www.czasopisma.pan.pl PAN www.journals.pan.pl

GEODESY AND CARTOGRAPHY
Vol. 67, No. 1, 2018, pp. 99–116

DOI: https://doi.org/10.24425/118699

Original research paper

Using a GEOBIA framework for integrating different data sources and classification methods in context of land use/land cover mapping

Anna Osmólska[1]*, Paweł Hawryło[2]

[1]University of Agriculture in Krakow
Faculty of Animal Science
24/28 Mickiewicza Av., 30-059 Kraków
e-mail: osmolska.anna@gmail.com, ORCID: https://orcid.org/0000-0001-6834-0886

[2]University of Agriculture in Krakow
Faculty of Forestry
46/29 Listopada Av., 31-425 Kraków
e-mail: p.hawrylo@ur.krakow.pl, ORCID: https://orcid.org/0000-0002-0170-8866

*Corresponding author: Anna Osmólska

Received: 27 January 2018 / Accepted: 10 May 2018

Abstract: Land use/land cover (LULC) maps are important datasets in various environmental projects. Our aim was to demonstrate how GEOBIA framework can be used for integrating different data sources and classification methods in context of LULC mapping. We presented multi-stage semi-automated GEOBIA classification workflow created for LULC mapping of Tuszyma Forestry Management area based on multi-source, multi-temporal and multi-resolution input data, such as 4 bands- aerial orthophoto, LiDAR-derived nDSM, Sentinel-2 multispectral satellite images and ancillary vector data. Various classification methods were applied, i.e. rule-based and Random Forest supervised classification. This approach allowed us to focus on classification of each class 'individually' by taking advantage from all useful information from various input data, expert knowledge, and advanced machine-learning tools. In the first step, twelve classes were assigned in two-steps rule-based classification approach either vector-based, ortho- and vector-based or ortho- and Lidar-based. Then, supervised classification was performed with use of Random Forest algorithm. Three agriculture-related LULC classes with vegetation alternating conditions were assigned based on aerial orthophoto and Sentinel-2 information. For classification of 15 LULC classes we obtained 81.3% overall accuracy and kappa coefficient of 0.78. The visual evaluation and class coverage comparison showed that the generated LULC layer differs from the existing land cover maps especially in relative cover of agriculture-related classes. Generally, the created map can be considered as superior to the existing data in terms of the level of details and correspondence to actual environmental and vegetation conditions that can be observed in RS images.

Keywords: data fusion, random forest, supervised classification, Sentinel-2

1. Introduction

Since ecosystems include interrelated biotic and abiotic elements (Begon et al., 2006), investigation and description of these elements and interactions between them needs broad integrative approach. Thus, recent ecological research have become more interdisciplinary. Associations with other science fields like probability theory, geoscience (geodesy, cartography, geology), physics, hydrology, climatology etc. (Guisan and Zimmermann, 2000; van Dam et al., 2013; de Vienne, 2016; Oktavia and Prativi, 2017; Garcia, 2014; Stephens et al., 2016) are often considered and modern methods, technologies, approaches or solutions are used. The ones of increasing importance are geographic information system (GIS) and remote sensing (RS) technologies. RS and GIS can provide comprehensive and objective information related to land cover and topography used as the input data for such research as species distribution models. These technologies can be also considered as a source of methods and solutions for data analysis and integration.

Landscape and conservation research or biodiversity monitoring based on spatial analyses and modelling process requires the high quality input data. The one with particular application in environmental monitoring, biodiversity research, nature conservation or natural-resources management is land use/land cover (LULC) map. Robust ecological studies in fine scale like species-habitat relationship investigation requires high quality and high resolution LULC data (Zanariah et al., 2012). Unfortunately, in Polish conditions, the updated fine-scale LULC maps for the land use-diversified areas with great share of agriculture landscape (i.e. Tuszyma Forestry Management area) are not available. Thus, the performance of any profound ecological research requires new high quality and high resolution LULC map to be produced. However, this could be challenging, since the geodata from various sources like remote sensing (high-resolution true color and false color composite aerial orthophotos, multispectral satellite images, LiDAR point clouds), ground truth campaign or existing databases could not be singly enough to perform reliable image classification, with high level of details. Thus, the data fusion approach could be markedly helpful.

One of the image classification method which allows data fusion is Geographic Object Based Image Analysis (GEOBIA). GEOBIA or OBIA is a method that partition remote sensing imagery into meaningful image-objects and assess their characteristics through spatial, spectral and temporal scale (Hay and Castilla, 2006). In this image classification method, instead of analyzing and classifying single pixels, image objects are generated through different segmentation methods (Hoffmann et al., 2011). Image-objects (or segments) exhibit many useful features like diversified spectral information and spatial features of objects (shape, distances, neighbourhood, topologies) (Hay and Castilla, 2006; Blaschke, 2010). GEOBIA simulates the human way of visual landscape or image interpretation. First, the spectral information of the pixels are aggregating into objects and then the objects are classified based on the predefined expert knowledge (Hoffmann et al., 2011).

One of the main advantage of GEOBIA is the opportunity to integrate data and knowledge from vast array of disciplines. It allows the geospatial data beyond images, or data of different quality, spatio-temporal scales or resolutions to be used (Blaschke et al., 2014). It can be easily integrated with commonly used GIS environment. The data integration approach has been proved to increase classification accuracy (Nordkvist et al., 2012, Huang and Zhu, 2013). The image classification schemes are often based on the dataset including ALS (airborne laser scanning) data combined with multispectral airborne-imagery (Syed et al., 2005, Hermosilla et al., 2010, Szostak et al., 2014; Piiroinen et al., 2015), or with hyperspectral satellite imagery (Augilar et al., 2012; Huang and Zhu, 2013). Additionally, the vector data can be incorporated in image classification framework (Frontoni et al., 2010, Tiede et al., 2010; Wężyk et al., 2016). Smith & Morton (2010) proposed to use the best existing real world feature datasets as the starting point for segmentation. GEOBIA approach allows not only the data fusion but also integration of different classification methods, such as combined object- and pixel- based supervised image classification framework (Bernadini et al., 2010; Salehi et al., 2013).

