Developing and applying a multi-purpose land cover validation dataset for Africa

Tsendbazar, N-E[1*]., Herold, M[1]., de Bruin, S[1]., Lesiv, M[2]., Fritz, S[2]., Van De Kerchove, R[3]., Buchhorn, M[3]., Duerauer, M[2]., Szantoi, Z[4,5]., and Pekel, J-F[4]

[1] Wageningen University & Research, Laboratory of Geo-Information Science and Remote Sensing, Droevendaalsesteeg 3, 6708 PB, Wageningen, the Netherlands

[2] International Institute for Applied Systems Analysis (IIASA), Schlossplatz 1, A-2361 Laxenburg, Austria.

[3] Flemish Institute for Technological Research (VITO), Mol, Belgium

[4] European Commission, Joint Research Centre, Directorate for Sustainable Resources, 21027 Ispra, Italy

[5] Department of Geography and Environmental Studies, Stellenbosch University, Stellenbosch 7602, South Africa

[*] Corresponding author: nandin.tsendbazar@wur.nl

Developing and applying a multi-purpose land cover validation dataset for Africa

Abstract

The production of global land cover products has accelerated significantly over the past decade thanks to the availability of higher spatial and temporal resolution satellite data and increased computation capabilities. The quality of these products should be assessed according to internationally promoted requirements e.g., by the Committee on Earth Observation Systems-Working Group on Calibration and Validation (CEOS-WGCV) and updated accuracy should be provided with new releases (Stage-4 validation). Providing updated accuracies for the yearly maps would require considerable effort for collecting validation datasets. To save time and effort on data collection, validation datasets should be designed to suit multiple map assessments and should be easily adjustable for a timely validation of new releases of land cover products. This study introduces a validation dataset aimed to facilitate multi-purpose assessments and its applicability is demonstrated in three different assessments focusing on validating discrete and fractional land cover maps, map comparison and user-oriented map assessments. The validation dataset is generated primarily to validate the newly released 100m spatial resolution land cover product from the Copernicus Global Land Service (CGLS-LC100). The validation dataset includes 3617 sample sites in Africa based on stratified sampling. Each site corresponds to an area of 100m×100m. Within site, reference land cover information was collected at 100 subpixels of 10m×10m allowing the land cover information to be suitable for different resolution and legends. Firstly, using this dataset, we validated both the discrete and fractional land cover layers of the CGLS-LC100 product. The CGLS-LC100 discrete map was found to have an overall accuracy of 74.6+/-2.1% (at 95% confidence level) for the African continent. Fraction cover products were found to have mean absolute errors of 9.3, 8.8, 16.2, and 6.5% for trees, shrubs, herbaceous vegetation and bare ground, respectively. Secondly, for user-oriented map assessment, we assessed the accuracy of the CGLS-LC100 map from four user groups' perspectives (forest monitoring, crop monitoring, biodiversity and climate modelling). Overall accuracies for these perspectives vary between 73.7% +/-2.1% and 93.5% ±0.9%, depending on the land cover classes of interest. Thirdly, for map comparison, we assessed the accuracy of the Globeland30-2010 map at 30m spatial resolution. Using the subpixel level validation data, we derived 15252 sample pixels at 30m spatial resolution. Based on these sample pixels, the overall accuracy of the Globeland30-2010 map was found to be 66.6 ±2.4% for Africa. The three assessments exemplify the applicability of multi-purpose validation datasets which are recommended to increase map validation efficiency and consistency. Assessments of subsequent yearly maps can be conducted by augmenting or updating the dataset with sample sites in identified change areas.

Keywords: Land cover validation, Validation data, Multi-purpose assessments, Discrete and fractional land cover, Map comparison and User specific accuracies.

1. Introduction

Land cover mapping at continental and global scales provides valuable information on the earth's surface and is used for many applications aiming to understand and to adapt to the changing environment (Verburg et al. 2011). As such, good quality land cover maps are required by multiple institutions, governments and researchers related to climate change, biodiversity and conservation, and zero-hunger efforts (Romijn et al. 2016).

The first satellite-based global land cover map dates back to 1994 (DeFries and Townshend 1994). Over the past decades numerous global land cover maps were produced using medium resolution satellite data (Arino et al. 2007; Bartholomé and Belward 2005; Friedl et al. 2002; Land Cover CCI. 2014; Tateishi et al. 2011). Pioneering the productions of higher resolution land cover mapping at large scale, researchers have created global and continental scale land cover products using Landsat (Chen et al. 2015; Gong et al. 2013; Hansen et al. 2013) and Sentinel-2 data (CCI Land Cover 2017a). Our understanding of the changing environment is further enhanced with the recent land cover change products namely annual LC-CCI land cover maps (CCI Land Cover 2017b), Global Surface Water Explorer (Pekel et al. 2016), Global Human Settlement Layers (Pesaresi et al. 2016) and Global Forest Change datasets (Hansen et al. 2013).

Advancements in land cover mapping at global or continental scales are being made continuously thanks to open access high spatial and temporal resolution remote sensing data and increased processing capabilities such as cloud computing. This is evident in the acceleration of developments of new land cover products over the current decade (Herold et al. 2016) and in the emerging high resolution land cover products generated using cloud computing facilities such as the Google Earth Engine (Gorelick et al. 2017). Complementing the higher resolution (~30m) large scale land cover mapping (e.g., CCI Land Cover (2017a) and Chen et al. (2015)), Copernicus Global Land Service (CGLS) aims to provide an operational global land cover mapping by focusing on yearly mapping from 2015 onwards with flexible thematic detail. The first product was generated for Africa at 100m resolution and it includes discrete (fixed legend) and fractional (vegetation continuous field layers providing estimates of fractions of land cover types: trees, shrubs, herbaceous vegetation and bare soil) maps (Copernicus Global Land Service 2017).

Although, the validation of global land cover products has become a common activity for assessing their quality and usability (Herold et al. 2016), validation activities should adjust to the emergence of new or subsequent products without much additional effort. Most global land cover validation datasets are collected via visual interpretation (Chen et al. 2015; Tsendbazar et al. 2015b; Xiong et al. 2017), a labour intensive task requiring efforts of multiple mapping and image interpretation experts (Defourny et al. 2011; Mayaux et al. 2006; Scepan et al. 1999). To guarantee the independence from the training data and the consistency of the validation results (as well as to save time and effort), such datasets should be designed to be suitable for multiple map assessments and could be re-used, to provide timely quality assessments on the new and subsequent land cover products.

However, most existing validation datasets were generated to validate a single land cover map and their characteristics such as sample site areas and thematic legends are not suitable to be used for validating multiple maps. For example, a validation dataset (with some 150 000 sample locations) for the Globeland30 map (Chen et al. 2015) is limited to assessing other maps having similar resolution as the Globeland30. Similarly, the validation dataset developed for the GlobCover 2009 map (Defourny et al. 2011) is constrained to be used for assessing maps with medium resolution (~300m) (CCI Land Cover 2017b). A recent review of metadata on global land cover validation datasets found that re-using a validation dataset to assess another map usually comes at a cost, namely loss of spatial and thematic detail (Tsendbazar et al. 2015b). This restricts the usage of validation datasets for purposes such as assessing fraction maps, map comparisons and map assessments from different users' perspectives. For example, most validation datasets represent the reference land cover as discrete classes according to fixed legends. Therefore they do not record land cover fraction information (e.g., tree cover fractions). As such their utility for validating land cover maps is limited (Tsendbazar et al. 2015b).

The call for a validation dataset suitable for multiple map validation was initiated by an international community, i.e., the Global Observations of Forest and Land Dynamics (GOFC-GOLD) (Herold et al.

2009). GOFC-GOLD emphasizes the importance of inter-operability and comparability of global land cover maps to help map users select the most suitable maps for their needs (Herold et al. 2008). A statistical comparison of several land cover maps requires a validation dataset that has been acquired by transparent means and that is suitable for multiple map assessments in terms of spatial resolution and thematic legends. For example, the class "forest" can have different definitions (e.g., >30% or >60% forest density)(Jung et al. 2006), thus the validation dataset used for comparison should be able to accommodate such differences. Therefore, GOFC-GOLD and the working group on calibration and validation of the Committee on Earth Observation Satellites (GEOS-WGCV) proposed a multi-purpose validation dataset (Herold et al. 2009) which was further detailed in Olofsson et al. (2012). For improved re-usability, the dataset was designed to be flexible in terms of sample selection, sample unit area and thematic detail (Olofsson et al. 2012). For example, the reference land cover in a sample unit area (5km × 5km) is generated from classifications of very high resolution (2m) images and this makes the dataset suitable for assessing maps with different resolutions up to 5km × 5km. Fractional coverage of land cover types within the sample unit area can also be estimated with this dataset. The initial sample comprised 500 sites and could be increased if required (Stehman et al. 2012). The dataset has been published by the United States Geological Survey (Pengra et al. 2015). However, thematically it only comprises four land cover categories, i.e., trees, water, bare, and other.

