



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

Tsendbazar, N. E., Herold, M., de Bruin, S., Lesiv, M., Fritz, S., Van De Kerchove, R., ... Pekel, J. F.

This is a "Post-Print" accepted manuscript, which has been published in "Remote Sensing of Environment"

This version is distributed under a non-commercial no derivatives Creative Commons



(CC-BY-NC-ND) user license, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and not used for commercial purposes. Further, the restriction applies that if you remix, transform, or build upon the material, you may not distribute the modified material.

Please cite this publication as follows:

Tsendbazar, N. E., Herold, M., de Bruin, S., Lesiv, M., Fritz, S., Van De Kerchove, R., ... Pekel, J. F. (2018). Developing and applying a multi-purpose land cover validation dataset for Africa. *Remote Sensing of Environment*, 219, 298-309. DOI: 10.1016/j.rse.2018.10.025

You can download the published version at:

<https://doi.org/10.1016/j.rse.2018.10.025>

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

2 Tsendbazar, N-E^{1*}, Herold, M¹., de Bruin, S¹., Lesiv, M²., Fritz, S²., Van De Kerchove, R³., Buchhorn, M³.,
3 Duerauer, M²., Szantoi, Z^{4,5}., and Pekel, J-F⁴

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

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

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

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

9 ⁵ Department of Geography and Environmental Studies, Stellenbosch University, Stellenbosch 7602, South
10 Africa

11 * Corresponding author: nandin.tsendbazar@wur.nl

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

13 **Abstract**

14 The production of global land cover products has accelerated significantly over the past decade thanks
15 to the availability of higher spatial and temporal resolution satellite data and increased computation
16 capabilities. The quality of these products should be assessed according to internationally promoted
17 requirements e.g., by the Committee on Earth Observation Systems-Working Group on Calibration and
18 Validation (CEOS-WGCV) and updated accuracy should be provided with new releases (Stage-4
19 validation). Providing updated accuracies for the yearly maps would require considerable effort for
20 collecting validation datasets. To save time and effort on data collection, validation datasets should be
21 designed to suit multiple map assessments and should be easily adjustable for a timely validation of new
22 releases of land cover products. This study introduces a validation dataset aimed to facilitate multi-
23 purpose assessments and its applicability is demonstrated in three different assessments focusing on
24 validating discrete and fractional land cover maps, map comparison and user-oriented map assessments.
25 The validation dataset is generated primarily to validate the newly released 100m spatial resolution land
26 cover product from the Copernicus Global Land Service (CGLS-LC100). The validation dataset
27 includes 3617 sample sites in Africa based on stratified sampling. Each site corresponds to an area of
28 100m×100m. Within site, reference land cover information was collected at 100 subpixels of 10m×10m
29 allowing the land cover information to be suitable for different resolution and legends. Firstly, using this
30 dataset, we validated both the discrete and fractional land cover layers of the CGLS-LC100 product.
31 The CGLS-LC100 discrete map was found to have an overall accuracy of 74.6±2.1% (at 95%
32 confidence level) for the African continent. Fraction cover products were found to have mean absolute
33 errors of 9.3, 8.8, 16.2, and 6.5% for trees, shrubs, herbaceous vegetation and bare ground, respectively.
34 Secondly, for user-oriented map assessment, we assessed the accuracy of the CGLS-LC100 map from
35 four user groups' perspectives (forest monitoring, crop monitoring, biodiversity and climate modelling).
36 Overall accuracies for these perspectives vary between 73.7% ±2.1% and 93.5% ±0.9%, depending on
37 the land cover classes of interest. Thirdly, for map comparison, we assessed the accuracy of the
38 Globeland30-2010 map at 30m spatial resolution. Using the subpixel level validation data, we derived
39 15252 sample pixels at 30m spatial resolution. Based on these sample pixels, the overall accuracy of the
40 Globeland30-2010 map was found to be 66.6 ±2.4% for Africa. The three assessments exemplify the
41 applicability of multi-purpose validation datasets which are recommended to increase map validation
42 efficiency and consistency. Assessments of subsequent yearly maps can be conducted by augmenting or
43 updating the dataset with sample sites in identified change areas.

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

46 1. Introduction

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

52 The first satellite-based global land cover map dates back to 1994 (DeFries and Townshend 1994). Over
53 the past decades numerous global land cover maps were produced using medium resolution satellite data
54 (Arino et al. 2007; Bartholomé and Belward 2005; Friedl et al. 2002; Land Cover CCI. 2014; Tateishi
55 et al. 2011). Pioneering the productions of higher resolution land cover mapping at large scale,
56 researchers have created global and continental scale land cover products using Landsat (Chen et al.
57 2015; Gong et al. 2013; Hansen et al. 2013) and Sentinel-2 data (CCI Land Cover 2017a). Our
58 understanding of the changing environment is further enhanced with the recent land cover change
59 products namely annual LC-CCI land cover maps (CCI Land Cover 2017b), Global Surface Water
60 Explorer (Pekel et al. 2016), Global Human Settlement Layers (Pesaresi et al. 2016) and Global Forest
61 Change datasets (Hansen et al. 2013).

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

72 cover types: trees, shrubs, herbaceous vegetation and bare soil) maps (Copernicus Global Land Service
73 2017).

74 Although, the validation of global land cover products has become a common activity for assessing their
75 quality and usability (Herold et al. 2016), validation activities should adjust to the emergence of new or
76 subsequent products without much additional effort. Most global land cover validation datasets are
77 collected via visual interpretation (Chen et al. 2015; Tsendbazar et al. 2015b; Xiong et al. 2017), a labour
78 intensive task requiring efforts of multiple mapping and image interpretation experts (Defourny et al.
79 2011; Mayaux et al. 2006; Scean et al. 1999). To guarantee the independence from the training data
80 and the consistency of the validation results (as well as to save time and effort), such datasets should be
81 designed to be suitable for multiple map assessments and could be re-used, to provide timely quality
82 assessments on the new and subsequent land cover products.

83 However, most existing validation datasets were generated to validate a single land cover map and their
84 characteristics such as sample site areas and thematic legends are not suitable to be used for validating
85 multiple maps. For example, a validation dataset (with some 150 000 sample locations) for the
86 Globeland30 map (Chen et al. 2015) is limited to assessing other maps having similar resolution as the
87 Globeland30. Similarly, the validation dataset developed for the GlobCover 2009 map (Defourny et al.
88 2011) is constrained to be used for assessing maps with medium resolution (~300m) (CCI Land Cover
89 2017b). A recent review of metadata on global land cover validation datasets found that re-using a
90 validation dataset to assess another map usually comes at a cost, namely loss of spatial and thematic
91 detail (Tsendbazar et al. 2015b). This restricts the usage of validation datasets for purposes such as
92 assessing fraction maps, map comparisons and map assessments from different users' perspectives. For
93 example, most validation datasets represent the reference land cover as discrete classes according to
94 fixed legends. Therefore they do not record land cover fraction information (e.g., tree cover fractions).
95 As such their utility for validating land cover maps is limited (Tsendbazar et al. 2015b).

