



Jacek Kozak

**DOKŁADNOŚĆ GLOBALNYCH MAP UŻYTKOWANIA ZIEMI
I POKRYCIA TERENU NA PRZYKŁADZIE
GLOBAL LAND COVER CHARACTERIZATION DATA BASE**

**ACCURACY OF GLOBAL LAND COVER DATASETS:
EXAMPLE OF GLOBAL LAND COVER CHARACTERIZATION
DATA BASE**

*Instytut Geografii i Gospodarki Przestrzennej Uniwersytetu Jagiellońskiego
Zakład Systemów Informacji Geograficznej
Institute of Geography and Spatial Management, Jagiellonian University
Geographic Information Systems Laboratory*

STRESZCZENIE. W badaniach oceniono dokładność wyznaczenia powierzchni europejskich lasów w bazie danych Global Land Cover Characterization (GLCC). Jako dane referencyjne wykorzystana została mapa CORINE Major Land Cover Types of Europe. Dla danych zagregowanych do dwóch klas: lasy i obszary nieleśne obliczono wskaźniki dokładności. Przestrzenne zróżnicowanie dokładności badano w polach podstawowych 100×100 km.

Powierzchnie lasów w obu bazach danych są zbliżone, co wynika z kompensowania się błędów niedoszacowania i przeszacowania powierzchni leśnej w bazie danych GLCC. Lesistość szacowana za pomocą GLCC może znacząco różnić się od wartości faktycznych. Przeszacowanie powierzchni leśnej cechuje Skandynawię, niedoszacowanie niziny zachodniej i środkowej Europy. Współczynnik Kappa zmienia się od 0 do 70%. Wartości wskaźników dokładności są zależne od faktycznej lesistości.

Błędy w oszacowaniu powierzchni leśnej danych GLCC wynikają przede wszystkim z wykorzystania danych wejściowych o małej rozdzielczości przestrzennej dla obszarów o znacznej heterogeniczności użytkowania ziemi i pokrycia terenu. Błędy te mogą być korygowane za pomocą prostych technik kalibracyjnych.

SŁOWA KLUCZOWE: dane globalne, lasy, Europa, dokładność

WSTĘP

Globalne mapy użytkowania ziemi i pokrycia terenu są ważnym źródłem informacji pozwalającym na zrozumienie złożoności geosystemu Ziemi, stanowiąc jedną z wielu warstw tematycznych, otrzymywanych dzięki dynamicznie rozwijającej się technologii satelitarnych badań Ziemi (Belward i in., 1999; DeFries, Belward, 2000). W latach 90. ubiegłego wieku globalne mapy użytkowania ziemi i pokrycia terenu tworzono na

podstawie danych skanera AVHRR (Advanced Very High Resolution Radiometer). Przykładem jest Global Land Cover Characterization (GLCC) Data Base, utworzona w ramach projektu kierowanego przez International Geosphere Biosphere Programme Data and Information System (IGBP-DIS), dostępna poprzez Internet od 1997 roku dla społeczności naukowej świata i szeroko wykorzystywana w różnego rodzaju opracowaniach (Brown i in., 1999; Teillet i in., 2000). Do niedawna była to jedyna tego typu baza danych obok tzw. globalnej mapy użytkowania ziemi Uniwersytetu Maryland (Hansen, Reed, 2000). Obecnie szybki rozwój satelitarnego monitoringu powierzchni Ziemi owocuje kolejnymi globalnymi lub kontynentalnymi mapami użytkowania ziemi i pokrycia terenu, tworzonymi na przykład w oparciu o dane sensora MODIS (Friedl i in., 2002) lub też SPOT Vegetation (Mayaux i in., 2002).

Zagadnieniem dużej wagi jest dokładność globalnych map użytkowania ziemi i pokrycia terenu wykonanych w oparciu o dane satelitarne (Scepan, 1999; Cihlar, 2000). Błędy w tematycznych mapach pochodzących z przetworzenia zdjęć satelitarnych mogą wynikać z własności obrazu satelitarnego i jego jakości (np. zachmurzenia), metodyki przetworzenia danych satelitarnych oraz niezgodności przestrzennych i czasowych z danymi referencyjnymi wykorzystywanymi w czasie przetworzenia danych (Congalton, Green, 1993). W przypadku danych o małej rozdzielczości, błędy wynikają także z niejednorodności informacji spektralnej, rejestrowanej w chwilowym polu widzenia skanera, a pochodzącej z różnych typów pokrycia terenu (Moody, Woodcock, 1994). Charakterystyczną cechą map tematycznych o małej rozdzielczości jest przeszacowanie powierzchni klas dominujących (Moody, Woodcock, 1996; Mayaux, Lambin, 1997). Zjawisko to jest zarówno efektem niejednorodnej odpowiedzi spektralnej, jak i może być traktowane jako efekt generalizacji danych, związanej ze wzrostem wielkości pola podstawowego.