Map users may require different thematic classes depending on the purpose of applications using land cover maps (Tsendbazar et al. 2016a). For instance, confusion between bare land and natural grassland may not be important for users who are only interested in cropland areas. The overall map accuracy of cropland/non-cropland areas would be different than the overall accuracy reported by the map producers that report confusion errors for all classes. To report map accuracy from different users' perspective, a validation dataset needs to be compatible with multiple legends. Tsendbazar et al. (2016b) used a re-interpreted version of the GlobeCover-2005 validation dataset for validating and comparing three global land cover maps for 2005 from different users' perspective. Although this dataset's thematic detail is compatible with multiple maps, it is only suitable for validating medium resolution (~300-500m) global land cover maps (Defourny et al. 2011). Pengra et al. (2015) and Tsendbazar et al. (2016b) showed that more efforts are needed to create validation datasets that match different spatial and thematic detail as well as different users' perspectives.

Subsequent releases of land cover products should be provided with updated independent validation reports according to the Stage 4 validation requirements of the CEOS-WGCV (Herold et al. 2009). Most currently available global land cover products do not meet this requirement. Apart from the CCI-2015, which was validated using the GlobCover-2009 validation dataset (CCI Land Cover 2017b), none of the yearly CCI-LC land cover products has been validated. The same applies to the MODIS land cover maps for which only the accuracy of the 2005 map was assessed (Friedl et al. 2010). Validation of new land cover products would benefit from a validation dataset that is updated using less demanding efforts, such as re-interpreting and adding additional sample locations in identified change areas. Stehman et al. (2012) recommended using stratified sampling to facilitate sample augmentation.

In this work, we aim (i) to develop a flexible validation dataset suitable for assessments of multiple land cover maps, and (ii) to illustrate its applicability for multiple-purposes in three different assessments namely validation of discrete and fractional land cover maps, map validations from user's perspectives and validating a different resolution map for a comparison purpose. It builds on an independent validation activity of the CGLS Dynamic Land Cover product (CGLS-LC100) (Tsendbazar et al. 2017). The CGLS-LC100 is a part of a framework for operational implementation of yearly global land cover mapping. We describe the design and production of the CGLS-LC100 land cover validation data for Africa suitable for assessing land cover maps at 10-100m resolution. Applicability of the validation dataset for multiple purposes is demonstrated for three different assessments requiring different accuracy metrics, legends and resolutions. Firstly, we calculated different accuracy metrics appropriate for assessing the discrete versus cover fraction CGLS-LC100 maps of Africa for the reference year of 2015. Secondly, to compare with the CGLS-LC100 accuracy, we used the validation dataset to assess the accuracy of 30 m resolution Globeland30 2010 map for Africa. Lastly, we assessed the accuracy of the CGLS-LC100 from different users' perspectives requiring varying legends. While the current study focuses on validation data at African continental scale, the dataset design can be expanded to global scale which can be used for assessing global land cover maps.

2. Methods and materials

2.1. Validation data collection

2.1.1. Sampling design

A probability sampling scheme was used to allow design-based inference of map accuracies. The sample selection scheme had to be suitable for validating the CGLS-LC100 maps and other land cover maps. Therefore, appropriate choices for sample size, sample selection scheme and sample unit size (spatial support) were considered given constraints imposed by allowable error (Foody 2009; Olofsson et al. 2012).

Considering the efforts required to collect the validation dataset (expert training, interpretation and quality checking: see Section 2.1.2) a sample size of 2700 sites was considered feasible. Such sample size is similar or larger than those used for statistical assessments of large scale land cover maps (Bontemps et al. 2011; Mayaux et al. 2006; Tateishi et al. 2014).

The criterion of statistical probability sampling with known and non-zero inclusion probabilities was followed. Due to its efficiency and ease of accommodating modifications such as an increase in sample size (Olofsson et al. 2012), we used stratified random sampling. We used a global stratification by Olofsson et al. (2012) that is independent from any land cover maps. This stratification is based on Köppen climate zones and human population density following the assumption that current land cover is influenced by climate as natural driver and human disturbances as anthropogenic driver (Olofsson et al. 2012). The stratification according to Olofsson et al. (2012), originally at 5km resolution, was resampled to 100m resolution for this study. For Africa there are 15 strata to which a water stratum was added (Figure 1).

The sample allocation process focused on strata in which some land cover classes that are more likely to be misclassified (Olofsson et al. 2012). Since, the Sahel and dry savannah's heterogeneous landscapes in Africa are known to have lower map accuracies (Tsendbazar et al. 2015a), more sample sites were allocated to these heterogeneous areas and to the populated strata (Figure1). The sample sizes per stratum are listed in Table S1(Supplementary Materials). At each sample site location, reference land cover of an area of 100m × 100m was identified. This support size coincides with the pixel size of the Proba-V satellite data used to generate the CGLS-LC100 land cover products.

To increase the sample representation in rare classes such as wetland and urban, an additional set of sample sites was collected. For this, the minimum required sample size per class was set to 250. If the sample size for a specific mapped class was smaller than 250, additional sample sites were collected to meet the requirement. This additional collection mostly focused on urban, wetland vegetation, water and shrubs areas based on the CGLS-LC100 discrete land cover map. Therefore, the augmented sample sites were selected independently of the initial stratification of Olofsson et al. (2012). For each stratum, sample sites were randomly selected as shown in Figure 1. The obtained sample size amounted to 3617 sites including the initial 2700 sample sites.

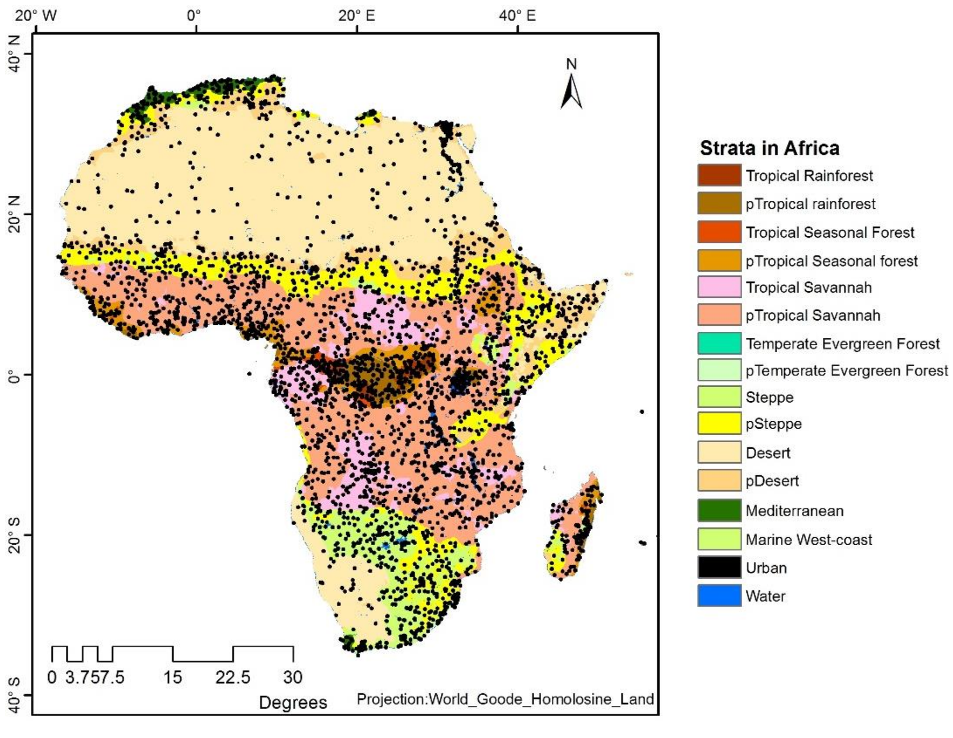

Figure 1: Spatial distribution of all validation sample sites and the stratification by Olofsson et al 2012: 'p' before the strata names denote populated part of climate zone.

2.1.2. Response design

To allow multi-purpose assessments of land cover maps, the spatial and thematic representations of the validation dataset are designed to be compatible for maps with different resolutions and legends. For this, similar to the training data collection used for the CGLS-LC100 product (Lesiv et al. 2016a), each sample site (100m × 100m) was divided into 10×10 small blocks (10m × 10m) and reference land cover was collected at the subpixel level. This makes the validation dataset compatible for assessing maps with 10-100m resolutions. For the thematic representation, we labelled the land cover in terms of generic elements dominating the 10m × 10m subpixels. Land cover elements include trees (different leaf and phenology types), shrubs, grass, crops, built-up areas, bare area, water body, snow &ice and regularly flooded herbaceous area (wetlands). The land cover elements were defined according to the United Nations Land Cover Classification System (UN-LCCS) (Di Gregorio 2005). This allows the validation dataset to be thematically compatible for multiple maps by using different combinations of the land cover elements based on legend definition requirements of multiple maps.

To collect reference land cover data for validation, we have developed a dedicated web-interface through the Geo-Wiki platform (Fritz et al. 2011). The interface provides access to different remote sensing data and allows labelling land cover (Figure 2). The data sources for interpretation include Google and Bing maps as well as Sentinel-2 (Level1C single-date) images with acquisition dates around 2015. Historic time series of NDVI profiles based on MODIS, Landsat and Proba-V data were used for plant phenology identification (Figure 2).