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

98 2009). GOFC-GOLD emphasizes the importance of inter-operability and comparability of global land
99 cover maps to help map users select the most suitable maps for their needs (Herold et al. 2008). A
100 statistical comparison of several land cover maps requires a validation dataset that has been acquired by
101 transparent means and that is suitable for multiple map assessments in terms of spatial resolution and
102 thematic legends. For example, the class “forest” can have different definitions (e.g., >30% or >60%
103 forest density)(Jung et al. 2006), thus the validation dataset used for comparison should be able to
104 accommodate such differences. Therefore, GOFC-GOLD and the working group on calibration and
105 validation of the Committee on Earth Observation Satellites (GEOS-WGCV) proposed a multi-purpose
106 validation dataset (Herold et al. 2009) which was further detailed in Olofsson et al. (2012). For improved
107 re-usability, the dataset was designed to be flexible in terms of sample selection, sample unit area and
108 thematic detail (Olofsson et al. 2012). For example, the reference land cover in a sample unit area (5km
109 × 5km) is generated from classifications of very high resolution (2m) images and this makes the dataset
110 suitable for assessing maps with different resolutions up to 5km × 5km. Fractional coverage of land
111 cover types within the sample unit area can also be estimated with this dataset. The initial sample
112 comprised 500 sites and could be increased if required (Stehman et al. 2012). The dataset has been
113 published by the United States Geological Survey (Pengra et al. 2015). However, thematically it only
114 comprises four land cover categories, i.e., trees, water, bare, and other.

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

125 more efforts are needed to create validation datasets that match different spatial and thematic detail as
126 well as different users' perspectives.

127 Subsequent releases of land cover products should be provided with updated independent validation
128 reports according to the Stage 4 validation requirements of the CEOS-WGCV (Herold et al. 2009). Most
129 currently available global land cover products do not meet this requirement. Apart from the CCI-2015,
130 which was validated using the GlobCover-2009 validation dataset (CCI Land Cover 2017b), none of the
131 yearly CCI-LC land cover products has been validated. The same applies to the MODIS land cover maps
132 for which only the accuracy of the 2005 map was assessed (Friedl et al. 2010). Validation of new land
133 cover products would benefit from a validation dataset that is updated using less demanding efforts,
134 such as re-interpreting and adding additional sample locations in identified change areas. Stehman et al.
135 (2012) recommended using stratified sampling to facilitate sample augmentation.

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

152 2. Methods and materials

153

154 2.1. Validation data collection

155 2.1.1. Sampling design

156 A probability sampling scheme was used to allow design-based inference of map accuracies. The sample
157 selection scheme had to be suitable for validating the CGLS-LC100 maps and other land cover maps.
158 Therefore, appropriate choices for sample size, sample selection scheme and sample unit size (spatial
159 support) were considered given constraints imposed by allowable error (Foody 2009; Olofsson et al.
160 2012).

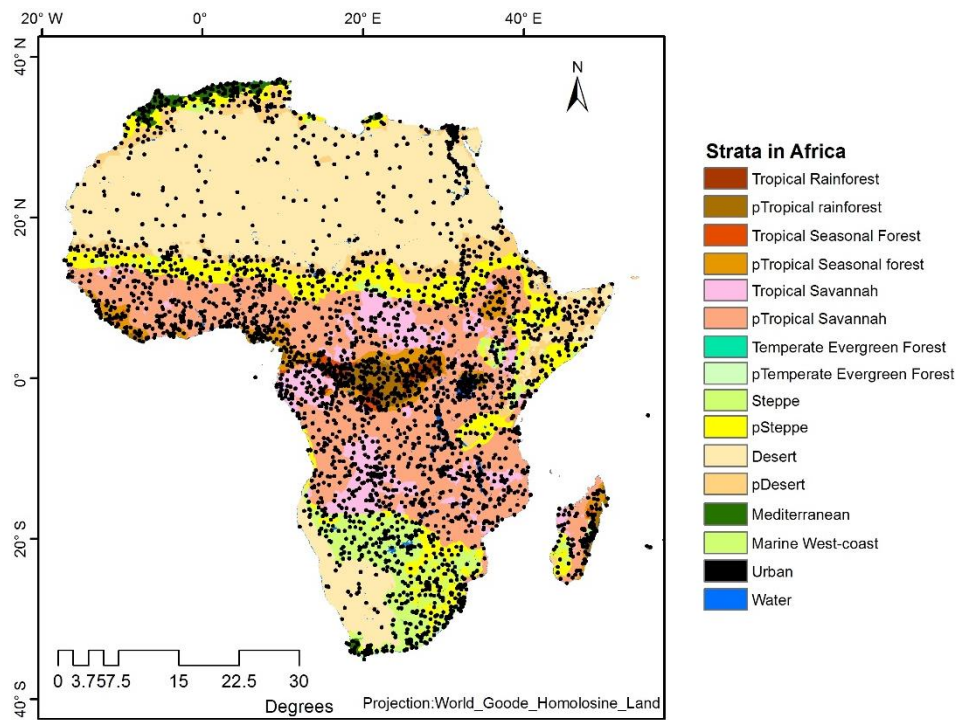
161 Considering the efforts required to collect the validation dataset (expert training, interpretation and
162 quality checking: see Section 2.1.2) a sample size of 2700 sites was considered feasible. Such sample
163 size is similar or larger than those used for statistical assessments of large scale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

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

174 The sample allocation process focused on strata in which some land cover classes that are more likely
175 to be misclassified (Olofsson et al. 2012). Since, the Sahel and dry savannah's heterogeneous landscapes
176 in Africa are known to have lower map accuracies (Tsendbazar et al. 2015a), more sample sites were
177 allocated to these heterogeneous areas and to the populated strata (Figure1). The sample sizes per
178 stratum are listed in Table S1(Supplementary Materials). At each sample site location, reference land

179 cover of an area of 100m × 100m was identified. This support size coincides with the pixel size of the
180 Proba-V satellite data used to generate the CGLS-LC100 land cover products.

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



189
190 **Figure 1: Spatial distribution of all validation sample sites and the stratification by Olofsson et al 2012: ‘p’**
191 **before the strata names denote populated part of climate zone.**

192 2.1.2. Response design

193 To allow multi-purpose assessments of land cover maps, the spatial and thematic representations of the
194 validation dataset are designed to be compatible for maps with different resolutions and legends. For
195 this, similar to the training data collection used for the CGLS-LC100 product (Lesiv et al. 2016a), each
196 sample site (100m × 100m) was divided into 10×10 small blocks (10m × 10m) and reference land cover

197 was collected at the subpixel level. This makes the validation dataset compatible for assessing maps
 198 with 10-100m resolutions. For the thematic representation, we labelled the land cover in terms of generic
 199 elements dominating the 10m × 10m subpixels. Land cover elements include trees (different leaf and
 200 phenology types), shrubs, grass, crops, built-up areas, bare area, water body, snow & ice and regularly
 201 flooded herbaceous area (wetlands). The land cover elements were defined according to the United
 202 Nations Land Cover Classification System (UN-LCCS) (Di Gregorio 2005). This allows the validation
 203 dataset to be thematically compatible for multiple maps by using different combinations of the land
 204 cover elements based on legend definition requirements of multiple maps.