The aim of this paper was to demonstrate how GEOBIA framework can be used for integrating different data sources and classification methods in context of land use/land cover mapping. For this reason, we presented multi-stage semi-automated GEOBIA classification workflow created for LULC mapping of Tuszyma Forestry Management area. Data integration approach allowed utilization of multi-source, multi-temporal and multi-resolution input data such as orthophoto imagery, nDSM (normalized Digital Surface Model), satellite imagery and vector layers. Using diversified input data types allowed incorporation of different information related to multispectral reflectance, terrain surface or exact location and shape of some land use objects into segmentation and classification processes. It was also demonstrated that in GEOBIA approach one classification workflow can involve rule-based classification for some land use classes assignment and supervised classification for other, in this case agriculture-related classes. The second purpose of presented research was to obtain a high quality and resolution LULC map for the study area that can provide a significant input for various ecological and landscape analysis, which can be useful for vast array of biodiversity, conservation, wildlife and land management purposes.

2. Methods

2.1. Study area

We have created the GEOBIA classification workflow for Tuszyma Forestry Management area located in South-Eastern Poland in Podkarpackie Province (Fig. 1) that covers approx. 53,000 ha. It is slightly ridged plain elevated 150 to 230 m a.s.l. The area is located within Wisła basin, in two rivers – Wisłoka and Breń watersheds and the river system is of dendritic pattern. Highly diversified soil conditions within the study area impact on highly diversified habitat and land-use conditions with extensive and small woodlands, more agriculture-dominated landscape in the North and with more shrublands and wastelands in the South. The farmland fragmentation is high with the average area of a single farm below 4.5 ha. There is only one big city within the study area – Mielec (approx. 47 km^2, 60.6 thousand inhabitants). Other settlements are small towns or villages. The southern part of study area is crossed by the A4 motorway with east-west direction. There are three nature reserves located within Tuszyma Forest District with overall area of approx. 70 ha that preserve the most precious parts of forest, water or grassland-cropland ecosystems.

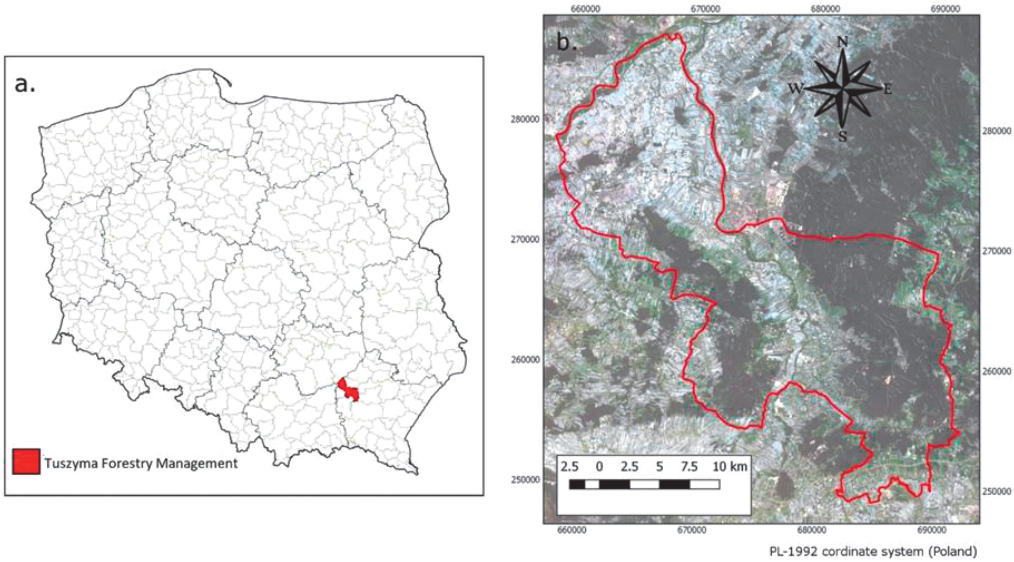

Fig. 1. Location of the study area in south-eastern Poland (a). The Tuszyma Forestry Management borders on Sentinel-2 RGB composite background (b)

2.2. Datasets and pre-processing

To achieve the research goals the variety of data types were used and integrated in one classification workflow (Table 1). All input layers covered whole study area. High resolution aerial orthophoto acquired in June 2015 included four spectral bands (NIR, Red, Green, Blue) with 0.25 m spatial resolution. In order to decrease the input layers size and required computing power, the aerial orthophotos were resampled to 1 m spatial resolution using bilinear interpolation technique. Resampling allowed to obtain orthoimages with the same spatial resolution as nDSM layer and still representing high level of detail. It has not big influence to the process of classification, since land use objects with area smaller than 1 m was not considered. nDSM layer with 1 m spatial resolution was generated from Airborne Laser Scanning point clouds acquired in September 2014. Both, orthoimages and ALS (LiDAR) data were collected from Polish National Geodesy Office database. Supervised classification was performed with use of multispectral satellite imageries from Sentinel-2 Earth observation mission obtained in two points of time: March 2016 and August 2016. These two dates correspond to significant periods in vegetation season and can be useful in distinguishing various agricultural classes. Ten spectral bands of Sentinel-2 images were used (bands: 2, 3, 4, 5, 6, 7, 8, 8A, 11, 12) with 10 m or 20 m spatial resolution, all sourced from European Space Agency (ESA) database. Spatial information from remotely sensed images were supported by GIS vector data including six layers: buildings, roads, cadastral parcels, land use classes, ditches and flowing water collected from National and Regional Geodesy Offices from BDOT database. Since the confrontation of roads and buildings vector data with orthophoto images showed that these layers needed updating, the corresponding layers from OpenStreetMap database were obtained (updated to March 2017). Datasets from both sources were merged. Based on the information about different road subclasses included in obtained datasets, the final roads layer subclasses were grouped into three classes: Motorway, Paved roads and Ground roads. Each input layer was clipped to the study area extent and if needed reprojected to one coordination system (PL-1992; EPSG:2180). All of the pre-processing was carried out in QGIS and ArcGIS (Esri) environment.

