Copernicus Global Land Operations "Vegetation and Energy"

"CGLOPS-1"

Framework Service Contract N° 199494 (JRC)

PRODUCT USER MANUAL

MODERATE DYNAMIC LAND COVER 100M VERSION 2

Issue I2.10

Organization name of lead contractor for this deliverable:

Book Captain:	Dr. Marcel Buchhorn (VITO)
Contributing Authors:	Bruno Smets (VITO)
	Luc Bertels (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:	Dr. Marcel Buchhorn	Sign	Date 10/05/2019
Approval:	Roselyne Lacaze	Sign	Date
Endorsement:	Mark Dowell	Sign	Date
Distribution:	Public		

Change Record

Issue/Rev	Date	Page(s)	Description of Change	Release
Draft	24.06.2017	All	Initial Version	D1.00
Draft	25.07.2017	All	Finished all paragraphs	D1.10
	10.10.2017	All	Revision after internal review	I1.20
Draft	29.04.2019	All	Revision for algorithm version 2	D2.00
	10.05.2019	All	Revision after internal review	I2.10

TABLE OF CONTENTS

List of Figures

List of Tables

List of Acronyms

Acronym	Meaning
AGC	Alaska Geobotany Center
ANIR	Angle at NIR
ARD	Analysis Ready Data
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AUC	Area Under the Curve
BYTE	8-bit value
CEOS-LPV	Committee of Earth Observation Satellites - Land Product Validation
CCI	Climate Change Initiative
CF V1.6	Climate & Forecast conventions compliant with version 1.6
CGLS	Copernicus Global Land service
CAVM	Circumpolar Arctic Vegetation Map
DDI	Data Density Indicator
DEM	Digital Elevation Model
EO	Earth Observation
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization of the united nation
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
GHS	Global Human Settlement
GSD	Ground Sampling Distance
GSHHG	Global Self-consistent, Hierarchical, High-resolution Geography database
GSW	Global Surface Water
GUF+	Global Urban Footprint plus
GWHR	Global scale Water History Record
GZD	Grid Zone Designators
HANTS	Harmonic ANalysis of Time Series
HMC5	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 100 m resolution
LCCS	Land Cover Classification System
madHANTS	Median Absolute Deviations of HANTS
MAE	Mean Absolute Error
MC10	10-daily Median Composite
MC5	5-daily Median Composite
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NASA JPL	NASA Jet Propulsion Laboratory
NAUC	Normalized Area Under the Curve
NDMI	Normalized Difference Moister Index
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
PVC	Percentage Vegetation Cover
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 Mask
SMAC	Simplified Method for the Atmospheric Correction
SRTMGL1	Shuttle Radar Topography Mission Global 1 arc second V003 data
SVP	Service Validation Plan
SWIR	Short Wave Infra-Red reflectance
tGAP	Time series Gap
TOA	Top Of Atmosphere
TOC	Top Of Canopy
UN	United Nations
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
VI's	Vegetation Indices
VITO	Vlaamse Instelling voor Technologisch Onderzoek (Flemish Institute for Technological Research), Belgium
WGS84	World Geodetic System 1984
WSF	World Settlement Footprint

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 timely manner and are complemented by the constitution of long-term time series.

Since January 2013, the Copernicus Global Land Service is continuously providing Essential 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 (CGLS-LC100) is a new product in the portfolio of the CGLS and delivers 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 at three levels, 12 classes at level 1 up to 23 classes at level 3, classes according to the LCCS scheme. Next to these discrete classes, the product also includes continuous field layers for all basic land cover classes that provide proportional estimates for vegetation/ground cover for the land cover types. 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.)

The first Land Cover map (V1.0) was 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.

This second Land Cover map (V2.0) is provided for the 2015 reference year over the entire Globe, derived from the PROBA-V 100 m time-series, a database of high quality land cover training sites and several ancillary datasets, reaching an accuracy of 80 % at Level1. This second collection will be extended in the next few months with change demonstration maps from 2016-2018 over the African continent and a full independent validation report including spatial accuracies.

It is planned to update the change maps to cover the entire Globe by early 2020 and the yearly updates from 2020 will be continued through the use of a Sentinel time-series.

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 the Annex.

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 moderate land cover product

CGLOPS1_TrainingDataReport_LC100m_V2	Report presenting the training data set used for the dynamic moderate land cover product for Version 2 (upcoming)
CGLOPS1_ATBD_LC100_V2	Algorithm Theoretical Basis Document of the 100 m dynamic moderate land cover product for Version 2 (upcoming)
CGLOPS1_VR_LC100_V2	Report describing the results of the scientific quality assessment of the 100 m dynamic moderate land cover product for Version 2 (upcoming)

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 pre-processing including atmospheric & geometric correction and data cleaning and (temporal) outlier detection techniques,
2. Applying data fusion techniques at multiple levels,
3. Supervised classification, and
4. Including established third party datasets via expert rules.

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

1. PROBA-V UTM Analysis Ready Data (ARD) generation,
2. data cleaning & compositing,
3. data fusion as well as quality indicator generation,
4. metrics extraction,
5. training data generation,
6. ancillary datasets preparation,
7. classification / regression,
8. land cover map and cover fraction layers generation plus final quality layer assembling.

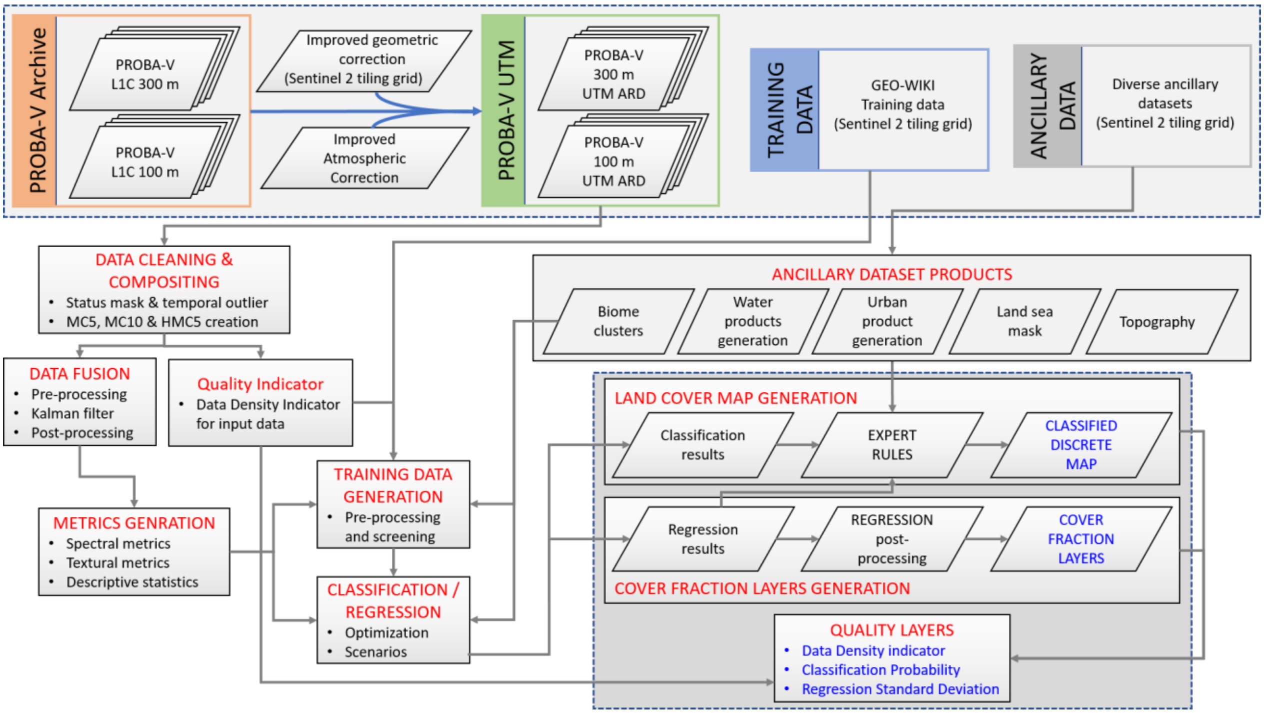


Figure 1: Workflow diagram for the CGLS Dynamic Land Cover 100 m (CGLS-LC100) products

To reduce distortion in the High North, to make our land cover products better usable with other data and to allow continuity of the service, the PROBA-V archive used as current main input data source was reprocessed with a new geometric correction and an improved atmospheric correction. The complete PROBA-V archive was in this way translated into the PROBA-V UTM ARD which is produced in Universal Transverse Mercator (UTM) coordinate system and fully aligned with the Sentinel-2 tiling grid in tiling naming as well as tile dimensions.

The PROBA-V UTM ARD main product, the 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 UTM daily multi-spectral image data with a GSD of ~0.003 degree (~300 m) secondarily. Next to a Status Mask (SM) 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. From this cleaned and outlier screened data a Data Density Indicator (DDI) is calculated which is used as a input quality indicator in the supervised learning process.

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 and textural metrics are generated. Overall, 270 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 UTM 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. The optimized training data together with the input data quality indicators (DDI dataset) are then used in a supervised classification/regression using Random Forest (RF) techniques.

Finally, we build upon the success of previous global mapping efforts and/or other ancillary datasets. Therefore, external datasets are resampled/warped to PROBA-V UTM 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 (UN) and Land 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 continuous cover fields that allow proportional estimates of cover fractions (also known as Percentage Vegetation Cover (PVC) for vegetation) of all main land cover classes. 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 only a summary of the retrieval method. More details can be found in [CGLOPS1_ATBD_LC100_V2].

2.2 THE RETRIEVAL METHODOLOGY

2.2.1 PROBA-V UTM ARD (Analysis Ready Data) generation

To avoid the geometric distortion given by data in the World Geodetic System 1984 (WGS84) coordinate system as used in version 1 of the product, we reprocessed the entire PROBA-V L1C data archive (2013 to present) into the Universal Transverse Mercator (UTM) projection. We started with the PROBA-V multi-spectral L1C satellite data since in this processing step the data was only radiometrically corrected to Top Of Atmosphere (TOA) reflectance and is still in a unprojected state given per strip/camera (Francois et al., 2014). During the new geometric correction only data on land masses was processed and fully aligned with the Sentinel-2 tiling grid in tiling naming as well as tile dimensions (110 km x 110 km) to allow continuity of the service when PROBA-V will go out of service (by end of 2020) (Figure 2).

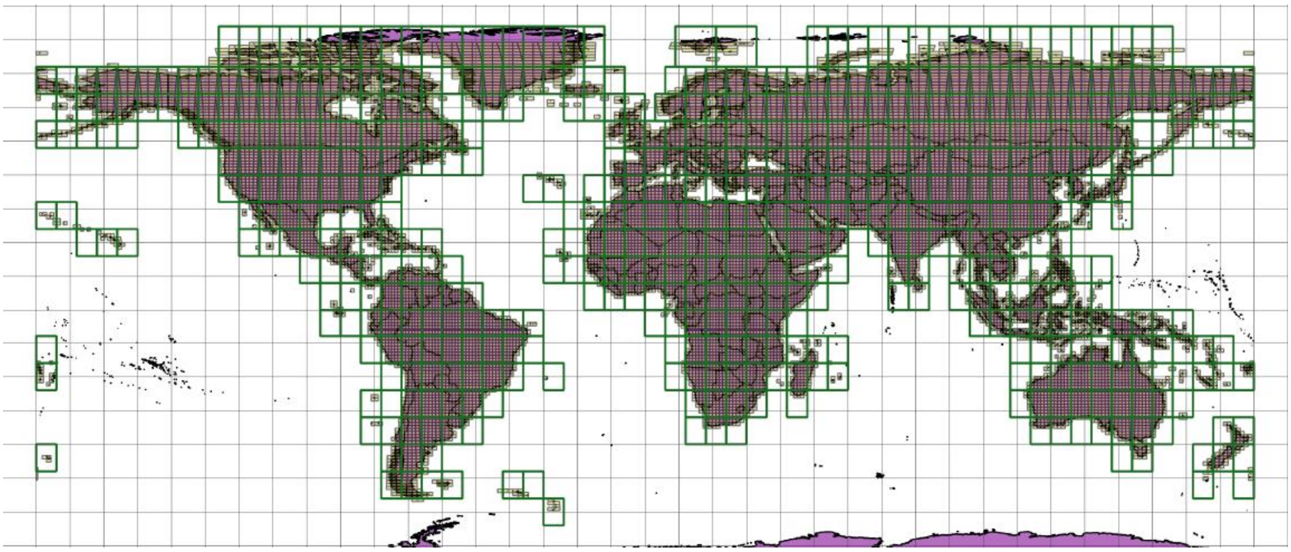


Figure 2: PROBA-V UTM tiling grid. Green rectangles show 6x8 deg UTM Grid Zone Designators (GZD's) overlaid on single 100,000-meter grid squares within the GZD's which form the single tiles.

Subsequently an improved atmospheric correction based on the original PROBA-V atmospheric correction (Sterckx et al., 2014; Dierckx et al., 2014; Wolters et al., 2017) is applied to obtain Top Of Canopy (TOC) reflectance values. Next to the improved SMAC (Simplified Method for the Atmospheric Correction) algorithm based on Rahman & Dedieu (1994), a bright surface and permanent snow mask was created based on a yearly temporal analysis to avoid errors over bright surfaces such as salt planes, permanent glaciers etc.

Afterwards all single L2B EO images were post-processed to generate daily (S1) synthesis on 100 m and 300 m GSD. The complete PROBA-V archive was in this way translated into the PROBA-V UTM ARD archive which allows more flexibility and faster data processing. The ARD generation follows the suggestions described by Dwyer et al. (2018).

2.2.2 Data cleaning & compositing

To improve further the data quality of the PROBA-V UTM ARD archive, the TOC 4-band reflectance data in the blue, red, Near Infra-Red (NIR), and Short Wave Infra-Red (SWIR) wavelength spectrum of PROBA-V UTM is outlier cleaned and composited to reduce noise. PROBA-V UTM ARD data delivers quality indicators for each pixel via a SM. The SM not only include information regarding the radiometric quality of the pixel, but also the information of the PROBA-V UTM cloud detection and retrieval algorithm (Sterckx et al., 2014; Dierckx et al., 2014; Wolters et al., 2017).

