DE GRUYTER
OPEN

Polish Cartographical Review
Vol. 47, 2015, no. 4, pp. 203–212
DOI: 10.1515/pcr-2015-0018

MARTYNA GOLENIA, BOGDAN ZAGAJEWSKI, ADRIAN OCHTYRA
Department of Geoinformatics, Cartography and Remote Sensing
Faculty of Geography and Regional Studies, University of Warsaw
martyna.golenia@student.uw.edu.pl, bogdan@uw.edu.pl
AGATA HOŚCIŁO
Remote Sensing Centre, Institute of Geodesy and Cartography, Warsaw
agata.hoscilo@igik.edu.pl

Semiautomatic land cover mapping according to the 2nd level of the CORINE Land Cover legend

Abstract. Actual land cover maps are a very good source of information on present human activities. It increases value of actual spatial databases and it is a key element for decision makers. Therefore, it is important to develop fast and cheap algorithms and procedures of spatial data updating. Every day, satellite remote sensing deliver vast amount of new data, which can be semi-automatically classified.

The paper presents a method of land cover classification based on a fuzzy artificial neural network simulator and Landsat TM satellite images. The latest CORINE Land Cover 2012 polygons were used as reference data. Three satellite images acquired 21 April 2011, 5 June 2010, 27 August 2011 over Warsaw and surrounding areas were processed. As an outcome of classification procedure, the maps, error matrices and a set of overall, producer and user accuracies and a kappa coefficient were achieved. The classification accuracy oscillates around 76% and confirms that artificial neural networks can be successfully used for forest, urban fabric, arable land, pastures, inland waters and permanent crops mapping. Low accuracies were obtained in case of heterogenic land cover units.

Keywords: classification, Corine Land Cover, Landsat, artificial neural networks, Warsaw

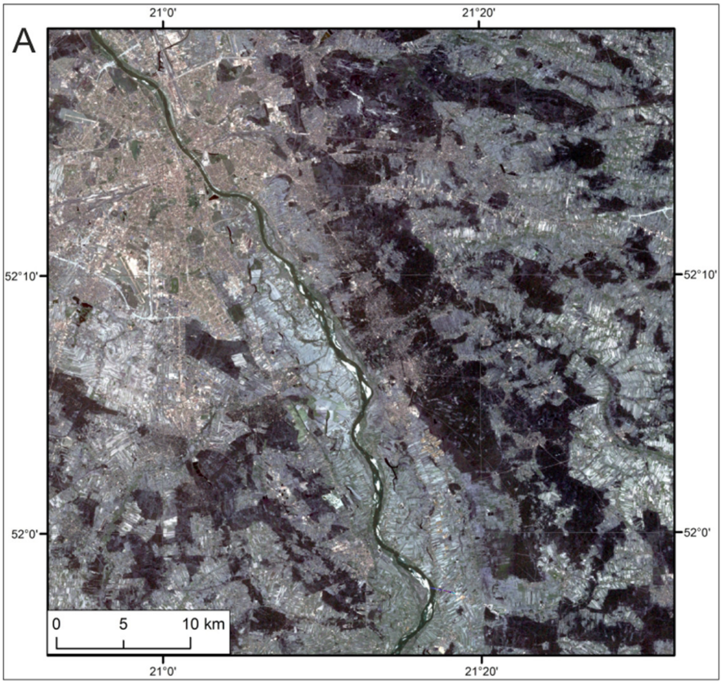
1. Introduction

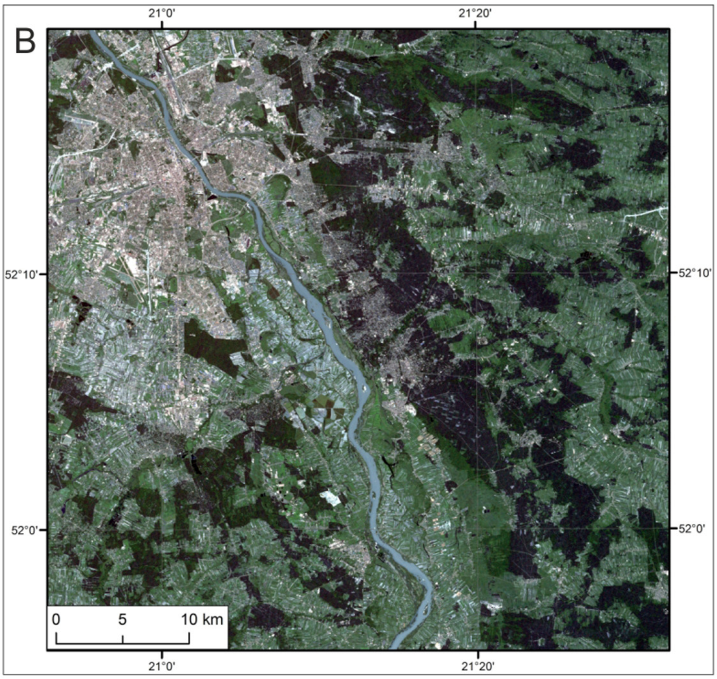
Development of a current land cover map is a time-consuming procedure, because due to vegetation growth and fast urbanization processes such maps must be frequently updated. Satellite images collected daily or in the interval of several days are one of the sources of data used for updating of land cover maps. Until now Corine Land Cover (CLC) databases for Poland have been created as a result of an expert visual interpretation of satellite images. Visual photointerpretation is time-consuming, in the case of CLC2012 covering the area of Poland it lasts almost one and half years (A. Hościło, M. Tomaszewska 2015), as well as costly and subjective. Thus it is important that an automatic or semi-automatic methods of land cover mapping are developed. Such methods are more objective and less expensive, and the process of classification takes less time than in the case of visual interpretation. Nevertheless there is still problem with mapping various stages of vegetative growth. Application of a fuzzy artificial neural network simulator fuzzy ARTMAP (G.A. Carpenter, S. Grossberg et al. 1992) for generation of land cover maps allows for automation of the classification process.

