

Production of Global Land Cover Data – GLCNMO2013

Toshiyuki Kobayashi¹, Ryutarō Tateishi¹, Bayan Alsaaidh¹, Ram C. Sharma¹, Takuma Wakaizumi¹, Daichi Miyamoto¹, Xiulian Bai¹, Bui D. Long¹, Gegentana Gegentana¹, Aikebaier Maitiniyazi¹, Destika Cahyana¹, Alifu Haireti¹, Yohei Morifuji¹, Gulijianati Abake¹, Rendy Pratama¹, Naijia Zhang¹, Zilaitigu Alifu¹, Tomohiro Shirahata¹, Lan Mi¹, Kotaro Iizuka¹, Aimaiti Yusupujiang¹, Fedri R. Rinawan¹, Richa Bhattarai¹ & Dong X. Phong¹

¹ Center for Environmental Remote Sensing, Chiba University, Chiba, Japan

Correspondence: Toshiyuki Kobayashi, Center for Environmental Remote Sensing, Chiba University, 1-33 Yayoi-cho, Inage, Chiba 263-8522, Japan. Tel: 81-43-290-2964. E-mail: kobayashit125@gmail.com

Received: April 23, 2017

Accepted: May 23, 2017

Online Published: June 25, 2017

doi:10.5539/jgg.v9n3p1

URL: <http://dx.doi.org/10.5539/jgg.v9n3p1>

Abstract

Global land cover products have been created for global environmental studies by several institutions and organizations. The Global Mapping Project coordinated by the International Steering Committee for Global Mapping (ISCGM) has been periodically producing global land cover datasets as one of the eight basic global datasets. It has produced a new fifteen-second (approximately 500 m resolution at the equator) global land cover dataset – GLCNMO2013 (or GLCNMO version 3). This paper describes the method of producing GLCNMO2013. GLCNMO2013 has 20 land cover classes, and they were mapped by improved methods from GLCNMO version 2. In GLCNMO2013, five classes, which are urban, mangrove, wetland, snow/ice, and water were independently classified. The remaining 15 classes were divided into 4 groups and mapped individually by supervised classification. 2006 polygons of training data collected for GLCNMO2008 were used for supervised classification. In addition, about 3000 polygons of new training data were collected globally using Google Earth, MODIS Normalized Difference Vegetation Index (NDVI) seasonal change patterns, existing regional land cover maps, and existing four global land cover products. The primary data of this product were Moderate Resolution Imaging Spectroradiometer (MODIS) data of 2013. GLCNMO2013 was validated at 1006 sampled points. The overall accuracy of GLCNMO2013 was 74.8%, and the overall accuracy for eight aggregated classes was 90.2%. The accuracy of the GLCNMO2013 was not improved compared with the GLCNMO2008 at heterogeneous land covers. It is necessary to prepare the training data for mosaic classes and heterogeneous land covers for improving the accuracy.

Keywords: decision tree, Global Mapping, land cover, MODIS, training data

1. Introduction

Several attempts to map global land cover have been made to date. Global land cover datasets is used for environmental studies. We can know the distribution and the change of land covers on a global scale using those datasets. The GlobCover 2009 V2.3 global land cover map, the newest version of GlobCover, was derived from the Medium Resolution Imaging Spectrometer Instrument (MERIS) Fine Resolution (FR) surface reflectance mosaics for the year 2009 (Arino et al., 2008). It was developed using an automatic and regionally-tuned unsupervised classification technique except for urban and wetland areas (Arino et al., 2008; European Space Agency [ESA] and Université catholique de Louvain [UCL], 2011). The Terra and Aqua MODIS Collection 5.1 Land Cover Type product (MCD12Q1) was produced using the data from MODIS at annual time step since 2001. It was derived through a supervised decision-tree classification algorithm. The Moderate Resolution Imaging Spectroradiometer (MODIS) Nadir BRDF-Adjusted Reflectance (NBAR) data for bands 1-7 and Land Surface Temperature (LST) data were used as inputs (Friedl et al., 2010). The Global Land Cover 2000 (GLC2000) product was produced using the SPOT/VEGETATION data for the year 2000 (Bartholomé & Belward, 2010). The classification was carried out region by region. Recently, two versions of 30 m global land cover maps were produced based on the Landsat Thematic Mapper (TM) / Enhanced Thematic Mapper Plus (ETM+) images (Gong et al., 2013; J. Chen et al., 2015).

At the “Earth Summit” in 1992, the “Agenda21” was adopted, and the “Global Mapping” project was proposed to contribute to global environmental conservation. It is an international project to develop eight basic global

datasets including land cover through cooperation of the National Mapping Organizations (NMOs) from about 180 countries/regions. In 1996, the International Steering Committee for Global Mapping (ISCGM) was established. It worked for the “Global Mapping” project. The land cover product is called the Global Land Cover by National Mapping Organizations (GLCNMO) (Tateishi et al., 2011).

This paper describes the method we produced a new global land cover dataset, GLCNMO2013 (or GLCNMO version 3) under the “Global Mapping” project. The land covers of the GLCNMO2013 were mapped class by class. They were overlaid at the final step. The main parts for mapping GLCNMO2013 are to: (1) improve the mapping method used in GLCNMO2008; (2) add new training data to those collected for the GLCNMO2008 (or GLCNMO version 2) using Google Earth, MODIS Normalized Difference Vegetation Index (NDVI) seasonal change patterns, and existing regional land cover maps; and (3) modify the map based on the reports from the 19 countries participating in the Global Mapping project (Appendix A).

2. Data used

2.1 MODIS Data

The main data used in the GLCNMO2013 mapping project are the Global MODIS 500 m data processed by the Center for Environmental Remote Sensing (CEReS), Chiba University (Hoan, Tateishi, & Al-Bilbisi, 2013). The data observed in 2013 was used in the GLCNMO2013 mapping. The source MODIS data of this dataset were MCD43A4, “the MODIS/Terra+Aqua Nadir Bi-directional Reflectance Distribution Function (BRDF)-Adjusted Reflectance 16-day L3 Global 500 m SIN Grid V005”. The data are 16 day composite, 7 bands, 500 m data. The geometric accuracy of this dataset ranged from 96 m to 200 m in RMSEs when compared with the Landsat images downloaded from the GLCF of the University of Maryland. The geometric accuracy in Oceania was 264 m and 344 m in RMSEs for the east-west direction and north-south direction, respectively (Hoan et al., 2013). The no data values for MODIS 7 bands were replaced by the new values through linear interpolation from two values before and after the time (Hoan et al., 2013). The values of 2012 and 2014 were used when there was no data values more than three months.

2.2 Other Data

2.2.1 Global Satellite Data

The 30 arc second spatial resolution Defense Meteorological Satellite Program (DMSP) - Operational Linescan System (OLS) data were available from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI) website [w1]. The 30 m spatial resolution Landsat ETM+ data were available at no charge from the US Geological Survey (USGS) Earth Resources Observation and Science Center (EROS) website [w2]. The 1.6 arc second spatial resolution global Phased Array type L-band Synthetic Aperture Radar (PALSAR) orthorectified mosaic data were available from the Japan Aerospace Exploration Agency (JAXA) website [w3]. The global DMSP-OLS data and the global Landsat ETM+ data were used for urban mapping. The Global PALSAR mosaic data were used for mapping forest classes.

