Copernicus Global Land Operations
"Vegetation and Energy"

"CGLOPS-1"

Framework Service Contract N° 199494 (JRC)

PRODUCT USER MANUAL

MODERATE DYNAMIC LAND COVER 100M
VERSION 1

Issue I1.00

Organization name of lead contractor for this deliverable:

Book Captain: Bruno Smets (VITO)

Contributing Authors: Marcel Buchhorn (VITO)
Myroslava Lesiv (IIASA)
Nandin-Erdene Tsendbazar (WUR)

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document Release Sheet

Book captain:	Bruno Smets	Sign	Date 28.09.2017
Approval:	Roselyne Lacaze	Sign	Date 10.10.2017
Endorsement:	Michael Cherlet	Sign	Date
Distribution:	Public		

Change Record

Issue/Rev	Date	Page(s)	Description of Change	Release
	25.07.2017	All	First released draft	D1.11
D1.11	28.09.2017	All	First issue	I1.00

TABLE OF CONTENTS

List of Figures

List of Tables

List of Acronyms

Acronym	Meaning
ATBD	Algorithm Theoretical Basis Document
CEOS-LPV	Committee of Earth Observation Satellites - Land Product Validation
CCI	Climate Change Initiative
CF	Climate & Forecast conventions
CGLS	Copernicus Global Land service
DEM	Digital Elevation Model
EO	Earth Observation
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
EVI	Enhanced Vegetation Index
FAO	Food and agriculture organization of the united nation
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
GEZ	Global Ecological Zone
GHS	Global Human Settlement
GLIMPSE	GLobal IMage Processing SoftwarE
GSD	Ground Sampling Distance
GUF+	Global Urban Footprint plus
HANTS	Harmonic ANalysis of Time Series
HCM5	Harmonized 5-daily median composite
HSV	Hue Saturation Value colour system
HUE	Chromacity
JRC	Joint Research Center
LAI	Leaf Area Index
LC	Land Cover
LC100	Land Cover map at 100m resolution
LCCS	Land Cover Classification System
LCML	Land Cover Meta Language
madHANTS	Median Absolute Deviations of HANTS
MAE	Mean Absolute Error
MC10	10-daily median composite
MC5	5-daily median composite
MESA	Monitoring Environment and Security in Africa
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Adminstration
NBR	Normalized Burn Ratio
NDVI	Normalized Difference Vegetation Index
netCDF	Network Common Data Form
NIR	Near Infra Red reflectance
NIRv	Near-Infrared reflectance of vegetation

NOVO	Number of valid observations
NPP	Net Primary production
QGIS	Quantum Geographic Information System
PFTs	Plant Function Types
PROBA-V	Vegetation instrument on board of PROBA satellite
PUM	Product User Manual
R&D	Research and Development
REDD+	Reducing Emissions from Deforestation and forest Degradation
RF	Random Forest classifier
RMSE	Root Mean Square Error
SIPI	Structure Intensive Pigment Index
SM	Status Map
SRTM	Shuttle Radar Topography Mission
SVP	Service Validation Plan
SWIR	Short Wave Infra Red reflectance
TOC	Top Of Canopy
UN	United Nations
USGS	United States Geological Survey

EXECUTIVE SUMMARY

The Copernicus Global Land Service (CGLS) is earmarked as a component of the Land service to operate “a multi-purpose service component” that provides a series of bio-geophysical products on the status and evolution of land surface at global scale. Production and delivery of the parameters take place in a reliable, automatic and timely manner and are complemented by the constitution of long-term time series.

From 1st January 2013, the Copernicus Global Land Service is providing continuously Essential Climate Variables like the Leaf Area Index (LAI), the Fraction of Absorbed Photosynthetically Active Radiation absorbed by the vegetation (FAPAR), the surface albedo, the Land Surface Temperature, the soil moisture, the burnt areas, the areas of water bodies, and additional vegetation indices, are generated every hour, every day or every 10 days on a reliable and automatic basis from Earth Observation satellite data.

The Dynamic Land Cover map at 100 m resolution is a new product in the portfolio of the CGLS and targets to deliver a yearly global land cover map at 100 m spatial resolution. Land cover plays a major role in the climate and biogeochemistry of the Earth system. The CGLS Land Cover product provides a primary land cover scheme with 18 classes. Next to these classes, the product also includes a set of four vegetation continuous field layers that provide proportional estimates for vegetation cover for the land cover types forest, herbaceous vegetation, shrub and bare ground. This continuous classification scheme may depict areas of heterogeneous land cover better than the standard classification scheme and, as such, can be tailored for application use (e.g. forest monitoring, crop monitoring, biodiversity and conservation, monitoring environment and security in Africa, climate modelling, etc.)

This first Land Cover map (V1.0) is provided for the 2015 reference year over the African continent, derived from the PROBA-V 100 m time-series, a database of high quality land cover training sites and several ancillary datasets.

1 BACKGROUND OF THE DOCUMENT

1.1 SCOPE AND OBJECTIVES

This Product User Manual (PUM) is the primary document that users have to read before handling the products.

It gives an overview of the product characteristics, in terms of algorithm, technical characteristics, and main validation results.

1.2 CONTENT OF THE DOCUMENT

This document is structured as follows:

- Chapter 2 summarizes the retrieval methodology,
- Chapter 3 describes the technical properties of the product,
- Chapter 4 summarizes the results of the quality assessment,
- Chapter 5 lists all references to cited literature

The users' requirements are recalled in Annex 1: Review of Users Requirements.

1.3 RELATED DOCUMENTS

1.3.1 Applicable documents

AD1: Annex I – Technical Specifications JRC/IPR/2015/H.5/0026/OC to Contract Notice 2015/S 151-277962 of 7th August 2015

AD2: Appendix 1 – Copernicus Global land Component Product and Service Detailed Technical requirements to Technical Annex to Contract Notice 2015/S 151-277962 of 7th August 2015

AD3: GIO Copernicus Global Land – Technical User Group – Service Specification and Product Requirements Proposal – SPB-GIO-3017-TUG-SS-004 – Issue I1.0 – 26 May 2015.

1.3.2 Input

Document ID	Descriptor
CGLOPS1_SSD	Service Specifications of the Global Component of the Copernicus Land Service.
CGLOPS1_SVP	Service Validation Plan of the Global Component of the Copernicus Land Service
CGLOPS1_URD_LC100m	User Requirements Document of the Dynamic land

	cover 100m product
CGLOPS1_TrainingDataReport_LC100m	Report presenting the training data set used for the Dynamic land cover 100m product
CGLOPS1_ATBD_LC100-V1	Algorithm Theoretical Basis Document of the Dynamic land cover 100m product
CGLOPS1_VR_LC100_V1	Report describing the results of the scientific quality assessment of the Dynamic land cover 100m product

1.3.3 External documents

PROBA-V	http://proba-v.vgt.vito.be/
PROBA-V User Manual	User Guide of the PROBA-V data, available on http://www.vito-eodata.be/PDF/image/PROBAV-Products_User_Manual.pdf

2 ALGORITHM

2.1 OVERVIEW

The CGLS Dynamic Land Cover Map at 100 m resolution (CGLS LC100) product is generated by combining several proven individual methodologies through:

1. Data cleaning and outlier detection techniques,
2. Applying data fusion techniques at multiple levels,
3. Supervised classification through collecting reference data, including crowdsourcing techniques,
4. Including established third party datasets via expert rules.

The workflow, shown in Figure 1, can be divided into the following sections:

1. data cleaning & compositing,
2. data fusion,
3. metrics extraction,
4. training data generation,
5. ancillary datasets preparation,
6. classification / regression,
7. cover fraction layer generation,
8. land cover map generation.

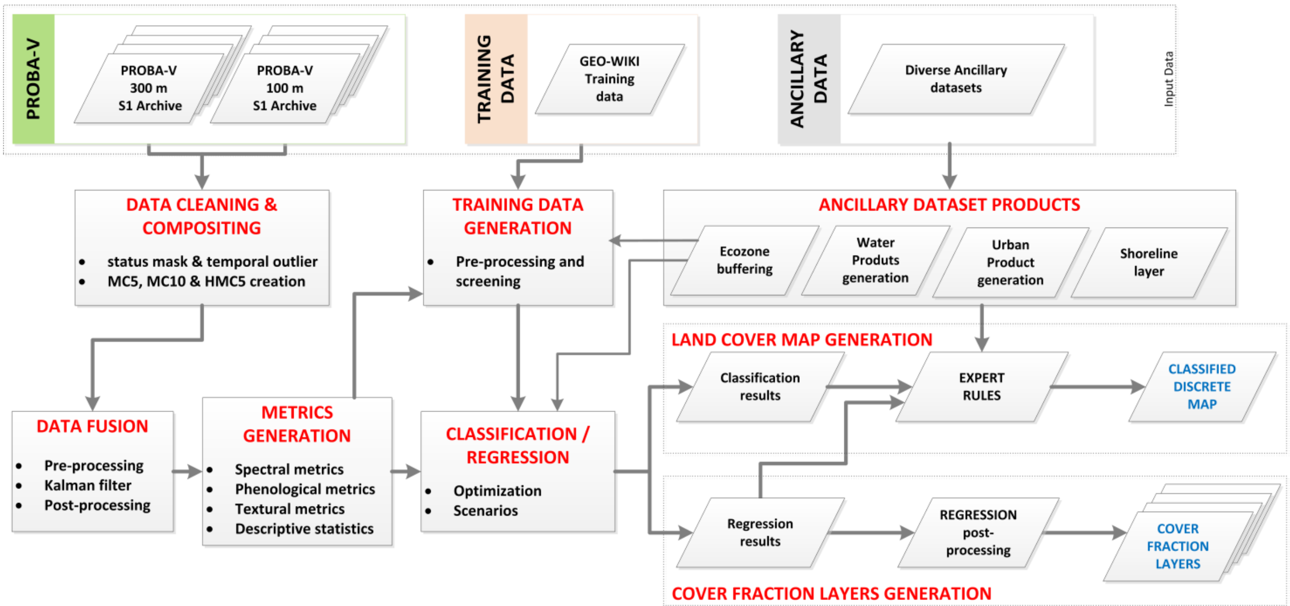


Figure 1: Workflow diagram for the CGLS Dynamic Land Cover 100m product for Africa 2015

To generate the product, 5-daily PROBA-V multi-spectral image data with a ground sampling distance (GSD) of ~0.001 degree (~100 m) is used as primary earth observation (EO) data, and PROBA-V daily multi-spectral image data with a GSD of ~0.003 degree (~300 m) secondarily. Next to a status mask cleaning using the internal quality flags of the PROBA-V EO data, a temporal cloud and outlier filter built on a Fourier transformation is applied to clean the data. Next, the 5-daily PROBA-V 100 m and daily 300 m datasets are fused using a Kalman filtering approach. The Kalman-filled 100 m data set is then automatically checked for consistency before extracting several metrics. Therefore, a harmonic model is fitted through each of the reflectance bands of the time series data as well as each of the additional derived vegetation indices for each time series step. Next to the parameters of the harmonic model which are used as metrics for the overall level and seasonality of the time series, descriptive statistics are extracted for the reference year as well as for the vegetation season and off-season within that reference year using phenological parameters (e.g. start- and end of season) extracted from the harmonic model itself. Overall, 392 metrics are extracted from the PROBA-V EO data.

The training data is collected through manual classification using Google Maps and Bing images at 10 m spatial resolution using the Geo-Wiki Engagement Platform (http://www.geo-wiki.org/). Therefore the training data not only includes the land cover type, but also the cover fractions of the main land cover classes in PROBA-V 100 m resolution. In the classification preparation, the metrics of the training points are analysed for intra- and inter- specific outliers, as well as screened for the best metrics combinations to run an optimized classification. The optimized training data is then used in a supervised classification using Random Forest techniques.

Finally, we build upon the success of previous global mapping efforts and/or other ancillary datasets. Therefore, the external datasets are resampled to PROBA-V 100 m spatial resolution and included via expert rules in the land cover map generation step. The produced land cover map uses a hierarchical legend based on the United Nations and Cover Classification System (LCCS). Compatibility with existing global land cover products is hereby taken into account. A novelty of this product is the generation of vegetation continuous fields that provide proportional estimates for vegetation cover for trees, herbaceous vegetation, shrub and bare ground. The input are the cover fractions collected for all training points which are used in a Random Forest regression. The validation is performed according to CEOS-LPV protocols.

This chapter provides a summary of the retrieval method. More details can be found in [CGLOPS1_ATBD_LC100-V1].

2.2 THE RETRIEVAL METHODOLOGY

2.2.1 Data cleaning

Input data for the generation of the CGLS Dynamic Land Cover Map at 100 m resolution is EO data from the PROBA-V multi-spectral satellite (Francois et al., 2014). PROBA-V products include processed top-of-canopy (TOC) 4-band reflectance data in the blue, red, near-infrared (NIR), and shortwave infrared (SWIR) wavelength spectrum. TOC reflectance data is also already geometric and radiometric corrected and delivers quality indicators for each pixel via a status mask (SM). The SM not only include information regarding the radiometric quality of the pixel, but also the information of the PROBA-V cloud detection and retrieval algorithm (Sterckx et al., 2014; Dierckx et al., 2014; Wolters et al., 2017).

In the first step, PROBA-V multi-spectral TOC image data with a ground sampling distance (GSD) of ~0.001 degree (~100 m) and PROBA-V daily multi-spectral TOC image data with a GSD of ~0.003 degree (~300 m) is retrieved from the S1 (daily) Collection 1 archive for the African continent for the reference year 2015 plus 3 months before and after the reference year. They are first cleaned using the status mask information to remove the pixels flagged as noise, cloud, or sea. Then, an additional temporal filter called madHANTS, built on a Fourier transformation based on HANTS (Harmonic ANalysis of Time Series) algorithm (Verhoef, 1996; Roerink et al., 2000) and outlier test based on median absolute deviations (Walker, 1931), is applied to clean the time series from remaining haze and undetected clouds (see Figure 2).

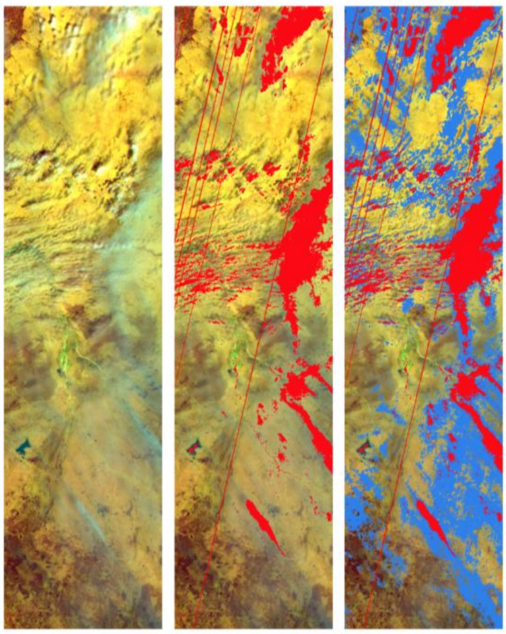


Figure 2: Example for data cleaning process of PROBA-V 100m image from 2016-03-06. Image is shown as false color composite (RGB = swir, nir, blue) for a sample area in tile X18Y06 (Nigeria) – left: raw image; middle: status masked cleaned image (quality flagged areas are shown in red); right: final cleaned image (additional pixels which are flagged as outliers or clouds are shown in blue).

