Climate Change Initiative Extension (CCI+) Phase 1

New Essential Climate Variables (NEW ECVS)

High-Resolution Land Cover ECV (HR_LandCover_cci)

Product User Guide

(PUG)

Prepared by:

Università degli Studi di Trento
Fondazione Bruno Kessler
Università degli Studi di Pavia
Università degli Studi di Genova
Université Catholique de Louvain
Politecnico di Milano
Université de Versailles Saint Quentin
CREAF
e-GEOS s.p.a.
Planetek Italia
GeoVille

Changelog

Issue	Changes	Date
1.0	First version.	21/04/2019
1.1	Final version including the modifications required by the RIDs	04/05/2020
1.2	Modified Historical HR Land Cover Map name description and added annex for reprojection tool. Update of summary product table and of legend table. Minor modifications to naming of Static and Dynamic products to align them with actual implementation.	04/02/2021
1.3	Added a section to access WMS and WCS services (paragraph 10.1) and modified section on archive structure (paragraph 2.5)	03/05/2021
2.0	Full update of the document that documents the project final products delivered with CRDP_v3.0. Removed section with access to WMS and WCS services, service is no longer available.	08/11/2022
2.1	Final update including sections reporting benchmarking and validation accuracy figures.	22/02/2023
2.2	Included section 2.4 about products download, and minor corrections.	08/04/2024

Detailed Change Record

Issue	RID	Description of discrepancy	Sections	Change
1.1	FR-01	The LC class legend reported in the table 2 is the one generated by the climate group and as per AI#41 Climate group to circulate changed legend to the users to have feedback?	2.6 Legend for classification products	The legend related to AI#41 is the LCC legend (as coming from the Users survey). The legend has been updated with latest version and a clarification on the mapping between pixels value and legend codes has been added.
1.2	FR-02	During production, it has been established that Historical products should follow the MGRS distribution as for Static products for analysis' compatibility. A reprojection tool has to be delivered too.	4.1.2 legend for naming convention 9: Annex 1	Changed naming convention as for actual production specification. Added paragraph 9 for reprojection tool description

Contents

Applicable documents

Ref.	Title, Issue/Rev, Date, ID
[AD1]	CCI HR Technical Proposal, v1.1, 16/03/2018
[AD2]	CCI Extension (CCI+) Phase 1 – New ECVs – Statement of Work, v1.3, 22/08/2017, ESA-CCI-PRGM-EOPS-SW-17-0032
[AD3]	Data Standards Requirements for CCI Data Producers, v2.1, 02/08/2019, CCI-PRGM-EOPS-TN-13-0009
[AD4]	CCI_HRLC_Ph1-D1.1_URD_v3.1
[AD5]	CCI_HRLC_Ph1-D1.2_PSD_v4.0
[AD6]	CCI_HRLC_Ph1-D2.1_PVASR_v3.1
[AD7]	CCI_HRLC_Ph1-D2.2_ATBD_v4.0
[AD8]	CCI_HRLC_Ph1-D2.3_E3UB_v4.0
[AD9]	CCI_HRLC_Ph1-D2.5_PVP_v4.2
[AD10]	CCI_HRLC_Ph1-D4.1_PVIRa_v2.1
[AD11]	CCI_HRLC_Ph1-D4.1_PVIRb_v2.2
[AD12]	CCI_HRLC_Ph1-D4.1_CRDP_v3.0
[AD13]	CCI_HRLC_Ph1-D5.1_CAR_v3.0

Reference documents

N/A

Acronyms and abbreviations

API	Application Programming Interface
ASAR	Advanced Synthetic Aperture Radar
ATTO	Amazon Tall Tower Observatory
BFAST	Breaks For Additive Seasonal and Trend
CCI	Climate Change Initiative
CMC	Climate Modelling Community
CMUG	Climate Modelling User Group
CRDP	Climate Research Data Package
CRS	Coordinate Reference System
ECV	Essential Climate Variable
ENVISAT	Environmental Satellite
EO	Earth Observation
ERS	European Remote Sensing
ESA	European Space Agency
FAO	Food and Agriculture Organization
GCOS	Global Climate Observing System
GeoTIFF	Geographic Tagged Image File Format
HR	High Resolution
HRLC	High Resolution Land Cover
JECAM	Joint Experiment for Crop Assessment and Monitoring
LC	Land Cover
LCCS	Land Cover Classification System

LCCS3	Land Cover Classification System, version 3
LCML	Land Cover Meta Language
MGRS	Military Grid Reference System
MR	Medium Resolution
NASA	National Aeronautics and Space Administration
NetCDF	Network Common Data Form
ORCHIDEE	Organising Carbon and Hydrology In Dynamic Ecosystems
PCC	Post Classification Comparison
PFT	Plant Functional Type
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SFT	Surface Functional Type
SITS	Satellite Images Time Series
SODEEP	Study of the development of extreme events over permafrost areas
SoW	Statement of Work
UNFCCC	United Nations Framework Convention on Climate Change
UTM	Universal Transverse Mercator
WFS	Web Feature Service
WGS84	World Geodetic System, version 1984
WMS	Web Map Service
WP	Work Package
ZOTTO	Zotino Tall Tower Observation Facility

1 Introduction

1.1 Purpose and scope

The purpose of the Product User Guide (PUG) is to present the product to users through detailed information regarding thematic content, data formats, filenames and conventions and metadata related to the project production outcome delivered with CRDP_v3.0 [AD12].

The Phase-I of the Climate Change Initiative (CCI) High Resolution Land Cover (HRLC) project generated a suite of products, with subcontinental spatial extent, consisting of:

- **HRLC10:** High Resolution Land Cover Maps at 10m spatial resolution for year 2019 (also referred to as static maps).
- **HRLC30:** High Resolution Land Cover Maps at 30m spatial resolution for years 1990, 1995, 2000, 2005, 2010, 2015, 2019 (also referred to as historical maps).
- **HRLCC30:** High Resolution Land Cover Change Maps at 30m spatial resolution of yearly changes. A map every 5 years (1990-1995, 1995-2000, 2000-2005, 2005-2010, 2010-2015,2015-2019) is provided which reports (high priority) changed pixels and their year within the 5-years temporal interval.
- Associated uncertainty products for each of the classification product lines HRLC10, HRLC30 and HRLCC30.