In the first step, the 100 m and 300 m S1 observations are cleaned using the SM information to remove the pixels flagged as noise, cloud, or sea. Then, an additional temporal filter called Median Absolute Deviation of Harmonic Analysis of Tile Series (madHANTS), built on a Fourier transformation based HANTS 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 3).

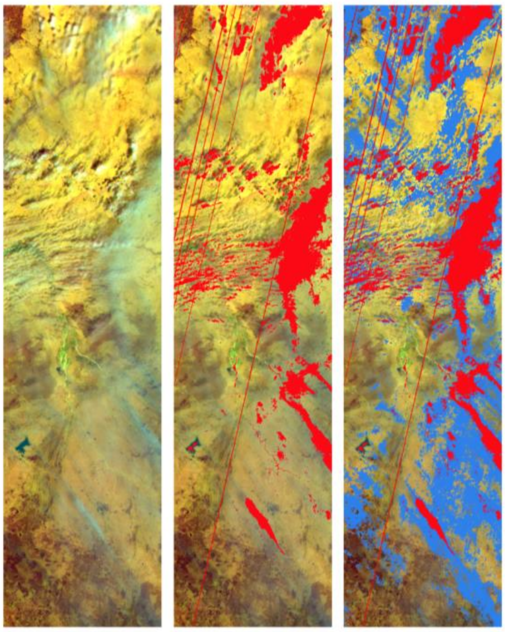


Figure 3: Example for advanced data cleaning process of PROBA-V UTM ARD 100 m image from 2016-Mar-06. Image is shown as false color composite (RGB = SWIR, NIR, blue) for a sample area in Nigeria – left) raw image, middle) status masked cleaned image (quality flagged areas are shown in red), right) final cleaned image (additional pixels which are flagged as outliers or clouds are shown in blue).

The next step includes the generation of 5-daily Median Composite images for the 100 m, and MC5 as well as 10-daily Median Composite (MC10) for the 300 m cleaned data for the epoch. 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 UTM ARD data archive (2013 – present) is produced. Therefore, the harmonized time series output produced by the HANTS algorithm is used.

After the additional data cleaning and compositing the archive is upgraded into the PROBA-V UTM ARD+ status where the plus (+) indicates the additional temporal outlier/cloud screening.

2.2.3 Data fusion

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

In the data fusion pre-processing, small gaps (5 – 10 days) in the 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 4). This is needed in order to guide the Kalman filtering approach in cases where no PROBA-V UTM ARD+ 100 m and 300 m MC5 data is available for more than 1 month for a pixel.

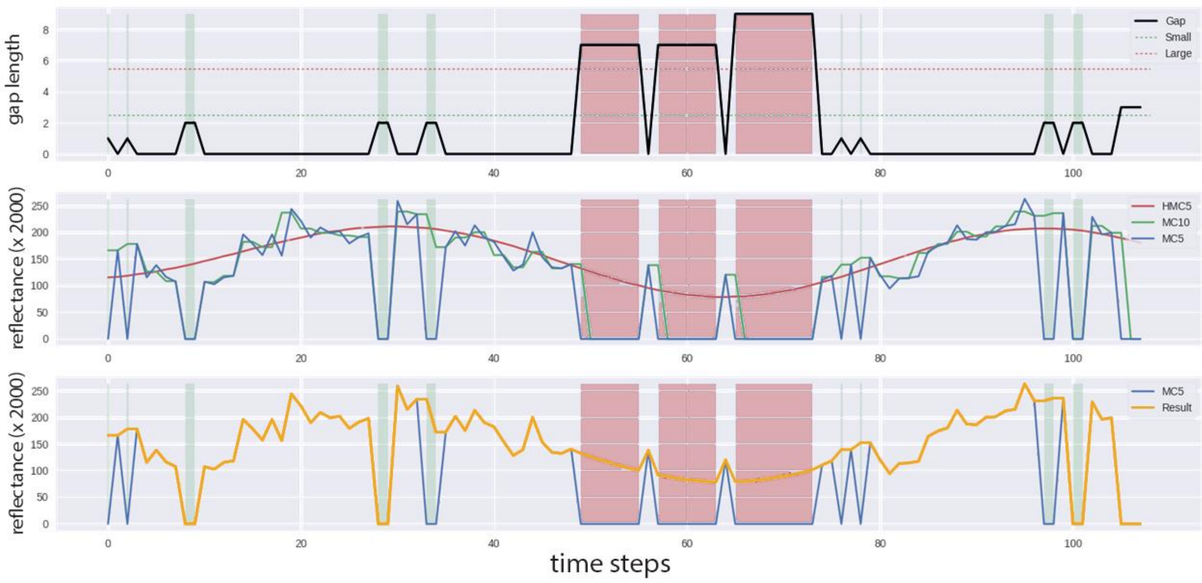


Figure 4: Example for data fusion pre-processing results for PROBA-V UTM ARD+ 300 m MC5 time series for pixel location 9.459° lon, 6.562° lat. top) continuous gap length in the time series (dashed green line indicates threshold for small gaps, dashed red line indicates threshold for big gaps), middle) 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 UTM ARD archive), bottom) 300 m time series after pre-processing (blue line shows the original 5-daily median composite time series, orange line shows the final pre-processed time series which will be used for the data fusion).

In the data fusion, the coarser resolution data (PROBA-V UTM ARD+ 300 m gap filled MC5 data) is used to estimate the pixel values in the gaps of the finer resolution data (PROBA-V UTM ARD+ 100 m MC5 data) via predictor-corrector algorithm for solving the numerical problem of the gap to the existing values outside the gap in the fine resolution data. The Kalman filter algorithm used in our workflow was introduced by Sedano et al. (2014) and was updated by VITO to improve the algorithm performance over heterogenous areas.

The last step in the data fusion is the post-processing in which the Kalman filled PROBA-V UTM ARD+ 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 epoch in 5-days intervals and with gap filled TOC reflectance data in the blue, red, NIR and SWIR wavelength region. An example for a tile in Nigeria before and after applying the Kalman filtering approach is shown in Figure 5.

After the data fusion the archive is upgraded into the PROBA-V UTM ARD++ status where the double plus (++) indicates the implemented data fusion to increase the temporal data density.

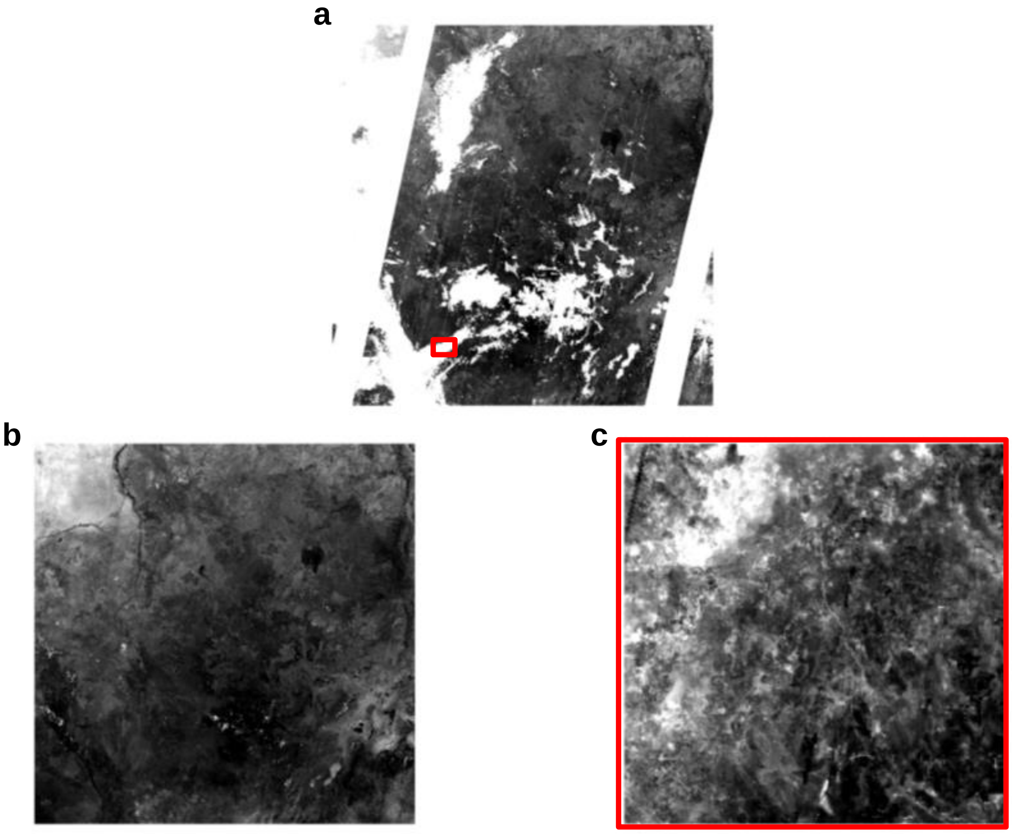


Figure 5: a) PROBA-V UTM ARD+ 100 m MC5 pre-processed red reflectance image (areas with missing data are shown in white), b) PROBA-V UTM ARD++ 100 m MC5 image after gap filling by Kalman approach, c) zoom in to full PROBA-V UTM 100 m resolution of left image of red box shown in top figure (red box had no data at all and shows after Kalman filter approach consistent image data). Example over test area in tile in Nigeria, on 2016-Mar-06.

2.2.4 Quality Indicator generation

To account in the workflow for areas with a low temporal data density in the EO input data as well as to provide the user with a quality indicator for the input data, we developed the Data Density Indicator (DDI). The DDI is calculated from the PROBA-V UTM ARD+ MC5 archive for 100 m and 300 m, so before the data fusion. Overall 19 single EO input data tests are combined via a scoring approach into the DDI which ranges from 0 to 100, where 0 indicates no existing input data and 100 the best data availability in the MC5 100 m and 300 m time series.

The single data input tests include:

- score for number of valid observations in epoch of 100 m;
- score for maximum possible observations in epoch of 100 m (length of time series);
- score for real observation ratio (valid versus all time steps in epoch of 100 m);
- score for maximum theoretical possible observations in epoch of 100 m (accounting for High Latitudes acquisitions shut off in winter);
- score for theoretical observation ratio in epoch of 100 m (valid versus theoretical possible observations);
- score for theoretical observations versus maximum observations ratio in epoch of 100 m;
- observation frequency score in epoch of 100 m (based on the number of valid observations and invalid observations);
- gap weight score in epoch of 100 m (based on the gap histogram);
- score for longest concurrent gap in days in epoch of 100 m time series;
- tGAP (time series Gap) mask score (single quality indicator from algorithm version 1);
- score for number of valid observations in epoch of 300 m;
- score for maximum possible observations in epoch of 300 m;
- score for real observation ratio epoch of 300 m;
- score for maximum theoretical possible observations in epoch of 300 m;
- score for theoretical observation ratio in epoch of 300 m;
- score for theoretical observations versus maximum observations ratio in epoch of 300 m;
- observation frequency score in epoch of 300 m;
- gap weight score in epoch of 300 m;
- score for longest concurrent gap in days in epoch of 300 m time series;

The final DDI score is used in the training of the supervised classification/regression models as an weight factor to account for the input data quality at training data locations.

2.2.5 Metrics extraction

Before the metrics extraction form the times series profiles can be carried out, a pre-processing steps is needed in which the PROBA-V UTM ARD++ reflectance data is used to generate additional Vegetation Indices (VI's) for each time step in the MC5 time series. Overall 10 additional VI's are generated and added as multi band images to the PROBA-V UTM ARD++ MC5 100 m archive:

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 Difference Moister Index (NDMI) using the NIR and the SWIR reflectance bands (Gao, 1996),
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).
8. Angle at NIR (ANIR) using the red, NIR and SWIR reflectance bands (Khanna, et al. 2007),
9. Area Under the Curve (AUC) using the blue, red, NIR and SWIR reflectance bands,
10. Normalized Area Under the Curve (NAUC) using the blue, red, NIR and SWIR reflectance bands.

For the metrics extraction, a harmonic model is fitted through each of the reflectance bands as well as each of the additional derived VI's multi band images of the PROBA-V UTM ARD++ MC5 time series sets at 100 m. 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, sum, median, 10^{th} percentile, 90^{th} percentile, 10^{th} – 90^{th} percentile range, time step (date) of the first minimum appearance, and time step (date) of the first maximum appearance are extracted for the epoch. These overall 11 descriptive metrics are again extracted for each of the four reflectance bands time series profiles 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 (4 reflectance bands, 10 VI's). 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 height, slope, aspect, and purity derived at 100 m from the NASA Shuttle Radar Topography Mission Global 1 arc second V003 data (SRTMGL1) (NASA JPL, 2013) and ASTER global DEM V2 (NASA/METI/AIST/Japan Spacesystems, and U.S./Japan ASTER Science Team, 2009) for locations in the High Latitudes.

Overall, 270 metrics (12 descriptive metrics and 7 harmonic metrics for the 14 time series sets (4 reflectance bands, 10 VI's) plus 4 topographic metrics) are generated from the PROBA-V UTM

ARD++ archive at 100 m and are input in the classification/regression step of the automated processing chain.

2.2.6 Training data generation

Training data has been collected through the Geo-Wiki engagement platform. A specific branch of Geo-Wiki (http://geo-wiki.org/) has been consolidated for collecting reference data at the required resolution and grid (PROBA-V UTM 100 m 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 10 m x 10 m 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, cropland, wetland, 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 "core" distribution of sample sites is systematic, with the same distance between sample sites, which is approximately 35 km. In addition to the "core" training data set, we collected more training sample sites for rare classes (e.g. wetland) and for areas of low accuracy (e.g. Africa).

In total, the experts have classified 150,405 unique locations. 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, and disagreeing locations with Australian land cover map and Corine Land Cover for Europe. Training sites that were impossible to identify a land cover class for by visual interpretation were removed. Final training dataset consists of approximately 141,000 sample sites.

