Acta geographica Slovenica, 62-3, 2022, 63–83

INTEGRATED METHOD FOR GLOBAL LAND COVER PRODUCTS' VALIDATION ON THE EXAMPLE OF BULGARIA

Vanya Stamenova, Stefan Stamenov


STEFAN STAMENOV

Pliska field at the National historical and archaeological reserve, Northeast Bulgaria.

DOI: https://doi.org/10.3986/AGS.10929
UDC: 711.14(497.2)"2009/2010"
659.2:004:91
COBISS: 1.01

Vanya Stamenova[1], Stefan Stamenov[1]

Integrated method for global land cover products' validation on the example of Bulgaria

ABSTRACT: The article presents the validation of four global land cover datasets for 2009–2010 (GlobeLand30, GlobCover, Forest/Non-forest map, Tree canopy cover) for the territory of Bulgaria as a part of a task led by the Group on Earth Observation (GEO). An integrated method for validation, combining the GIS processing capabilities, various data and the expert knowledge-based method of visual interpretation was developed. It aims to be an effective, time-saving approach, allowing validation of several datasets in one environment. The results are presented in vector format with structured attribute data for all validated datasets. The analysis reveals that the highest accuracy is observed for Tree canopy cover (96.1%), the lowest for GlobCover 2009 (51.0%), and the other products' accuracy is considerably high, about 80.0%.

KEY WORDS: land cover validation, global land cover products, integrated method, GIS processing, visual interpretation, Bulgaria

Integrirana metoda za validacijo globalnih produktov pokrovnosti tal na primeru Bolgarije

POVZETEK: V članku je predstavljena validacija štirih globalnih podatkovnih nizov pokrovnosti tal za obdobje 2009–2010 (GlobeLand30, GlobCover, Forest/Non-forest map, Tree canopy cover) za ozemlje Bolgarije v okviru naloge, ki jo je vodila organizacija GEO. Razvita je bila integrirana metoda za validiranje, ki združuje zmogljivosti obdelave GIS, različne podatke in metodo vizualne interpretacije, temelječe na ekspertnem znanju. Gre za učinkovit in časovno varčen pristop, ki omogoča potrjevanje več podatkovnih nizov v enem okolju. Rezultati so predstavljeni v vektorski obliki s strukturiranimi atributnimi podatki za vse potrjene podatkovne nize. Analiza je pokazala, da je največja natančnost ugotovljena za Tree canopy cover (96,7 %), najmanjša pa za GlobCover 2009 (51,0 %). Natančnost ostalih podatkovnih slojev je precej visoka, približno 80 %.

KLJUČNE BESEDE: validacija pokrovnosti tal, globalni produkti pokrovnosti tal, integrirana metoda, GIS obdelava, vizualna interpretacija, Bolgarija

The article was submitted for publication on June 30th, 2022.
Uredništvo je prejelo prispevek 30. junija 2022.

[1] Bulgarian Academy of Sciences, Space Research and Technology Institute, Sofia, Bulgaria
vanya_stamenova@yahoo.com (https://orcid.org/0000-0002-8933-5426), stamenovstefan@yahoo.bg (https://orcid.org/0000-0001-6010-8142)

1 Introduction

Land cover is defined as the observed (bio-)physical cover of the Earth's surface (Di Gregorio and Jansen 1998). It includes various types of vegetation cover, man-made objects, bare land and inland water surfaces. Global, continental and regional land cover products serve as essential source of information for various applications, services, and global natural and anthropogenic processes and phenomenon as global climate changes, global land cover changes, deforestation, soil degradation, anthropogenization; policies and policy making at global, regional and national level as sustainable development, climate change monitoring, security, food and water security, land management, conservation, etc. (European Commission and Eurostat 1999; Grekousis, Mountrakis and Kavouras 2015; Tsendbazar, de Bruin and Herold 2015; Herold et al. 2016; Chen et al. 2016a; 2017).

During the past decades, several global land cover products with spatial resolution varying from 1 km to 25 m were produced using different remote sensing data – 1 km and 500 m global land cover (GLC) products from AVHRR and MODIS data (Hansen et al. 2000; Loveland et al. 2000; Grekousis, Mountrakis and Kavouras 2015), European Space Agency (ESA) global land cover product GlobCover with 300 m resolution (Bontemps et al. 2011) and products with finer resolutions such as GlobeLand30 product of National Geomatics Center of China (NGCC) with 30 m (Chen, Yifang and Songnian 2014; Ran and Li 2015), the datasets developed within the Global Forest Watch and Global Forest Monitoring projects – 30 m (Hansen et al. 2013), and a Forest/Non-forest dataset with 25 m resolution of Japan Aerospace Exploration Agency (JAXA) (Shimada et al. 2014).

The validation process is a suite of techniques, which are used for determining the quality of a certain product (Strahler et al. 2006). According to the definition of the Committee on Earth Observing Satellites Working Group on Calibration and Validation the term validation is defined as »The process of assessing, by independent means, the quality of the data products derived from the system outputs« (Achard et al. 2011, 13). The overall structure of the validation process typically includes several stages.

The statistically based or quantitative method of validation requires selection of the sampling design and allocation, selection of reference data and visual interpretation, and accuracy assessment and interpretation of the results from the validation of the land cover products (Strahler et al. 2006). The sampling method refers to the rules used for allocation of sample units (points or polygons) within the area which will be validated. The stratification, commonly part of the sampling design, allows to specify the samples allocated to each stratum, where the strata are the land cover classes (Olofsson et al. 2012). Many different sampling methods exist, such as simple random sampling, stratified random sampling, cluster sampling, etc. and there is no universal sampling design suitable for all global validations and assessments (Strahler et al. 2006). To improve the spatial sampling design and allocation, a landscape shape index-based approach was developed and implemented in the online validation tool for GlobeLand30 product: GeoVal – GLC Validation Platform (Chen et al. 2016b). It takes into consideration the spatial heterogeneity of the land cover in estimation of the sample size and their spatial distribution.

The sample units for a validation of a certain territory are visually interpreted using interpretation keys and various reference data and sources (Mayaux et al. 2006; Achard et al. 2011; Olofsson et al. 2012; Congedo and Munafò 2012). The prior information and knowledge for the geographical region which is validated is also important. The knowledge-based verification uses information about the natural environment and human activities which affect the distribution of the land cover and defines rules, taking into consideration three types of knowledge – natural, cultural and temporal (Chen et al. 2015; Zhang et al. 2016).

Accuracy assessment refers to the process which determines the quality of a product or map, created from remote sensing data after the validation is made (Congedo and Munafò 2012). There are many statistical methods and parameters applied for assessment of the classification accuracy and its representation, including, error matrix, overall accuracy, user's and producer's accuracy, Kappa statistics, spatial accuracy, etc. (Mayaux et al. 2006; Strahler et al. 2006; Achard et al. 2011; Congedo and Munafò 2012; Brovelli et al. 2015; Tsendbazar et al. 2015).

The current study aims to presents an integrated method for validation of global land cover products, developed and applied to the validation of four global land cover datasets for the territory of Bulgaria. It is focused on the application of GIS processing and visual interpretation, and proposing a validation and decision-making approach which allows validation of several datasets in one environment and generating as a final result a single file with the validation information for all GLC products.

The GLC validation for the territory of Bulgaria has been carried out within the international task Data Validation of Global Land Cover Datasets with 30 m resolution, which is part of the Group on Earth Observation (GEO) component SB-02-C2 Global Land Cover Validation and User Engagement (https://www.earthobservations.org). The global land cover datasets which have been validated within the GEO-led international task were GlobeLand30 dataset in 30 m resolution from NGCC, GlobCover dataset in 300 m resolution from ESA, Europe, Forest/Non-Forest Map dataset in 25 m resolution from Japan, and Global Forest Watch dataset in 30 m resolution from Maryland University, USA. The global land cover products selected for validation are for one and the same time period – 2009/2010, providing valuable and credible information for the land cover of a given moment in time and being very useful for studying processes in time.

2 Methods and Data

2.1 Study area

The validation was conducted for the territory of Bulgaria. Bulgaria is a Southeastern European country, located at the Balkan Peninsula and covering an area of 110,994 km² (Figure 1). It is characterized by significant diversity, heterogeneous landscapes and a variety of landforms. It lies at the southern parts of the temperate climatic zone at the transition to the Mediterranean climate in the southern part which determines the transitivity of characteristics of the natural components (Penin 2000). The SRTM digital terrain model for Bulgaria (Figure 1) gives an overall impression for the diversity of landforms and the segmentation of the relief. The plains and lowlands, situated in the northern part of the country are part of Danubian plain, where the steppe character of the vegetation can be observed. Significant part of the territory of Bulgaria is occupied by mountains: the Stara planina mountain chain and Sredna Gora Mountains, crossing the country in the middle from west to east, the Rila–Rhodope Mountain areas in the southwest and south part of Bulgaria. Bulgaria is also one of the richest in biodiversity European countries and 34.9% of its territory falls into the European ecological network Natura 2000 according to information published at Ministry of Environment and Water of the Republic of Bulgaria website (https://www.moew.government.bg/en/nature/).

Due to the specifics of the natural component the land cover of Bulgaria is strongly heterogeneous. In Bulgaria, as well as in whole Europe the land cover is rather segmented and the main land cover types which could be observed are typical for European territory – urban areas, agricultural areas, forests (coniferous, deciduous and mixed), which covers 38% of the territory of Bulgaria, bare lands, water areas and wetlands, etc. (Figure 1).