The article presents the methods of land cover classification according to the Corine Land Cover legend using satellite images from Landsat TM[1] and the fuzzy ARTMAP simulator. A research area covers a part of Warsaw and the sub Warsaw area (fig. 1).

The join NASA/USGS Landsat program has provided images of the Earth since 1972 (8 satellites have been launched so far). Table 1 presents the parameters of Landsat satellites. In this article, the Landsat 5 TM images with a spatial resolution of 30 m were analyzed. Landsat images were used to develop Corine Land Cover 1990 and 2000 (E. Bielecka, A. Ciołkosz 2004).

[1] Project Corine Land Cover, www.igik.edu.pl/pl/corine-projekt


A
21°0'
21°20'
52°10'
52°0'
0
5
10 km


B
21°0'
21°20'
52°10'
52°0'
0
5
10 km

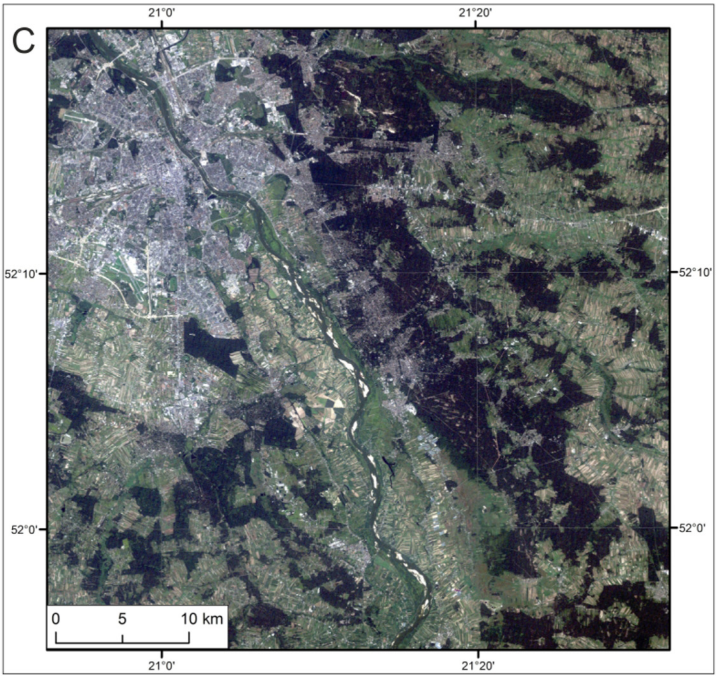


Fig. 1. Study area the Landsat 5 TM satellite, RGB composition: A – 21.04.2011, B – 05.06.2010, C – 27.08.2011

Tab. 1. Landsat missions

Satellite	**Time of data acquisition**	**Spectral resolution (μm)**	**Spatial resolution (m)**
Landsat 1	1972–1978	RBV: 0.48–0.83 (3 spectral bands), MSS: 0.5–1.1 (4 spectral bands)	RBV: 80 m; MSS: 80 m
Landsat 2	1975–1982	RBV: 0.48–0.83 (3 spectral bands), MSS: 0.5–1.1 (4 spectral bands)	RBV: 80 m; MSS: 60 × 80 m
Landsat 3	1978–1983	RBV: 0.505–0.750 (3 spectral bands), MSS: 0.5–12.6 (5 spectral bands)	RBV: 40 m; MSS: 60 × 80 m
Landsat 4	1982–1993	MSS: 0.5–1.1 (4 spectral bands), TM 0.45–12.5 (7 spectral bands)	MSS: 60 × 80 m (MS); TM: 30 m (MS); 120 m (TIR)
Landsat 5	1984–2013	MSS: 0.5–1.1 (4 spectral bands), TM 0.45–12.5 (7 spectral bands)	MSS: 80 m (MS); TM: 30 m (MS) and 120 m (IR)
Landsat 6	1993 (Launch failure)		
Landsat 7	1999–2013	ETM+:0.45–12.50, PAN: 0.5–0.9 (8 spectral bands)	ETM+: 15 m (PAN), 30 m (MS) and 60 m (TIR)
Landsat 8	11 February 2013	OLI: 0.433–12.5 (11 spectral bands)	OLI: 15 m (PAN), 30 m (MS, IR), TIRS: 100 m