Celem niniejszej pracy jest określenie dokładności i wskazanie potencjalnych źródeł błędów bazy danych GLCC. Badania przeprowadzono z punktu widzenia użytkowników globalnych modeli użytkowania ziemi i pokrycia terenu, zainteresowanych odpowiedzią na pytanie o dokładność predykcji powierzchni klas użytkowania ziemi i pokrycia terenu w skali regionalnej. Skoncentrowano się przede wszystkim na wpływie przeszacowania powierzchni klasy dominującej na dokładność mapy, co jest problemem uniwersalnym dotyczącym wszystkich bez wyjątku map tematycznych, opartych na danych satelitarnych niskiej rozdzielczości.

W badaniach wykorzystano wycinek przestrzenny bazy GLCC, obejmujący Europę Zachodnią i Środkową. Z uwagi na fakt, iż kompleksowe badanie dokładności różnych klas jest dość złożone, do badań wybrano lasy, tym bardziej, iż baza GLCC jest jednym z najszerzej używanych, kluczowych źródeł informacji o lasach w skali globalnej (Janetos, Ahern, 1997). Ponieważ europejskie lasy zostały skartowane z dużą dokładnością na mapie użytkowania ziemi, która powstała w ramach projektu CORINE Land Cover (Baranowski, Ciołkosz, 1997; Håme i in., 2001), do oceny dokładności bazy danych GLCC wykorzystano mapę użytkowania ziemi i pokrycia terenu CORINE.

MATERIAŁY

Global Land Cover Characterization Data Base składa się z kilku warstw rastrowych o rozdzielczości przestrzennej 1 km, uzupełnionych o dodatkowe dane atrybutowe.

Wykorzystane dane są udostępnione na serwerze U.S. Geological Survey EROS Data Center pod adresem <http://edcwww.cr.usgs.gov>. Baza danych jest wynikiem interpretacji danych AVHRR, zebranych w okresie jednego roku, od kwietnia 1992 do marca 1993. W czasie analizy opartej o metodę klasyfikacji bezwzorcowej wydzielono względnie homogeniczne regiony (Seasonal Land Cover Regions), którym później przypisywano odpowiednią etykietę zgodną z globalną klasyfikacją ekosystemów Olsona, interpretowaną następnie według przyjętych w bazie danych GLCC klasyfikacji użytkowania ziemi i pokrycia terenu (Loveland i in., 1999; *Global ...*, 2000). Klasyfikacje te różnią się pod względem liczby klas i potencjalnych zastosowań, a powierzchnia poszczególnych klas użytkowania ziemi i pokrycia terenu może się różnić. W większości klasyfikacji stosowanych w bazie danych GLCC, w tym także w klasyfikacji U.S. Geological Survey Land Use / Land Cover System wykorzystanej w pracy, występuje pięć klas lasów, reprezentujących obszary zdominowane przez drzewa o wysokości co najmniej 2 m i zwarciu koron wynoszącym 60% lub więcej (Scepan, 1999). Szacunki powierzchni leśnej są najbardziej jednorodne we wszystkich klasyfikacjach (Loveland i in., 1999).

Badaniom poddane zostało drugie wydanie GLCC. W porównaniu z wersją pierwszą, w Europie poprawiona została delimitacja lasów, w szczególności na obszarach zdominowanych przez mozaikę użytków leśnych i nieleśnych (Brown i in., 1999).

Dokładność pierwszego wydania GLCC była już przedmiotem badań (Scepan, 1999). Oceniono ją na 73.5%, jednakże wartość ta różni się dla poszczególnych klas oraz wykazuje znaczące zróżnicowanie przestrzenne. Dla różnych typów lasów dokładność delimitacji zmienia się od 40% (lasy liściaste, zrzucające liście) do 84% (wiecznie zielone lasy iglaste). Najniższe wartości dokładności uzyskano w Ameryce Północnej (63%) oraz w Eurazji (69%). Niskie wartości w Eurazji tłumaczone są długą i złożoną historią użytkowania i wpływu człowieka (Loveland i in., 2000).