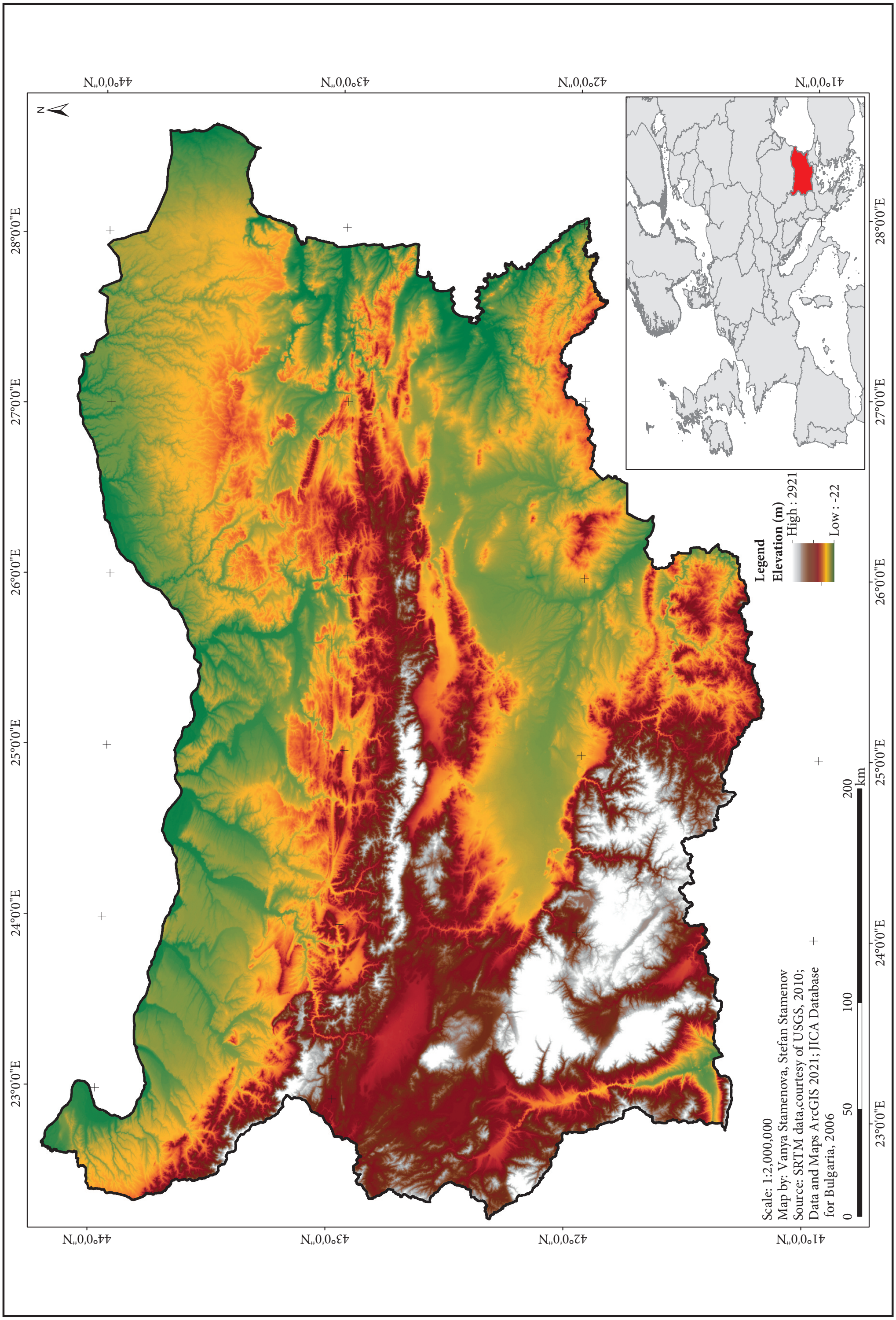
2.2 Global land cover datasets

Four GLC datasets for the period 2009–2010 have been validated for the territory of Bulgaria: GlobeLand30 dataset for 2010 (30 m), GlobCover 2009 dataset (300 m), Forest/Non-Forest Map for 2010 (25 m) and tree canopy cover dataset for 2010 (30 m) (Table 1).

Table 1: Used global land cover datasets.

Name of the dataset	Source	Year of production	Website
GlobeLand30	NGCC	2010	http://www.globallandcover.com
GlobCover	ESA, Europe	2009	http://due.esrin.esa.int/page_projects.php
Forest/Non-Forest Map	JAXA, Japan	2010	http://www.eorc.jaxa.jp/ALOS
Tree canopy cover	Maryland University, USA	2010	http://glad.umd.edu/dataset

Figure 1: Location and general view of Bulgaria. ➤ p. 67


N
44°0'0"N
43°0'0"N
42°0'0"N
41°0'0"N
28°0'0"E
27°0'0"E
26°0'0"E
25°0'0"E
24°0'0"E
23°0'0"E
Legend
Elevation (m)
High : 2921
Low : -22
Scale: 1:2,000,000
Map by: Vanya Stamenova, Stefan Stamenov
Source: SRTM data,courtesy of USGS, 2010;
Data and Maps ArcGIS 2021; JICA Database
for Bulgaria, 2006
0
50
100
200
km

GlobeLand30 product (30 m) is a NGCC's product derived by classification of multispectral Landsat TM and ETM+ images and multispectral images of the Chinese Environmental Disaster Alleviation Satellite (HJ-1) without clouds for 2010, ranging plus/minus 1 year (National Geomatics … 2014). It consists of 10 land cover classes: 10 – cultivated land, 20 – forest, 30 – grassland, 40 – shrubland, 50 – wetland, 60 – water bodies, 70 – tundra, 80 – artificial surfaces, 90 – bare land and 100 – permanent snow and ice.

The GlobCover project is an initiative of ESA, started in 2005 in partnership with JRC, EEA, FAO, UNEP, GOFC-GOLD and IGBP, aiming to develop a service capable of delivering global composites and land cover maps using as input observations from the 300 m resolution MERIS sensor on board the ENVISAT satellite mission (Bontemps et al. 2011). The GlobCover 2009 land cover product is derived from a global MERIS FR mosaic for the year 2009. This is the second product generated within the GlobCover project, after its first release in 2005, which has spatial resolution of 300 m and its map projection is a Plate-Carrée (WGS84 ellipsoid). The land cover classification scheme of the GlobeCover2009 includes 22 land cover classes defined with the United Nations (UN) Land Cover Classification System (LCCS) (Di Gregorio and Jansen 1998; Bontemps et al. 2011).

The global forest/non-forest map is a free dataset generated by JAXA through classification of the backscattering intensity values of the global PALSAR-2/PALSAR mosaic dataset with 25 m resolution. The strong backscattering values in HV-polarization of the global PALSAR-2/PALSAR mosaic dataset are classified as »forest« (shown in green color), while the low backscatter as »non-forest« (colored in yellow) (Japan Aerospace … 2016). In the forest/non-forest dataset, the class »forest« is defined as the natural forest with the area larger than 0.5 ha and forest cover over 90%, corresponding to the Food and agriculture organizations of the United Nation's definition (FAO-FRA 2000; Food … 2012). The dataset includes three types of classes – forest, non-forest and water.

The tree canopy cover dataset for 2010 is a global product of tree canopy cover including all vegetation taller than 5 m in height, represented as a percentage per grid cell 30 × 30 meter in range from 0 to 100. Percent tree cover is defined as the density of tree canopy coverage of the land surface, and it estimates the percent maximum of the tree canopy for 2010 (the peak of the growing season) from cloud-free annual growing season composite Landsat 7 ETM+ data. A regression tree model estimating per pixel percent tree canopy cover was applied to annual composites from 2000 to 2012 inclusive (Hansen et al. 2013). The datasets were generated by GLAD (Global Land Analysis & Discovery) lab at the University of Maryland, US Geological Survey (USGS), NASA, and other partners of the Global Forest Watch Initiative and the data for Bulgaria were downloaded from the website of USGS (https://landcover.usgs.gov/glc/).

2.3 Reference layers and datasets

Different reference data have been used for the validation of the global land cover datasets – high resolution satellite images, aerial photos, and additional information about land cover of the territory of Bulgaria (Table 2). The reference layers and datasets must be synchronous in time with the data being validated.

The main source of information used for the validation are color orthophoto maps of Bulgaria, with spatial resolution 0.5 m and 0.4 m, acquired in 2006 and 2010–2011, respectively, provided by the Ministry of agriculture, food and forests of Republic of Bulgaria for scientific use. High resolution satellite imagery basemaps layers available in ArcGIS online platform and the Corine Land Cover data for 2012 have been used additionally for reference source of information.

Table 2: Used reference datasets.

Name of the dataset	Source	Year of production	Website
Orthophoto maps for Bulgaria	Ministry of agriculture, food and forests, Bulgaria	2010–2011	
Corine Land Cover	Copernicus Portal	2012	http://land.copernicus.eu
High resolution satellite imagery basemap layers	ArcGIS Online	–	https://www.arcgis.com/

2.4 Validation process

For the purpose of GLC products validation an integrated method was developed and applied, which gives the possibility for validation of several datasets in one environment and producing one file with the results, not several. The method was aimed to be as simple as possible, effective and time saving. The workflow process of the proposed integrated method consists of four main steps named – Preparation, Online Validation, Processing in GIS Environment and Validation, and Accuracy Assessment and Representation (Figure 2).

The first stage is the preparation, where the four GLC datasets which cover the study area have been downloaded and extracted by mask and the reference data were chosen and collected. The sampling procedure was chosen to fulfill two main requirements – to be suitable for areas with high landscape heterogeneity and the points to be evenly distributed over the whole territory of Bulgaria.