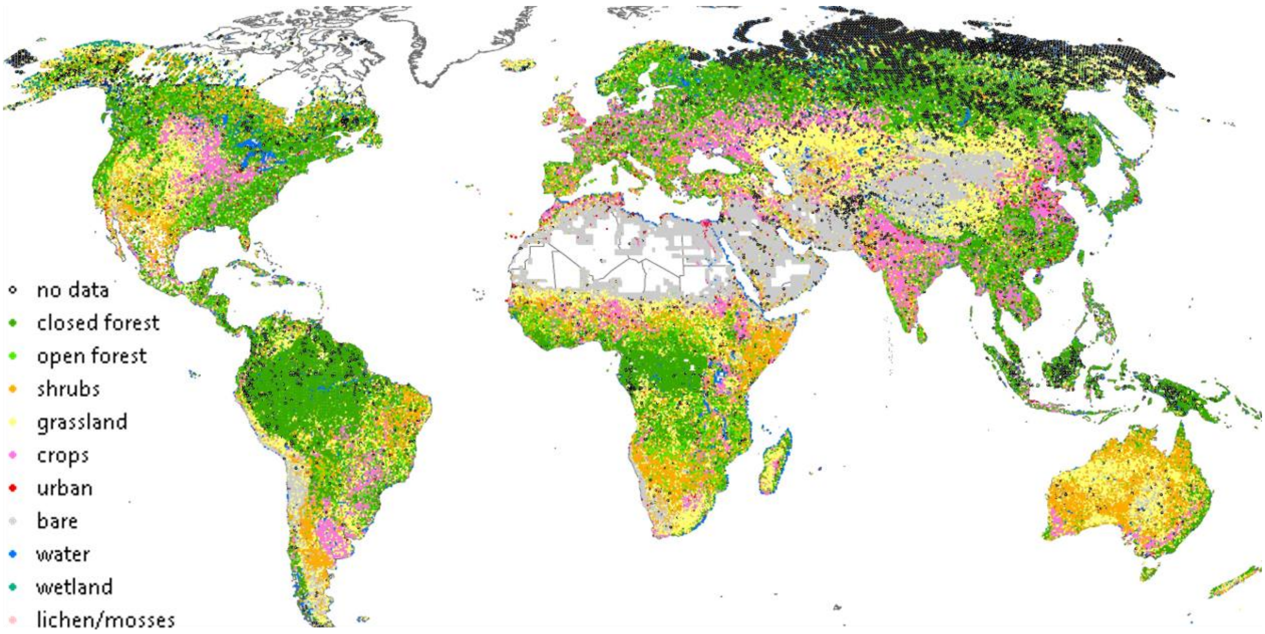


Figure 6: Training dataset used in the CGLS-LC100 workflow showing the ~141K training locations.

2.2.7 Ancillary dataset preparation

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

1. Global land/sea mask,
2. Global Surface Water (GSW) layers, i.e. permanent and seasonal water and maximum water extent,
3. Urban mask.

The global land/sea mask 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) imprinted on top of the Global Self-consistent, Hierarchical, High-resolution Geography database (GSHHG) provided by Wessel & Smith (1996). The combined global land/sea mask was only resampled to the PROBA-V UTM 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 (see below).

The global surface water layers contain the permanent and temporary water body layers on a yearly basis. 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 UTM 100 m spatial resolution. The location of temporary water bodies and herbaceous wetlands was solely computed by using Bertels et al. (2016) algorithm (called also WetProducts algorithm).

The urban (built-up) mask was generated exclusively by down-sampling the World Settlement Footprint (WSF) layer (former also known as DLR's Global Urban Footprint Plus layer (GUF+)) for 2015 (Marconcini et al., 2017a, Marconcini et al., 2017b) to the PROBA-V UTM 100 m resolution. The WSF 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.

2.2.8 Classification / regression

In order to adapt the classification/regressor algorithm to sub-continental and continental patterns, the classification/regression of the data is carried out in biome clusters defined by the combination of several global ecological layers. Overall, 73 global biome clusters have been defined merging the information of the following data sets:

- Ecoregions 2017 global dataset (Dinerstein et al., 2017)
- Global tree line layer (AGC, 2019)
- Geiger-Koeppen global ecozones after Olofsson update (Olofsson et al., 2012)
- Global land/sea mask for PROBA-V UTM

- PROBA-V UTM tiling grid and extent
- global DEM (SRTMGL1 / ASTER datasets)
- Global ecological zones for FAO forest reporting - update 2010 (FAO, 2012)
- Circumpolar Arctic Bioclimate Subzones of the Circumpolar Arctic Vegetation Map (CAVM) (CAVM Team, 2003)

For every training point location in the training dataset the 270 metrics were extracted. Next, the training data for pure endmembers (100 % coverage of a 100 x 100 pixel by one LC class) for each biome cluster was screened for inter-class outliers by analysing the spectral angle as well as the Root Mean Square Error (RMSE) of the metrics of all "pure" 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. In the next step, all final "pure" training points plus the "non-pure" training points in the biome clusters are used to rank all metrics by separability of all land cover classes as well within one land cover class for the regression. The last optimization step is the hyper-parameter search for each training data set (biome cluster specific). 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 (RF) classification is conducted for each biome cluster independently using the biome cluster 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. This class probability can be used as a quality indicator for the performed classification and is provided as a quality layer in the final product. Overall three RF classification scenarios for each biome cluster with different settings are carried out:

1. "pure class" scenario: in this setting only training points with a cover percentage of 100 % 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 100 % coverage).
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 a 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, shrubland, herbaceous vegetation, cropland, moss & lichen, and bare ground. The input are the cover fractions collected for all training points which are used in a RF regression. Overall six (Forest, Shrubland, Herbaceous vegetation, Cropland, Moss & Lichen, Bare ground) regression scenarios for each biome cluster are carried out using, in the regression model, the respective vegetation cover percentages of the training points. Moreover, the standard deviations for the six cover fraction regressions is added in the product as additional quality layers. Note that in the final product the cover fraction layers for snow, built-up (urban), permanent inland water bodies and seasonal inland water bodies are added.

2.2.9 Cover fraction layers generation

Next to the discrete land cover map, 10 cover fraction layers, where 9 explain the distribution of the main land cover classes for each pixel and one (the seasonal inland water body layer) can be used to describe the seasonal water influence on a certain area of the pixel, 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 RF regression results (see section 2.2.8). The main post-processing step after the scenario-specific regressions is a check for pixels in which the overall coverage of the 9 base cover fraction layers doesn't reach 100 %. For these pixels a normalization approach using the regression quality indicators (standard deviations of the single regressions) was developed. In a final step, metadata attributes of the Climate & Forecast conventions compliant with version 1.6 (CF V1.6) are injected. Figure 7 shows the ten cover fraction layers in a collage on global scale.

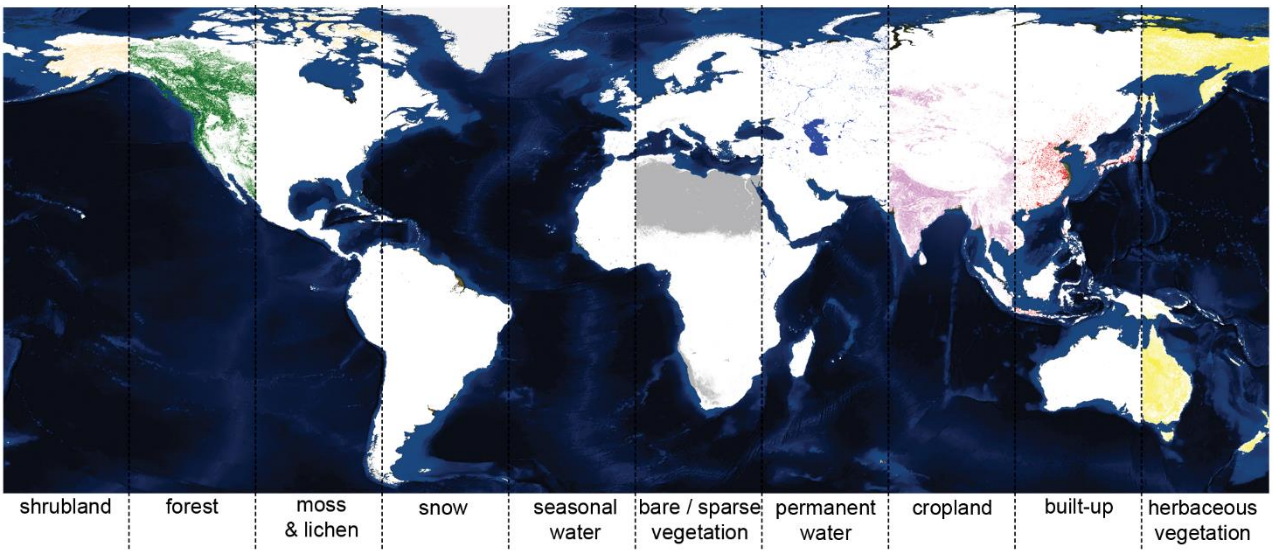


Figure 7: The cover fraction layers for the 9 base land cover classes and the seasonal inland water cover fraction of the CGLS Dynamic Land Cover product at 100 m for epoch 2015 (shown as a collage on global scale).

2.2.10 Land Cover map generation and final quality layer assembling

Expert rules are applied to combine the existing knowledge represented by the ancillary datasets (section 2.2.7) and the classification and regression results (section 2.2.8). In order to incorporate the vegetation cover fraction layers, a discrete map was generated by applying the training data rules on the 9 cover fraction layers naming forest, shrubland, herbaceous vegetation, moss & lichen, bare/sparse vegetation, cropland, built-up, snow, and permanent water bodies. In detail, during the training data generation (see section 2.2.6) 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_LC100m_V2].

The following datasets are used to generate the final CGLS-LC100 discrete map:

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 nine normalized cover fraction layers;
7. Number Of Valid Observations (NOVO) mask showing pixels with no PROBA-V UTM 100 m observations in the whole epoch (novo mask);
8. WetProduct layer with wetland mask (WetProducts); and
9. PROBA-V UTM 100 m Land-Sea-mask (Shoreline).

The predicted class probabilities are used as thresholds in the decision tree designed expert rules in order to generate the discrete LC map with 23 classes. 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 60 % from the "discrete class" classification are directly used. The discrete map generated from the nine cover fraction layers is used as a decision layer in areas were the predicted class probability from the "discrete class" classification is under 60 % and equals with the cover fraction discrete map class. Hard masks like the NOVO mask, Land-Sea-mask, and snow 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, but marked with a lower class probability. The "forest type" classification results were then used to separate the discrete classes "closed forest" and "open forest" into the different forest type classes.

In a final step, metadata attributes compliant with version 1.6 of the Climate & Forecast conventions (CF V1.6) and the colour tables 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. This classification quality layer was bundled together with the cover fraction quality layers and the Data Density Indicator layer as quality indicator for the input data as overall product quality layers.

Figure 8 shows an overview of the discrete map with 23 classes on global scale, where Figure 9 shows the legend in more detail.

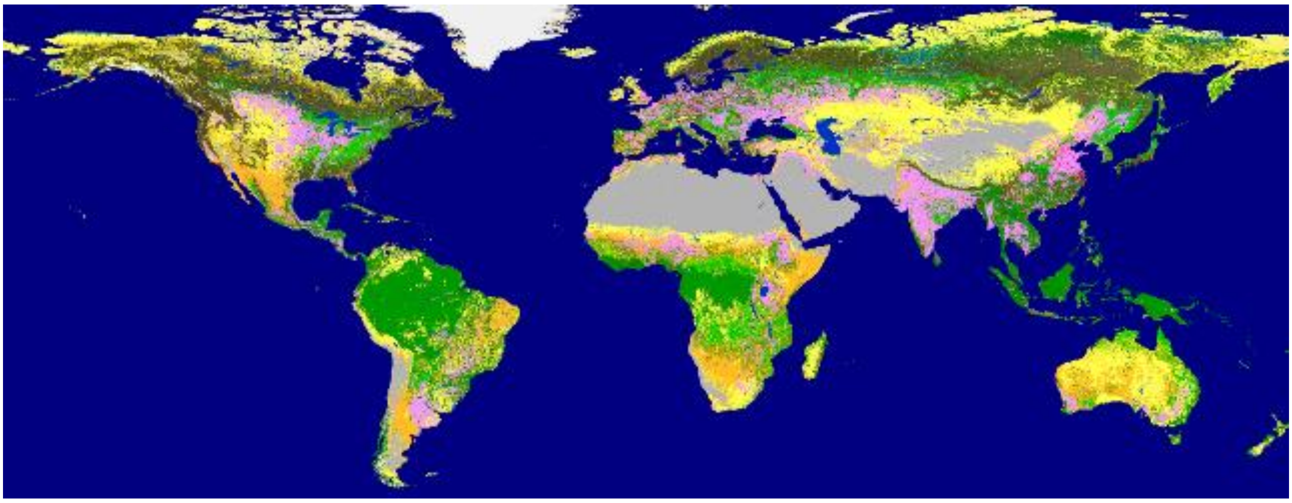


Figure 8: The CGLS Dynamic Land Cover Map at 100 m for epoch 2015 with 23 discrete classes (detailed legend in Figure 9)

Figure 9: Legend for the 23 discrete classes of the CGLS Dynamic Land Cover Map at 100 m

2.3 Limitations of the Method

- Due to the used input data the product is limed to a latitudinal extent between 60° south and 78.25° north.
- Remaining shadowed pixels in the time series not filtered out during the data cleaning process can lead to misclassifications.
- Fires (mainly appearing as dark burned areas) and other dark surfaces (e.g. volcanic soils) were not yet taken into account and therefore could lead to misclassifications.