Mapa użytkowania ziemi CORINE Major Land Cover Types of Europe, którą wykorzystano jako dane referencyjne, została opublikowana w 1997 roku. Mapa, obejmująca większą część Europy, ma postać warstwy rastrowej o rozdzielczości przestrzennej 250 m. Raster został utworzony z danych wektorowych, będących wynikiem interpretacji zdjęć satelitarnych, przede wszystkim Landsat TM i SPOT, przeprowadzonej przez grupy ekspertów z krajów uczestniczących w projekcie CORINE. Zdjęcia wykonywane były dla różnych krajów w różnym czasie, w większości pod koniec lat 80. i na początku lat 90. ubiegłego wieku (Baranowski, Ciołkosz, 1997), a więc w terminie zbliżonym do wykonania serii zdjęć AVHRR, będącej podstawą dla map bazy danych GLCC. Ważną cechą mapy CORINE jest określenie minimalnej wielkości homogenicznego poligonu na 25 hektarów (Perdigão, Annoni, 1997), co powoduje generalizację danych wejściowych oraz utratę informacji o obiektach o wielkości mniejszej od progowej. Legendę mapy CORINE Major Land Cover Types of Europe tworzy siedem kategorii, w tym jedna kategoria lasów. Jest to wynik agregacji szczegółowej legendy wykorzystanej w projekcie CORINE, na którą składają się 44 klasy, pogrupowane w trójstopniowym systemie hierarchicznym (Perdigão, Annoni, 1997). Lasy definiowane są jako obszary zajęte przez drzewa o wysokości co najmniej 5 m i zwarciu koron co najmniej 30% (Bossard i in., 2000).

Z uwagi na znacznie wyższą rozdzielczość przestrzenną materiałów źródłowych wykorzystanych przy tworzeniu mapy CORINE, można przyjąć, iż jest ona znacznie dokładniejsza od map bazy danych GLCC, i dlatego też może być wykorzystana jako materiał referencyjny. Podejście takie prezentowane jest w innych pracach, w których wykorzystywane są dane AVHRR (Mücher i in., 2000; Håme i in., 2001). Jakkolwiek wyniki oceny dokładności jakiegokolwiek zbioru danych mogą być zniekształcone przez błędy danych referencyjnych, w szczególności tych, które wytworzono na podstawie zdjęć satelitarnych (Foody, 2002), zagadnienie to nie będzie rozważane w niniejszej pracy.

METODY

Obie mapy zostały przekształcone do równopowierzchniowego odwzorowania Lamberta ze środkiem o współrzędnych 55° szerokości geograficznej północnej i 20° długości geograficznej wschodniej. Następnie dokonano agregacji klas na mapach do dwóch kategorii: lasów i obszarów nieleśnych, po czym nałożono mapy na siebie.

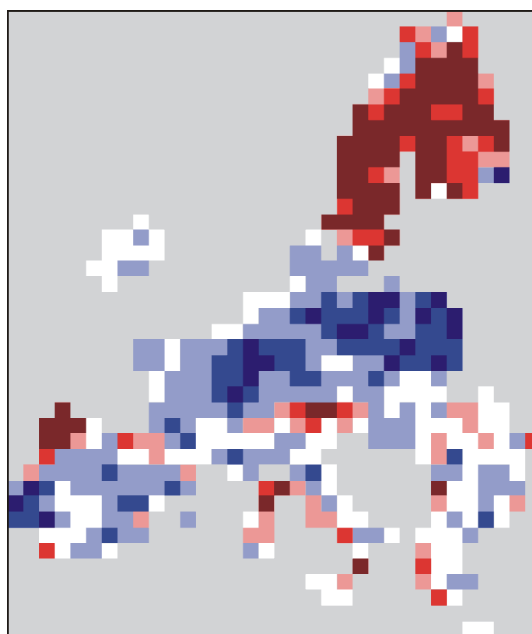
Nałożenie map pozwoliło na obliczenie globalnych wskaźników dokładności dla danych GLCC: udziału obszarów sklasyfikowanych poprawnie, współczynnika Kappa, dokładności producenta oraz dokładności użytkownika dla klasy lasów. Miary te stanowią standardowe wskaźniki dokładności map tematycznych, generowane z macierzy błędów (załącznik 1). Współczynnik Kappa jest miarą ostrzejszą niż udział obszarów sklasyfikowanych poprawnie, gdyż nie uwzględnia zgodności mapy weryfikowanej i danych referencyjnych, wynikającej z losowego dopasowania (Stehman, Czaplewski, 1998; Foody, 2002).