Table 1. Multi-source, multi-temporal and multi-resolution input data used in presented semi-automated LULC classification workflow

Data	Type	Spatial resolution	Source	Date of acquisition	Pre-processing
Aerial orthophoto	4 bands (NIR, Red, Green, Blue)	0.25 m resampled to 1 m	National Geodesy Office	June 2015	Resampling; clipping to study area extent
Lidar data	nDSM	1 m	National Geodesy Office	September 2014	Clipping
Satellite imagery – Sentinel 2	Bands: 2, 3, 4, 5, 6, 7, 8	Bands: 2, 3, 4, 8: 10 m; 5, 6, 7, 8A, 11, 12: 20 m	ESA	March 2016, August 2016	Reprojecting; clipping
Vector data	6 layers: buildings, roads, parcels, land use LU, ditches, flowing water		National and Regional Geodesy Offices, OSM		Reprojecting; databases unification; clipping; buildings and roads layer: two-sources layers merging; roads: subclasses grouping

2.3. Workflow

The multi-stage GEOBIA classification workflow was applied as presented in Figure 2. First two steps of semi-automated LULC classification involved data collection and preprocessing performed in QGIS and ArcGIS software. For multistep image object classification the eCognition Developer (TRIMBLE GeoSpatial) software was used. The GEOBIA classification procedure includes two main stages: the segmentation and classification of resulting segments. Generally, classification stage is either expert knowledge-based set of rules or supervised classification method based on training samples. Classification workflow applied in our study integrates both approaches and is comprised of two steps: rule-based classification and then supervised classification with Radom Forest method. Spectral information from aerial orthophoto bands was expanded by appending temporary layers of two vegetation indices (VIs): Normalized Difference Vegetation Index (NDVI) and Green Chlorophyll Index (CIGreen). These VIs proved to be accurate and linear estimator of canopy chlorophyll and N content (Clevers and Gitelson, 2013), thus can be helpful in vegetation classification. Additionally, the Normalized Difference Water Index was used (Kaplan and Avdan, 2017).

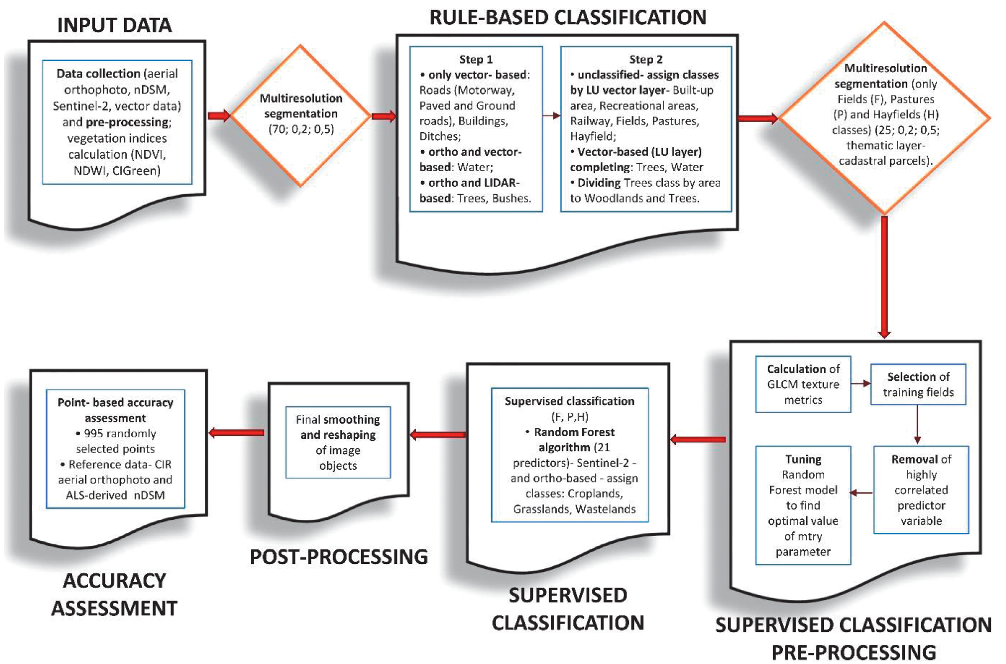

Fig. 2. Multi-stage GEOBIA classification workflow for Tuszyma Forestry Management area

The primary step of image-object analysis is segmentation, the process of grouping neighboring pixels into homogenous segments. During this process, the user can select the size and homogeneity criteria of the objects. In this research, we used multiresolution segmentation algorithm from eCogniton Developer software, which is proven to be one of the most successful image segmentation algorithms (Witharana and Civco, 2014). It is an optimization procedure which, for a given number of image objects, minimizes the average heterogeneity and maximizes their respective homogeneity (Baatz and Schape, 2000). The scale parameter is considered to be one of the most important variables because it controls the relative size of the image objects, thus has a direct impact on the subsequent classification steps (Ma et al., 2015). For spectral information it is possible to define weights for each spectral band. The shape parameter defines the value of compactness and smoothness. Optimal segmentation parameters were determined based on a trail-and-error approach and previous experience of operators. In the presented workflow, the following parameters of multiresolution segmentation were used:

– scale parameter: 20
– shape parameter: 0.2
– spectral bands weights: Red, Green, Blue, NDWI-1; NIR, NDVI, CIGreen, nDSM-2
– compactness/smoothness: 0.5.

First step of rule-based classification included only vector-based class labelling for some LU classes (Motorway, Paved roads, Ground roads, Buildings, Ditches), ortho and vector-based for others (Water) or ortho and LiDAR-based (Trees, Bushes). The roads layer was used to create new segments with exact location of roads and then assign three roads classes (Motorway, Paved roads, Ground roads) based on the attributes of vector layer. According to them, the segments were extended with use of pixel-based object resizing algorithm. Similarly, the Buildings segments extraction and class labelling was only vector-based. Water was assigned based on spectral information from orthoimages, specifically on a mean value of NDWI index and standard deviation of Blue band. The flowing water vector layer was used for completing Water class. The unclassified objects with mean nDSM value above 1.5 m and NDVI >0.05 were assigned to Trees class. Similar process was performed for Bushes class (mean nDSM above 1 m). The Ditches class segmentation and assignment was carried out only vector-based.

The first step of rule-based classification described above, produced a certain mask comprised of eight classes. In the next step, for remaining unclassified objects the following classes were assigned based on LU vector layer: Built-up area, Recreational area, Railway, Fields, Pastures, Hayfields. The three latter classes representing agricultural land use were submitted to supervised classification in the subsequent step and reclassified to: Croplands, Grasslands and Wastelands. Built-up area class was expanded by classifying the selected segments neighboring to Buildings class. LU layer was also used for expanding and complementing Water and Trees classes. Once all Trees were classified and segments were merged, two classes were distinguished by area: objects above 0.05 ha were assigned as Woodlands while below 0.05 ha as Trees.