2.2.2 Global Land Cover Data

Four kinds of existing global land cover datasets were used to choose candidates for training data. The used products were the Global Land Cover 2000 (GLC2000) product by the Joint Research Centre [w4], the MCD12Q1 product by Boston University, the GLCNMO2008 product by ISCGM [w5], and the GlobCover2009 V2.3 product by the European Space Agency (ESA) [w6].

2.2.3 Regional Land Cover Maps

In the mapping of the GLCNMO2003 and the GLCNMO2008, existing regional land cover data were used to check intermediate results of classification. They were also used to choose candidates for training data (Tateishi et al., 2011; Hoan et al., 2013). The list of these datasets is available from the CEReS website [w7]. The same data were also used in the mapping process of the GLCNMO2013. In addition to the above maps, the most recent products of the above maps were also used in the GLCNMO2013.

2.2.4 Reference Data

High-resolution imagery displayed on Google Earth was used for reference data to confirm intermediate classification results. It was also used to collect training data for reference. We used many other data for mapping individual classes. A 30 m spatial resolution Advanced Spaceborne Thermal Emission (ASTER) Global Digital Elevation Model (GDEM) Version 2 data were used for water mapping. The GTOPO30, a 30 arc second (about 1km) spatial resolution DEM data, available from USGS were used for mangrove mapping. Population density data were used for urban mapping. The other data used for mapping are listed in “Appendix B”.

3. Land Cover Classification

3.1 Legend

The legend of land cover for GLCNMO2013 is the same as the former versions of GLCNMOs. Table 1 shows the land cover legend for the GLCNMO2013. The definition of the legend was given by Tateishi et al. (2011). It was based on the Land Cover Classification System (LCCS)

3.2 Method of Mapping

For the GLCNMO2013, the following 9 maps were produced individually.

- Sparse vegetation and bare area map (class code 10, 16 and 17 in Table 1)
- Herbaceous vegetation and shrub map (class code 7, 8 and 9 in Table 1)
- Forest map (class code 1, 2, 3, 4, 5 and 6 in Table 1)
- Agricultural map (class code 11, 12 and 13 in Table 1)
- Wetland map (class code 15 in Table 1)
- Mangrove map (class code 14 in Table 1)
- Snow/Ice map (class code 19 in Table 1)
- Urban map (class code 18 in Table 1)
- Water map (class code 20 in Table 1)

For the GLCNMO2008, 6 classes were mapped individually. However, herbaceous areas, forests and agricultural areas were not mapped well. We mapped these three classes separately for the GLCNMO2013. This method is an improved method of mapping the GLCNMO2008 (Tateishi et al., 2014).

Table 1. Land cover legend for GLCNMO

Code	GLCNMO land cover class	LCC Label
1	Broadleaf Evergreen Forest	Broadleaved Evergreen Closed to Open (100-40%) Trees
2	Broadleaf Deciduous Forest	Broadleaved Deciduous Closed to Open (100-40%) Trees
3	Needleleaf Evergreen Forest	Needleleaved Evergreen Closed to Open (100-40%) Trees
4	Needleleaf Deciduous Forest	Needleleaved Deciduous Closed to Open (100-40%) Trees
5	Mixed Forest	Broadleaved Closed to Open Trees and Needleleaved Closed to Open (100-40%) Trees
6	Tree Open	Open (40-(20-10)%) Trees (Woodland)
7	Shrub	Closed to Open Shrubland (Thicket)
8	Herbaceous	Closed to Open Herbaceous Vegetation, Single Layer
9	Herbaceous with Sparse Tree / Shrub	Closed to Open Herbaceous Vegetation with Trees and Shrubs
10	Sparse Vegetation	Sparse Herbaceous Vegetation // Sparse Woody Vegetation
11	Cropland	Herbaceous Crop(s)
12	Paddy field	Graminoid Crops
13	Cropland / Other Vegetation Mosaic	Cultivated and Managed Terrestrial Area(s), and Natural and Semi-Natural Primarily Terrestrial Vegetation // Cultivated Aquatic or Regularly Flooded Area(s), and Natural and Semi-Natural Primarily Terrestrial Vegetation
14	Mangrove	Closed to Open Woody Vegetation with Water Quality: Saline Water
15	Wetland	Closed to Open Woody Vegetation with Water Quality: Fresh Water // Closed to Open Woody Vegetation with Water Quality: Brackish Water // Closed to Open Herbaceous Vegetation with Water
16	Bare area, consolidated (gravel, rock)	Consolidated Material(s)
17	Bare area, unconsolidated (sand)	Unconsolidated Material(s)
18	Urban	Artificial Surfaces and Associated Area(s)
19	Snow / Ice	Perennial Snow // Perennial Ice
20	Water bodies	Artificial Waterbodies // Natural Waterbodies

The classification was done by continental basis (Eurasia, North America, South America, Africa, and Oceania). The final global map was produced by mosaicking five maps at the final step. The main features of the method are as follows:

- i. Modification of the individual mapping methods of the GLCNMO2008 for five classes (wetland, mangrove, snow/ice, urban, and water);
- ii. Dividing into four groups for the remaining 15 classes and mapping separately;
- iii. Reuse of the training data used for the former versions of GLCNMO after verification;
- iv. Collection of the new training data using Google Earth, MODIS NDVI seasonal change patterns, and the existing regional land cover maps, with reference of the existing four global land cover products.
- v. Modification of the map based on the reports from 21 countries, a land cover map for Viet Nam and aerial photos for New Zealand.

Figure 1 shows the whole flow of the mapping for GLCNMO2013.