Wynik nałożenia na siebie map zagregowanych poddano dalszej analizie w celu zbadania przestrzennego zróżnicowania dokładności mapy GLCC. Wyznaczono 493 kwadratowe pola podstawowe, o boku 100 km i powierzchni lądowej większej lub równej 1000 km^2 . Dla każdego z pól podstawowych obliczono dwa wskaźniki: różnicę pomiędzy lesistością szacowaną na podstawie GLCC, a lesistością obliczoną na podstawie danych CORINE oraz współczynnik Kappa. Pierwszy wskaźnik określa, na ile w danym polu podstawowym szacunek powierzchni leśnej jest zgodny ze stanem faktycznym. Ponieważ szacunek taki może wynikać ze wzajemnego kompensowania się błędów niedoszacowania i przeszacowania powierzchni leśnej, obliczono także współczynnik Kappa, określając tym samym dla każdego pola podstawowego poprawność wyznaczenia obszaru leśnego w bazie danych GLCC względem danych CORINE.

Lesistość określoną na podstawie danych GLCC oraz współczynnik Kappa odniesiono do lesistości określonej na podstawie danych CORINE, aby zbadać związek pomiędzy udziałem powierzchniowym lasów, a spodziewaną dokładnością ich wyznaczenia na mapie GLCC.

WYNIKI

Powierzchnia lasów w analizowanej części Europy określona na podstawie danych GLCC wynosi $1.17 \cdot 10^6 \text{ km}^2$, 30.1% powierzchni analizowanego obszaru. Prawie identyczna jest powierzchnia określona na podstawie danych referencyjnych – $1.18 \cdot 10^6 \text{ km}^2$, także 30.1%. Zgodność szacunków powierzchni lasów jest jednak wynikiem znacznych, znoszących się wzajemnie błędów niedoszacowania i przeszacowania powierzchni leśnej.

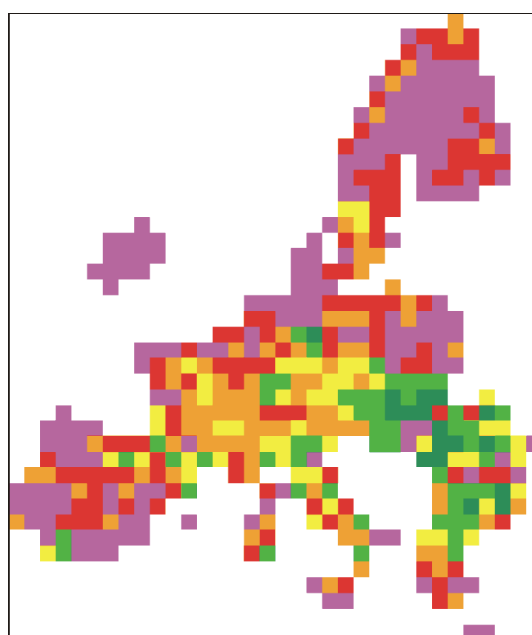


Ryc. 1. Różnice pomiędzy szacowaną (GLCC) a faktyczną (CORINE) lesistością.

- 1 - < -25%;
- 2 - -25% : -15%;
- 3 - -15% : -5%;
- 4 - -5% : 5%;
- 5 - 5% : 15%;
- 6 - 15% : 25%;
- 7 - >= 25%

Fig. 1. Differences between predicted (GLCC) and actual (CORINE) forest proportion.

- 1 - < -25%;
- 2 - -25% : -15%;
- 3 - -15% : -5%;
- 4 - -5% : 5%;
- 5 - 5% : 15%;
- 6 - 15% : 25%;
- 7 - >= 25%



Ryc. 2. Przestrzenne zróżnicowanie współczynnika Kappa.

- 1 - < 10%;
- 2 - 10% : 20%;
- 3 - 20% : 30%;
- 4 - 30% : 40%;
- 5 - 40% : 50%;
- 6 - >= 50%

Fig. 2. Kappa coefficient.

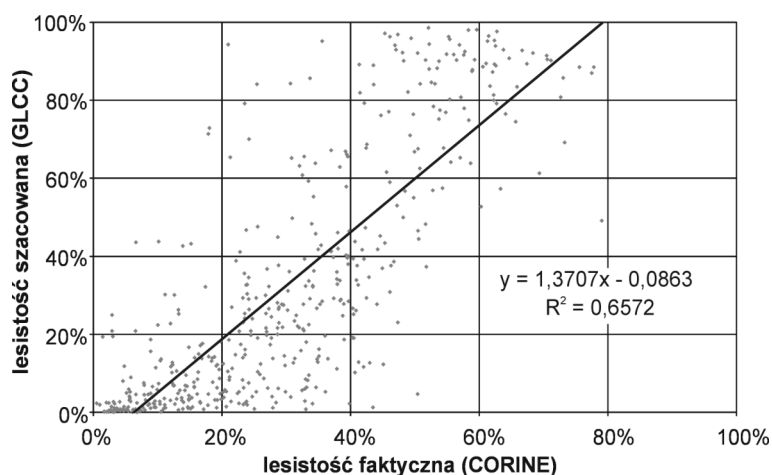
- 1 - < 10%;
- 2 - 10% : 20%;
- 3 - 20% : 30%;
- 4 - 30% : 40%;
- 5 - 40% : 50%;
- 6 - >= 50%



Określone na podstawie analizy map zagregowanych do dwóch klas dokładności użytkownika i producenta dla klasy lasów GLCC wynoszą odpowiednio 56.9% oraz 56.8%. Obie wartości są relatywnie niskie. Udział obszarów sklasyfikowanych poprawnie wynosi 74.0%, jednakże współczynnik Kappa tylko 38.2%.