Initially the validation of GlobeLand30 product has been performed in its own online validation platform GeoVal where the sample points for the territory of Bulgaria has been generated. The sampling method used for the generation of the sample points is Landscape Shape Index (LSI), selection method of random

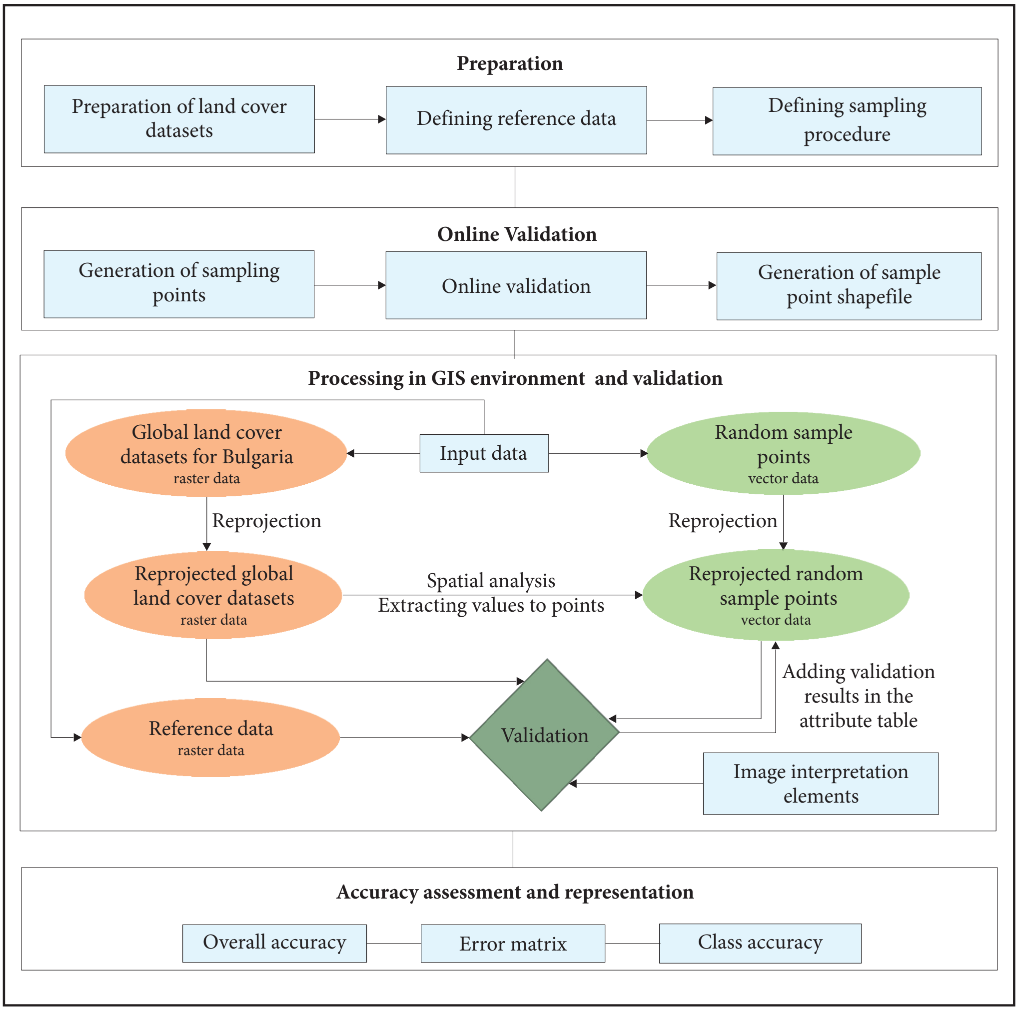


Figure 2: Overall structure of the validation process.

sampling at 95% confidence level, and plausibility judgment as sample judgment method. The LSI method is suitable for territories with spatial heterogeneity of the land cover, since it takes into consideration this heterogeneity and provides for better spatial sample allocation and distribution (Chen et al. 2021). The validation has been made using 382 randomly generated sample points for the whole territory of Bulgaria (Figure 3). The generation of the sample points has been made based on the Globeland30 land cover product and the same points have been used for the validation of the other three land cover dataset which provides for the comparability of results. The sample points were exported from the online validation platform and the validation has been continued in ArcGIS environment.

The next step of the GLC validation is the data processing in GIS environment and performing the validation using visual interpretation method. GIS allows all data to be stored, manipulated, processed and analyzed. All raster and vector data (GLC datasets, validation sample points and reference data) are reprojected in unified coordinate system, which in this case is UTM WGS 84, zone 35N. To the attribute table of the sample point shapefile new columns are added to store the validation results. The information about the land cover type values for each of the datasets have been extracted using the Spatial Analyst tools in ArcGIS – Extract Multi Values to Points and stored in the sample points shapefile attribute table lctype column. The process of validation is conducted after extraction through visual interpretation of the reference images and assigning the correct land cover values for each sample point. In the process of interpretation direct (shape, size, tone etc.) and indirect (association and location) interpretation elements are used (Akovetsky 1983; Lillesand and Kiefer 1999) and the results are stored in the shapefile table. The final product of the validation process is a shapefile with attributive table containing all the information for the land cover classes classified and ground truth, their index numbers and names. The designed structure of the sample point shapefile attribute table contains columns storing information for the land cover types for each land cover dataset. For each dataset three attribute columns have been generated with exception of Tree canopy cover dataset where the data are presented as a percentage value and are stored in two columns:

- column storing the index values for the land cover types,
- column with the names of the land cover types,
- column with the land cover codes assigned during the validation.

3 Results

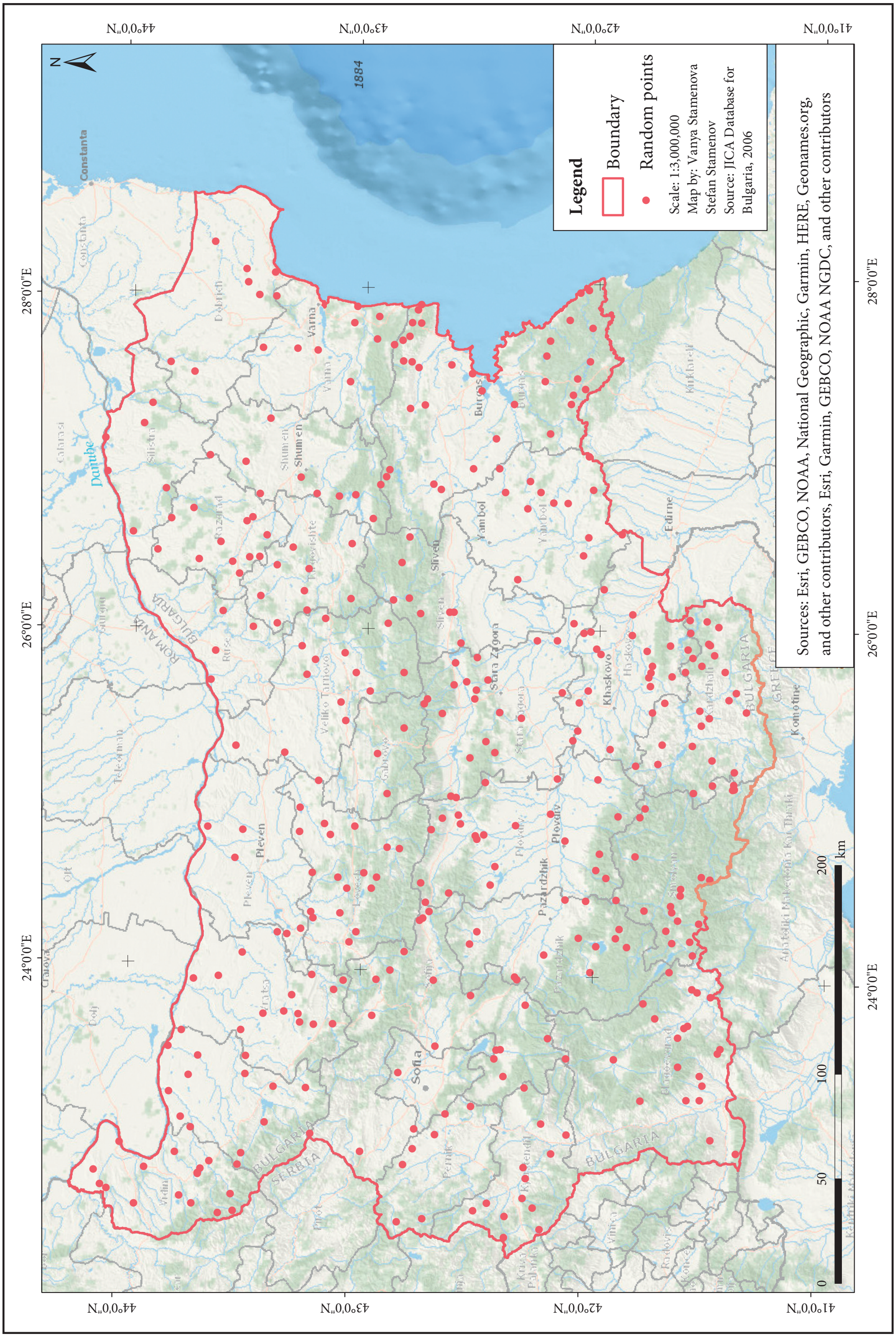
In this section the results obtained from the GLC products' validation and the estimated accuracies using the proposed integrated method are presented. For validating the products, a geodatabase for the territory of Bulgaria has been designed and built (Figure 4). The geodatabase has file structure and is built using ArcGIS 10.5 software. It contains raster global land cover datasets, reference datasets, vector layers and tables.