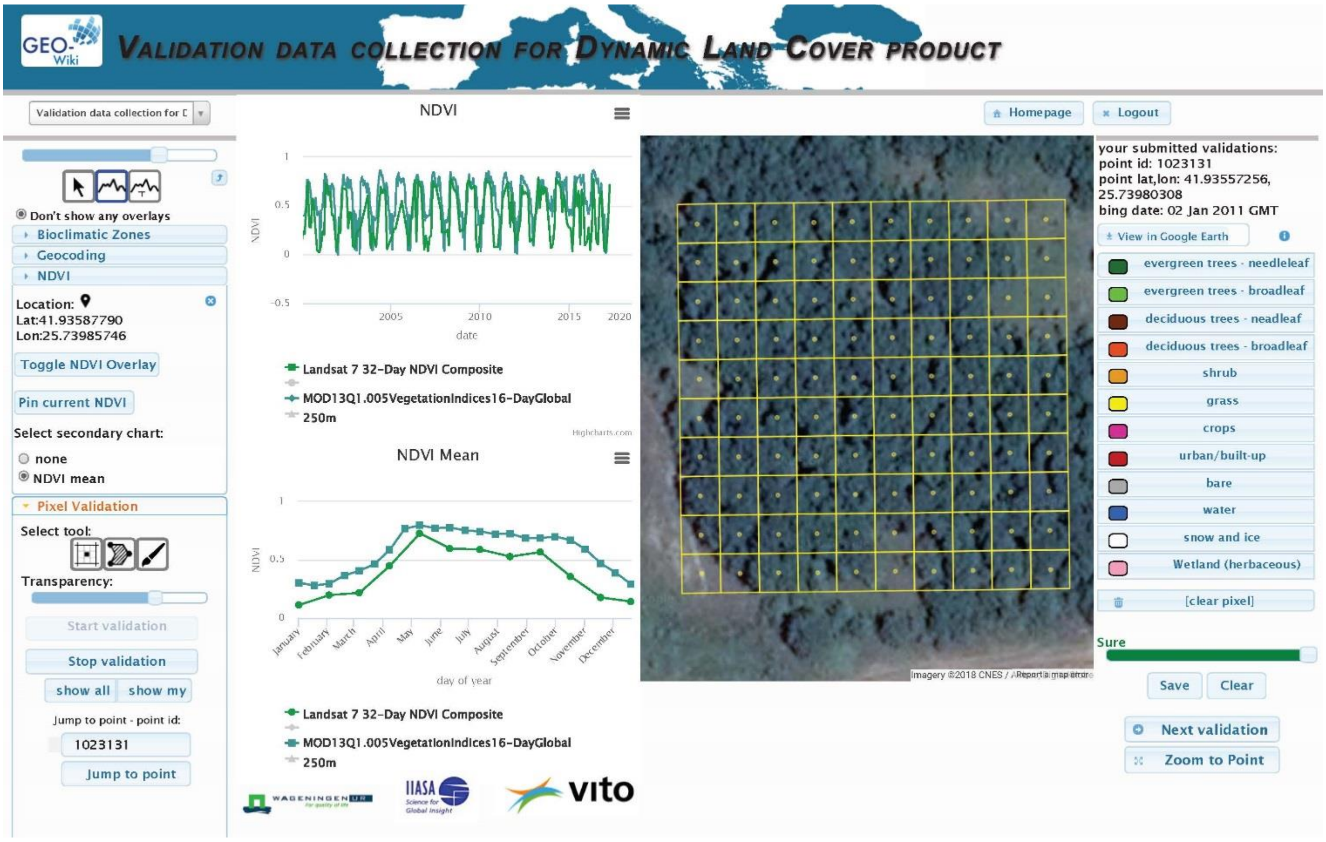

Figure 2: Screen shot of Geo-Wiki based interface for land cover validation

An example of labelling the land cover in a sample site is provided in Figure 3.

Figure 3: A screenshot of an example sample interpretation (green – trees, orange – shrubs, yellow – grassland)

Land cover at each site was visually interpreted by a single expert. In total there were six experts who contributed remotely for different regions in Africa. All experts have experience in satellite based land cover analysis and image interpretation. The GOFC-GOLD regional network was used for recruiting some of the experts. Table 1 provides a list of the regional experts who contributed to data collection. On average, one expert interpreted 80-100 sample sites per day. Overall, validation data collection and quality control took three months. The experts' efforts were financially compensated depending on the work load.

Table 1: Selected regional experts for sample interpretation

	Name	Country	Region	Affiliation
1	Andre Mazinga	DRC	Central and Western Africa	OSFAC, DRC
2	Ifo Suspence	Republic of Congo	Central Africa	Marien Ngouabi University, Brazzaville, République du Congo.
3	Elias Buzayane	Ethiopia	Eastern Africa	HoLiN Training and Consultancy Services PLC
4	Natasha Ribeiro	Mozambique	Southern Africa	Universidade Eduardo Mondlane and MIOMBO and GOFC-GOLD network
5	Matthias Herkt	Germany	Southern and Eastern Africa	Institute of Experimental Ecology, University of Ulm, Germany
6	Emmanuel Amoah Boakye	Ghana	Western Africa	WASCAL, Accra, Ghana

Different quality control measures were applied to obtain a reliable and good quality reference dataset for validation. Firstly, in addition to a tutorial on land cover interpretation, a training workshop was organized for the global land regional land cover mapping experts in January 2017 at IIASA, Laxenburg, Austria. The aim of the workshop was to reduce interpretation discrepancies among the experts. The experts were asked to interpret the same 30 sample sites (100m x 100m) and feedback on any discrepancy was provided upon examination by global land cover mapping experts. The global land cover mapping experts were independent from the CGLS-LC100 product generation. Secondly, depending on the available sources of information (e.g., high resolution images and NDVI profiles) and complexity of landscapes (e.g., small holder cultivation areas), the confidence in the interpretation can be different. Therefore, we recorded the interpretation confidence levels (i.e., unsure, bit sure, quite sure, sure). Three percent of the sample sites were tagged as “unsure” or “bit sure”. Lastly, all the interpretations including these unsure interpretations were checked by global land cover mapping experts and feedback on each interpretation was provided to the experts. The regional experts either rebutted the feedback or corrected their interpretations where necessary.

2.2. Land cover products

To demonstrate applicability of the validation dataset for multiple applications, we selected two land cover maps at different spatial resolutions and different legends: (1) the CGLS-LC100 V1.0 at 100m resolution provided for the 2015 reference year over Africa (Buchhorn et al. 2017); (2) the Globeland30 2010 map (Chen et al. 2015).

The CGLS-LC100 V1.0 at 100m resolution product, provided for the 2015 reference year over Africa (Buchhorn et al. 2017), is a new product in the CGLS portfolio. The CGLS-LC100 is based on the Proba-V 100m data archive (Dierckx et al. 2014), a high quality land cover training dataset (Lesiv et al. 2016a) and several ancillary datasets. More description of the map generation is detailed in Buchhorn et al. (2017). Apart from a discrete land cover type map, the product includes four vegetation continuous field layers providing estimates of fractions (0 - 100%) for the land cover types: trees, shrub, herbaceous vegetation and bare ground.

Table 2 lists the land cover classes and their definitions (Lesiv et al. 2016b).

Table 2: Land cover classes accounted for in CGLS dynamic land cover map

Code	Land cover classes	Definitions according to UN LCCS
11	Closed Forest	Lands dominated by woody plants with a percent cover >70% and height exceeding 5 meters. Exception: a woody plant with a clear physiognomic aspect of trees can be classified as trees even if the height is lower than 5 m but more than 3 m. Depending on the phenology and leaf type, forest can be divided into evergreen, deciduous, needleleaf and broadleaf forests.
12	Open Forest	Lands dominated by woody plants with a percent cover 15-70% and height exceeding 5 meters. Exception: a woody plant with a clear physiognomic aspect of trees can be classified as trees even if the height is lower than 5 m but more than 3 m. Depending on the phenology and leaf type, forest can be divided into evergreen, deciduous, needleleaf and broadleaf forests.
20	Shrubs	These are woody perennial plants with persistent and woody stems and without any defined main stem being less than 5 m tall. The shrub foliage can be either evergreen or deciduous.
30	Herbaceous vegetation	Plants without persistent stem or shoots above ground and lacking definite firm structure. Tree and shrub cover is less than 10%.
40	Cropland	Lands covered with temporary crops followed by harvest and a bare soil period (e.g., single and multiple cropping systems). Note that perennial woody crops will be classified as the appropriate forest or shrub land cover type.
50	Urban/built up	Land covered by buildings and other man-made structures
60	Bare/sparse vegetation	Lands with exposed soil, sand, or rocks and never has more than 10% vegetated cover during any time of the year
70	Snow and Ice	Lands under snow or ice cover throughout the year.
80	Open water	Oceans, seas, lakes, reservoirs, and rivers. Can be either fresh or salt-water bodies.
90	Wetland herbaceous vegetation	Lands that have free water at or on the surface for at least the major part of the growing season. Wetland vegetation include open wetlands, permanent and seasonally flooded wetland herbaceous vegetation. Note that wetland woody vegetation are classified as the appropriate forest or shrub land cover type.

We also assessed the Globeland30 map (Chen et al. 2015) for comparison. The Globeland30 project of China's Ministry of Science and Technology produced global land cover maps for the year 2000 and 2010. The maps were produced at 30m resolution using Landsat TM and ETM+ and the Chinese Environmental Disaster Alleviation Satellite (HJ-1) data. We used the 2010 map for Africa. This map has ten land cover classes of which eight occur in Africa (cultivated land, forest, grassland, shrubland, wetland, water bodies, artificial surfaces and bare land) (Globeland30 2016). The overall map accuracy has been reported to be 79.26% at global level (Chen et al. 2015) but no accuracy information is available for Africa.

