Copernicus Global Land Operations "Vegetation and Energy"

"CGLOPS-1"

Framework Service Contract N° 199494 (JRC)

PRODUCT USER MANUAL

MODERATE DYNAMIC LAND COVER

100M

VERSION (COLLECTION) 3

Issue I3.3

Organization name of lead contractor for this deliverable: VITO

Book Captain: Dr. Marcel Buchhorn (VITO)

Contributing Authors: Bruno Smets (VITO)
Luc Bertels (VITO)
Bert De Roo (VITO)
Myroslava Lesiv (IIASA)
Nandin-Erdene Tsendbazar (WUR)
Linlin Li (WUR)
Agnieszka Tarko (WUR)

Please cite this publication the following:

Buchhorn, M.; Smets, B.; Bertels, L.; De Roo, B.; Lesiv, M.; Tsendbazar, N.E., Linlin, L., Tarko, A. (2020): Copernicus Global Land Service: Land Cover 100m: Version 3 Globe 2015-2019: Product User Manual; Zenodo, Geneve, Switzerland, September 2020; doi: 10.5281/zenodo.3938963.

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 10/07/2020
Approval:	Roselyne Lacaze	Sign	Date 02.12.2019
Endorsement:	Michael Cherlet	Sign	Date
Distribution:	Public		

Change Record

Issue/Rev	Date	Page(s)	Description of Change	Release
	24.06.2017	All	Initial Version	D1.00
	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 for public release with products	I2.10
	10.11.2019	All	Revision after release ATBD and addition of dataset DOI	I2.20
Draft	08.07.2020	All	Revision for product version 3	D3.02
	03.09.2020	All	Revision after internal review	I3.3

TABLE OF CONTENTS

List of Figures

List of Tables

List of Acronyms

Acronym	Meaning
AD	Applicable Documents
ARD	Analysis Ready Data
ATBD	Algorithm Technical Basis document
BFAST	Break for Additive Season and Trend
CCI	Climate Change Initiative
CF V1.6	Climate & Forecast conventions compliant with version 1.6
CGLS	Copernicus Global Land service
COG	Cloud-Optimized GeoTIFF
CRS	Coordinate Reference System
DDI	Data Density Indicator
DOI	Digital Object Identifier
EO	Earth Observation
EPSG	European Petroleum Survey Group
EU	European Union
FAO	Food and Agriculture Organization of the united nation
GeoTIFF	Geospatial Tagged Image File Format
GSD	Ground Sampling Distance
IIASA	International Institute for Applied Systems Analysis
JRC	Joint Research Center
LC	Land Cover
LC100	Land Cover map at 100 m resolution
LCCS	Land Cover Classification System
MC5	5-daily Median Composite
NDVI	Normalized Difference Vegetation Index
OECD	Organisation for Economic Cooperation and Development
PPSIM	Potential Permanent Snow and Ice Mask
PROBA-V	Vegetation instrument on board of PROBA satellite
PSD	Product Specifications Document
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
RGB	Red Green Blue
SDG	Sustainable Development Goal
SEEA	System of Environmental and Economic Accounting
SM	Status Mask
SSD	Service Specifications Document

SVP	Service Validation Plan
UN	United Nations
UNCCD	United Nations Convention to Combat Desertification
UNFCCC	United Nations Framework Convention on Climate Change
URD	Users Requirements Document
UTM	Universal Transverse Mercator
VI's	Vegetation Indices
VITO	Vlaamse Instelling voor Technologisch Onderzoek (Flemish Institute for Technological Research), Belgium
VR	Validation Report
WGS84	World Geodetic System 1984
WUR	Wageningen University and Research

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.

From 1st January 2013, the Copernicus Global Land Service is providing a series of bio-geophysical products describing the status and evolution of land surface at global scale, i.e Leaf Area Index (LAI), the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), the Land Surface Temperature, the soil moisture, and other vegetation indices on an hourly to 10 days interval from Earth Observation satellite data.

The CGLS delivers an annual dynamic global Land Cover product at 100 m spatial resolution (CGLS-LC100). 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 classification levels with class definitions according to the Land Cover Classification System (LCCS) scheme. Next to these discrete classes, the product also includes continuous field layers or "fraction maps" 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, rangeland management, crop monitoring, biodiversity and conservation, monitoring environment and security, climate modelling, etc.)

The first Land Cover map, produced with algorithm V1.0, was provided in July 2017, derived from the Vegetation instrument on board of PROBA satellite (PROBA-V) 100 m time-series, a database of high quality land cover training sites and several ancillary datasets. This collection 1 was solely covering the African continent and the mapping year 2015. For the global up-scaling of the product for the reference year 2015, the algorithm V2.0 was improved not only to map the entire globe, but also to improve the map quality, reaching now 80 % of accuracy at class level 1 on each continent, and its usability. Therefore, the production of the map has switched to the Universal Transverse Mercator (UTM) coordinate system aligned to the Sentinel-2 tiling grid in order to improve mapping quality in the high latitudes. Moreover, the Sentinel-2 tiling grid facilitates continuation in the production as soon the primary data input has to switch from PROBA-V to Sentinel-2. Collection 2 was delivered in May 2019. Algorithm version 3 adds the change detection methodology to the CGLS workflow. We put particular emphasis on the temporal consistency of the annual land cover maps. Specific algorithms were applied to increase the stability of the annual classifications with the effect that differences between the annual maps resulting from inconsistent classifications are reduced while areas of probable land cover change are considered as different classes in the various annual maps. Therefore, collection 3 of this product provides the annual global land cover maps and cover fraction layers for the reference years 2015 to 2019.

This Product User Manual describes the global Land Cover V3.0 product.

1 BACKGROUND OF THE DOCUMENT

1.1 SCOPE AND OBJECTIVES

This Product User Manual (PUM) is the primary document that users should read before handling CGLS-LC100 products of the third version (collection 3).

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 for the dynamic moderate land cover product

CGLOPS1_PSD_LC100m	Product Specification Document for the Version 2 of dynamic moderate land cover product.
CGLOPS1_TrainingDataReport_LC100m	Report presenting the training dataset used for the Dynamic Moderate Land cover product
CGLOPS1_ATBD_LC100_V3.0	Algorithm Theoretical Basis Document of the dynamic moderate land cover products in 100m in Version 3 (collection 3).
CGLOPS1_VR_LC100m_V3.0	Report describing the results of the scientific quality assessment of the dynamic moderate land cover products in 100m in Version 3 (collection 3).

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 ANNUAL PRODUCT GENERATION

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

1. Data pre-processing including atmospheric & geometric correction
2. Data cleaning by sensor specific status masks and (temporal) outlier detection techniques,
3. Applying data fusion techniques at multiple levels,
4. Supervised classification,
5. Including established third party datasets via expert rules, and
6. Temporal cleaning via break detection analysis of the landcover trends.

The workflow, shown in Figure 1, can be divided into two main parts: the generation of epoch-wise classification and cover fraction layers pre-products for the single reference years, and the temporal post-processing to generate consistent annual land cover maps and cover fraction layers. The temporal coverage of the annual products is one nominal year each (reference years) but includes three years of data as input for each map (epochs spanning +/- 1 year around the reference year).