The three sub-continental areas of interest are within Africa, South-America and Siberia, therefore a representation of very different ecosystems and climatic conditions.

1.2 Background of the project

The European Space Agency (ESA) CCI projects will deliver the next generation of satellite derived geophysical parameters, with quantified uncertainties that will allow each parameter to be assessed against requirements from the Global Climate Observing System (GCOS) for Essential Climate Variables (ECV) and the Climate Modelling Community (CMC), represented within the CCI program by the Climate Modelling User Group (CMUG).

The objective of the CCI is to realize the full potential of the long-term global Earth Observation (EO) archives that ESA together with its Member states have established over the last thirty years, as a significant and timely contribution to the ECV databases required by United Nations Framework Convention on Climate Change (UNFCCC). The programme is organized in 2 phases.

The CCI Phase-I provided a unique opportunity for the European EO science community to define and validate innovative approaches for continuously generating and updating a comprehensive and consistent set of ECV global satellite based data products in the long term – i.e. decades hence. The focus was on a major sustained, and coordinated scientific effort to review and improve underlying processing, retrieval and validation methods.

In this context the outcome of the CCI HRLC project is an ensemble of validated products that fully exploits the capabilities of a wide range of Remote Sensing data (Copernicus Sentinels, NASA Landsat and SAR missions, etc.), it is developed in sustainable way and scalable to global production, and it followed a cyclic development phase including feedbacks from the Climate modelers community.

1.3 Structure of the document

The remainder of the document is organized as follows. Firstly the main structure of the project is summarized. Next, some considerations about products requirements are given. Then, the main project production outcomes HRLC10, HRLC30, HRLCC30 are described. Finally, some details for data visualization are given.

2 CCI HRLC project

2.1 High resolution land cover concept

A specific requirement for all CCI projects is the full exploitation of operational Copernicus services such as the Sentinels, and other valuable satellite data sources (historical archives of Landsat and SAR missions). Considering quality data availability requirements, this allowed a temporal span 1990-today, at the time of the project proposal, therefore 1990-2019, which includes data from Landsat-5/7/8, and Sentinel-1/2, ERS-1, ERS-2, ENVISAT-ASAR.

In view of the high spatial resolution of Sentinels and Landsats, which is in the order of tens of meters, the concept of land cover classes can be essentially improved with respect to the Medium Resolution (MR) products. This aspect fostered high expectations on the users and an in-depth user requirements analysis has been performed during the first year of project activity, as reported in the URD [AD4].

At its core, the proposed classification system is based on the LCCS approach upgraded to LCML/LCCS3, as this seemed very convenient for the, the LMCL system allows to: (i) characterize geographical features and describe phenomena related to specific bio-geo-physical processes and people activities by the means of valuable attributes, and, (ii) to better translate the LC classes in the Surface Functional Types used in climate models to describe processes and their feedbacks on water and carbon cycles. An agreement of a refined classification system was achieved after two years of cyclic feedback-based interaction involving the EO Science team, the Validation team and Climate team. The agreed classification system collects the demand of the climate modellers, the feasibility of the estimated EO skills classifiers and the provision and specifications of the samples for the validation procedures.

From the RS perspective, improving land cover products from medium resolution (300m) to high resolution (10/30m) required a change of paradigm in the data analysis and processing scheme that came from the intrinsic different characteristics of the data at higher resolution. In particular,

- For the production of MRLC maps (300m) images are available with finer temporal resolution (daily) than for 10m HRLC (2-5 days) and 30m HRLC (16 days).
- Sentinel data cannot serve for historical mapping, therefore the consortium decided to make two distinct lines of products: one at 10m based on Sentinel-1/2 data (only for year 2019), and one at 30m based on Landsat-5/7/8 data (historical series from 1990 to 2015 every five years and for year 2019).
- As a consistency requirement, the legend is defined to agree with classification capabilities of both production lines.

2.2 Key areas location

Considering the aims of Phase-I of the CCI project for which production can be bounded to sub-continental extents, the aim of the Consortium was therefore to select some key areas that could maximize the potential of intersecting HRLC products with as many as possible Climate-related (and other user communities) data sources.

As a result of the key users consultation, giving advantages to climate modellers demands (especially MPI and Met Office who participated both to the questionnaire and virtual meetings), a choice has been made for three zones that could include a large number of ecosystems and regions that are of interest for other ongoing studies, for example the region of the SODEEP project of MPI team or Equatorial Guinea for Met Office. Extensions in the latitudinal gradients were chosen rather than in longitudinal ones to cover a larger diversity of climatic zones in order to better assess climate-surface relationships. The locations were also optimized to include a maximum number of FluxNet stations (in situ site with eddy covariance flux measurements), which could be used to validate methods and models (example of ATTO and ZOTTO stations in Amazonia and Siberia respectively). All the details regarding user consultations and climatological assessment of the areas are available in URD [AD4] and PSD [AD5].

The geographical coordinates of the three selected regions are summarized hereafter and illustrated in Figure 1. In green are represented the areas for the HRLC10 products (the static ones) and in orange the areas for the HRLC30 and HRLCC30 (the historical products):

Africa: This region includes the Sahel band and East Africa which is a complex climatic region which experiences severe climatic events (droughts in the 70's and 80's and floods more recently) often attributed to climate warming.

- Static map: (0.1°S – 18.1°N; 9.9°E – 43.3°E),
- Historical LC and LCC maps: (3.5°N – 16.3°N; 27.0°E – 43.3°E)

Amazonia: This region includes the Amazon basin which has for several decades focused the attention of the scientific community due to large deforestation rates and potential associated large-scale climate impacts. Agricultural expansion and climate variability have become important agents of disturbance in the Amazon basin, mainly in the southern and eastern portions.