Dokładność delimitacji obszarów leśnych i nieleśnych wykazuje znaczne zróżnicowanie przestrzenne. Różnica pomiędzy szacowaną na podstawie GLCC oraz faktyczną lesistością może przewyższać 25 punktów procentowych (ryc. 1), co dotyczy prawie jednej piątej badanego terenu. Tylko 25% badanego obszaru wykazuje różnice mniejsze niż 5 punktów procentowych. Wyraźne są trzy obszary o zróżnicowanych błędach szacunku lesistości: pierwszy, z wyraźnym przeszacowaniem powierzchni leśnej, obejmujący Skandynawię, drugi, z wyraźnym niedoszacowaniem powierzchni leśnej, obejmujący niziny zachodniej i środkowej Europy oraz trzeci, silnie zróżnicowany, obejmujący regiony górskie oraz region śródziemnomorski.

Współczynnik Kappa nie przekracza w żadnym wypadku 70%. Wartości niższe niż 10% występują przede wszystkim w nizinnej części zachodniej i środkowej Europy, w regionie śródziemnomorskim oraz w Skandynawii (ryc. 2), łącznie obejmując 36% badanego obszaru. Wartości wyższe niż 40% występują w pasie ciągnącym się od Pirenejów na zachodzie po Rumunię i Bułgarię na wschodzie i stanowią 14% badanego obszaru.



Ryc. 3. Zależność pomiędzy faktyczną lesistością pól podstawowych a lesistością szacowaną
Fig. 3. Relation between actual and predicted forest proportion

Lesistość określona na podstawie bazy danych GLCC oraz współczynniki Kappa pozostają w związku z faktyczną lesistością pola podstawowego. Zależność lesistości szacowanej od faktycznej ma wyraźny liniowy charakter (ryc. 3), choć wartość R^2 wynosząca 0.66 wskazuje na znaczne lokalne różnice między GLCC i CORINE. Na podstawie równania regresji stwierdzić można, iż lesistość szacowana jest przeciętnie niższa od faktycznej dla obszarów o lesistości niższej niż 23%, natomiast przeciętnie wyższa dla obszarów o lesistości przekraczającej 23%. Dla obszarów o lesistości przekraczającej 80%, szacunki lesistości GLCC wynoszą 100%. Największe odchylenia

lesistości szacowanej od faktycznej występują na obszarach o małym i dużym udziale lasów. Potwierdza to zależność wartości współczynnika Kappa od lesistości, który najwyższe wartości osiąga dla lesistości w zakresie od 20% do 50%, poza tym przedziałem wartości współczynnika Kappa są bliskie zeru.

DYSKUSJA

Otrzymane szacunki dokładności delimitacji lasów w bazie GLCC, oparte o udział powierzchni poprawnie zaklasyfikowanej, są zbliżone do wartości podawanych w literaturze (Sečan, 1999). Dokładności użytkownika i producenta dla klasy lasów są zbliżone, choć nieco niższe do wartości podawanych dla lasów Eurazji przez Lovelanda i in. (2000). Wartość współczynnika Kappa obliczona dla całego badanego wycinka mapy wskazuje natomiast na dość niską zgodność delimitacji lasów obu baz danych. Chociaż powierzchnia leśna obliczona na podstawie mapy GLCC odpowiada prawie dokładnie powierzchni leśnej mapy CORINE, wynika to wyłącznie z kompensowania się na całym badanym obszarze znacznych błędów przeszacowania i niedoszacowania. Błędy te natomiast nie kompensują się lokalnie, stąd też odchylenia szacunku lesistości mogą przekraczać 25 punktów procentowych. Relatywnie niska dokładność delimitacji lasów na mapie GLCC odpowiada poglądom wyrażanym w literaturze, że dokładność map tematycznych, wykonanych na podstawie danych AVHRR nie przekracza 70% (Mücher i in., 2000).