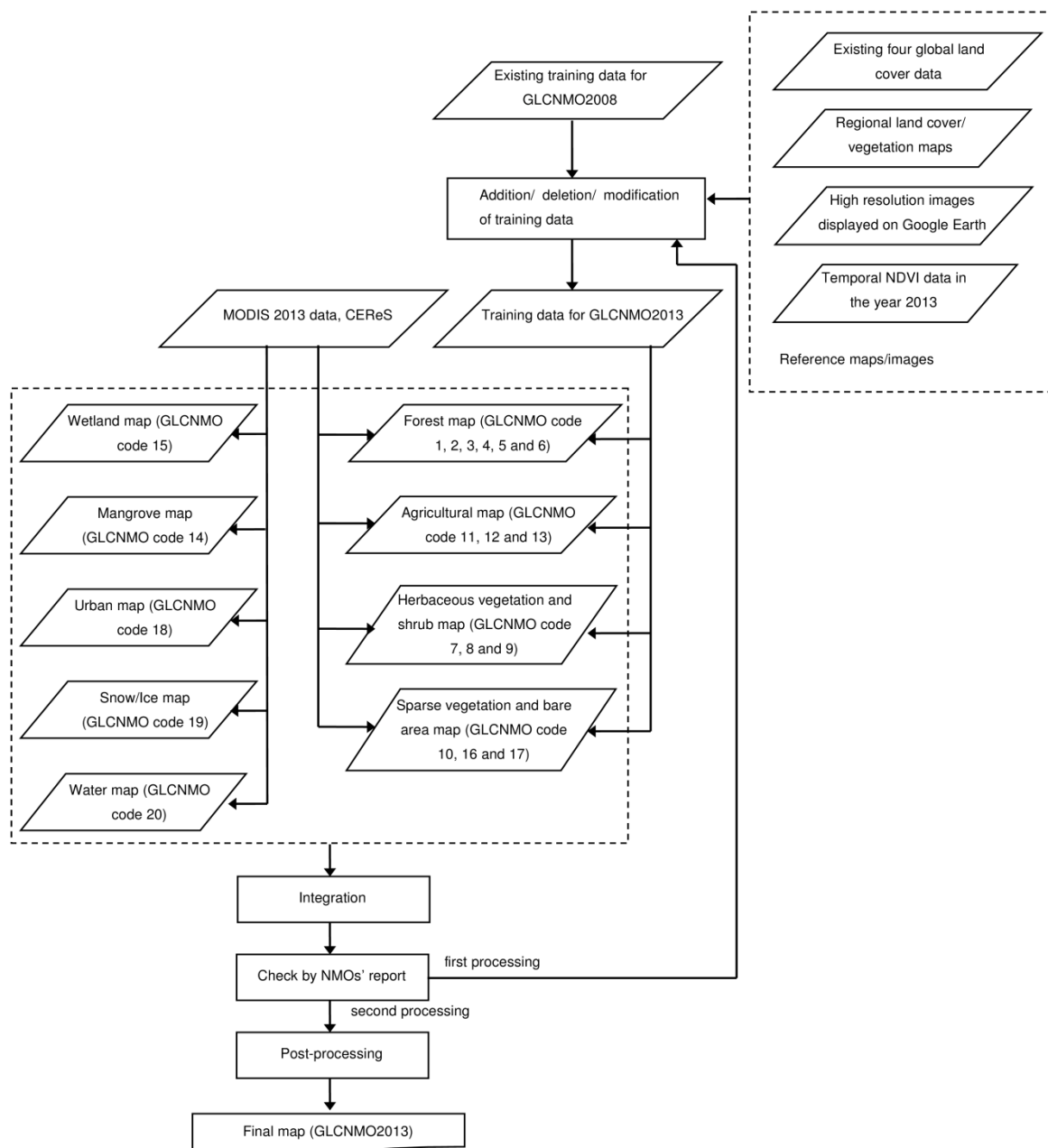


Figure 1. Flow of GLCNMO2013 mapping

3.3 Sampling Procedures

The existing training data used for the previous versions of GLCNMO were used as training data for mapping 15 land cover classes: five kinds of forest classes, tree open, shrub, two kinds of herbaceous vegetation classes, three kinds of agricultural classes, sparse vegetation, and two kinds of bare area classes. These data were updated by examining whether the land covers were changed or not between 2008 and 2013, using the following references:

- Four kinds of existing global land cover maps;
- High resolution images around 2008 and 2013 displayed on Google Earth;
- Existing regional land cover maps;
- NDVI seasonal change patterns of the year 2008 and 2013 calculated from the global MODIS data;
- Color composite images of MODIS data.

287 535 pixels from 2006 polygons from the previous GLCNMO version were used for mapping 15 land cover classes (Figure 2).

In addition to the above training data, new training data were collected for mapping 15 land cover classes in the GLCNMO2013. The candidate areas for collecting training data were selected and examined with the reference of the data used for updating the existing training data. The training data in open forests were also collected in this step, because there was no available training data in open forests in the previous versions of GLCNMO. These new training data were individually collected for producing four kinds of maps (forest map, agricultural field map, herbaceous vegetation and shrub map, and sparse vegetation and bare area map). The training data collected for producing forest map, those for producing agricultural map, those for producing herbaceous vegetation and shrub map, and those for producing sparse vegetation and bare area map were also used for mapping other classes. More than 3,000 polygons for training were newly collected. In total, more than 5000 training polygons were used.

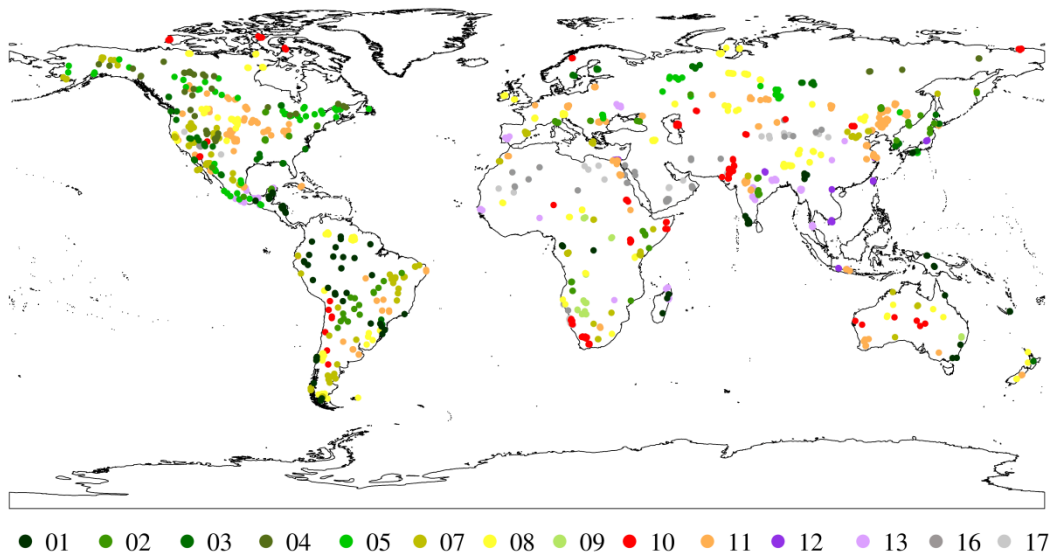


Figure 2. Distribution of existing 2,006 training polygons updated for GLCNMO2013 mapping

3.4 Classification

Nine kinds of land cover maps were produced independently. Several mapping methods were applied for each class, and the method showed the best result were selected as the final mapping method for each class.

3.4.1 Sparse Vegetation and Bare Area Map (GLCNMO Class Code 10, 16 and 17)

Sparse vegetation and bare areas were mapped using a hierarchical method. First, sparse vegetation and non-vegetation areas were extracted using threshold values of NDVI. Second, two types of bare areas and sparse vegetation were mapped using maximum likelihood method. Inputs used for maximum likelihood classification were 11 periods of MODIS bands 1, 2, 5 and 7, NDVI, the ratio of band 7 to band 6, and Bare Soil Index (BSI). The ratio of band 7 to band 6 can detect vegetation water content (Guerschman et al., 2009). It is able to identify

dry vegetation from bare soil, due to cellulose scattering (Guerschman et al., 2009). The BSI was proposed to derive bare lands by W. Chen et al. (2004). The BSI was calculated from MODIS data as follows:

$$BSI = \frac{(b_1 + b_4) - (b_2 + b_3)}{(b_1 + b_4) + (b_2 + b_3)} \quad (1)$$

In those equations, b_i = reflectance of MODIS band i .