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

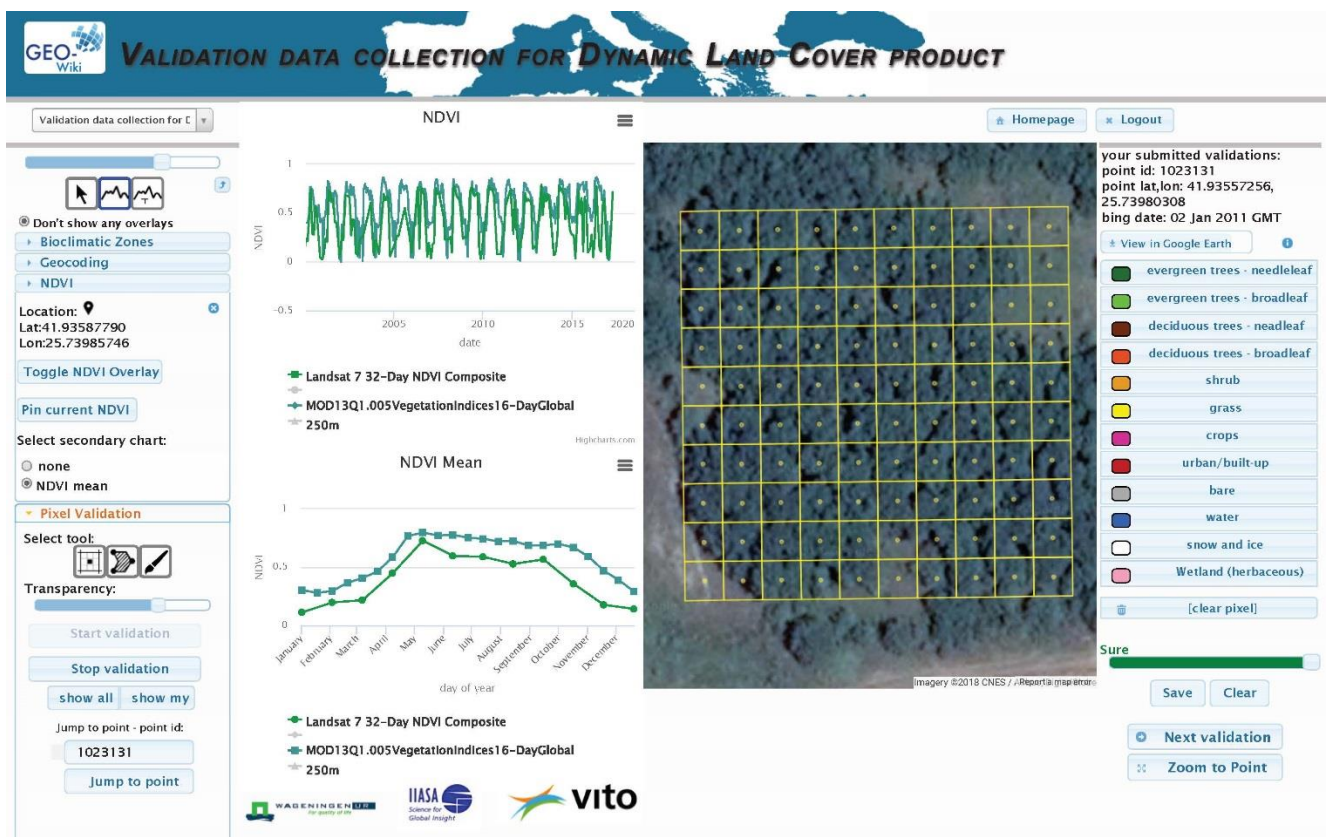
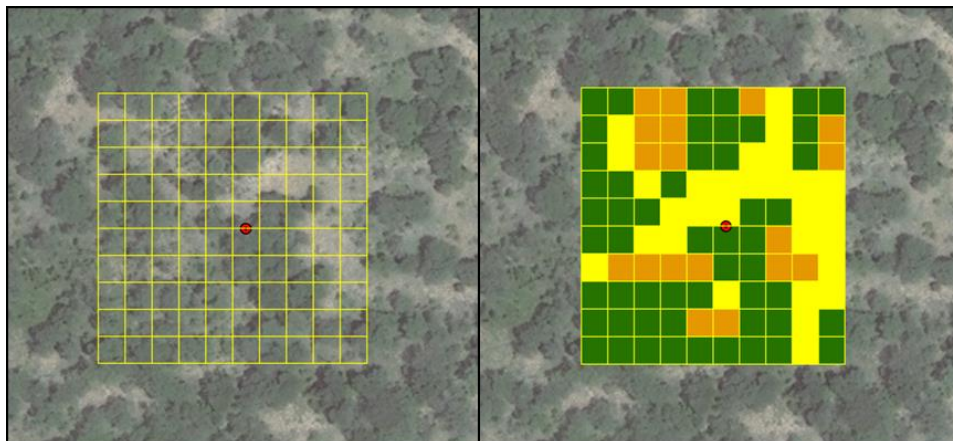


Figure 2: Screen shot of Geo-Wiki based interface for land cover validation

211
212

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



214

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

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

224

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 GOF-C-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

225

226 Different quality control measures were applied to obtain a reliable and good quality reference dataset
227 for validation. Firstly, in addition to a tutorial on land cover interpretation, a training workshop was
228 organized for the global land regional land cover mapping experts in January 2017 at IIASA, Laxenburg,
229 Austria. The aim of the workshop was to reduce interpretation discrepancies among the experts. The

230 experts were asked to interpret the same 30 sample sites (100m x 100m) and feedback on any
231 discrepancy was provided upon examination by global land cover mapping experts. The global land
232 cover mapping experts were independent from the CGLS-LC100 product generation. Secondly,
233 depending on the available sources of information (e.g., high resolution images and NDVI profiles) and
234 complexity of landscapes (e.g., small holder cultivation areas), the confidence in the interpretation can
235 be different. Therefore, we recorded the interpretation confidence levels (i.e., unsure, bit sure, quite sure,
236 sure). Three percent of the sample sites were tagged as “unsure” or “bit sure”. Lastly, all the
237 interpretations including these unsure interpretations were checked by global land cover mapping
238 experts and feedback on each interpretation was provided to the experts. The regional experts either
239 rebutted the feedback or corrected their interpretations where necessary.

240 2.2. Land cover products

241 To demonstrate applicability of the validation dataset for multiple applications, we selected two land
242 cover maps at different spatial resolutions and different legends: (1) the CGLS-LC100 V1.0 at 100m
243 resolution provided for the 2015 reference year over Africa (Buchhorn et al. 2017); (2) the Globeland30
244 2010 map (Chen et al. 2015).