2.3. Validation of discrete and fractional land cover map

To assess the discrete CGLS-LC100 map, the land cover elements of 10 × 10 subpixels were summed for each sample site to derive fractions of land cover types per validation site (e.g. 70% trees and 30%

grass = = 70 subpixels trees and 30 subpixels grass). This information was then translated to the CGLS-LC100 discrete legend using the UN-LCCS as a basis. For homogeneous sample sites, land cover fractions were directly converted to land cover classes (e.g., 100% water proportion corresponds to water body class). Approximately 37% of the sample sites were homogeneous (100% covered by a single land cover type). In heterogeneous sample sites where conditions can concurrently meet definitions of multiple land cover types, a priority rule was applied, similar to the CGLS-LC100 training data translation approach (Lesiv et al. 2016a). In such cases, the preferential order was open water, urban, cropland, closed forest, open forest, shrubs, wetland, herbaceous vegetation and bare/sparse vegetation, respectively. In the legend translation, +/- 5% deviations from the legend definition thresholds were allowed. This aimed to consider the geolocation error of Google and Bing Map images which were used for land cover interpretation.

To estimate the accuracy of the land cover maps, we accounted for unequal inclusion probabilities between different strata because sample sites were not allocated proportionally to the strata areas (Olofsson et al. 2012; Wickham et al. 2010). Based on Pengra et al. (2015), the inclusion probability for stratum h is $\pi_h=k_h/K_h$, where k_h is number of sample sites in stratum h and K_h is the population size for stratum h (see Table S1 for inclusion probabilities per stratum). Number of sites is based on the 100m × 100m units. Inclusion probability for the additional sample sites were calculated based on the population of possible sample sites within the rare classes of the CGLS-LC100 map. The estimation weight, the inverse of inclusion probability ($\omega_h=1/\pi_h$), was then calculated and used to construct the confusion matrix accounting for unequal sample inclusion probabilities following the methods described in Stehman et al. (2003) and Wickham et al. (2010). We then estimated the overall and class specific accuracies and their confidence intervals (at 95% confidence level) following Stehman (2014) which specifically addresses estimating map accuracies when the sampling strata are different from the map classes. Thus, by appending three rare class strata to the original stratification, 19 strata were used in the calculations.

Validation data does not contain information on temporary waterbody areas because of limited availability on multiple high resolution images per year for each sample location. Thus, we merged the mapped classes of permanent and temporary waterbody for the accuracy assessment. Owing to the limited sample size for combinations of forest density (closed and open forest) and forest phenology, the accuracy estimation focused on generic classes without taking specific forest phenology into account.

To assess the fraction cover layers, fraction information of the land cover types in the validation dataset was directly used. For each cover fraction layer, the mean absolute error (MAE) and root mean square error (RMSE) were calculated (Foody 1996; Pengra et al. 2015).

$$RMSE_c = \sqrt{\frac{\sum_{i=1}^{n} \omega_i (p_i - v_i)^2}{\sum_{i=1}^{n} \omega_i}} \quad \text{(Eq.1)}$$

where $RMSE_c$ is the root mean squared error of class c, v_i is the reference fraction of class c (in percent), p_i is the mapped fraction of class c, ω_i represents the estimation weight for the sample site and n is the total number of sample sites.

$$MAE_c = \frac{\sum_{i=1}^{n} \omega_i |p_i - v_i|}{\sum_{i=1}^{n} \omega_i} \quad \text{(Eq.2)}$$

where MAE_c is the mean absolute error of class c.

2.4. Accuracy comparison with other datasets at different spatial resolution

For map comparison, the validation dataset should be suitable for the maps being compared in terms of thematic legend and spatial resolution. The CGLS validation dataset can be used to assess land cover maps with 10-100m resolutions. Information on generic land cover elements of this dataset also makes it suitable for maps with different legends. To compare the accuracy of the CGLS-LC100 discrete map, the validation dataset was used to assess the accuracy of the Globeland30-2010 map (Chen et al. 2015). This map was selected because its pixel size is smaller than the spatial support of the CGLS-LC100 validation dataset.

To make the validation dataset compatible with 30m resolution Globeland30 map, we extracted pixel values of the Globeland30 map over each subpixel area (10×10m) of the validation dataset. Using the subpixel centroid locations, we selected Globeland30 pixels that spatially overlap with the subpixels of the validation dataset (at least nine subpixel centre points of the validation dataset). The reference land covers over nine subpixels were aggregated to derive reference land cover for 30m pixels. For homogeneous areas, the land cover elements were directly converted to land cover classes. In heterogeneous areas that can have multiple possible land cover types, we used the dominant land cover type as reference land cover. Sample pixels which did not have a clear dominance (e.g., four sub-pixels of trees, four sub-pixels of shrubs and one sub-pixel of water), totalling to 1037 cases, were excluded from the assessment. A total of 15252 sample pixels were available at 30m resolution.

Next, the Globeland30 map was evaluated using a stratified one-stage cluster approach (Pengra et al. 2015) because multiple 30m sample pixels within the 100m × 100m sites were used for the assessment. Calculation of inclusion probabilities, accuracy estimates and confidence intervals followed the stratified one-stage cluster approach described in Pengra et al. (2015) and Stehman et al. (2003).

2.5. Map validation from different users’ perspectives

We assessed the accuracy of the CGLS-LC100 product from the perspective of four user groups (forest monitoring, crop monitoring, biodiversity and climate modelling). User requirements in terms of map accuracy, spatial and thematic details were defined for the CGLS-LC100 product by the European Commission’s Copernicus Global Land Monitoring program (Lesiv et al. 2016b). We adopted these requirement specifications and derived lists of land cover classes that were deemed to be of interest to the user groups.

<u>Forest monitoring</u>

Researchers and analysts engaged in forest monitoring need information on forest land cover classes. These include closed forests, mixed forests or mosaics of forests with other land cover types, for example, landscapes that are common in Savannah regions in Africa.

The current legend of the CGLS-LC100 discrete map includes closed forests (>70% tree cover) and open forests (15-70% tree cover) classes. A tree cover mosaic class (30 – 70% tree cover) is also widely used in forest monitoring applications (e.g., TREES3 dataset) (Achard et al. 2002; Mayaux et al. 2013), We used the tree cover fraction layer of the CGLS-LC100 product to separate the open forests class in the discrete map into two different classes (tree cover mosaic (30-70% tree cover) and open tree cover mosaic (15-30% tree cover)). Figure 4a depicts a map with seven forest-related classes differing in terms of phenology and tree cover densities based on the CGLS-LC100 discrete LC map and tree cover fraction layer.

A similar procedure as specified in Section 2.3, was followed to translate the reference data and to assess the accuracy.

Crop monitoring

Cropland/non-cropland masks are useful for crop monitoring applications. We created a cropland mask based on the 'cropland class' of the CGLS-LC100 discrete map and assessed its accuracy from crop monitoring perspective (Figure 4b). Area estimates of this class were also calculated for the whole of Africa.