Corine Land Cover (CLC) program (COoRdination de l'Information sur l'Environnement / CO-oRdination of INformation on Environment) resulted from an initiative of the European Union in 1985. Its objective was to gather harmonized information about the state of environment in areas of priority for all E.U. countries and to synchronize efforts in data gathering at the national and pan-European level (Commission of the European Communities 1995). CLC legend includes 44 classes of land cover and is divided into 3 levels of detail. The first level, of least detail, contains 5 main classes, the second – 15 classes, and the third – 44 classes. CLC database in Poland includes 33 classes of land cover within the third level of detail. CLC database is updated in a 6 year cycle; the latest was created for 2012. Because of the cyclicity of CLC databases' elaboration it is possible to analyze changes in land cover between periods (J. Feranec et al. 2007). CLC2012 database was used for selection of training and verification polygons (J.B. Campbell, R.H. Wynne 2011) necessary for classification, i.e. assigning the pixels of similar spectral reflection into a particular cluster, which relates to a defined group of objects (J.R. Anderson et al. 1976).

Classification using the artificial neural networks method was based on image spectral parameters and non-parametric components,

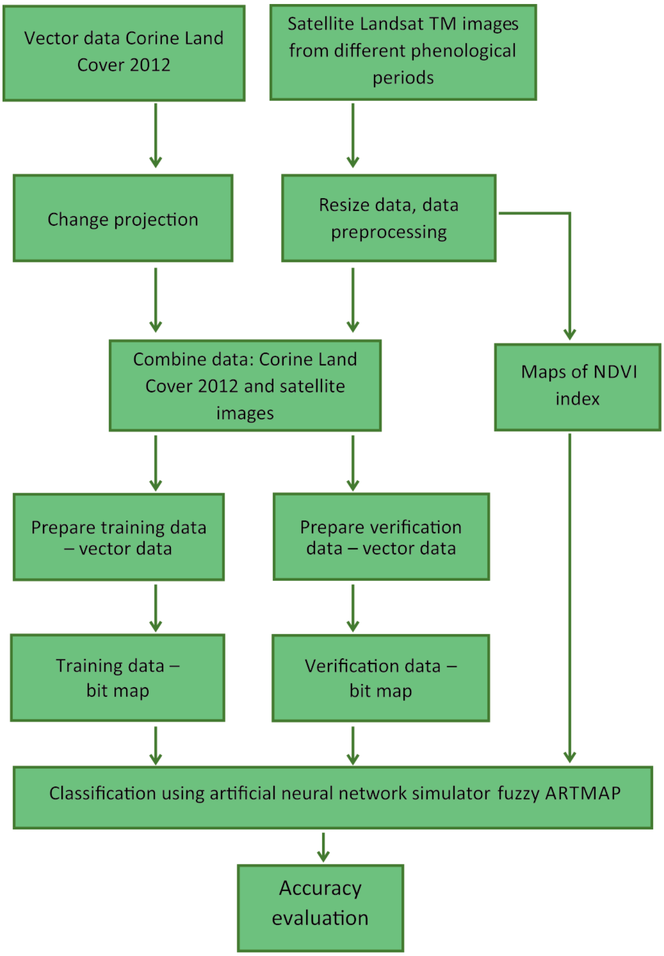


Fig. 2. Research schema

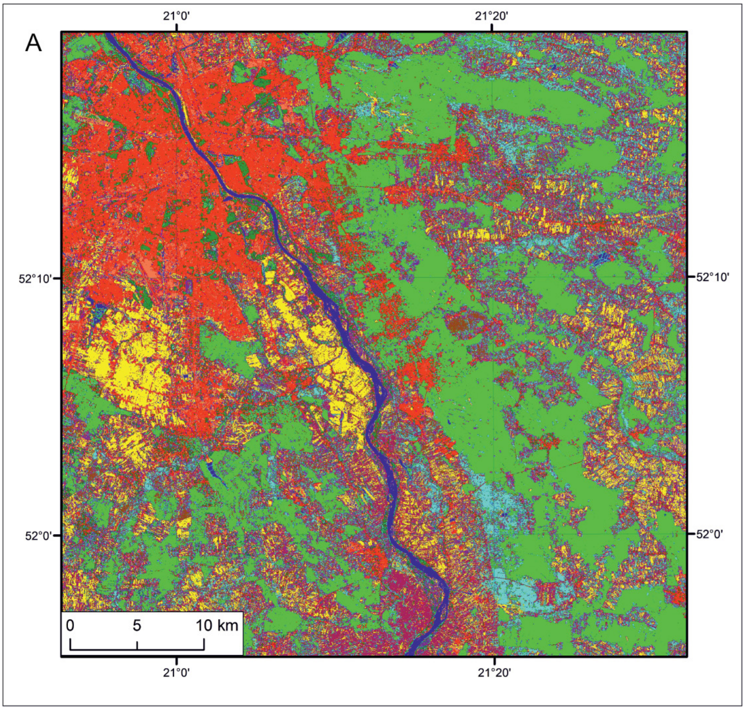
Tab. 2. Error matrix of the image classification acquired on 21 April 2011

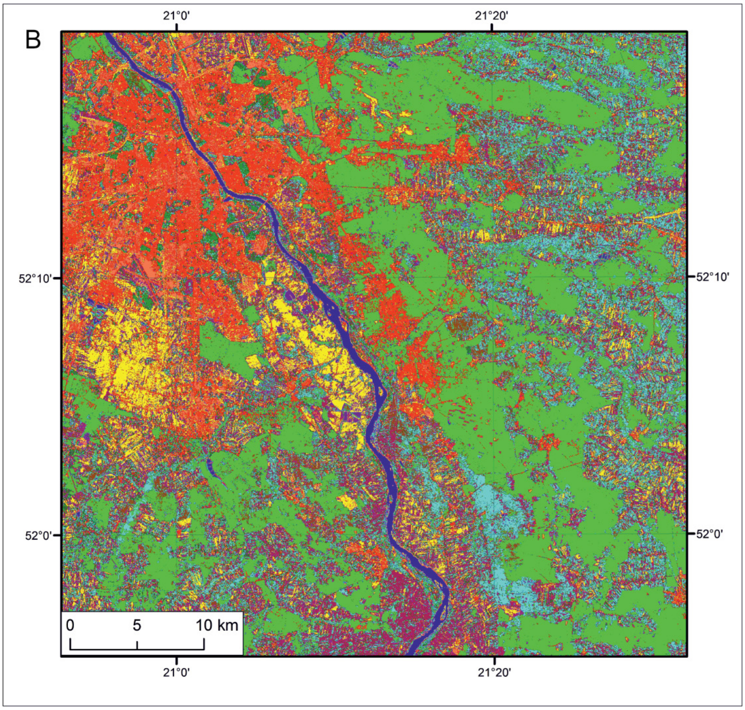
	User Acc.	Land cover types (number of classified pixels)									
		3.1	**1.1**	**2.1**	**2.3**	**5.1**	**2.2**	**3.2**	**2.4**	**1.4**	**1.2**
3.1 – Forests	92.13%	7084	0	0	67	0	9	458	71	0	0
1.1 – Urban fabric	86.93%	15	4424	33	36	15	17	6	56	192	295
2.1 – Arable land	77.30%	0	146	3653	98	15	299	215	201	29	70
2.3 – Pastures	68.50%	324	0	57	2460	0	156	347	202	30	15