The next step includes the generation of 5-daily (MC5) and 10-daily (MC10) median composites for the 100 m and 300 m cleaned data for the reference year. This step is needed to gather regular time steps in the time series. For the 300 m PROBA-V data also a harmonized 5-daily median composite (HMC5) for the whole PROBA-V data archive (2013 – 2017) is produced. Therefore, the harmonized time series output produced by the HANTS algorithm is used.

2.2.2 Data fusion

High seasonal cloud coverage in several African regions are challenging for all optical based land cover classification approaches. In order to overcome the low data density and therefore data gaps in the PROBA-V 100 m MC5 time series product, PROBA-V 300 m data which has a daily revisit time is fused in via a Kalman filtering (Kalman, 1960).

In the data fusion pre-processing, small gaps (5 – 10 days) in the 100 m and 300 m MC5 time series products are filled with the pixel values of the MC10 time series products for the corresponding time steps. In a second step, bigger gaps (> 25 days) in the 300 m MC5 time series product are filled via interpolation with the HMC5 300 m long term trend product for the corresponding time series steps (see Figure 3). This is needed in order to guide the Kalman filtering approach in cases where no PROBA-V 100 m and 300 m MC5 data is available for more than 1 month in the row for a pixel.

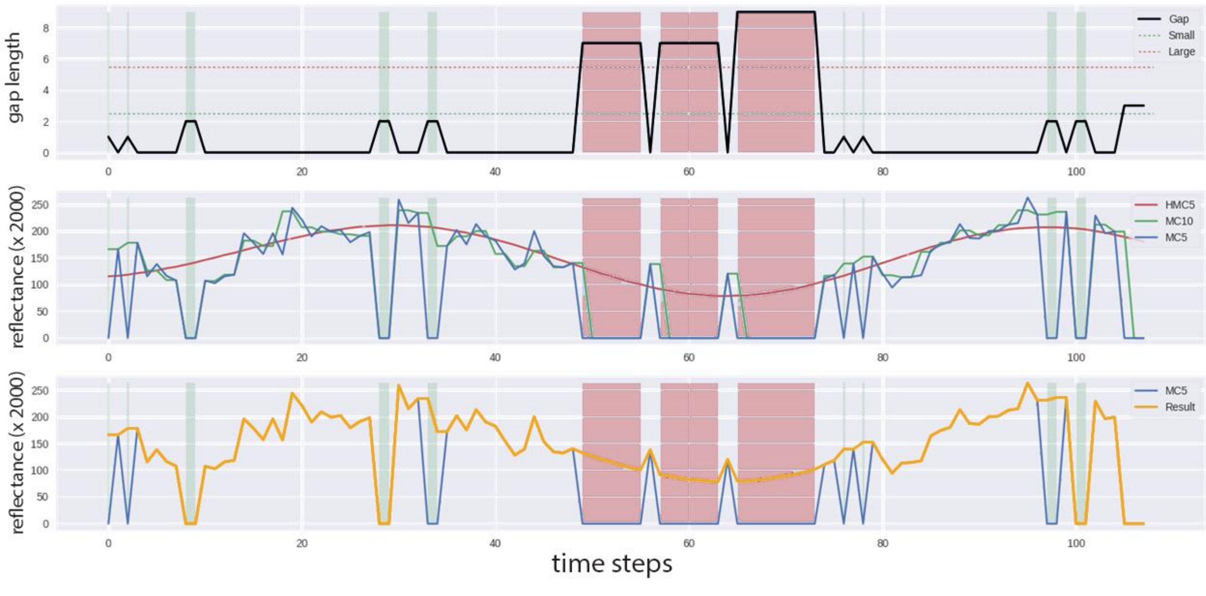


Figure 3: Example for data fusion pre-processing results for PROBA-V 300m MC5 time series for pixel location 9.459° lon, 6.562° lat. Top: continuous gap length in the time series (dashed green line indicates threshold for small gaps, dashed red line indicates threshold for big gaps); middle: time series before pre-processing (blue line shows original 5-daily median composite time series, green line shows original 10-daily median composite time series, red line shown the long term harmonized time series for the full PROBA-V archive); bottom: 300m time series after pre-processing (orange line shows the final pre-processed time series which will be used for the data fusion).

The last step in the data fusion is the post-processing in which the Kalman filled PROBA-V 100 m MC5 data is screened using the madHANTS algorithm to remove introduced outliers. Output of the data fusion is a consistent 100 m time series for the reference year in 5-days intervals and with gap filled TOC reflectance data in the blue, red, NIR and SWIR wavelength region. An example for tile X18Y06 (Nigeria) before and after applying the Kalman filtering approach is shown in Figure 4.

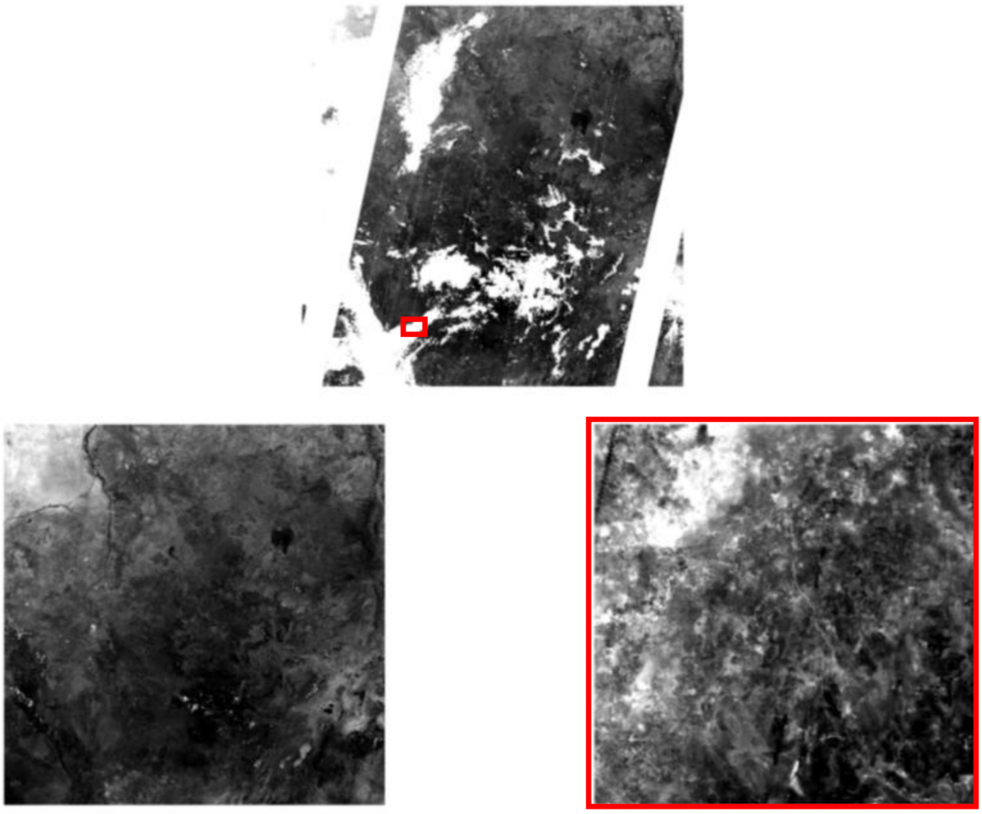


Figure 4: top) PROBA-V 100 m MC5 pre-processed red reflectance image (areas with missing data are shown in white), bottom-left) PROBA-V 100 m MC5 image after gap filling by Kalman approach, bottom-right) zoom in to full PROBA-V resolution of left image of red box shown in top figure (red box had no data at all and shows after Kalman filter approach consistent image data). Example over test area in tile X18Y06 (Nigeria), on 2016-03-06.

2.2.3 Metrics extraction

Before the metrics extraction, two additional pre-processing steps are needed. In the first, additional vegetation indices for each time step in the PROBA-V 100 m Kalman filled MC5 time series is generated:

1. Normalized Difference Vegetation Index (NDVI) using the red and the NIR reflectance bands (Tucker, 1979),

2. Enhanced Vegetation Index (EVI) using the blue, red, and NIR reflectance bands (Huete, et al., 2002),
3. Structure Intensive Pigment Index (SIPI) using the blue, red, and NIR reflectance bands (Blackburn, 1998),
4. Normalized Burn Ratio (NBR) using the NIR and the SWIR reflectance bands (Key and Benson, 2005),
5. HUE of the HSV transformation using the red, NIR and SWIR reflectance bands (Pekel et al., 2014),
6. VALUE of the HSV transformation using the red, NIR and SWIR reflectance bands (Pekel et al., 2014),
7. Near-Infrared reflectance of vegetation (NIRv) using the red and the NIR reflectance bands (Badgley et al., 2017).

In second, three masks are generated:

1. Phenology mask for each pixel showing start and end date of up to two seasons within the reference year using GLIMPSE module PHENOdef which is part of the SPIRITS (Software for the Processing and Interpretation of Remotely sensed Image Time Series) software (Eerens et al., 2014).
2. Time series gap mask (tGAPmask) for the PROBA-V 100m MC5 data (before Kalman filling) showing areas with data gaps longer than 30 – 60 days in the vegetation season, data gaps longer than 60 – 90 days in the off-vegetation season, or data gaps longer than 90 days in the reference year. This mask is later used to evaluate the influence of the Kalman filling approach on the data quality.
3. Number of valid observations in reference year (NOVO) mask which shows for each pixel how many valid PROBA-V 100 m MC5 images (cloud/outlier free observations) exists in the time series for the reference year before applying the Kalman filter (max. 72).

For the metrics extraction, a harmonic model is fitted through each of the reflectance bands of the PROBA-V 100 m Kalman filled MC5 time series data as well as each of the additional derived vegetation indices time series sets. The harmonic model is again based on the HANTS algorithm using a Fourier transformation (Verhoef, 1996; Roerink et al., 2000). The seven parameters of the harmonic model are used as metrics for the overall level and seasonality of the time series.

Moreover, descriptive statistics of the time series, such as the mean, standard deviation, minimum, maximum, minimum-maximum range, sum, median, 10^{th} percentile, and 90^{th} percentile, are extracted for the reference year as well as for the vegetation season and off-season within that reference year using the phenology mask. These overall 27 descriptive metrics are extracted for each of the four reflectance bands of the PROBA-V 100 m Kalman filled MC5 time series data as well as for each of the additional derived vegetation indices time series sets. An additional descriptive metric is calculated using a 3x3 moving window calculating the standard deviation of the box for all calculated median statistics for the full reference year (4 reflectance bands, 7 vegetation indices). This metric can be interpreted as textural metric representing the uniformity of the pixel in its box (low values show a homogeneous area, high values a more heterogeneous land cover).

Additional metrics include the phenological parameters for start and end of season (for maximal two seasons in the reference year), the seasonality index indicating if a pixel has a seasonality overall, and the length of the vegetation season (if more than one vegetation season exists than these lengths are combined). Moreover, the Water Bodies Potential Mask (WBPM) (Bertels et al., 2016) is used as a topographic parameters/metric indicating if a pixel could be possible an water body.

Overall, 392 metrics (28 descriptive metrics and 7 harmonic metrics for the 11 time series sets (4 reflectance bands, 7 vegetation indices) plus 7 additional metrics) have been generated from the PROBA-V 100 m Kalman filled MC5 time series data and are input in the classification/regression step of the automated processing chain.

2.2.4 Training data generation

Training data has been collected through the Geo-Wiki engagement platform. A new branch of Geo-Wiki (http://geo-wiki.org/) was developed for collecting reference data at the required resolution and grid (PROBAV-100m pixels). It shows the pixels to be interpreted on top of Google Earth and Bing imagery, where each pixel is further subdivided into 100 sub-pixels of approximately 10m x 10m each. Using visual interpretation of the underlying very high resolution imagery, experts interpret each sub-pixel based on the land cover type visible, which includes trees, shrubs, water objects, arable land, burnt areas, etc. This information is then translated into different legends using the UN LCCS (United Nations Land Cover Classification System) as a basis [CGLOPS1_URD_LC100].

The distribution of sample sites is systematic, with the same distance between sample sites, which is approximately 35 km. However, land cover data are not collected at every sample site as the frequency depends on the heterogeneity of land cover types by region and availability of valid PROBA-V 100m imagery.

In total, the experts have classified almost 24,000 unique locations [CGLOPS1_TrainingDataReport_LC100]. The quality of the data has been checked by revisiting locations that were either inter – or intra- land cover class outliers from remote sensing perspective. Classifications that were wrong or those impossible to identify land cover class by visual interpretation were removed. Final training dataset consists of circa 20,000 sample sites.

2.2.5 Ancillary dataset preparation

Three ancillary datasets are included next to the classification and regression results in the CGLS LC100 land cover map generation:

1. Shoreline vector layer,

2. Permanent water body mask, temporarily water body mask, and herbaceous wetland mask (called WetProducts dataset),
3. Urban mask.

The shoreline layer is mainly used to distinguish between open land water and open sea water. We used the 30 m shoreline vector layer of the U.S. Geological Survey (USGS) which was produced from Landsat 7 EO data for the Africa Ecosystem Project (Sayre et al., 2013). The shoreline layer was only resampled to the PROBA-V 100 m spatial resolution. Since this vector layer cannot be used as a definitive shoreline boundary, we included this dataset also in the WetProducts generation algorithm in order to detect shoreline changes as well as false detections.

The WetProducts dataset contains the permanent and temporary water body mask as well as the herbaceous wetland mask. The automatic detection of permanent water bodies following Bertels et al. (2016) algorithm was improved using the water seasonality layer from the Global scale Water History Record (GWHR) (Pekel et al., 2016) while the maximum water extent (all the locations ever detected as water over the Landsat data archive period) was mainly used as an exclusion layer for non-water areas. Both layers were first resampled from the 30 m Landsat resolution to the PROBA-V 100 m spatial resolution. The location of temporary water bodies and herbaceous wetlands was solely computed by using Bertels et al. (2016) algorithm.

The urban mask was generated through the combination of DLR's Global Urban Footprint Plus layer (GUF+) for 2015 (Marconcini et al., 2017a, Marconcini et al., 2017b) and JRC's Global Human Settlement Layer (GHS) for 2014 (Pesaresi et al., 2015). The GUF+ layer used mainly Sentinel-1 radar data in combination with Landsat-8 multispectral optical data to detect urban structures with a spatial resolution of 10 m, where the GHS built-up grid used the 30 m Landsat EO data archive to generate a human settlement layer with a spatial resolution of 38 m. Both layers had to be resampled to the PROBA-V 100 m spatial resolution in the first step. Secondly, the GUF+ and GHS layers have been fused whereby missing urban areas in the GUF+ layer have been incorporated from the GHS layer (mainly needed for islands).

2.2.6 Classification / regression

In order to adapt the classification/regressor algorithm to continental patterns, the classification/regression of the data is carried out in ecozones defined by the global ecological zone (GEZ) dataset for 2010 of the Food and Agriculture Organization of the United Nations (FAO) (FAO, 2012). The GEZ layer was also used to subset the training data. Moreover, the tGAPmask was applied to ensure that all training points have sufficient time series data.

Next, the training data for each ecozone are screened for inter-class outliers by analysing the spectral angle as well as the root mean square error (RMSE) of the metrics of all training points within their land cover class and compared to all other land cover classes. All inter-class outliers with an impact score over 50% are removed from further processing. The last optimization step is the hyper-parameter search for each training data set. For that, we used a combined grid and random search with a five folded cross-validation to identify the optimal model parameter for each training data set.