Analiza przestrzennego zróżnicowania dokładności delimitacji lasów wykazuje, iż najistotniejszy wpływ na błędy bazy danych GLCC ma przeszacowanie powierzchni klasy dominującej, typowa cecha map tematycznych o małej rozdzielczości (Moody, 1998). Faktycznym wartościom lesistości rzędu 80% i więcej odpowiadają szacowane na podstawie GLCC wartości zbliżone do 100%, natomiast wartościom niższym niż 10% – wartości zbliżone do zera. Oznacza to, iż w bazie danych GLCC zaniżany jest udział obszarów nieleśnych na obszarach o znacznej lesistości (np. w Skandynawii), oraz lasów na obszarach o małej lesistości (niziny Europy Zachodniej i Środkowej).

Przeszacowanie klasy dominującej dotyczy nie tylko danych GLCC, ale także innych map tematycznych wykonanych w oparciu o dane AVHRR. Podobny związek pomiędzy faktyczną lesistością a lesistością szacowaną na podstawie danych AVHRR stwierdzili dla lasów tropikalnych Mayaux i Lambin (1995) oraz Achard i in. (2001). W przypadku danych GLCC ta tendencja jest dodatkowo wzmocniona przez metodologię obróbki danych AVHRR, opartą o klasyfikację bezwzorcową (Mücher i in., 2000).

Jakkolwiek stwierdzone lokalne odchylenia szacunków powierzchni wybranych klas są znaczne, należy stwierdzić, iż nie prowadzi to do wniosku o braku przydatności map tematycznych opartych o dane małej rozdzielczości do badań w skali regionalnej lub lokalnej. Problemy z oszacowaniem powierzchni klas mogą być bowiem rozwiązywane za pomocą różnego typu technik kalibracyjnych (Mayaux, Lambin, 1995; Moody, Woodcock, 1996; Mayaux, Lambin, 1997; Moody, 1998). Najprostsze z nich pozwalają na korekcję szacunku udziału danej klasy na podstawie równań regresji, otrzymywanych dzięki odniesieniu analizowanych danych tematycznych do dokładnych danych referencyjnych. Na pokonanie trudności związanych z szacowaniem powierzchni lub udziału klas pozwalają także techniki umożliwiające określenie udziału danej klasy w pikselu

mapy tematycznej. Zamiast jakościowego i jednoznacznego określenia przynależności piksela mapy do klasy, udział danej klasy w pikselu określany jest jako zmienna ciągła z przedziału od 0 do 1. Mapy takie pozwalają na dokładniejsze szacowanie powierzchni klas, dobrze też oddają naturę mieszanego charakteru odpowiedzi spektralnej pola podstawowego o dużych rozmiarach na silnie heterogenicznym obszarze. Metoda ta jest coraz częściej wykorzystywana do tworzenia map regionalnych lub globalnych (DeFries i in., 1997; Håme i in., 2001; Hansen i in., 2002).

WNIOSKI

Baza danych GLCC pozwala na względnie dokładne szacowanie powierzchni leśnej zarówno w skali globalnej, jak i regionalnej. W tym drugim wypadku należy się liczyć jednak z możliwością występowania błędów, wynikających przede wszystkim z małej rozdzielczości przestrzennej materiałów źródłowych, wykorzystanych przy tworzeniu mapy. W szczególności przeszacowana zostaje powierzchnia klasy dominującej, co w dużym stopniu uznać można za efekt analogiczny do tego, który występuje przy generalizacji mapy tematycznej o początkowo dużej rozdzielczości przestrzennej. Prowadzi to do przeszacowania powierzchni leśnej na obszarze o dużym udziale lasów i niedoszacowania jej na obszarze o małej lesistości. Wielkość błędu szacunku lesistości przekracza niejednokrotnie 25 punktów procentowych.

Błędy tego typu mogą być w większości wypadków korygowane za pomocą prostych technik kalibracyjnych. Nie zostaną one natomiast zlikwidowane dzięki poprawie jakości technicznej materiałów źródłowych, gdyż zależą przede wszystkim od rozdzielczości przestrzennej. Tak więc będą one występowały także na mapach tematycznych utworzonych na podstawie danych satelitarnych nowszej niż AVHRR generacji, takich jak MODIS lub SPOT Vegetation.

PODZIĘKOWANIA

Badania wykonano w ramach grantu KBN nr 6 P04 020 19. Dane CORINE Major Land Cover Types of Europe uzyskano z European Topic Centre on Land Cover – European Environment Agency.