- Static map: (23.6°S - 0°S; 42.9°W – 62.1°W)
- Historical LC and LCC maps: (23.6°S – 11.7°S; 46.7°W – 62.1°W)

Siberia: The region is situated in the northern high latitudes, in Siberia, for which future climate changes are expected to be particularly strong, a phenomenon known as polar amplification. In Siberia, complex climate feedbacks over land, implicating natural and human factors, may further amplify these changes and make this region as a possible hot spot of future climate changes.

- Static map: (51.3°N – 75.7°N; 64.4°E – 93.4°E),
- Historical LC and LCC maps: (59.4°N – 73.9°N; 64.8°E – 87.4°E).

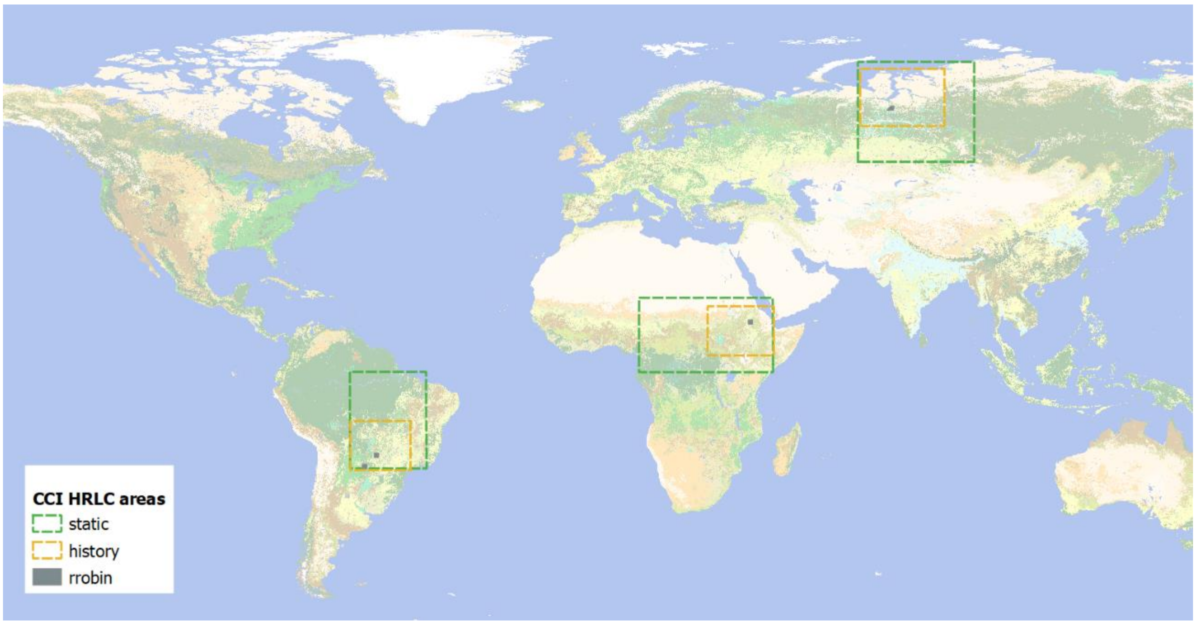


Figure 1. Graphic representation of the location of the 3 study areas in Amazon, Africa and Siberia.

2.3 Project outputs

The outputs of the CCI HRLC Phase-I project concern sub-continental land cover products as detailed in Table 1 with structured uncertainty products associated with them. All data being delivered along with metadata. The outputs also include software systems description, products documentation and benchmarking and validation reports.

Table 1. Summary table of the HRLC products.

	HRLC10 (static maps)	HRLC30 (historical maps)	HRLCC30 (change detection maps)
Products	Classification Map	Classification Map	Change Detection Map
Resolution (meters)	10	30	30
Source Data	Sentinel 2 Sentinel 1	Landsat 5-7-8 ASAR, ERS 1-2	Landsat 5-7-8
Years	2019	1990, 1995, 2000, 2005, 2010, 2015, 2019	1990-2019
Projection	UTM (MRGS products) WGS84 (Mosaics)	UTM (MRGS products) WGS84 (Mosaics)	UTM (MRGS products) WGS84 (Mosaics)
Grid	MGRS	MGRS	MGRS
Uncertainty products available	Yes	Yes	Yes
Encoding	Multiple GeoTIFF files, one for each MGRS tile. NetCDF for mosaicked products.	Multiple GeoTIFF files, one for each MGRS tile. NetCDF for mosaicked products.	Multiple GeoTIFF files, one for each MGRS tile. NetCDF for mosaicked products.

2.4 Download directory structure

Project page in the CEDA Archive: https://catalogue.ceda.ac.uk/uuid/b057708eec1042238fb333ab02ec772e

```
Root:
https://data.ceda.ac.uk/neodc/esacci/high_resolution_land_cover/data/land_cover_maps

/A01_Africa
/A02_Amazonia
/A03_Siberia

      /historical
            /v1.2
                  /geotiff
                        /HRLC30
                              /mosaic
                              /tiles
                                    /1990
                                    /1995
                                    /2000
                                    /2005
                                    /2010
                                    /2015
                                    /2019
                        /HRLCC30
                              /mosaic
                              /tiles
                                    /1990_1995
                                    /1995_2000
```

```
                                        /2000_2005
                                        /2005_2010
                                        /2010_2015
                                        /2015_2019
        /static
                /v1.2
                        /geotiff
                                /HRLC10
                                        /mosaic
                                        /tiles
                                                /2019
```

/01_Africa/static/v1.2/geotiff/HRLC10
https://catalogue.ceda.ac.uk/uuid/f107a4ce186844bb8adf8cd1f2f6d552
/02_Amazonia/static/v1.2/geotiff/HRLC10
https://catalogue.ceda.ac.uk/uuid/0bc7042123984c69aa45cb6788bfdaa0
/03_Siberia/static/v1.2/geotiff/HRLC10/
https://catalogue.ceda.ac.uk/uuid/e7864129084c4baaa34be3a1cfaaa13d
/01_Africa/historical/v1.2/geotiff
https://catalogue.ceda.ac.uk/uuid/a3fb75aa46db4711ab587f3fa3ca01fe
/02_Amazon/historical/v1.2/geotiff
https://catalogue.ceda.ac.uk/uuid/b053b51e854d484a9657f6bfb5ebd516
/03_Siberia/historical/v1.2/geotiff
https://catalogue.ceda.ac.uk/uuid/854cc98dbc634cdb8afa8835994428f5