245 The CGLS-LC100 V1.0 at 100m resolution product, provided for the 2015 reference year over Africa
246 (Buchhorn et al. 2017), is a new product in the CGLS portfolio. The CGLS-LC100 is based on the Proba-
247 V 100m data archive (Dierckx et al. 2014), a high quality land cover training dataset (Lesiv et al. 2016a)
248 and several ancillary datasets. More description of the map generation is detailed in Buchhorn et al.
249 (2017). Apart from a discrete land cover type map, the product includes four vegetation continuous field
250 layers providing estimates of fractions (0 - 100%) for the land cover types: trees, shrub, herbaceous
251 vegetation and bare ground.

252 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.

254

255 We also assessed the Globeland30 map (Chen et al. 2015) for comparison. The Globeland30 project of
 256 China's Ministry of Science and Technology produced global land cover maps for the year 2000 and
 257 2010. The maps were produced at 30m resolution using Landsat TM and ETM+ and the Chinese
 258 Environmental Disaster Alleviation Satellite (HJ-1) data. We used the 2010 map for Africa. This map
 259 has ten land cover classes of which eight occur in Africa (cultivated land, forest, grassland, shrubland,
 260 wetland, water bodies, artificial surfaces and bare land) (Globeland30 2016). The overall map accuracy
 261 has been reported to be 79.26% at global level (Chen et al. 2015) but no accuracy information is available
 262 for Africa.

263 2.3. Validation of discrete and fractional land cover map

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

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

277 To estimate the accuracy of the land cover maps, we accounted for unequal inclusion probabilities
278 between different strata because sample sites were not allocated proportionally to the strata areas
279 (Olofsson et al. 2012; Wickham et al. 2010). Based on Pengra et al. (2015), the inclusion probability for
280 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
281 stratum h (see Table S1 for inclusion probabilities per stratum). Number of sites is based on the 100m
282 \times 100m units. Inclusion probability for the additional sample sites were calculated based on the
283 population of possible sample sites within the rare classes of the CGLS-LC100 map. The estimation
284 weight, the inverse of inclusion probability ($\omega_h = 1 / \pi_h$), was then calculated and used to construct the
285 confusion matrix accounting for unequal sample inclusion probabilities following the methods described
286 in Stehman et al. (2003) and Wickham et al. (2010). We then estimated the overall and class specific
287 accuracies and their confidence intervals (at 95% confidence level) following Stehman (2014) which
288 specifically addresses estimating map accuracies when the sampling strata are different from the map
289 classes. Thus, by appending three rare class strata to the original stratification, 19 strata were used in the
290 calculations.

291 Validation data does not contain information on temporary waterbody areas because of limited
292 availability on multiple high resolution images per year for each sample location. Thus, we merged the

293 mapped classes of permanent and temporary waterbody for the accuracy assessment. Owing to the
294 limited sample size for combinations of forest density (closed and open forest) and forest phenology,
295 the accuracy estimation focused on generic classes without taking specific forest phenology into
296 account.

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

$$300 \quad RMSE_c = \sqrt{\frac{\sum_{i=1}^n \omega_i (p_i - v_i)^2}{\sum_{i=1}^n \omega_i}} \quad (\text{Eq.1})$$

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

$$304 \quad MAE_c = \frac{\sum_{i=1}^n \omega_i |p_i - v_i|}{\sum_{i=1}^n \omega_i} \quad (\text{Eq.2})$$

305 where MAE_c is the mean absolute error of class c .

306 [2.4. Accuracy comparison with other datasets at different spatial resolution](#)

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

314 To make the validation dataset compatible with 30m resolution Globeland30 map, we extracted pixel
315 values of the Globeland30 map over each subpixel area (10×10m) of the validation dataset. Using the
316 subpixel centroid locations, we selected Globeland30 pixels that spatially overlap with the subpixels of

317 the validation dataset (at least nine subpixel centre points of the validation dataset). The reference land
318 covers over nine subpixels were aggregated to derive reference land cover for 30m pixels. For
319 homogeneous areas, the land cover elements were directly converted to land cover classes. In
320 heterogeneous areas that can have multiple possible land cover types, we used the dominant land cover
321 type as reference land cover. Sample pixels which did not have a clear dominance (e.g., four sub-pixels
322 of trees, four sub-pixels of shrubs and one sub-pixel of water), totalling to 1037 cases, were excluded
323 from the assessment. A total of 15252 sample pixels were available at 30m resolution.

324 Next, the Globeland30 map was evaluated using a stratified one-stage cluster approach (Pengra et al.
325 2015) because multiple 30m sample pixels within the 100m × 100m sites were used for the assessment.
326 Calculation of inclusion probabilities, accuracy estimates and confidence intervals followed the
327 stratified one-stage cluster approach described in Pengra et al. (2015) and Stehman et al. (2003).

328 [2.5. Map validation from different users' perspectives](#)

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

335 Forest monitoring

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

339 The current legend of the CGLS-LC100 discrete map includes closed forests (>70% tree cover) and
340 open forests (15-70% tree cover) classes. A tree cover mosaic class (30 – 70% tree cover) is also widely
341 used in forest monitoring applications (e.g., TREES3 dataset) (Achard et al. 2002; Mayaux et al. 2013),
342 We used the tree cover fraction layer of the CGLS-LC100 product to separate the open forests class in

343 the discrete map into two different classes (tree cover mosaic (30-70% tree cover) and open tree cover
344 mosaic (15-30% tree cover)). Figure 4a depicts a map with seven forest-related classes differing in terms
345 of phenology and tree cover densities based on the CGLS-LC100 discrete LC map and tree cover
346 fraction layer.