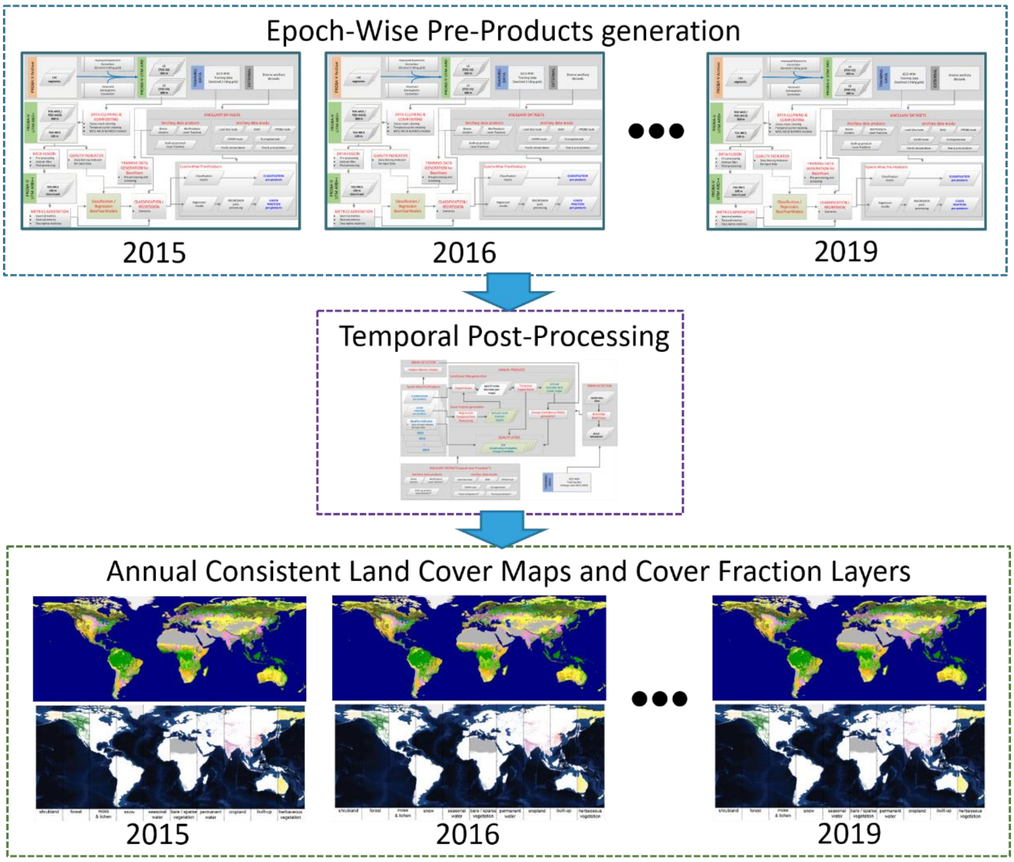

Figure 1: Workflow diagram for the annual CGLS Dynamic Land Cover layers at 100 m resolution (CGLS-LC100 products) generation.

Figure 2 shows in detail the epoch-wise generation of the pre-products (classification results and cover fraction layers) for each reference year (upper part of Figure 1) which can be divided into the following sections:

1. PROBA-V UTM Analysis Ready Data (ARD) generation for full PROBA-V archive,
2. data cleaning & compositing for full ARD archive,
3. yearly quality indicator generation for PROBA-V UTM ARD+ input data,
4. data fusion of PROBA-V UTM 100 m and 300 m ARD+ archive,
5. metrics generation/extraction per three year epoch around reference year,
6. yearly ancillary datasets products generation,
7. training data generation for BaseYear (if needed),
8. classification / regression using BaseYear models,
9. epoch-wise classification and cover fraction layers pre-products generation.

To reduce distortion in the High North, to make the land cover products better usable with other data sets and to allow continuity of the service, the PROBA-V archive used as current main input data source and normally distributed in the geographic coordinate system World Geodetic System 1984 (WGS84) – known as EPSG:4326 - was reprocessed with a new geometric correction and an improved atmospheric correction. The complete PROBA-V L1C archive (October 2013 – March 2020) (unprojected, Top-of-Atmosphere data) was in this way translated into the PROBA-V UTM Analysis Ready Data (ARD) archive which is projected to the Universal Transverse Mercator (UTM) coordinate system, fully aligned with the Sentinel-2 tiling grid in tiling naming as well as tile dimensions, and corrected to surface reflectance (Top-of-Canopy).

The PROBA-V UTM ARD main product, the PROBA-V multi-spectral daily-synthesis surface reflectance Top Of Canopy (TOC) data (L3 or also known as TOC-S1) with a Ground Sampling Distance (GSD) of 100 m is used as primary Earth Observation (EO) data, and PROBA-V UTM multi-spectral daily-synthesis surface reflectance data with a GSD of 300 m as secondarily. Next to a Status Map (SM) cleaning using the internal quality flags of the PROBA-V UTM L3 data, a temporal cloud and outlier filter built on a Fourier transformation is applied to clean the data. Moreover, a median compositing is carried out in order to archive regular 5-daily time steps in the 100 m and 300 m PROBA-V EO time series. This is needed since the PROBA-V satellite provides a daily global coverage for the 300 m L3 product corresponding to 5-daily for the 100 m L3 product. This cleaning and compositing step upgrades the PROBA-V UTM ARD archive into the PROBA-V UTM ARD+ archive. From the cleaned and outlier screened 5-daily time series a Data Density Indicator (DDI) is calculated for each reference year which is used as an quality indicator for the used input data as well as weight factor in the supervised learning process of the model generation.

Next, to improve the data density in the 5-daily 100 m time series, the 100 m and 300 m EO datasets are fused using a Kalman filtering approach. The Kalman-filled 100 m time series are then automatically checked for consistency. This upgrades the PROBA-V UTM ARD+ archive into the PROBA-V UTM ARD++ archive, which gives temporal cleaned, consistent and dense 5-daily image stacks for all global land masses at 100 m resolution (Note: 300 m data is not existing in the ARD++).

Following, several Vegetation Indices (VI's) are generated out of the PROBA-V UTM ARD++ surface reflectance data for each time step in the archive. Since this data would be too much to be used in a supervised learning, the time dimension in the data stack has to be condensed. Again, for the generation of the land cover maps and cover fraction layers for each reference year, the data of three years (+/- 1 year around the reference year) is used. The three year period around the reference year is called "epoch" in this documentation. Therefore, several metrics explaining the time series of an epoch are extracted as input for the supervised classification and regression for each reference year. Thus for each epoch a harmonic model is fitted through the time steps of each of the derived vegetation indices. Next to the parameters of the harmonic model which are used as metrics for the overall level and seasonality of the time series of the corresponding epoch, descriptive statistics and textural metrics are generated. Overall, 183 metrics are extracted from the PROBA-V UTM ARD++ archive for each epoch (Note: to distinguish the epochs in the documentation they are labelled by their reference year e.g. epoch2015, epoch2016, …).