The steps for mapping sparse vegetation and bare areas (GLCNMO class code 10, 16 and 17) were as follows.

Step 1: 23 periods of NDVI, 11 periods of the ratio of band 7 to band 6, and BSI were calculated.

Step 2: Training data were collected (Section 3.3).

Step 3: Non-vegetation areas (including sparse vegetation) were extracted using the maximum value of NDVI in 23 periods as threshold values.

Step 4: Sparse vegetation and two kinds of bare areas were mapped by maximum likelihood method using 77 variables (11 periods of MODIS bands 1, 2, 5 and 7, NDVI, BSI, and the ratio of band 7 to band 6).

From the visual check as a reference of Google Earth, it was found that small areas of sparse vegetation were misclassified as consolidated bare areas.

3.4.2 Herbaceous Vegetation and Shrub Map (GLCNMO Class Code 7, 8 and 9)

The potential maps were produced for mapping herbaceous vegetation and shrub areas, which showed the agreement of four global land cover products. The reliable areas of three classes (shrub, herbaceous, and herbaceous with sparse tree/ shrub) were mapped as each class in GLCNMO2013 because the areas classified as the same class in all four or three existing products means that these areas were most probably correctly classified. Classification was conducted only for the areas except reliable areas of three classes.

Herbaceous vegetation and shrub areas were mapped by the following steps.

Step 1: "Potential map of herbaceous vegetation and shrub areas" and "potential map of sparse vegetation, bare areas and cropland" were produced by overlaying four kinds of existing global land cover maps.

Step 2: Reliable areas and unreliable areas were extracted from the potential map produced at Step 1.

Step 3: Training data of each land cover class were collected from unreliable areas of herbaceous vegetation and shrub areas (Section 3.3).

Step 4: Three classes of land cover (shrub, herbaceous, and herbaceous with sparse tree/ shrub) in unreliable areas were mapped by maximum likelihood method using 12 periods of MODIS bands 1, 2, 4, 6, 7, and NDVI data.

Step 5: Training data of each land cover class were collected from unreliable areas of sparse vegetation, bare areas and cropland.

Step 6: Land covers of sparse vegetation, bare areas and cropland in unreliable areas of sparse vegetation, bare areas and cropland were mapped by maximum likelihood method using 12 periods of MODIS bands 1, 2, 4, 6, 7, and NDVI data.

Step 7: The areas of sparse vegetation, bare areas and cropland mapped at Step 6 were removed from the map produced at Step 4 if the areas classified as shrub, herbaceous, or herbaceous with sparse tree/ shrub in the map produced at Step 4 were also mapped as sparse vegetation, bare areas or cropland in the map produced at Step 6.

Step 8: Reliable areas of three classes (shrub, herbaceous, herbaceous with sparse tree/ shrub) were overlaid on the map produced at Step 7.

3.4.3 Forest Map (GLCNMO Class Code 1, 2, 3, 4, 5 and 6)

For forest mapping, decision tree modelling was accomplished using a commercial software (See5; RuleQuest Research). The theory of See5 was developed by Quinlan (1993). The method for combining multiple tree models (committee models) was used in mapping. Predictor variables used for decision tree modelling were 176 variables obtained from MODIS data, and 4 variables obtained from PALSAR data. 176 predictor variables obtained from MODIS data were;

- The 12 composite periods (1, 3, 5, ..., 23) of the data in 2013 for MODIS seven bands, NDVI, Green and Red ratio Vegetation Index (GRVI) (Falkowski, Gessler, Morgan, Hudak, & Smith, 2005), and Land Surface Water Index (LSWI) (Gao, 1996; Xiao et al., 2002) (totaling $10 \times 12 = 120$ variables),

- MODIS seven bands, NDVI, GRVI, and LSWI values at the date with the maximum NDVI values (totaling 10 variables),
- Annual mean, maximum, minimum, and the difference of maximum and minimum values for MODIS seven bands, NDVI, GRVI, and LSWI (totaling $4 \times 10 = 40$ variables), and
- The six phenological variables based on NDVI, which are the length of the growing season, rate of NDVI increase at the beginning of the vegetation season, rate of NDVI decrease at the end of the vegetation season, time for the start of the growing season, time for the end of the growing season, and the integral of NDVI values at the vegetation season. These variables were calculated using TIMESAT software automatically (Eklundh & Jönsson, 2012; Jönsson & Eklundh, 2002; Jönsson & Eklundh, 2004), available from TIMESAT home page [w8].

The NDVI, GRVI and LSWI were calculated from MODIS data as follows:

$$\text{NDVI} = \frac{b_2 - b_1}{b_2 + b_1} \quad (2)$$

$$\text{GRVI} = \frac{b_4 - b_1}{b_4 + b_1} \quad (3)$$

$$\text{LSWI} = \frac{b_2 - b_6}{b_2 + b_6} \quad (4)$$

In those equations, b_i = reflectance of MODIS band i . The GRVI has a stronger correlation with crown closure than NDVI (Falkowski et al., 2005). The LSWI is effective for classifying forests (Gao, 1996). The four variables obtained from PALSAR data were backscattering coefficients of HH polarization (H polarization transmit and H polarization receive) and HV polarization, the difference value of HH and HV (HH-HV), and the ratio value of HH and HV (HH/HV). As the pixel size for PALSAR data (25 m) differed for MODIS data, PALSAR data were resampled into the same size of MODIS data using mean of all pixels within MODIS pixels. For collecting training data, the map of terrestrial ecoregions of the world produced by Olson et al. (2001) was used as a reference, in addition to the reference data described in the Section 3.3.

The steps for mapping forest classes (GLCNMO class code 1, 2, 3, 4, 5 and 6) were as follows.

Step 1: 180 predictor variables for mapping were calculated.

Step 2: Training data (forest, tree open, and other classes) were collected (Section 3.3).

Step 3: The decision tree model for classifying into 3 classes (forest, tree open, and other classes) was produced using See5 software, and forest and tree open areas were mapped.

Step 4: Training data for forest classes were grouped into several hundreds of sub-classes (e.g. around 350 classes for Eurasia) from the temporal profiles of NDVI data, and they were mapped using decision tree method. The mapped classes were labeled as 5 classes, broadleaf evergreen forest, broadleaf deciduous forest, needleleaf evergreen forest, needleleaf deciduous forest, or mixed forest, using reference data.

Step 5: The final forest map (GLCNMO class code 1, 2, 3, 4, 5 and 6) was obtained by integrating two maps produced at the above steps.

3.4.4 Agricultural Map (GLCNMO Class Code 11, 12 and 13)

For agricultural area mapping, “potential cropland (including paddy) map” was produced to collect training data beforehand. It was produced by overlaying four existing global land cover products. It shows the agreement of four products or reliability of the mapped result for cropland. For example, the areas classified as cropland in all four existing products means that these areas are most probably cropland. The detailed method of producing “potential map” was given by Zhang & Tateishi (2013) and Tateishi et al. (2014). Predictor variables used for decision tree modelling were 23 periods of MODIS bands 1-7 reflectance, NDVI data, Normalized Difference Flood Index 2 (NDFI2) data (Boschetti, Nutini, Manfron, Brivio, & Nelson, 2014), and Normalized Difference Soil Index (NDSI) data (Rogers & Kearney, 2004). Though the different acronyms of NDSI and NDFI2 have been proposed in the remote sensing community, the NDFI2 and NDSI in this paper were calculated from MODIS data as follows:

$$\text{NDFI2} = \frac{b_1 - b_7}{b_1 + b_7} \quad (5)$$

$$NDSI = \frac{b_6 - b_2}{b_6 + b_2} \quad (6)$$

In those equations, b_i = reflectance of MODIS band i .