The aim of next step of the workflow was to distinguish three classes – Croplands, Grasslands and Wastelands – using the supervised classification approach. Firstly, new temporary layers only Sentinel-based were created: NDVI and CIGreen indices for each date. Then, the multiresolution segmentation was performed for only three classes: Fields, Pastures, Hayfields. The parameters of the multiresolution segmentation were as follows: scale: 25; shape: 0.2; compactness: 0.5; spectral bands weights: CIGreen, NDVI: 2, Blue, Green, NIR, Red: 1. Parcels vector layer was used in segmentation. The created segments with associated attributes were then exported to a vector layer and analysed in the R software environment. The segments' attributes were both ortho- and Sentinel-2-based and consisted of: mean values of all spectral bands of aerial orthophoto and two Sentinel-2 images; mean and standard deviation values of calculated VIs; selected Haralic texture metrics calculated from aerial orthophoto (GLCM Ang. 2nd moment, GLCM Contrast, GLCM Correlation, GLCM Dissimilarity, GLCM Entropy, GLCM Homogeneity, GLCM Mean, GLCM Standard Deviation). The values of these variables were calculated for each segment which allowed to integrate information from datasets with different spatial resolution. As training samples 450 segments were selected (150 segments for each class). Samples were labelled based on ground truth data from field visits and visual interpretation of high resolution 0.25 m orthophoto CIR composite and Sentinel-2 satellite images performed by trained operator.

Segments' attributes exported from eCognition software were used as predictor variables in supervised classification. Prior to classification execution the highly correlated predictors were removed. First, the correlation matrix based on Pearson correlation coefficient was created for all used predictor variables. Then, if two variables had a high correlation (>0.9) the variable with the higher mean absolute correlation to all other variables was removed.

Supervised classification was performed using Random Forest (RF) classification method. Many key advantages of RF algorithm which include their non-parametric nature, high classification accuracy and capability to determine variable importance (Rodriguez-Galiano et al., 2012) cause that it is of increasing interest and use in LULC classification projects. RF is an ensemble method using defined number of simple decision trees. Each tree is independently determined using a bootstrap sample of the data. The most popular class from the trees' votes is considered as final prediction. To reduce the correlation among trees the algorithm randomly selects defined number of predictors at each split (Breiman, 2001). The number of trees for the forest in the study was set to 1000 while the optimal number of randomly selected predictors for splits was tuned from the following values: 2, 6, 10, 14, 18 using 10-fold cross-validation method. The predictive model was created using the randomForest package for R (Liaw and Wiener, 2002). The variable importance was calculated as mean decrease in classification accuracy after permutation of a certain variable. For each tree, the error rate for classification was computed on the out-of-bag portion of the data. Then the error rate was calculated after permuting each predictor variable. The difference between the obtained error rates were then averaged over all trees, and normalized by the standard deviation of the differences (Liaw and Wiener, 2002).

Rule-based and supervised classification were followed by post-processing step. It included manual editing process involved visual checking of the results and corrections in case of explicit errors in class assignment, final smoothing and reshaping in order to obtain more compacted image objects, merging segments within one class, or removing objects smaller than user-defined minimum area. Most corrections of explicit errors was performed among classes related to secondary succession i.e. Wastelands, Bushes and Trees, as well as Trees vs. Woodlands classes. After verification of entire LULC layer, the improved results underwent accuracy assessment.

Accuracy assessment was performed on 995 randomly selected points. Although point-based accuracy assessment can be considered as unstable and inferior to area-based method (Ma et al., 2017), point-based approach was chosen for this evaluation because land cover could be unambiguously determined for each reference point using a combination of aerial orthophoto and nDSM. The number of required test points to generate an error matrix was calculated using the equation for a multinomial distribution provided by Congalton and Green (2009). Number of test points for each class was then calculated based on the proportion of the area covered by this class however, the minimum number of points for each class was set to 20. Calculated numbers of points for each class was then randomly placed within the study area. Reference data for each test point was collected through on-screen visual interpretation of 0.25 m resolution CIR aerial orthophoto and 1 m resolution nDSM. An error matrix with producer (PA; the ratio between the number of samples correctly classified in that category and the total number of samples observed as belonging to that category (Congalton, 1991)) and user accuracies (UA; the ratio between the number of samples correctly classified in that category and the total number of samples classified to respective category (Congalton, 1991)) was generated along with the overall accuracy (OA) and Kappa coefficient.

3. Results

We performed multi-stage semi-automated GEOBIA classification workflow integrating rule-based and supervised classification methods and based on multi-source, multi-temporal and multi-resolution input data. As the result, the LULC vector layer for Tuszyma Forestry Management area was generated. The results of classification along with overall accuracy and Kappa coefficient was shown in Table 2. The overall layer covers 642.26 km^2 and objects are classified into 15 different LULC classes. There are three LULC classes predominating in the study area: Woodlands which covers 31.89%, Croplands (26.47%) and Wastelands (18.9%). However, the agricultural land defined as the area used for farming (in this case the Croplands and Grasslands classes) covers over 33% of the entire study area and it slightly exceeds the Woodland class. Three classes which represents the land out of the agricultural use as well as secondary forest succes-

Table 2. The results of multi-stage GEOBIA classification workflow: the LULC classes surface coverage along with overall accuracy and Kappa coefficient

Overall accuracy = 81.3% Kappa coefficient = 0.78		
LULC class	**Area [km^2]**	**Area [%]**
Buildings	5.42	0.85
Built-up	31.56	4.97
Bushes	3.20	0.50
Croplands	168.08	26.47
Ditches	5.02	0.79
Grasslands	43.92	6.91
Ground roads	1.56	0.25
Motorway	0.54	0.09
Paved roads	12.94	2.04
Railway	0.29	0.05
Recreational	0.86	0.14
Trees	31.70	4.99
Wastelands	120.01	18.90
Water	7.47	1.18
Woodlands	202.51	31.89
Overall	635.08	100.00

sion which are Wastelands, Bushes and Trees together cover over 24% of the study area. The Built-up and Buildings classes cover less than six percent, indicating the fact the built-up area within Tuszyma Forestry Management area is relatively small and sparse relating to occurrence of only one big city and more small towns and villages within the area. All roads represent 2.37% of the study area.