The final product generated as a result of the performed validation process is the creation of a vector layer, containing the whole information from the validation for each global product. Using the attribute data for each of the validated dataset error matrix tables have been generated.

3.1 Validation of GlobeLand30 dataset

The land cover map of Bulgaria, based on the GlobeLand30 dataset for 2010 (Figure 5), shows that the most prevalent land cover class is cultivated land, followed by the class forest. The GlobeLand30 dataset was validated first in the online validation platform GeoVal, where google maps were integrated, and then it was checked again together with the validation of the other three land cover datasets in GIS. The results of the validation are presented as error matrix in Table 3. The results of validation of the Globeland30 showed an overall accuracy of 79.8%. Table 3 provides information about the producer and user accuracy for each of the land cover classes. Highest producer accuracy (over 80.0%) is observed for the cultivated land, forest and water bodies and artificial areas classes, while shrubland and wetland land cover classes have lowest results. The land cover class with highest user accuracy is forest (92.2%), followed by the artificial areas, while lowest user accuracy is estimated for the wetland land cover class (Chen et al. 2021).

Figure 3: Distribution of the randomly selected sample points for the territory of Bulgaria. ➤ p. 71


Legend
Boundary
Random points
Scale: 1:3,000,000
Map by: Vanya Stamenova
Stefan Stamenov
Source: JICA Database for
Bulgaria, 2006
Sources: Esri, GEBCO, NOAA, National Geographic, Garmin, HERE, Geonames.org,
and other contributors, Esri, Garmin, GEBCO, NOAA NGDC, and other contributors
0 50 100 200 km
44°0'0"N
43°0'0"N
42°0'0"N
41°0'0"N
24°0'0"E
26°0'0"E
28°0'0"E
Sofia
Plovdiv
Varna
Burgas
Stara Zagora
Pleven
Shumen
Sliven
Yambol
Haskovo
Pazardzhik
Danube
Constanta
BULGARIA
ROMANIA
SERBIA
1884

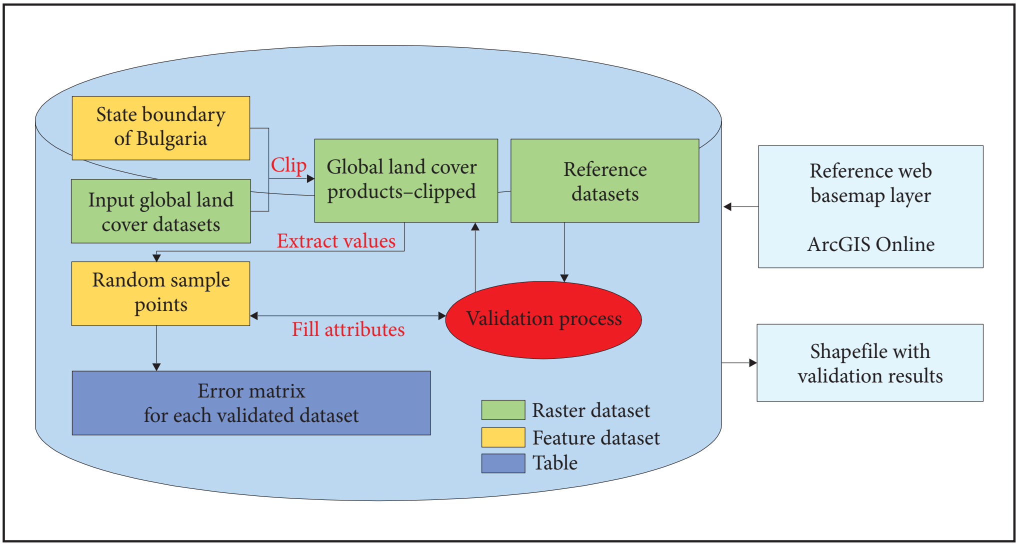


Figure 4: Geodatabase for GLC validation.

3.2 Validation of the GlobCover2009 product

The GlobCover2009 land cover dataset has 22 land cover classes, 17 are presented on the territory of Bulgaria (Figure 6). The result from validation is presented in Table 4, and the estimated overall accuracy is 51.04%. Many of the sample points fall within areas with mixed land cover, so during the interpretation the prevailing land cover type for each 300 m pixel was taken into consideration and the most appropriate land cover type was assigned.

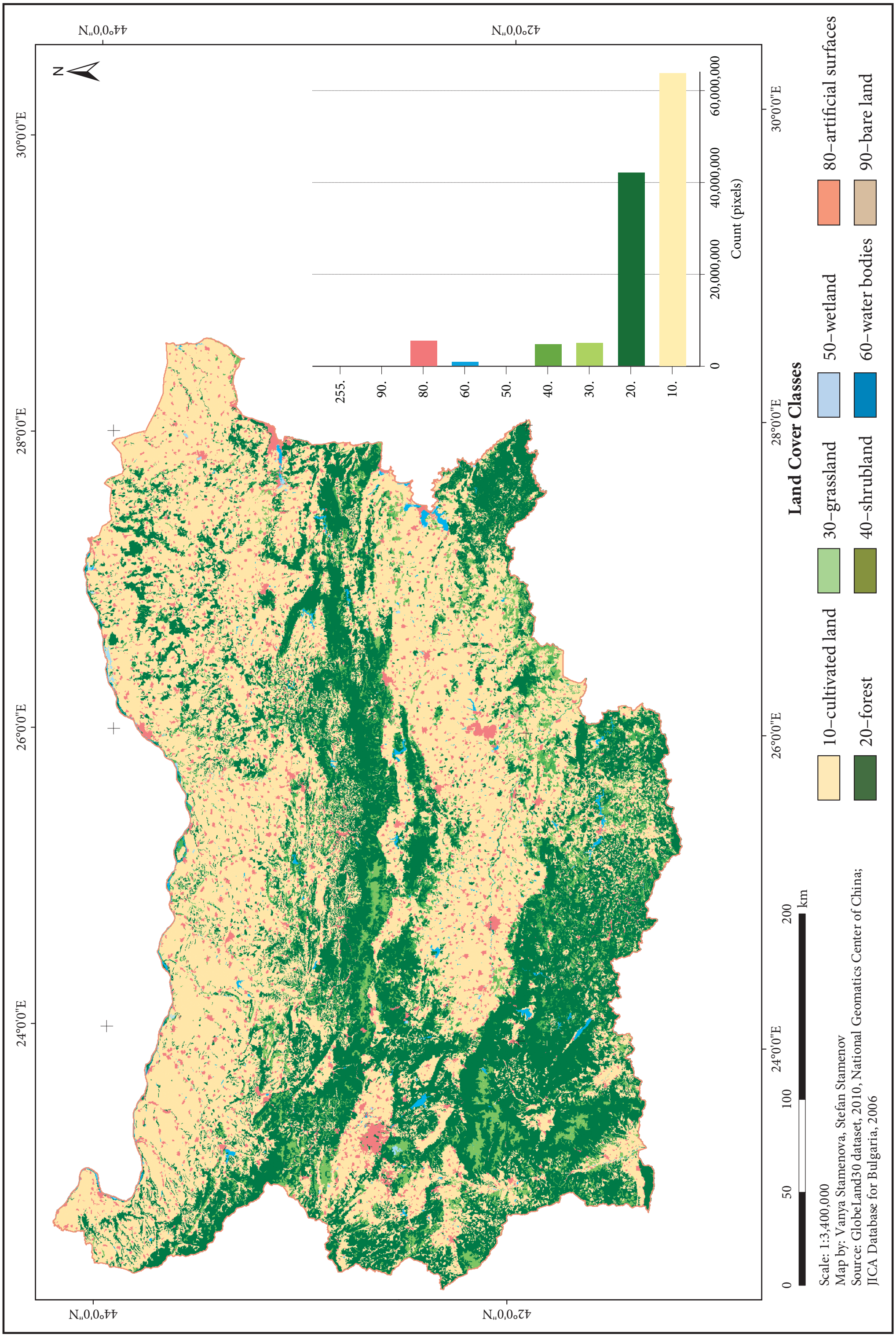
The land cover classes with user accuracy over 80% are unitary classes – class 70 and class 190, while the lowest user accuracy (0%) is received for the compound vegetation land cover classes – class 110, class 120, class 140 and the class 180 (Figure 6, Table 4). The producer accuracy shows similar results. Highest values are observed for three simple classes, which include only one type of land cover – needleleaved evergreen forest, shrubland and sparse (<15%) vegetation (e.g. land cover classes 70, 130, 150). Worst producer accuracy results (0%) have four classes, the same compound land cover classes with low user accuracy – classes 110, 120, 140 and 180. The lower overall accuracy obtained for the GlobCover 2009 dataset is due to the large number of mixed pixels.