The agricultural area mapping was carried out as follows.

Step 1: “Potential map of cropland (including paddy)” and “potential map of cropland/ other vegetation mosaic” were produced by overlaying four kinds of existing global land cover maps.

Step 2: Reliable areas and unreliable areas were extracted from potential maps produced in the Step 1.

Step 3: Training data of paddy, cropland, cropland/ other vegetation mosaic, and other land cover classes were collected as a reference of reliable and unreliable areas of three classes (Section 3.3).

Step 4: Three classes of land cover (cropland and paddy, cropland/ other vegetation mosaic, and other land covers) were mapped using decision tree method. 23 periods of MODIS bands 1-7, NDVI and NDSI data were used for predictor variables (totaling 207 variables).

Step 5: The map produced at the above step was visually checked. Three classes were mapped again region by region, by adding newly collected training data, if the result was not good.

Step 6: Cropland and paddy areas were classified by decision tree method using 23 periods of MODIS bands 1-7, NDVI, NDSI and NDFI2 data as predictor variables (totaling 230 variables). The classified results were visually checked and modified.

3.4.5 Wetland Map (GLCNMO Class Code 14)

Wetlands larger than 500 km² by Ramsar Sites Database were mapped for GLCNMO2013, while only wetlands larger than 1000 km² were mapped for GLCNMO2008. The methodology used for mapping wetlands for GLCNMO2013 was the same as that for GLCNMO2008. The threshold method was used for extracting wetland areas. The best index and period from 23 periods of MODIS tasseled cap indices were manually selected for each wetland mapping. The method of wetland mapping was published by Tana & Tateishi (2013).

3.4.6 Mangrove Map (GLCNMO Class Code 15)

Mangrove areas in 2013 were mapped by maximum likelihood method using three indices calculated from MODIS data and DEM data. The method of mangrove mapping was published by Alsaaidh, Al-Hanbali, Tateishi, Kobayashi, & Hoan (2013).

3.4.7 Snow/Ice Map (GLCNMO Class Code 19)

The permanent Snow/ Ice map in 2013 was produced using the same method used for mapping Snow/ Ice areas of GLCNMO2008. It was produced by threshold method using brightness index and wetness index calculated from MODIS data in 2013 by Tasseled cap transformation. The threshold values of Tasseled cap wetness index and Tasseled cap brightness index used for GLCNMO2013 were different from those for GLCNMO2008. Permanent Snow/ Ice areas were extracted when the number of snow periods was more than 20 among 23 periods and the areas was not water cover at the other periods. The method to produce Snow/ Ice map was described in Tateishi et al. (2014).

3.4.8 Urban Map (GLCNMO Class Code 18)

Urban map in 2013 was produced using population data in 2012, DMSP-OLS night-time light data in 2010, Impervious Surface Area (ISA) data in 2010, and MODIS NDVI data in 2013 by thresholding method. The values for thresholding were decided based on the Gross Domestic Product (GDP) per capita. The method to produce urban map was published by Phong, Nguyen, Kobayashi, & Tateishi (2013).

3.4.9 Water Map (GLCNMO Class Code 20)

The water map in 2013 was produced using Superfine Water Index (SWI) developed by Sharma, Tateishi, Hara, & Nguyen (2015). The SWI were calculated from MODIS data as follows:

$$SWI = \frac{Sat_{(RGB)} - 7 \times b_2}{Sat_{(RGB)} + 7 \times b_2} \quad (7)$$

In those equations, b_i = reflectance of MODIS band i , and $Sat_{(RGB)}$ was obtained from the Hue-Saturation-Value (HSV) transformation of the RGB composite (bands 1, 3 and 4) of the MODIS data. Region specific (10 degrees × 10 degrees) thresholding of trimonthly color composite images and masking of mountainous shadows using ASTER GDEM elevation data were also conducted for mapping. The detail method to produce water map was

described in another paper (Sharma et al. 2015).

3.5 Integration and Post-processing for Final Mapping

20 classes classified at Section 3.4 were integrated as shown in Figure 3. It means that “Water” had higher priority than the class “Bare areas”. The order was decided from the accuracy of the maps. GLCNMO2008 product was used as a base map, because there were areas where none of 20 classes were assigned. The integrated five continental maps were checked by 19 NMOs (Appendix A). The methods for analysis were different for NMOs. GLCNMO2013 map was modified based on the reports from NMOs. This step was important because training data in some areas could not be collected without local knowledge.

As the post-processing, continental data were combined to global data. The area of 80-90 degree north was added using the GlobCover2009 V2.3 product. This was because the MODIS data covered up to only 80 degree in north latitude. For Antarctica, “land” and “lake” in the Scientific Committee on Antarctic Research (SCAR) Antarctic Digital Database were overlaid on GLCNMO2013. By these steps, the global land cover data, GLCNMO2013, was completed.