The obtained overall accuracy of the classification process defined as the ratio between the number of correctly classified samples and the total number of samples equals 81.3% and the Kappa coefficient value equals 0.78. The accuracy assessment results are presented in confusion matrix (Tab. 3) with given Producer's Accuracy (PA) and User's Accuracy (UA) for each class. The values of PA ranged from 0.4 for Trees class to 1 for Motorway class and the average Producer's Accuracy reached value 0.8. The UA ranged from 0.57 for Ground roads class up to 1 for Motorway class. The average User's Accuracy value equals 0.79. Particular attention should be paid to the results of accuracy assessment for three classes assigned by Random Forest supervised classification which are Croplands, Grasslands and Wastelands. Best results were obtained for Croplands class with PA = 0.94 and UA = 0.88, for Grasslands class it equals 0.72 and 0.8, respectively and for Wastelands class both PA and UA values equal 0.73. Both average Producer's and User's accuracy for these three classes equal 0.8 and it was equal to values obtained for all classes. Overall, most confusion was observed within Bushes and

Table 3. Confusion matrix

		Buildings	Built-up	Bushes	Croplands	Ditches	Grasslands	Ground roads	Motorway	Paved roads	Railway	Recreational	Trees	Wastelands	Water	Woodlands	PA
Reference data	Buildings	19	9	0	1	0	0	0	0	0	0	0	0	0	0	0	0.66
	Built-up	1	28	1	2	1	1	2	0	1	2	3	0	2	0	0	0.64
	Bushes	0	0	15	0	0	0	0	0	0	0	0	4	5	1	2	0.56
	Croplands	0	1	0	180	1	2	1	0	0	0	0	1	6	0	0	0.94
	Ditches	0	0	0	1	14	0	0	0	0	0	0	0	0	0	0	0.93
	Grasslands	0	0	0	13	1	43	1	0	0	0	0	0	2	0	0	0.72
	Ground roads	0	0	0	0	0	0	34	0	5	0	0	0	1	0	0	0.85
	Motorway	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	1
	Paved roads	0	0	0	0	0	0	14	0	43	0	0	0	0	0	0	0.75
	Railway	0	0	0	0	1	0	0	0	0	15	0	0	0	0	0	0.94
	Recreational	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	1
	Trees	0	1	1	0	1	0	1	0	4	1	0	27	21	2	8	0.4
	Wastelands	0	0	3	8	1	7	1	0	2	2	0	5	107	1	9	0.73
	Water	0	0	0	0	0	0	0	0	0	0	1	0	0	16	0	0.94
	Woodlands	0	0	0	0	0	1	6	0	5	0	0	2	3	0	232	0.93
	UA	0.95	0.72	0.75	0.88	0.7	0.8	0.57	1	0.72	0.75	0.8	0.69	0.73	0.8	0.92	

Tree classes. First class was often confused with Trees and Wastelands and the latter with Wastelands and Woodlands.

The results of variable importance analysis presented in Fig. 3 showed that the contribution of mean ortho-based NDVI was the highest with the value of mean decrease of accuracy equal 46.6. High values were also reached by other vegetation indices variables: Sentinel-based CIGreen Standard Deviation for August (37.5) and ortho-based CIGreen Standard Deviation (35.3). Two Sentinel Vegetation Red-Edge bands (5 and 6), along with SWIR band (11) showed relatively high importance with values of mean decrease of accuracy ranged from 33.3 to 28.9. Interestingly, the variables that contributed the least were Sentinel-based VI's Standard Deviations with the lowest value equal 9 for NDVI Standard Deviation for August.

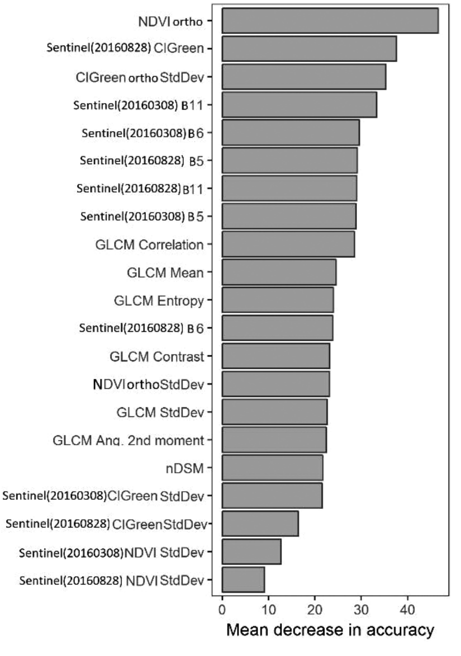

Fig. 3. Importance of predictor variables used for supervised Random Forest classification in terms of mean decrease in accuracy

The visual "on-screen" interpretation (Figure 4) as well as the tabulation of surface coverage of specified land use classes (Table 4) of the generated LULC layer and the existing land cover maps were performed. Selected region of the study area presented in Figure 4 shows that the CORINE Land Cover (CLC) layer from year 2012, apart from being not-updated is very generalized and only indicates wide general classes with large objects and low level of details. There is no roads class or any class related to ditches, bushes or small trees patches included in this layer. The land use vector layer available from BDOT database is much more detailed in terms of number of distinguished classes and patch size compared to CLC map, however it often reflects rather the legal status of the area under consideration than the actual present land cover conditions not including

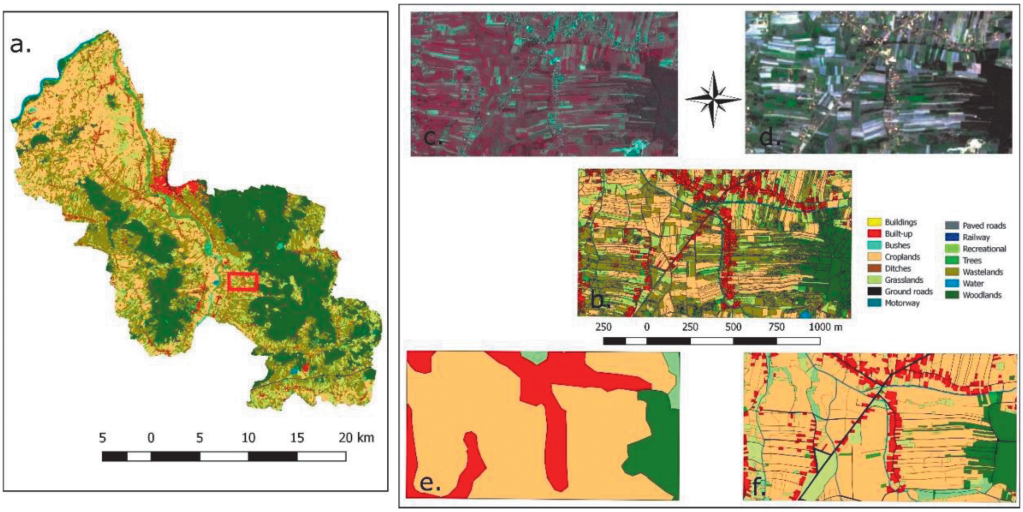

Fig. 4. LULC map of Tuszyma Forestry Management area (a); zoom to OBIA LULC map of selected region (b); orthophoto CIR composite (c); Sentinel-2 RGB composite (d); CORINE Land Cover 2012 (e); land use layer from BDOT database (f)