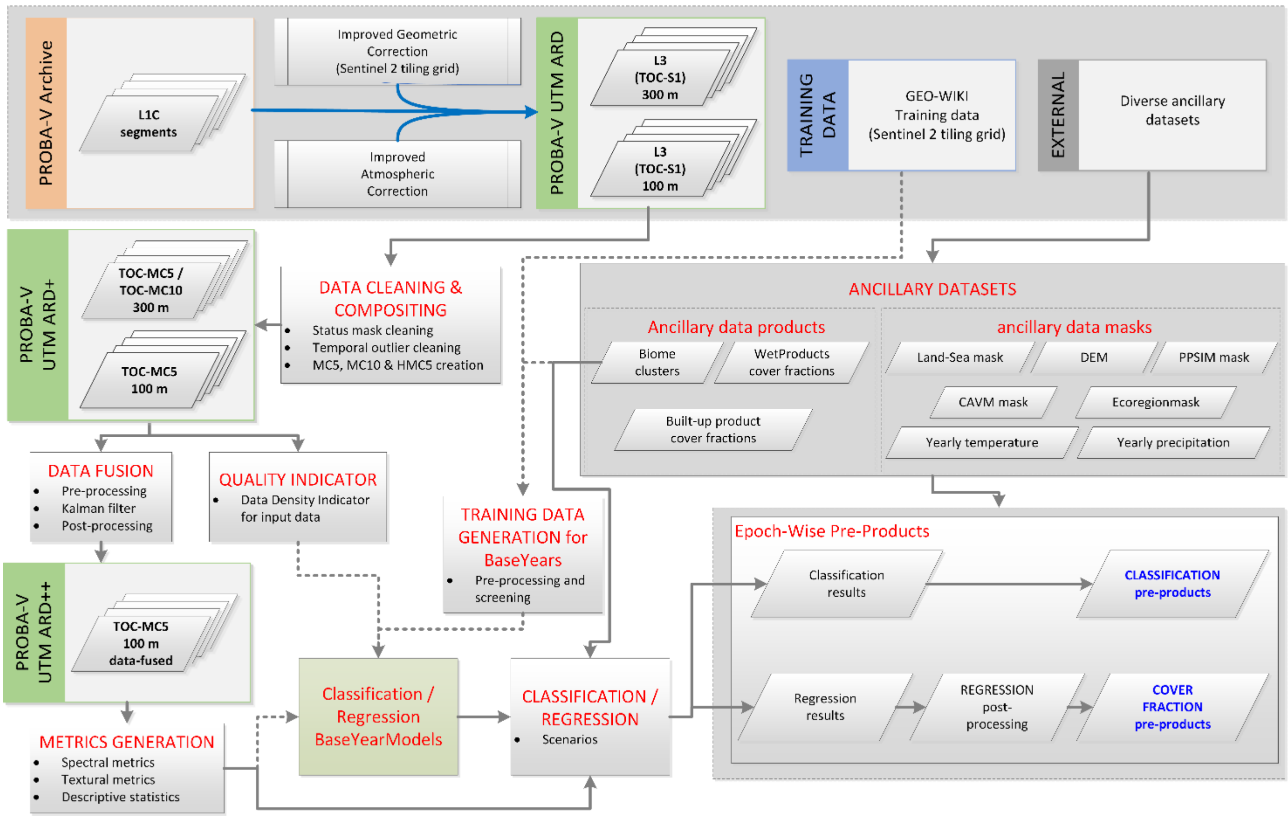

Figure 2: Workflow diagram for Part A of the CGLS-LC100 products – generation of the epoch-wise classification and cover fraction layers pre-products.

To use external ancillary data in the classification/regression, the datasets have been checked for consistency and if needed re-warped to the UTM coordinate system, resampled to 100 m, retiled into the Sentinel-2 tiling grid, and post-processed to usable ancillary data products. Some of the used external ancillary data is valid for all reference years were some have to be generated for each reference year. The training data for the supervised learning 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 can be derived for 100 m resolution. The training data was initially collected for the reference year 2015 and will be updated every 5 years. Therefore the 5-yearly intervals are called BaseYears (e.g. BaseYear2015 represents the training data / model for the reference year 2015 with a duration of 1 year). To optimize the CGLS-LC100 products to regional patterns, the classification/regression are handled eco-region specific (called biome clusters). In the classification and regression preparation, the metrics of each biome cluster of the training points of a BaseYear are analysed for inter-specific outliers in the pure endmembers, as well as screened for the best metrics combinations (best band selection) in order to reduce redundant information. The optimized training data together with the quality indicator of the input data (DDI dataset) for the corresponding epoch of the BaseYear are input in the training of the supervised classifier/regressor using Random Forest (RF) techniques. The generated biome cluster specific models are optimized via a 5-folded cross-validation in order to estimate the optimal classifier/regressor parameters. These models are therefore only produced in 5-year intervals for the BaseYears and used for the classification/regression of all reference years starting with the BaseYear up but not including the next BaseYear (e.g. the models of BaseYear2015 is used for the processing of the reference years 2015 – 2019). Finally, after applying the biome cluster specific classification/regression models and running a post-processing, epoch-wise pre-products are generated which include the epoch-wise classification results and epoch-wise cover fraction layers.

Since all epoch-wise pre-products are generated independently for each reference year, a second processing flow (lower part of Figure 1) is needed to generate consistent annual land cover maps and cover fraction layers over time (Figure 3). Therefore, the pre-products (classification and cover fraction layers) of every epoch are post-processed in several steps:

1. Generation of break-maps with BFAST algorithm;
2. Generation of secondary break-map with a Hidden Markov model;
3. Temporal post-processing of epoch-wise cover fraction layer pre-products;
4. Applying expert rules to generate annual land cover maps and applying temporal expert rules to make them consistent.

In the first pre-processing step, break maps are created based on a Breaks For Additive Season and Trend (BFAST) algorithm. These break maps show areas where change occurred between years. The BFAST family of break detection algorithms are unsupervised change detection methods that use a univariate VI time series with irregular timesteps as input. From this input, several components are computed: a monotonously increasing trend component, a nominal season component that indicates the season that each observation belongs to, and harmonic components that are sine and cosine functions for each harmonic order. The second pre-processing step consist of creating a break map based on a hidden Markov model (HMM). This model takes the cover fractions of all years and predict from these cover fractions the discrete map. The differences in the discrete maps between the years are then used as break map. BFAST was used as our main predictor for breaks and HMM was used as ancillary data set to improve the confidence for all detected breaks.

Next, in the temporal regression post-processing (TPR) all epoch-wise cover fraction pre-products are stacked together to obtain a cover fraction time series. These time series (one for each LC class) are then evaluated in terms of stability based on the deviation of the cover fractions from their respective means. Depending on the stability certain rules are applied in order to smooth the cover fraction depending on whether or not BFAST breaks has been detected. Applying these rules implies consistent cover fraction over all year unless BFAST detected some breaks.

In a second processing step, we build upon the success of previous global mapping efforts and/or other ancillary datasets which are included via expert rules in the annual land cover maps generation. Based on the classification pre-products of the single epochs and on the annual temporal cleaned cover fraction, expert rules are used to obtain annual land cover pre-maps. To improve our temporal stability, the change in the land cover maps is aligned with the prediction of BFAST and imprinted in the final annual land cover maps. Also a 3-level confidence mask is generated by implementing the HMM break maps.