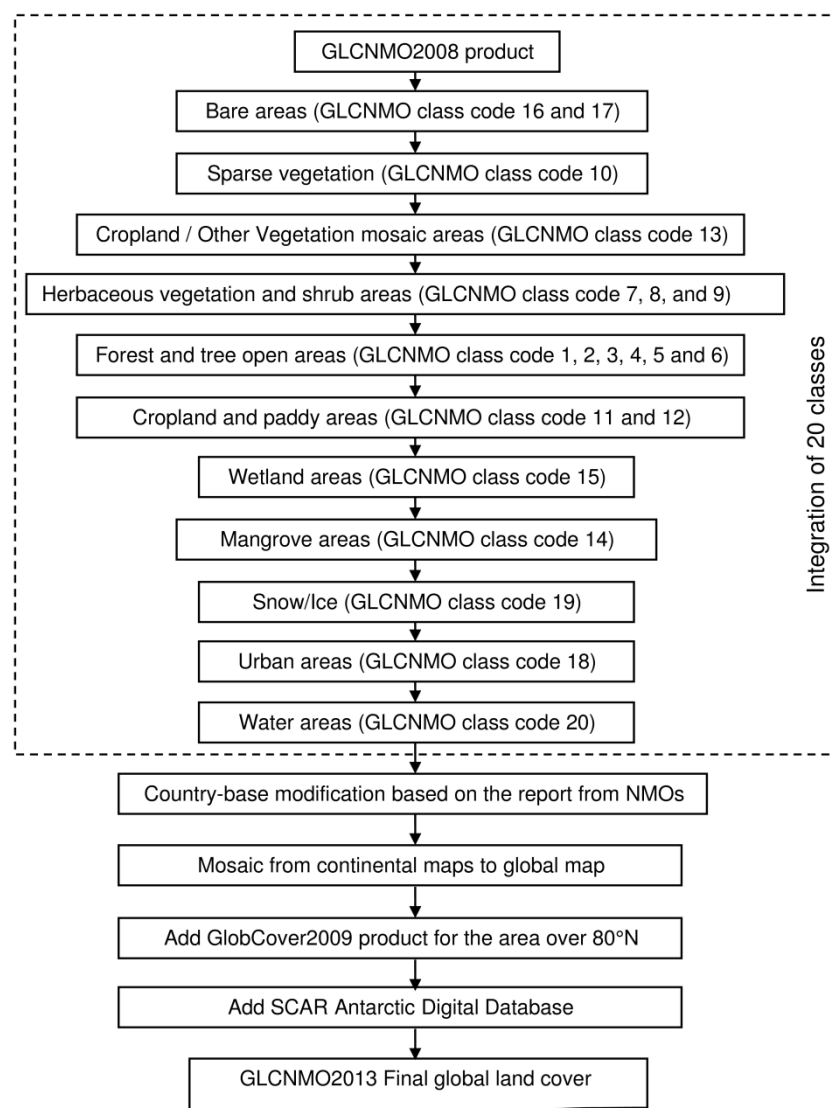


Figure 3. Flow of integration and post-processing

4. Results

Figure 4 shows the final map product, the GLCNMO2013.

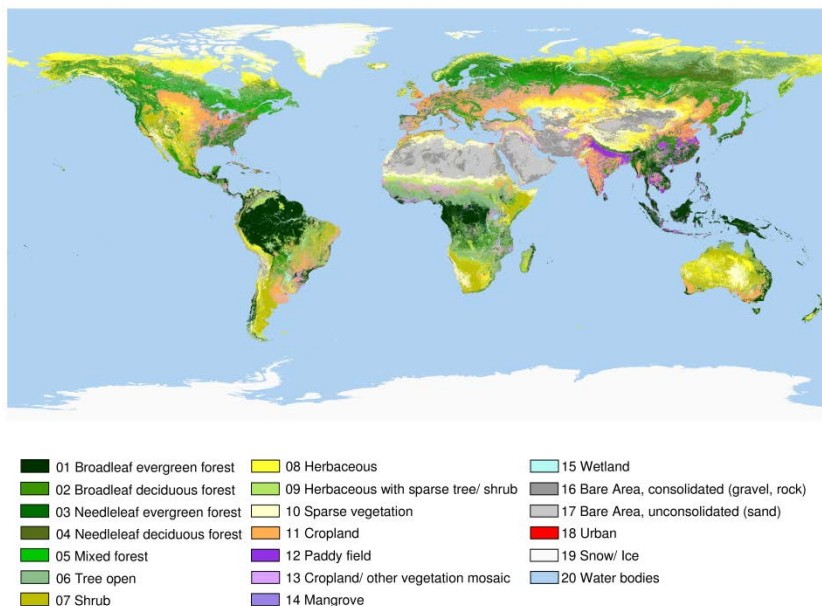


Figure 4. GLCNMO2013 (GLCNMO version 3)

The GLCNMO2013 was validated at the points where the GLCNMO2008 was validated. Several points were excluded from validation of the GLCNMO2013 because there was the possibility of land cover change from the year 2008. They were identified using the reference data described at Section 3.3. In those cases, we added the new points for validation. In total, 1006 points were used for validation globally. A confusion matrix of the GLCNMO2013 is shown in Table 2. The overall accuracy was 74.8%. The average user’s accuracy and the average producer’s accuracy was 75.2% and 74.0%, respectively. The confusion matrix for aggregated classes is shown in Table 3. Similar classes were aggregated to eight classes in Table 3: Forest, Other natural vegetation, Cropland, Wetland, Bare area/ sparse vegetation, Urban, Snow/ Ice, and Water. The aggregated overall accuracy, the average of user’s accuracy, and the average of producer’s accuracy were 90.2%, 93.0% and 90.3%, respectively. The overall accuracy increased about 15.4% by aggregation of classes from 20 to 8.

Table 2. Confusion matrix of GLCNMO2013

Code ¹	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Total	User’s accuracy (%)
1	68	1	2	0	1	5	0	0	0	0	0	0	10	3	1	0	0	0	0	0	81	84
2	1	31	0	4	6	2	0	0	1	0	0	0	0	0	2	0	0	0	0	0	47	66
3	1	1	33	2	0	3	0	0	0	0	0	0	0	0	2	0	0	0	0	0	42	79
4	0	1	5	29	7	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	45	64
5	0	5	9	6	30	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	52	58
6	1	1	3	2	1	32	5	1	11	0	1	0	5	0	4	0	0	0	0	1	68	47
7	0	0	0	0	0	1	35	3	7	2	1	0	2	0	1	0	0	0	0	0	52	67
8	0	0	0	0	0	0	1	27	4	7	0	0	0	0	0	2	0	0	1	0	42	64
9	0	0	0	0	0	1	5	11	23	2	0	0	3	0	1	0	0	0	0	0	46	50
10	0	0	0	0	0	0	1	7	0	36	0	0	0	0	2	5	1	0	0	2	54	67
11	0	2	0	0	0	3	2	3	2	0	48	4	6	0	0	0	0	0	0	0	70	69
12	0	0	0	0	0	0	0	0	0	6	36	1	0	0	0	0	0	0	0	0	43	84
13	1	0	0	0	0	1	1	3	0	0	9	4	27	0	0	0	0	1	0	0	47	57
14	0	0	0	0	0	0	0	0	0	0	0	0	0	39	1	0	0	0	0	0	40	98
15	0	0	0	0	0	0	0	0	0	0	0	2	0	1	26	0	0	0	0	1	30	87
16	0	0	0	0	0	0	1	0	0	4	0	0	0	0	0	31	4	0	0	0	40	78
17	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	2	40	0	0	0	45	89
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	53	0	0	53	100
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	56	0	57	98
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	52	52	100
Total	72	42	52	43	45	51	51	55	48	54	65	47	45	43	40	41	45	54	57	56	1006	
Producer’s accuracy (%)	94	74	63	67	67	63	69	49	48	67	74	77	60	91	65	76	89	98	98	93		

Overall accuracy = 74.8%; average of user's accuracy = 74.0%; average of producer's accuracy = 75.2%.

1 "Code" corresponds to the GLCNMO2013 code shown in Table 1.

Table 3. Confusion matrix of aggregated classes of GLCNMO2013

Class name	Corresponding class code before aggregation	1	2	3	4	5	6	7	8	Total	User's accuracy (%)
1. Forest	1, 2, 3, 4, 5, 6	296	18	8	12	0	0	0	1	335	88.4
2. Other natural vegetation	7, 8, 9	2	116	6	2	13	0	1	0	140	82.9
3. Cropland	11, 12, 13	7	11	141	0	0	1	0	0	160	88.1
4. Wetland	14, 15	0	0	2	67	0	0	0	1	70	95.7
5. Bare area/ Sparse vegetation	10, 16, 17	0	9	0	2	126	0	0	2	139	90.6
6. Urban	18	0	0	0	0	0	53	0	0	53	100.0
7. Snow/ Ice	19	0	0	0	0	1	0	56	0	57	98.2
8. Water	20	0	0	0	0	0	0	0	52	52	100.0
Total		305	154	157	83	140	54	57	56	1006	
Producer's accuracy (%)		97.0	75.3	89.8	80.7	90.0	98.1	98.2	92.9		

Overall accuracy = 90.2%; average of user's accuracy = 90.3%; average of producer's accuracy = 93.0%.