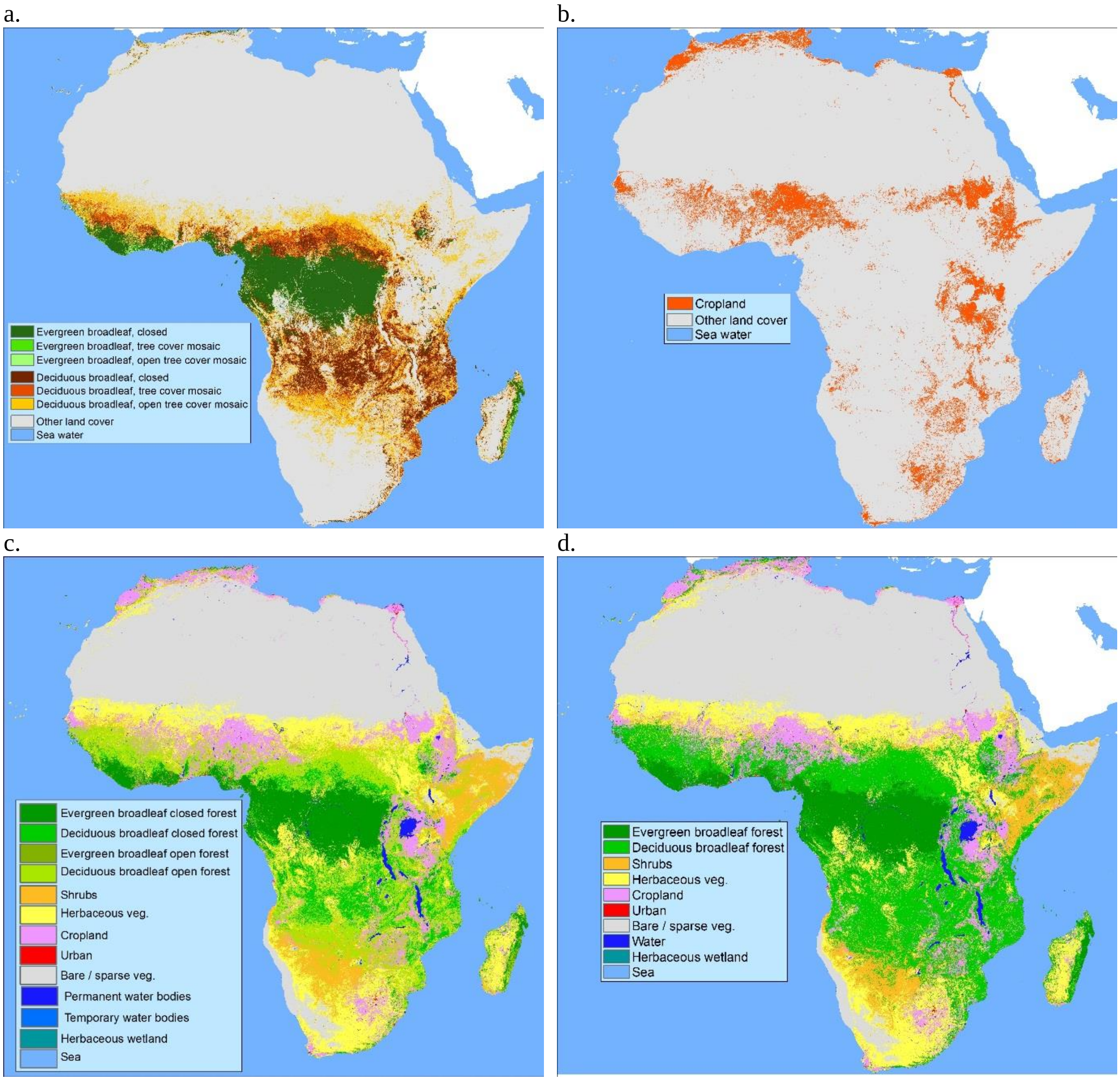

Figure 4. User specific maps based on the CGLS-LC100 products for (a) forest monitoring, (b) crop monitoring, (c) biodiversity and (d) climate modelling

Biodiversity

Land cover maps provide base information for many studies involving biodiversity and conservation (Tuanmu and Jetz 2014). In addition to land cover classes referred to in Section 2.3, we considered different forest type classes as useful classes for biodiversity assessments. Similar to Section 2.3, the temporary waterbody class was merged with the permanent waterbody class. Figure 4c depicts the CGLS-LC100 map with eleven classes that were deemed useful for biodiversity assessments.

Climate modelling

According to the user requirements of the CGLS-LC100 product, the savannah class that is similar to the open forest class is not distinctive for climate modelling purposes (Lesiv et al. 2016b). Thus, open forest was merged with closed forest while only evergreen and deciduous forest types were separated (Figure 4d). Similar to Section 2.3, the temporary waterbody class was merged with the permanent waterbody class.

3. Results

3.1. Validation of discrete and fractional land cover maps

The CGLS-LC100 V1 product (the discrete map and four fraction layers) was assessed using the validation dataset described in Section 2.1. The count-based confusion matrix before correcting for un-equal inclusion probabilities is provided in Table S2 (Supplementary Materials).

The estimated confusion matrix incorporating unequal inclusion probabilities is shown in Table 3.

Overall map accuracy of the CGLS-LC100 discrete map amounts to 74.6% ±2.1% (confidence interval at 95% confidence level)(Table 3).

Table 3: Confusion matrix for the discrete CGLS-LC100 map for Africa, expressed in percentages.

		Reference class												
		Closed forest	Open forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/Sparse veg.	Water	Wetland	**Sample count**	Total	**User's accuracy**	Confidence interval +/-
Mapped class	Closed forest	**11.89**	1.96	0.24	0.13	0.13			0.03	0.15	730	14.5	**81.8**	3.6
	Open forest	1.68	**11.04**	1.49	1.54	1.19		0.02	0.02	0.58	584	17.6	**62.9**	4.3
	Shrubs	0.07	2.19	**5.90**	0.92	0.43	0.03	0.25	0.00	0.09	253	9.9	**59.7**	9.0
	Herbaceous veg.	0.23	2.07	2.00	**10.92**	0.87	0.04	0.70	0.07	0.25	517	17.1	**63.7**	6.3
	Croplands	0.05	1.18	0.59	1.39	**5.48**	0.00	0.07	0.35	0.10	412	9.2	**59.4**	6.5
	Urban		0.03	0.00	0.03	0.00	**0.17**	0.00	0.00		250	0.2	**70.4**	5.7
	Bare/Sparse veg.		0.02	0.39	1.27	0.15		**28.29**	0.28		309	30.4	**93.1**	3.2
	Water		0.01	0.01	0.01	0.01		0.00	**0.87**	0.03	312	0.9	**93.3**	2.8
	Wetland		0.00	0.00	0.00	0.00	0.00		0.01	**0.07**	250	0.1	**78.0**	5.1
Sample count		695	645	292	554	383	180	284	343	241	**3617**			
Total		13.9	18.5	10.6	16.2	8.3	0.3	29.3	1.6	1.3		100		
Producer's accuracy		**85.4**	**59.7**	**55.6**	**67.4**	**66.3**	**68.8**	**96.4**	**53.2**	**5.3**			**74.6**	**2.1**
Confidence interval +/-		3.4	4.9	8.4	5.8	6.2	29.4	2.5	20.0	1.7				

The closed forest and bare/sparse vegetation classes are mapped with relatively high accuracy while the accuracies for open forest, herbaceous vegetation and cropland classes are relatively low. Among the natural vegetation classes, shrubs have the lowest accuracy. The producer's accuracy of the wetland class is particularly low. Substantial wetland areas are omitted in the CGLS-LC100 map since they are confused with the open forest and herbaceous vegetation classes (Table 3).

Table 4 lists the MAE and RMSE for the fraction cover maps.

Table 4. Accuracy of the cover fraction layers expressed in percentages.

	Mean absolute error (MAE)	Root mean square error (RMSE)
Tree fraction	9.32	16.75
Shrub fraction	8.83	15.09
Herbaceous vegetation fraction	16.21	24.84
Bare fraction	6.56	14.85

The bare area fraction map has the lowest error with a MAE of 6.5% and a RMSE of 14.8% while the herbaceous vegetation fraction has the highest error with a MAE of 16.2% and a RMSE of 24.8%.

Upon visual inspection, the deviation from the validation dataset tends to be higher in regions bordering The Sahara desert, The Congo basin and The Horn of Africa.

3.2. Accuracy comparison with other datasets at different spatial resolution

Based on the 15 252 sample pixels, the overall accuracy of the Globeland30 2010 for Africa was assessed at 66.6% ±2.4 % (at 95% confidence level) (Table 5).

Table 5: Confusion matrix for the Globeland30 2010 for Africa, expressed in percentages

		Reference class								**Sample count**	Total	**User's accuracy**	Confidence interval +/-
		Cultivated areas	Forest	Grassland	Shrubland	Wetland	Water bodies	Artificial surfaces	Bareland				
Mapped class	Cultivated areas	**3.84**	0.39	1.45	0.24	0.07	0.09	0.04	0.11	1408	6.23	**61.6**	6.6
	Forest	0.61	**13.20**	2.33	0.86	0.31	0.02	0.00	0.02	3491	17.35	**76.1**	3.2
	Grassland	2.22	5.01	**16.68**	5.16	0.62	0.05	0.05	2.83	4567	32.62	**51.1**	3.8
	Shrubland	0.31	1.05	2.97	**1.48**	0.26	0.02	0.00	1.3	1114	7.40	**20.0**	5.8
	Wetland	0.008	0.913	0.25	0.009	**0.39**	0.155	0	0.025	940	1.75	**22.5**	9.6
	Water bodies	0.004	0.08	0.001	0.00	0.04	**1.39**	0	0.02	1673	1.54	**90.3**	5.5
	Artificial surfaces	0.024	0.10	0.12	0.00	0	0.001	**0.17**	0.251	712	0.66	**25.9**	12.9
	Bareland	0.16	0.006	2.12	0.067	0.039	0.59	0.06	**29.41**	1347	32.45	**90.6**	4.4
Sample count		1453	4040	3693	942	1212	1739	534	1639	**15252**			
Total		7.18	20.75	25.91	7.83	1.73	2.31	0.32	33.96		**100**		
Producer's accuracy		**53.4**	**63.6**	**64.4**	**18.9**	**22.8**	**60.2**	**53.1**	**86.6**			**66.6**	2.4
Confidence interval +/-		6.5	3.5	4.5	6.1	10.8	24.9	23.5	4				

Bareland has relatively high class accuracy, followed by the forest class. The forest class is greatly confused with the grassland class and Globeland30 tends to map substantial forested areas as grasslands (Table 5). Cultivated areas and shrubland are also under-estimated due to over-estimation of grasslands. The shrubland and wetland class have the lowest accuracies compared to other classes.

The count-based confusion matrix for the Globeland30 map can be found in Table S3 (Supplementary Materials).

3.3. Map validation from different users' perspectives

The accuracy of the CGLS-LC100 map from different user's perspective is summarized in Table 6. The detailed confusion matrices are provided in Table S4-S7.

Overall map accuracy for forest monitoring was estimated at 81.3% ± 1.4% (Table 6). The confusion matrix and class specific accuracies show that closed forests types (evergreen broadleaf and deciduous broadleaf) are mapped with higher accuracy (Table S4). Closed evergreen broadleaf forest is mapped with good accuracy (>90%). The accuracy of the tree cover mosaic and the open tree cover mosaic classes are low.

The overall accuracy of the cropland mask was found to be 93.5 ±0.9% % (Table 6). The class specific accuracies of the cropland class are 59.4 % and 66.3% for user's and producer's accuracy respectively (Table S5).