For the supervised classification, the Random Forest classification was conducted for each ecozone independently using the GEZ dataset to split up the input data, and using the ecozone-specific generated training data sets and hyper-parameters. Next to the classification results showing the discrete class for each pixel, also the predicted class probability for each pixel is generated. Overall three Random Forest classification scenarios for each ecozone with different settings are carried out:

1. "pure class" scenario: in this setting only training points with a cover percentage over 95% in the bare, herbaceous vegetation, shrub, forest, or agriculture class are used. This setting can be interpreted as endmember selection (extreme sample reduction in terms of purity) and classification. A pixel's metrics/spectral profile is matched to the metric/spectral signature of a specified land cover type (endmember). By incorporating the predicted class probability, the pixels with "pure" land cover classes can be identified (e.g., a pixel classified as forest with 90% predicted class probability would mean that the classifier is to 90% certain that the pixel is forest with a minimum of 95% of forest cover).
2. "discrete class" scenario: in this setting all training points are used.
3. "forest type" scenario: in this setting only training points with a forest cover percentage over 15% and a valid forest type attribute are used. The resulting map is therefore forest type map and later used to subdivide the forest class.

A novelty of the CGLS LC100 product is the generation of vegetation continuous fields that provide proportional estimates for vegetation cover for trees, herbaceous vegetation, shrub and bare ground. The input are the cover fractions collected for all training points which are used in a Random Forest regression. Overall five (Forest, Shrub, Herbaceous vegetation, Bare ground, Agriculture) regression scenarios for each ecozone are carried out using, in the regression model, the respective vegetation cover percentages of the training points. Note that the agriculture cover fraction map is only used to create an agriculture mask and will be not delivered as a cover fraction layer in the final product.

2.2.7 Cover fraction layer generation

Next to the discrete land cover map, the cover fraction layers for forest, shrubs, herbaceous vegetation, and bare ground are part of the CGLS LC100 product. These cover fraction layers, indicating the proportional estimates of land cover for the specific land cover type, are generated out of the Random Forest regression results (see section 2.2.6). The main processing step is a linear normalization for pixels with a cover fraction sum of the five regression results with more than 100 %. In detail, pixels with an overall sum over 100 % of the single cover fractions results in the forest, shrub, herbaceous vegetation, agriculture and bare ground regression cycles are proportional scaled that their sum is 100 %. Moreover, the permanent water body mask and urban mask are incorporated by setting the pixel values for all cover fractions in these areas to 0 %.

In a final step, CF V1.6 metadata attributes compliant with version 1.6 of the Climate & Forecast conventions (CF V1.6) and the colour bars translating the vegetation continuous fields code into the legend are injected. Figure 5 shows the four cover fraction layers for the land cover types forest, shrub, herbaceous vegetation and bare.

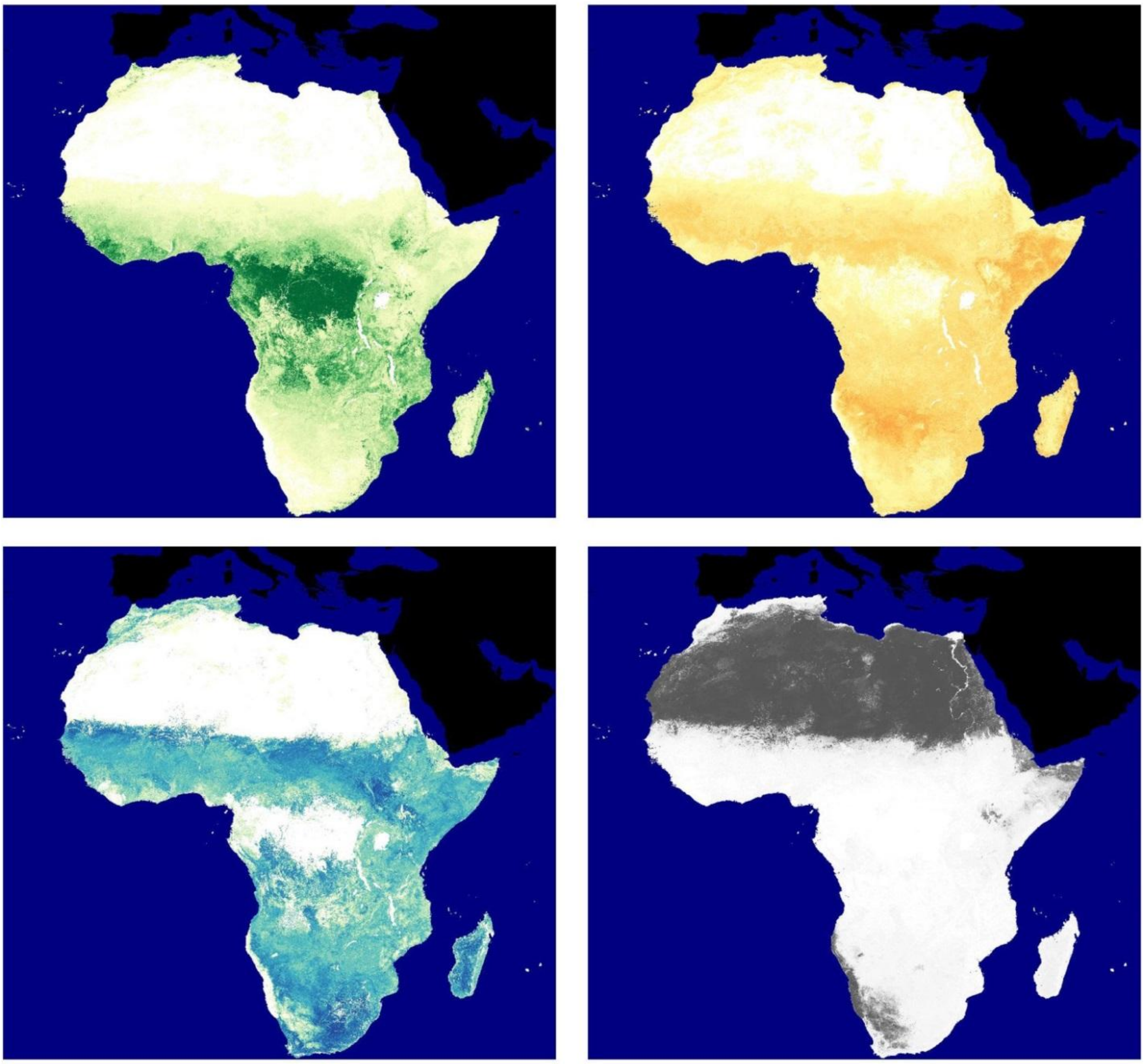


Figure 5: The cover fraction layers for the forest, shrubs, herbaceous vegetation and bare land cover classes of the CGLS Dynamic Land Cover Map at 100 m product for Africa 2015 (shown at continental scale).

2.2.8 Land Cover map generation

Expert rules are applied to combine the existing knowledge represented by the ancillary datasets (section 2.2.5) and the classification and regression results (section 2.2.6). In order to incorporate the vegetation cover fraction layers, a discrete map was generated by applying the training data rules on the forest, herbaceous vegetation, shrub, bare ground and agriculture cover fraction layers. In detail, during the training data generation (see section 2.2.4) a set of rules has been established to assign a training point with its cover fraction percentages to a discrete class (e.g. training point with cover percentages of 65 % forest and 35 % shrubs is classified as an "open forest" training point) [CGLOPS1_TrainingDataReport_LC100]. Moreover, the agriculture cover fraction layer was used to generate an agriculture mask by applying a threshold.

The following datasets are used to generate the final CGLS Dynamic Land Cover Map at 100 m resolution (CGLS LC100):

1. Random Forest classification result of the "pure class" scenario;
2. Predicted class probability layer of the Random Forest classification result of the "pure class" scenario;
3. Random Forest classification result of the "discrete class" scenario;
4. Predicted class probability layer of the Random Forest classification result of the "discrete class" scenario;
5. Random Forest classification result of the "forest type" scenario;
6. Discrete map generated from the five normalized cover fraction layers;
7. Agriculture mask generated from the normalized agriculture cover fraction layer;
8. Maximum natural vegetation mask generated from the normalized cover fraction layers for forest, herbaceous vegetation, shrub and bare ground;
9. "number of valid observations" mask showing pixels with no PROBA-V 100 m observations in the whole reference year (novo mask);
10. PROBA-V 100m urban mask (urban mask);
11. WetProduct layer including the permanent water body mask, temporary water body mask, and herbaceous wetland mask (WetProducts); and
12. PROBA-V 100m Shoreline mask (Shoreline).

The predicted class probabilities are used as thresholds in the decision tree designed expert rules in order to generate the 18 class discrete map product. Therefore, only pixels with a predicted class probability over 90 % are used from the "pure class" classification, and pixels with a predicted class probability over 50 % from the "discrete class" classification. The discrete map generated from the five cover fraction layers is used in areas were the predicted class probability from the "discrete class" classification is under 50 % and equals with the cover fraction discrete map class. Hard masks like the agriculture mask, NOVO mask, permanent water body mask, temporary water body mask, herbaceous wetland mask, and urban mask were used directly as they are and could overwrite the classification results. Remaining land mass pixels with no discrete class assigned by the decision tree were filled with the results of the "discrete class" classification. The "forest type" classification results were then used to separate the discrete classes "closed forest" and "open forest” into the different forest type classes. The shoreline layer was mainly used to separate inland water bodies from the open sea class.

In a final step, metadata attributes compliant with version 1.6 of the Climate & Forecast conventions (CF V1.6) and the colour bars translating the discrete class code into the legend are injected. Moreover, the probability layer indicating the classifier certainty was produced out of the predicted class probabilities of the classification results. Figure 7 shows a screenshot of the legend for the discrete map with 18 classes, where Figure 6 shows an overview of the product on continental scale.

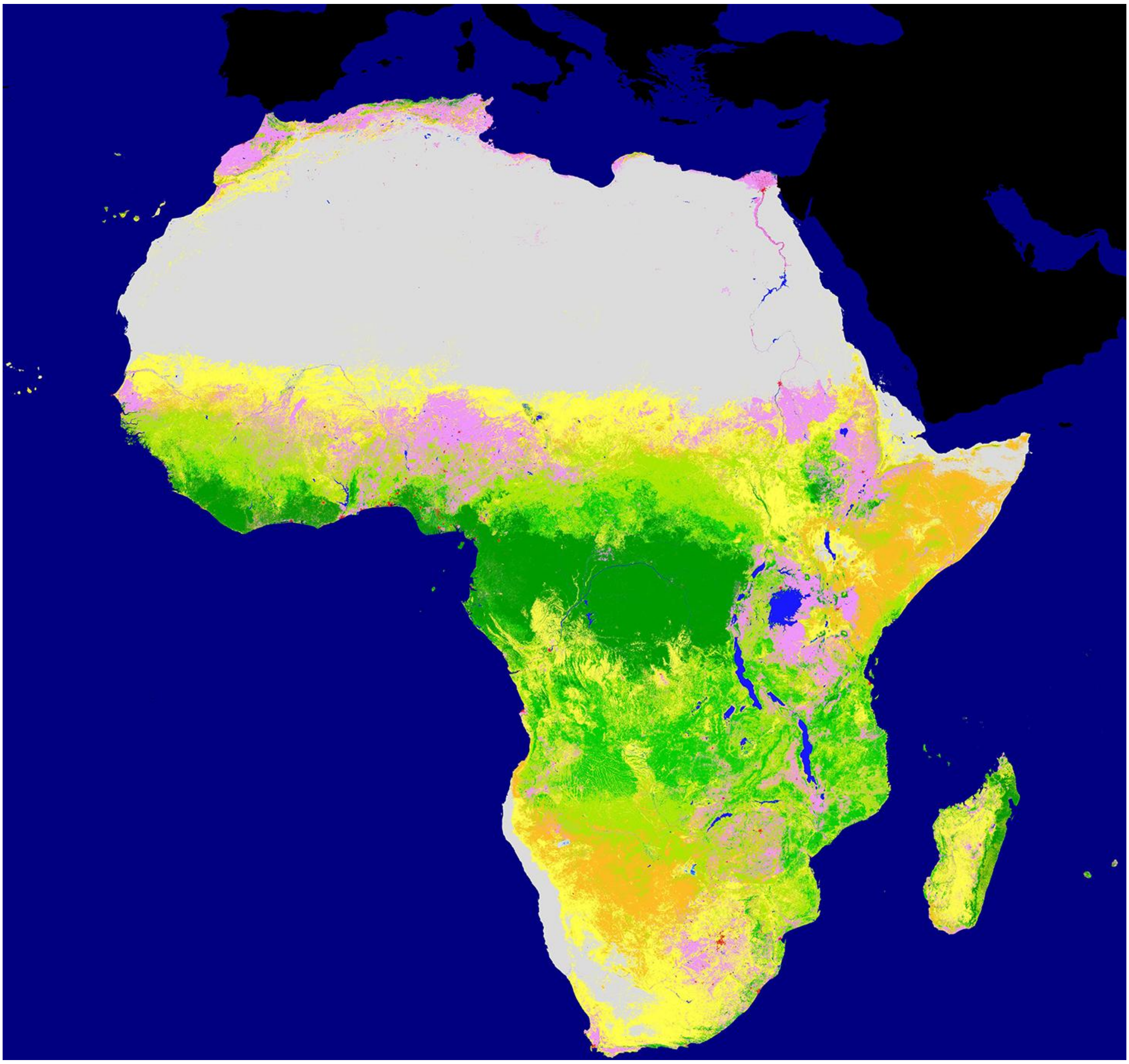


Figure 6: The CGLS Dynamic Land Cover Map at 100 m for Africa 2015 with 18 discrete classes (shown at continental scale).

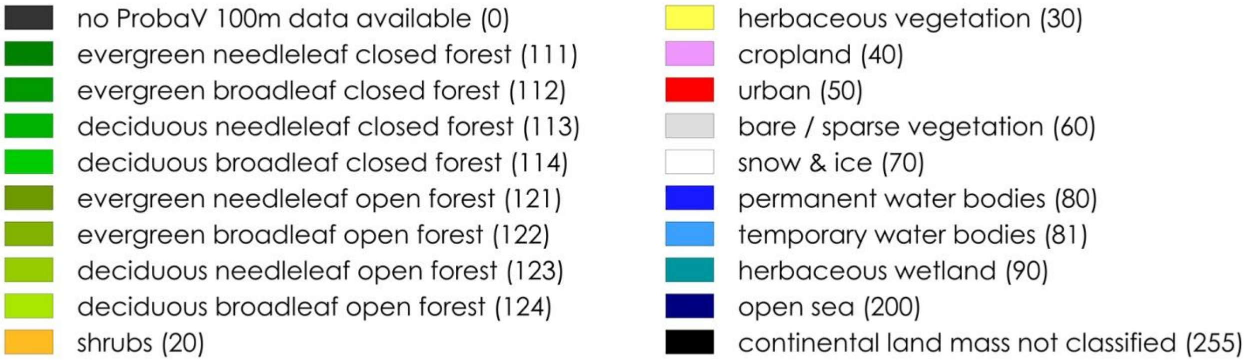


Figure 7: Legend for the 18 discrete classes of the CGLS Dynamic Land Cover Map at 100 m for Africa 2015. Note: the number in brackets represents the numerical code for a land cover class.