2.5 Independent validation

All HRLC products have subcontinental extent, which at the HR scale of 10m is measurable in the order Gpixels. To address the complexity of a proper validation exercise , given the large size of the data, two independent validation methodologies have been used in the project: i) large scale validation on the basis of "reliable" labelled samples extracted from the agreement of a large number of available HRLC products (called intercomparison); and ii) traditional validation on a reduced set of samples extracted by photo-interpretation. The inter-comparison activity consisted of providing a very significant number (in the order of billions) of "reliable" labelled pixels obtained through systematic intersection and analysis of several available HRLC products. The photo-interpretation activity focused on providing a much smaller number (in the order of thousands) of labelled samples through manual interpretation by experts. The main advantage of the former is the huge number of samples considered; the disadvantage is related to the possible errors in the agreement obtained and on the bias toward agreement classes. The main advantage of the latter is that it also allows to analyse seasonal classes (which are not present in the products used for inter-comparison); the disadvantage is the very small number of samples that results in high uncertainty on the considered area. All the details of the planned activities related to the two validation strategies and their outcomes can be checked in PVP [AD9], PVIRa [AD10], and PVIRb [AD11].

The outcomes of the two strategies mentioned above, are intended to provide a global overview and a preliminary assessment on the products quality.

The HRLC10 products map LC on the three static areas for the year 2019. Given the recent target date, a lot of information has been gathered for benchmarking in the inter-comparison activity. Validation based on photo-interpretation presented more difficulties because, at the unprecedented resolution of 10m, sample labelling required understanding of the landscape, which makes production of verified samples a slow process. The main consequence is limited confidence in the accuracy estimations because of the small number of samples available.

It is planned for the next phase activities to increment the number of photo-interpreted samples by at least one order of magnitude.

Due to the large period investigated for HRLC30 and HRLCC30, the historical products mapping LC and LC change from 1990 to 2019, reference data availability has been a critical aspect for both the benchmarking and validation activities. Related methodologies required adaptations steps to properly harmonize available HR LC information for building intercomparison maps and photo-interpretation. Given the difficulty in gathering photo-interpreted information for such a long span of data, especially past in time where data are scarce, validation information can only be considered partial. The most critical open issues for validating HR LC for historical products are specifically illustrated in PVIRb [AD11].

2.5.1 Inter-comparison with existing products

To maximize the usefulness of already existing HR LC products (gathered by the joint efforts of the Consortium during the whole project duration) and simplify the interpretation of the results, a new approach to product benchmarking was introduced that relies on the concept of "map of agreement". The approach consists of intersecting all the available existing HRLC products (gathered during the activity of the project) in a region and extracting coherent information from them. The result of the intersection is a map that we call the map of agreement. The idea behind the approach is that every land cover map aims at representing material on the Earth's surface as accurately as possible. When multiple existing datasets are compared among themselves, the area in which they all show consistent information is the area with the highest probability that they are correct. Since no existing HRLC product currently available has the same level of detail in terms of LC classes as those produced in HR_LandCover_cci, the definition of map of agreement requires aggregating some classes to make them match those obtained by intersection of existing products. More on the products used and on the definition of the map of agreement approach can be found in PVP [AD9].

2.5.2 Validation by photo-interpretation

A statistical validation framework was established in the context of the GlobCover and CCI Moderate Resolution MRLC projects. This statistical validation framework is organized in four steps as described in Figure 4. These steps were adjusted to the CCI HRLC mapping activities.

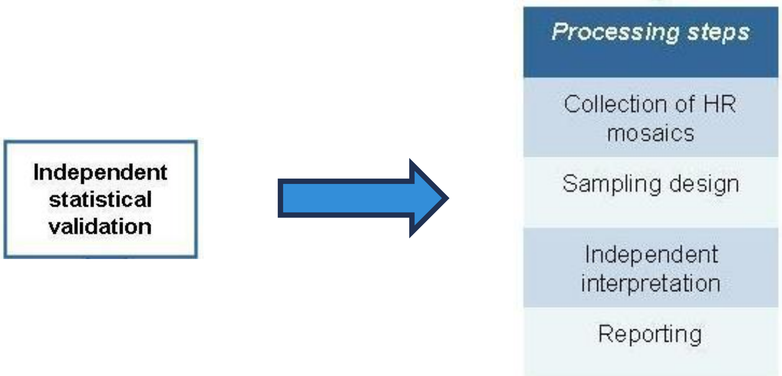


Figure 4. Main components of the independent statistical validation procedure.

LC product accuracy is assessed by comparison with independent data sources such as ground based in-situ measurements or higher resolution satellite data. The collection of higher resolution satellite as "ground truth" is considered as the best option to support the validation of remote sensing products in general and of the HR LC maps in particular. Reference data samples are collected on specific areas designed to be statistically valid for accuracy assessment of the HRLC products, to address the issues of rare classes with a strong impact on the climate system (urban areas, wetlands, etc.) and to bias a specific sampling for the LC change validation. The reference data sources are then intended to be interpreted over each validation sample through an international network of experts in a standardized manner. Then, the classification is compared to the reference data, and the extent to which these two classifications agree is defined as accuracy. The full methodology and the accuracy figures are documented in the PVP [AD9] and complete results are available in PVIRb [AD11].

3 Technical specifications

3.1 File format considerations

Regarding the format of the products, they are finally delivered in GeoTiff format and this choice comes from important considerations implied by the average size of a HRLC product:

GeoTIFF:

- Is portable but has not a strict way of storing metadata and there is not a common convention. Anyway, metadata can be added.
- Coordinates are derived by the context (top left pixel position, resolution in x and y).
- Has the concept band which is less general that a sub-dataset as it inherits from the file encoding (e.g. it does not support different types).