Table 6 : A summary of the considered land cover classes and their accuracies for the users

User groups	User specific maps and remarks	Overall accuracy (area adjusted) / Estimate with 95% confidence intervals
General user (producer)	Discrete land cover map with 9 general classes	74.6% ±2.1%
Forest monitoring	A map with 6 forest related classes (Figure 4a)	81.3% ± 1.4%
Crop monitoring	Cropland and non-cropland mask (Figure 4b)	93.5 ±0.9% Cropland class: User's accuracy: 59.4 ±6.5 % Producer's accuracy: 66.3 ±6.2%
Biodiversity	Discrete land cover map with 11 classes (Figure 4c)	73.7 % ± 2.1%
Climate Modelling	Discrete land cover map with 9 classes (Figure 4d) Fractional land cover maps for trees, shrubs, herbaceous vegetation and bare areas	77.3% ± 2.1% MAE: 9, 8.8, 16, and 6.5%, respectively RMSE: 16.7, 15, 24.8, and 14.8% , respectively

The overall accuracy was assessed at 73.7 % ± 2.1% for biodiversity related use. The class accuracies and the confusion matrix are provided in Table S6. The class accuracies are similar to those presented in Table 3. The producer's accuracy of the open forest, evergreen broadleaf class is low since this class is mostly confused with closed forest evergreen broadleaf and open forest deciduous broadleaf classes.

For climate modelling users, the map overall accuracy was determined to be 77.3% ± 2.1% (Table 6). The class-specific accuracies and the confusion matrix can be found in Table S7. For the evergreen broadleaf forest class, the user's and producer's accuracies are 95% and 89.6% respectively. This class appears to be slightly under-represented. The deciduous broadleaf forest is slightly over-represented with users and producer's accuracy of 72.9% and 74% respectively. In addition to the accuracy of the discrete map from the climate modelling perspective, the accuracy of the cover fraction layers provided in Table 6 can be important as climate modellers are often interested in land cover information related to plant functional types and fraction information on the main land cover types are very useful towards this.

4. Discussion

4.1. The multi-purpose validation dataset development and use

We designed and developed a protocol and validation dataset for independent and multi-purpose assessments of land cover products, and we applied it to different land cover maps (discrete and fractional) of Africa. Particularly, the dataset can address multi-purpose assessments of land cover maps namely (1) validating discrete and fractional land cover maps, (2) map comparability, (3) user oriented accuracy reporting, and (4) updated validation of subsequent land products and cost effectiveness for data collections (Defourny et al. 2011; Herold et al. 2008; Mayaux et al. 2006; Tsendbazar et al. 2016b). The results obtained in this study exemplify the first three purposes mentioned above. The last purpose, updated validation of subsequent land products was not specifically demonstrated in this study. However, the current design of the dataset should be suitable for this purpose as explained in this section.

Recording the reference land cover information at 10×10m sub-pixel level facilitated the following:

(i) To extract class fraction information within the sample site areas;

(ii) To collect information on the land cover elements such as trees and buildings to be used for different legends; and

(iii) To validate land cover maps at finer resolution (e.g. at Sentinel-2 and Landsat scale)

These characteristics make this dataset suitable for multiple map validations requiring different legends, resolutions and requiring different accuracy metrics.

A design of multi-objective accuracy assessment was previously introduced for National Land Cover Data of the United States of America (Stehman et al. 2008). This design addresses different aims of accuracy assessments such as class-specific accuracies, land cover proportion accuracies and net change detection accuracy. This design is limited to one map with a fixed legend and resolution and it is for the extent of the United States of America. The CGLS validation dataset is produced for the African continent and the proposed approach can be expanded to global scale applications thanks to the global stratification derived from Köppen climate zones and population density (Olofsson et al. 2012). The current setup for data collection in the African continent (Section 2.1) can be replicated to other continents to collect validation dataset at global scale. If the similar numbers of sample sites were collected for the five other continents, the total sample size would be larger than 20 000. A stratification independent from the target land cover maps allows collecting the validation data while the target map is being produced, thus reducing the lag between map production and its accuracy assessment. Regardless of the stratification chosen, the accuracy estimates will be unbiased for the true accuracy of each map. However, the precision of the accuracy estimates computed from a stratification independent of the target map will be lower than if that map itself would be used for stratification.

Thanks to the flexibility of the stratified sampling, the number of sample sites could also be increased if required (Stehman et al. 2012). Increasingly, this characteristic is important to provide timely and updated validation of subsequent land cover products following the requirements of the CEOS-WGCV State 4 validation. For subsequent maps, temporary sets of sample sites can be added to the original (permanent) sample to better represent modified or change recorded areas. A potential strategy would be to re-interpret only part of the permanent sample sites rather than all of them assuming no changes occurred in the sites not re-interpreted. The statistical implications of these adjustments need to be further addressed.

4.2. Validation of the discrete and fractional land cover maps

We assessed the accuracy of the CGLS-LC100 discrete and fractional land cover maps using the validation dataset described in Section 2.1. The overall accuracy of the discrete map was found to be 74.6% ±2.1%. This overall accuracy is comparable with the reported accuracy for the CCI-LC-2015 map at global scale (75.3% using only homogeneous sample sites) (Land Cover CCI. 2017). At the African continental scale, Tsendbazar et al. (2015a) found overall correspondences of 50-63% for four global land cover maps (GlobCover 2009, CCI-LC 2010, MODIS-2010 and Globeland 2010). Similarly, the overall accuracy of the Globeland30-2010 map obtained in the current study was assessed at 66.6 ±2.4% for Africa (Table 4). These results suggest that the CGLS-LC100 discrete map has higher overall accuracy compared to Globeland30 map (Table 5) and other land cover maps for Africa (Tsendbazar et al. 2015a).

Closed forest and bare/sparse vegetation classes have higher class specific accuracies followed by the open forest, herbaceous vegetation and cropland classes (Table 3). Among the natural vegetation classes, shrubs are mapped with the lowest accuracies owing to high confusion with open-forest and herbaceous vegetation classes. Confusion between open forests, herbaceous vegetation and shrubs is a known problem for land cover mapping in savannah ecosystems where different vegetation layers (woody and herbaceous vegetation) co-exist (Huttich et al. 2011; Jung et al. 2006). The CGLS-LC100 map slightly over-represents the bare/sparse vegetation class at the cost of herbaceous and shrubs areas, particularly in border regions of the Sahara and Namib deserts (Table 3). The cropland class is confused with open forest and herbaceous vegetation (Table 3). This can be attributed to the difficulty of separating cropland from herbaceous vegetation, and small-scale cultivation in heterogeneous landscapes (Xiong et al. 2017).

The producer's accuracy of water and wetland classes are low, although 85% and 81% of the corresponding validation sites showed agreement in the count-based confusion matrix (Table S2). The confusion was mostly with herbaceous vegetation, croplands, open forest and bare sparse vegetation. The very low producer's accuracy of the wetland class indicates omission of wetland areas in the CGLS-LC100 map. The main wetland regions such as Okavango Delta in Botswana, and the Sudd in South Sudan are under-represented in this map. Therefore, further improvements are needed particularly for mapping the wetland and shrubs classes.

Among the land cover fraction maps, bare area has the lowest errors (MAE 6.5% and RMSE of 14.8%), while, herbaceous vegetation has the highest errors. This can be attributed to the difficulty of separating herbaceous vegetation from other land cover types. This is confirmed by Table 3 where the herbaceous vegetation class is mostly confused with other classes. For the tree cover fraction map, there is no direct comparison available for Africa. However, compared to reported errors in other regions, the tree cover fraction of the CGLS-LC100 has similar or slightly lower errors (MAE 9.3% and RMSE 16.7%). For example, in South-America, the Landsat based tree cover 2010 product by Hansen et al. (2013) was found to have a MAE of 9.39% (Pengra et al. 2015). A Landsat based rescaled version of the MODIS Vegetation Continuous Field percent tree cover product was reported to have 17% RMSE when compared against LiDAR measurements of four regions in North America (Sexton et al. 2013).

In contrast to discrete land cover maps whose accuracies are often reported using overall and class accuracies calculated using confusion matrices (Mayaux et al. 2006; Olofsson et al. 2014), cover fraction maps (e.g., trees, shrubs and herbaceous vegetation) are assessed in terms of the deviation from the reference fraction commonly represented by mean error, MAE and RMSE. Since most validation datasets represent the reference land cover as discrete classes (Tsendbazar et al. 2015b), these datasets cannot be used for assessing cover fraction layers unless the cover fraction layers are hardened (applying a threshold to create discrete classes). Recording reference land cover at higher resolution (e.g., 10m) allowed estimating fraction of main land cover types, thus making this validation dataset suitable for assessing cover fraction layers. This way of collecting reference information could complement substantially the validation datasets created by classification of very high resolution images (Pengra et al. 2015) and LiDAR based measurements in limited locations (Sexton et al. 2013) referred in the previous paragraph. In addition to the four land cover types assessed in this study, other thematic cover fraction layers could also be assessed.

4.3. Accuracy comparison with other datasets at different spatial resolution

We also assessed the 30m resolution Globeland30 2010 map for Africa using our validation dataset. This demonstrates the suitability of our validation dataset for assessing a higher resolution map for comparison. Based on 15252 sample pixels derived from our validation data, the overall accuracy of the Globeland30 was estimated at 66.6 ±2.4% (Table 5). This accuracy is lower than the accuracy reported by the map producers (79.26% ± 0.2%) (Chen et al. 2015). However the accuracy by Chen et al. (2015) is for the entire globe and while the results obtained in this study are for Africa, a continent that tends to have lower map accuracy than other continents (Tsendbazar et al. 2016a). We used the dominant land cover type for validation, because details in the legend definition of some classes were not clear for this map. However, if more detailed information on the legend thresholds is made available, validation could also be done based on the legend definition of the Globeland30 map. There is a 5-year difference in the reference year of the Globeland30 2010 map and this might have an influence on the lower overall accuracy for this map. The assessment of this map serves here to demonstrate our validation data applicability for maps having different spatial resolutions. It should be noted that temporal discrepancies between validation data and maps to be assessed should be kept at a minimum.