2.3 Limitations of the Method

- Remaining shadowed pixels in the time series not filtered out during the data cleaning process can lead to misclassifications.
- Fires (burned areas) were not yet taken into account and therefore could lead to misclassifications.
- Artefacts in the phenological cycle detection can lead to misclassifications.
- Artefacts at boundaries of ecozone can appear due to the used ecozone vector layer as well as the ecozone-specific generated hyper-parameter for the Random Forest classifier and regressor.
- Highly fragmented landscapes, in particular mixed areas with very small cropland fields (less < 0.5 ha), are very difficult to map because of the resolution of 100m (i.e. Nigeria, Ghana). This could lead to overestimate the croplands.
- Areas with low cropland fragmentation (very sparse cropland fields of a very small size) are difficult to map because of the resolution of 100m. This could lead to underestimate the croplands.
- Very small African villages are difficult to map, especially when not detected by the GUF+ layer at 12 m resolution, which could lead to an underestimate of urban.
- Some limitations are due to the legend or class definition:
  - In the southern part of Africa, there are huge areas with kind of tundra type of vegetation, NDVI values are very low in these areas and can confuse the classifier to misclassify between herbaceous vegetation or bare land.
  - In Africa, there are a lot of riparian forests which are evergreen. A lot of pixels were noticed with mixed deciduous trees and riparian evergreen forest which can confuse the classifier to misclassify the forest type.

2.4 DIFFERENCES WITH THE PREVIOUS VERSION

This is the initial map, so no previous versions exist.

2.5 ROADMAP

The Collection 100m of LC product is currently generated from the PROBA-V sensor data.

The Copernicus Global Land service will continue the 100m production through the Sentinel-2 mission. The adaptation of the retrieval methodology to Sentinel-2 is planned and will be performed in 2018 using both Sentinel-2A and Sentinel-2B data. In a first phase, the Sentinel-2 data will be merged with the PROBA-V data. Once the PROBA-V data is not available anymore, the classification will be based on Sentinel-2 data to maintain the yearly classification at global scale.

In the near future, it is planned to provide tools to enable the user to customize the classification to their application, as is:

- QGIS plug-in to perform Plant Function Type conversion.
- QGIS plug-in to create a distinct discrete classification based on the available layers.

3 PRODUCT DESCRIPTION

3.1 FILE NAMING

The **dynamic moderate Land Cover (LC)** Collection **100m** product follows the naming standard:

c_gls_LC100-<layer>_<YYYYMMDDHHMM>_<AREA>_<SENSOR>_<VERSION>

where

- <layer> gives the name of the actual represented image (see 3.2)
- <YYYYMMDDHHMM> gives the temporal location of the file. YYYY, MM, DD, HH, and MM denote the year, the month, the day, the hour, and the minutes, respectively. The reference date is the first day of the year (01 Jan).
- <AREA> gives the spatial coverage of the file. In our case, <AREA> is AFRI, short name for Africa continent.
- <SENSOR> gives the name of the sensor used to retrieve the product, with PROBAV referring to PROBA-V
- <VERSION> shows the processing line version used to generate this product. The version denoted as M.m.r (e.g. 1.0.1), with 'M' representing the major version (e.g. V1), 'm' the minor version (starting from 0) and 'r' the production run number (starting from 1) (see Table 1).

Table 1: Explanation in version numbering and recommendations for using efficiently the products.

Versions	Differences	Recommendations
Major	Significant change to the algorithm.	Do not mix various major versions in the same applications, unless it is otherwise stated.
Minor	Minor changes in the algorithm	Can be mixed in the same applications, but require attention or modest modifications
Run	Fixes to bugs and minor issues. Later run automatically replaces former	Consider it as a drop-in replacement

3.2 FILE FORMAT

The LC100 product is delivered as a set of separate files. It has been decided not to combine multiple layers into a single file due to the size when moving to global scale.

The LC100 product is provided in two different formats:

- single-band Network Common Data Form version 4 (netCDF4) file with metadata attributes compliant with version 1.6 of the Climate & Forecast conventions (CF V1.6).
- single-band GEOTIFF file with metadata attributes compatible with the netCDF4 attributes

The LC100 product contains the following files (LAYER):

- LCCS, the discrete Land Cover Classification System according the definition of FAO
- LCCS-PROB, the probability per pixel of the discrete classification
- LCCS-QFLAG, a binary mask to indicate if the pixel was filled with 300m data
- COV-BARE, the cover fraction layer for bare
- COV-FOREST, the cover fraction layer for forest
- COV-GRASSLAND, the cover fraction layer for grass
- COV-SHRUB, the cover fraction layer for shrub
- OCCUR-WB, the occurrence of water bodies

Next to the image files, some ancillary data files are also part of the LC100 product:

- an xml file containing the metadata conform to INSPIRE2.1
- for a more user-friendly view of the XML contents in a browser, an XSL transformation file can downloaded at http://land.copernicus.eu/global/sites/default/files/xml/c_gls_ProductSet.xsl.
This file should be placed in the same folder as the XML file.
- a quicklook in a coloured GeoTIFF format. The quicklook sub-sampled to 5% in both horizontal and vertical direction from the LCCS layer.

3.3 PRODUCT CONTENT

3.3.1 Data File

As explained in section 3.2, the LC100 product consists of set of separate files, each representing one information layer. The following paragraphs describe these layers in more detail.

3.3.1.1 Discrete classification (LCCS)

The discrete classification map was defined using the Land Cover Classification System (LCCS) developed by the United Nations (UN) Food and Agriculture Organization (FAO) providing 18 classes (Table 2) and is coded as byte. The UN-LCCS system was designed as a hierarchical classification, which allows adjusting the thematic detail of the legend to the amount of information available. The “level 1” legend, as presented in Table 2, contains these classes associated with the ten values code (i.e. class codes of 10, 20, 30, etc.). The “level 2”, also known as regional legend, provides a higher level of detail and is associated with non-ten values (i.e. 11, 12, 21, 22, etc.). The “level 3” further details some legends and is associated to hundred-type values (i.e. 111 – 114 and 121 – 124 are used to further distinguish the forest types).

Table 2: Discrete classification coding

LC100 code	UN LCCS level	Land Cover Class	Definition according UN LCCS	Color code (RGB)
0	-	No PROBAV data available		51, 51, 51
-	A12A3A20B2	Forest	Lands dominated by woody plants with a percent cover >15% 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.	
111	A12A3A10B2D2E1	Closed forest, evergreen needle leaf	tree canopy >70%, almost all needle leaf trees remain green all year. Canopy is never without green foliage.	0, 130, 0
112	A12A3A10B2D1E1	Closed forest, evergreen, broad leaf	tree canopy >70%, almost all broadleaf trees remain green year round. Canopy is never without green foliage.	0, 153, 0
113	A12A3A10B2D2E2	Closed forest, deciduous needle leaf	tree canopy >70%, consists of seasonal needle leaf tree communities with an annual cycle of leaf-on and leaf-off periods	0, 179, 0
114	A12A3A10B2D1E2	Closed forest, deciduous broad leaf	tree canopy >70%, consists of seasonal broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.	0, 204, 0
121	A12A3A11B2D2E1	Open forest, evergreen needle leaf	top layer- trees 15-70% and second layer- mixed of shrubs and grassland, almost all needle leaf trees remain green all year. Canopy is never without green foliage.	112, 153, 0
122	A12A3A11B2D1E1	Open forest, evergreen broad leaf	top layer- trees 15-70% and second layer- mixed of shrubs and grassland, almost all broadleaf trees remain green year round. Canopy is never without green foliage.	131, 179, 0
123	A12A3A11B2D2E2	Open forest, deciduous needle leaf	top layer- trees 15-70% and second layer- mixed of shrubs and grassland, consists of seasonal needle leaf tree communities with an annual cycle of leaf-on and leaf-off periods	150, 204, 0
124	A12A3A11B2D1E2	Open forest, deciduous broad leaf	top layer- trees 15-70% and second layer- mixed of shrubs and grassland, consists of seasonal broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.	169, 230, 0
20	AA12A4A20B3(B9)	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.	255, 187, 34
30	A12A2(A6)A20B4	Herbaceous vegetation	Plants without persistent stem or shoots above ground and lacking definite firm structure. Tree and shrub cover is less than 10%.	255, 255, 76
40	A11A3	Cultivated and managed vegetation/agriculture (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.	240, 150, 255
50	B15A1	Urban / built up	Land covered by buildings and other man-made structures	255, 0, 0
60	B16A1(A2)	Bare / sparse vegetation	Lands with exposed soil, sand, or rocks and never has more than 10% vegetated cover	220, 220, 220

LC100 code	UN LCCS level	Land Cover Class	Definition according UN LCCS	Color code (RGB)
			during any time of the year	
70	B28A2(A3)	Snow and Ice	Lands under snow or ice cover throughout the year.	255, 255, 255
80	B28A1B1	Permanent water bodies	lakes, reservoirs, and rivers. Can be either fresh or salt-water bodies.	25, 25, 255
81	B28A1B2	Temporary water bodies		60, 160, 255
90	A24A2A20	Herbaceous wetland	Lands with a permanent mixture of water and herbaceous or woody vegetation. The vegetation can be present in either salt, brackish, or fresh water.	0, 150, 160
200	B28A1B1[1]	Open sea	Oceans, seas. Can be either fresh or salt-water bodies.	0, 0, 128
255	-	Not classified		0, 0, 0

The LCCS-PROB layer provides the probability of the discrete class and represents a number between 0 (uncertain) and 100 (certain) and is coded as byte. All values above 100 are not valid: all "Open sea" pixels are set to 200 and all missing values to 255.

The LCCS-QFLAG layer provides a binary mask: if a pixel was originated from 100m input data only (0=non_filled) or fused with 300m input data (1=filled). It is coded as byte. In general, since filling may introduce noise, the quality of the classification is less certain for pixels with LCCS_QFLAG=1. All values above 1 are not valid: all "Open sea" pixels are set to 200 and all missing values to 255.

3.3.1.2 Occurrence layer (OCCUR-WB)

The water bodies occurrence layer is a data layer which describes the presence of a water body during the reference year, and is coded as a byte. It represents a number between 0 (no detection) and 100 (permanent). All values above 100 are not valid: all "Open sea" pixels are set to 200 and all missing values to 255.

Note that the pixels with OCCUR-WB=100 correspond to pixels with LCCS=80.

3.3.1.3 Cover layers (COV)

The cover layers are data layers which describe the percent of a pixel which is covered by a class (Forest, herbaceous vegetation, Shrub, and bare soil), and are coded as a byte. As such it complements the LCCS layer through providing not only the default class, but also the percentage of pixel covered by these four classes.

The fraction cover layers provide a number ranging from 0 to 100 in steps of 1. The number represents the % of the 100m pixel filled with the cover. All values above 100 are not valid: all "Open sea" pixels are set to 200 and all missing values to 255.

[1] Note a distinction is made between Open sea (oceans) = 200 and other permanent water bodies = 80, despite they're mapped to the same UN LCCS layer legend.

Note that the pixels with COV-SHRUB=100 correspond to pixels with LCCS=20, the pixels with COV-GRASSLAND=100 correspond to pixels with LCCS=30 and the pixels with COV-BARE=100 correspond to pixels with LCCS=60.

3.3.1.4 Attributes

The netCDF files contain a number of metadata attributes according to the CF1.6 convention:

- on the file-level (Table 3);
- on the layer-level (Table 4);
- at the level of the standard dimension variables for latitude ('lat') and longitude ('lon'), holding one value per row or column respectively (Table 5);
- at the level of the grid mapping (spatial reference system) variable ('crs') (Table 6).

The GEOTIFF files provide the metadata attributes as key value pairs from the CF1.6 convention.

Table 3: Description of netCDF file attributes

Attribute name	Description	Data Type	Example(s)
Conventions	Version of the CF-Conventions used	String	CF-1.6
title	A description of the contents of the file	String	Dynamic Land Cover Map 100M: AFRI 2015-01-01T00:00:00Z
institution	The name of the institution that produced the product	String	VITO NV
source	The method of production of the original data	String	Derived from EO satellite imagery
history	A global attribute for an audit trail. One line, including date in ISO-8601 format, for each invocation of a program that has modified the dataset.	String	Processing line LC100: 2017-06-14
references	Published or web based references that describe the data or methods used to produce it.	String	http://land.copernicus.eu/global/products/lc
archive_facility	Specifies the name of the institution that archives the product	String	VITO
product_version	Version of the product (VM.m.r)	String	V1.0.1
time_coverage_start	Start date and time of the total coverage of the data for the product.	String	2014-10-01T00:00:00Z
time_coverage_end	End date and time of the total coverage of the data for the product.	String	2016-03-31T23:59:59Z
platform	Name(s) of the orbiting platform(s)	String	Proba-V
sensor	Name(s) of the sensor(s) used	String	VEGETATION
identifier	Unique identifier for the product	String	urn:cgls:global:lc100_v1_100m:LC100-LCCS_201501010000_AFRI_PROBAV_V1.0.1
parent_identifier	Identifier of the product collection (time series) for the product in Copernicus Global Land Service metadata catalogue.	String	urn:cgls:global:lc100_v1_100m
long_name	Extended product name	String	Land Cover
orbit_type	Orbit type of the orbiting platform(s)	String	LEO
processing_level	Product processing level	String	L4
processing_mode	Processing mode used when generating the product (Near-Real Time, Consolidated or Reprocessing)	String	Offline
copyright	Text to be used by users when referring to the data source of this product in publications (copyright notice)	String	Copernicus Service information 2017

Table 4: Description of netCDF layer attributes.

Attribute	Description	Data Type	Examples for LCCS layer
CLASS	Dataset type		
standard_name	A standardized name that references a description of a variable's content in CF-Convention's standard names table. Note that each standard_name has corresponding unit (from Unidata's udunits).	String	LCCS
long_name	A descriptive name that indicates a variable's content. This name is not standardized. Required when a standard name is not available.	String	Land Cover Classification
valid_range	Smallest and largest values for the variable. Missing data is to be represented by one or several values outside of this range.	Same as data variable	0, 200
_FillValue	Single value used to represent missing or undefined data and to pre-fill memory space in case a non-written part of data is read back. Value must be outside of valid_range.	Same as data variable	0
missing_value	Single value used to represent missing or undefined data, for applications following older versions of the standards. Value must be outside of valid_range.	Same as data variable	255
grid_mapping	Reference to the grid mapping variable	String	Crs
flag_values	Provides a list of the flag values. Used in conjunction with flag_meanings.	Same as data variable	111, 112, 113, 114, 121, 122, 123, 124, 20, 30, 40, 50, 60, 70, 80, 81, 90, 200
flag_meanings	Descriptive words or phrases for each flag value.	String (blank separated list)	The land cover classes as defined in Table 2

Table 5: Description of netCDF attributes for coordinate dimensions (latitudes and longitudes).