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

349 Crop monitoring

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

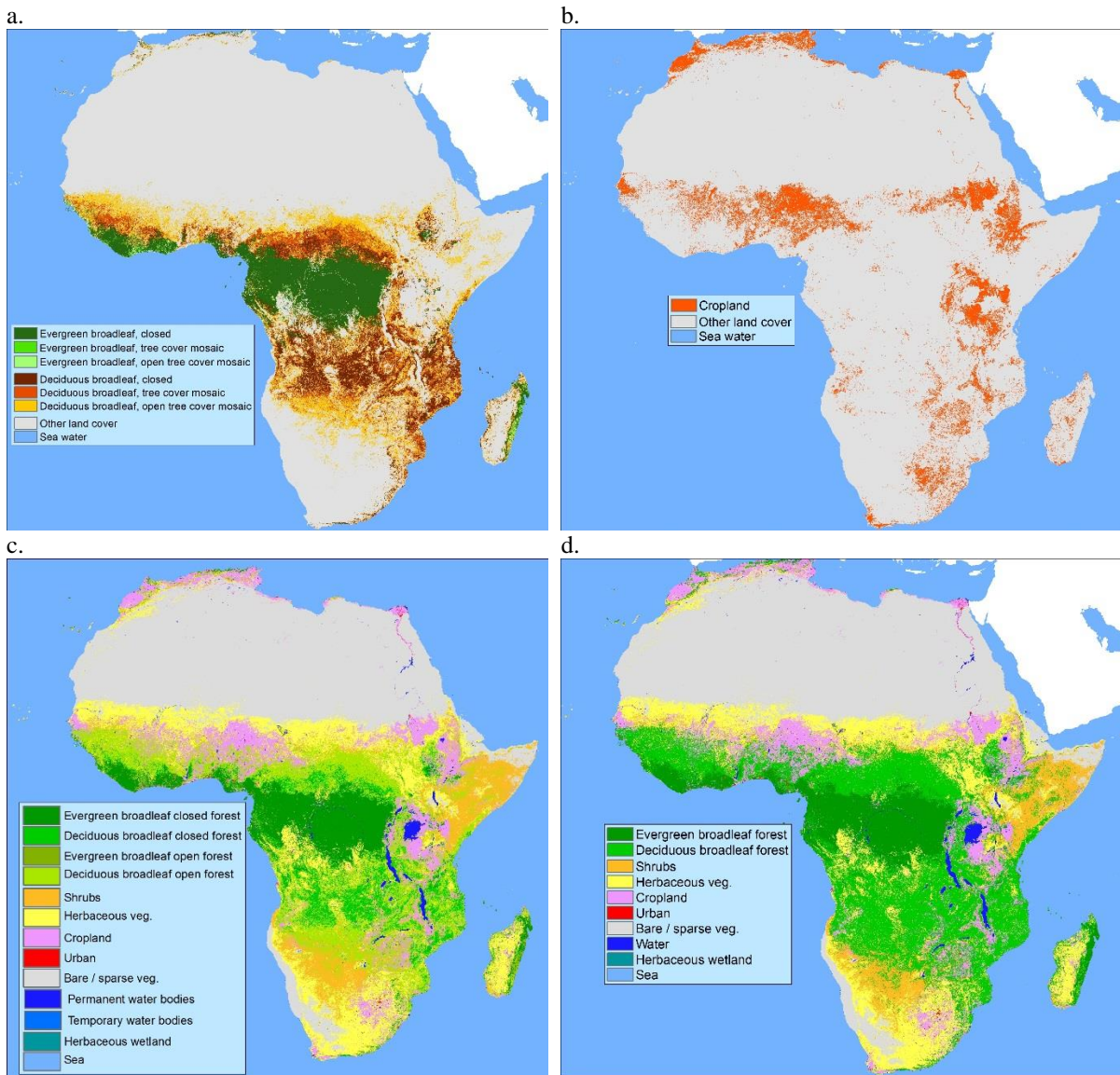


Figure 4. User specific maps based on the CGLS-LC100 products for (a) forest monitoring, (b) crop monitoring, (c) biodiversity and (d) climate modelling

354
355

356 Biodiversity

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

362 Climate modelling

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

368 3. Results

369 3.1. Validation of discrete and fractional land cover maps

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

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

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

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

		Reference class									Sample count	Total	User's accuracy	Confidence interval +/-
		Closed forest	Open forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/Sparse veg.	Water	Wetland				
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				

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

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

383 **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

384

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

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

389 [3.2. Accuracy comparison with other datasets at different spatial resolution](#)

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

392

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				

394

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

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

401 3.3. Map validation from different users' perspectives

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

404 Overall map accuracy for forest monitoring was estimated at $81.3\% \pm 1.4\%$ (Table 6). The confusion
405 matrix and class specific accuracies show that closed forests types (evergreen broadleaf and deciduous
406 broadleaf) are mapped with higher accuracy (Table S4). Closed evergreen broadleaf forest is mapped

407 with good accuracy (>90%). The accuracy of the tree cover mosaic and the open tree cover mosaic
 408 classes are low.

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

412 **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% \pm 2.1%
Forest monitoring	A map with 6 forest related classes (Figure 4a)	81.3% \pm 1.4%
Crop monitoring	Cropland and non-cropland mask (Figure 4b)	93.5 \pm 0.9% Cropland class: User's accuracy: 59.4 \pm 6.5 % Producer's accuracy: 66.3 \pm 6.2%
Biodiversity	Discrete land cover map with 11 classes (Figure 4c)	73.7 % \pm 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% \pm 2.1% MAE: 9, 8.8, 16, and 6.5%, respectively RMSE: 16.7, 15, 24.8, and 14.8% , respectively

413
 414 The overall accuracy was assessed at $73.7\% \pm 2.1\%$ for biodiversity related use. The class accuracies
 415 and the confusion matrix are provided in Table S6. The class accuracies are similar to those presented
 416 in Table 3. The producer's accuracy of the open forest, evergreen broadleaf class is low since this class
 417 is mostly confused with closed forest evergreen broadleaf and open forest deciduous broadleaf classes.

418 For climate modelling users, the map overall accuracy was determined to be $77.3\% \pm 2.1\%$ (Table 6).
 419 The class-specific accuracies and the confusion matrix can be found in Table S7. For the evergreen
 420 broadleaf forest class, the user's and producer's accuracies are 95% and 89.6% respectively. This class
 421 appears to be slightly under-represented. The deciduous broadleaf forest is slightly over-represented
 422 with users and producer's accuracy of 72.9% and 74% respectively. In addition to the accuracy of the
 423 discrete map from the climate modelling perspective, the accuracy of the cover fraction layers provided
 424 in Table 6 can be important as climate modellers are often interested in land cover information related

425 to plant functional types and fraction information on the main land cover types are very useful towards
426 this.

427 4. Discussion

428 4.1. The multi-purpose validation dataset development and use

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

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

- 439 (i) To extract class fraction information within the sample site areas;
- 440 (ii) To collect information on the land cover elements such as trees and buildings to be used for
441 different legends; and
- 442 (iii) To validate land cover maps at finer resolution (e.g. at Sentinel-2 and Landsat scale)

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

445 A design of multi-objective accuracy assessment was previously introduced for National Land Cover
446 Data of the United States of America (Stehman et al. 2008). This design addresses different aims of
447 accuracy assessments such as class-specific accuracies, land cover proportion accuracies and net change
448 detection accuracy. This design is limited to one map with a fixed legend and resolution and it is for the
449 extent of the United States of America. The CGLS validation dataset is produced for the African
450 continent and the proposed approach can be expanded to global scale applications thanks to the global

451 stratification derived from Köppen climate zones and population density (Olofsson et al. 2012). The
452 current setup for data collection in the African continent (Section 2.1) can be replicated to other
453 continents to collect validation dataset at global scale. If the similar numbers of sample sites were
454 collected for the five other continents, the total sample size would be larger than 20 000. A stratification
455 independent from the target land cover maps allows collecting the validation data while the target map
456 is being produced, thus reducing the lag between map production and its accuracy assessment.
457 Regardless of the stratification chosen, the accuracy estimates will be unbiased for the true accuracy of
458 each map. However, the precision of the accuracy estimates computed from a stratification independent
459 of the target map will be lower than if that map itself would be used for stratification.