3.3 Validation of Forest/Non-Forest dataset

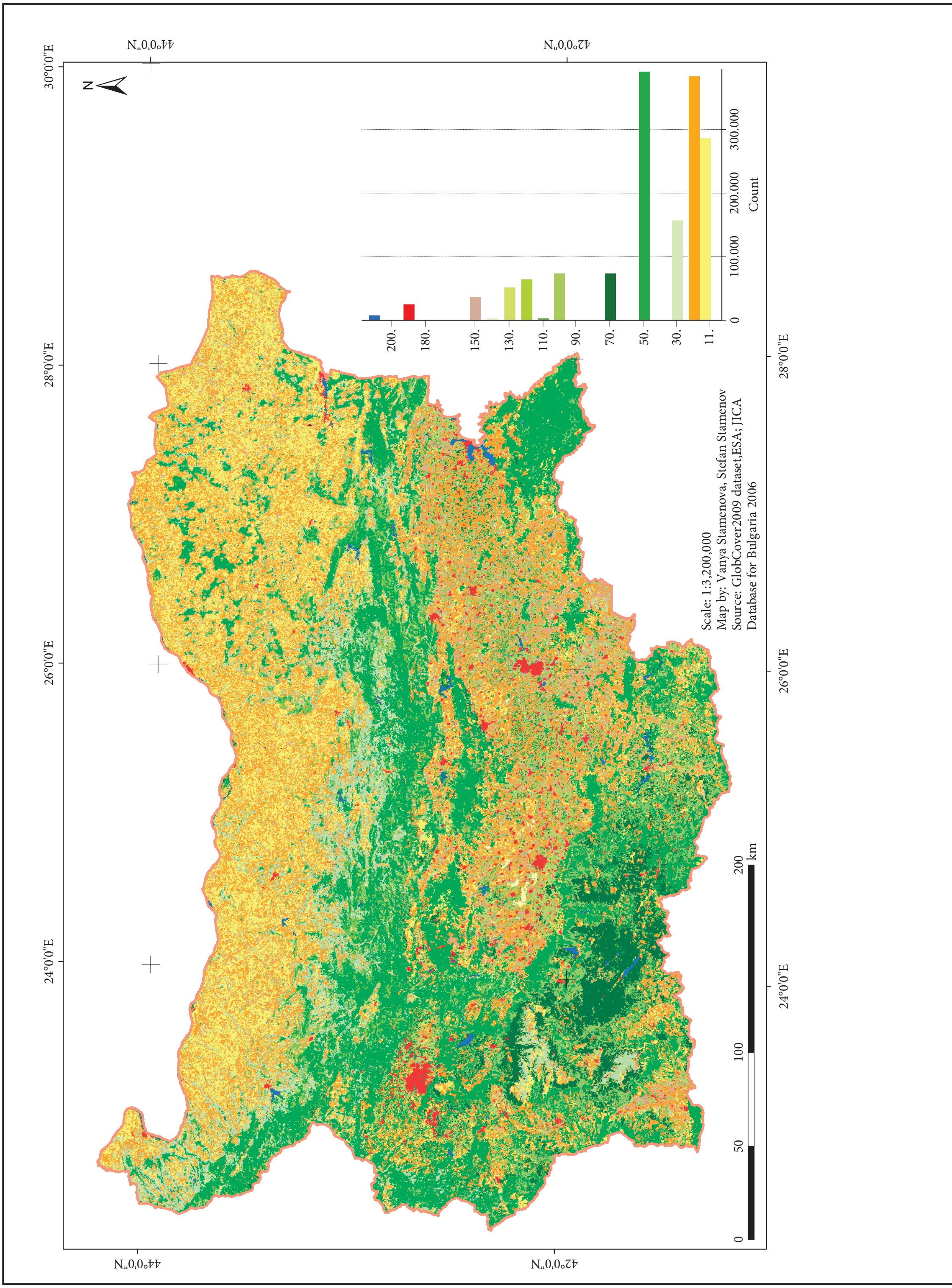
The global forest/non-forest product of the JAXA has smallest pixel size of 25 m and contains three land cover classes – forest, non-forest and water areas. The map of the forest and non-forest areas for the territory of Bulgaria, based on the JAXA's global product for 2010 is shown on Figure 7. The results of the validation are presented as the error matrix in Table 5.

The estimated overall accuracy for this product is 80.1%. The calculated user accuracy is higher than the producer accuracy for all the classes. The lowest value of per class accuracy is observed for the producer accuracy of water class.

Figure 5: Land cover map of Bulgaria based on the GlobeLand30 product. ➤ p. 73
Figure 6: Land cover map of Bulgaria based on the GlobCover 2009 product. ➤ p. 74–75
Figure 7: Forest / Non-forest map of Bulgaria for 2010. ➤ p. 77



Land Cover Classes
10–cultivated land
20–forest
30–grassland
40–shrubland
50–wetland
60–water bodies
80–artificial surfaces
90–bare land
Count (pixels)
0
20,000,000
40,000,000
60,000,000
10.
20.
30.
40.
50.
60.
80.
90.
255.
24°0'0"E
26°0'0"E
28°0'0"E
30°0'0"E
42°0'0"N
44°0'0"N
N
0
50
100
200
km
Scale: 1:3,400,000
Map by: Vanya Stamenova, Stefan Stamenov
Source: GlobeLand30 dataset, 2010, National Geomatics Center of China;
JICA Database for Bulgaria, 2006



44°0'0"N
42°0'0"N
30°0'0"E
28°0'0"E
26°0'0"E
24°0'0"E
N
Scale: 1:3,200,000
Map by: Vanya Stamenova, Stefan Stamenov
Source: GlobCover2009 dataset,ESA; JICA
Database for Bulgaria 2006
0
50
100
200
km
200.
180.
150.
130.
110.
90.
70.
50.
30.
11.
0
100.000
200.000
300.000
Count

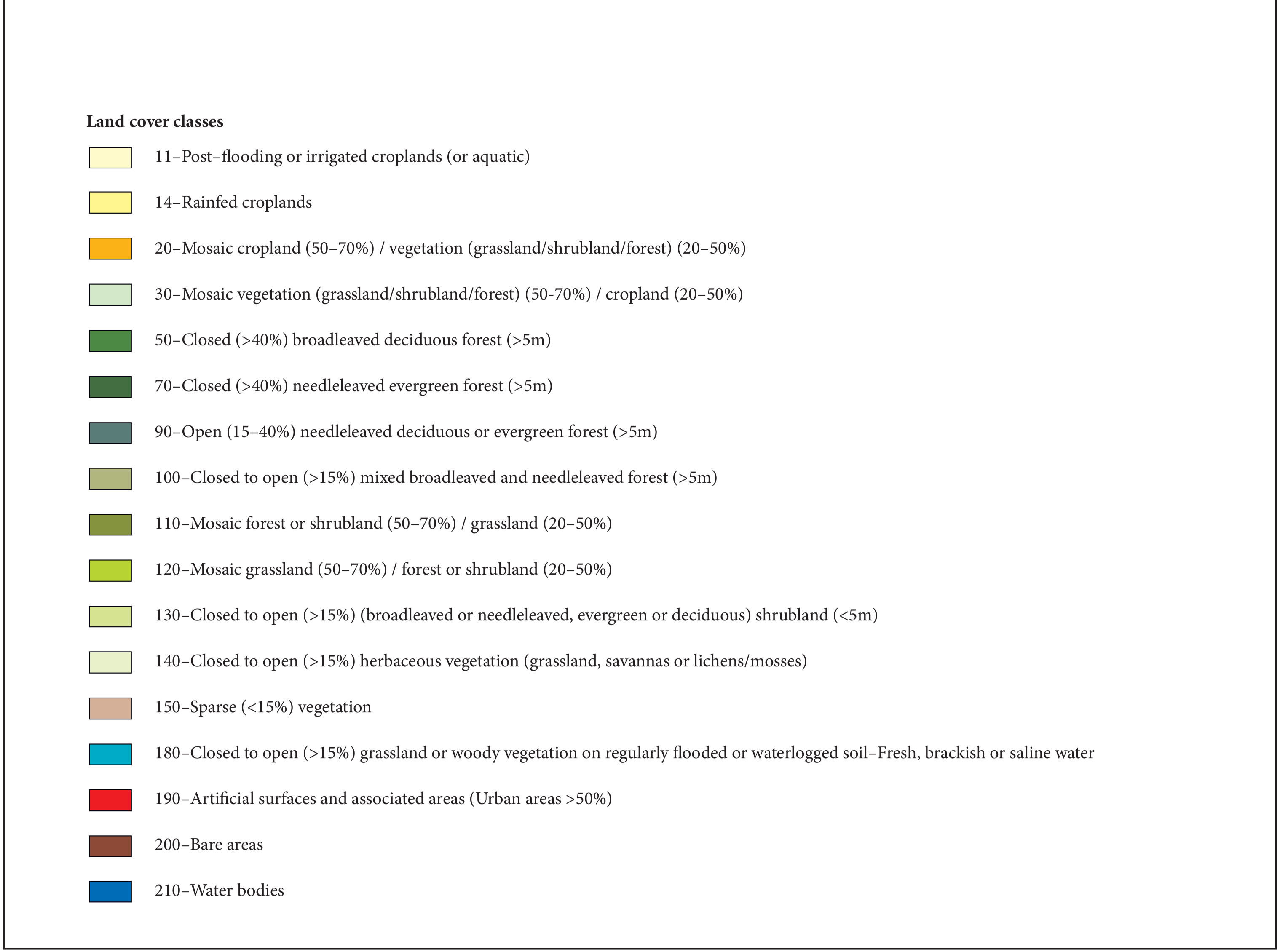
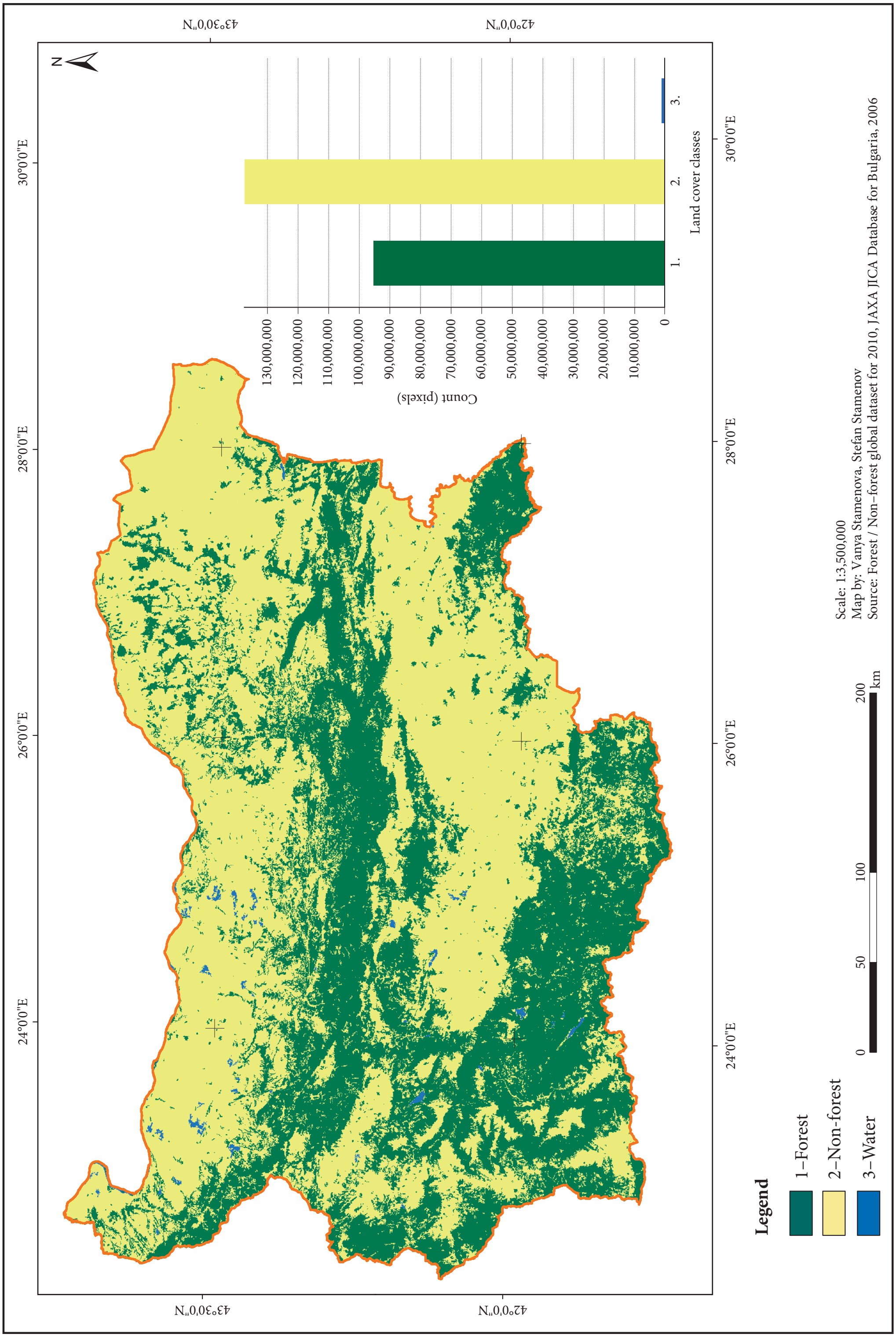