Attribute	Description	Data Type	Example
CLASS	Dataset type	String	DIMENSION_SCALE
DIMENSION_ LABELS	Label used in netCDF4 library	String	lon
NAME	Short name	String	lon
standard_name	A standardized name that references a description of a variable's content in CF-Convention's standard names table. Note that each standard_name has corresponding unit (from Unidata's udunits).	String	longitude
long_name	A descriptive name that indicates a variable's content. This name is not standardized. Required when a standard name is not available.	String	longitude
units	Units of a the variable's content, taken from Unidata's udunits library.	String	degrees_east
axis	Identifies latitude, longitude, vertical, or time axes.	String	X
_CoordinateAxis Type	Label used in GDAL library	String	Lon

Table 6: Description of netCDF attributes for the grid mapping variable.

Attribute	Description	Data Type	Example
GeoTransform	Six coefficients for the affine transformation from pixel/line space to coordinate space, as defined in GDAL's GeoTransform	String	-30.000496031746032, 0.000992063492063, 0.0, 45.000496031746032 0.0, -0.000992063492063
_CoordinateAxisTypes	Label used in GDAL library	String, blank separated	GeoX GeoY
_CoordinateTransform Type	Type of transformation	String	Projection
grid_mapping_name	Name used to identify the grid mapping	String	latitude_longitude
inverse_flattening	Used to specify the inverse flattening (1/f) of the ellipsoidal figure associated with the geodetic datum and used to approximate the shape of the Earth	Float	298.257223563
long_name	A descriptive name that indicates a variable's content.	String	coordinate reference system
longitude_of_prime_m eridian	Specifies the longitude, with respect to Greenwich, of the prime meridian associated with the geodetic datum	Float	0.0

semi_major_axis	Specifies the length, in metres, of the semi-major axis of the ellipsoidal figure associated with the geodetic datum and used to approximate the shape of the Earth	Float	6378137.0
spatial_ref	Spatial reference system in OGC's Well-Known Text (WKT) format	String	GEOGCS["WGS 84",DATUM["WGS_1984",… AUTHORITY["EPSG","4326"]]

3.3.2 Quicklook

The quicklook is a GEOTIFF file. The spatial resolution is sub-sampled, using nearest neighbour resampling, to 5% in both directions, hence a quicklook is 1/400th of the size of the original image.

The quicklook is coded from the LCCS layer and follows the same colour scheme as described in Table 2.

3.4 Product Characteristics

3.4.1 Projection and Grid Information

The product is displayed in a regular latitude/longitude grid (plate carrée) with the ellipsoid WGS 1984 (Terrestrial radius=6378 km). The resolution of the grid is 1/1008°.

The reference is the centre of the pixel. It means that the longitude of the upper left corner of the pixel is (pixel_longitude – angular_resolution/2.)

3.4.2 Spatial Information

The LC100 product is provided from longitude -30°E to +60°W and latitude +45°N to -35°S.

3.4.3 Temporal Information

The LC100 product is provided on yearly basis from 01 January to 31 December. However, it uses also the data 3 months prior and pastor the reference year. As such the temporal coverage provides a start date of 01 October Year-1 to 31 March Year+1.

3.5 Data Policies

All users of the Global Land service products benefit from the free and open access policy as defined in the European Union's Copernicus regulation (N° 377/2014 of 3 April 2014) and

Commission Delegated Regulation (N° 1159/2013), available on the Copernicus programme's web site, http://www.copernicus.eu/library/detail/248). Products from legacy R&D projects are also provided with free and open action.

This includes the following use, in so far that is lawful:

a) reproduction;
b) distribution;
c) communication to the public;
d) adaptation, modification and combination with other data and information;
e) any combination of points (a) to (d).

EU law allows for specific limitations of access and use in the rare cases of security concerns, protection of third party rights or risk of service disruption.

By using Service Information the user acknowledges that these conditions are applicable to him/her and that the user renounces to any claims for damages against the European Union and the providers of the said Data and Information. The scope of this waiver encompasses any dispute, including contracts and torts claims that might be filed in court, in arbitration or in any other form of dispute settlement.

Where the user communicates to the public on or distributes the original LC100 products, he/she is obliged to refer to the data source with (at least) the following statement (included as the copyright attribute of the <file format e.g. netCDF> file):

Copernicus Service information [Year]

With [Year]: year of publication

Where the user has adapted or modified the products, the statement should be:

Contains modified Copernicus Service information [Year]

For complete acknowledgement and credits, the following statement can be used:

"*The products were generated by the Global Land Service of Copernicus, the Earth Observation programme of the European Commission. The research leading to the current version of the product has received funding from various European Commission Research and Technical Development programs. The product is based on PROBA-V data provided by ESA and distributed by VITO.*

The user accepts to inform Copernicus about the outcome of the use of the above-mentioned products and to send a copy of any publications that use these products to the following address: helpdeskticket@vgt.vito.be

3.6 ACCESS AND CONTACTS

The LC100 products are available through the Copernicus Global Land Service website at the address: http://land.copernicus.eu/global/products/lc with contact information (helpdesk) on http://land.copernicus.eu/global/contactpage.

Accountable contact: European Commission Directorate – General Joint Research Center

Email address: copernicuslandproducts@jrc.ec.europa.eu

Scientific & Technical Contact e-mail address: helpdeskticket@vgt.vito.be.

4 VALIDATION RESULTS

CGLS LC100 discrete and proportion maps were assessed qualitatively and quantitatively.

For the qualitative evaluation, the CGLS LC100 discrete map was compared visually using Google Map and other global scale land cover products (e.g. LC-CCI 2015 and Globeland30 2010).

For the quantitative quality evaluation, the land cover products were assessed using an independent validation dataset. The detailed information on the sample selection and reference data collection can be found in the Service Validation Plan document [CGLOPS1_SVP]. A brief summary of the validation dataset is provided here.

The validation dataset contains 3617 sample sites across Africa after two rounds of reference data collection. The first round of collection was based on a stratification based on Köppen climate zones and human population density (Olofsson et al., 2012). The second round of the reference data collection focuses on the rare land cover types such as urban, waterbody and wetland. A requirement of at least 250 sample sites per land cover type were used and was collected using the LC100 discrete V1 map [CGLOPS1_VR_LC100_V1].

Sample unit areas match with a PROBA-V data pixel (100 x 100 m at the Equator). At each sample unit, 10x10 subpixels were created and reference information on the land cover was collected by experts that have knowledge on African regional landscapes and remote sensing products. Since, reference land cover was interpreted at each sub pixel, fractions of land cover types were calculated. Land cover type fractions were then translated to LC100 discrete map legend based on the legend descriptions. This was used to validate LC100 discrete map. Land cover type fraction information was directly used to assess other fraction layers.

Based on this validation dataset, the overall accuracy of the CGLS LC100 discrete map is 74.3+/-1.8%. Fraction cover layers have mean absolute error of 11, 8, 16, and 6% for trees, shrubs, herbaceous vegetation and bare ground, respectively.

More details on this can be found in the full validation report [CGLOPS1_VR_LC100_V1].

4.1 QUALITATIVE ASSESSMENT

A visual comparison shows that CGLS LC100 discrete map is quite good compared to other available global land cover maps such as LC-CCI 2015 and Globeland30 2010. Due to its similarity in the legend, the general pattern of the land cover types in the CGLS-LC100 discrete map is similar to that of the Globeland30 2010 map. The LC-CCI 2015 is different because it characterizes more land cover types.

The separation of natural vegetation is very good with the CGLS LC100 map. For example, central part of Madagascar was mapped as 'mosaics of shrubs and grassland' in the LC-CCI 2015 while it is mostly herbaceous vegetation as shown in the CGLS LC100 and Globeland30 2010 (Figure 8).

Similarly, in Botswana, some shrub regions were mapped in the LC-CCI 2015 with boundaries that do not correspond to natural landscapes while it does in the CGLS LC100. A close zoom in on the Google map confirms the omission of shrubs in the Globeland30 2010 map (Figure 8).

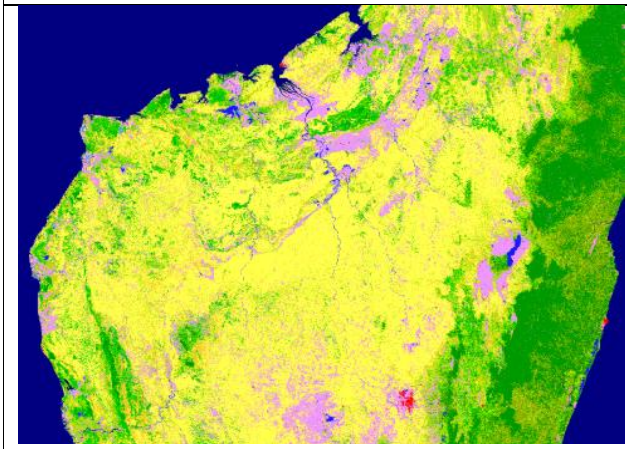
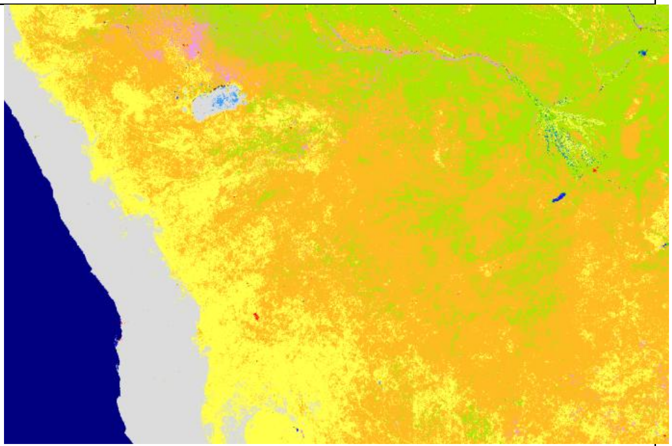
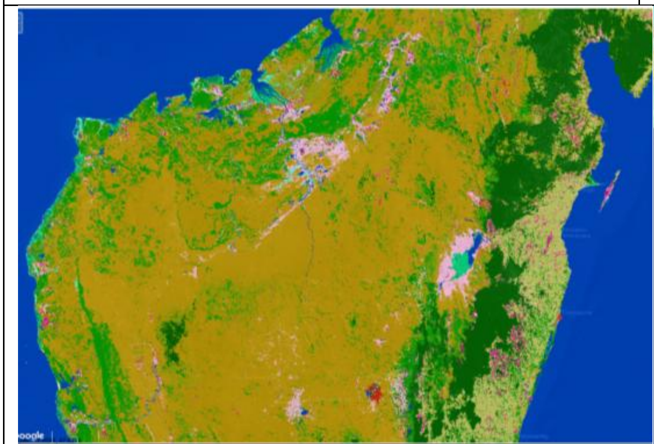
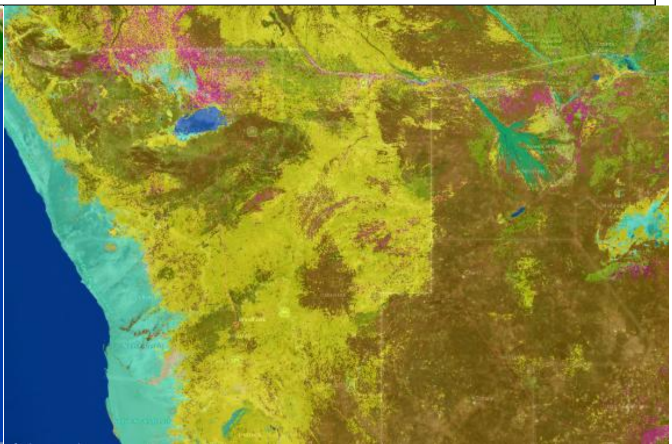
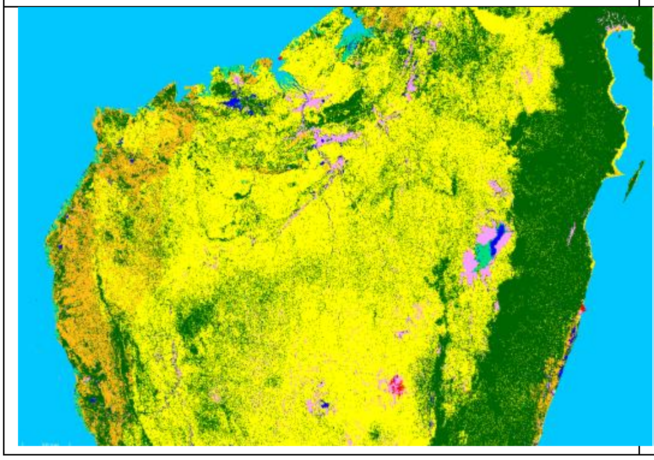
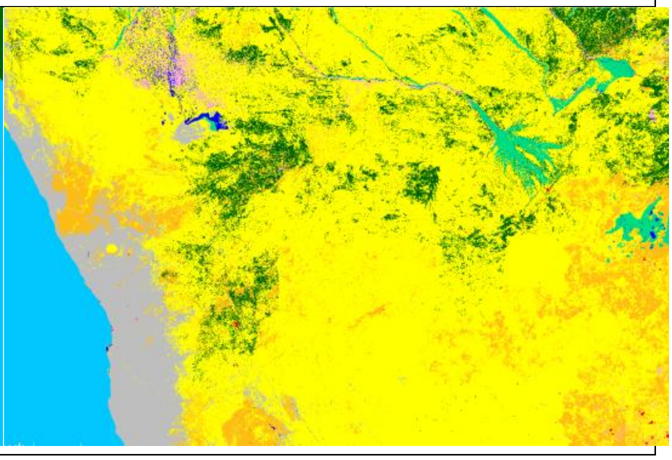
CGLS LC100 Madagascar	CGLS LC100 Botswana
	
LC-CCI 2015 Madagascar	LC-CCI 2015 Botswana
	
Globeland30 2010 Madagascar	Globeland30 2010 Botswana
	

Figure 8: Comparison of three maps in Madagascar and Botswana (legends are shown in Annex 2: Legends of the other global land cover maps)