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

468 4.2. Validation of the discrete and fractional land cover maps

469 We assessed the accuracy of the CGLS-LC100 discrete and fractional land cover maps using the
470 validation dataset described in Section 2.1. The overall accuracy of the discrete map was found to be
471 74.6% \pm 2.1%. This overall accuracy is comparable with the reported accuracy for the CCI-LC-2015
472 map at global scale (75.3% using only homogeneous sample sites) (Land Cover CCI. 2017). At the
473 African continental scale, Tsendbazar et al. (2015a) found overall correspondences of 50-63% for four
474 global land cover maps (GlobCover 2009, CCI-LC 2010, MODIS-2010 and Globeland 2010). Similarly,
475 the overall accuracy of the Globeland30-2010 map obtained in the current study was assessed at 66.6
476 \pm 2.4% for Africa (Table 4). These results suggest that the CGLS-LC100 discrete map has higher overall

477 accuracy compared to Globeland30 map (Table 5) and other land cover maps for Africa (Tsendbazar et
478 al. 2015a).

479 Closed forest and bare/sparse vegetation classes have higher class specific accuracies followed by the
480 open forest, herbaceous vegetation and cropland classes (Table 3). Among the natural vegetation classes,
481 shrubs are mapped with the lowest accuracies owing to high confusion with open-forest and herbaceous
482 vegetation classes. Confusion between open forests, herbaceous vegetation and shrubs is a known
483 problem for land cover mapping in savannah ecosystems where different vegetation layers (woody and
484 herbaceous vegetation) co-exist (Huttich et al. 2011; Jung et al. 2006). The CGLS-LC100 map slightly
485 over-represents the bare/sparse vegetation class at the cost of herbaceous and shrubs areas, particularly
486 in border regions of the Sahara and Namib deserts (Table 3). The cropland class is confused with open
487 forest and herbaceous vegetation (Table 3). This can be attributed to the difficulty of separating cropland
488 from herbaceous vegetation, and small-scale cultivation in heterogeneous landscapes (Xiong et al.
489 2017).

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

497 Among the land cover fraction maps, bare area has the lowest errors (MAE 6.5% and RMSE of 14.8%),
498 while, herbaceous vegetation has the highest errors. This can be attributed to the difficulty of separating
499 herbaceous vegetation from other land cover types. This is confirmed by Table 3 where the herbaceous
500 vegetation class is mostly confused with other classes. For the tree cover fraction map, there is no direct
501 comparison available for Africa. However, compared to reported errors in other regions, the tree cover
502 fraction of the CGLS-LC100 has similar or slightly lower errors (MAE 9.3% and RMSE 16.7%). For

503 example, in South-America, the Landsat based tree cover 2010 product by Hansen et al. (2013) was
504 found to have a MAE of 9.39% (Pengra et al. 2015). A Landsat based rescaled version of the MODIS
505 Vegetation Continuous Field percent tree cover product was reported to have 17% RMSE when
506 compared against LiDAR measurements of four regions in North America (Sexton et al. 2013).

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

520 [4.3. Accuracy comparison with other datasets at different spatial resolution](#)

521 We also assessed the 30m resolution Globeland30 2010 map for Africa using our validation dataset.
522 This demonstrates the suitability of our validation dataset for assessing a higher resolution map for
523 comparison. Based on 15252 sample pixels derived from our validation data, the overall accuracy of the
524 Globeland30 was estimated at $66.6 \pm 2.4\%$ (Table 5). This accuracy is lower than the accuracy reported
525 by the map producers ($79.26\% \pm 0.2\%$) (Chen et al. 2015). However the accuracy by Chen et al. (2015)
526 is for the entire globe and while the results obtained in this study are for Africa, a continent that tends
527 to have lower map accuracy than other continents (Tsendbazar et al. 2016a). We used the dominant land
528 cover type for validation, because details in the legend definition of some classes were not clear for this
529 map. However, if more detailed information on the legend thresholds is made available, validation could

530 also be done based on the legend definition of the Globeland30 map. There is a 5-year difference in the
531 reference year of the Globeland30 2010 map and this might have an influence on the lower overall
532 accuracy for this map. The assessment of this map serves here to demonstrate our validation data
533 applicability for maps having different spatial resolutions. It should be noted that temporal discrepancies
534 between validation data and maps to be assessed should be kept at a minimum.

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

546 4.4. Map validation from different users' perspectives

547 The land cover fraction information of the validation dataset allowed the assessment of the CGLS-
548 LC100 product from different users' perspectives. We created four maps with different legends (Figure
549 4) reflecting users' preferences on different land cover types (Lesiv et al. 2016b), and our results showed
550 varying overall accuracies (73.7% ±2.1% for biodiversity to 93.5±0.9 for crop monitoring) (Table 6).
551 Differences can partly be attributed to the number of land cover classes considered in the assessments
552 but also the class combinations used matter. Accuracy for the biodiversity users was lower due to the
553 number of classes used in the assessment. The higher overall accuracy for crop monitoring (cropland
554 and non-cropland map) was to be expected since internal confusions among non-cropland classes are
555 discarded in this assessment. Thus, the cropland class accuracies are also important map quality
556 measures in this case. The overall accuracy of the CGLS-LC100 cropland mask is similar, but the

557 cropland class accuracies are lower compared to the Landsat based nominal cropland mask (Xiong et al.
558 2017). In the forest monitoring applications, a map with more forest classes was created by combining
559 the discrete and tree cover fraction map of the CGLS-LC100 product. This further illustrates the
560 suitability of the CGLS-LC100 product towards creating user-tuned maps. Note that the users of land
561 cover maps are not restricted to the user groups identified in this study and the overall map accuracies
562 will differ for different applications, i.e., different classes of interest are considered.

563 4.5. Lessons learnt on the multi-purpose validation data development

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

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

580 To support the use of this validation data for other map assessments, future work can focus on
581 developing a service to provide instantly validated user-tuned land cover maps of the CGLS-LC100
582 products. The validation results could also be provided when other land cover maps are uploaded to the

583 service. This ensures the validation dataset is used for validation purpose rather than training or
584 calibration purposes.

585 5. Summary

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

591 We collected a validation dataset consisting of 3 617 sample sites for Africa using a global stratification
592 independent from any land cover map. Reference land cover of the sample sites (100m × 100m area)
593 was recorded at 10m × 10m subpixels by visual interpretation on a dedicated branch of the Geo-Wiki
594 platform with contributions from several regional experts from Africa. Several quality measures were
595 applied to ensure data quality. The response design of this validation dataset facilitates flexibility
596 towards multi-purpose applications. For example, the ability to assess maps of different resolution (10-
597 100m) is gained by subpixel level reference land cover information. The validation data also supports
598 assessment of maps with different legends. As opposed to creating legend categories by merging certain
599 classes, subpixel level reference land cover data allow specifically targeting classes defined by user-
600 specific composition thresholds. Furthermore, the stratified sampling scheme enables sample
601 augmentation for classes of interest (Stehman et al. 2012).

602 The applicability of the validation dataset was demonstrated for (1) validation of discrete and fractional
603 land cover maps (CGLS-LC100 product: overall accuracy 74.6% ±2.1% for the discrete and MAE
604 6.15%-16% for the fraction cover layers); (2) map comparison (Globeland30-2010 map: overall
605 accuracy 66.6 ±2.4%); and (3) user oriented accuracy reporting (CGLS-LC100 product users: overall
606 accuracy: 73.7% +/-2.1% to 81.3% +/- 1.4%).