Land cover classes
11–Post–flooding or irrigated croplands (or aquatic)
14–Rainfed croplands
20–Mosaic cropland (50–70%) / vegetation (grassland/shrubland/forest) (20–50%)
30–Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20–50%)
50–Closed (>40%) broadleaved deciduous forest (>5m)
70–Closed (>40%) needleleaved evergreen forest (>5m)
90–Open (15–40%) needleleaved deciduous or evergreen forest (>5m)
100–Closed to open (>15%) mixed broadleaved and needleleaved forest (>5m)
110–Mosaic forest or shrubland (50–70%) / grassland (20–50%)
120–Mosaic grassland (50–70%) / forest or shrubland (20–50%)
130–Closed to open (>15%) (broadleaved or needleleaved, evergreen or deciduous) shrubland (<5m)
140–Closed to open (>15%) herbaceous vegetation (grassland, savannas or lichens/mosses)
150–Sparse (<15%) vegetation
180–Closed to open (>15%) grassland or woody vegetation on regularly flooded or waterlogged soil–Fresh, brackish or saline water
190–Artificial surfaces and associated areas (Urban areas >50%)
200–Bare areas
210–Water bodies

Table 3: Error matrix for GlobeLand30 dataset (Chen et al. 2021).

Land cover (number of points)	10	20	30	40	50	60	70	80	90	100	All	User accuracy (%)
10	75	10	7	10	0	0	0	4	0	0	106	70.8
20	5	154	4	4	0	0	0	0	0	0	167	92.2
30	2	9	26	7	0	1	0	1	0	0	46	56.5
40	1	3	2	23	0	0	0	0	0	0	29	79.3
50	0	0	0	1	1	0	0	0	0	0	2	50.0
60	0	0	1	0	1	5	0	0	0	0	7	71.4
70	0	0	0	0	0	0	0	0	0	0	0	0.0
80	1	0	1	1	0	0	0	21	0	0	24	87.5
90	0	1	0	0	0	0	0	0	0	0	1	0.0
100	0	0	0	0	0	0	0	0	0	0	0	0.0
All	84	177	41	46	2	6	0	26	0	0	382	0.0
Producer accuracy (%)	89.3	87.0	63.4	50.0	50.0	83.3	0.0	80.8	0.0	0.0	0.0	79.8

Table 4: Error matrix for GlobCover2009 dataset.

Land cover (number of points)	14	20	30	50	70	100	110	120	130	140	150	180	190	210	Total	User Accuracy (%)
14	23	4	5	7	1	0	6	2	0	0	0	0	5	0	53	43.4
20	12	24	4	6	1	0	11	4	0	0	0	0	6	0	68	35.3
30	4	7	16	8	0	0	9	1	0	0	0	0	4	0	49	32.7
50	2	2	9	85	0	4	10	8	1	1	0	0	1	0	123	69.1
70	0	0	1	0	15	0	0	0	0	0	0	1	0	1	18	83.3
100	0	1	0	3	1	15	5	0	0	0	0	0	0	0	25	60.0
110	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0.0
120	0	3	1	8	0	0	5	0	0	0	0	0	0	0	17	0.0
130	0	0	1	1	0	0	1	1	5	0	0	0	0	1	10	50.0
140	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2	0.0
150	1	1	0	0	0	0	0	0	0	0	1	0	0	0	3	33.3
180	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.0
190	0	1	1	0	0	0	0	0	0	0	0	0	8	0	10	80.0
210	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	100.0
Total	42	44	38	120	18	19	47	16	6	1	1	1	24	5	382	0.0
Producer Accuracy (%)	54.8	54.5	42.1	70.8	83.3	78.9	0.0	0.0	83.3	0.0	100.0	0.0	33.3	60.0	0.0	51.0



Legend
1–Forest
2–Non-forest
3–Water
Count (pixels)
Land cover classes
1.
2.
3.
0
10,000,000
20,000,000
30,000,000
40,000,000
50,000,000
60,000,000
70,000,000
80,000,000
90,000,000
100,000,000
110,000,000
120,000,000
130,000,000
24°0'0"E
26°0'0"E
28°0'0"E
30°0'0"E
42°0'0"N
43°30'0"N
N
0
50
100
200
km
Scale: 1:3,500,000
Map by: Vanya Stamenova, Stefan Stamenov
Source: Forest / Non-forest global dataset for 2010, JAXA JICA Database for Bulgaria, 2006

Table 5: Error matrix for Forest/Non-Forest dataset.

Land cover class (number of points)	Forest	Non-forest	Water	Total	User Accuracy (%)
Forest	156	41	0	197	79.2
Non-forest	32	149	3	184	81.0
Water	0	0	1	1	100.0
Total	188	190	4	382	0.0
Producer Accuracy (%)	83.0	78.4	25.0	0.0	80.1

3.4 Validation of the Tree Canopy Cover 2010 dataset

The tree canopy cover product for 2010 shows the tree cover in percentages from 0 to 100, for trees with height higher than 5 m. The map of the forests for the territory of Bulgaria generated using the tree canopy cover dataset is presented on Figure 8.