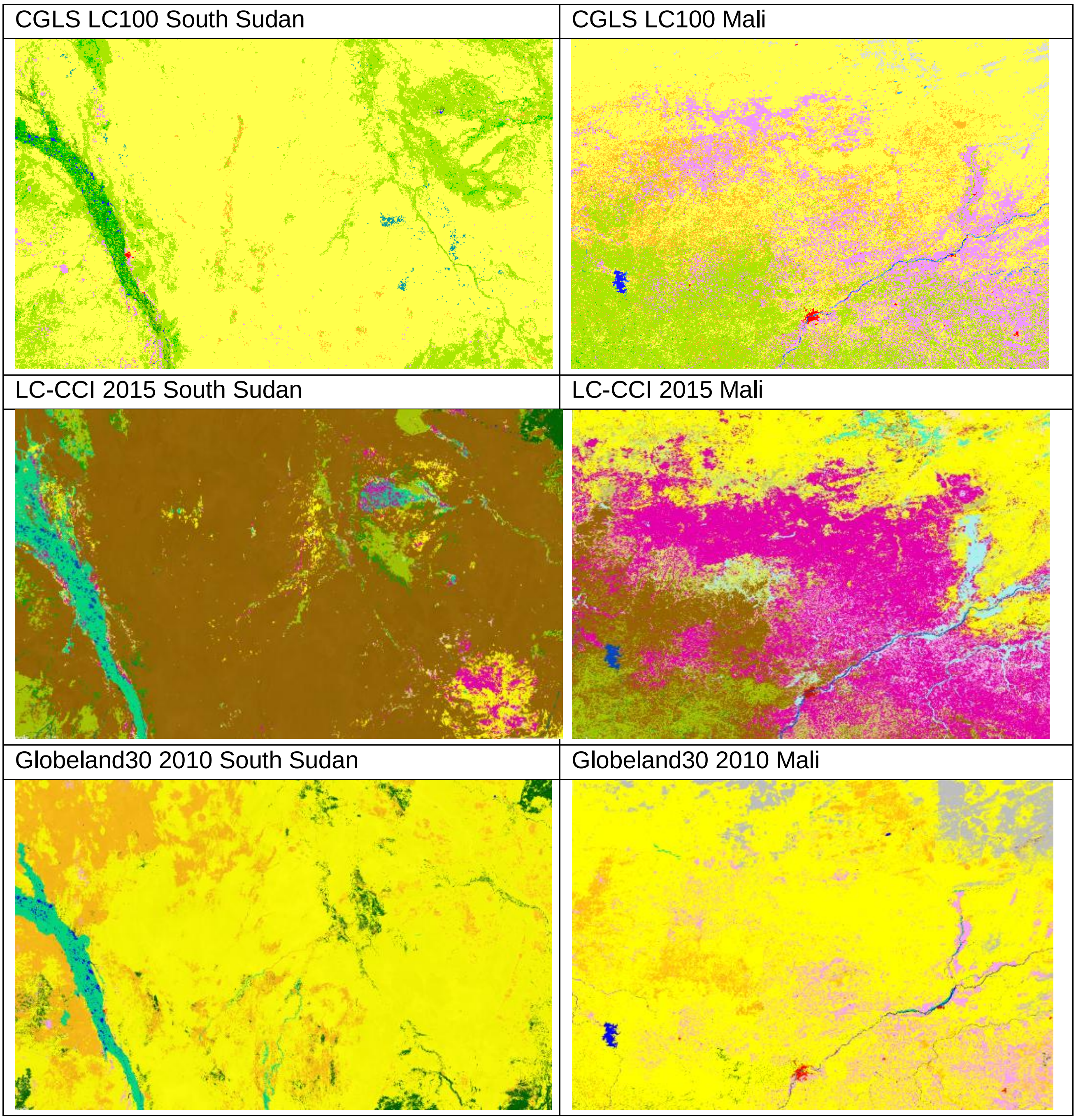


Figure 9: Comparison of three maps in South Sudan and Mali (legends are shown in Annex 2)

Figure 9 shows the comparisons in South Sudan and Mali. In South Sudan, most areas of grassland were captured correctly in the CGLS LC100 map and the Globeland30 2010 map, while this area is mapped mostly as shrubs in the LC-CCI 2015. However, the CGLS-LC100 maps much less areas as wetland vegetation as compared with the LC-CCI 2015 and Globeland30 2010 maps. In Mali, the Globeland30 2010 map mapped very small area as crops, while the LC-CCI 2015 maps considerably large areas as crops. The CGLS-LC100 map delineates larger areas as cropland as compared with the Globeland30 map which maps most of these regions as grassland.

At smaller scale, the three land cover maps are visually compared with high resolution Google Map images. Based on a visual interpretation of high resolution Google Map images and the three land cover maps, the discrete CGLS-LC100 map shows improvements in the land cover identification as compared with the LC-CCI 2015 and the Globeland30 maps. A few examples are shown in Figure 10 - Figure 12.

As shown in Figure 10, the CGLS LC100 map delineated the irrigated cropland area well, while LC-CCI 2015 and Globeland30 2010 maps missed out some crop areas.

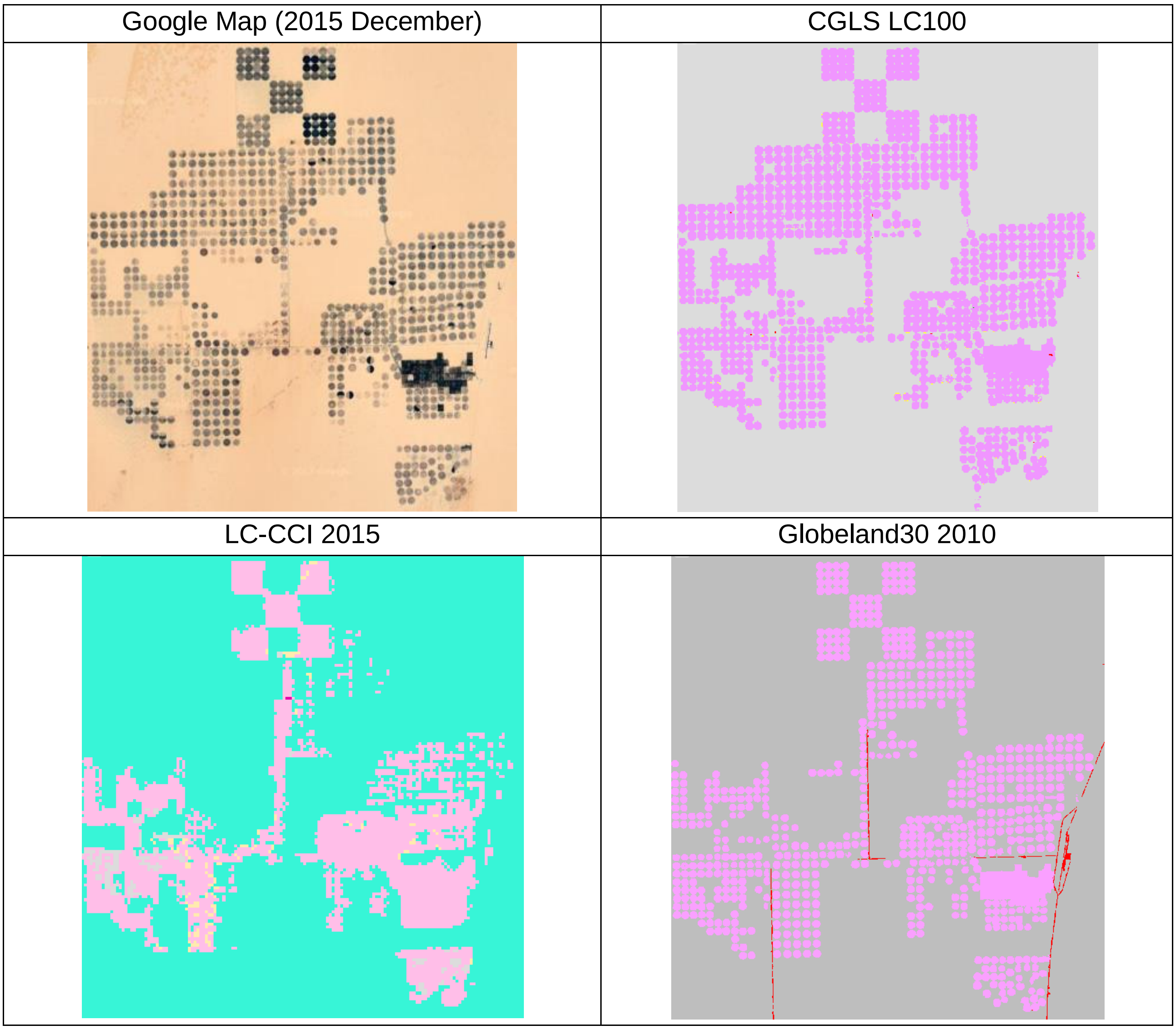


Figure 10: Visual comparison of the maps in Egypt (Lat: 22.65, Lon: 28.51) (legends are shown in Annex 2: Legends of the other global land cover maps)

Open forest areas in Southern part of Somalia were mapped as cropland in the LC-CCI 2015 and shrubs in the Globeland30 2010 maps (Figure 11). On the other hand, visual comparison of the Google Map confirms that there are more open forests as it is mapped in the CGLS LC100 map.

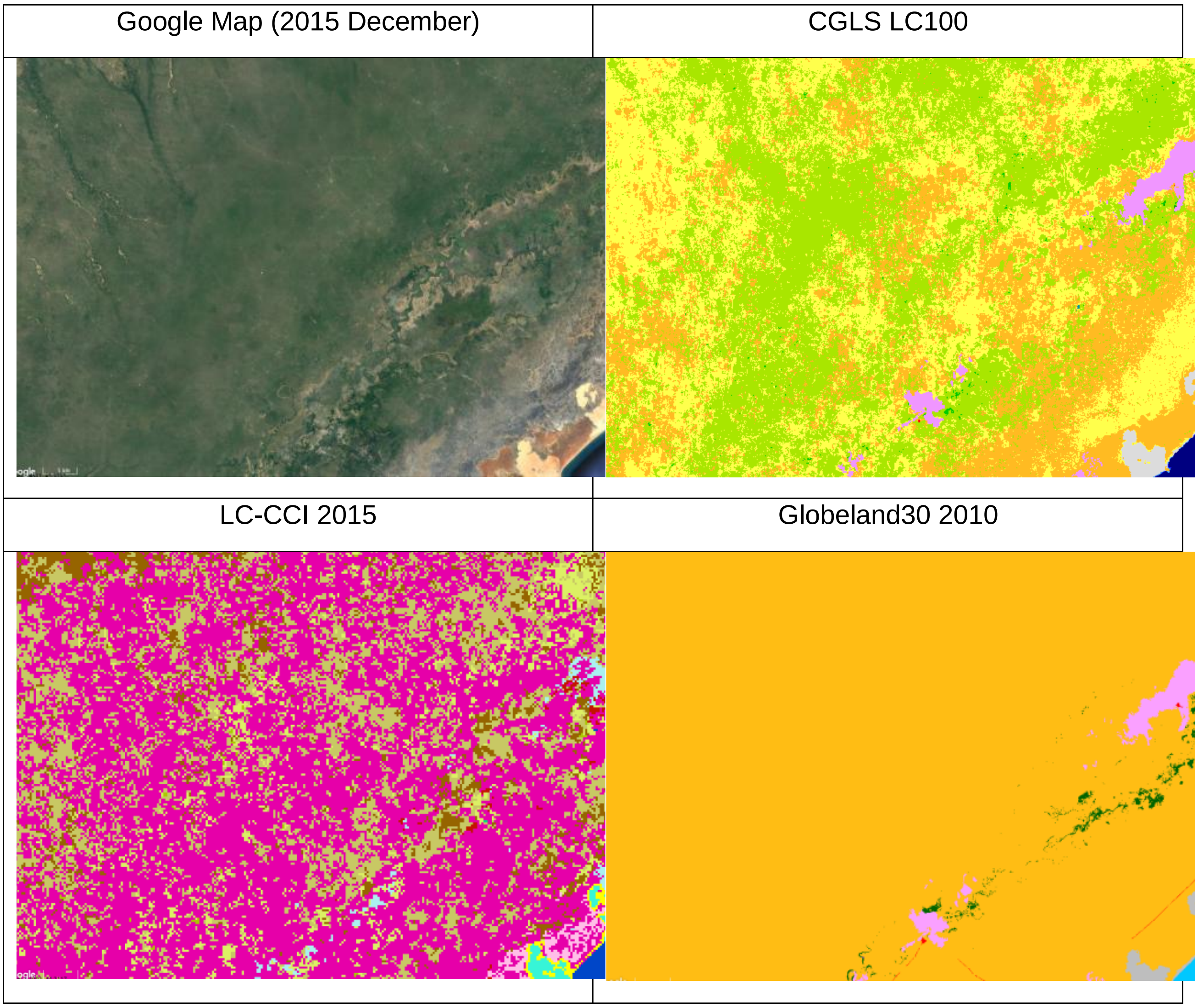


Figure 11: Visual comparison of the maps in Somalia (Lat: 1.1679, Lon: 43.4928) (legends are shown in Annex 2: Legends of the other global land cover maps)

In Ivory Coast, east of San-Pedró city, the LC-CCI 2015 map overestimated considerably cropland areas (Figure 12). This area was mapped mostly as forest in the Globeland30 2010 map. However, the Globeland30 2010 map introduced artificial boundaries transitioning from forest to grassland. This map also tends to miss some crop areas. The CGLS-LC100 also delineated some cropland areas, which appear to be overestimated a bit when compared with visual interpretation the Google map images.

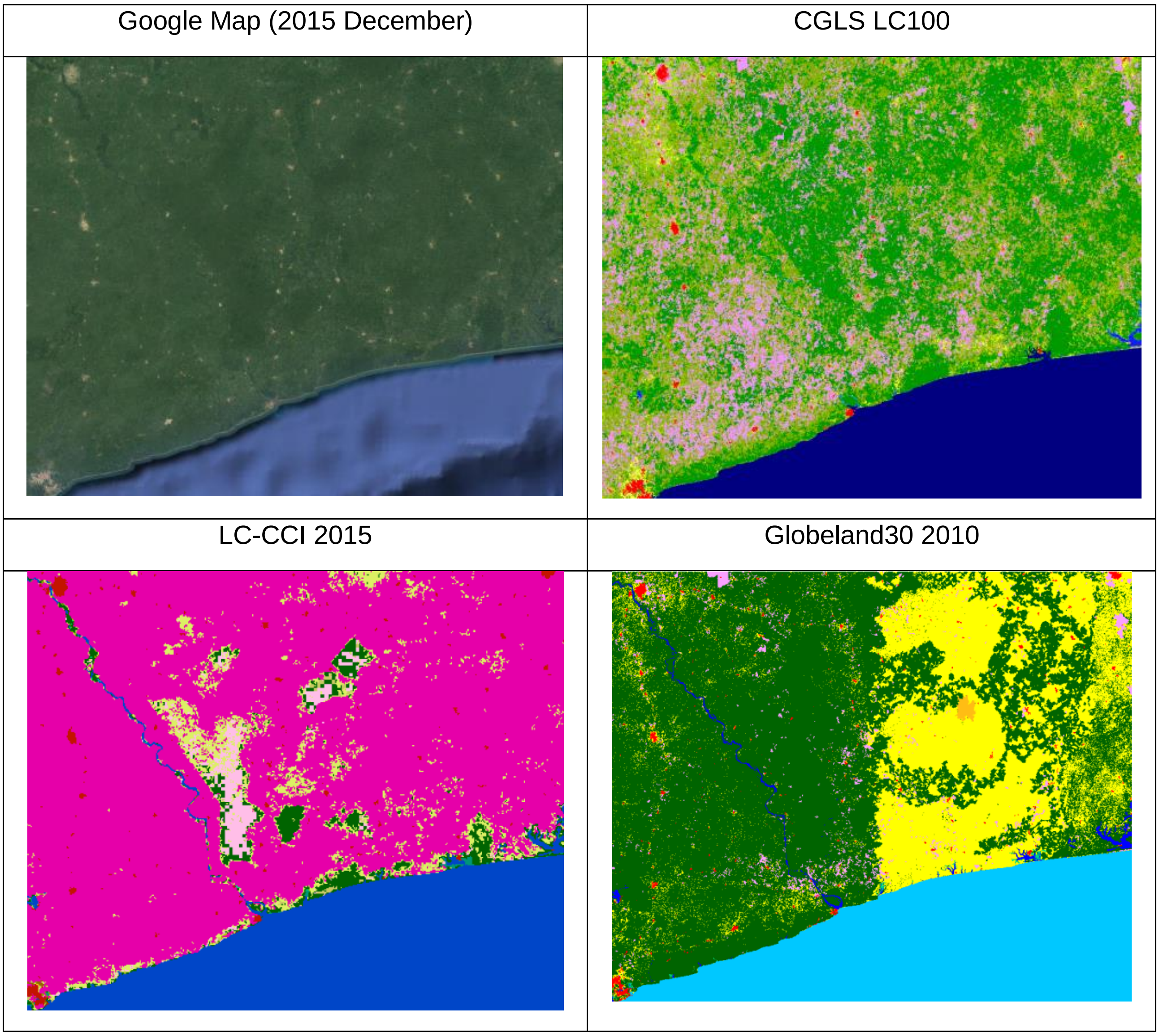


Figure 12: Visual comparison of the maps in Ivory Coast (Lat: 5.43, Lon: -6.24) (legends are shown in Annex 2: Legends of the other global land cover maps)

4.2 Accuracy Assessment (Quantitative Assessment)

The validation of CGLS LC100 products were based on the Service Validation Plan of the Dynamic Land Cover product of the Copernicus Global Land Service [CGLOPS1_SVP].