the changes resulting for example from farmland abandonment or secondary forest succession. Thus, this layer may also be considered as out-of-date. Additionally, the class borders often correspond to cadastral parcels borders and do not reflect the environmental conditions. The LULC map generated in this study, although based on datasets from different points of time, generally can be considered as up-to-date, since input data used for classification of fast alternating LU classes like buildings, roads, or agriculturally related classes were most updated. General visual evaluation shows that objects properly correspond to the shape of land use patches and the assigned classes correctly reflects the land cover conditions.

The comparison of surface coverage of specified LU classes from the three LULC maps mentioned above was presented in Table 3. The greatest differences were observed within Wastelands, Grasslands and Croplands classes. According to BDOT LULC map, Croplands class covers 43.45% of the study area, Grasslands covers 12.69% and Wastelands relative surface area is only 0.22%. In the CLC layer the relative surface area of Croplands is the highest (47.86%) from all three layers, Wastelands (6.71%) is higher than from BDOT but lower than from OBIA classification, and Grasslands for CLC is the lowest (4.71%). The coverage of other classes in OBIA classification and BDOT LULC maps are comparable. The mentioned three agriculture-related LULC classes with the greatest differences between-LULC layers were those which we paid particular attention to in classification process. They were classified with use of supervised RF classification method and with Sentinel-2 data from two time periods. The reason was that we wanted to capture as most updated vegetation conditions as possible, since agricultural landscape within the study area has been changing due to i.a. farmland abandonment, which is shown by high percentage of Wastelands class coverage in our LULC map.

Table 4. The surface coverage of specified land use classes from LULC layers derived from OBIA classification workflow, BDOT database and CLC 2012 database

	OBIA classification		BDOT		CLC12	
	Area [km^2]	Area [%]	Area [km^2]	Area [%]	Area [km^2]	Area [%]
Built-up	36.98	5.82	30.70	4.49	48.26	7.61
Bushes	3.20	0.50	3.52	0.52		
Croplands	168.08	26.47	296.85	43.45	303.49	47.86
Ditches	5.02	0.79	5.34	0.78		
Grasslands	43.92	6.91	86.70	12.69	29.85	4.71
Roads	15.04	2.37	19.25	2.82		
Railway	0.29	0.05	0.54	0.08		
Recreational	0.86	0.14	1.19	0.17	0.63	0.10
Trees	31.70	4.99	32.62	4.77		
Wastelands	120.01	18.90	1.52	0.22	42.57	6.71
Water	7.47	1.18	7.19	1.05	2.92	0.46
Woodlands	202.51	31.89	197.82	28.95	206.39	32.55
Overall	635.08	100.00	683.25	100.00	634.12	100.00

4. Discussion

Data and methods fusion approach in presented GEOBIA classification framework allowed us to accomplish overall accuracy of 81.3% and kappa coefficient of 0.78. The obtained results are satisfactory and comparable to those achieved by other authors who also used ancillary vector data in classification workflow (Castillejo-González et al., 2009; Tiede et al., 2010; Wężyk et al., 2016). Average Producer's accuracy was 0.8, however for only one class (Trees) it was below 0.5 and for six classes PA value was over 0.9. User's accuracy values were less spread and there was no value below 0.5 and for three classes UA exceeded 0.9. The most confusion comes from misidentification of Trees class often confused with Wasteland and Woodlands classes which can be explained by the fact that Trees and Woodlands classes were divided only according to area condition and by secondary forest succession process within Wastelands class. This process can also be related to confusion between Bushes and Wastelands classes.

Object-based classification approach has been proved to outperform pixel-based classification yielding higher accuracy results by many authors (Syed et al., 2005; Castillejo-González et al., 2009; Varga et al, 2014; Belgiu and Csillik, 2018). Additionally, fast development of GIS and RS technologies and continuously increasing availability of RS and vector datasets give the possibilities of using multisource datasets in image classification framework. Data integration approach which allows to increase the classification accuracy by incorporating more useful information, has been proved to be advantageous in LULC mapping by many authors (Nordkvist et al., 2012, Huang and Zhu, 2013; Szostak et al., 2014; Jia, 2015).

GEOBIA classification framework, besides of data fusion, allows also to integrate different classification methods. Some authors took advantage from this opportunity and developed combined object- and pixel- based image classification framework (Bernadini et al., 2010; Salehi et al., 2013) or integrated a rule-based expert system classifier and a neural network classifier (Liu et al., 2002). In our study besides of integrating various datasets also two different classification methods were incorporated: rule-based and Random Forest supervised classification algorithm. This approach allowed us to focus on classification of each LU class 'individually' by taking advantage from all useful information from various input data, expert knowledge, and available machine-learning methods.

Li et al. (2016) showed that Random Forest algorithm, is highly suitable for GEOBIA classification in agricultural areas. In this study, Random Forest supervised classification was perform in order to assign three agriculture-related classes with vegetation alternating conditions: Croplands, Grasslands and Wastelands. Obtained PA accuracy values from 0.72 to 0.94 and UA accuracy from 0.73 to 0.88 for these classes can be considered as satisfactory. Immitzer et al. (2016) performed object-based crop types classification with Sentinel-based Random Forest algorithm and obtained overall accuracy of 76.8% and PA varied from 0.281 to 0.963, UA from 0.624 to 0.881 for various crop types. Inglada et al. (2015) investigated the opportunities of supervised classification methods for crop type mapping at global scale. For twelve test sites all over the world they obtained OA values for RF classifier above 0.8 for seven sites, and only three of them were under 0.7.

