5	Pixels	53404	65990	338928	6011819	845160	30642	7345943
	%	0.288	0.355	1.825	32.379	4.552	0.165	39.564
Razem <i>Total</i>	Pixels	3122726	742479	1476517	11434689	1578321	212495	18567227
	%	16.818	3.999	7.952	61.585	8.500	1.144	100.000

Z analizowanych klas pokrycia utworzonych według EGiB, najlepiej były klasyfikowane piksele klasy E1 (las i zadrzewienia), gdyż biorąc pod uwagę wszystkie warianty, w 92,1–97,7% zostały zaliczone do klasy lasów i zadrzewień. Rozkład wyników klasyfikacji klasy E2 (wody) można tłumaczyć udziałem małych cieków, wzdłuż których z reguły są położone tereny zadrzewione. Korony drzew w takich przypadkach w dużej części zasłaniają lustro wody i wypadkowe charakterystyki spektralne są podobne do charakterystyk zadrzewień. Podobnie jest z gruntami zabudowanymi (klasa E3), które zapewne z tych samych powodów były klasyfikowane głównie albo do lasów i zadrzewień (kl. T1) albo do gruntów bez pokrywy roślinnej (kl. T4). Wyniki klasyfikacji klasy E5 są też w zasadzie oczywiste – ponieważ są to grunty rolne z pokrywą roślinną (łąki, pastwiska, sady), wykazywały powinowactwo spektralne do lasów i zadrzewień (kl. T1) oraz do gruntów rolnych z pokrywą roślinną (kl. T5). Podobnie jest z nieużytkami (E6), które jako w większości pokryte zadrzewieniami, zostały zaliczone głównie do kl. T1 (las i zadrzewienia). Grunty rolne (bez Ł, Ps, S), tj. klasa E4 według EGiB, w momencie wykonywania obrazów satelitarnych składały się z gruntów z pokrywą roślinną (oziminy i już rozwinięte inne uprawy) lub bez pokrywy roślinnej, tj. z jeszcze odsłoniętą glebą. Jak wspomniano, nie było możliwości utworzenia na podstawie danych EGiB oddzielnych warstw referencyjnych dla tych zasadniczo różnych (z perspektywy satelitarnej) gruntów. Znalazło

to odzwierciedlenie w wynikach klasyfikacji – grunty zostały zaliczone do odpowiadających im spektralnie klas: T4 (grunty bez pokrywy roślinnej) oraz T5 (grunty z pokrywą roślinną).

### Dyskusja wyników i wnioski

Należy zwrócić uwagę na istotne utrudnienia dla wykonywanych analiz, polegające na niepełnej zgodności zastosowanych w klasyfikacji obrazów tzw. teledetekcyjnych (T) klas pokrycia terenu oraz klas użytków gruntowych (E) według EGiB. Do klasy E5 mapy referencyjnej, tj. gruntów rolnych z pokrywą roślinną, wprost zaliczyć można było jedynie łąki, pastwiska oraz sady. Nie było możliwości wskazania w EGiB tych gruntów rolnych, które były pokryte w momencie wykonywania obrazów satelitarnych ozimiami i roślinami uprawnymi będącymi w zaawansowanym stadium rozwoju (a więc powinny być zaliczone do klasy E5). Tak samo nie było możliwości wskazania w danych EGiB tych gruntów rolnych, które w tym czasie nie miały pokrywy roślinnej.

Dalsze utrudnienia wynikały z braków lub błędów w EGiB polegających na niesklasyfikowaniu części gruntów lub zaliczeniu niektórych równocześnie do różnych rodzajów użytków. Obszary te (1,4% powierzchni powiatu) wyeliminowano z dalszych analiz. W klasie E6 znalazły się nieużytki, o których wyłączeniu do osobnej

klasy zdecydował brak informacji o ich pokryciu. Dla tej kategorii spodziewane było w większości pokrycie roślinnością drzewiastą i krzewami, a więc o charakterze zbliżonym do lasów.

Istotna jest także odmienna rozdzielczość przestrzenna porównywanych materiałów. Mapa EGiB wykonywana jest w skali 1:5000, podczas gdy zasób informacyjny obrazów o pikselach wielkości 10×10 m pozwala na konstruowanie map o skali co najwyżej 1:100 000 (Będkowski, Piekarski 2017: 34–36). Przetworzenie mapy EGiB z formatu wektorowego na rastrowy, odpowiadający rozdzielczości obrazów satelitarnych, prowadzi do zgubienia wielu obiektów o małej powierzchni lub o małej szerokości (głównie dróg i mniejszych cieków wodnych). Jest to wyraźnie widoczne w danych z Tabeli 6 i dotyczy praktycznie wszystkich klas EGiB: E1 – lasy i zadrzewienia, E2 – wody, E3 – tereny zabudowane, E5 – grunty orne z pokrywą roślinną oraz E6 – nieużytki. Terenów niesklasyfikowanych jest dużo więcej, po przetworzeniu warstwy wektorowej w warstwę rastrową, z powodu różnic w sposobie obliczania powierzchni w tych warstwach. Dokładniejsze są obliczenia wykonywane na danych wektorowych, ponadto zaznaczył się wspomniany wcześniej problem „zgubienia” w rastrze (o pikselach 10×10 m) obiektów o małej powierzchni lub długich i wąskich.