4.2.1 Accuracy assessment of CGLS LC100 discrete map

At each validation sample location, mapped land cover types were extracted. Based on the mapped and reference land cover types, a confusion or an error matrix was calculated. This error matrix was corrected by proportions of the mapped land cover types in African continent. This aims to reduce area bias due to different sampling densities for the land cover types.

Area weighted confusion matrix for 9 general land cover types is shown in Table 7. More detailed explanation on the accuracy assessment can be found in the full validation report [CGLOPS1_VR_LC100_V1].

Table 7: A confusion matrix for discrete CGLS LC100 map as proportions of African continent area

		Reference class												
		Closed forest	Open forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/Sparse veg.	Water	Wetland	**Correct proportion**	Total proportion	**User's accuracy**	Confidence interval +/-
Mapped class	Closed forest	**12.38**	1.43	0.22	0.14	0.20	0	0	0.04	0.1	**12.38**	14.50	**85.3**	2.6
	Open forest	1.57	**10.96**	1.60	1.51	1.13	0	0.03	0.03	0.47	**10.96**	17.30	**63.4**	3.9
	Shrubs	0.08	1.61	**4.04**	0.75	0.39	0.03	0.08	0	0.06	**4.04**	7.04	**57.3**	6.1
	Herbaceous veg.	0.27	1.83	1.49	**10.42**	0.88	0.03	0.43	0.09	0.30	**10.42**	15.75	**66.2**	4.1
	Croplands	0.17	0.92	0.46	1.43	**6.65**	0.02	0.10	0.15	0.1	**6.65**	10.00	**66.5**	4.6
	Urban	0	0.03	0.002	0.03	0.005	**0.17**	0.001	0.001	0	**0.17**	0.25	**70.4**	5.7
	Bare/Sparse veg.	0	0.11	0.88	3.31	0.66	0	**28.72**	0.44	0	**28.72**	34.14	**84.1**	4.1
	Water	0	0.01	0.01	0.01	0.01	0	0.003	**0.87**	0.03	**0.87**	1	**93.3**	2.8
	Wetland	0	0.003	0	0.00	0.002	0.0003	0	0.01	**0.07**	**0.07**	0.09	**78.0**	5.1
Correct proportion		**12.38**	**10.96**	**4.04**	**10.42**	**6.65**	**0.17**	**28.72**	**0.87**	**0.07**	**74.3**			
Total proportion		14.47	16.91	8.70	17.61	9.92	0.26	29.36	1.63	1.12		**100**		
Producer's accuracy		**85.5**	**64.8**	**46.4**	**59.2**	**67.0**	**67.6**	**97.8**	**53.4**	**6.0**			**74.3**	**1.8**
Confidence interval +/-		2.8	3.2	5.3	4.5	5.2	24.8	0.9	15.2	1.8				

Overall map accuracy before area bias correction was 73.9%. With area bias correction, overall accuracy of the CGLS LC100 discrete map increased slightly to 74.3% +/-1.8% (confidence intervals at 95% confidence level). The increase is expected mainly due to large proportion of bare/sparse vegetation areas in Africa that have higher class specific accuracies. However, only a slight increase in the overall accuracy resulted because significant number of sample sites for urban and water areas (240 and 312 sample sites) influenced positively the overall accuracy before area bias correction, but their influence was low when corrected by area bias due to their small margins in area in African land.

In terms of class specific accuracies, closed forest and bare/sparse vegetation mapped with higher accuracy while accuracies for open forest, herbaceous vegetation and cropland classes were relatively lower. Among the natural vegetation classes, shrubs had the lowest accuracies. The producer's accuracy of the wetland class is particularly very low suggesting high omission error. This shows significant wetland areas are omitted in the CGLS-LC100m-V1 with high confusion with open forests and herbaceous vegetation classes.

Before area bias correction, the producer's accuracies of urban and water and wetland classes were 98%, 91% and 81%, respectively. After area bias correction, the producer's accuracy of the urban and water classes reduced due to confusions of a few sample sites in large-area classes namely herbaceous vegetation and croplands. This also influenced the higher confidence intervals for these classes.

A brief comparison of the CGLS LC100 discrete map with other available maps were made based on available reference datasets in Africa compiled by Tsendbazar et al. (2015). The authors of this research combined six available reference datasets in Africa and checked their correspondence with Globcover 2009, LC-CCI 2010, MODIS2010 and Globeland30 maps. They compared 3887 reference points with the four global land cover maps and reported the overall agreements of the maps in Africa as shown in Table 8.

Table 8: Agreement of combined reference points with four maps in Africa

	Overall agreement %
Globcover 2009	50.8
LC-CCI 2010	55.5
MODIS 2010	63
Globeland 2010	57.2

Although there are quite some differences with respect to the reference years and legend harmonization, and differences in the reference dataset (Tsendbazar et al., 2015), we used these reference data to check the agreement of the CGLS LC100 discrete product. The overall agreement of the CGLS LC100 product with the combined reference data was 66% which is higher than any of the maps compared in Tsendbazar et al., (2015).

Comparison of the CGLS LC100 with other global land cover map, namely the Globeland30-2010using our validation dataset is be provided in the full validation report [CGLOPS1_VR_LC100_V1].

4.2.2 Accuracy assessment of CGLS LC100 cover fraction maps

The CGLS LC100 cover fraction maps namely tree cover, shrub cover, herbaceous vegetation cover and bare/sparse vegetation areas were assessed using the cover fraction information in the validation dataset. Table 9 lists the mean absolute error and root mean square error for the fraction cover maps.

Table 9: Accuracy of the cover fraction products.

	Tree fraction	Shrub fraction	Herbaceous vegetation fraction	Bare fraction
Mean absolute error (MAE)	11.1 %	8.3 %	16.3 %	6.03 %
Root mean square error (RMSE)	18.3 %	13.8 %	24.5 %	14.1 %

Among the cover fraction products, bare area fraction map showed the lowest error with an average deviation from the reference data of 6% and RMSE of 14%. Herbaceous vegetation fraction product had the highest error with MAE of 16.3% and RMSE of 24.5%. This can be due to difficulty in separating herbaceous vegetation from other land cover types. This is also observed in Table 7 where herbaceous vegetation had higher confusions with other classes. In general, tree cover fraction product tends to underestimate the areas with higher tree cover densities (higher range of tree cover fractions). This is also observed over fraction layer for natural vegetation.

Figure 13 shows the distribution of reference points along with the differences in the reference and mapped land cover fractions. The deviation from the reference data tended to be higher on bordering regions of main land cover types.

Tree fraction product tends to underestimate more near the Gulf of Guinea coastal regions. Shrub fraction product tends to mostly underestimate in Ethiopia, Somalia and Namibia. Accordingly, at the cost for shrub proportions, herbaceous vegetation fractions are overestimated in those areas. While in the Sahel transition, regions near the rain forest areas and in Madagascar, herbaceous fraction tend to be underestimated. Bare fraction product tended to overestimate bare fractions in the Sahel transition areas, while it tends to underestimate bare fractions in the Horn of Africa.

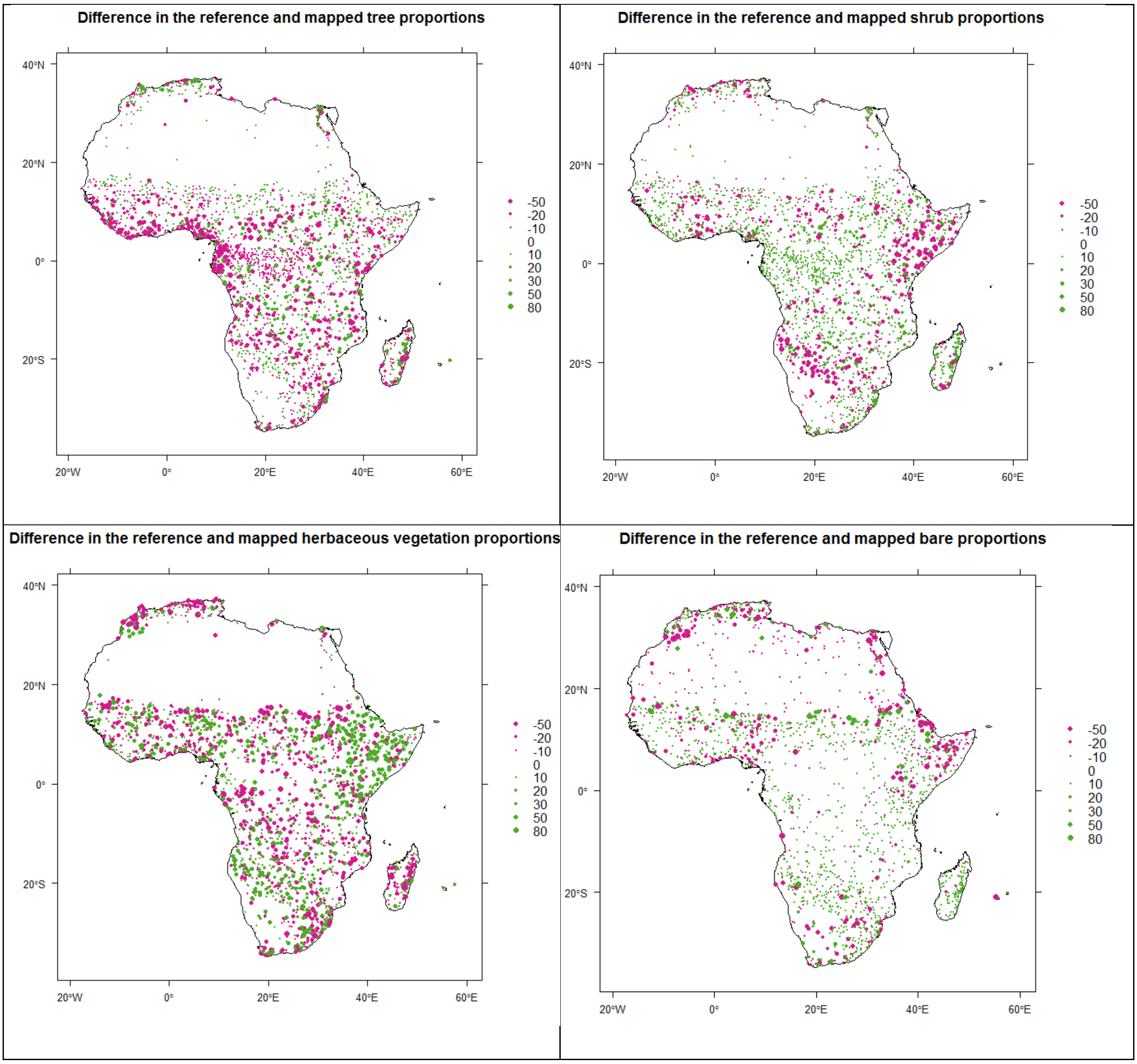


Figure 13: Differences in the validation data and LC100 fraction layers in percentage (right to left: trees, shrub, herbaceous vegetation and bare fractions; pink colour indicates under-estimation, green indicates over-estimation)

4.3 Conclusion

Our assessments showed that the CGLS-LC100m discrete map is mapped with 74.3+/-1.8% accuracy. Closed forest and bare/sparse vegetation classes are mapped with higher class specific accuracies while shrubs and wetland classes are mapped with lowest accuracies. Among the four cover fraction layers, bare fraction is mapped with lowest errors, MAE and RMSE (6% and 14% respectively). On the other hand, herbaceous vegetation cover has the highest errors.

Although the overall accuracy of the CGLS-LC100m discrete map does not meet the requirement of 80% by the main users, our comparison shows that the CGLS-LC100 discrete map has higher accuracy than other recent global land cover map in Africa. Our visual comparison of the CGLS-LC100m, the LC-CCI 2015 and the Globeland30 2010 maps shows that the CGLS-LC100m characterizes natural vegetation classes better. Quantitative comparison using a combined reference data shows the CGLS-LC100m map has 3% higher agreement than the other global land cover maps in Africa. These comparison results show significant improvement in characterizing land cover in Africa as compared with other global land cover maps.

More detailed assessments on spatial map uncertainty and map accuracies from different user perspectives are provided in [CGLOPS1_VR_LC100_V1].

5 REFERENCES

Badgley, G., Field, C. B., & Berry, J. A. (2017). Canopy near-infrared reflectance and terrestrial photosynthesis. Science Advances, 3(3), e1602244.

Bertels, L., Smets, B., Wolfs, D. (2016). Dynamic Water Surface Detection Algorithm Applied on PROBA-V Multispectral Data. Remote Sens., 8, 1010.

Blackburn, G.A. (1998). Spectral indices for estimating photosynthetic pigment concentrations: a test using senescent tree leaves. International Journal of Remote Sensing, 19 (4), 657-675.

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

Copernicus Global Land Operations Lot 1 (C-GLOPS1) consortium (2017). Input to service validation plan – dynamic land cover – Issue I1.20. Project report, JRC, Italy.

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, 35:7, 2589-2614, DOI: 10.1080/01431161.2014.883097

Eberenz, J., Verbesselt, J., Herold, M., Tsendbazar, N.-E., Sabatino, G., Rivolta, G. (2016). Evaluating the Potential of PROBA-V Satellite Image Time Series for Improving LC Classification in Semi-Arid African Landscapes. Remote Sens., 8, 987.

Eerens, H., Haesen, D., Rembold, F., Urbano, F., Tote, C., and Bydekerke, L. (2014). Image time series processing for agriculture monitoring. Environ. Model. Soft. 53, 154–162. doi: 10.1016/j.envsoft.2013.10.021

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO) (2012). Global ecological zones for FAO forest reporting: 2010 Update. Forest Resources Assessment Working Paper 179, FAO, Rome, Italy.

Francois, M., Santandrea, S., Mellab, K., Vrancken, D., Versluys, J. (2014). The PROBA-V mission: the space segment, International Journal of Remote Sensing, 35:7, 2548-2564, DOI: 10.1080/01431161.2014.883098

Huete, A., et al. (2002). Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. Remote Sensing of Environment, 83,195–213.

Hughes, G. F. (1968). On the Mean Accuracy of Statistical Pattern Recognizers. IEEE Transactions on Information Theory, IT-14:55-63.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. Journal of Basic Engineering. 82: 35. doi:10.1115/1.3662552.

Kempeneers, P., Sedano, F., Piccard, I., Eerens, H. (2016). Data Assimilation of PROBA-V 100 and 300 m. In IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 7, pp. 3314-3325. DOI: 10.1109/JSTARS.2016.2527922

Key, C., Benson, N. (2005). Landscape Assessment: Remote Sensing of Severity, the Normalized Burn Ratio; and Ground Measure of Severity, the Composite Burn Index. In FIREMON: Fire Effects Monitoring and Inventory System, RMRS-GTR, Ogden, UT: USDA Forest Service, Rocky Mountain Research Station.