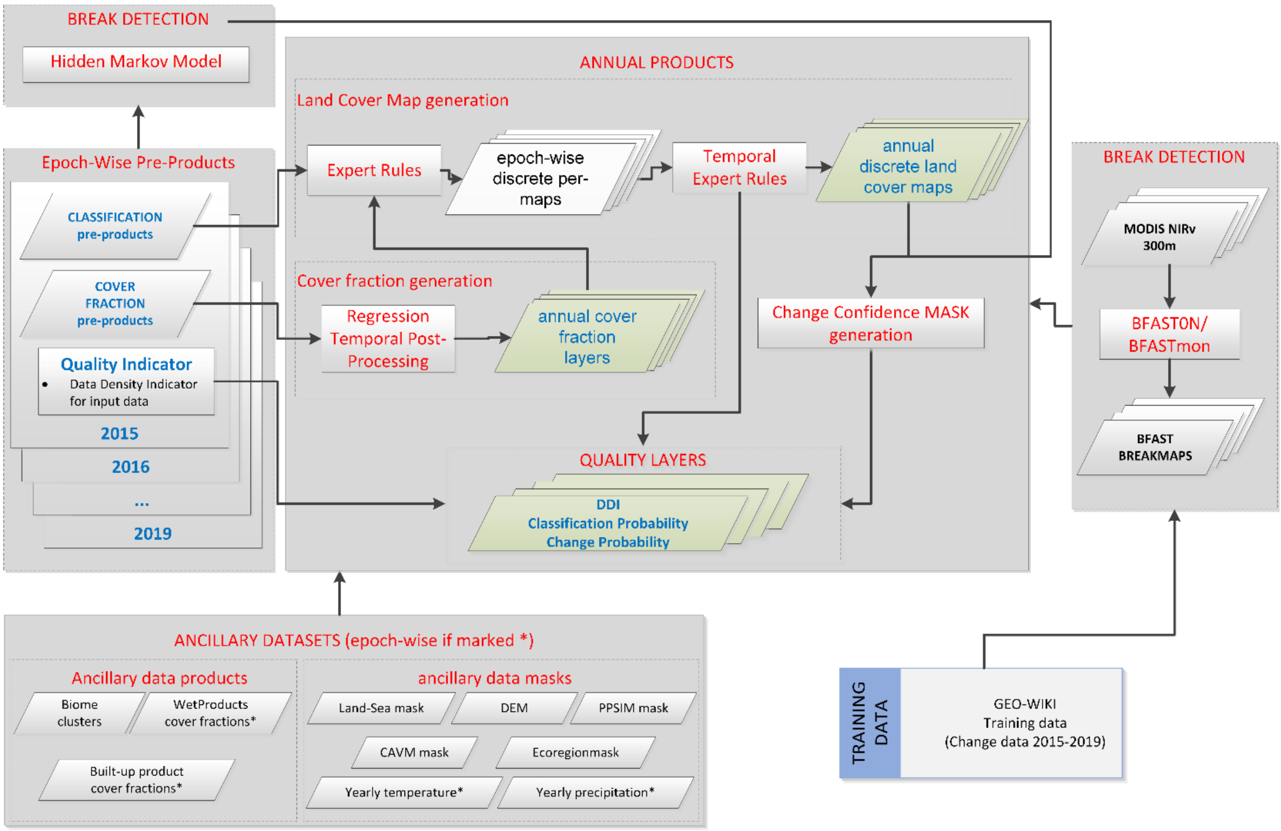

Figure 3: Workflow diagram for Part B of the CGLS-LC100 products – temporal post-processing to generate consistent annual land cover maps and cover fraction layers with the corresponding Quality layers.

The produced annual land cover maps use 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)) of all main land cover classes for each year.

Details of the retrieval method can be found in the ATBD [CGLOPS1_ATBD_LC100_V3.0].

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 for each annual discrete map was produced out of the predicted class probabilities of the classification results. This classification quality layer was bundled together with the annual Data Density Indicator layer as quality indicator for the input data and the annual change confidence layers as overall product quality layers.

Figure 4 shows an overview of the discrete map with 23 classes on global scale, where Figure 5 shows the legend in more detail. Figure 6 shows the 10 provided CGLS_LC100 cover fraction layers in a collage. The shown colours for the cover fractions are the ones integrated as RGB colour bars in the metadata of the cover fraction product layers.

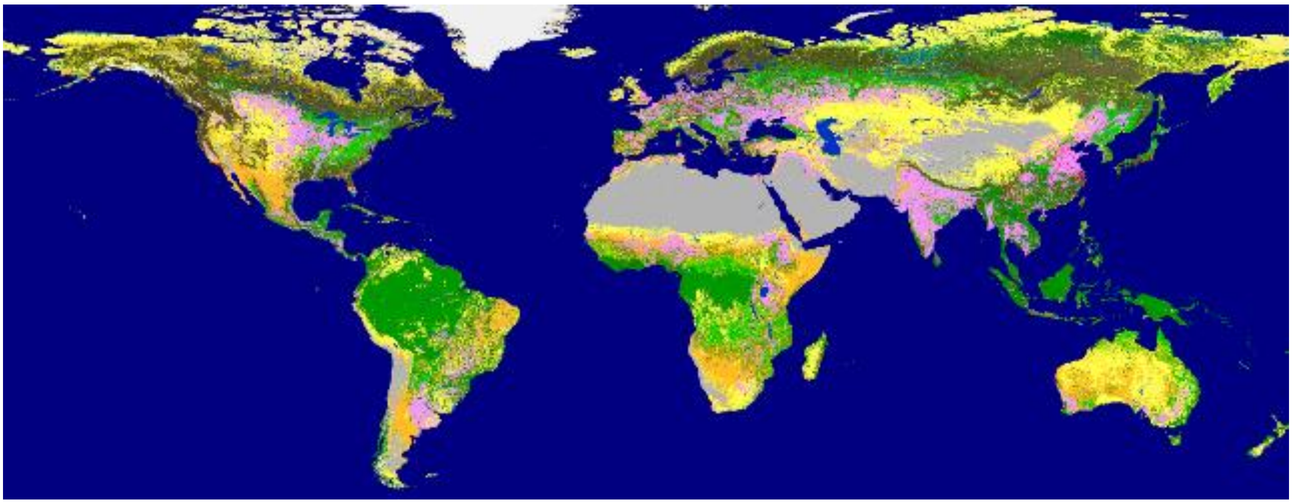

Figure 4: The CGLS Dynamic Land Cover Map at 100 m for epoch 2015 with 23 discrete classes (detailed legend in Figure 5)

Figure 5: Legend for the 23 discrete classes of the CGLS Dynamic Land Cover Map at 100 m shown in legend level 3.

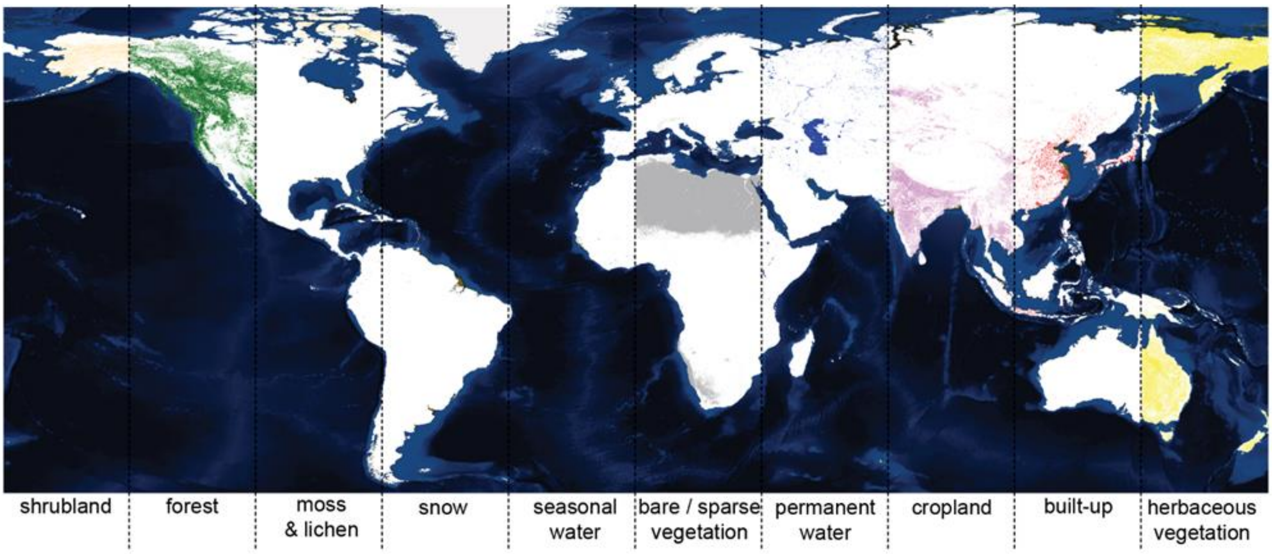

Figure 6: The cover fraction layers for the 9 base land cover classes and the seasonal inland water cover fraction of the CGLS_LC100 product for epoch 2015 (shown as a collage on global scale).