PIŚMIENNICTWO

- Achard F., Eva H., Mayaux P., 2001. Tropical forest mapping from coarse spatial resolution satellite data: production and accuracy assessment issues. *International Journal of Remote Sensing* 22, 14: 2741–2762.
- Baranowski M., Ciołkosz A., Nowa mapa użytkowania ziemi w Polsce jako pochodna bazy danych „CORINE Land Cover”. *Polski Przegląd Kartograficzny* 29, 4: 219–228.
- Belward A. S., Estes J. E., Kline K. D., 1999. The IGBP-DIS Global 1-Km Land-Cover Data Set DISCover: A Project Overview. *Photogrammetric Engineering & Remote Sensing* 65, 9: 1013–1020.
- Bossard M., Feranec J., Otahel J., 2000. The revised and supplemented Corine land cover nomenclature. Technical report 38, European Environmental Agency, 110 ss.
- Brown J. F., Loveland T. R., Ohlen D. O., Zhu Z., 1999. The Global Land-Cover Characteristics Database: The Users' Perspective. *Photogrammetric Engineering & Remote Sensing* 65, 9: 1069–1074.

- Cihlar J., 2000. Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing* 21, 6&7: 1093–1114.
- Congalton R. G., Green K., 1993. A Practical Look at the Sources of Confusion in Error Matrix Generation. *Photogrammetric Engineering & Remote Sensing* 59, 5: 641–644.
- DeFries R. S., Belward A. S., 2000. Global and regional land cover characterization from satellite data: an introduction to the Special Issue. *International Journal of Remote Sensing* 21, 6&7: 1083–1092.
- DeFries R. S., Hansen M., Steining M., Dubayah R., Sohlberg R., Townshend J., 1997. Subpixel Forest Cover in Central Africa from Multisensor, Multitemporal Data. *Remote Sensing of Environment* 60: 228–246.
- Foody G.M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment* 80: 185–201.
- Friedl M. A., McIver D. K., Hodges J. C. F., Zhang X. Y., Muchoney D., Strahler A. H., Woodcock C. E., Gopal S., Schneider A., Cooper A., Baccini A., Gao F., Schaaf C., 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment* 83: 287–302.
- Global Land Cover Characteristics Data Base Documentation, 2000. U.S. Geological Survey EROS Data Center. Internet: <http://edcwww.cr.usgs.gov>.
- Håme T., Stenberg P., Andersson K., Rauste Y., Kennedy P., Folving S., Sarkela J., 2001. AVHRR-based forest proportion map of the Pan-European area. *Remote Sensing of Environment* 77: 76–91.
- Hansen M. C., DeFries R. S., Townshend J. R. G., Sohlberg R., Dimiceli C., Carroll M., 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS data. *Remote Sensing of Environment* 83: 303–319.
- Hansen M. C., Reed B., 2000. A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products. *International Journal of Remote Sensing* 21, 6&7: 1365–1373.
- Janetos, A.C., Ahern, F., 1997, CEOS Pilot Project: Global Observation of Forest Cover (GOFC). Ottawa Workshop Report, July 7–10, 1997.
- Loveland T. R., Reed B. C., Brown J. C., Ohlen D. O., Zhu Z., Yang L., Merchant J. W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing* 21, 6&7: 1303–1330.
- Loveland T. R., Zhu Z., Ohlen D. O., Brown J. C., Reed B. C., Yang L., 1999. An Analysis of the IGBP Global Land-Cover Characterization Process. *Photogrammetric Engineering & Remote Sensing* 65, 9: 1021–1032.
- Mayaux P., Bartholome E., Massart M., Belward A. S., 2002. The land cover of Africa for the year 2000. *LUCC (Land Use and Land Cover Change) Newsletter* 8: 4–6.
- Mayaux P., Lambin E. F., 1995. Estimation of Tropical Forest Area from Coarse Spatial Resolution Data: A Two-Step Correction Function for Proportional Errors Due to Spatial Aggregation. *Remote Sensing of Environment* 53:1–15.
- Mayaux P., Lambin E. F., 1997. Tropical Forest Area measured from Global Land-Cover Classifications: Inverse Calibration Models Based on Spatial Textures. *Remote Sensing of Environment* 59: 29–43.
- Moody A., Woodcock C. E., 1994. Scale-dependent errors in the estimation of land-cover proportions – implications for global land-cover datasets. *Photogrammetric Engineering & Remote Sensing*, 60, 5: 585–594.
- Moody A., Woodcock C. E., 1996. Calibration-Based Models for Correction of Area Estimates Derived from Coarse Resolution Land-Cover Data. *Remote Sensing of Environment* 58: 225–241.
- Moody A., 1998. Using Landscape Spatial Relationships to Improve Estimates of Land-Cover Area from Coarse Resolution Remote Sensing. *Remote Sensing of Environment* 64: 202–220.
- Mücher C. A., Steinnocher K. T., Kressler F. P., Heunks C., 2000. Land cover characterization and change detection for environmental monitoring of pan-Europe. *International Journal of Remote Sensing* 21, 6&7: 1159–1181.