607 In addition, the validation dataset is compatible with CGLS's focus on operational monitoring of land
608 cover with yearly releases of global maps. The global stratification used in the sampling facilitates

609 expanding to global scale by replicating the current setup for data collection in African continent for
610 other continents to collect validation dataset at global scale, with additional resources and expert
611 involvements. The flexibility of stratified sampling allows augmenting the validation dataset for
612 validating subsequent new maps to meet the updated validation requirement of the CEOS-WGCV Stage
613 4 validation. For the latter purpose, sampling needs to be densified in change area by re-interpreting
614 additional sample sites. On the contrary, validation sites in no-change areas can be re-used with little
615 effort by re-interpreting only a part of sample sites in no-change areas.

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

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

627 **Acknowledgments**

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

632 6. References

- 633 Achard, F., Eva, H.D., Stibig, H.J., Mayaux, P., Gallego, J., Richards, T., & Malingreau, J.P. (2002). Determination of deforestation rates of
634 the world's humid tropical forests. *Science*, 297, 999-1002
- 635 Arino, O., Leroy, M., Ranera, F., Gross, D., Bicheron, P., Nino, F., Brockman, C., Defourny, P., Vancutsem, C., & Achard, F. (2007).
636 GLOBCOVER-A Global Land Cover Service with MERIS. In, *Envisat Symposium 2007* (pp. 23-27). Montreux, Switzerland
- 637 Bartholomé, E., & Belward, A. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International
638 Journal of Remote Sensing*, 26, 1959-1977
- 639 Bontemps, S., Defourny, P., Van Bogaert, E., Kalogirou, V., & Arino, O. (2011). *GLOBCOVER 2009: Products Description and Validation
640 Report*: UCLouvain and ESA
- 641 Buchhorn, M., Bertels, L., Smets, B., Lesiv, M., & Tsendbazar, N. (2017). *Copernicus Global Land Operations "Vegetation and Energy":
642 Algorithm Theoretical Basis Document for Moderate Dynamic Land
643 Cover*.[^]https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS1_ATBD_LC100m-V1_IL00.pdf
- 644 CCI Land Cover. (2017a). *CCI LAND COVER - S2 prototype Land Cover 20m map of Africa 2016*
645 <http://2016africallandcover20m.esrin.esa.int/viewer.php>
- 646 CCI Land Cover. (2017b). *Release of a 1992-2015 time series of annual global land cover maps at 300 m*.[^][https://www.esa-landcover-
647 cci.org/index.php?q=webfm_send/88](https://www.esa-landcover-cci.org/index.php?q=webfm_send/88)
- 648 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
649 resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 7-27
- 650 Copernicus Global Land Service. (2017). *Copernicus Global Land Service- Land Cover*: VITO.2018, March
651 [12, https://land.copernicus.eu/global/products/lc](https://land.copernicus.eu/global/products/lc)
- 652 Defourny, P., Bontemps, S., Schouten, L., Bartalev, S., Cacetta, P., de Wit, A., di Bella, C., Gérard, B., Giri, C., Gond, V., Hazeu, G.,
653 Heinimann, A., Herold, M., Jaffrain, G., Latifovic, R., Lin, H., Mayaux, P., Mûcher, S., Nonguierma, A., Stibig, H., Y. Shimabakuro, Van
654 Bogaert, E., Vancutsem, C., Bicheron, P., Leroy, M., & Arino, O. (2011). GLOBCOVER 2005 and GLOBCOVER 2009 validation : learnt
655 lessons In, *GOFC-GOLD Global Land Cover & Change Validation Workshop*. Laxenburg, Austria
- 656 DeFries, R., & Townshend, J. (1994). NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing*,
657 15, 3567-3586
- 658 Di Gregorio, A. (2005). *Land cover classification system: classification concepts and user manual: LCCS*. Rome, Italy: Food and
659 Agriculture Organization of the United Nations
- 660 Dierckx, W., Sterckx, S., Benhadj, I., Livens, S., Duhoux, G., Van Achteren, T., Francois, M., Mellab, K., & Saint, G. (2014). PROBA-V
661 mission for global vegetation monitoring: standard products and image quality. *International Journal of Remote Sensing*, 35, 2589-2614
- 662 Foody, G.M. (1996). Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data.
663 *International Journal of Remote Sensing*, 17, 1317-1340
- 664 Foody, G.M. (2009). Sample size determination for image classification accuracy assessment and comparison. *International Journal of
665 Remote Sensing*, 30, 5273-5291
- 666 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., &
667 Cooper, A. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83, 287-302
- 668 Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X.M. (2010). MODIS Collection 5 global
669 land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114, 168-182
- 670 Fritz, S., McCallum, I., Schill, C., Perger, C., See, L., Schepaschenko, D., van der Velde, M., Kraxner, F., & Obersteiner, M. (2011). Geo-
671 Wiki: An online platform for improving global land cover. *Environmental Modelling & Software*, 31, 13
- 672 Globeland30. (2016). *Globeland30 product introduction* 10 Nov 2017, <http://www.globeland30.org/home/Enbackground.aspx>
- 673 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,
674 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.,
675 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
676 resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of
677 Remote Sensing*, 34, 2607-2654
- 678 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial
679 analysis for everyone. *Remote Sensing of Environment*, 202, 18-27