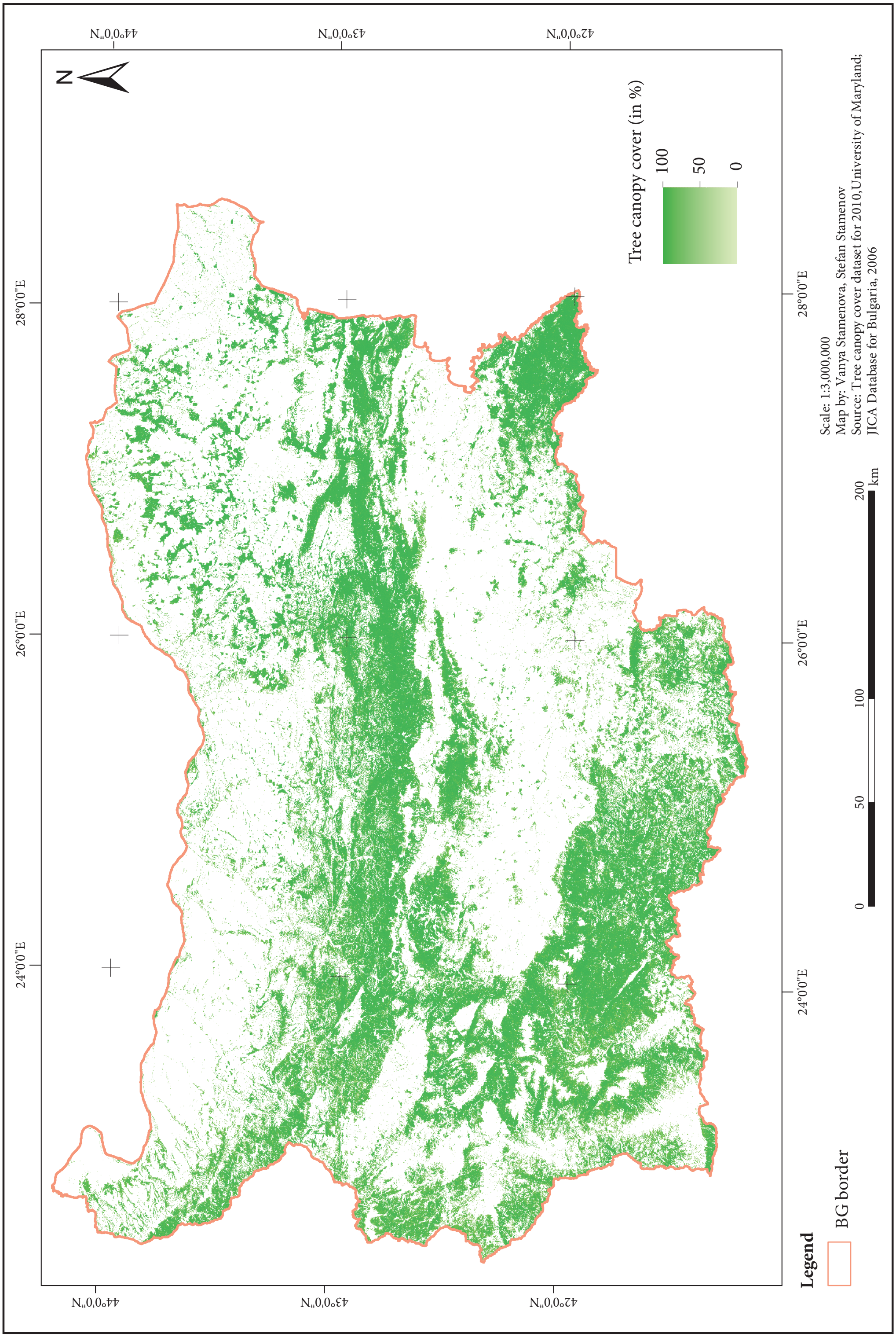
The validation for this dataset was made through visual interpretation and assigning true or false values (respectively 1 and 2) depending on the availability of forest for each sample unit. The estimated overall accuracy for the Tree canopy cover product is very high – 96.1%.

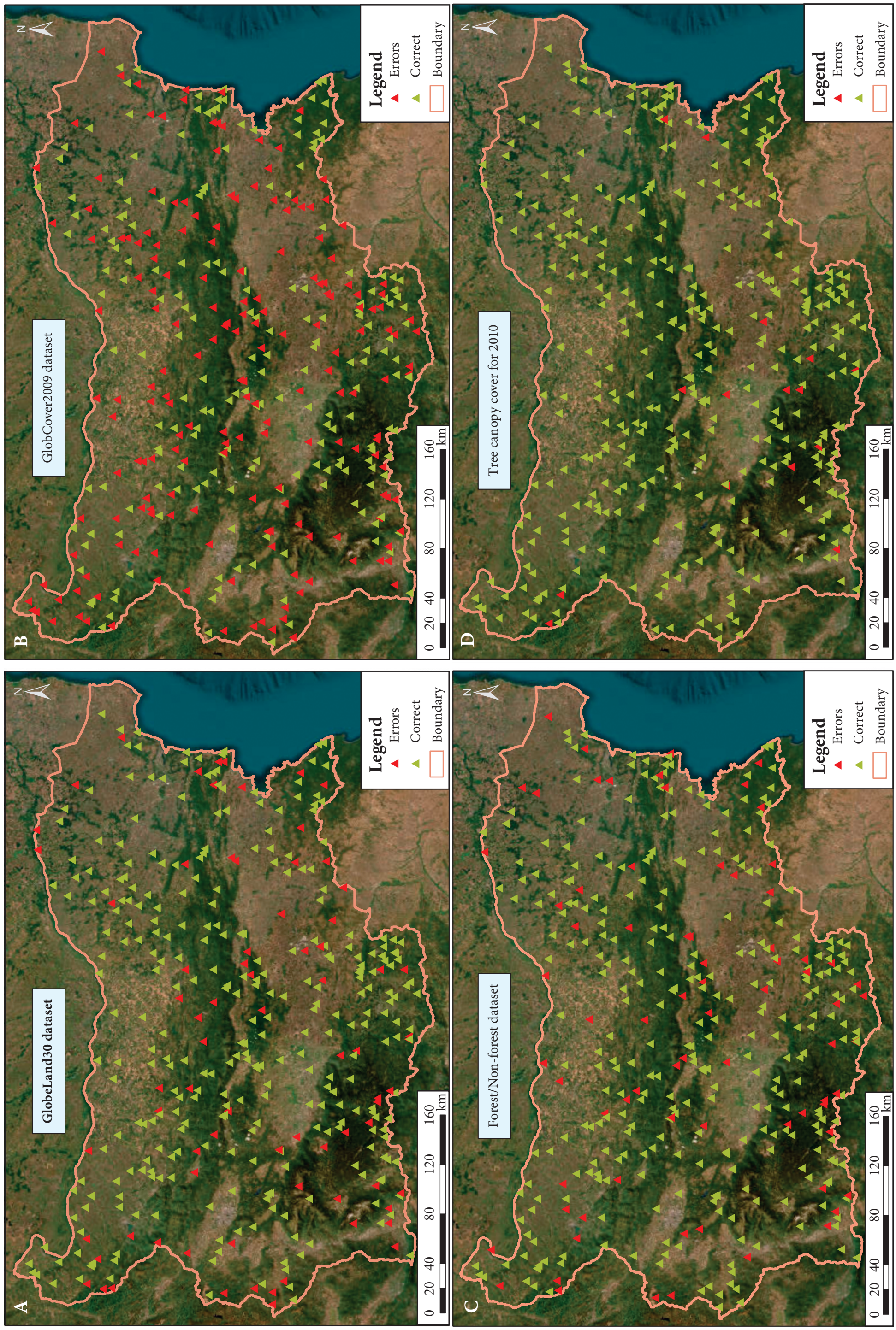
4 Discussion

The validation results of the four global land cover datasets show that the highest accuracy is observed for Tree canopy cover (96.1%), the lowest for GlobCover 2009 (51.0%), and the other products' accuracy is considerably high – about 80%. Each of the four GLC products, which have been validated, has its own specifics, spatial resolution and thematic classes' structure and content of the classes. The products differ in their spatial resolution (from 25 to 300 m), in the used land cover classification system and the number of land cover classes. The pixel size affects the details of the land cover products, there is a functional relationship between the pixel size and number of the details at the land cover product. It is obvious that the bigger the pixel size, the lower the number of details will be. On the other hand, the validation is made with satellite and aerial images with high and very high spatial resolution, where a detailed representation of the earth surface is observed. The pixel size influence on the accuracy of the GLC, since when the pixel size of the product is bigger each pixel covers area, which is much more likely to be heterogeneous than for the products which have smaller pixel size and is more likely to decrease the accuracy of the product. Also, when we have a product with more classes it represents more complex information about the Earth surface but at the same time it is more likely the product to have mixed pixels, and thus to result in higher inaccuracies in the assignment of the land cover classes.

Differences exist also in the way of determining the classes for the different land cover products. The vegetation and cultivated land cover classes of the ESA GLC product are defined as compound classes with percent ratios between the different types of vegetation while the other three products have only unitary land cover classes. In the cases when we have compound classes, the validation using high resolution reference data for the mixed pixels is a reason for ambiguity in how to assess the percent presence of each presented class and to which compound class to assign the pixel. The lower estimated accuracy for the GlobCover 2009 product is related to the large number of mixed pixels and their incorrect assignment to related land cover class. The results of the validation confirm that registering lower per class accuracies (user and producer) for the compound land cover classes (Table 4). The validation of the product with unitary classes is characterized with more unambiguous assignment to land cover classes and more often, the mistakes observed in their classifications are due to the misclassified pixels from classes, which are very close and difficult to distinguish, e.g. grassland and cultivated land.

Figure 8: Map of the tree canopy cover for the territory of Bulgaria for 2010. ➤ p. 79
Figure 9: Distribution of the correctly classified pixels and error pixels on the territory of Bulgaria for each of the global land cover datasets: A – GlobeLand30 dataset; B – GlobCover2009 dataset; C – Forest/Non-Forest dataset and D – Tree canopy cover dataset. ➤ p. 80


N
Tree canopy cover (in %)
100
50
0
Scale: 1:3,000,000
Map by: Vanya Stamenova, Stefan Stamenov
Source: Tree canopy cover dataset for 2010,University of Maryland;
JICA Database for Bulgaria, 2006
0
50
100
200
km
Legend
BG border
44°0'0"N
43°0'0"N
42°0'0"N
24°0'0"E
26°0'0"E
28°0'0"E



Scale: 1: 4,000,000 Map by: Vanya Stamenova, Stefan Stamenov Source: JICA Database for Bulgaria, 2006

The spatial distribution of the correctly classified and error pixels for each of the validated GLC datasets has been presented on Figure 9. As it's seen, the areas with high concentration of error pixels are the mountain and low-mountain regions in south and west Bulgaria – Rhodope, Kraishte and the central part of Stara planina (especially for Figure 9A and 9C), which are characterized with significant heterogeneity of the landscape and land cover. The error pixels for the GlobCover2009 are distributed comparatively equally over the territory of Bulgaria (Figure 9B), due to the large number of mixed pixels for this product. The errors for tree canopy cover are located also in the Rhodope Mountain area (Figure 9D).