- Ice sheets (especially over lakes in the High North) and not detected bright surfaces in the bright surface mask (e.g. salt planes or parts of it) could lead to misclassifications.
- Artefacts at boundaries between biome clusters can appear due to the manual generation of the biome cluster vector layer as well as the biome cluster 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 100 m (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 100 m. This could lead to underestimate the croplands.
- Areas with a high frequency of cloud cover (e.g. Central Africa) have a low temporal data density and are therefore sensitive to misclassifications.
- New water bodies which are not in the maximum water extent layer of the GSW product will not be detected and will result in omission errors.
- Some limitations are due to the legend or class definition of the discrete classes:
    - 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

Collection	Coverage	Status	Main characteristics
V1	Africa 2015	Demonstration	PROBA-V S1 time-series (plate carree) Random Forest
V2	Global 2015 Africa 2016-2018 (upcoming)	Operational Demonstration	PROBA-V L1C time-series (gridded to Sentinel-2 UTM) Random Forest
V3 (upcoming)	Global 2015-2019	Operational	PROBA-V L1C time-series (gridded to Sentinel-2 UTM) Random Forest

2.5 ROADMAP

The Collection 100 m of LC product is generated from the PROBA-V 100 m and 300 m sensor data and will cover the years 2015-2019.

The Copernicus Global Land service will continue the 100 m production, from 2020, through the combination of Sentinel-1 and Sentinel-2 mission. The adaptation of the retrieval methodology to the Sentinels is currently under test.

3 PRODUCT DESCRIPTION

The Global 100m Land Cover version 2 products are provided in **ZIP** files, per 20 x 20 degree tile (Figure 10), each containing a set of **GeoTIFF** files for each of the following **20 layers**:

- Discrete classification
- Classification probability (quality indicator) of the discrete classification
- Forest type
- Data Density Indicator
- Fractional cover for 10 classes
- Standard deviation (quality indicator) for the fractional cover of six (6) classes, that were computed via regression of the percentage of cover

3.1 FILE NAMING

The ZIP files and GeoTIFF files they contain follow this naming convention:

<TILE>_<SENSOR>_LC100_epoch<YEAR>_global_<VERSION>_<LAYER>_<CRS>.tif

where

- <TILE> the designation of the 20 x 20 degree tile, composed of the 3-digit longitude and 2-digit latitude of the top-left corner (see Figure 10)
  Example: W180N80 for the tile covering the area from 180W to 160W and 80N to 60N.
- <SENSOR> the EO sensor used, “ProbaV”.
- “LC100” indicates this is a 100 m resolution Land Cover product
- epoch< YEAR> indicates the epoch year in four digits.
- “global” indicates that the tile is part of a Land Cover product that covers the global land surface.
- <VERSION> shows the processing line version used to generate this product. The version denoted as vM.m.r (e.g. v2.0.1), with ‘M’ representing the major version (e.g. v2), ‘m’ the minor version (starting from 0) and ‘r’ the production run number (starting from 1) (Table 1).
- <LAYER> gives the name of the data layer (see Table 2)
- <CRS> is the coordinate reference system used. The current tiles are provided in EPSG:4326, geographic latitude/longitude CRS.

Figure 10: Scheme of the 20 x 20 degree tiles

Table 1: Version numbering and recommendations for handling version updates

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, updates in input data.	Later run is a drop-in replacement of all former runs.

Table 2: Land Cover layer names in the filename

Layer in filename	Description
discrete-classification	Main discrete classification according to FAO LCCS scheme
discrete-classification-proba	Classification probability, a quality indicator for the discrete classification
forest-type-layer	Forest type for all pixels where tree cover fraction is bigger than 1 %
bare-coverfraction-layer	Fractional cover (%) for the bare and sparse vegetation class
crops-coverfraction-layer	Fractional cover (%) for the cropland class
grass-coverfraction-layer	Fractional cover (%) for the herbaceous vegetation class
moss-coverfraction-layer	Fractional cover (%) for the moss & lichen class
shrub-coverfraction-layer	Fractional cover (%) for the shrubland class
snow-coverfraction-layer	Fractional cover (%) for the snow & ice class
tree-coverfraction-layer	Fractional cover (%) for the forest class

Layer in filename	Description
urban-coverfraction-layer	Fractional cover (%) for the built-up class
water-permanent-coverfraction-layer	Fractional cover (%) for the permanent inland water bodies class
water-seasonal-coverfraction-layer	Fractional cover (%) for the seasonal inland water bodies class
bare-coverfraction-StdDev crops-coverfraction-StdDev grass-coverfraction-StdDev moss-coverfraction-StdDev shrub-coverfraction-StdDev tree-coverfraction-StdDev	Quality indicator (standard deviation) of the percentage vegetation cover regression, for the respective class.
DataDensityIndicator	Data density indicator showing quality of the EO input data between 0 – 100 (o = bad, 100 = perfect data)

3.2 File Format

The Land Cover 100 m layers are provided as single-band GeoTIFF files that are internally compressed, include overviews on levels 2, 4, 8 and 16 for faster loading in GIS and include standard metadata attributes.

Note that this format may not be fully compliant with Cloud-Optimized GeoTIFF (COG) requirements. The GeoTIFF format will be further improved towards COG in upcoming Land Cover products.

3.3 Product Content

All land cover layers are stored as single bytes per pixel, without scaling or offset.

3.3.1 Discrete classification

The discrete classification map provides 23 classes (Table 3) and is defined using the Land Cover Classification System (LCCS) developed by the United Nations (UN) Food and Agriculture Organization (FAO). 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 contains classes with codes that are multiples of ten (10, 20, 30, etc.).
- The "level 2", also known as regional legend, has class codes of two digits that is not a multiple of ten (i.e. 11, 12 are sub-classes of 10, and so on).
- The "level 3" classes have three digits (i.e. 111 – 116 and 121 – 126) and are used to further distinguish the forest types (sub-classes of 11 – open forest and 12 – closed forest).

The discrete map is coded with special values 200 for sea pixels and 0 signifying missing input data (i.e. not observed by PROBA-V sensor).

Table 3: Discrete classification coding

Map code	UN LCCS level	Land Cover Class	Definition according UN LCCS	Color code (RGB)
0	-	No input data available	-	40, 40, 40
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.	88, 72, 31
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	112, 102, 62
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
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
115	A12A3A10	Closed forest, mixed	Closed forest, mix of types	78, 117, 31
116	A12A3A10	Closed forest, unknown	Closed forest, not matching any of the other definitions	0, 120, 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.	102, 96, 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	141, 116, 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.	141, 180, 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.	160, 220, 0
125	A12A3A12	Open forest, mixed	Open forest, mix of types	146, 153, 0
126	A12A3A12	Open forest, unknown	Open forest, not matching any of the other definitions	100, 140, 0
20	A12A4A20B3(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
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
100	A12A7	Moss and lichen	Moss and lichen	250, 230, 160

Map code	UN LCCS level	Land Cover Class	Definition according UN LCCS	Color code (RGB)
60	B16A1(A2)	Bare / sparse vegetation	Lands with exposed soil, sand, or rocks and never has more than 10 % vegetated cover during any time of the year	180, 180, 180
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	250, 0, 0
70	B28A2(A3)	Snow and Ice	Lands under snow or ice cover throughout the year.	240, 240, 240
80	B28A1B1	Permanent water bodies	lakes, reservoirs, and rivers. Can be either fresh or salt-water bodies.	0, 50, 200
200	B28A1B1[1]	Open sea	Oceans, seas. Can be either fresh or salt-water bodies.	0, 0, 128

3.3.2 Fractional Cover layers (cover fractions)

The Fractional Cover layers, also referred to as cover fractions, give the percentage of a 100 m pixel that is filled with a specific land cover class (forest/trees, herbaceous vegetation, shrub, etc.). As such it provides more detailed information than the dominant class that is shown in the discrete classification.

The Fractional Cover layers are coded as a number between 0 and 100, in steps of 1 %. All values above 100 are invalid, missing values set to 255.

3.3.3 Quality layers

3.3.3.1 Probability of the discrete classification

The probability of the discrete classification indicates the quality of the discrete classification and is provided as a number between 0 and 100, in steps of 1 %. All values above 100 are invalid and value 255 is used for missing values.

3.3.3.2 Standard Deviation for the Fractional Cover

The Standard Deviation of the percentage cover regression indicates the quality of the associated Fractional Cover layer. It is provided as a number between 0 and 100, in steps of 1 %. All values above 100 are invalid and value 255 is used for missing values.

[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.

3.3.3.3 Data Density Indicator (DDI)

The Data Density Indicator indicates the availability of input data from the PROBA-V UTM ARD+ MC5 archive for 100 m and 300 m resolutions. It is a score between 0 = no input data available and 100 = best data availability. All values above 100 are invalid and missing DDI values are coded as 255.

3.3.4 Forest type layer

The Forest Type layer provides discrete values per type of forest (see Table 4), for all pixels where the tree (forest) cover fraction exceeds 1 %.

Table 4: Forest type coding

Value	Short name	Description
0	Unknown	Doesn't match any of the other types
1	ENF	Evergreen needle leaf forest
2	EBF	Evergreen broad leaf forest
3	DNF	Deciduous needle leaf
4	DBF	Deciduous broad leaf
5	Mixed	Mix of forest types

3.3.5 Metadata attributes

The GEOTIFF files provide the metadata attributes as key value pairs, according to the Climate and Forecast Convention (CF, version 1.6):

- on the file-level (Table 5);
- on the band-level, with an example values given for the main discrete classification layer (Table 6.

Table 5: Description of GEOTIFF file attributes

Attribute name	Description	Example(s)
archive_facility	Specifies the name of the institution that archives the product	VITO NV
copyright	Text to be used when referring to this product in publications (copyright notice)	Copernicus Service information 2019
creator	Principal investigator of the processing algorithm	Dr. Marcel Buchhorn (VITO)
delivered_product_crs	Land Cover product is delivered in this Coordinate Reference System	WGS84 (EPSG:4326)
delivered_product_grid	Land Cover product is delivered in this tile grid	global 20x20 deg tiling grid
file_creation	File creation timestamp	Fri Apr 19 11:46:08 2019
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.	2019-04-26 Processing line LC100
Info	Additional comment on the processing history.	MasterTile W160N20 for product discrete-classification of CGLOPS LC100 layers for epoch 2015.
institution	The name of the institution that produced the product	VITO NV
long_name	Extended product name	Land Cover
orbit_type	Orbit type of the orbiting platform(s)	LEO
platform	Name(s) of the orbiting platform(s)	Proba-V
processing_level	Product processing level	L3
processing_mode	Processing mode used when generating the product (Near-Real Time, Consolidated or Reprocessing)	offline
production_crs	Coordinate Reference System used for the pre-processed input data and during the different production steps	UTM
production_grid	Grid used for the pre-processed input data and during the different production steps	MGRS (Sentinel-2 tiling grid)
product_version	Version of the product	V2.0.1
references	Published or web references with more product information.	https://land.copernicus.eu/global/products/lc
region_name	Name of the geographic area covered, e.g. name of the 20x20 degree tile	W160N20
sensor	Name(s) of the sensor(s) used	VEGETATION
source	The method of production of the original data	Derived from 100m EO satellite imagery
time_coverage_end	End date and time of the temporal coverage of the input data.	2016-12-31T23:59:59Z
time_coverage_start	Start date and time of the temporal coverage of the input data.	2014-01-01T00:00:00Z
title	A description of the contents of the file	Dynamic Land Cover Map 100m 2015

Table 6: Description of GEOTIFF band attributes.

Attribute	Description	Examples for LCCS layer
CLASS	Dataset type	DATA
band_crs	Coordinate Reference System used for this GeoTIFF band.	WGS84 (EPSG:4326)
flag_meanings	Description for each flag value	unknown, ENF_closed, EBF_closed, DNF_closed, DBF_closed, mixed _closed, unknown_closed, ENF_open, EBF_open, DNF_open, DBF_open, mixed _open, unknown_open, shrubland, herbaceous_vegetation, cropland, built-up, bare_sparse_vegetation, snow_ice, permanent_inland_water, herbaceous_wetland, moss_lichen, sea
flag_values	Provides a list of the specific values used.	0, 111, 112, 113, 114, 115, 116, 121, 122, 123, 124, 125, 126, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200
long_name	A non-standardized, descriptive name that indicates a variable's content.	Land Cover Classification
missing_value	Single value, outside of valid_range, used to represent missing or undefined data, for applications following older versions of the standards.	255
short_name	A shortened, non-standardized name.	discrete-classification
unit	Physical unit. None or omitted when the data is dimensionless.	None
valid_range	Smallest and largest valid values.	0, 254

3.4 PRODUCT CHARACTERISTICS

3.4.1 Projection and Grid Information

The Land Cover products are delivered in a regular latitude/longitude grid (EPSG:4326) with the ellipsoid WGS 1984 (Terrestrial radius=6378 km). The resolution of the grid is 1°/1008 or approximately 100 m at equator.

3.4.2 Spatial Information

The Land Cover products cover the geographic area from longitude 180°E to 180°W and latitude 78.25°N to 60°S. They are provided in 20 x 20 degree tiles (see Figure 10).

3.4.3 Temporal Information

The Land Cover product is provided with yearly updates, each representing the land cover for the epoch or reference calendar Year (from 01 January to 31 December). The data 1 year prior and pastor the reference year is used in its processing. As such, the temporal coverage provides a start date of 01 January Year-1 to 31 December Year+1.

3.5 DATA POLICIES

EU law[2] grants free and open access to Copernicus Sentinel Data and Service Information, which includes Global Land Service products, for the purpose of the following use in so far as it 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 (Sentinel Data or) 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.

[2] European Commission, Regulation (EU) No 377/2014 and Commission Delegated Regulation (EU) No 1159/2013.

Where the user communicates to the public on or distributes the **original** Land Cover products, he/she is obliged to refer to the data source with (at least) the following statement (included as the copyright metadata item):

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 product was generated by the Global component of the 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 Belgian Science Policy Office (BELSPO) 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 scientific & technical support (help desk) contact specified in the next section.

3.6 ACCESS AND CONTACTS

The Land Cover products are available through the Global Land Cover viewer, available at https://land.copernicus.eu/global/lcviewer (see Figure 11). It displays the various land cover layers (discrete map, cover fractions, false-colour combinations of cover fractions) on a map, allows to download the data in 20x20 degree tiles and reports on land cover statistics per administrative area.