2.2 PRODUCT GENERATION MODES

The annual CGLS-LC100 products are generated using an 'epoch', where an epoch consists of 3 years of input data (one year before and one year after the reference year). The classification/regressor models are generated on the **Base** reference year. The machine learning models are re-used in the subsequent generation of the change maps. As shown in Figure 7, the actual land cover can change over time and dependent on when the change happens it can be detected. The yearly land cover map updates are delivered at the end of the 1st quarter of the year-end, and hence changes during the latter part of the year can be detected but the new class can not yet be identified (depicted in red crosses). This mode is called **Near-Real-Time**.

Therefore the yearly map is re-generated the year after, known as **Consolidated**, and enough data is available to identify all new classes including the land covers which changed at late stage.

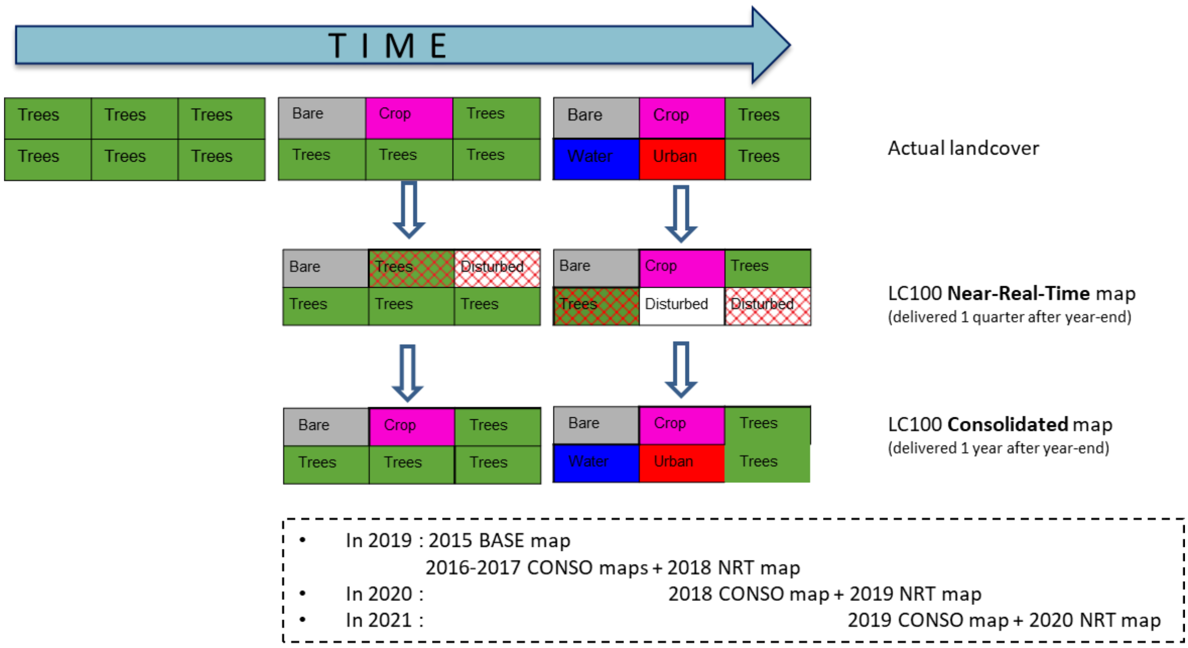

Figure 7 : Processing modes of land cover change maps

The result of this 'mode' processing implies that during the beginning of a new year (Yr), two maps are provided:

1. Near-Real-Time (NRT) map of previous year (Yr-1)
2. Consolidated (CONSO) map of the year before (Yr-2)

The latter map consolidates the detected changes of the NRT map and hence substitutes the former map.

2.3 LIMITATIONS

Although minimization of omission and commission errors is achieved by the usage of ancillary dataset, they are sometimes inevitable. An overview of reasons for omission and commission errors include:

- Due to the usage of optical RS data, classifications in areas with high yearly cloud cover can have lower classification accuracies.
- Artefacts at boundaries of biome clusters can appear due to the used biome cluster vector layer as well as the biome cluster-specific generated hyper-parameter for the Random Forest classifier and regressor.
- Remaining shadowed pixels in the time series not filtered out during the data cleaning process can lead to misclassifications.
- Fires and more specifically burned areas were not yet taken into account and therefore could lead to misclassifications.
- 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 (e.g. Nigeria, Ghana). This could lead to an overestimation of 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 an underestimation of croplands.
- Very small villages or bigger houses are difficult to map, especially when not detected by the WSF layer at 10 m resolution or the OpenStreetMap layers, which could lead to an underestimate of built-up areas.
- Through the usage of the OpenStreetMap layer all limitation of the layer apply, e.g. wrongly mapped buildings in location or size. In certain areas urban commission errors were detected due to scaling issues in the OSM layer. This can lead also to errors in the change detection between years.
- Some limitations are due to the legend or class definition, e.g.:
    - In the southern part of Africa, there are huge areas with kind of tundra type of vegetation, NDVI values are very low in these areas and can confuse the classifier to misclassify between grassland or bare land.
    - In Africa, there are a lot of riparian forests, which are evergreen. A lot of pixels where noticed with mixed deciduous trees and riparian evergreen forest which can confuse the classifier to misclassify the forest type.
- The permanent snow and ice class was limited by the PPSIM layer in order to reduce permanent snow&ice misclassifications especially over temporary frozen lakes (mostly in the High Latitudes where high cloud cover is frequent and therefore images of the thawed lakes are rare). This can led to underestimation of newly formed ice&snow areas outside the PPSIM.
- The CGLS-LC100 permanent snow and ice cover fraction layer shows only cover fractions of 0 % or 100 % at the moment. This is due to the fact that not sufficient amounts of training

points with a reported permanent snow&ice cover fraction were available. Therefore, we focused on areas with 100 % covered permanent snow&ice. This can lead to an underestimation of permanent snow and ice in partly covered 100 m pixels.

- The usage of PROBA-V multi-spectral data limits the possibilities to successfully distinguish between herbaceous vegetation and shrubland, as well as between open forest and shrubland, especially when the class borders within a pixel are not clear (meaning the cover fraction of the neighbouring discrete classes are close and therefore the voting in the classifier is difficult). We implemented an extra rule part in the Expert Rules section to deal with these classification confusions, but still misclassifications between these classes are possible.
- Since the PROBA-V time series is limited at the moment (2014 to 2020), misclassifications can happen especially in temporary ultra-dry and temporary ultra-wet areas. This is due to the fact that some regional effects with a low frequency (e.g. El Nino and La Nina) could not correctly captured with the training data reference base of 2015.
- Internal change process from one class to another and back (mainly shrub - grass) can appear due to the usage of pre-trained classification / regression models. Since the models are trained on 2015 reference data and if this was an unexpected wetter or dryer year than usual in a region, this can cause confusion in shrub – grass detection in certain regions (e.g. Central Australia). Therefore this leads to false change detection.
- Due to the pixel-based classification and regression, spatial artefacts can occur. We encourage users to apply cluster/sieving techniques in non-fragmented areas to further improve the regional classification results.
- Due to the low training data density in the High Latitudes, the classification results above 70 degree North can be less accurate.
- Small-scale changes may be missed due to MODIS data being used for change detection, as its MMU is 300 m (MODIS 250 m regridded to Proba-V 300 m grid).
- The change detection algorithms were calibrated using training data that was created using very high resolution image interpretation. This may not always scale well to what is visible at the 300 m level, as changes indicated from the former are not visible (and cannot be detected) in the time series of the latter. This impacts the statistics from BFAST calibration.
- A change detection approach based on a vegetation index time series cannot detect transitional stages in degradation or regrowth, e.g. the succession of grasses to shrubs to trees would always get detected as two breaks: before the start of the regrowth and after the start of the regrowth. These breaks may be offset from the corresponding land cover change times, depending on the speed of regrowth.
- As a global parametrisation for break detection is used, its performance may vary over different continents or ecozones.
- Detecting changes close to the end of the time series may be less accurate in cases where a rapid change is ongoing. For example, the end of deforestation may not be detected as a break because the model can assume that negative vegetation index values are possible, and it takes a longer time series to make sure that the new situation is stable.