- 680 Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland,
681 T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., & Townshend, J.R.G. (2013). High-Resolution Global Maps of 21st-Century
682 Forest Cover Change. *Science*, 342, 850-853
- 683 Herold, M., Mayaux, P., Woodcock, C.E., Baccini, A., & Schmullius, C. (2008). Some challenges in global land cover mapping: An
684 assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, 112, 2538-2556
- 685 Herold, M., See, L., Tsendbazar, N.-E., & Fritz, S. (2016). Towards an Integrated Global Land Cover Monitoring and Mapping System.
686 *Remote Sensing*, 8, 1036
- 687 Herold, M., Woodcock, C.E., Stehman, S.V., Nightingale, J., Friedl, M.A., & Schmullius, C. (2009). The GOF-C-GOLD/CEOS land cover
688 harmonization and validation initiative: technical design and implementation. In, 33rd ISRS. Stresa, Italy
- 689 Huttich, C., Herold, M., Wegmann, M., Cord, A., Strohbach, B., Schmullius, C., & Dech, S. (2011). Assessing effects of temporal
690 compositing and varying observation periods for large-area land-cover mapping in semi-arid ecosystems: Implications for global monitoring.
691 *Remote Sensing of Environment*, 115, 2445-2459
- 692 Jung, M., Henkel, K., Herold, M., & Churkina, G. (2006). Exploiting synergies of global land cover products for carbon cycle modeling.
693 *Remote Sensing of Environment*, 101, 534-553
- 694 Land Cover CCI. (2014). *CCI-LC Product User Guide*. Louvain-la-Neuve, Belgium: UCL-Geomatics
- 695 Land Cover CCI. (2017). *CCI-LC Product User Guide V2.0*. Louvain-la-Neuve, Belgium: UCL-Geomatics
- 696 Lesiv, M., Fritz, S., McCallum, I., Tsendbazar, N., Herold, M., Pekel, J.-F., Buchhorn, M., Smets, B., & Van De Kerchove, R. (2017).
697 Evaluation of ESA CCI prototype land cover map at 20m. *IIASA Working Paper, WP-14-021*
- 698 Lesiv, M., Fritz, S., Moorthy, I., Tsendbazar, N., Herold, M., Van De Kerchove, R., & Smets, B. (2016a). *CGLOPS1: Report describing the*
699 *training dataset used for Dynamic Land Cover 100m product*
- 700 Lesiv, M., Moorthy, I., Fritz, S., Herold, M., Tsendbazar, N.E., Smets, B., & Van De Kerchove, R. (2016b). *GCLOPS: User requirement*
701 *document - Dynamic Land Cover*.
- 702 Mayaux, P., Eva, H., Galleo, J., Strahler, A.H., Herold, M., Agrawal, S., Naumov, S., De Miranda, E.E., Di Bella, C.M., Ordoyne, C.,
703 Kopin, Y., & Roy, P.S. (2006). Validation of the global land cover 2000 map. *Geoscience and Remote Sensing, IEEE Transactions on*, 44,
704 1728-1739
- 705 Mayaux, P., Pekel, J.-F., Desclée, B., Donnay, F., Lupi, A., Achard, F., Clerici, M., Bodart, C., Brink, A., Nasi, R., & Belward, A. (2013).
706 State and evolution of the African rainforests between 1990 and 2010. *Philosophical Transactions of the Royal Society B: Biological*
707 *Sciences*, 368
- 708 Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., & Wulder, M.A. (2014). Good practices for estimating area and
709 assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42-57
- 710 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
711 land-cover validation data set, part I: fundamental design principles. *International Journal of Remote Sensing*, 33, 5768-5788
- 712 Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A.S. (2016). High-resolution mapping of global surface water and its long-term changes.
713 *Nature*, 540, 418-422
- 714 Pengra, B., Long, J., Dahal, D., Stehman, S.V., & Loveland, T.R. (2015). A global reference database from very high resolution commercial
715 satellite data and methodology for application to Landsat derived 30m continuous field tree cover data. *Remote Sensing of Environment*, 165,
716 234-248
- 717 Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A., Freire, S., Halkia, M., Julea, A., Kemper, T., Soille, P., & Syrris, V. (2016). Operating
718 procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. *Publ. Off.*
719 *Eur. Union*
- 720 Romijn, E., Herold, M., Mora, B., Briggs, S., Seifert, F.M., & Paganini, M. (2016). *Monitoring progress towards: Sustainable Development*
721 *Goals The role of land monitoring*. Wageningen,
722 Netherlands. http://www.gofcgold.wur.nl/documents/newsletter/Sustainable_Development_Goals-infobrief.pdf
- 723 Scepan, J., Menz, G., & Hansen, M.C. (1999). The DISCover validation image interpretation process. *Photogrammetric Engineering and*
724 *Remote Sensing*, 65, 1075-1081
- 725 Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C., &
726 Townshend, J.R. (2013). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous
727 fields with lidar-based estimates of error. *International Journal of Digital Earth*, 6, 427-448

- 728 Stehman, S.V. (2014). Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes.
729 *International Journal of Remote Sensing*, 35, 4923-4939
- 730 Stehman, S.V., Olofsson, P., Woodcock, C.E., Herold, M., & Friedl, M.A. (2012). A global land-cover validation data set, II: augmenting a
731 stratified sampling design to estimate accuracy by region and land-cover class. *International Journal of Remote Sensing*, 33, 6975-6993
- 732 Stehman, S.V., Wickham, J., Smith, J., & Yang, L. (2003). Thematic accuracy of the 1992 National Land-Cover Data for the eastern United
733 States: Statistical methodology and regional results. *Remote Sensing of Environment*, 86, 500-516
- 734 Stehman, S.V., Wickham, J.D., Wade, T.G., & Smith, J.H. (2008). Designing a multi-objective, multi-support accuracy assessment of the
735 2001 National Land Cover Data (NLCD 2001) of the conterminous United States. *Photogrammetric Engineering & Remote Sensing*, 74,
736 1561-1571
- 737 Tateishi, R., Hoan, N.T., Kobayashi, T., Alsaaidh, B., Tana, G., & Phong, D.X. (2014). Production of Global Land Cover Data –
738 GLCNMO2008. *2014*, 6
- 739 Tateishi, R., Uriyangqai, B., Al-Bilbisi, H., Ghar, M.A., Tsend-Ayush, J., Kobayashi, T., Kasimu, A., Hoan, N.T., Shalaby, A., Alsaaidh,
740 B., Enkhzaya, T., Gegentana, & Sato, H.P. (2011). Production of global land cover data - GLCNMO. *International Journal of Digital Earth*,
741 4, 22-49
- 742 Tsendbazar, N.-E., de Bruin, S., & Herold, M. (2016a). Integrating global land cover datasets for deriving user-specific maps. *International*
743 *Journal of Digital Earth*, 1-19
- 744 Tsendbazar, N., Herold, M., Fritz, S., & Lesiv, M. (2017). *Copernicus Global Land Operations: Validation Report for Dynamic Land Cover*
745 *100m product: Copernicus Global Land Operations*
- 746 Tsendbazar, N.E., de Bruin, S., Fritz, S., & Herold, M. (2015a). Spatial Accuracy Assessment and Integration of Global Land Cover
747 Datasets. *Remote Sensing*, 7, 15804-15821
- 748 Tsendbazar, N.E., de Bruin, S., & Herold, M. (2015b). Assessing global land cover reference datasets for different user communities. *ISPRS*
749 *Journal of Photogrammetry and Remote Sensing*, 103, 93-114
- 750 Tsendbazar, N.E., de Bruin, S., Mora, B., Schouten, L., & Herold, M. (2016b). Comparative assessment of thematic accuracy of GLC maps
751 for specific applications using existing reference data. *International Journal of Applied Earth Observation and Geoinformation*, 44, 124-135
- 752 Tuanmu, M., & Jetz, W. (2014). A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and*
753 *Biogeography*, 23, 1031-1045
- 754 Verburg, P.H., Neumann, K., & Nol, L. (2011). Challenges in using land use and land cover data for global change studies. *Global Change*
755 *Biology*, 17, 974-989
- 756 Wickham, J., Stehman, S., Fry, J., Smith, J., & Homer, C. (2010). Thematic accuracy of the NLCD 2001 land cover for the conterminous
757 United States. *Remote Sensing of Environment*, 114, 1286-1296
- 758 Xiong, J., Thenkabail, P., Tilton, J., Gumma, M., Teluguntla, P., Oliphant, A., Congalton, R., Yadav, K., & Gorelick, N. (2017). Nominal 30-
759 m Cropland Extent Map of Continental Africa by Integrating Pixel-Based and Object-Based Algorithms Using Sentinel-2 and Landsat-8
760 Data on Google Earth Engine. *Remote Sensing*, 9, 1065
- 761