Porównując wyniki klasyfikacji obrazów satelitarnych z warstwą referencyjną EGiB (Tab. 7) łatwo zauważyć, mimo niepełnej zgodności znaczeniowej porównywanych klas, że na badanym obszarze wykazano znaczny odsetek lasów i zadrzewień (zależnie od metody i wariantu 23,177–34,208% powierzchni powiatu), podczas gdy odpowiadająca im według EGiB klasa E1 zajmuje tylko 16,818%. Do lasów i zadrzewień zostały zaklasyfikowane głównie grunty kl. E1 (lasy i zadrzewienia) oraz klasy E4, tj. grunty rolne (bez Ł, Ps, S) – z pokrywą roślinną lub bez pokrywy roślinnej. Trafiała tu także znaczna część nieużytków (kl. E6). Wynik wskazuje, że część upraw rolnych, zapewne tych z pokrywą roślinną, miała charakterystyki spektralne podobne do charakterystyk lasów. W tej liczbie mogą być także ukryte grunty rolne, które zaprzestano uprawiać i samorzutnie pokryły się zadrzewieniami oraz zadrzewione nieużytki. Uzyskany wynik jest zgodny z badaniami innych autorów. Jabłoński (2015) wskazuje, na przykładzie wybranych powiatów, na znaczne niekiedy różnice pomiędzy wielością obszarów leśnych ujętych w EGiB i lasami występującymi w rzeczywistości. Wielostronna analiza przeprowadzona przez Hościło i in. (2016) oraz Hościło i Ziółkowskiego (2016) oparta na zawartości kilku baz danych, również wykazała, że rzeczywista powierzchnia lasów w Polsce jest wyższa od ujętej w oficjalnych bazach danych. Wysocka-Fijorek i Kaliszewski (2017) piszą, że problem pominięcia w ewidencji dotyczy w przeważającej mierze gruntów rolnych zalesionych w drodze naturalnej sukcesji. Jak zauważa natomiast Krawczyk (2014), zjawisko to jest pomijane, co może mieć wpływ na wyniki prognoz dotyczących wzrostu lesistości w Polsce. Inaczej jest z gruntami rolnymi (Tab. 7, suma klas T4 i T5) – tutaj, zależnie od zastosowanej metody klasyfikacji uzyskano od 34,247%

(metoda MD) do 71,931% (metoda ML) powierzchni powiatu, podczas gdy dane referencyjne EGiB (suma klas E4 i E5) wskazują na udział tych gruntów wynoszący 70,085%.

Klasyfikacje prowadzone metodami MD oraz ML na czterech (2, 3, 4, 8) oryginalnych kanałach Sentinel-2 oraz na czterech komponentach uzyskanych za pomocą metody PCA dają analogiczne rezultaty, co jest w pełni zrozumiałe, gdyż w obydwu przypadkach korzysta się z pełnej informacji, jedynie inaczej zorganizowanej. Dopiero redukcja liczby komponentów uczestniczących w klasyfikacji, co oznacza rezygnację z części informacji, powoduje nieznaczną zmianę wyników. We wszystkich wariantach klasyfikacji, wyniki najbardziej zgodne z danymi referencyjnymi uzyskano za pomocą algorytmu największego prawdopodobieństwa (ML).

Przeprowadzone analizy upoważniają do sformułowania następujących wniosków ogólnych:

1. Redukcja liczby komponentów PCA użytych w klasyfikacji prowadzi jedynie do niewielkich zmian uzyskiwanych wyników w porównaniu do wyników klasyfikacji prowadzonej na obrazach oryginalnych. Różnice są największe w przypadku zastosowania tylko dwóch komponentów, lecz i w tym przypadku sprowadzają się do odmiennej klasyfikacji stosunkowo niewielkich obszarów. Wynik upoważnia do stwierdzenia, że wykorzystywanie informacji teledetekcyjnej zawartej w pierwszych komponentach wystarcza do uzyskania w miarę poprawnych wyników klasyfikacji. W przypadku zastosowanych kanałów Sentinel-2 (2, 3, 4, 8), przetworzenie ich metodą PCA doprowadziło do skumulowania informacji w dwóch pierwszych komponentach. Wynik, o ile byłby potwierdzony dla innych scen Sentinel-2 z tej pory roku, miałby ważne znaczenie dla przetwarzania obrazów satelitarnych dużych obszarów, gdy celem jest testowanie różnych metod (lub wariantów) klasyfikacji – zamiast bowiem przetwarzać za każdym razem pełny komplet kanałów, można je najpierw przetworzyć do kilku komponentów.
2. Istnieją duże różnice między wynikami uzyskiwanymi za pomocą dwóch algorytmów – najmniejszej odległości (MD) oraz największego prawdopodobieństwa (ML). Te różnice są dużo większe od różnic wynikających z zastosowania w klasyfikacji różnej liczby komponentów PCA. Wybór metody klasyfikacji ma zatem bardziej istotny wpływ na wyniki, niż przetworzenie (skumulowanie) informacji do kilku pierwszych komponentów PCA. Można przypuszczać, że równie duży, a może i większy wpływ na wyniki mogą mieć przyjęty schemat klas pokrycia terenu oraz wybór lokalizacji pól treningowych, a także ich homogeniczność.
3. Obydwie metody klasyfikacji prowadzą do uzyskania wyników odbiegających od oficjalnych statystyk pokrycia terenu. Przyczyn tej rozbieżności można dopatrywać się m.in. w rozdzielczości przestrzennej obrazów satelitarnych, skali mapy ewidencyjnej, charakterystykach spektralnych obiektów nie pozwalających na wyodrębnienie z obrazów satelitarnych terenów odpowiadających rodzajowi użyt-