Since the CGLS-LC100 validation dataset has reference land cover information at 10m×10m subpixels for a spatial support of 100m×100m, the validation dataset can be used for assessing and comparing the accuracy of maps at 10-100m resolutions, including 10-20m resolution Sentinel-2 based land cover maps. Recently, Lesiv et al. (2017) assessed the prototype version of the Sentinel-2 based CCI20 African land cover map (CCI Land Cover 2017a) using the CGLS-LC100 training and validation datasets. In this case, the reference land cover corresponding to the Sentinel-2 20m×20m pixels was extracted based on 2×2 subpixels of the validation dataset. This further emphasizes the applicability of our validation dataset for higher resolution map assessments. Using the validation data for higher spatial resolution maps increases the number of sample sites (e.g., 15252 for 30m resolution Globeland30). However, in this situation, the accuracy statistics should be estimated using cluster sampling equations (Pengra et al. 2015; Stehman et al. 2003) since otherwise standard errors would be underestimated.

4.4. Map validation from different users' perspectives

The land cover fraction information of the validation dataset allowed the assessment of the CGLS-LC100 product from different users' perspectives. We created four maps with different legends (Figure 4) reflecting users' preferences on different land cover types (Lesiv et al. 2016b), and our results showed varying overall accuracies (73.7% ±2.1% for biodiversity to 93.5±0.9 for crop monitoring) (Table 6). Differences can partly be attributed to the number of land cover classes considered in the assessments but also the class combinations used matter. Accuracy for the biodiversity users was lower due to the number of classes used in the assessment. The higher overall accuracy for crop monitoring (cropland and non-cropland map) was to be expected since internal confusions among non-cropland classes are discarded in this assessment. Thus, the cropland class accuracies are also important map quality measures in this case. The overall accuracy of the CGLS-LC100 cropland mask is similar, but the cropland class accuracies are lower compared to the Landsat based nominal cropland mask (Xiong et al. 2017). In the forest monitoring applications, a map with more forest classes was created by combining the discrete and tree cover fraction map of the CGLS-LC100 product. This further illustrates the suitability of the CGLS-LC100 product towards creating user-tuned maps. Note that the users of land cover maps are not restricted to the user groups identified in this study and the overall map accuracies will differ for different applications, i.e., different classes of interest are considered.

4.5. Lessons learnt on the multi-purpose validation data development

Although the validation dataset was successfully developed and utilized for assessing multiple land cover maps, the dataset also has some limitations to be fully compatible for other map assessments. Sample stratification is focused on the heterogeneity of landscapes (natural and human influenced) (Olofsson et al. 2012) and it was not specifically designed to validate changes in the land cover, which may be a prominent issue when the aim is to estimate change areas for each land cover type. For this, additional strata (e.g., change areas for the corresponding period) need to be added to better represent changed areas and the inclusion probability of the augmented sample sites need to be calculated accordingly (Stehman et al. 2012).

Furthermore, the CGLS-LC100 validation dataset is based on the Proba-V grid at 100m and this could be problematic for validating another map at 100m resolution in which the pixel alignment (grid) may mismatch the Proba-V one. In contrast, for higher resolution maps e.g., 20-30m, the full coverage of the reference land cover over the target pixel can be calculated and used as reference. For 10m resolution map assessments, geolocation errors may have a bigger impact. To reduce impact of such errors, assessment units of 2×2 or 3×3 pixels can be used after resampling the map to 20m or 30m resolution. This also implies accuracy is evaluated at 20m or 30m resolution rather than at the original 10m. Such approach has been used by Mayaux et al. (2006) and Land Cover CCI. (2014).

To support the use of this validation data for other map assessments, future work can focus on developing a service to provide instantly validated user-tuned land cover maps of the CGLS-LC100 products. The validation results could also be provided when other land cover maps are uploaded to the service. This ensures the validation dataset is used for validation purpose rather than training or calibration purposes.

5. Summary

This study designed and developed a multi-purpose validation dataset that aims to be applicable for multiple map assessments. The dataset was developed as part of an independent assessment of the CGLS-LC100 land cover product for Africa. We demonstrated the applicability of the validation dataset for multi-purpose assessments requiring different legends and spatial resolution and requiring different accuracy metrics.

We collected a validation dataset consisting of 3 617 sample sites for Africa using a global stratification independent from any land cover map. Reference land cover of the sample sites (100m × 100m area) was recorded at 10m × 10m subpixels by visual interpretation on a dedicated branch of the Geo-Wiki platform with contributions from several regional experts from Africa. Several quality measures were applied to ensure data quality. The response design of this validation dataset facilitates flexibility towards multi-purpose applications. For example, the ability to assess maps of different resolution (10-100m) is gained by subpixel level reference land cover information. The validation data also supports assessment of maps with different legends. As opposed to creating legend categories by merging certain classes, subpixel level reference land cover data allow specifically targeting classes defined by user-specific composition thresholds. Furthermore, the stratified sampling scheme enables sample augmentation for classes of interest (Stehman et al. 2012).

The applicability of the validation dataset was demonstrated for (1) validation of discrete and fractional land cover maps (CGLS-LC100 product: overall accuracy 74.6% ±2.1% for the discrete and MAE 6.15%-16% for the fraction cover layers); (2) map comparison (Globeland30-2010 map: overall accuracy 66.6 ±2.4%); and (3) user oriented accuracy reporting (CGLS-LC100 product users: overall accuracy: 73.7% +/-2.1% to 81.3% +/- 1.4%).

In addition, the validation dataset is compatible with CGLS's focus on operational monitoring of land cover with yearly releases of global maps. The global stratification used in the sampling facilitates expanding to global scale by replicating the current setup for data collection in African continent for other continents to collect validation dataset at global scale, with additional resources and expert involvements. The flexibility of stratified sampling allows augmenting the validation dataset for validating subsequent new maps to meet the updated validation requirement of the CEOS-WGCV Stage 4 validation. For the latter purpose, sampling needs to be densified in change area by re-interpreting additional sample sites. On the contrary, validation sites in no-change areas can be re-used with little effort by re-interpreting only a part of sample sites in no-change areas.

Although the validation dataset was demonstrated to be suitable for multiple purposes of land cover map assessments, there are remaining aspects that require further attention. The validation dataset was not specifically designed to validate changes in the land cover. Thus, if the aim is to estimate change areas for each land cover type, additional strata (e.g., likely change areas) need to be added to better represent those areas. Furthermore, as validation data collection is a collective work, significant effort is needed to maintain a dataset up-to-date. Therefore, to maintain a full utility of a validation dataset, the importance of updating should be recognized. To better understand the importance of the experts' contribution, interpretation variability and its cause are currently being investigated in separate study.

Finally, to provide timely assessments of new and yearly global land cover products, map producers are encouraged to improve the efficiency of validation datasets given the available resources. In this respect, the proposed design of the validation dataset can serve as a basis to improve upon.

Acknowledgments

This work was supported by the European Commission - Copernicus programme. The authors thank the regional experts from Africa; Andre Mazinga, Ifo Suspence, Elias Buzayane, Natasha Ribeiro, Matthias Herkt, and Emmanuel Amoah Boakye for their contributions in collecting the validation dataset. We thank the anonymous reviewers for their constructive comments that helped us improve the manuscript.

6. References

Achard, F., Eva, H.D., Stibig, H.J., Mayaux, P., Gallego, J., Richards, T., & Malingreau, J.P. (2002). Determination of deforestation rates of the world's humid tropical forests. *Science, 297*, 999-1002

Arino, O., Leroy, M., Ranera, F., Gross, D., Bicheron, P., Nino, F., Brockman, C., Defourny, P., Vancutsem, C., & Achard, F. (2007). GLOBCOVER-A Global Land Cover Service with MERIS. In, *Envisat Symposium 2007* (pp. 23-27). Montreux, Switzerland

Bartholomé, E., & Belward, A. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing, 26*, 1959-1977

Bontemps, S., Defourny, P., Van Bogaert, E., Kalogirou, V., & Arino, O. (2011). *GLOBCOVER 2009: Products Description and Validation Report*: UCLouvain and ESA

Buchhorn, M., Bertels, L., Smets, B., Lesiv, M., & Tsendbazar, N. (2017). *Copernicus Global Land Operations "Vegetation and Energy": Algorithm Theoretical Basis Document for Moderate Dynamic Land Cover*.^https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_ATBD_LC100m-V1_I1.00.pdf

CCI Land Cover. (2017a). *CCI LAND COVER - S2 prototype Land Cover 20m map of Africa 2016* http://2016africalandcover20m.esrin.esa.int/viewer.php

CCI Land Cover. (2017b). *Release of a 1992-2015 time series of annual global land cover maps at 300 m*.^https://www.esa-landcover-cci.org/index.php?q=webfm_send/88

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., & Lu, M. (2015). Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing, 103*, 7-27