5.1 – Inland waters	96.41%	3	41	39	4	3300	0	0	4	0	32
2.2 - Permanent crops	74.15%	50	78	197	145	0	2389	25	160	85	93
3.2 – Scrub and/or herbaceous vegetation associations	51.66%	49	3	93	477	0	142	1245	348	44	9
2.4 – Heterogeneous agricultural areas	18.33%	58	132	500	214	0	388	448	428	133	34
1.4 – Artificial, non-agricultural vegetated areas	65.83%	15	220	12	66	0	188	3	169	1308	6
1.2 – Industrial, commercial and transport units	50.03%	0	698	27	0	58	2	0	17	4	807
	Prod. Acc.	93.24%	77.05%	79.22%	68.97%	97.40%	66.55%	45.32%	25.85%	71.67%	59.29%

User Acc. – user accuracy
Prod. Acc. – producer accuracy

i.e. repeated pixel arrangements (R. Tadeusiewicz 2007). It relies on the physiology of human perception of objects which uses not only color but also object structure and texture (R. Lula, R. Tadeusiewicz 2001). A human is capable of recognizing objects by these features (B. Zagajewski 2010). An artificial neural network also uses such features and thanks to them and the input data it generates new information. Main research on neural networks was conducted in 1960-1980 and evolved with the development of information technology. The breakthrough came with association memory and creation of ART type network (G.A. Carpenter, S. Grossberg et al. 1992). Development of artificial neural networks resulted in their more common application in the process of classification (J.F. Mas 2004, H. Yuan et al. 2009, J. Olczyk 2014). M. Heinl et al. (2009) compared the outcome of land cover classification performed using artificial neural networks (total accuracy 86%) with the method of maximum likelihood (total accuracy 75%) and discriminant analysis (total accuracy 75%). Growing significance of artificial neural networks for classification of urban areas is worth noting, because of a wide range of spectral reflection of such areas (M.J. Aitkenhead, I.H. Aalders 2008; A. Iwaniak et al. 2002; M. Krówczyńska 2004).