- Perdigão V., Annoni A., 1997. Technical and methodological guide for updating CORINE Land Cover data base. European Commission, EUR 17288, 140 pp.
- Scepan J., 1999. Thematic Validation of High-Resolution Global Land-Cover Data Sets. *Photogrammetric Engineering & Remote Sensing* 65, 9: 1051–1060.
- Stehman S. V., Czaplewski R. L., 1998. Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles. *Remote Sensing of Environment* 64: 331–344.
- Teillet P. M., El Saleous N., Hansen M. C., Eidenshink J. C., Justice C. O., Townshend J. R. G., 2000. An evaluation of the global 1-km AVHRR land dataset. *International Journal of Remote Sensing* 21, 10: 1987–2021.

ACCURACY OF GLOBAL LAND COVER DATASETS: EXAMPLE OF GLOBAL LAND COVER CHARACTERIZATION DATA BASE

S u m m a r y

The study attempts to estimate the accuracy of the delimitation of European forests within the Global Land Cover Characterization (GLCC) Data Base. As a reference, CORINE Major Land Cover Types of Europe data set was used. The analysis was carried out for data aggregated to two classes: forests and non-forest areas. Matrix overlays allowed computation of accuracy indices. Kappa coefficient and the difference of actual (CORINE) and predicted (GLCC) forest proportion were mapped for spatial units 100×100 km.

The area of forests derived from the CORINE data equals $1.18 \cdot 10^6$ km², and the value predicted is almost exactly the same, $1.17 \cdot 10^6$ km². User's and producer's accuracies of forest delimitation equal 57%. Local differences between actual and predicted forest proportion may exceed 25 percentage points. Overestimation of the forest area occurs in Scandinavia, while underestimation in lowland Western and Central Europe. Kappa coefficient varies from 0 up to almost 70%. Low values were found in lowland Central and Western Europe, Ireland, the Mediterranean and Scandinavia. Values exceeding 40% were observed in a belt stretching from northern Spain to Bulgaria. Differences between actual and predicted forest proportion, as well as Kappa coefficient depend on the actual forest proportion.

Errors of forest delimitation within the GLCC data set are related mainly to the classification of coarse resolution data in areas of highly diverse land cover pattern. The overestimation of dominant land cover classes in the GLCC leads to locally inaccurate predictions, which should be calibrated for the purposes of regional research.

KEYWORDS: global data, forests, Europe, accuracy

Recenzent: dr inż. Regina Tokarczyk, AGH Kraków.

ZAŁĄCZNIKI

Załącznik 1. Miary dokładności mapy tematycznej (wg Stehman, Czaplewski, 1998; Foody, 2002).

Mapa tematyczna o q klasach została poddana weryfikacji w n punktach kontrolnych. Poniższa tabela przedstawia macierz błędów (*confusion matrix*), będącą wynikiem weryfikacji. Wartość n_{ij} , gdzie $i, j = 1, \dots, q$, określa liczebność punktów kontrolnych (np. pikseli) klasyfikowanych na weryfikowanej mapie tematycznej jako klasa i , a należących faktycznie do klasy j .

		Dane referencyjne				Σ
		Klasa 1	Klasa 2	...	Klasa q	
Mapa weryfikowana	Klasa 1	n_{11}	n_{12}	...	n_{1q}	n_{1+}
	Klasa 2	n_{21}	n_{22}	...	n_{2q}	n_{2+}

	Klasa q	n_{q1}	n_{q2}	...	n_{qq}	n_{q+}
	Σ	n_{+1}	n_{+2}	...	n_{+q}	n

Dla tak określonej macierzy, miary dokładności oraz miary błędów weryfikowanej mapy tematycznej określić można w następujący sposób:

$$\begin{aligned} \text{udział obszarów sklasyfikowanych poprawnie:} & \dots \frac{\sum_{k=1}^q n_{kk}}{n} \\ \text{dokładność użytkownika (user's accuracy), dla klasy } i & \dots \frac{n_{ii}}{n_{i+}} \\ \text{dokładność producenta (producer's accuracy), dla klasy } j & \dots \frac{n_{jj}}{n_{+j}} \\ \text{współczynnik Kappa} & \dots \frac{n \sum_{k=1}^q n_{kk} - \sum_{k=1}^q n_{k+} n_{+k}}{n^2 - \sum_{k=1}^q n_{k+} n_{+k}} \\ \text{błąd przeszacowania (commission error), dla klasy } i & \dots 1 - \frac{n_{ii}}{n_{i+}} \\ \text{błąd niedoszacowania (ommission error), dla klasy } j & \dots 1 - \frac{n_{jj}}{n_{+j}} \end{aligned}$$