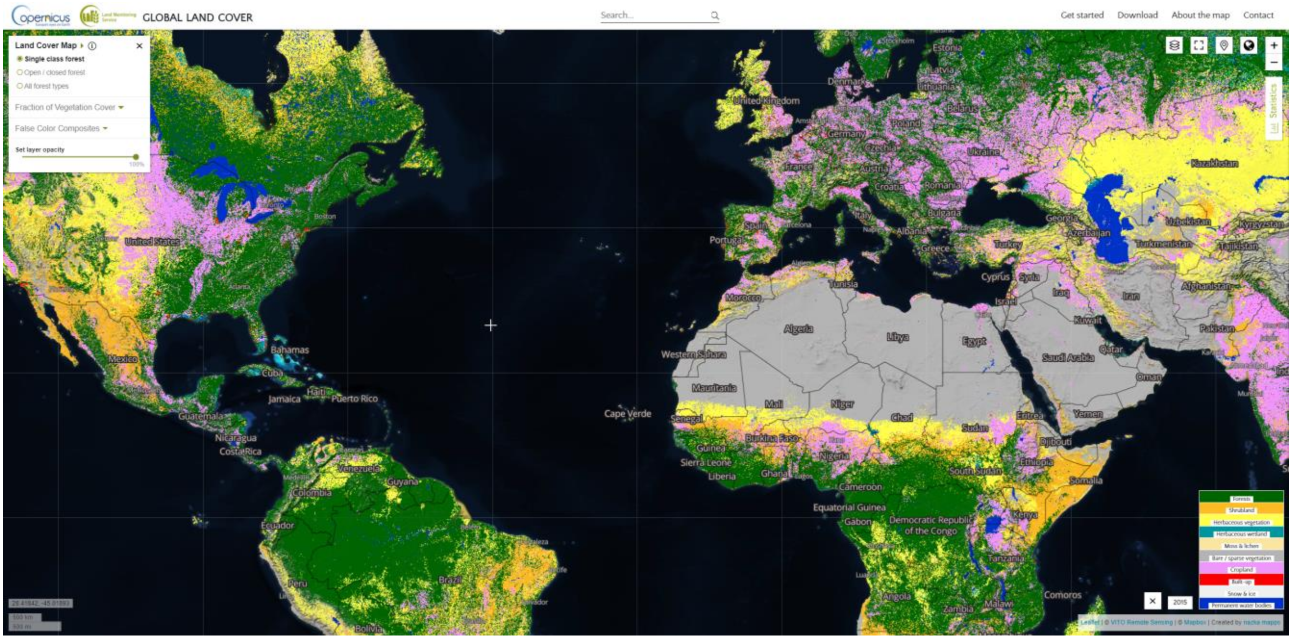


Figure 11 Screenshot of the Global Land Cover viewer

More information and documentation about the product is available from the Copernicus Global Land Service web site at https://land.copernicus.eu/global/products/lc

Accountable Contact: European Commission, Directorate-General Joint Research Centre, Italy Email address: copernicuslandproducts@jrc.ec.europa.eu

Scientific & Technical support contact e-mail address:
https://land.copernicus.eu/global/contactpage.

4 VALIDATION RESULTS

CGLS-LC100 discrete map and cover fraction layers were assessed quantitatively and qualitatively.

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 CGLS-LC100 land cover maps were assessed using an independent validation dataset. The detailed information on the sample selection and reference data collection can be found in Tsendbazar et al. (2018) which explains the validation data collection and application for Africa. The validation data collection in other continents followed the same approach as Africa (Figure 12). Due to its exceptionally large size, Asia was divided into 2 sub-continents. Northern Eurasia included Russia (both European and Asian part), Kazakhstan and Mongolia.

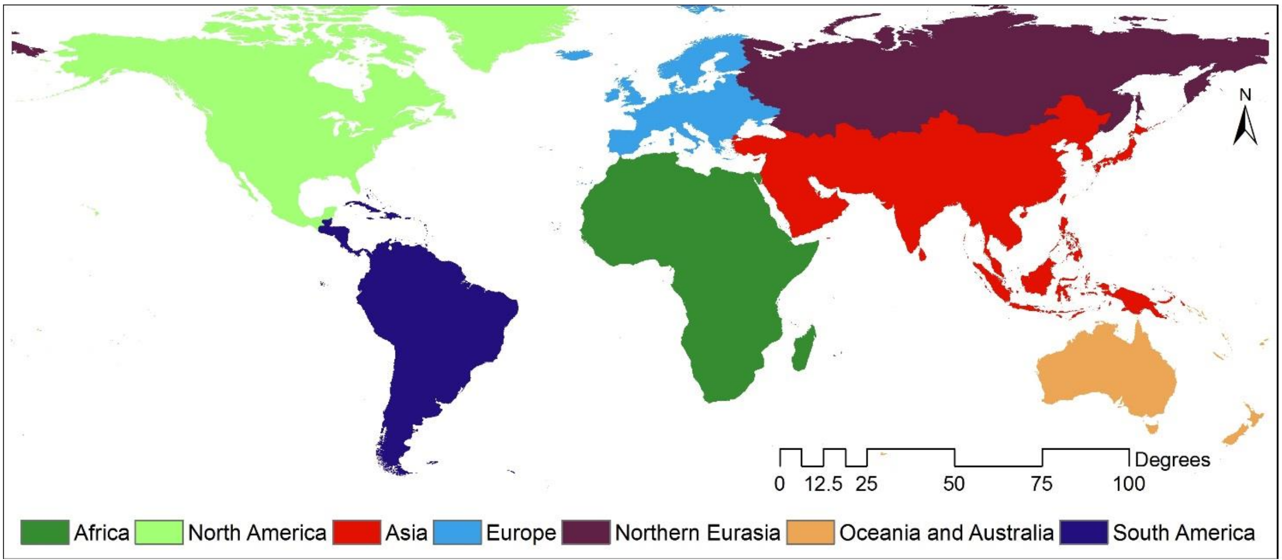


Figure 12. Continents used for validation data collection

The validation dataset contains 20,019 sample sites across the world with sample sites ranging from 2,600 to 3,600 for each continent (Figure 12). The validation sites were selected based on a global stratification stratification using Köppen climate zones and human population density (Olofsson et al., 2012). Stratification map including the selected validation sites are depicted in Figure 13.

Additional validation sites yet to be collected for rare land cover types such as urban, waterbody and wetland. The final results on the validation including the extra validation sites for rare land cover types will be included in the full validation report [CGLOPS1_VR_LC100_V2].

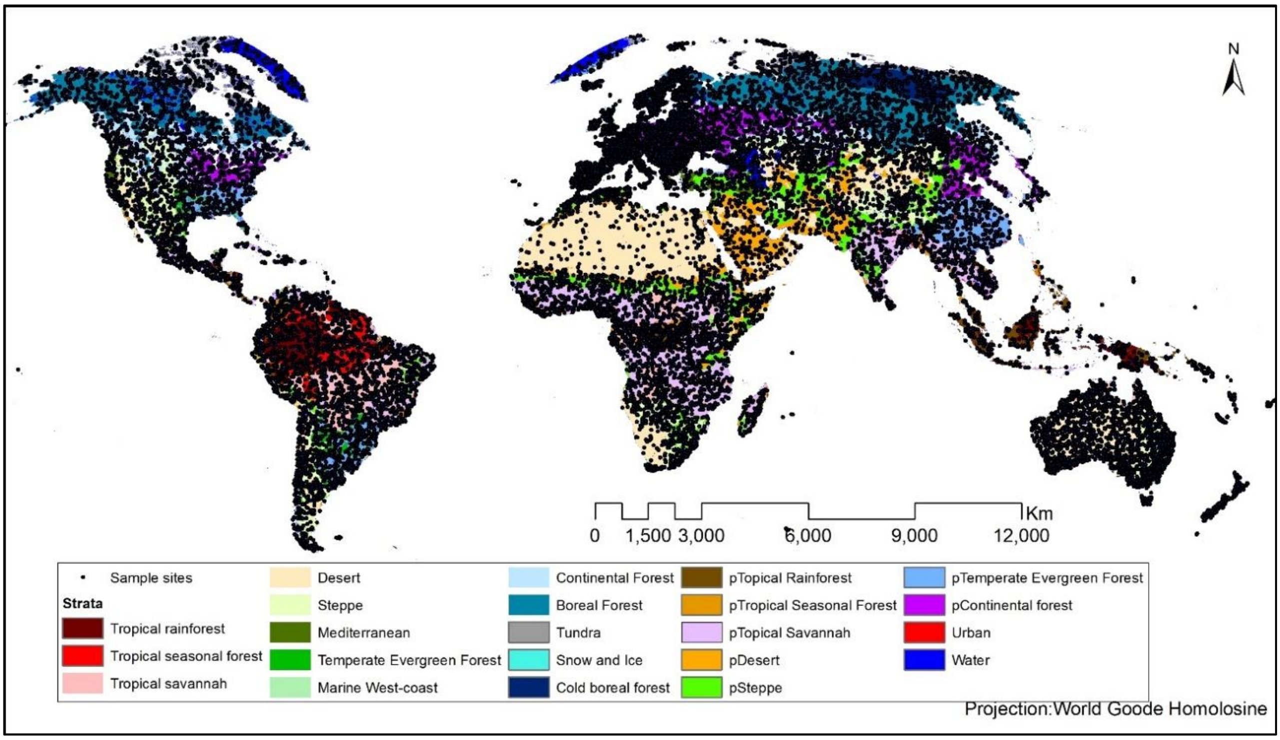


Figure 13. Stratification and selected validation sites

Sample unit areas match with a UTM based PROBA-V data pixel (100 m x 100 m). At each sample unit, 10 x 10 subpixels were created and reference information on the land cover was collected by around 30 experts that have knowledge on different regional landscapes of the world 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 CGLS-LC100 discrete map legend based on the legend descriptions. This was used to validate the CGLS-LC100 discrete map. Land cover type fraction information was directly used to assess other fraction layers.

Based on the current validation dataset, the overall accuracy of the CGLS-LC100 discrete map at Level 1 is **80.1+/-0.7 %**. Fraction cover layers have mean absolute error of 0.1 – 17.3 for different land cover types.

The results presented further in this chapter summarizes the main results. More details on this will be included in the full validation report [CGLOPS1_VR_LC100_V2].

4.1 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].

The validation dataset used for statistical accuracy assessment is shown in Figure 14.

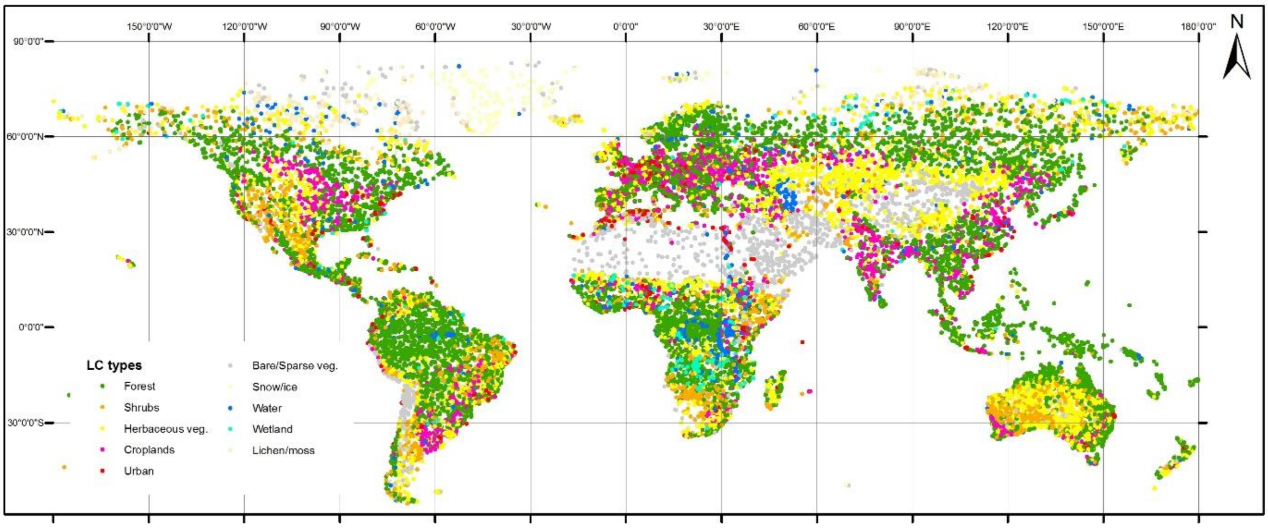


Figure 14. Distribution of validation sites and their respective land cover types

4.1.1 Accuracy assessment of CGLS LC100 discrete map

At each validation sample location, mapped land cover types were extracted. Out of 20,019 validation sites, 56 validation sites had no data in the CGLS-LC100 product. These location are mostly in the extreme north, outside the mapped region of the CGLS-LC100 product. Therefore, overall 19,863 sites were used to validate the map.

Based on the mapped and reference land cover types, a confusion or an error matrix was calculated. This error matrix was corrected by sample inclusion probabilities (Tsendbazar et al., 2018).

Confusion matrix for 10 general land cover types at Level 1 is shown in Table 7. More detailed explanation on the accuracy assessment will be included in the full validation report [CGLOPS1_VR_LC100_V2].