Fuzzy ARTMAP simulator used by the authors was prepared by the team of Professor Paolo Gamba at the University of Pavia (G. Trianni 2007). It is a supervised network, which uses an earlier prepared model of maps with separately prepared training and verification polygons for classification. When creating polygons it is important that verification and training polygons do not overlap; such overlapping can have negative effect on the result of classification. Two patters are input into the network: the first – learning pattern, and the second – verification pattern. Parameters concerning neuron signal choice, network learning speed, probability factor and iteration number are also defined.


A
21°0'
21°20'
52°10'
52°0'
0
5
10 km


B
21°0'
21°20'
52°10'
52°0'
0
5
10 km

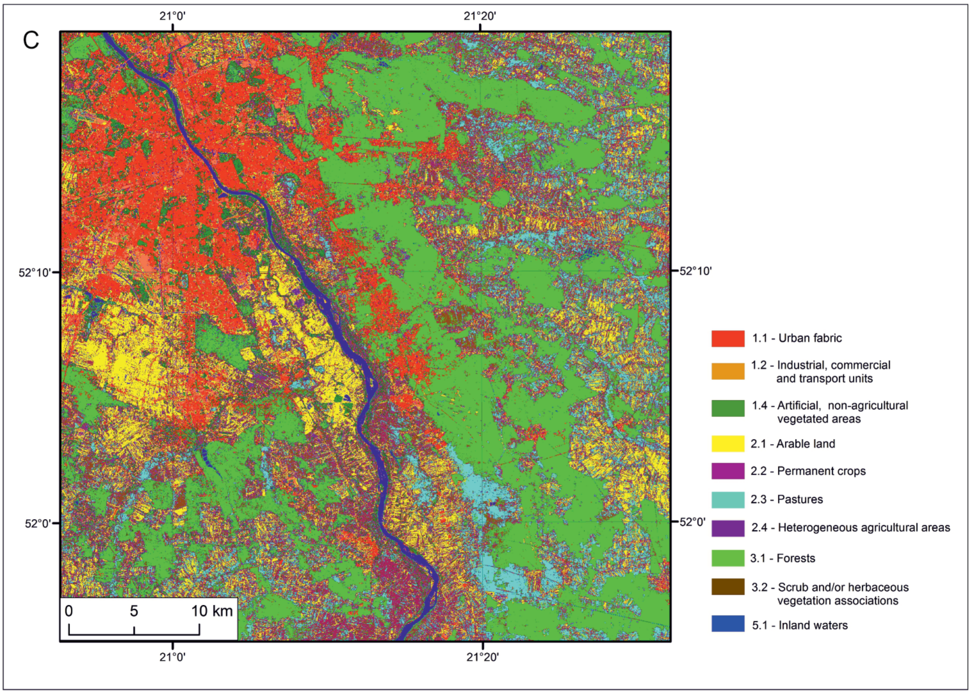


Fig. 3. Classification map of the image acquired on: A – 21.04.2011, B – 05.06.2010, C – 27.08.2011