- Changes in the time period January – February of a year could be recorded in the change of the year before due to BFAST break detection (time shifting problem).

2.4 DIFFERENCES TO THE PREVIOUS VERSION

Table 1: Successive versions (collections) of the CGLS-LC100 products. The V3.0 described in this PUM is highlighted in bold.

Version	Coverage	Status	Main characteristics
V1.0	Africa 2015	Demonstration	PROBA-V S1 time-series (plate carree) Random Forest for classification
V2.0	Global 2015	Operational	PROBA-V L1C time-series (gridded to Sentinel-2 UTM) Random Forest for classification
V2.1	Africa 2015-base, 2016-conso, 2017-conso, 2018-NRT	Demonstration	PROBA-V L1C time-series (gridded to Sentinel-2 UTM) Random Forest for classification BFAST for break detection linear regression
V3.0	Global 2015-base, 2016-conso, 2017-conso, 2018-conso, 2019-nrt	Operational	PROBA-V L1C time-series (gridded to Sentinel-2 UTM) Random Forest for classification Optimized BFAST for break detection

2.5 ROADMAP

The Copernicus Global Land service will continue the 100m production, from 2020, through the combination of Sentinel-1 and Sentinel-2 mission. The adaptation of the retrieval methodology to the Sentinels missions data has been successfully tested showing a good consistency with the Collection-3 maps. The system is ready to produce the 2019-conso and 2020-nrt map once all Sentinel input data is available.

3 PRODUCT DESCRIPTION

The CGLS-LC100 products in version 3 (collection 3) are provided as **GeoTIFF** files per 20 x 20 degree tile (Figure 8) and mapping period and layer. Meaning: each mapping period contains 14-15 layers:

- One discrete land cover map;
- Fractional cover maps for 10 classes (bare/sparse vegetation, cropland, herbaceous grassland, moss&lichen, shrubland, tree, permanent snow&ice, built-up, permanent water, seasonal water);
- One forest type map
- One discrete classification probability map (quality indicator of the classifier)
- Data Density Indicator for PROBA-V UTM 100m input data
- Change Confidence Layer showing the certainty that a change occurred compared to the previous year (only delivered for maps produced in conso or nrt mode)

3.1 FILE NAMING

The GeoTIFF files follow this naming convention:

<TILE>_<SENSOR>_LC100_<AREA>_<VERSION>_<YEAR>-<MODE>_<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 8)
  Example: W180N80 for the tile covering the area from 180W to 160W and 80N to 60N.
- <SENSOR> the EO sensor used. Here is "PROBAV".
- "LC100" indicates that this is a 100 m resolution Land Cover product
- <AREA> indicates which spatial extent the whole product has. Here is "global".
- <VERSION> shows the product version. The version denoted as M.m.r (e.g. v3.0.1), with 'M' representing the major version (collection), 'm' the minor version (starting from 0) and 'r' the production run number (starting from 1) (Table 2).
- < YEAR>-<MODE> indicates the reference year in four digits of the product layer and its processing mode (base, conso, or nrt).
- <LAYER> gives the name of the data layer (see Table 3)
- <CRS> is the coordinate reference system used. The current tiles are provided in EPSG:4326, geographic latitude/longitude CRS.

Figure 8: Scheme of the 20 x 20 degree tiles

Table 2: 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 3: Land Cover layer names in the filename

Layer in filename	Description
Discrete-Classification-map	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
MossLichen-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
BuiltUp-CoverFraction-layer	Fractional cover (%) for the built-up class
PermanentWater-CoverFraction-layer	Fractional cover (%) for the permanent inland water bodies class
SeasonalWater-CoverFraction-layer	Fractional cover (%) for the seasonal inland water bodies class
DataDensityIndicator	Data density indicator showing quality of the EO input data between 0 – 100 (0 = bad, 100 = perfect data)
Change-Confidence-layer	Quality layer regarding the change detection between 0-3 (0 = no change, 3 = high confidence). Note: this layer is only available for products produced in production mode conso or nrt.

3.2 File Format

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

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 4) 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 4: 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

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 (Table 3). 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 %. The sum of all fractional cover layers for a given pixel is 100. Missing values are set to 255.

3.3.3 Forest type layer

The Forest Type layer provides discrete values per type of forest (see Table 5), for all pixels where the tree (forest) cover fraction exceeds 1 %. Value 255 is used for missing values.

Table 5: 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

[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.

3.3.4 Quality layers

3.3.4.1 Probability of the discrete classification

The probability of the discrete classification is provided as a number between 0 and 100, in steps of 1 %. The higher the probability value, the higher is the confidence that the discrete classification of this pixel is correct. Value 255 is used for missing values

3.3.4.2 Data Density Indicator (DDI)

The Data Density Indicator indicates the availability of input data from the 5-daily PROBA-V UTM ARD+ composites for 100 m and 300 m resolutions. It is a score between 0 = no input data available and 100 = best data availability. Missing DDI values are coded as 255.

3.3.4.3 Change Confidence Layer (only CONSO and NRT maps)

The Change Confidence layer provides the user with a quality layer regarding the change detection of the current mapped year to the previous mapped year. The confidence mask for the break detection has been created based on the amount of evidence there is for the break detection. It is a 3 level confidence mask for all CONSO and NRT LC maps with value definitions stated in Table 6.

Table 6: Change Confidence level coding

Value	Short name	Description
0	No change	No change in discrete class between year and previous year detected
1	Potential change	BFASTmon detected break in second half of NRT year – potential change
2	Medium confidence	Imprint of urban, permanent water, snow or wetland OR change detected by BFAST but HMM model didn't confirm this break in higher resolution OR change detected by BFASTmon in the first half of NRT year
3	High confidence	BFAST detected a change and HMM confirmed this change in higher resolution.

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 7);
- on the band-level, with an example values given for the main discrete classification layer (Table 8).

Table 7: Description of GEOTIFF file attributes

Attribute name	Description	Example(s)
archive_facility	Name of the institution that archives the product	VITO NV
copyright	Ccopyright notice	Copernicus Service information 2020
creator	Principal investigator of the 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
delivered_product_tile_name	Name of the 20x20 degree tile	W160N20
doi	Digital Object Identifier	10.5281/zenodo.3939038
file_creation	File creation timestamp	Fri Apr 19 11:46:08 2019
history	A global attribute for an audit trail.	2020-07-10 Processing line LC100 v4.1
Info	Additional comment on the processing history.	W160N20 for product Discrete-Classification-map of Copernicus Global Land Service LC100 layers for epoch 2015-base
institution	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, Offline 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	V3.0.1
references	Web reference with more product information	https://land.copernicus.eu/global/products/lc
region_name	Name of the geographic area covered. The 20x20° tiles are part of a global product	global
sensor	Name(s) of the sensor(s) used	VEGETATION
source	Method of production of the original data	Derived from 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-base

Table 8: 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-map
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 CGLS-LC100 products in version 3 (collection 3) 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 CGLS-LC100 products in version 3 (collection 3) 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 8).

The position of a pixel is by standard, through GeoTIFF format, given by the "upper left corner", as marked with the GDALMD_ARE_OR_POINT[2] geotransform metadata entry.

[2] GDAL identifies GDALMD_AOP_AREA as "geotransform is the position of the upper left corner of the area spanned by the pixel, which center is the upper left point in the dataset if interpreted as DTM".

3.4.3 Temporal Information

The CGLS-LC100 products in version 3 (collection 3) represents the land cover for a given reference calendar year (from 01 January to 31 December). The data 1 year prior and past of the reference year (called three year epoch) is used in its processing to improve the classification accuracy. As such, the temporal coverage provides a start date of 01 January Year-1 to 31 December Year+1.