Random Forest algorithm allows for the variables importance assessment. It is extremely useful for variable selection in the classification of complex areas where large multisource data sets with a large number of variables are used. In our study, the highest importance was shown by vegetation indices, followed by Red-Edge and shortwave infrared (SWIR) Sentinel-2 bands. Our results are similar to outcomes of Ramoelo et al., (2015) or Immitzer et al. (2016) who performed Sentinel-based classification of forest and agriculture sites and confirmed the high value of the red-edge and SWIR bands for vegetation mapping. Schuster et al. (2012) proved that incorporation of red-edge information can increase classification accuracy.

According to accuracy assessment results, the multi-stage semi-automatic GEOBIA classification workflow presented in this study can be considered as a suitable LULC mapping tool for the study area. However, there are some shortcomings of its application. Firstly, although the incorporation of multi-source input data improves the classification accuracy and allows for class-focused classification approach, data collection and pre-processing including i.a. unification of all databases and input layers can be very effort- and time-consuming. Secondly, the input data used in this workflow are multitemporal i.e. they were acquired or updated to different time points, which could impact on classification accuracy. The classification framework integrated rule-based and supervised methods. The first one involves creating expert knowledge-based set of rules which is labour and time-consuming (Liu et al., 2002; Stephenson, 2010) and unfortunately it cannot be easily transferable to other region. Taking into account its proven classification robustness (Li et al., 2016; Kulkarni and Lowe, 2016; Ma et al., 2017) Random Forest algorithm was chosen for supervised classification step but no comparison of different machine-learning methods like Support Vector Machines, Neural Networks or Decision Trees was presented in this study.

5. Conclusions

In face of fast developing GIS and RS technologies, rapidly increasing amount of remote sensing images and GIS vector data available from databases of various agencies as well as open source and the fact that modern image classification methods and techniques are still under investigation with very promising results, there is a vast array of LULC mapping opportunities. On the other hand, there is an increasing demand for improvement of existing LULC maps. To address this demand, different available datasets and tools can be used separately or taking advantage from data and methods fusion approach which is currently of increasing importance. The presented GEOBIA classification workflow can be a good example of such LULC mapping tool. We have shown that one multi-stage GEOBIA classification framework can integrate multi-source, multi-temporal and multi-resolution input data, rule-based and supervised RF classification methods in order to take advantage from all useful available information, as well as expert knowledge and advanced machine-learning tools for LU class-focused classification. Since the LULC maps are used for the protection of the habitats, agriculture and forest policy, hydrology modelling, sustainable management of the environment, monitoring of the influence of climate change, etc., the quality of land use data are extremely important Therefore, we recommend to use the variability of available spatial data and classification tools for LULC mapping, which can be integrated in GEOBIA framework, as we have presented in this paper.

Acknowledgments

This paper is based on own research conducted at the Faculty of Animal Science, University of Agriculture in Krakow.

References

Baatz, M. and Schape, A. (2000). Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation. *J. Photogramm. Remote Sens.*, 58, 12–23.

Begon, M., Townsend, C.R. and Harper, J.L. (2006). *Ecology: From individuals to ecosystems*. Malden, MA: Blackwell Pub, p. 499.

Belgiu, M. and Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sensing of Environment*, 204, 509–523; DOI: 10.1016/j.rse.2017.10.005.

Bernardini, A., Frontoni, E., Malinverni, E., Mancini, A., Tassetti, A. and Zingaretti, P. (2010). Pixel, object and hybrid classification comparisons. *Journal of Spatial Science*, 55, 43–54. DOI: 10.1080/ 14498596.2010.487641.

Blaschke, T. (2010). Object Based Image Analysis for Remote Sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65 (1), 2–16. DOI: 10.1016/j.isprsjprs.2009.06.004.

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F. and Tiede, D. (2014). Geographic Object-Based Image Analysis – Towards a New Paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 180–191. DOI: 10.1016/j.isprsjprs.2013.09.014.

Breiman, L., (2001). Random Forests. *Mach. Learn.*, 45, 5–32. DOI: 10.1023/A:1010933404324.

Castillejo-González, I.L., López-Granados, F., García-Ferrer, A., Peña-Barragán, J.M., Jurado-Expósito, M., de la Orden, M.S. and González-Audicana, M. (2009). Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. *Comput. Electron. Agric.*, 68, 207–215.

Clevers, J.G.P.W. and Gitelson, A.A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformation*, 23, 344–351.

Congalton, R.G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37, 35–46.

Congalton, R.G. and Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data Principles and Practices*, 2nd ed., Mapping Science. CRC Press, Taylor & Francis Group, Boca Raton. DOI: 10.1201/9781420048568.fmatt.

de Vienne, D.M. (2016). Lifemap: Exploring the Entire Tree of Life. *PLoS Biol.*, 14 (12), e2001624. https://doi.org/10.1371/journal.pbio.2001624.

Garcia, M.G. (2014). Nonequilibrium Statistical Physics in Ecology: Vegetation Patterns, Animal Mobility and Temporal Fluctuations. PhD Thesis. IFISC, Institute for Cross Disciplinary Physics and Complex Systems, Spain.

Guisan, A. and Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186.

Hay, G.J. and Castilla, G. (2006). Object-Based Image Analysis: Strengths, Weaknesses, Opportunities and Threats (SWOT). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36, 4.

Hofmann, P., Blaschke, T. and Strobl, J. (2011). Quantifying the robustness of fuzzy rule sets in object-based image analysis. *International Journal of Remote Sensing*, 32 (22), 7359–7381. http://dx.doi.org/10.1080/01431161.2010.523727.

Huang, R. and Zhu, J. (2013). Using Random Forest to integrate LiDAR data and hyperspectral imagery for landcover classification. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Melbourne, Australia, 21–26 July 2013; pp. 3978–3981.

Immitzer, M., Vuolo, F. and Atzberger, C. (2016). First Experience With Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.*, 8 (3), 166. DOI: 10.3390/rs8030166.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P. and Koetz, B. (2015). Assessment of an Operational System for Crop Type Map Production Using High Temporal and Spatial Resolution Satellite Optical Imagery. *Remote Sens.*, 7, 12356–12379; DOI: 10.3390/rs70912356.

Jia, Y. 2015. Object-based Land Cover Classification with Orthophoto and LIDAR Data. Master of Science Thesis, Royal Institute of Technology (KTH) Stockholm, Sweden.