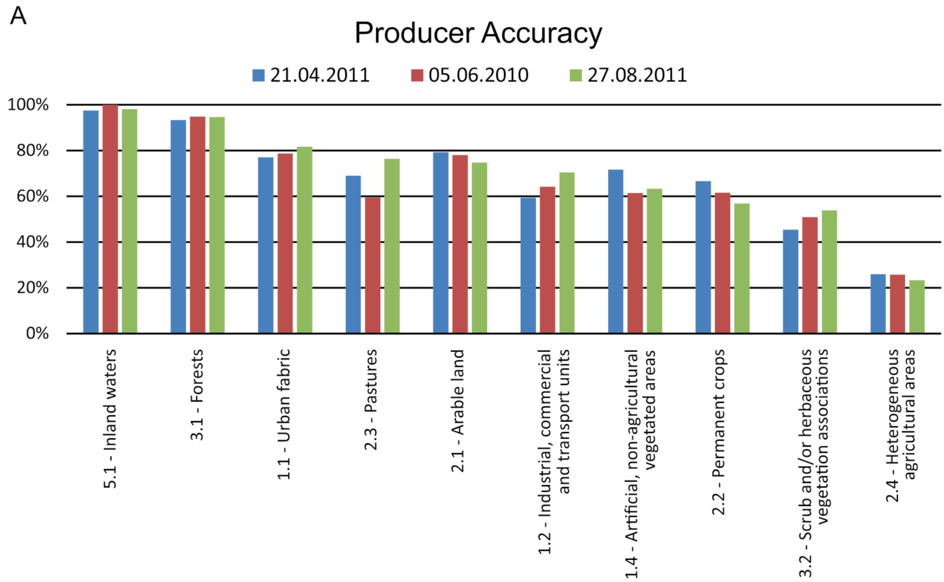
The network is supposed to learn the relations between testing layers, multispectral images and additional files, e.g. NDVI value layer (B. Zagajewski 2010). Fuzzy ARTMAP simulator is also a competing network. Within the network there are two simulators which simultaneously analyzed data. The image which is classified better is presented as the final result of network's operation (B. Zagajewski 2010). In their works B. Zagajewski (2010), U. Pytlak (2013) and M. Kacprzyk (2013) classified satellite images using fuzzy ARTMAP simulator. B. Zagajewski (2010), working on hyperspectral DAIS 7915 data, obtained producer accuracy of 86% and user accuracy of 75%. U. Pytlak (2013), classifying a Landsat satellite image, obtained total accuracy of 91%, and M. Kacprzyk (2013), also from a Landsat satellite image, obtained a total accuracy of 68%.

2. Research methods

Classification was done at the second level of Corine Land Cover legend, using 10 classes. Classification and evaluation of its accuracy were performed according to the following scheme (fig. 2).

Images acquired by Landsat 5 on 21 April 2011, 5 June 2010 and 27 August 2011 were obtained from the USGS' Earth Explorer (fig. 1A, 1B, 1C). At the first stage the CLC2012 was transformed from the PUWG 1992 coordinate system into UTM WGS-84 model using ArcMap ESRI 10.2 software, to comply with Landsat satellite images. Finally, the study area was subset.

In the next step the training and verification polygons were determined basing on CLC2012 database. In total, 10 classes of land cover were established in the analyzed area according to the second level of CLC legend: 1.1 urban fabric, 1.2 industrial, commercial and transport units, 1.4 artificial, non-agricultural vegetated areas, 2.1 arable land, 2.2 permanent crops, 2.3 pastures, 2.4 heterogeneous agricultural areas, 3.1 forests, 3.2 scrub and/or herbaceous vegetation, and 5.1 inland waters. Basing on the Landsat images, an NDVI normalized green vegetation index was calculated based on red



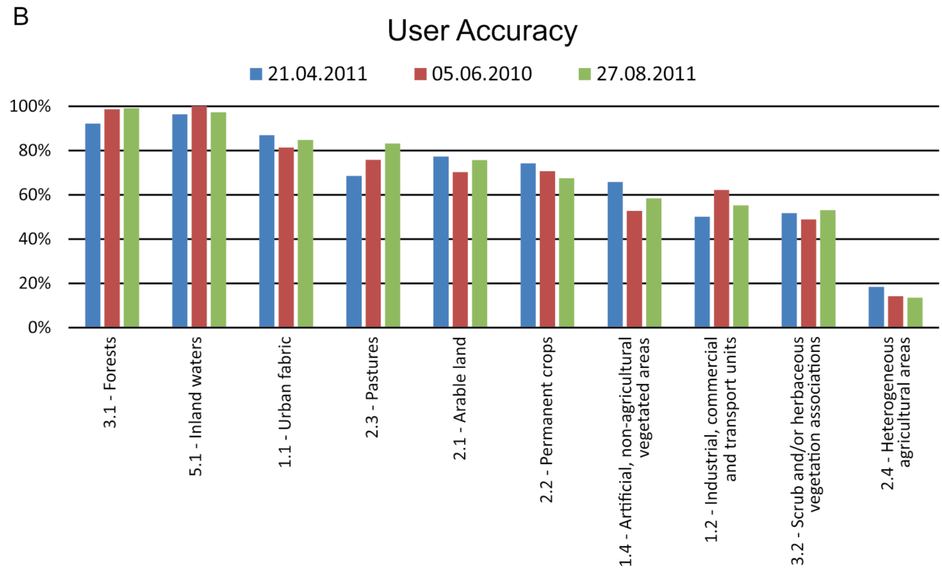


Fig. 4. A – producer accuracy of the second level CLC2012, B – user accuracy of the second level CLC2012

and infrared bands. NDVI determine the density and condition of the vegetation, thus can be useful to differentiate various vegetation types. The classification with the fuzzy ARTMAP simulator was performed using six bands of Landsat images (except thermal band), NDVI data and training and verification polygons in bitmap format. A post-classification image and an accuracy chart were obtained as a result of artificial neural network operation.