ków gruntowych według EGiB. Można zauważyć, że wynik klasyfikacji metodą ML dla obrazów składających się z oryginalnych kanałów jest najbardziej zbliżony do danych EGiB.

4. Podjęta próba porównania pokrycia terenu ustalonego za pomocą klasyfikacji obrazów satelitarnych z klasami pokrycia, które zostały wyodrębnione z referencyjnej mapy EGiB wykazała, że przetworzenie mapy z postaci wektorowej na rastrową wpływa istotnie na uzyskiwane wyniki.

## Bibliografia

- Adamczyk J., Będkowski K., 2007, Metody cyfrowe w teledetekcji, Wydawnictwo SGGW, Warszawa.
- Babenko D., Marmanis H., McIlwraith D., 2017, Inteligentna sieć. Algorytmy przyszłości, Wydawnictwo Helion, Gliwice.
- Będkowski K., Piekarski E., 2017, Podstawy fotogrametrii i teledetekcji dla leśników, Wydawnictwo SGGW, Warszawa.
- Bhattacharya A.K., Majumdar T.J., Pal S.K., 2007, ERS-2 SAR and IRS-1C LISS III data fusion: A PCA approach to improve remote sensing based geological interpretation, *Photogrammetry & Remote Sensing*, 61: 281–297.
- GUS 2014: Bank Danych Lokalnych (BDL), grupa: Powierzchnia Geodezyjna Kraju (Dane GUGiK), podgrupa: Powierzchnia geodezyjna kraju według kierunków wykorzystania.
- Hościło A., Mirończuk A., Lewandowska A., 2016, Określenie rzeczywistej powierzchni lasów w Polsce na podstawie dostępnych danych przestrzennych, *Sylvan*, 160(8): 627–634.
- Hościło A., Ziółkowski D., 2016, Las z satelity Sentinel, *Głos Lasu*, 7–8: 35–37.
- Jabłoński M., 2015, Powierzchnia gruntów leśnych – przyczyny zmian i spójność źródeł danych, *Wiadomości Statystyczne*, 11: 54–67.
- Krawczyk R., 2014, Zalesienia a sukcesja wtórna, *Leśne Prace Badawcze*, 75(4): 423–427.
- Uchwała 2016: Uchwała Rady Miasta Płocka nr 438/XXV/2016 z dnia 29 listopada 2016 roku: Program ochrony środowiska dla miasta Płocka.
- Uchwała 2017: Uchwała Rady Powiatu Płockiego nr 273/XXIX/2017 z dnia 29 listopada 2017 roku: Program ochrony środowiska dla powiatu płockiego.
- Ustawa 1989: Ustawa z dnia 17 maja 1989 roku: Prawo geodezyjne i kartograficzne (Dz.U., 2017, poz. 2101; 2019, poz. 725, 730).
- Wysocka-Fijorek E., Kaliszewski A., 2017, Przyczyny i kierunki przeklasyfikowywania gruntów nieleśnych na leśne w świetle badań ankietowych, *Sylvan*, 161(6): 460–466.

## Źródła internetowe

- Europejska Agencja Kosmiczna (ESA), Paryż, Francja, <https://scihub.copernicus.eu> (dostęp: 22.09.2019).
- Główny Urząd Geodezji i Kartografii, Warszawa, <https://www.geoportal.gov.pl> (dostęp: 30.06.2019).
- Główny Urząd Statystyczny, Warszawa, <https://stat.gov.pl/> (dostęp: 08.10.2020).
- PCI Geomatics, Markham, Kanada, <https://www.pcigeomatics.com> (dostęp: 20.10.2019).
- Sejm Rzeczypospolitej Polskiej, Warszawa, <http://prawo.sejm.gov.pl> (dostęp: 15.08.2019).
- Sentinel-2 ESA's Optical High-Resolution Mission for GMES Operational Services, [https://sentinel.esa.int/documents/247904/349490/S2\\_SP-1322\\_2.pdf](https://sentinel.esa.int/documents/247904/349490/S2_SP-1322_2.pdf) (dostęp: 30.12.2020).
- Starostwo Powiatowe w Płocku, Płock, <http://powiat-plock.geoportal2.pl> (dostęp: 14.05.2019).