Kaplan, G. and Avdan, U. (2017). Object-based water body extraction model using Sentinel-2 satellite imagery. *European Journal of Remote Sensing*, 50 (1), 137–143.

Kulkarni, A.D. and Lowe, B. (2016). Random Forest Algorithm for Land Cover Classification. *International Journal on Recent and Innovation Trends in Computing and Communication*, 4 (3), 58–63.

Li, M., Ma, L. and Blaschke, T. (2016). A systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. *Int. J. Appl. Earth Obs.*, 49, 87–98. DOI: 10.1016/j.jag.2016.01.01.

Liaw, A. and Wiener, M., (2002). Classification and Regression by random Forest. *R News*, 2, 18–22.

Liu, X.H., Skidmore, A.K. and Oosten, H.V. (2002). Integration of classification methods for improvement of land cover map accuracy. *ISPRS J. Photogramm. Remote Sens*., 56, 257–268.

Ma, L., Fu, T., Blaschke, T., Li, M., Tiede, D., Zhou, Z., Ma, X. and Chen, D. (2017). Evaluation Of Feature Selection Methods For Object-based Land Cover Mapping Of Unmanned Aerial Vehicle Imagery Using Random Forest And Support Vector Machine Classifiers. *ISPRS International Journal of Geo-information*, 6 (2), 51. DOI: 10.3390/ijgi6020051.

Ma, L., Li, M., Ma, X., Cheng, L., Du, P. and Liu, Y. (2017). A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130, 277–293. http://dx.doi.org/10.1016/j.isprsjprs.2017.06.001.

Nordkvist, K., Granholm, A.H., Holmgren, J., Olsson H. and Nilsson, M. (2012). Combining optical satellite data and airborne laser scanner data for vegetation classification. *Remote Sensing Letters*, 3 (5), 393–401. http://dx.doi.org/10.1080/01431161.2011.606240.

Oktavia, D. and Prativi, S.D. (2017). Integrated Geological and Ecological Studies of Heath Forest: ecological restoration at the former tin mining land. Conference: IUFRO 2017, At Freiburg, Germany.

Ramoelo, A., Cho, M., Mathieu, R. and Skidmore, A.K. (2015). Potential of Sentinel-2 spectral configuration to assess rangeland quality. *J. Appl. Remote Sens.*, 9, 094096. DOI: 10.1117/1.JRS.9.094096.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J.P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.*, 67, 93–104.

Salehi, B., Zhang, Y., and Zhong, M. (2013). A combined object-and pixel-based image analysis framework for urban land cover classification of VHR imagery. *Photogrammetric Engineering & Remote Sensing*, 79 (11), 999–1014. DOI: 10.14358/PERS.79.11.999.

Schuster, C., Förster, M. and Kleinschmit, B. (2012). Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *Int. J. Remote Sens.*, 33, 5583–5599. DOI: 10.1080/01431161.2012.666812.

Smith, G.M. and Morton, R.D. (2010). Real World Objects in GEOBIA through the Exploitation of Existing Digital Cartography and Image Segmentation. *Photogrammetric Engineering & Remote Sensing*, 76(2), 163–171. DOI: 10.14358/PERS.76.2.163.

Stephens, P.A., Mason, L.R., Green, R.E., Gregory, R.D., Sauer, J.R., Alison, J., Aunins, A., Brotons, L., Butchart, S.H.M., Campedelli, T., Chodkiewicz, T., Chylarecki, P., Crowe, O., Elts, J., Escandell, V., Foppen, R.P.B., Heldbjerg, H., Herrando, S., Husby, M., Jiguet, F., Lehikoinen, A., Lindström, Å., Noble, D.G., Paquet, J.-Y., Reif, J., Sattler, T., Szép, T., Teufelbauer, N., Trautmann, S., van Strien, A.J., van Turnhout, C.A.M., Vorisek P. and Willis, S.G. (2016). Consistent response of bird populations to climate change on two continents. *Science*, 352 (6281), 84–87.

Stephenson, G. (2010). A comparison of supervised and rule-based objectorientated classification for forest mapping. Postgraduate. Stellenbosch University.

Syed, S., Dare, P. and Jones, S. (2005). Automatic classification of land cover features with high resolution imagery and lidar data: an object-oriented approach. Proceedings of SSC2005 Spatial Intelligence, Innovation and Praxis: The national biennial Conference of the Spatial Sciences Institute, September, 2005. Melbourne: Spatial Sciences Institute. ISBN 0-9581366-2-9.

Szostak, M., Wężyk, P. and Tompalski, P. (2014). Aerial Orthophoto and Airborne Laser Scanning as Monitoring Tools for Land Cover Dynamics: A Case Study from the Milicz Forest District (Poland). *Pure and Applied Geophysics*, 171 (6), 857–866, DOI: 10.1007/s00024-013-0668-8.

Tiede, D., Lang, S., Albrecht, F. and Hölbling, D. (2010). Object-Based Class Modeling for Cadastre-Constrained Delineation of Geo-Objects. *Photogrammetric Engineering & Remote Sensing*, 76 (2), 193–202. DOI: 10.14358/PERS.76.2.193.

van Dam, A.A., Kipkemboi, J., Rahman, M.M. and Gettel, G.M. (2013). Linking Hydrology, Ecosystem Function, and Livelihood Outcomes in African Papyrus Wetlands Using a Bayesian Network Model. *Wetlands*, 33, 381–397.

Varga, O.G., Szabó, S. and Túri, Z. (2014). Efficiency Assessments of GEOBIA in Land Cover Analysis, NE Hungary. *Bulletin of Environmental and Scientific Research*, 3 (4), 1–9.

Wężyk, P., Hawryło, P., Szostak, M., Pierzchalski, P. and de Kok, R. (2016). Using GEOBIA and data fusion approach for land use and land cover mapping. *Quaestiones Geographicae*, 35 (1), pp. 93–104.

Witharana, C. and Civco, D.L. (2014). Optimizing multi-resolution segmentation scale using empirical methods: exploring the sensitivity of the supervised discrepancy measure Euclidean distance 2 (ED2). *ISPRS J. Photogramm. Remote Sens.*, 87, 108–121.

Zanariah, Z., Apan, W.N., Le Brocque, A.F. and Maraseni, T.N. (2012). Fine-Scale Habitat Modelling of Wildlife Species Using Spatial Information Tools. In: 2nd Malaysian Postgraduate Conference (MPC 2012), 7–9 July 2012, Gold Coast, Australia.