Finally, an evaluation of total accuracy, which shows the proportion of correctly classified pixels to their total number, was performed. For this purpose, an error matrix was built with verses containing number of properly classified pixels of each kind of land cover types according to the validation set and columns in which results of particular pixels representing analyzed classes originating from classification are recorded (B. Zagajewski 2010). The matrix informs how accurate (i.e. with what accuracy) the pixels in particular verification areas were classified (J.R. Jensen 2005).

3. Classification results

As a result of the classification procedure using polygons drawn on the base of CLC2012 map a set of post-classification images was obtained (fig. 3A, 3B, 3C). Error matrices presenting accuracy were also generated (table 2).

Total accuracies achieved for all post-classification images are comparable, about 70%. The best result (76%) was achieved for the image acquired in August (fig. 1C) and the worst for the image from June (fig. 1B, 68,6%).

The best classification was achieved classifying the image from 21 April 2011 (fig. 1A), where 75.1% pixels were correctly classified (fig. 3A). The error matrix (table 2a), presenting the accuracy of the classification, shows that 4 out of 10 classes of land cover are classified very well (user and producer accuracies 80–100%). These classes are: 5.1 inland waters (96.41–97.40%), 3.1 forests (92.13–93.24%), 1.1 urban fabric (77.05–86.93%), and 2.1 arable land (77.30–79.22%). Class 2.4 – heterogeneous agricultural areas were not classified properly, accuracies below 30% (producer accuracy 13.45%, user accuracy 23.3%). Remaining classes are matched with accuracy at the level of 50–70%.

Achieved results are in line with the results of the other researchers i.e. M. Łyko (2012) – total accuracy: 62%, J.F. Mas et al. (2004) – 74%, H. Yuan et al. (2009) – 88%, M.J. Aitkenhead and I.H. Aalders (2008) – 85%. Similar results were achieved by M. Kacprzyk (2013) for natural lowland ecosystems. It was also noted that the achieved accuracy depends on the type of land cover (its heterogeneity). U. Pytlak (2013) achieved the highest accuracy for large agriculture areas, where bigger inland waters scored the best results (98.1–99.9%), and homogenous arable land (94.9–96.3%) and forests (89.9–97.4%) were classified very well.

4. Conclusion

Basing on the achieved results (accuracy at the level of 68–76%) it was determined that the artificial neural networks can be a useful tool for elaboration of selected land cover types. Classification using non-automatic methods gives similar accuracy but requires more time and effort. Another advantage of artificial neural networks is that they lower the influence of error resulting from the lack of knowledge of land cover. The simulator classifies terrain based on satellite images. Land cover maps in various countries are elaborated by the same standards and parameters. Nevertheless it should be noted that results generated by artificial neural networks are cartographically incorrect and further rendering by adding additional elements is required.

Conclusions:

- The highest total accuracy was achieved on a Landsat 5 image acquired on 27 August 2011 (76.6%); the lowest accuracy (68.8%) was achieved on the spring image acquired on 21 April 2011.
- Despite similar values of the total accuracy for all images (about 70%), user and producer accuracy vary even by several dozen percent.
- Land cover classes like inland waters, forests and urban fabric were classified with the highest accuracy. For these classes artificial neural networks are a very good classification tool.
- The mixed crop classes were not classified correctly with the artificial neural networks.

Acknowledgements

The authors would like to thank the Main Inspectorate for Environmental Protection for granting access to the Corine Land Cover 2012 database. The unit responsible for the Corine Land Cover 2012 in Poland, within the Copernicus GIO Land Monitoring program financed by the E.U., was the Institute for Geodesy and Cartography, acting as one of the national EIONET reference centers for land cover (NRC Land Cover).

Literature

Anderson J.R., Hardy E.E., Roach J.T., Witmer R.E., 1976, *A land use and land cover classification system for use with remote sensor data*. „Geological Survey Proffesional Paper" No. 964, pp. 3–10.

Aitkenhead M.J., Aalders I.H., 2008, *Classification of Landsat Thematic Mapper imagery for land cover using neural networks*. „Intern. Journal of Remote Sensing" Vol. 29, no. 7, pp. 2075–2084.

Baranowski M., Ciołkosz A., 1997, *Opracowywanie baz danych pokrycia terenu dla Polski*. „Prace Instytutu Geodezji i Kartografii" T. 44, z. 95, pp. 7–28.

Bielecka E., Ciołkosz A., 2004, *Metodyczne i realizacyjne aspekty aktualizacji bazy Corine Land Cover*. „Prace Instytutu Geodezji i Kartografii" T. 50, z. 108, pp. 73–92.

Bielecka E., Ciołkosz A., 2005, *Baza danych Corine Land Cover*. Biblioteka Monitoringu Środowiska. Warszawa: Inspekcja Ochrony Środowiska, 76 pp.