For this reason, as a general behaviour, the GeoTIFF cannot store different datatypes (e.g. int and float together) so it is impossible to store in the same file both a classification (e.g. integer 8 bit) and a posterior probability (e.g. float 32 bit) or at least a transformation has to be performed like converting the probability in an integer value from 0 to 100 discarding some of the precision.

For data types, please consider the following references:

- "Band types of Byte, UInt16, Int16, UInt32, Int32, Float32, Float64, CInt16, CInt32, CFloat32 and CFloat64 are supported for reading and writing. Paletted images will return palette information associated with the band" from https://gdal.org/drivers/raster/gtiff.html
- Data Type page from NetCDF official documentation: https://www.unidata.ucar.edu/software/netcdf/documentation/NUG/data_type.html

In the case of GeoTIFF, palette information will be reported whenever possible with information about the legend (see paragraph 3.4).

3.2 File naming convention

In order to enhance the user experience with HRLC data, a unique file naming convention is used across all products. Table 2 shows the naming pattern with all its fields.

Table 2. Naming structure of the HRLC products.

```
ESACCI-HRLC-L4-{product_type}-{product_id}-{tiling_id}-{spatial_res}-{temporal_freq}-{epoch}-fv01.0.{file_extension}
```

Table 3 shows the values that can be assigned to the placeholders of the naming pattern, according to their meaning, for each HRLC product.

Table 3. Detail of field values for the naming of the different HRLC products.

field	values	explanation	notes
product_type	MAP	Classification product (CL01)	
	UNCERT	Uncertainty product (CL02,PS01,PS02,IQIX)	
	CHANGE	Change detection product (CDET)	

product_id	CL01	First class (accord. to classifier ranking)	Further subdivision of product types. A distinct code for each product available.
	CL02	Second class (accord. to classifier ranking)	
	PS01	Posterior probability associated to CL01	
	PS02	Posterior probability associated to CL02	
	IQIX	Input quality index	
	CDET	Change detection	
tiling_id	<A>NN<T>MMXXX	For tile products	
	<A>NN<MOSAIC>	For mosaic products	
spatial_res	10m / 30m	Spatial resolution	
temporal_freq	P1Y / P5Y	Temporal resolution, in years	
epoch	YYYY	Epoch of the product (e.g., 1990,2000,2019)	
file_extension	.tif / .nc	File format (.tif for tile products and .nc for mosaic products)	

With:

NN: 01=Africa, 02=Amazon, 03=Siberia

MMXXX: tiling nomenclature a 100x100 km2 tile in the MGRS format: e.g., 45XVC.

3.3 Global attributes (metadata)

The table below contains the description of the global attributes as defined in CCI Data Standards with an example reported in the table.

Table 4. Global attributes of HRLC products and example values.

Global Attribute	**Example**
title	ESA CCI High Resolution Land Cover L3P derived from Sentinel 2 L2A and Sentinel 1 L1GRD
Institution	ESA CCI HRLC Consortium
institution_abbreviation	
source	CCI High Resolution Land Cover Static Map Processing Chain 1.0
history	2020-09-01T14:17:10 UTC -- Creation from processor version v1.0
references	https://climate.esa.int/en/projects/high-resolution-land-cover/
Conventions	CF-1.6, ACDD-1.3, ISO 8601
product_version	1.0
summary	Classification product derived from processing of Sentinel 1 and Sentinel 2 with resolution of 10 m over specific Area of Interest with Legend based on a derivation of LCCS
keywords	satellite, earth observation, land cover

Id	ESACCI-HRLC-L4-MERGED-SAHEL-29PPL-00010m-P1Y-2019-fv01.0.nc
naming authority	Egeos
keywords_vocabulary	NASA Global Change Master Directory (GCMD) Science
keywords	land cover classification
cdm_data_type	Grid
comment	The products are generated as part of ESA CCI HRLC project
date_created	20200901T120000Z
date_modified	20200901T120000Z
creator_name	e-GEOS
creator_url	https://www.e-geos.it
creator_email cersat@ifremer.fr	cci-hrlc@e-geos.it
creator_institution	e-GEOS
project	Climate Change Initiative - European Space Agency
geospatial_lat_min	4
geospatial_lat_max	16
	degrees_north
geospatial_lon_min	27
geospatial_lon_max	43.5
geospatial_lon_units	degrees_east
geospatial_vertical_min	0
geospatial_vertical_max	1180
time_coverage_start	20190101T000000Z
time_coverage_end 20170130T154708Z	20191231T000000Z
time_coverage_duration	PT1Y
time_coverage_resolution	PT1Y
geospatial_bounds	POLYGON ((27. 4., 27. 16., 43.5 16., 43.5 4., 27. 4.))
standard_name_vocabulary	NetCDF Climate and Forecast (CF) Metadata Convention version 1.6
license	ESA CCI Data Policy: free and open access
platform	Sentinel 2, Sentinel 1
sensor	Multi Spectral Instrument (MSI)
spatial_resolution	10 m

netcdf_version_id	4.3.0 of Jul 8 2013 12:17:12 $
acknowledgement	Please acknowledge the use of these data with the following statement: these data were obtained from the ESA CCI High Resolution Land Cover project
format_version processing_level	Data Standards Requirements for CCI Data Producers, v2.1, 2 August 2019
processing_software	CCI High Resolution Land Cover Static Map Processing Chain 1.0 deployed on e-GEOS Processing System
processing_level	L3P
key_variables	class
geospatial_lat_units	meters
geospatial_lon_units	meters
geospatial_lon_resolution	10
geospatial_lat_resolution	10

3.4 Legend of the classification products

The Food and Agriculture Organization (FAO) Land Cover Classification System (LCCS) was found pertinent to support the description of the CCI HRLC maps. During the initial phase of user requirements definition, the community of key users was surveyed and technical considerations based on Remote Sensing (RS) data properties were made. The outcome of this feedback process was an adaptation of the FAO LCCS framework that resulted in the 16 classes for HRLC mapping detailed hereafter. More details are available in PSD [AD5].