Table 7: Confusion matrix for discrete CGLS-LC100 map at Level 1

	Forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/sparse veg.	Snow/ice	Water	Wetland	Lichen/ moss	Correct	Total	User's accuracy	Confidence interval ±
Forest	6585	376.6	235.5	117.4	8.6	0.3		8.8	17.9	19.4	**6585**	7369.6	**89.4**	0.8
Shrubs	304.1	1144.4	279.1	52	1.3	29.5		3.4	21.1	10.6	**1144.4**	1845.4	**62**	3.2
Herbaceous veg.	243.9	465.8	2742.8	127.6	4.4	99.3		10.3	65.9	205	**2742.8**	3965	**69.2**	1.8
Croplands	227.6	88.9	322.9	1624.9	10.8	6.2		11.1	21.4		**1624.9**	2313.6	**70.2**	2.1
Urban	13.7	1.9	7.4	3.1	99.1	0.2					**99.1**	125.5	**79**	7
Bare/sparse veg.	6.4	61	157.5	7.4	5	2647.8	1.4	3.6	2.7	9	**2647.8**	2901.9	**91.2**	1.8
Snow/ice						22.6	382	0.2			**382**	404.8	**94.4**	3.7
Water	9.2	1.7	1.9	1.9		7.3		418.7	1.6		**418.7**	442.2	**94.7**	2.2
Wetland	16.2	13	55	5.4	0.1	0.4		14.8	85.1	17.2	**85.1**	207.2	**41.1**	7.9
Lichen/moss			30.5			73.1	1.4	5.4		177.4	**177.4**	287.8	**61.6**	7.2
Correct	**6585**	**1144.4**	**2742.8**	**1624.9**	**99.1**	**2647.8**	**382**	**418.7**	**85.1**	**177.4**				
Total	7406.1	2153.3	3832.7	1939.8	129.2	2886.6	384.9	476.3	215.6	438.6				
Producer's accuracy	**88.9**	**53.1**	**71.6**	**83.8**	**76.7**	**91.7**	**99.3**	**87.9**	**39.4**	**40.5**				
Confidence interval ±	0.8	2.9	1.9	2	8.3	1.5	0.8	3.4	7.9	5.5				
Overall accuracy													**80.1**	0.7

Overall map accuracy is 80.1 % +/-0.7 % (confidence intervals at 95 % confidence level).

In terms of class specific accuracies, forest, bare/sparse vegetation, snow/ice and water classes are mapped with very high accuracies. The class accuracies of herbaceous vegetation, croplands and urban are moderate, while wetlands, lichen/moss and shrubs class have lower class accuracies.

The overall accuracies and confusion matrices were also calculated for each continent (Table 8). All continents are mapped with overall accuracies around 80 % with lowest 76.7 % for North America and highest 83.4 % for Asia at Level 1. Confusion matrices of at continental levels are detailed in Annex 2.

Table 8. Overall accuracy of the CGLS-LC100 discrete map L1 per continent

	N sample	Overall accuracy	Confidence intervals ±
Asia	2732	83.4	1.5
Eurasia	2673	79.4	1.7
Europe	2714	80.4	1.6
North America	2691	76.7	1.7
Oceania & Australia	2703	81.8	1.9
South America	2734	79.4	1.5
Africa	3616	80.1	2

The accuracy of the CGLS-LC100 discrete map at level 2 including the continental accuracies are shown in Appendix 2. At level 2, when open and closed forests are separated, the overall accuracy was 75 % +/-0.8 %, ranging from 70.6-79.6 for different continents.

4.1.2 Accuracy assessment of CGLS LC100 cover fraction maps

The CGLS LC100 cover fraction maps 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. Note that additional validation data collection for rare classes will have impact on the below accuracies particularly classes such as bare, snow, built up and water.

Table 9: Accuracy of the cover fraction products.

	Trees	Shrub	Herbaceous vegetation	Bare	Crops	Lichen /moss	Snow	Built-up	Water
Mean absolute error (MAE)	9.10	9.30	17.30	5.50	5.10	2.90	0.10	0.80	0.80
Root mean square error (RMSE)	18.00	17.40	28.10	15.60	15.80	15.30	3.20	5.60	5.90

Among the cover fraction products, snow, built-up, water and lichen/moss fraction maps have lowest errors. On the other hand, herbaceous vegetation fraction layer had the highest error with MAE of 17.3 and RMSE of 28. 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. At continental scale, mean absolute error of each cover fraction layers was also calculated (Table 10).

Table 10 Mean absolute error of the cover fraction layers

	asia	eurasia	europe	north america	oceania	south america	africa
tree	7.8	11.8	11.6	8.4	9	9.7	7.9
shrub	5.9	8.3	7.8	9.6	11.8	11.4	11.4
grass	12.6	23.5	17.9	20.2	18.5	15.6	15.6
crops	7.4	2.7	10.7	3.7	2.3	4.9	5.7
moss	0.1	5.9	1.1	11	0.1	0.2	0
bare	5.5	3.8	2.7	8.1	9.1	4.9	4.9
snow	0.1	0	0.1	0.6	0	0.1	0
built.up	1.6	0.2	2.7	0.6	0.2	0.3	0.7
water	0.8	1.1	0.7	1.3	0.4	0.8	0.4

4.2 Qualitative Assessment

A visual comparison shows that the CGLS-LC100 discrete map agrees with the LC-CCI 2015 and Globeland30 2010 in terms of the spatial distribution of the main land cover types. 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. As expected, there are also differences between the products in some parts of the world.

The separation of natural vegetation appears to be 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 15). Similarly, in Northern Amazonia in South America, large grassland regions were captured in the CGLS-LC100 and Globeland30 2010, while the area was mapped as mostly shrubs in the LC-CCI 2015 (Figure 15). The CGLS-LC100 mapped less area as wetland than the other two maps. This could be partly under-estimation of wetland areas and partly due to the legend difference between the maps as the CCI-LC2015 and Globeland30 2010 wetland classes include flooded trees and shrubs, while the CGLS-LC100 depicts only flooded grassland.

Figure 16 shows the comparisons in Rondonia and Mato Grosso, Brazil and regions around Uzbekistan. In Rondonia and Mato Grosso, Brazi, most areas of grassland were captured correctly in the CGLS-LC100 map and the Globeland30 2010 map, while this area is mapped mostly as 'mosaic of cropland and natural vegetation' in the LC-CCI 2015. The spatial pattern of cropland delineation shows agreement between the products ('rainfed cropland; for LC-CCI 2015). Similarly, the spatial pattern of cropland type also agrees in regions around Uzbekistan. However, there are disagreements in delineation of sparse vegetation in LC-CCI 2015 and grassland class in the CGLS-LC100 and Globeland30 maps. Furthermore, shrubland areas in Kyzyl Kum are delineated in the CGLS-LC100 and LC-CCI 2015 while it was underestimated in the Globeland30 map. Shrubland areas appear to be overestimated in Southern part of Kazakhstan in the LC-CCI 2015 map when compared against a high resolution Google Map images.

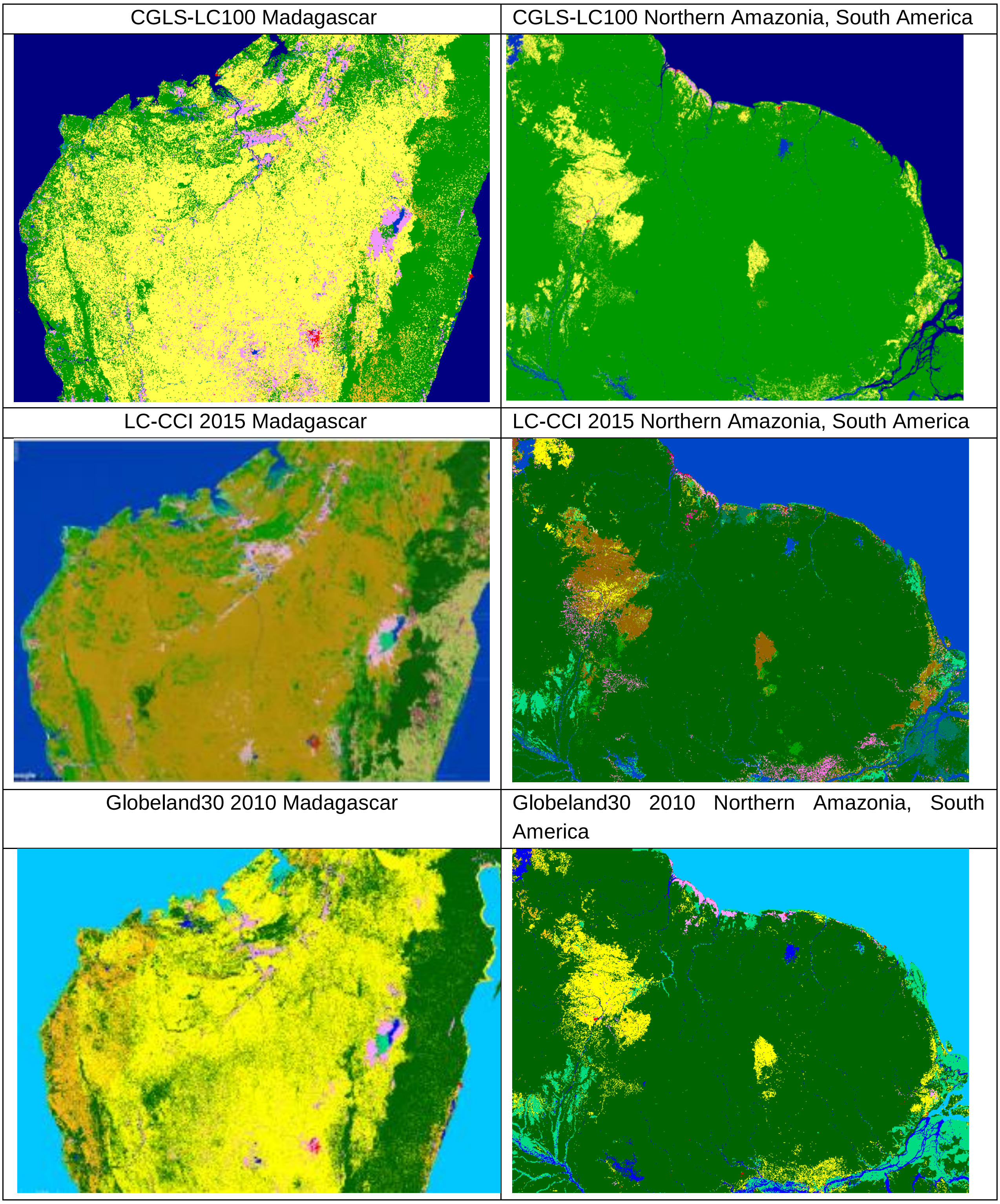


Figure 15: Comparison of three maps in Madagascar and Northern Amazonia, South America (legends are shown in Annex 3)

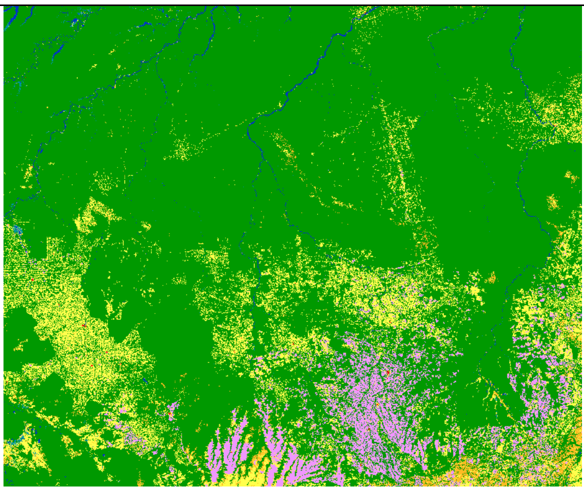
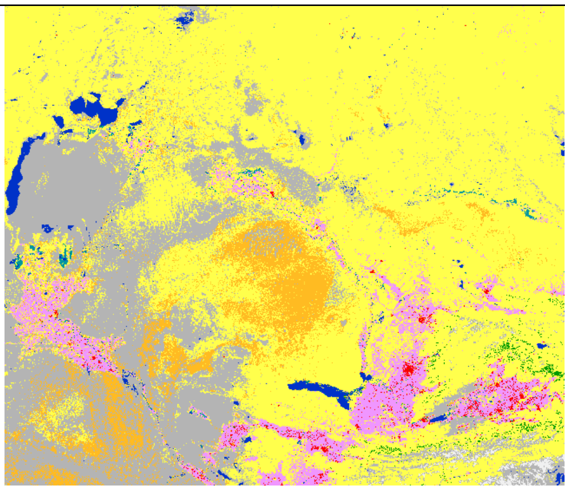
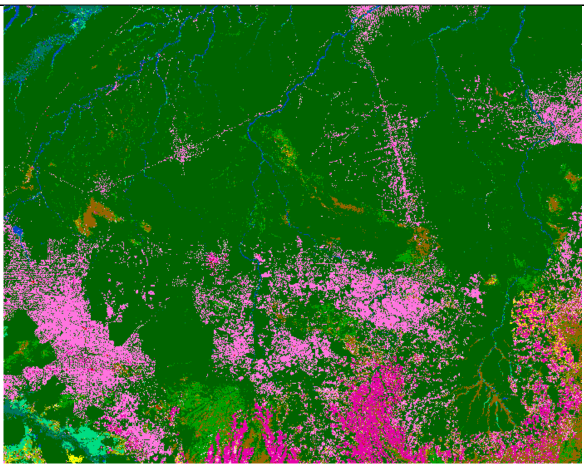
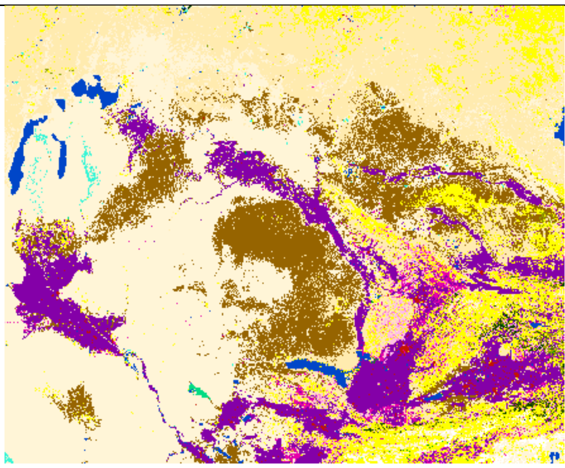
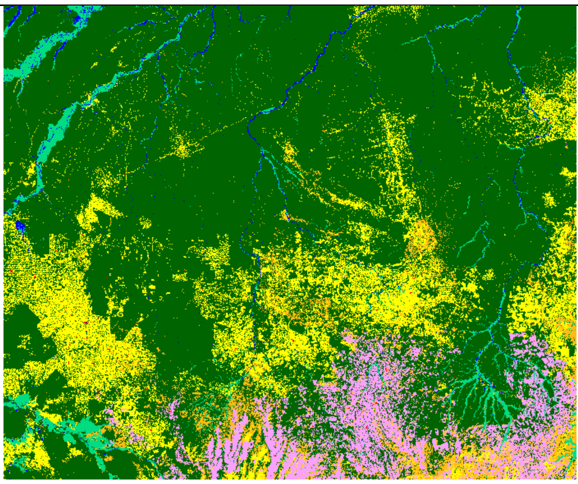
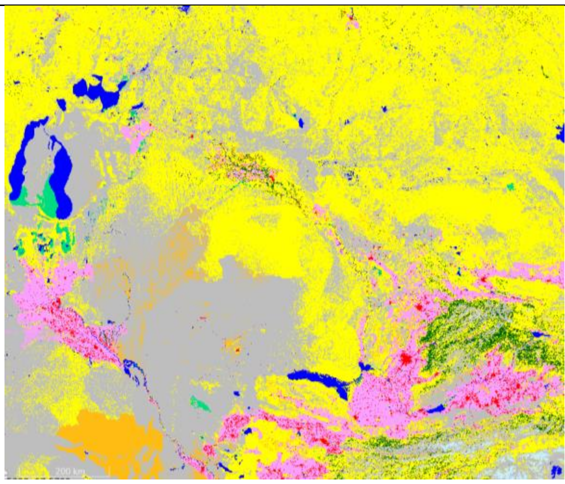
CGLS-LC100 Rondonia and Mato Grosso, Brazil	CGLS-LC100 Uzbekistan
	