Campbell J.B., Wynne R.H., 2011, *Introduction to Remote Sensing.* Fifth ed. United States of America, The Guilford Press, pp. 335–382.

Carpenter G.A., Grossberg S., Markuzon N., Reynolds J.H., Rosen D.B., 1992, *Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps*. „IEEE Transactions on Neural Networks" Vol. 3, no. 5, pp. 698–713.

Commission of the European Communities, 1995, *Corine Land Cover*, The European Environment Agency (EEA), http://www.eea.europa.eu/publications/COR0-landcover (access: 21.12.2015 r.).

Feranec J., Hazeu G., Christensen S., Jaffrain G., 2007, *Corine land cover change detection in Europe (case studies of the Netherlands and Slovakia)*. „Land Use Policy" No. 24, pp. 234–247.

Heinl M. i in., 2009, *Classifiers vs. input variables – The drivers in image classification for land mapping*. „International Journal of Applied Earth Observation and Geoinformation" Vol. 11, no. 6, pp. 423–430.

Hościło A., Tomaszewska M., 2015, *CORINE Land Cover 2012 – 4th CLC inventory completed in Poland*. „Geoinformation Issues" Vol. 6, no. 1(6), pp. 49–58.

Iwaniak A., Krówczyńska M., Paluszyński W., 2002, *Użycie sieci neuronowych do klasyfikowania obszarów miejskich na zdjęciach satelitarnych*. „Geodesia et Descriptio Terrarum" Nr 1(1-2), pp. 5–13.

Jensen, J.R., 2005, *Introductory digital image processing: a remote sensing perspective.* Third ed. United States of America, Prentice Hall.

Kacprzyk M. 2013, *Klasyfikacja pokrycia terenu Podlasia na podstawie zdjęć Landsat TM i symulatora sztucznych sieci neuronowych fuzzy ARTMAP*. Warszawa: Wydział Geografii i Studiów Regionalnych UW, unpublished licentiate thesis, 83 pp.

Krówczyńska M., 2004, *Wykorzystanie spektralnych i strukturalnych cech obiektów odwzorowanych na zdjęciach satelitarnych w kartowaniu użytkowania ziemi*. Warszawa: Wydział Geografii i Studiów Regionalnych UW, unpublished doctor dissertation.

Lula P., Tadeusiewicz R., 2001, *Wprowadzenie do sieci neuronowych*. Kraków, Statsoft Polska.

Łyko M., 2012, *Opracowanie mapy pokrycia/użytkowania terenu w oparciu o klasyfikację danych wielospektralnych LANDSAT i IKONOS*. AGH, unpublished master thesis, 76 pp.

Mas J.F., 2004, *Mapping land use/cover in a tropical coastal area using satellite sensor data, GIS and artificial neural networks.* „Estuarine Coastal and Shelf Science" No. 59, pp. 219–230.

Olczyk J., 2014, *Klasyfikacja pokrycia terenu województwa warmińsko-mazurskiego na podstawie zdjęć Landsat TM i symulatora sztucznych sieci neuronowych*. Warszawa: Wydział Geografii i Studiów Regionalnych UW, unpublished master thesis.

Pytlak U. 2013, *Klasyfikacja pokrycia terenu Wielkopolski na podstawie zdjęć Landsat TM i symulatora sztucznych sieci neuronowych fuzzy ARTMAP*. Warszawa: Wydział Geografii i Studiów Regionalnych UW, unpublished licentiate thesis, 30 pp.

Tadeusiewicz R., 2007, *Sieci neuronowe*. 2nd ed., Warszawa: Akadmicka Oficyna Wydawnicza. RM.

Trianni G., 2007, *Techniques for fusion of remotely sensed data over urban environments*, Universita Degli Studi Di Pavia, Dottorato Di Ricerca in Ingegneria Elettronica Elettrica ed Informatica, XX Ciclo, Doctor dissertation, promotor Prof. dr Paolo Gamba, Pavia, 140 pp.

Yuan H., Van Der Wiele C.F., Khorram S., 2009, *An autoated neural network system for land use/land Cover classification from Landsat TM imagery*. „Remote Sensing" No. 1, pp. 243–265.

Zagajewski B., 2010, *Ocena przydatności sieci neuronowych i danych hiperspektralnych do klasyfikacji Tatr Wysokich*. „Teledetekcja Środowiska" T. 43, 112 pp.

Zagajewski B., Jarocińska A., Olesiuk D., 2009, *Metody i techniki badań geoinformatycznych*. Warszawa: Wydział Geografii i Studiów Regionalnych UW, 117 pp.

Żmudzka E., 2012, *Wieloletnie zmiany zasobów termicznych w okresie wegetacyjnym i aktywnego wzrostu roślin w Polsce*, „Woda – Środowisko – Obszary Wiejskie" T. 12, z. 2, pp. 377–389.