Figure 2 summarizes the final legend used for all classification products delivered in this project along with its internally coded pixel values, default colours and class names. Table 5 provides a thematic description of the HRLC classes in the FAO LCCS standard.

HRLC CLASSES			
CODE	DESCRIPTION		
0	No data		
10	Tree cover evergreen broadleaf		
20	Tree cover evergreen needleleaf		
30	Tree cover deciduous broadleaf		
40	Tree cover deciduous needleleaf		
50	Shrub cover evergreen		
60	Shrub cover deciduous		
70	Grasslands		
80	Croplands		
90	Woody vegetation aquatic or regularly flooded		
100	Grassland vegetation aquatic or regularly flooded		
110	Lichens and mosses		
120	Bare areas		
130	Built-up		
140	**Open water**	141	Open water seasonal
		142	Open water permanent
150	Permanent snow and/or ice		

Figure 2. The legend common to all HRLC classification products.

Table 5. FAO LCCS description of the classes of the CCI HRLC products.

Code	Label	Description
10	**Tree cover evergreen broadleaf**	Primarily vegetated areas with a tree canopy cover of more than 50 % at the time of fullest development. Snow and/or ice, open water or built-up areas cover less than 50% of the area. A tree is a woody, perennial plant with a simple and well-defined stem, bearing a more or less defined crown and a minimum height of 5 m. Tree canopy cover is composed of trees that are never entirely without green foliage. Trees are broadleaved and come from the Angiospermae group.
20	**Tree cover evergreen needleleaf**	Primarily vegetated areas with a tree canopy cover of more than 50 % at the time of fullest development. A tree is a woody, perennial plant with a simple and well-defined stem, bearing a more or less defined crown and a minimum height of 5 m. Tree canopy cover is composed of trees that are never entirely without green foliage. Trees carry typical needle-shaped leaves and come from the Gymnospermae group.
30	**Tree cover deciduous broadleaf**	Primarily vegetated areas with a tree canopy cover of more than 50 % at the time of fullest development. Snow and/or ice, open water or built-up areas cover less than 50% of the area. A tree is a woody, perennial plant with a simple and well-defined stem, bearing a more or less defined crown and a minimum height of 5 m. Tree canopy cover

		composed of trees that are leafless for a certain period during the year. Trees are broadleaved and come from the Angiospermae group.
40	**Tree cover deciduous needleleaf**	Primarily vegetated areas with a tree canopy cover of more than 50 % at the time of fullest development. Snow and/or ice, open water or built-up areas cover less than 50% of the area. A tree is a woody, perennial plant with a simple and well-defined stem, bearing a more or less defined crown and a minimum height of 5 m. Tree canopy cover is composed of trees that are leafless for a certain period during the year. Trees carry typical needle-shaped leaves and come from the Gymnospermae group.
50	**Shrub cover evergreen**	Primarily vegetated areas with a shrub canopy cover of more than 50 % at the time of fullest development. Snow and/or ice, open water or built-up areas cover less than 50% of the area. A shrub is a woody perennial plant with persistent woody stems and without any defined main stem, being less than 5 m tall. Shrub canopy cover composed of shrubs that are never entirely without green foliage.
60	**Shrub cover deciduous**	Primarily vegetated areas with a shrub canopy cover of more than 50 % at the time of fullest development. Snow and/or ice, open water or built-up areas cover less than 50% of the area. A shrub is a woody perennial plant with persistent woody stems and without any defined main stem, being less than 5 m tall. Shrub canopy cover is composed of shrubs that are leafless for a certain period during the year.
70	**Grassland**	Primarily vegetated areas with an herbaceous cover of more than 50% at the time of fullest development. Snow and/or ice, open water or built-up areas cover less than 50% of the surface. Herbaceous plants are defined as plants without persistent stems or shoots above ground and lacking definite firm structures.
80	**Croplands**	Primarily vegetated areas with a herbaceous cover of more than 50 % at the time of fullest development. Snow and/or ice, open water or built-up areas cover less than 50%. Croplands are mainly herbaceous plants that are sowed/planted and harvestable at least once within the 12 months after the sowing/planting date. Herbaceous plants are defined as plants without persistent stems or shoots above ground and lacking definite firm structures. Cropland includes rainfed crops, irrigated crops, aquatic crops and annual pastures. It is an adaptation of the Joint Experiment for Crop Assessment and Monitoring (JECAM) cropland definition. Croplands exclude permanent crops like woody plantations that are part of the tree or shrub classes.
90	**Woody vegetation aquatic or regularly flooded**	Primarily vegetated areas with trees and/or shrubs, grasslands or lichens and mosses covering more than 50 % of the area flooded by water for more than 4 months throughout the year. The water can be saline, fresh or brackish.
100	**Grassland vegetation aquatic or regularly flooded**	Primarily vegetated areas with grasslands and/or lichens and mosses covering more than 50 % of the area flooded by water for more than 4 months throughout the year. The water can be saline, fresh or brackish.
110	**Lichen and mosses**	Primarily vegetated areas with a cover of more than 50% at the time of fullest development. Snow and/or ice, open water or built-up areas cover less than 50% of the surface. Mosses are a group of photo-autotrophic land plants without true leaves,

		stems or roots. Lichens are composite organisms formed from the symbiotic association of fungi and algae.
120	**Bare areas**	Areas where the sum of vegetation cover is less than 50% at the time of fullest development. Snow and/or ice, open water or built-up areas cover less than 50% of the surface. Bare rock areas, sands and deserts are classified as bare areas. Extraction sites (open mines and quarries) and salt flats covered by water for less than 5 months are classified as bare areas.
130	**Built-up**	Areas where any predominant type of linear and non-linear artificial surface covers at least 50%. Snow and/or ice, and open water cover less than 50% of the surface. Built-up areas include buildings, roads, airports, greenhouses, etc. but may, however, exclude temporary settlements.
141	**Open Water seasonal**	Areas where open water covers at least 50% of the surface and remains between 5 and 9 months a year, except in special circumstances (particularly dry years, construction of dams, etc.). Snow and/or ice and built-up areas cover less than 50% of the surface. Water bodies can be natural or artificial. Water can be saline, fresh or brackish.
142	**Open Water permanent**	Areas where open water covers at least 50% of the surface and remains for more than 9 months a year, except in special circumstances (particularly dry years, construction of dams, etc.). Snow and/or ice and built-up areas cover less than 50% of the surface. Water bodies can be natural or artificial. Water can be saline, fresh or brackish.
150	**Permanent snow and ice and/or Ice**	Areas where snow and/or ice cover at least 50% of the surface for more than 9 months a year. Built-up areas and open water cover less than 50% of the surface.