LC-CCI 2015 Rondonia and Mato Grosso, Brazil	LC-CCI 2015 Uzbekistan
	
Globeland30 2010 Rondonia and Mato Grosso, Brazil	Globeland30 2010 Uzbekistan
	

Figure 16: Comparison of three maps in Rondonia and Mato Grosso, Brazil and Uzbekistan (legends are shown in Annex 3)

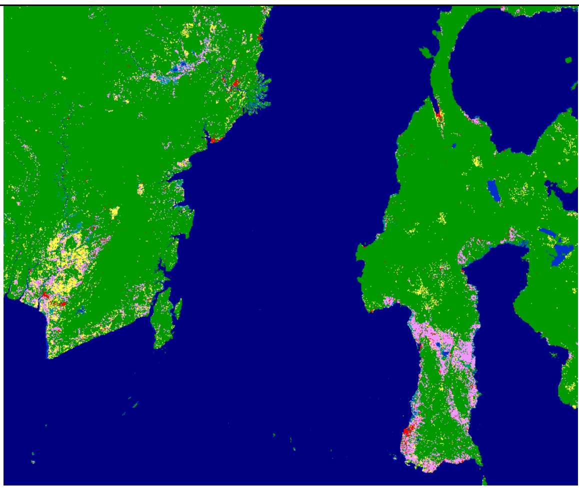
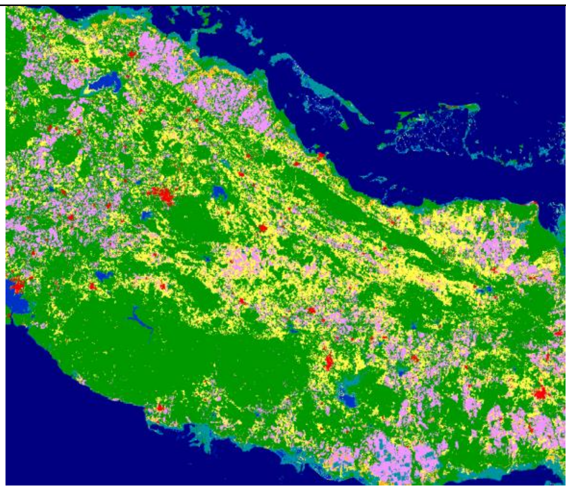
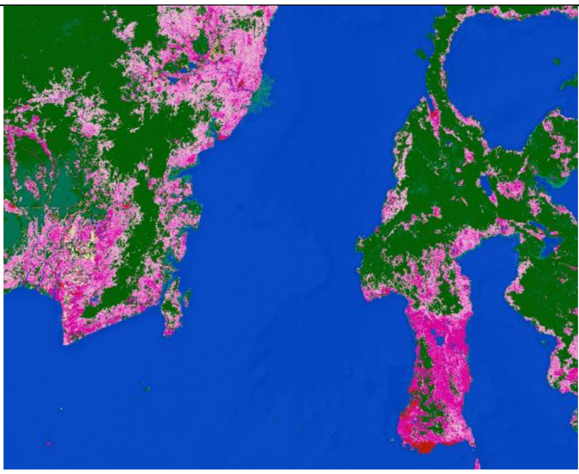
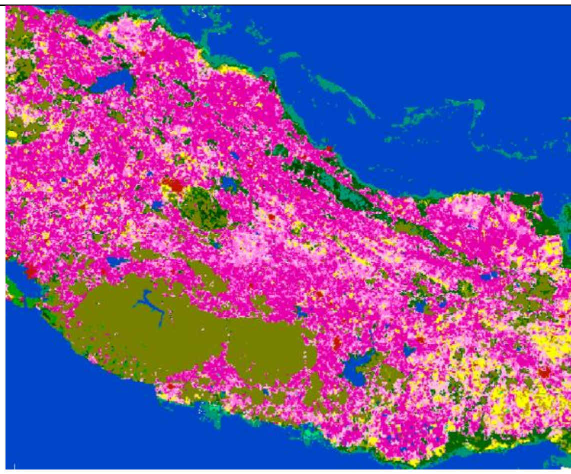
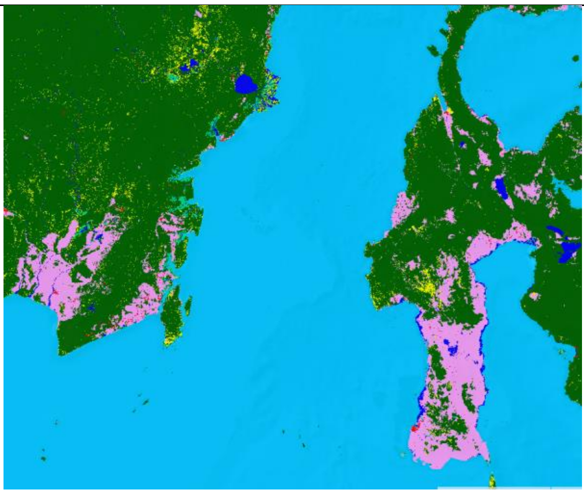
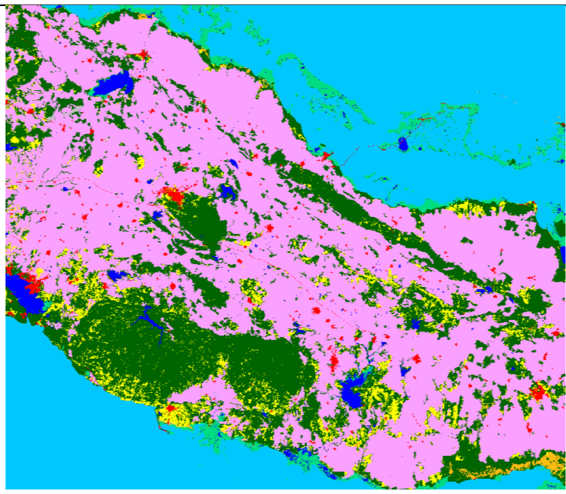
CGLS-LC100 Kalimantan and Sulawesi, Indonesia	CGLS-LC100 Cuba
	
LC-CCI 2015 Kalimantan and Sulawesi, Indonesia	LC-CCI 2015 Cuba
	
Globeland30 2010 Kalimantan and Sulawesi, Indonesia	Globeland30 2010 Cuba
	

Figure 17: Comparison of three maps in Kalimantan and Sulawesi, Indonesia, and Cuba (legends are shown in Annex 3)

Figure 17 shows the comparisons in Kalimantan and Sulawesi, Indonesia and Cuba. In Indonesia, the CGLS-LC100 map depicts less cropland areas than the other two products, while the LC-CCI 2015 tend to overestimate cropland areas in this part. Cropland underestimation in the CGLS-LC100 is particularly apparent in the north of the city of Banjamarsin and southern tip of Sulawesi Selatan, as compared to the Globeland30. Similar tendency of depicting less cropland areas is also visible in Cuba. However, visual inspection of images in this area in Google map suggests that the CGLS-LC100 better separates cropland areas from pasture areas while Globeland30 cropland definition include pasture as cropland. The CGLS-LC100 map tend to overestimate forests in Cuba as compared to the other two products.

4.3 CONCLUSION

Our assessments shows that the CGLS-LC100 discrete map is mapped with 80.1 % +/-0.7 % accuracy. Forest, bare/sparse vegetation, snow/ice and water classes are mapped with very high accuracies. The class accuracies of herbaceous vegetation, croplands, urban are moderate, while wetlands, lichen and moss and shrubs class have lower class accuracies.

Among the cover fraction layer, snow, built-up, water and lichen/moss fraction maps show lowest errors, followed by crops and bare fraction types. On the other hand, herbaceous vegetation fraction product have the highest error. Globally, the overall accuracy of the CGLS-LC100 discrete map with 10 generic classes reached the targeted accuracy 80 %. For the continents, accuracies were also around 80 % for the most of the continents, with exceptions to North America (76.7 % ±1.7 %) and Asia (83.4 % ±1.5 %).

Our visual comparison of the CGLS-LC100, the LC-CCI 2015 and the Globeland30 2010 maps shows good characterization of natural vegetation classes and croplands in the CGLS-LC100. Classes such as wetland and cropland tend to have less areas in the CGLS-LC100 as compared to the other maps and this can be attributed to the differences in the class definitions. The CGLS-LC100 better distinguishes pasture areas from cropland areas.

5 REFERENCES

Alaska Geobotany Center: Circumpolar Arctic Coastline and Treeline Map, http://arcticatlas.geobotany.org/catalog/entries/5421-circumpolar-arctic-coastline-and-treeline-map

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.

CAVM Team. 2003. Circumpolar Arctic Vegetation Map. Scale 1:7,500,000. Conservation of Arctic Flora and Fauna (CAFF) Map No. 1. U.S. Fish and Wildlife Service, Anchorage, AlaskaCopernicus 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

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E.C., Jones, B., Barber, C.V., Hayes, R., Kormos, C., Martin, V., Crist, E., Sechrest, W., Price, L., Baillie, J.E.M., Weeden, D., Suckling, K., Davis, C., Sizer, N., Moore, R., Thau, D., Birch, T., Potapov, P., Turubanova, S., Tyukavina, A., Souza, N.d., Pintea, L., Brito, J.C., Llewellyn, O.A., Miller, A.G., Patzelt, A., Ghazanfar, S.A., Timberlake, J., Klöser, H., Shennan-Farpón, Y., Kindt, R., Lillesø, JP.B., Breugel, P.v., Graudal, L., Voge, M., Al-Shammari, K.F., Saleem, M. (2017). An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm, BioScience, Volume 67, Issue 6, 1 June 2017, Pages 534–545, https://doi.org/10.1093/biosci/bix014

Dwyer, J.L., Roy, D.P., Sauer, B., Jenkerson, C.B., Zhang, H.K., Lymburner, L. (2018). Analysis Ready Data: Enabling Analysis of the Landsat Archive. Remote Sens, 10, 1363.

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

Gao, B. C. (1996). NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58: 257–266.

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

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

Khanna, S., Palacios-Orueta, A., Whiting, M., Ustin, S., Riano, D., Litago, J. (2007). Development of angle indexes for soil moisture estimation, dry matter detection and land-cover discrimination, Remote Sensing of Environment, 109(2), 154–165. [doi: 10.1016/j.rse.2006.12.018].

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

NASA JPL (2013). NASA Shuttle Radar Topography Mission Global 1 arc second [Data set]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MEaSUREs/SRTM/SRTMGL1.003

NASA/METI/AIST/Japan Spacesystems, and U.S./Japan ASTER Science Team (2009). ASTER Global Digital Elevation Model [Data set]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/ASTER/ASTGTM.002

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

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

Ragman, H., Dedieu, G. (1994). SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum, International Journal of Remote Sensing, 15:1, 123-143, DOI: 10.1080/01431169408954055

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.

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.

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

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.Wessel, P., Smith, W. H. F. (1996), A global, self-consistent, hierarchical, high-resolution shoreline database, J. Geophys. Res., 101(B4), 8741–8743, doi:10.1029/96JB00104

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: 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 300 m and/or 100 m 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 100 m and/or 300 m 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 11 provides a summary of the major requirements from the stakeholders, while Table 12 shows an overview of the requested classes to be covered by the mapping.