Copernicus Global Land Service. (2017). *Copernicus Global Land Service- Land Cover*: VITO.2018,March 12,https://land.copernicus.eu/global/products/lc

Defourny, P., Bontemps, S., Schouten, L., Bartalev, S., Cacetta, P., de Wit, A., di Bella, C., Gérard, B., Giri, C., Gond, V., Hazeu, G., Heinimann, A., Herold, M., Jaffrain, G., Latifovic, R., Lin, H., Mayaux, P., Mücher, S., Nonguierma, A., Stibig, H., Y. Shimabakuro, Van Bogaert, E., Vancutsem, C., Bicheron, P., Leroy, M., & Arino, O. (2011). GLOBCOVER 2005 and GLOBCOVER 2009 validation : learnt lessons In, *GOFC-GOLD Global Land Cover & Change Validation Workshop*. Laxenburg, Austria

DeFries, R., & Townshend, J. (1994). NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing, 15*, 3567-3586

Di Gregorio, A. (2005). *Land cover classification system: classification concepts and user manual: LCCS*. Rome, Italy: Food and Agriculture Organization of the United Nations

Dierckx, W., Sterckx, S., Benhadj, I., Livens, S., Duhoux, G., Van Achteren, T., Francois, M., Mellab, K., & Saint, G. (2014). PROBA-V mission for global vegetation monitoring: standard products and image quality. *International Journal of Remote Sensing, 35*, 2589-2614

Foody, G.M. (1996). Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing, 17*, 1317-1340

Foody, G.M. (2009). Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing, 30*, 5273-5291

Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., & Cooper, A. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment, 83*, 287-302

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X.M. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment, 114*, 168-182

Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F., & Obersteiner, M. (2011). Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software, 31*, 13

Globeland30. (2016). *Globeland30 product introduction* 10 Nov 2017,http://www.globeland30.org/home/Enbackground.aspx

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., & 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

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18-27

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*, 850-853

Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., & Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment, 112*, 2538-2556

Herold, M., See, L., Tsendbazar, N.-E., & Fritz, S. (2016). Towards an Integrated Global Land Cover Monitoring and Mapping System. *Remote Sensing, 8*, 1036

Herold, M., Woodcock, C.E., Stehman, S.V., Nightingale, J., Friedl, M.A., & Schmullius, C. (2009). The GOFC-GOLD/CEOS land cover harmonization and validation initiative: technical design and implementation. In, *33rd ISRSE*. Stresa, Italy

Huttich, C., Herold, M., Wegmann, M., Cord, A., Strohbach, B., Schmullius, C., & Dech, S. (2011). Assessing effects of temporal compositing and varying observation periods for large-area land-cover mapping in semi-arid ecosystems: Implications for global monitoring. *Remote Sensing of Environment, 115*, 2445-2459

Jung, M., Henkel, K., Herold, M., & Churkina, G. (2006). Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment, 101*, 534-553

Land Cover CCI. (2014). *CCI-LC Product User Guide*. Louvain-la-Neuve , Belgium: UCL-Geomatics

Land Cover CCI. (2017). *CCI-LC Product User Guide V2.0*. Louvain-la-Neuve , Belgium: UCL-Geomatics

Lesiv, M., Fritz, S., McCallum, I., Tsendbazar, N., Herold, M., Pekel, J.-F., Buchhorn, M., Smets, B., & Van De Kerchove, R. (2017). Evaluation of ESA CCI prototype land cover map at 20m. *IIASA Working Papaer, WP-14-021*

Lesiv, M., Fritz, S., Moorthy, I., Tsendbazar, N., Herold, M., Van De Kerchove, R., & Smets, B. (2016a). *CGLOPS1: Report describing the training dataset used for Dynamic Land Cover 100m product*

Lesiv, M., Moorthy, I., Fritz, S., Herold, M., Tsendbazar, N.E., Smets, B., & Van De Kerchove, R. (2016b). *GCLOPS: User requirement document - Dynamic Land Cover.*

Mayaux, P., Eva, H., Gallego, J., Strahler, A.H., Herold, M., Agrawal, S., Naumov, S., De Miranda, E.E., Di Bella, C.M., Ordoyne, C., Kopin, Y., & Roy, P.S. (2006). Validation of the global land cover 2000 map. *Geoscience and Remote Sensing, IEEE Transactions on, 44*, 1728-1739

Mayaux, P., Pekel, J.-F., Desclée, B., Donnay, F., Lupi, A., Achard, F., Clerici, M., Bodart, C., Brink, A., Nasi, R., & Belward, A. (2013). State and evolution of the African rainforests between 1990 and 2010. *Philosophical Transactions of the Royal Society B: Biological Sciences, 368*

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., & Wulder, M.A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment, 148*, 42-57

Olofsson, P., Stehman, S.V., Woodcock, C.E., Sulla-Menashe, D., Sibley, A.M., Newell, J.D., Friedl, M.A., & Herold, M. (2012). A global land-cover validation data set, part I: fundamental design principles. *International Journal of Remote Sensing, 33*, 5768-5788

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A.S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature, 540*, 418-422

Pengra, B., Long, J., Dahal, D., Stehman, S.V., & Loveland, T.R. (2015). A global reference database from very high resolution commercial satellite data and methodology for application to Landsat derived 30m continuous field tree cover data. *Remote Sensing of Environment, 165*, 234-248

Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A., Freire, S., Halkia, M., Julea, A., Kemper, T., Soille, P., & Syrris, V. (2016). Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. *Publ. Off. Eur. Union*

Romijn, E., Herold, M., Mora, B., Briggs, S., Seifert, F.M., & Paganini, M. (2016). *Monitoring progress towards: Sustainable Development Goals The role of land monitoring*. Wageningen, Netherlands.^http://www.gofcgold.wur.nl/documents/newsletter/Sustainable_Development_Goals-infobrief.pdf

Scepan, J., Menz, G., & Hansen, M.C. (1999). The DISCover validation image interpretation process. *Photogrammetric Engineering and Remote Sensing, 65*, 1075-1081

Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C., & Townshend, J.R. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth, 6*, 427-448

Stehman, S.V. (2014). Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. *International Journal of Remote Sensing, 35*, 4923-4939

Stehman, S.V., Olofsson, P., Woodcock, C.E., Herold, M., & Friedl, M.A. (2012). A global land-cover validation data set, II: augmenting a stratified sampling design to estimate accuracy by region and land-cover class. *International Journal of Remote Sensing, 33*, 6975-6993

Stehman, S.V., Wickham, J., Smith, J., & Yang, L. (2003). Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: Statistical methodology and regional results. *Remote Sensing of Environment, 86*, 500-516

Stehman, S.V., Wickham, J.D., Wade, T.G., & Smith, J.H. (2008). Designing a multi-objective, multi-support accuracy assessment of the 2001 National Land Cover Data (NLCD 2001) of the conterminous United States. *Photogrammetric Engineering & Remote Sensing, 74*, 1561-1571

Tateishi, R., Hoan, N.T., Kobayashi, T., Alsaaideh, B., Tana, G., & Phong, D.X. (2014). Production of Global Land Cover Data – GLCNMO2008. *2014, 6*

Tateishi, R., Uriyangqai, B., Al-Bilbisi, H., Ghar, M.A., Tsend-Ayush, J., Kobayashi, T., Kasimu, A., Hoan, N.T., Shalaby, A., Alsaaideh, B., Enkhzaya, T., Gegentana, & Sato, H.P. (2011). Production of global land cover data - GLCNMO. *International Journal of Digital Earth, 4*, 22-49

Tsendbazar, N.-E., de Bruin, S., & Herold, M. (2016a). Integrating global land cover datasets for deriving user-specific maps. *International Journal of Digital Earth*, 1-19

Tsendbazar, N., Herold, M., Fritz, S., & Lesiv, M. (2017). *Copernicus Global Land Operations: Validation Report for Dynamic Land Cover 100m product*: Copernicus Global Land Operations

Tsendbazar, N.E., de Bruin, S., Fritz, S., & Herold, M. (2015a). Spatial Accuracy Assessment and Integration of Global Land Cover Datasets. *Remote Sensing, 7*, 15804-15821

Tsendbazar, N.E., de Bruin, S., & Herold, M. (2015b). Assessing global land cover reference datasets for different user communities. *ISPRS Journal of Photogrammetry and Remote Sensing, 103*, 93-114

Tsendbazar, N.E., de Bruin, S., Mora, B., Schouten, L., & Herold, M. (2016b). Comparative assessment of thematic accuracy of GLC maps for specific applications using existing reference data. *International Journal of Applied Earth Observation and Geoinformation, 44*, 124-135

Tuanmu, M., & Jetz, W. (2014). A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography, 23*, 1031-1045

Verburg, P.H., Neumann, K., & Nol, L. (2011). Challenges in using land use and land cover data for global change studies. *Global Change Biology, 17*, 974-989

Wickham, J., Stehman, S., Fry, J., Smith, J., & Homer, C. (2010). Thematic accuracy of the NLCD 2001 land cover for the conterminous United States. *Remote Sensing of Environment, 114*, 1286-1296

Xiong, J., Thenkabail, P., Tilton, J., Gumma, M., Teluguntla, P., Oliphant, A., Congalton, R., Yadav, K., & Gorelick, N. (2017). Nominal 30-m Cropland Extent Map of Continental Africa by Integrating Pixel-Based and Object-Based Algorithms Using Sentinel-2 and Landsat-8 Data on Google Earth Engine. *Remote Sensing, 9*, 1065