For instance, the temporal coverage of the global 2015 CGLS-LC100 product spans from 1st January 2014 to 31 December 2016.

3.5 DATA POLICIES

EU law[3] 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.

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 general statement can be used:

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

"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."

For citation of a specific dataset in scientific publications, the following statement can be used*:*

Buchhorn, M.; Smets, B.; Bertels, L.; Lesiv, M.; Tsendbazar, N.-E.; Masiliunas, D.; Linlin, L.; Herold, M.; Fritz, S. (2020). Copernicus Global Land Service: Land Cover 100m: Collection 3: epoch <YEAR>: Globe (Version V3.0.1) [Data set]. Zenodo. DOI: <DOI>;

With <YEAR>, the actual year of the dataset used, i.e. 2015, 2016, 2017, 2018 or 2019; and <DOI> the digital identifier number of the version as shown in Table 9.

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 9). It displays the various land cover layers (discrete map, cover fractions, change occurrence and change processes) on a map, allows downloading the data in 20x20 degree tiles and reports on land cover statistics per administrative area.

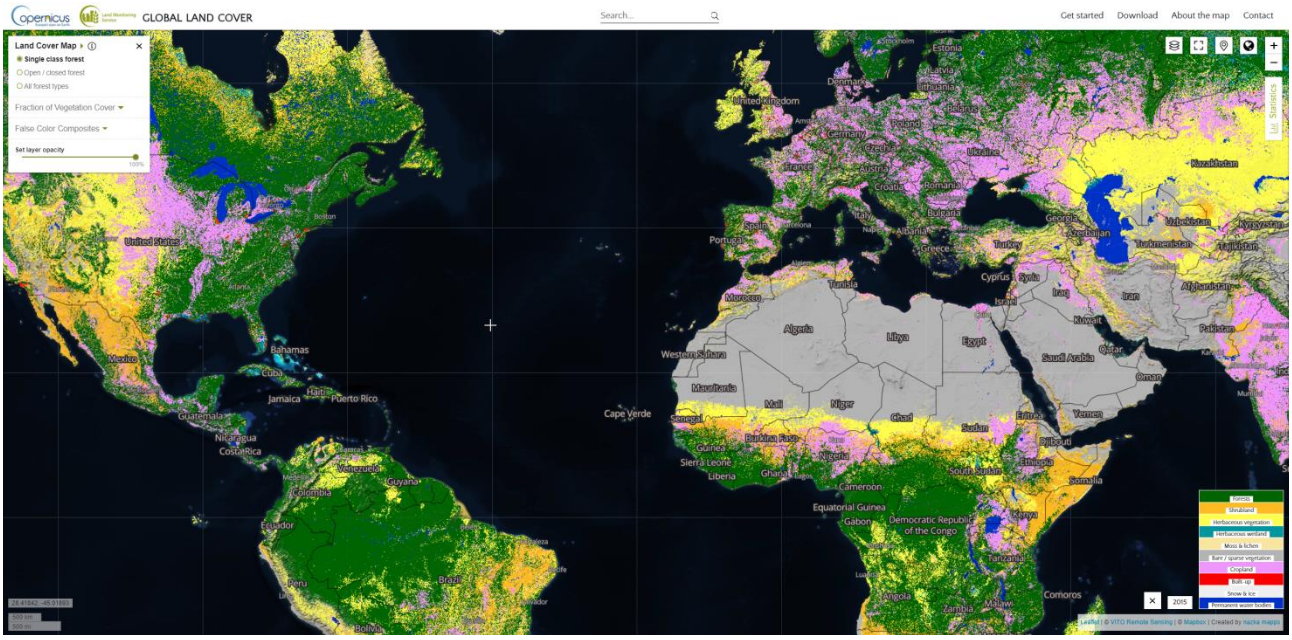

Figure 9: Screenshot of the Global Land Cover viewer

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

Support contact details: https://lcviewer.vito.be/contact or https://land.copernicus.eu/global/contact

Direct access to the data can also be achieved through the Digital Object Identifier (DOI), following the information as shown in Table 9 below.

Table 9: Digital Object Identifier information

Dataset name	Collection 1	Collection 2	Collection 3
Dataset version	V1.0.2	V2.0.2	V3.0.1
Temporal range	2015	2015	2015 - 2019
Geographical range	Africa	Global	Global
Processor algorithm version for Dataset	V1.0	V2.2	V4.1
Concept DOI (always resolves in newest version)	-	10.5281/zenodo.3243508	2015: 10.5281/zenodo.3243508 2016: 10.5281/zenodo.3518025 2017: 10.5281/zenodo.3518035 2018: 10.5281/zenodo.3518037 2019: 10.5281/zenodo.3939049
Version DOI	-	10.5281/zenodo.3243509	2015: 10.5281/zenodo.3939038 2016: 10.5281/zenodo.3518026 2017: 10.5281/zenodo.3518036 2018: 10.5281/zenodo.3518038 2019: 10.5281/zenodo.3939050
Citation	-	Marcel Buchhorn, Bruno Smets, Luc Bertels, Myroslava Lesiv, Nandin-Erdene Tsendbazar, Martin Herold, & Steffen Fritz. (2019). Copernicus Global Land Service: Land Cover 100m: Collection 2: epoch 2015 (Version V2.0.2) [Data set]. Zenodo. DOI: 10.5281/zenodo.3243509	Marcel Buchhorn, Bruno Smets, Luc Bertels, Myroslava Lesiv, Nandin-Erdene Tsendbazar, Dainius Masiliunas, Lin Linlin, Martin Herold, & Steffen Fritz. (2020). Copernicus Global Land Service: Land Cover 100m: Collection 3: epoch <YEAR>: Globe (Version V3.0.1) [Data set]. Zenodo. DOI: <VERSION DOI FOR YEAR>
Direct access	-	https://land.copernicus.eu/global/lcviewer	https://land.copernicus.eu/global/lcviewer
Data layers per epoch	8	20	14 for BASE years; 15 for CONSO and NRT years
File Size	6.6 GB	63.3 GB	187.4 GB

4 VALIDATION RESULTS

The global CGLS-LC100 discrete land cover V3.0 map and the nine cover fraction layers were validated using an independent validation dataset containing around 21 700 points generated in collaboration with regional experts. In addition, the CGLS-LC100 V3.0 discrete map was qualitatively and quantitatively compared against other existing global land cover maps. The validation procedure and the detailed results are presented into the Validation Report [CGLOPS1_VR_LC100m_V3.0]. The main outcomes are summarized below.

Assessments show that the CGLS-LC100 discrete map Level 1 is mapped with 80.6+/-0.4% accuracy. In terms of land cover types, bare/sparse vegetation, snow/ice and permanent water are mapped with high accuracies, while shrubs and herbaceous wetland classes are mapped with lowest accuracies. The yearly (2016-2019) maps are assessed with around 80.3-80.5% accuracy. Overall accuracy at continental level is around 80%, with highest accuracy of 83.7% for Asia and the lowest accuracy of 77.6% for North America. At Level 2, when closed and open forests classes are separated, global overall accuracy is 75.4% +/-0.4% for 2015 while for the other years (2016-2019) it ranges between 75.1-75.2%. Overall accuracies at global and continental levels show consistency in the quality of the yearly maps. Among the cover fraction layer, snow/ice, built-up, water and lichen/moss fraction maps show lowest errors, followed by crops and bare/sparse vegetation fraction types. On the other hand, herbaceous vegetation fraction product has the highest error.