Table 11: 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 12: 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		

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
	22	A12A4A20B3(B9)XXE2	Deciduous shrubs			X		
30		A12A2(A6)A20B4	Herbaceous vegetation			X	X	X
		A12A6A10 // A11A1A11B4XXXXXXF2F4F7G4-F8	Pasture/managed grassland					X
		A122(A6)A10	Natural grassland			X		X
		A12A2	Grass types for Western Africa			X		
		A12A3A11B2XXXXXXF2F4F7G4-A12; A12A3A11B2-A13; A12A1A11	Savannas			X		
40		A11A3	Cultivated and managed vegetation/agriculture		X	X	X	X
	41	A11A3XXXXXXD3(D9)	Irrigated cropland		X			X
	42	A11A3XXXXXXD1	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

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
70		B28A2(A3)	Snow and Ice				X	X
80		B28A1	Open water				X	X
		A24A1(A2/A3/A4)	Wetland			X	X	X
		A24A3	Mangroves	X		X		

ANNEX 2: STATISTICAL VALIDATION RESULTS

Table 13: Confusion matrix for discrete CGLS-LC100 map at Level 1 for Asia

Asia	Forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/sparse veg.	Snow/ice	Water	Wetland	Correct	Total	User's accuracy	Confidence interval ±
Forest	739.9	31.9	26.4	22.3				1		739.9	821.5	90.1	2
Shrubs	6	38.8	9.6	1		1.4		1.4		38.8	58.3	66.5	13.5
Herbaceous veg.	18.9	41.4	284.3	15.5	0.4	6.9		2.3	3.1	284.3	372.7	76.3	4.9
Croplands	66.8	18.6	58.9	380	4.7	2.3		7.3	10.3	380	548.9	69.2	4
Urban	2.5	0.4		1.7	30.8					30.8	35.4	87.1	9.5
Bare/sparse veg.	2.5	17.6	51.7	5	2.6	745.8		1	1.2	745.8	827.5	90.1	2.5
Snow/ice						1.4	8.9			8.9	10.3	86.4	22.9
Water	1.9			1				50.4		50.4	53.3	94.5	6.2
Wetland	0.2		0.6	2.1				1.1			4		0
Correct	739.9	38.8	284.3	380	30.8	745.8	8.9	50.4					
Total	838.6	148.6	431.5	428.7	38.5	757.9	8.9	64.6	14.6				
Producer's accuracy	88.2	26.1	65.9	88.6	80	98.4	100	78					
Confidence interval ±	2.2	8.2	4.8	3.1	15	1	0	10.3					
Overall accuracy												83.4	1.5

Table 14: Confusion matrix for discrete CGLS-LC100 map at Level 1 for Eurasia

Eurasia	Forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/sparse veg.	Snow/ice	Water	Wetland	Lichen/ moss	Correct	Total	User's accuracy	Confidence interval ±
Forest	1073.7	71.2	59.2	3.9				3.8	1.1	8.3	1073.7	1221.2	87.9	2.1
Shrubs	11.1	43	10.3						4.2	2.8	43	71.4	60.2	13.1
Herbaceous veg.	39.2	64.3	652.2	16.8		20.3		0.8	19.2	76.1	652.2	889.1	73.4	3.1
Croplands	10.1	2.2	42.1	131.7							131.7	186	70.8	6.8
Urban	0.2		0.3	0.1	2.9						2.9	3.5	82.5	20.1
Bare/sparse veg.			6.8			74.1		1.1			74.1	81.9	90.4	6.3
Snow/ice							4.2				4.2	4.2	100	0
Water	2.6		1			2.7		84.9			84.9	91.2	93.1	5.4
Wetland	1.1	7.6	32.5	0.2				3	41.1	11.8	41.1	97.4	42.3	10.9
Lichen/moss			8.4			4		1.4		13.2	13.2	27	48.9	20.2
Correct	1073.7	43	652.2	131.7	2.9	74.1	4.2	84.9	41.1	13.2				
Total	1138	188.3	812.7	152.7	3	101.1	4.2	95	65.6	112.3				
Producer's accuracy	94.3	22.8	80.2	86.2	97.8	73.3	100	89.4	62.7	11.8				
Confidence interval ±	1.5	6.9	2.9	5.7	3.4	8.7	0	6.5	12.9	6.3				
Overall accuracy													79.4	1.7

Table 15: Confusion matrix for discrete CGLS-LC100 map at Level 1 for Europe

Europe	Forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/sparse veg.	Snow/ice	Water	Wetland	Lichen/ moss	Correct	Total	User's accuracy	Confidence interval ±
Forest	1111	39.5	41	43.7	9			1.1	1.3	2.3	1111	1248.9	89	1.8
Shrubs	31.9	19.1	10.8	3.7	0.1	0.1					19.1	65.7	29.1	10.9
Herbaceous veg.	39.9	27.5	238.8	27.4	3	17			1.4	23.8	238.8	378.7	63.1	5.1
Croplands	63.6	10.3	100	670.1	7.8				2.4		670.1	854.3	78.4	3
Urban	3.9	1	5.3		54.7						54.7	64.9	84.2	9.3
Bare/sparse veg.			0.2			13.3				1.1	13.3	14.5	91.5	14.2
Snow/ice						1	12.1				12.1	13	92.7	13.9
Water	1.6			1.1				56.6			56.6	59.2	95.6	5.4
Wetland	5.3		1.6					0.3	6.5		6.5	13.7	47.3	28.5
Lichen/moss						1						1		0
Correct	1111	19.1	238.8	670.1	54.7	13.3	12.1	56.6	6.5					
Total	1257.2	97.5	397.7	746	74.5	32.2	12.1	58	11.6	27.2				
Producer's accuracy	88.4	19.6	60	89.8	73.4	41.2	100	97.6	56.1					
Confidence interval ±	1.9	7.9	5	2.3	9.9	17.4	0	3.8	29.5	0				
Overall accuracy													80.4	1.6

Table 16: Confusion matrix for discrete CGLS-LC100 map at Level 1 for North America

North America	Forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/sparse veg.	Snow/ice	Water	Wetland	Lichen/ moss	Correct	Total	User's accuracy	Confidence interval ±
Forest	847.4	40	23.7	8.6	4.4			0.8	6.4	5.9	847.4	937.2	90.4	2
Shrubs	48.3	202.8	42.3	3.1	0.1				3.5	5.8	202.8	305.9	66.3	5.7
Herbaceous veg.	22.5	85.9	233.1	9.1	1.2	14.9		1.1	11.2	82.9	233.1	461.9	50.5	5
Croplands	34	6.8	40.1	211	0.1						211	292.1	72.3	5.1
Urban	1	0.6	0.9		6.9						6.9	9.5	73.1	24.2
Bare/sparse veg.	1.3	7.7	2.1			28.3				5.3	28.3	44.7	63.3	17.2
Snow/ice						14	281.4				281.4	295.4	95.3	3.8
Water						1.4		114.4	1.1		114.4	116.9	97.9	2.9
Wetland	1.3	1.1	6.4	0.9		0.1		2.3	10.8	2.6	10.8	25.4	42.4	21.5
Lichen/moss			16.3			53.7	1.1	3		127.9	127.9	202	63.3	7.7
Correct	847.4	202.8	233.1	211	6.9	28.3	281.4	114.4	10.8	127.9				
Total	955.8	344.8	364.8	232.8	12.7	112.3	282.5	121.7	33	230.5				
Producer's accuracy	88.7	58.8	63.9	90.7	54.6	25.2	99.6	94.1	32.6	55.5				
Confidence interval ±	2.2	5.5	5.2	3.7	24.9	9.6	0.8	4.8	18.6	7.2				
Overall accuracy													76.7	1.7

Table 17: Confusion matrix for discrete CGLS-LC100 map at Level 1 for Oceania and Australia

Oceania and Australia	Forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/sparse veg.	Water	Wetland	Correct	Total	User's accuracy	Confidence interval ±
Forest	691.6	74.4	22.4	2.9	0.3	0.6		1.5	691.6	793.7	87.1	2.9
Shrubs	48.9	358	89.5			2.2			358	498.6	71.8	5.7
Herbaceous veg.	50.5	119.7	1047.8	14.9		16.9		1.7	1047.8	1251.6	83.7	2.8
Croplands	2.1	3.1	32.3	90.2					90.2	127.8	70.6	8.6
Urban	0.2	0.3			3.1				3.1	3.6	88.6	15.9
Bare/sparse veg.			0.6			12.9			12.9	13.4	95.9	8.2
Water	0.8	0.6				5.5	6.3		6.3	13.1	48.1	22.7
Wetland	0.6		0.7							1.3		0
Correct	691.6	358	1047.8	90.2	3.1	12.9	6.3					
Total	794.6	556	1193.2	108	3.5	38	6.3	3.2				
Producer's accuracy	87	64.4	87.8	83.5	88.8	33.9	98.9					
Confidence interval ±	2.9	5.6	2.5	7.4	14.8	17.9	1.7					
Overall accuracy											81.8	1.9

Table 18: Confusion matrix for discrete CGLS-LC100 map at Level 1 for South America

South America	Forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/sparse veg.	Snow/ice	Water	Wetland	Correct	Total	User's accuracy	Confidence interval ±
Forest	1347.7	73	29.9	16.3				1.3	4.3	1347.7	1472.6	91.5	1.4
Shrubs	55.2	187.1	45.3	3.6	1	13.1		1.1	5.6	187.1	312.1	59.9	5.7
Herbaceous veg.	55.6	79.5	296.9	21.5	1	6.2		0.7	12.7	296.9	474.3	62.6	4.5
Croplands	28.8	10.9	41.3	162.7	0.9	1.7			0.7	162.7	247	65.9	5.8
Urban	0.8				9.2					9.2	10.1	91.5	13.4
Bare/sparse veg.		13.5	16		0.9	128.9	1.4	0.8	0.7	128.9	162.3	79.4	6.5
Snow/ice						2.5	3.5	0.2		3.5	6.1	56.8	40.2
Water	2.4	1.4	0.7					32.6		32.6	37.1	87.8	10.2
Wetland	2	2.7	3		0.1			1.5	3.2	3.2	12.5	25.6	24.9
Correct	1347.7	187.1	296.9	162.7	9.2	128.9	3.5	32.6	3.2				
Total	1492.5	368.2	433.2	204.2	13.1	152.4	4.9	38.2	27.3				
Producer's accuracy	90.3	50.8	68.5	79.7	70.5	84.6	70.6	85.3	11.8				
Confidence interval ±	1.5	5.2	4.5	5.6	22.1	6	32.8	10.8	12.7				
Overall accuracy												79.4	1.5

Table 19: Confusion matrix for discrete CGLS-LC100 map at Level 1

Africa	Forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/sparse veg.	Water	Wetland	Correct	Total	User's accuracy	Confidence interval ±
Forest	923.7	62.1	36.6	30.1				2.5	923.7	1055.1	87.5	2.2
Shrubs	109.3	349.7	90.5	34.1	0.1	10.8	0.1	5	349.7	599.7	58.3	6.4
Herbaceous veg.	41	64	345.7	29.6	0.2	22.1	3	8.7	345.7	514.4	67.2	6.7
Croplands	18.4	28	38.2	176.4	0.2	0.9	0.1	3.7	176.4	265.8	66.4	6.4
Urban	5.3		3.4	0.4	10.3	0.2			10.3	19.5	52.8	20.5
Bare/sparse veg.	0.8	9.2	45.1		0.2	1038.9		0.1	1038.9	1094.4	94.9	2.9
Water							36.7	0.1	36.7	36.9	99.6	0.6
Wetland	7.3	0.1	3.3	0.6		0.3	4.2	14.5	14.5	30.3	48	18.4
Correct	923.7	349.7	345.7	176.4	10.3	1038.9	36.7	14.5				
Total	1105.8	513.2	562.8	271.3	11	1073.1	44.3	34.6				
Producer's accuracy	83.5	68.2	61.4	65	94.1	96.8	83	42				
Confidence interval ±	2.6	6.6	6.5	7.2	3.9	2.1	10.3	18				
Overall accuracy											80.1	2

The accuracy of the the CGLS-LC100 discrete map at level 2 differentiating closed and open forest types in addition to other land cover classes are shown in Table 20.

Table 20: Confusion matrix for discrete CGLS-LC100 map at Level 2 on global scale

	Closed forest	Open forest	Shrubs	Herbaceous veg.	Croplands	Urban	Bare/sparse veg.	Snow/ice	Water	Wetland	Lichen/ moss	Correct	Total	User's accuracy	Confidence interval ±
Closed forest	**4413.3**	775.3	166.9	90.2	22.6	0.6			2.9	10.9	9.8	**4413.3**	5492.5	**80.4**	1.2
Open forest	235.2	**1161.3**	209.7	145.3	94.8	8.1	0.2		5.9	7	9.5	**1161.3**	1877.1	**61.9**	2.5
Shrubs	41.1	262.9	**1144**	279.1	52	1.3	29.5		3.4	21.1	10.6	**1144.4**	1845.4	**62**	3.2
Herbaceous veg.	27.4	216.5	465.8	**2742.8**	127.6	4.4	99.3		10.3	65.9	205	**2742.8**	3965	**69.2**	1.8
Croplands	38.2	189.3	88.9	322.9	**1624.9**	10.8	6.2		11.1	21.4		**1624.9**	2313.6	**70.2**	2.1
Urban		13.7	1.9	7.4	3.1	**99.1**	0.2					**99.1**	125.5	**79**	7
Bare/sparse veg.		6.4	61	157.5	7.4	5	**2647.8**	1.4	3.6	2.7	9	**2647.8**	2901.9	**91.2**	1.8
Snow/ice							22.6	**382**	0.2			**382**	404.8	**94.4**	3.7
Water	0.5	8.7	1.7	1.9	1.9		7.3		**418.7**	1.6		**418.7**	442.2	**94.7**	2.2
Wetland	0.1	16.1	13	55	5.4	0.1	0.4		14.8	**85.1**	17.2	**85.1**	207.2	**41.1**	7.9
Lichen/moss				30.5			73.1	1.4	5.4		**177.4**	**177.4**	287.8	**61.6**	7.2
Correct	**4413.3**	**1161.3**	**1144**	**2742.8**	**1624.9**	**99.1**	**2647.8**	**382**	**418.7**	**85.1**	**177.4**				
Total	4755.8	2650.2	2153	3832.7	1939.8	129.2	2886.6	384.9	476.3	215.6	438.6				
Producer's accuracy	**92.8**	**43.8**	**53.1**	**71.6**	**83.8**	**76.7**	**91.7**	**99.3**	**87.9**	**39.4**	**40.5**				
Confidence interval ±	0.8	2.1	2.9	1.9	2	8.3	1.5	0.8	3.4	7.9	5.5				
Overall accuracy														**75**	0.8

Overall accuracy at Level 2 is shown per continents in Table 21.

Table 21: Overall accuracy for discrete CGLS-LC100 map at Level 2 on continental level

	N sample	Overall accuracy	Confidence interval ±
Asia	2732	79.6	1.6
Eurasia	2673	70.6	1.9
Europe	2714	73.5	1.8
North America	2691	71.9	1.8
Oceania & Australia	2703	77.7	2
South America	2734	74	1.6
Africa	3616	76.5	2

ANNEX 3: 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 18: Legend of the global CCI-LC maps, based on LCCS.

Figure 19: Legend of the Globeland30 2010 map