The proposed method for validation of GLC datasets combines the capabilities of the online platform, the GIS processing techniques and functionality which is serious advantage and the interpretation of remote sensing and other supplementary data for assigning each of the sample units to the correct land cover class.

The main result product from the validation process is the created vector file, storing the information for the validation results of all examined datasets in its attribute table. This allows the obtained results to be stored in vector format, e.g. shapefile, which has small volume, high popularity and interoperability. This allows the available results from the already performed validation process to be used in future and allows the land cover classes for the sample units to be compared. In the land cover validation process visual interpretation stays as one of the most popular, effective and widespread methods. It is based on knowledge which every expert has (culture-based, nature-based knowledge), which provides the highest level of confidence (Chen et al. 2015; Zhang et al. 2016).

5 Conclusion

The presented article proposed an integrated method for land cover validation, which combines the GIS environment data integrity with classical expert knowledge-based method as visual interpretation with attempt to elaborate a determined and simple enough algorithm for visual interpretation, which can also be used in crowdsourcing approach for gathering and processing of information. The proposed method is designed as effective, time-saving and having a clear structure and sequence of the steps without being too complex.

ACKNOWLEDGMENT: The study has been realized by using an ArcGIS 10.5 Desktop software product, granted for non-commercial studies in the field of conservation by ESRI Conservation Program. The authors equally participated and contributed to data processing and validation and article preparation.

6 References

Achard, F., Bontemps, S., Herold, M., Mayaux, P., Arino, O., Kalogirou, V. 2011: Product validation plan. Project document, UCL-Geomatics. Belgium. Internet: http://www.esa-landcover-cci.org/?q=webfm_send/44 (17. 1. 2023).

Akovetsky, V. I. 1983: Deciphering images. Moscow.

Bontemps, S., Defourny, P., Bogaert, E. V., Kalogirou, V., Perez, J. R. 2011: GLOBCOVER 2009. Products description and validation report. Internet: https://epic.awi.de/id/eprint/31014/16/GLOBCOVER2009_Validation_Report_2-2.pdf (17. 1. 2023).

Brovelli, M., Molinari, M., Hussein, E., Chen, J., Li, R. 2015: The first comprehensive accuracy assessment of GlobeLand30 at a national level: Methodology and results. Remote Sensing 7-4. DOI: https://doi.org/10.3390/rs70404191

Chen J., Chen, L., Chen, F., Ban, Y., Li, S., Han, G., Tong, X., et al. 2021: Collaborative validation of GlobeLand30: Methodology and practices. Geo-spatial Information Science 24-1. DOI: https://doi.org/10.1080/10095020.2021.1894906

Chen, F., Chen, J., Wu, H., Hou, D., Zhang, W., Zhang, J., Zhou, X., et al. 2016b: A landscape shape index-based sampling approach for land cover accuracy assessment. Science China Earth Sciences 59-12. DOI: https://doi.org/10.1007/s11430-015-5280-5

Chen, J., Cao, X., Peng, S., Ren, H. 2017: Analysis and applications of GlobeLand30: A review. ISPRS International Journal of Geo-Information 6-8. DOI: https://doi.org/10.3390/ijgi6080230

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., et al. 2015: Global land cover mapping at 30m resolution: A POK-based operational approach. ISPRS Journal of Photogrammetry and Remote Sensing 103. DOI: https://doi.org/10.1016/j.isprsjprs.2014.09.002

Chen, J., Yifang, B., Songnian, L. 2014: China: Open access to Earth land-cover map. Nature 514. DOI: https://doi.org/10.1038/514434c

Chen, J., Zhang, J., Zhang, W. and Peng, S. 2016a: Continuous updating and refinement of land cover data product. National Remote Sensing Bulletin 20-5. DOI: https://doi.org/10.11834/jrs.20166250

Congedo, L., Munafò, M. 2012: Development of a methodology for land cover classification validation. Working paper. Rome. Internet: http://www.planning4adaptation.org/Docs/papers/Working%20Paper%20%20Development%20of%20Methodology%20for%20Land%20Cover%20Classification%20Validation.pdf (17. 1. 2023).

Di Gregorio, A., Jansen, L. J. 1998: Land Cover Classification System (LCCS): Classification concepts and user manual. Rome. Internet: https://www.fao.org/3/x0596e/x0596e00.htm (18. 1. 2023).

Food and agriculture organization of the United Nations 2012: FRA 2015: Terms and definitions. Rome. Internet: https://www.fao.org/3/ap862e/ap862e00.pdf (18. 1. 2023).

Food and agriculture organization of the United Nations 2000: FRA 2000. On definitions of forest and forest change. Rome. Internet: https://www.fao.org/3/ad665e/ad665e00.htm (18. 1. 2023).

Grekousis, G., Mountrakis, G., Kavouras, M. 2015. An overview of 21 global and 43 regional land-cover mapping products. International Journal of Remote Sensing 36-21. DOI: https://doi.org/10.1080/01431161.2015.1093195

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., Townshend, J. R. G. 2013: High-resolution global maps of 21st-century forest cover change. Science 342-6160. DOI: https://doi.org/10.1126/science.1244693

Hansen, M. C., Defries, R. S., Townshend, J. R. G., Sohlberg, R. 2000: Global land cover classification at 1 km spatial resolution using a classification tree approach. International Journal of Remote Sensing 21-6,7. DOI: https://doi.org/10.1080/014311600210209

Herold, M., See, L., Tsendbazar, N. E., Fritz, S. 2016: Towards an integrated Global Land Cover monitoring and mapping system. Remote Sensing 8-12. DOI: https://doi.org/10.3390/rs8121036

Japan aerospace exploration agency, Earth observation research center 2016: Global 25m resolution PALSAR mosaic and forest/non-forest map (FNF). Dataset description. Internet: https://www.eorc.jaxa.jp/ALOS/en/dataset/pdf/DatasetDescription_PALSAR_Mosaic_FNF_revO.pdf (18. 1. 2023).

European Commission, Eurostat 1999: Land cover and land use information systems for European Union policy needs, Proceedings of the seminar. Luxembourg. Internet: https://op.europa.eu/s/oCqL (18. 1. 2023).

Lillesand, T., Kiefer, R. 1999: Remote sensing and image interpretation. New York.

Loveland, T. R., Reed, B. C., Brown, J. F., 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. DOI: https://doi.org/10.1080/014311600210191

Mayaux, P., Eva, H., Gallego Pinilla, F. J., Strahler, A., Herold, M., Agrawal, S., Naumov, S., et al. 2006: Validation of the global land cover 2000 map. IEEE Transactions on Geoscience and Remote Sensing 44-7. DOI: https://doi.org/10.1109/TGRS.2006.864370

Morisette, J. T., Privette, J. L., Justice, C. O., Running S. W. (eds.) 1998: MODIS Land Validation Plan. Internet: https://modis.gsfc.nasa.gov/data/atbd/land_val.pdf (18. 1. 2023).

National geomatics center of China 2014: 30-meter global land cover dataset (GlobleLand30) product description. Internet: https://sustainable-caucasus.unepgrid.ch/documents/7/download (18. 1. 2023).

Olofsson, P., Stehman, S. V., Woodcock, C. E., Sulla-Menashe, D., Sibley, A. M., Newell, J. D., Friedl, et al. 2012: A global land-cover validation data set, part I: Fundamental design principles. International Journal of Remote Sensing 33-18. DOI: https://doi.org/10.1080/01431161.2012.674230

Penin, R. 2007: Natural geography of Bulgaria. Sofia.

Ran, Y., Li, X. 2015: First comprehensive fine-resolution global land cover map in the world from China – Comments on global land cover map at 30-m resolution. Science China Earth Sciences 58-9. DOI: https://doi.org/10.1007/s11430-015-5132-4

Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., Lucas, R. 2014: New global forest/non-forest maps from ALOS PALSAR data (2007–2010). Remote Sensing of Environment 155. DOI: https://doi.org/10.1016/j.rse.2014.04.014

Strahler, A. H., Boschetti, L., Foody, G. M., Friedl, M. A., Hansen, M. C., Herold, M., Mayaux, P. et al. 2006: Global land cover validation. Recommendations for evaluation and accuracy assessment of global land cover maps. Luxembourg. Internet: https://op.europa.eu/en/publication-detail/-/publication/52730469-6bc9-47a9-b486-5e2662629976 (18. 1. 2023).

Tsendbazar, N. E., de Bruin, S., Herold, M. 2015: Assessing global land cover reference datasets for different user communities. ISPRS Journal of Photogrammetry and Remote Sensing 103. DOI: https://doi.org/10.1016/j.isprsjprs.2014.02.008

Tsendbazar, N. E., de Bruin, S., Fritz, S., Herold, M. 2015: Spatial accuracy assessment and integration of global land cover datasets. Remote Sensing 7-12. DOI: https://doi.org/10.3390/rs71215804

Zhang, W., Chen, J., Liao, A., Han, G., Chen, X., Chen, L., Peng, S. et al. 2016: Geospatial knowledge-based verification and improvement of GlobeLand30. Science China Earth Sciences 59. DOI: https://doi.org/10.1007/s11430-016-5318-4