Land cover change between 2015 and 2018 were assessed at change and no change level to gain understanding of the consistency and differences of the annual maps. The overall accuracy is 99.6% of the change/no-change map. Here, the no-change class is mapped with very high accuracy, while change class is more likely to be committed than omitted (land cover change commission error 45.6%, omission error 36%). At continental level, land cover change class has higher accuracies in North Eurasia, North America, Australia and South America. This statistical assessment of land cover change offers the first statistical assessment of generic land cover change at global scale for the most recent time spans. Considering that land cover change detection is much more complex than land cover classification, based on our analysis, the CGLS-LC100 V3.0 yearly maps reflect reasonably well the land cover changes that occur in the recent years globally and offer sufficient stability and consistency in the land cover map accuracies for the yearly maps. Our visual analysis of areas that have changed reveals that they regularly correspond with changes that could be observed from higher resolution images such as Sentinel and Landsat. Users should use the annual land cover maps with confidence but should be careful and critical when doing detailed land cover change analysis since uncertainties (i.e. land change commission and omission errors) and related limitations vary for different world regions.

Comparison of the CGLS-LC100 V3.0 discrete map with the V2.0 discrete map showed that these two versions have similar accuracies, with V3.0 map having marginally higher accuracy (0.1%). Similar tendency applies at continental level. Our visual comparison confirms the similarity of the versions. It also highlights some improvements in the CGLS-LC100 V3.0 with respect to characterizing forest, cropland and permanent water classes. These results indicate that the global CGLS-LC100 V3.0 product is largely consistent with the V2.0 product for 2015.

Spatial uncertainty assessments on the three aggregated classes (forest, cropland and natural vegetation) reveals high level of accuracy in different regions of the world. This achievement is significant considering that the spatial uncertainty assessment is based on large number (>200 000) sample points that are fully independent from the map production. The assessment also highlights some regions where map quality is lower, possibly due to over-estimation of forest class. However, when analyzing three land cover classes, number of points for forest class is larger in comparison to the other two classes and therefore the spatial accuracy maybe biased towards forest rather than crop and natural vegetation.

The comprehensive assessment results conform high quality of the CGLS_LC100m V3.0 land cover product and its improvements as compared to the previous version V2.0. In Validation Report [CGLOPS1_VR_LC100m_V3.0] we further highlight some limitations and potential improvements.

5 REFERENCES

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

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

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

ANNEX: REVIEW OF USERS REQUIREMENTS

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

- **Definition**: Dynamic global land cover products at 300m and/or 100m resolution using UN Land Cover Classification System (LCCS)
- **Geometric properties**:
    - Pixel size of output data shall be defined on a per-product basis so as to facilitate the multi-parameter analysis and exploitation.
    - The baseline datasets pixel size shall be provided, depending on the final product, at resolutions of 100m and/or 300m and/or 1km.
    - The target baseline location accuracy shall be 1/3 of the at-nadir instantaneous field of view.
    - Pixel co-ordinates shall be given for centre of pixel.
- **Geographical coverage**:
    - geographic projection: lat long
    - geodetical datum: World Geodetic System 1984 (WGS84)
    - pixel size: 1/112° - accuracy: min 10 digits
    - 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_LC100m]. Larger consultations in 2017 and 2018 allowed collecting the requirements of wide user communities which were translated in product specifications [CGLOPS1_PSD_LC100m].

Table 10 summarizes the usefulness of information on LC and LC change processes for different international actions and programmes. Table 11 includes the LC classes that are required by different JRC units and are marked by ("X"), respectively. The last column provides the following information: either a class is included in the LC100m V2 product's legend, or it can be derived by users from the fraction layers, or additional R&D is needed.

Table 10: Usefulness of information on LC and LC change processes for different international actions and programmes.

	LC types	Related land change processes	UNFCCC	UNCCD	OECD	SEEA/FAO	SDGs
1	Urban/built-up areas	Urbanization	✓	✓	✓	✓	✓
2	Cropland	Crop expansion	✓	✓	✓	✓	✓
3	Cropland and other vegetation	Land abandonment	✓	✓	✓	✓	✓
4	Forest	Deforestation	✓	✓	✓	✓	✓
5	Forest	Reforestation	✓	✓	✓	✓	✓
6	Wetland	Wetland degradation	✓	✓	✓	✓	✓
7	Water body	Expansion of water surface			✓	✓	✓
8	Water body	Reduction of water surface			✓	✓	✓
9	Bare areas	Desertification			✓	✓	✓

Table 11: List of land cover classes requested by users.

Code Level 1	Code Level 2	UN LCCS level	Land cover class	Forest modelling/REDD+	Crop monitoring	Biodiversity	Monitoring Environment and Security in Africa	Climate modelling	Class included in the product
10		A12A3A20B2	Forest/tree cover	X		X	X	X	Yes
	11	A12A3A20B2D 2E1	Evergreen Needleleaf forest	X			X	X	Yes
	12	A12A3A20B2D 1E1	Evergreen Broadleaf forest	X			X	X	Yes
	13	A12A3A20B2D 2E2	Deciduous Needleleaf forest	X			X	X	Yes
	14	A12A3A20B2D 1E2	Deciduous Broadleaf forest	X			X	X	Yes

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	Class included in the product
	15	A12A3A20B2D 1D2	Mixed forest	X		X			Yes
	16	A12A3A10B2X XXX (assuming that an intact forest is a very dense forest)	Intact forest	X		X		X	To map these classes, addition R&D is required for methodology. Either we develop an expert rule based on other datasets such as: http://www.intactforests.org/data.ifl.html or new training and validation datasets. These classes could potentially be included in the next product evolutions.
	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	Yes
	21	A12A4A20B(B9)XXE1	Evergreen shrubs			X			These classes could be potentially included in the next product evolutions. We will have to collect corresponding training and validation data.
	22	A12A4A20B3(B9)XXE2	Deciduous shrubs			X			
30		A12A2(A6)A20B4	Herbaceous vegetation			X	X	X	Yes
		A12A6A10 // A11A1A11B4XXXXXXF2F4F7G4-F8	Pasture/managed grassland					X	To map these classes, addition R&D is required for methodology. Also we will have to develop new training and validation datasets.
		A122(A6)A10	Natural grassland			X		X	
		A12A2	Grass types for Western Africa			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	Class included in the product
									These classes could potentially be included in the next product evolutions.
		A12A3A11B2XXXXXXF2F4F7G4-A12; A12A3A11B2-A13; A12A1A11	Savannas			X			The LC100 V2 product includes such a LC class as open forest, which is a mix of trees (more than 15%), shrubs and grassland. This class only partly corresponds to savannas because a 100m x100m pixel may include less trees but still be considered as savanna. However, users are encouraged to use the fraction layers to produce their own savanna layer by applying specific thresholds for tree, shrub and grass cover.
40		A11A3	Cultivated and managed vegetation/agriculture		X	X	X	X	Yes
	41	A11A3XXXXXXD3(D9)	Irrigated cropland		X			X	To map these classes, addition R&D is required for methodology. Also we will have to develop new training and validation datasets.
	42	A11A3XXXXXXD1	Rainfed cropland		X			X	
	43	A11A3	Big and small farming/field size		X				
	44	A11A1-W8/A2	Permanent crops		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	Class included in the product
	45	A11A3	Row crops		X				These classes could potentially be included in the next product evolutions.
		A11A2	Crop types: long/short cycle or winter/summer crops		X				
		A11A2	Multiple crop cycles		X				
50		B15A1	Urban/built up			X	X	X	Yes
60		B16A1(A2)	Bare/sparse vegetation				X	X	Yes
70		B28A2(A3)	Snow and Ice				X	X	Yes
80		B28A1	Open water				X	X	Yes
90		A24A1(A2/A3/A4)	Wetland			X	X	X	Yes
		A24A3	Mangroves	X		X			To map this class, addition R&D is required for methodology. Also we will have to develop new training and validation datasets. These classes could potentially be included in the next product evolutions.