4 HRLC10: High Resolution Land Cover Map at 10m products

4.1 Product description

The CCI HRLC project provides to the climate community a land-cover map at the resolution of 10m (HRLC10) of three sub-continental areas: Africa (Sahel band), Amazon and Siberia, these are represented in Figure 3. The main sources of information from which the land cover class is inferred are Sentinel-2 optical imagery and Sentinel-1 for radar data. This focus is explained by the desired spatial resolution of 10m of the product. The HRLC10 maps represent land cover information for year 2019. The legend of the classification map is presented in Section 3.4. The HRLC10 product is available in two formats: i) GeoTIFF tiles (as specified in the naming) following the Sentinel 2 MGRS tiling scheme, and ii) NetCDF mosaics. The tile products use as Coordinate Reference System (CRS) the Universal Transverse Mercator (UTM) map projection system based on the World Geodetic System 84 (WGS84) reference ellipsoid. The mosaic products use as CRS a geographic coordinate system (i.e., latitude and longitude) based on the WGS84 reference ellipsoid. Note that UTM provides a constant distance relationship anywhere on the map. Instead, in angular coordinate systems like latitude and longitude, the distance covered by a degree of longitude differs as you move towards the poles and only equals the distance covered by a degree of latitude at the equator.

The HRLC10 product is illustrated in Figure 3, where the full mosaics over the three selected sub-continental areas are illustrated, and in Figure 4, where some sample tiles (MGRS scheme, 100km x 100km) are presented on some key areas within the static regions.

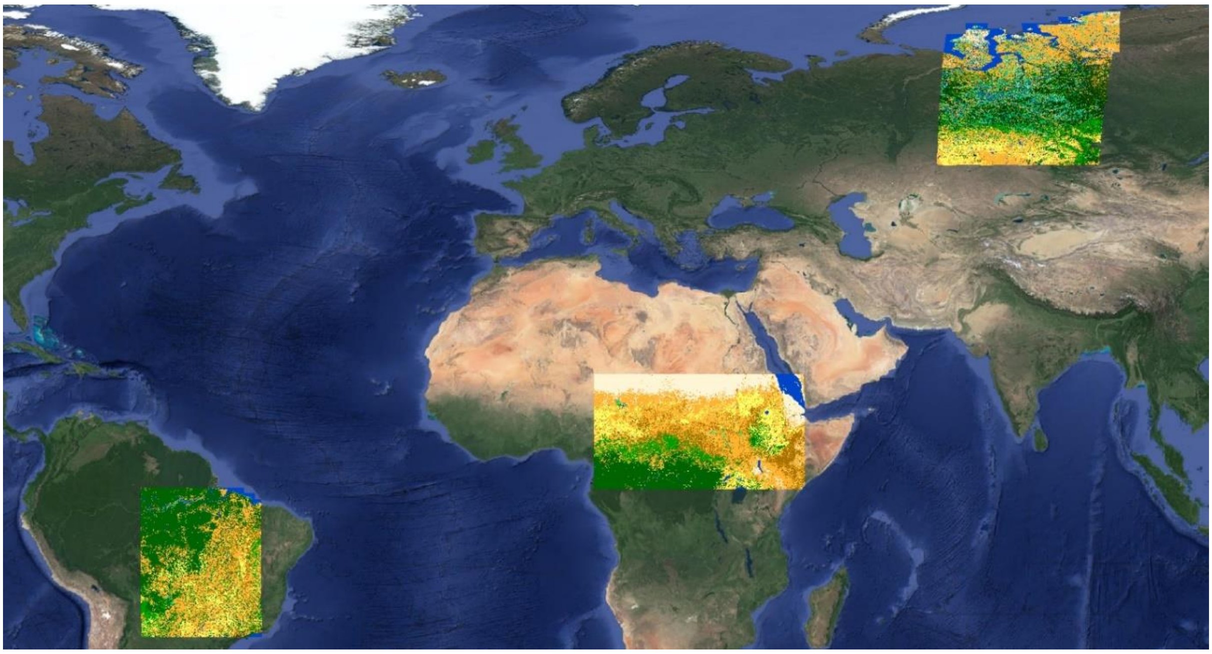
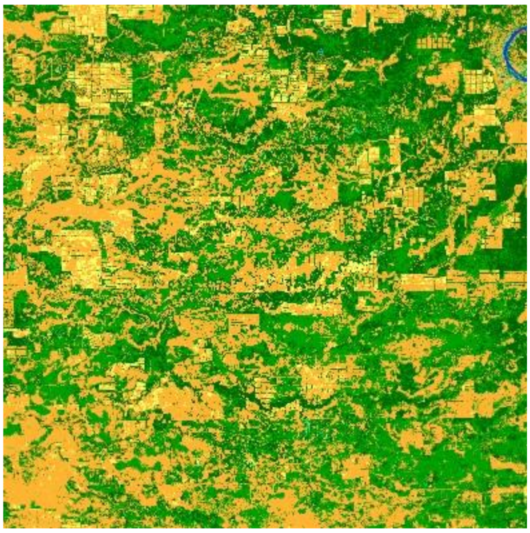
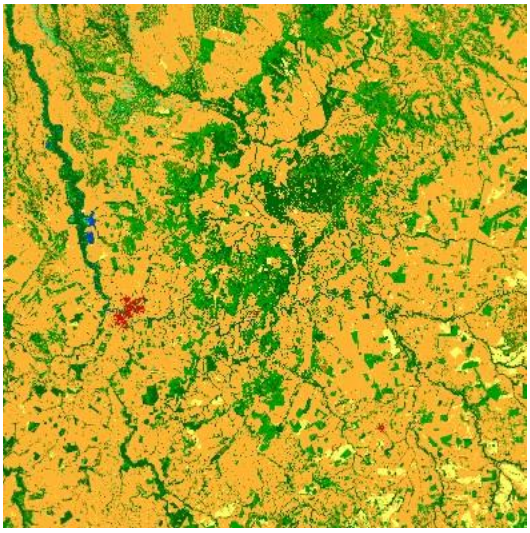