Kursa, M.B. (2017): Boruta. Webpage: https://m2.icm.edu.pl/boruta/. Retrieved: July 17, 2017.

Kursa, M.B., Rudnicki, W.R., (2010). Feature Selection with the Boruta Package. Journal of Statistical Software, [S.l.], v. 36, Issue 11, p. 1 - 13.

Marconcini, M., Uereyen, S., Esch, Y., Metz, A., Zeidler, J. (2017a). Mapping urban areas globally by jointly exploiting optical and radar imagery – the GUF+2015 layer. Presentation at the Worldcover 2017 Conference, ESA-ESRIN, Frascati, Italy, 14-16 March 2017.

Marconcini, M., Üreyen, S., Esch, T., Metz, A., Zeidler, J. and Palacios-Lopez, D. (2017b). “Outlining the urban side of the Earth – the GUF+2015”, Scientific Data (in preparation).

NASA JPL. (2013). NASA Shuttle Radar Topography Mission Global 3 arc second [Data set]. NASA LP DAAC. https://doi.org/10.5067/measures/srtm/srtmgl3.003

Olofsson, P., Stehman, S. V., Woodcock, C. E., Sulla-Menashe, D., Sibley, A. M., Newell, J. D., Herold, M. (2012). A global land-cover validation data set, part I: fundamental design principles. International Journal of Remote Sensing, 33(18), 5768-5788. doi: 10.1080/01431161.2012.674230

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, p. 2825-2830.

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. (doi:10.1038/nature20584)

Pekel, J.-F., Vancutsem, C., Bastin, L., Clerici, M., Vanbogaert, E., Bartholomé, E., & Defourny, P. (2014). A near real-time water surface detection method based on HSV transformation of MODIS multi-spectral time series data. Remote Sensing of Environment, 140, 704–716. doi:10.1016/j.rse.2013.10.008

Pesaresi, M., Ehrilch, D., Florczyk, A. J., Freire, S., Julea, A., Kemper, T.P., Syrris, V. (2015). GHS built-up grid, derived from Landsat, multitemporal (1975, 1990, 2000, 2014). European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-ghsl-ghs_built_ldsmt_globe_r2015b

Roerink, G. J.; Menenti, M.; Verhoef, W. (2000). Reconstructing cloudfree NDVI composites using Fourier analysis of time series. International Journal of Remote Sensing, 21 (9), pp. 1911-1917. DOI: 10.1080/014311600209814

Sayre, R., Comer,P., Hak, J., Josse, C., Bow, J., Warner, H., Larwanou, M., Kelbessa, E., Bekele, T., Kehl, H., Amena, R., Andriamasimanana, R., Ba, T., Benson, L., Boucher, T., Brown, M., Cress, J., Dassering, O., Friesen, B., Gachathi, F., Houcine, S., Keita, M., Khamala, E., Marangu, D., Mokua, F., Morou, B., Mucina, L., Mugisha, S., Mwavu, E., Rutherford, M., Sanou, P., Syampungani, S., Tomor, B., Vall, A., Vande Weghe, J., Wangui, E, Waruingi, L. (2013). A New Map of Standardized Terrestrial Ecosystems of Africa. Washington, DC: Association of American Geographers, 24 pages.

Scikit-learn (sklearn) (2017). User guide for ensemble module: ensemble methods. Webpage: http://scikit-learn.org/stable/modules/ensemble.html#forest. Retrieved: July 18, 2017.

Sedano, F., Kempeneers, P., Hurtt, G. (2014). A Kalman Filter-Based Method to Generate Continuous Time Series of Medium-Resolution NDVI Images. *Remote Sens.*, *6*, 12381-12408.

Sterckx, S., Benhadj, I., Duhoux, G., Livens, S., Dierckx, W., Goor, E., Adriaensen, S., Heyns, W., Van Hoof, K., Strackx, G., Nackaerts, K., Reusen, I., Van Achteren, T., Dries, J., Van Roey, T., Mellab, K., Duca, R., Zender, J. (2014). The PROBA-V mission: image processing and calibration. International Journal of Remote Sensing, 35:7, 2565-2588. DOI: 10.1080/01431161.2014.883094

Tsendbazar, N. E., de Bruin, S., Fritz, S., & Herold, M. (2015). Spatial Accuracy Assessment and Integration of Global Land Cover Datasets. Remote Sensing, 7(12), 15804-15821.

Tucker, C.J. (1979). 'Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. Remote Sensing of Environment, 8(2),127-150.

Verhoef, W. (1996). Application of Harmonic Analysis of NDVI Time Series (HANTS). In Fourier Analysis of T emporal NDVI in the Southern African and American Continents, edited by S. Azzali and M. Menenti, DLO Winand Staring Centre, Wageningen, The Netherlands, Report 108, pp. 19–24.

Wagner, J. E., Stehman, S. V. (2015). Optimizing sample size allocation to strata for estimating area and map accuracy. Remote Sensing of Environment 168 (2015) 126-133, doi:10.1016/j.rse.2015.06.027

Walker, H. (1931). Studies in the History of the Statistical Method. Baltimore, MD: Williams & Wilkins Co. pp. 24–25.

Welch, G., Bishop, G. (2006). An Introduction to the Kalman Filter. Report, University of North Carolina at Chapel Hill, Department of Computer Science,TR 95-041.

Wolters, E., Dierckx, W., Iordache, M.-D., Swinnen, E. (2017). PROBA-V Products User Manual v2.1. VITO N.V., authorized by ESA, 103 pp, available online at: http://proba-v.vgt.vito.be/sites/proba-v.vgt.vito.be/files/probav-products_user_manual_v2.11.pdf.

ANNEX 1: REVIEW OF USERS REQUIREMENTS

According to the applicable document [AD2] and [AD3], the user's requirements relevant for Dynamic Moderate Land Cover are:

- **Definition**: Dynamic global land cover products at 300m and/or 100m resolution using UN Land Cover Classification System (LCCS)
- **Geometric properties**:
  - Pixel size of output data shall be defined on a per-product basis so as to facilitate the multi-parameter analysis and exploitation.
  - The baseline datasets pixel size shall be provided, depending on the final product, at resolutions of 100m and/or 300m and/or 1km.
  - The target baseline location accuracy shall be 1/3 of the at-nadir instantaneous field of view.
  - Pixel co-ordinates shall be given for centre of pixel.
- **Geographical coverage**:
  - geographic projection: lat long
  - geodetical datum: WGS84
  - pixel size: 1/112° - accuracy: min 10 digits
  - coordinate position: pixel centre
  - global window coordinates:
    - Upper Left: 180°W-75°N
    - Bottom Right: 180°E, 56°S
- **Accuracy requirements**: Overall thematic accuracy of dynamic land cover mapping products shall be >80%. The overall accuracy assessment (including confidence limits) will be based on a stratified random sampling design and the minimum number of sampling points per land cover class relevant to the product shall be calculated as described in Wagner and Stehman, 2015.

Few workshops were held in 2016 to consult different stakeholders to understand users' needs for global land cover maps. A feasibility study was performed to define the guidelines to create the first LC100 map. More details can be found in [CGLOPS1_URD_LC100].

Table 10 provides a summary of the major requirements from the stakeholders, while Table 11 shows an overview of the requested classes to be covered by the mapping.

Table 10: Summary of stakeholder requirements

Land cover change information	
Forest modelling/REDD+	Forest change information is needed for identifying areas of tree loss and gain.
Crop monitoring	Static land cover maps of a high accuracy are of high priority
Biodiversity and conservation	Reliable information on the extent, location and change of habitats is needed for integration in a change alert system.
Monitoring Environment and Security in Africa	Depending on application, both types of maps are needed: change maps and static land cover map.
Climate modelling	Priority is given to stable land cover maps. Change maps are desirable as well, accompanied with a measure of reliability quantifying their statistical accuracy.
Resolution	
Forest modelling/REDD+	1-20 m – higher is better
Crop monitoring	100 m resolution is satisfactory for cropland mask
Biodiversity and conservation	1-20 m – higher is better
Monitoring Environment and Security in Africa	100 m is acceptable
Climate modelling	100 m resolution is very good to produce better PFT fraction estimations at coarser scales
Accuracy/error information	
All users	Overall thematic accuracy > 80% and should be based on stratified random sampling design, with a number of sample points per land cover class calculated (Wagner et al, 2015)
	Accuracy estimates should be not only overall, but also class specific.
	Accuracy has to be calculated at different geographical levels, e.g. regional, national, continental, global
	Minimum error has to be less than 15% or 20% at class level and at regional or national level (large country).
	Qualifying the error in a spatial manner is important, e.g. using covariance matrices, (Tsendbazar et al 2015).
Thematic requirements	
Forest modelling/REDD+	Mapping human impact on forest: primary and secondary forests, intactness, core/edge, managed/unmanaged, as well as forest

	parameters such as tree height and carbon stock/biomass, NPP, etc.
Crop monitoring	More classes on managed land/cultivated areas: irrigation, big/small farming, permanent crops, fallow, grassland (artificial, natural), some plantations
Biodiversity and conservation	Savannah, wooded shrubs, wetlands, natural vs man-made; Abandoned land; Infrastructure such as mines, roads, built infrastructure, including settlements, roads, electric lighting, canals and water control structures.
Monitoring Environment and Security in Africa	Forestry, Inland Waters, Pastoral Resources, Land Cover Change Assessment (including urbanization), Land Degradation, Natural Habitat Conservation Assessment, Monitoring and Assessment of Environmental Impacts of Mineral Resources Exploitation
Climate modelling from vegetation	Classes related to PFTs: trees vs shrubs vs grasses, C3 crops vsC4 crops vs irrigated crops; leaf types; managed vs natural classes, change vs phenology, etc.
All users	More land cover classes of Level 2. More details in a section below. UN LCCS should be used by default.
Projection	
All users	Commonly used projection (e.g. WGS 1984, EPSG: 4326), eventually easy to convert.
Access	
All users	Easy and open access, options for countries with slow connections, options to choose between global and regional products
Other requirements	
All users	Yearly updates and consistency among consecutive products.
	Continuity on nomenclature of the land cover products. Reprocess operations should be performed whenever the nomenclature evolves.
	A clear distinction should be made between “date of issue” and the “data used” (period).

Table 11: List of land cover classes requested by users

Code Level 1	Code Level 2	UN LCCS level	Land cover class	Forest modelling/REDD+	Crop monitoring	Biodiversity	Monitoring Environment and Security in Africa	Climate modelling
10		A12A3A20B2	Forest/tree cover	X		X	X	X
	11	A12A3A20B2D2 E1	Evergreen Needleleaf forest	X			X	X
	12	A12A3A20B2D1 E1	Evergreen Broadleaf forest	X			X	X
	13	A12A3A20B2D2 E2	Deciduous Needleleaf forest	X			X	X
	14	A12A3A20B2D1 E2	Deciduous Broadleaf forest	X			X	X
	15	A12A3A20B2D1 D2	Mixed forest	X		X		
	16	A12A3A10B2X XXX (assuming that an intact forest is a very dense forest)	Intact forest	X		X		X
	17	-	Secondary forest	X		X		X
	18	A11A1	Managed forest	X		X		X
		A11A1	Plantation forest/tree crops	X	X	X		X
		A11A1	Oil palm plantation	X	X			
		-	Forest logging	X	X	X		
		A12A3	Dominant tree species, e.g. spruce, pine, birch	X		X		
		A11A1(A2/A3)	Shifting cultivation system	X	X			X
20		AA12A4A20B3(B9)	Shrub			X	X	X
	21	A12A4A20B(B9)XXE1	Evergreen shrubs			X		
	22	A12A4A20B3(B	Deciduous shrubs			X		

Code Level 1	Code Level 2	UN LCCS level	Land cover class	Forest modelling/REDD+	Crop monitoring	Biodiversity	Monitoring Environment and Security in Africa	Climate modelling
		9)XXE2						
30		A12A2(A6)A20 B4	Herbaceous vegetation			X	X	X
		A12A6A10 // A11A1A11B4X XXXXXF2F4F7 G4-F8	Pasture/managed grassland					X
		A122(A6)A10	Natural grassland			X		X
		A12A2	Grass types for Western Africa			X		
		A12A3A11B2X XXXXXF2F4F7 G4-A12; A12A3A11B2-A13; A12A1A11	Savannas			X		
40		A11A3	Cultivated and managed vegetation/agriculture		X	X	X	X
	41	A11A3XXXXXX D3(D9)	Irrigated cropland		X			X
	42	A11A3XXXXXX D1	Rainfed cropland		X			X
	43	A11A3	Big and small farming/field size		X			
	44	A11A1-W8/A2	Permanent crops		X			X
	45	A11A3	Row crops		X			
		A11A2	Crop types: long/short cycle or winter/summer crops		X			
		A11A2	Multiple crop cycles		X			
50		B15A1	Urban/built up			X	X	X
60		B16A1(A2)	Bare/sparse vegetation				X	X
70		B28A2(A3)	Snow and Ice				X	X

Code Level 1	Code Level 2	UN LCCS level	Land cover class	Forest modelling/REDD+	Crop monitoring	Biodiversity	Monitoring Environment and Security in Africa	Climate modelling
80		B28A1	Open water				X	X
		A24A1(A2/A3/A4)	Wetland			X	X	X
		A24A3	Mangroves	X		X		

ANNEX 2: LEGENDS OF THE OTHER GLOBAL LAND COVER MAPS

Value	Label	Color
0	No Data	
10	Cropland, rainfed	
11	Herbaceous cover	
12	Tree or shrub cover	
20	Cropland, irrigated or post-flooding	
30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)	
40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)	
50	Tree cover, broadleaved, evergreen, closed to open (>15%)	
60	Tree cover, broadleaved, deciduous, closed to open (>15%)	
61	Tree cover, broadleaved, deciduous, closed (>40%)	
62	Tree cover, broadleaved, deciduous, open (15-40%)	
70	Tree cover, needleleaved, evergreen, closed to open (>15%)	
71	Tree cover, needleleaved, evergreen, closed (>40%)	
72	Tree cover, needleleaved, evergreen, open (15-40%)	
80	Tree cover, needleleaved, deciduous, closed to open (>15%)	
81	Tree cover, needleleaved, deciduous, closed (>40%)	
82	Tree cover, needleleaved, deciduous, open (15-40%)	
90	Tree cover, mixed leaf type (broadleaved and needleleaved)	
100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)	
110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)	
120	Shrubland	
121	Evergreen shrubland	
122	Deciduous shrubland	
130	Grassland	
140	Lichens and mosses	
150	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)	
151	Sparse tree (<15%)	
152	Sparse shrub (<15%)	
153	Sparse herbaceous cover (<15%)	
160	Tree cover, flooded, fresh or brakish water	
170	Tree cover, flooded, saline water	
180	Shrub or herbaceous cover, flooded, fresh/saline/brakish water	
190	Urban areas	
200	Bare areas	
201	Consolidated bare areas	
202	Unconsolidated bare areas	
210	Water bodies	
220	Permanent snow and ice	

Figure 14: Legend of the global CCI-LC maps, based on LCCS.

Figure 15: Legend of the Globeland30 2010 map