5. Discussion

The GLCNMO2013 (or GLCNMO version 3), a pixel size of 15 arc second global land cover product, has been produced using the MODIS 500 m data acquired in 2013. The GLCNMO2013 was produced by improved methods from the GLCNMO version 2. The main improved parts in the method are:

1. The 15 classes mapped together by supervised classification in GLCNMO2008 were divided into 4 groups and mapped individually;
2. More than 3000 polygons of new training data for supervised classification were collected globally in addition to about 2000 polygons of training data collected for GLCNMO2008.
3. Water map was produced by the new method developed by Sharma et al. (2015).
4. Smaller wetlands were mapped compared with GLCNMO2008.

The classification accuracy for "class code 5, 6, 9 and 13" in Table 1 was lower. The forest type was difficult to distinguish using high resolution images in Google Earth. We could not collect enough training data for mixed forest (code 5). We also could not collect enough training data for tree open. In the GLCNMO2008, training data for tree open was not collected, because tree open (code 6) was mapped using the continuous tree cover data. In addition, for collecting training and validation data, distinguishing code 6, 8, and 9 was difficult. Especially, mosaic classes (code 9 and 13) were difficult to distinguish with other classes. We think it was difficult to map those classes for MODIS resolution.

The accuracy of the GLCNMO2013 was not improved compared with the GLCNMO2008. The method used in the GLCNMO2013 needs to be improved, especially for heterogeneous land covers. Preparing the training data for mosaic classes and heterogeneous land covers is important for improving the accuracy of the GLCNMO. However, we think that one of the most important things for the global land cover mapping was producing the global land cover data continuously. GLCNMOs have been produced at an interval of five years.

It is also important to compare the GLCNMO2013 with other existing global land cover datasets. In this paper, we could not compare our result with other datasets, because the definition of land cover classes is different each other. The comparison among the land cover datasets is one of the future issues. In addition, there was a possibility that the validation points had biases because the points were not collected by random sampling. This point is also one of the future issues. We will revise the method of producing the global land cover map and compare our result with other datasets at the next step.

The product is available from ISCGM website and CERES website (Appendix C).

Acknowledgments

This work was conducted as an activity of the Global Mapping Project coordinated by the International Steering Committee for Global Mapping (ISCGM). The authors express their sincere gratitude to the collaborating National Geospatial Information Authorities (or National Mapping Organizations) (Appendix B) for their contribution of intermediate validation, and thank the Secretariat of ISCGM, Geospatial Information Authority of Japan (GSI), for their assistance to cooperate with NMOs, especially Masaki Suga of GSI for his valuable comments.

References

- Alsaaidh, B., Al-Hanbali, A., Tateishi, R., Kobayashi, T., & Hoan, N. T. (2013). Mangrove forests mapping in the southern part of Japan using Landsat ETM+ with DEM. *Journal of Geographic Information System*, *5*, 369-377. <https://doi.org/10.4236/jgis.2013.54035>
- Arino, O., Bicheron, P., Achard, F., Latham, J., Witt, R., & Weber, J. L. (2008). GLOBCOVER: The most detailed portrait of Earth. *European Space Agency Bulletin*, *136*, 24-31. Retrieved from <https://earth.esa.int/web/guest/-/globcover-the-most-detailed-portrait-of-earth-5910>
- Bartholomé, E., & Belward, A. S. (2010). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, *26*, 1959-1977. <https://doi.org/10.1080/01431160412331291297>
- Boschetti, M., Nutini, F., Manfron, G., Brivio, P. A., & Nelson, A. (2014). Comparative analysis of normalised difference spectral indices derived from MODIS for detecting surface water in flooded rice cropping systems. *PLOS ONE*, *9*, 1-21. <https://doi.org/10.1371/journal.pone.0088741>
- Chen, J., Chen J., Liao, A., Cao X., Chen, L., Chen, X., ... Mills, J. (2015). Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, 7-27. <https://doi.org/10.1016/j.isprsjprs.2014.09.002>
- Chen, W., Liu, L., Zhang, C., Wang, J., Wang, J., & Pan, Y. (2004). Monitoring the seasonal bare soil areas in Beijing using multi-temporal TM images. *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 3379-3382). Anchorage, USA.
- Eklundh, L., & Jönsson, P. (2012). *TIMESAT 3.1 software manual*, Sweden: Lund University.
- European Space Agency (ESA) and Université catholique de Louvain (UCL). (2011). *GLOBCOVER 2009 product description and validation report*. Retrieved from http://due.esrin.esa.int/page_globcover.php
- Falkowski, M.J., Gessler, P. E., Morgan, P., Hudak, A. T., & Smith, A. M. S. (2005). Characterizing and mapping forest fire fuels using ASTER imagery and gradient modeling. *Forest Ecology and Management*, *217*, 129-146. <https://doi.org/10.1016/j.foreco.2005.06.013>
- Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, *114*, 168-182. <https://doi.org/10.1016/j.rse.2009.08.016>
- Gao, B.C. (1996). NDWI – a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*, 257-266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., ... Chen, J. (2013). Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, *34*, 2607-2654. <https://doi.org/10.1080/01431161.2012.748992>
- Guerschman, J.P., Hill, M.J., Renzullo, L.J., Barrett, D.J., Marks, A.S., & Botha, E.J. (2009). Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sensing of Environment*, *113*, 928-945. <https://doi.org/10.1016/j.rse.2009.01.006>
- Hoan, N. T., Tateishi, R., & Al-Bilbisi, H. (2013). *Global MODIS 500m Data User's Manual*. Retrieved from <http://www.cr.chiba-u.jp/~database-jp/wiki/wiki.cgi?page=GEOinfoDB%5Fglobal>
- Jönsson, P., & Eklundh, L. (2002). Seasonality extraction and noise removal by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, *40*, 1824-1832. <https://doi.org/10.1109/TGRS.2002.802519>
- Jönsson, P., & Eklundh, L. (2004). TIMESAT - a program for analyzing time-series of satellite sensor data.

Computers and Geosciences, 30, 833-845.

- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., ... Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, 51, 933-938.
- Phong, D. X., Nguyen, H. T., Kobayashi, T., & Tateishi, R. (2013). A global 500-m urban map for GLCNMO version 2. *Proceedings of the International Symposium on Remote Sensing (ISRS)* (pp. 666-669). Chiba, Japan.
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Francisco, CA: Morgan Kaufmann Publishers.
- Rogers, A. S., & Kearney, M. S. (2004). Reducing signature variability in unmixing coastal marsh Thematic Mapper scenes using spectral indices. *International Journal of Remote Sensing*, 25, 2317-2335. <https://doi.org/10.1080/01431160310001618103>
- Sharma, R. C., Tateishi, R., Hara, K., & Nguyen, L. V. (2015). Developing superfine water index (SWI) for global water cover mapping using MODIS data. *Remote Sensing*, 7, 13807-13841. <https://doi.org/10.3390/rs71013807>
- Tana, G., & Tateishi, R. (2013). Mapping wetlands in Africa using MODIS tasseled cap indices. *Proceedings of the International Symposium on Remote Sensing (ISRS)* (pp. 654-657). Chiba, Japan.
- Tateishi, R., Bayaer, U., Al-Bilbisi, H., Aboel Ghar, M., Tsend-Ayush, J., Kobayashi, T., ... Sato, H. P. Production of global land cover data – GLCNMO. (2011). *International Journal of Digital Earth*, 4, 22-49. <https://doi.org/10.1080/17538941003777521>
- Tateishi, R., Hoan N.T., Kobayashi, T., Alsaaidh, B., Gegentana, & Phong, D. X. (2014). Production of global land cover data – GLCNMO2008. *Journal of Geography and Geology*, 6, 99-122. <https://doi.org/10.5539/jgg.v6n3p99>
- Xiao, X., Boles, S., Frohling, S., Salas, W., Moore, B., Li, C., ... Zhao, R. (2002). Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data. *Remote Sensing of Environment*, 82, 335-348. <https://doi.org/10.1080/01431160110107734>
- Zhang, N., & Tateishi, R. (2013). Integrated use of existing global land cover datasets for producing a new global land cover dataset with a higher accuracy: a case study in Eurasia. *Advances in Remote Sensing*, 2, 365-372. <https://doi.org/10.4236/ars.2013.24039>

Web References

w1 NOAA/NCEI website

<http://ngdc.noaa.gov/eog/dmsp.html>

w2 USGS/EROS Center website

<http://glovis.usgs.gov/>

w3 JAXA website

http://www.eorc.jaxa.jp/ALOS/palsar_fnf/fnf_jindex20140116.htm

w4 Joint Research Centre website

<http://forobs.jrc.ec.europa.eu/products/glc2000/glc2000.php>

w5 ISCGM website

<http://www.iscgm.org/gmd/download/glcnm.html>

w6 ESA website

http://due.esrin.esa.int/page_globcover.php

w7 CEReS website

http://www.cr.chiba-u.jp/~database-jp/wiki/wiki.cgi?page=GEOinfoDB_global

w8 TIMESAT website

<http://web.nateko.lu.se/personal/Lars.Eklundh/TIMESAT/login.asp>

Appendix A

Collaborating National Geospatial Information Authorities

The following 19 collaborating National Geospatial Information Authorities (NGIAs) collaborated with the authors at the validation process of intermediate classification.

- Algeria: National Institute of Cartography and Remote Sensing
- Australia: Geoscience Australia
- Botswana: Surveys and Mapping, Ministry Lands and Housing
- Brazil: Instituto Brasileiro de Geografia e Estatística
- Burundi: Institut Géographique du Burundi
- Chile: Instituto Geográfico Militar
- Colombia: Instituto Geográfico Agustín Codazzi
- Hong Kong, S.A.R., China: Lands Department, the Government of the Hong Kong Special Administrative Region
- Japan: Geospatial Information Authority of Japan
- Latvia: Latvian Geospatial Information Agency
- Macao, S.A.R., China: Direcção dos Serviços de Cartografia e Cadastro, Governo da Região Administrativa Especial de Macau
- Macedonia: Agency of Real Estate Cadastre (AREC)
- Madagascar: National Geographic and Hydrographic Institute
- Malaysia: Department of Survey and Mapping Malaysia
- Mexico: Instituto Nacional de Estadística y Geografía
- Romania: National Agency for Cadastre and Land Registration of Romania
- Senegal: Agence Nationale de l'Aménagement du Territoire-ANAT
- Sweden: Lantmäteriet – The Swedish Mapping, Cadastre and Land Registration Authority
- Thailand: Royal Thai Survey Department.

Appendix B

Additional data used in the study

The following additional data were used in the study.

1. Ramsar Sites Database: (used for wetland mapping)
The Ramsar Sites Database provides information of all wetlands of international importance. It is a searchable database, fully accessible through the internet with a password protected data entry system, and a reporting system for public use (<http://ramsar.wetlands.org/Database/AbouttheRamsarSitesDatabase/tabid/812/Default.aspx>).
2. Köppen-Geiger climate classification map: (used for wetland mapping)
It is a frequently used climate classification map of Wladimir Köppen, presented by Rudolf Geiger (Kottek et al., 2006) (<http://koeppen-geiger.vu-wien.ac.at/>).
3. GTOPO30: (used for wetland mapping and mangrove mapping)
Global DEM with a horizontal grid spacing of 30 arc seconds (approximately 1 km). The data is available from U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center.
4. U.S. National Wetlands Inventory data: (used for wetland mapping)
The National Wetlands Inventory database, produced by the U.S. Fish and Wildlife Service, provides digital wetland data for the United States (approximately 82% of the conterminous states) (Tiner, 1997). The vector data is available from the Fish and Wildlife Service National Wetlands Inventory website (<http://www.fws.gov/wetlands/>).

5. Canadian Wetland Inventory data: (used for wetland mapping)

The Environment Canada–Canadian Wildlife Service (CWS) produced the Canadian Wetland Inventory to provide digital wetland data for parts of Canada via the website of Duck Unlimited Canada (<http://maps.ducks.ca/cwi/>).

6. Land Cover, circa 2000-Vector (LCC2000-V) data: (used for wetland mapping)

It is the vectorized land cover data originating from classified Landsat 5 and Landsat 7 ortho-images for Canada. The LCC2000-V data were downloaded from the GeoBase website the Canadian Council on Geomatics (CCOG) (<http://www.geobase.ca/>).

7. GDEM: (used for global water mapping)

It is the 1 arc second (30m) DEM data produced by the Ministry of Economy, Trade, and Industry (METI) of Japan and the United States National Aeronautics and Space Administration (NASA). It is available from NASA Reverb (<http://reverb.echo.nasa.gov/reverb/>).

8. LandScan 2012™ population data: (used for urban mapping)

It shows the population distribution at about 1 km resolution (30 seconds). We used LandScan 2012™ High Resolution global Population Data Set copyrighted by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under Contract No. DE-AC05-00OR22725 with the United States Department of Energy. (<http://www.ornl.gov/landscan/>).

9. Global Distribution and Density of Constructed Impervious Surfaces Area (ISA): (used for urban mapping)

This data present the global inventory of the spatial distribution and density on 1 km² grids. ISA include roads, parking lots, buildings, driveways, sidewalks and other manmade surface (http://ngdc.noaa.gov/eog/dmsp/download_global_isa.html).

10. GDP per capita data: (used for urban mapping)

It is the gross domestic product based on purchasing power parity per capita data by the International Monetary Fund (IMF). (<http://www.imf.org/external/pubs/ft/weo/2008/02/weodata/index.aspx>).

Appendix C

Published products by this study

The following products were published.

- GLCNMO2013, land cover of 30 degree by 30 degree areas and global area from ISCGM website (<https://www.iscgm.org/gmd/>)
- GLCNMO2013, land cover of global and continental areas from CEReS website (<http://www.cr.chiba-u.jp/databases/GLP/database-GLP.html>)

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).