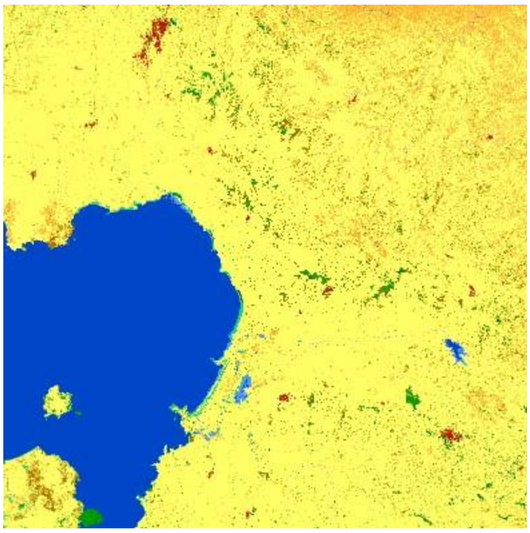
Figure 3. HRLC10 maps on the three sub-continental areas within Amazon, Africa and Siberia.



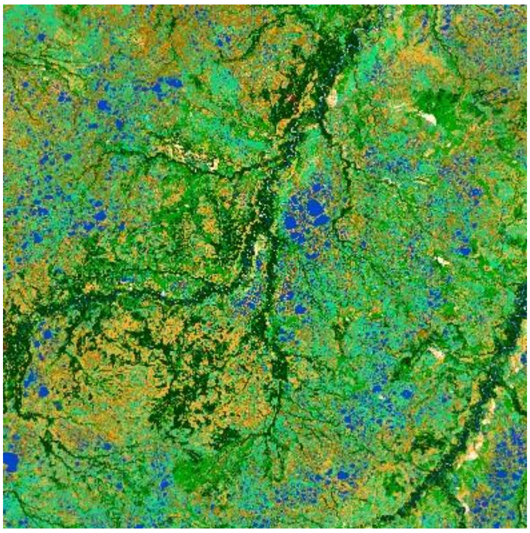
(a) 21KUQ – Amazon



(b) 21KXT – Amazon



(c) 37PCP – Africa



(d) 42WXS – Siberia

Figure 4. HRLC10 on few MGRS tiles. (a) and (b) are in Amazonia, (c) in Africa and (d) in Siberia.

4.2 Processing chain

The high-level workflow of the processing chain is presented in Figure 5. Sentinel-2 optical multispectral imagery is the main source of data as input for the classification. The SAR processing chain is implemented for Sentinel-1. The products obtained by the optical and the SAR processing chains are then integrated in the data fusion module in order to produce the final HRLC10 products. This design choice of fusion at the decision level makes it possible to develop advanced and ad hoc processing approaches for optical, SAR, and multisensor data, while keeping the system modular and scalable.

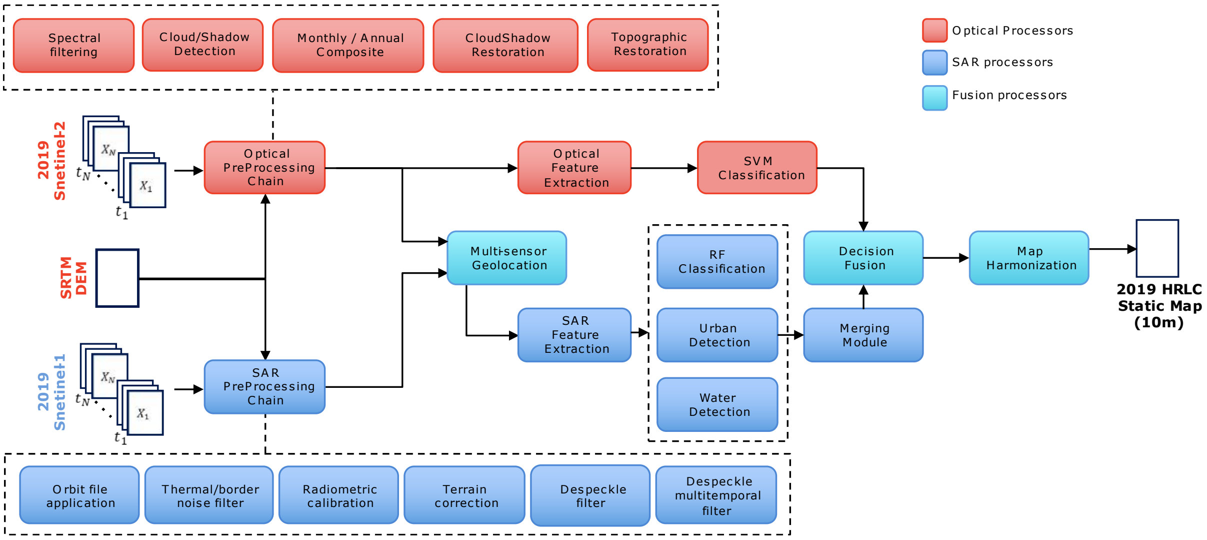


Figure 5. Block-based representation of the processing chain for the production of static HRLC10 maps.

The processing chain is described in greater detail in ATBD [AD7] and PVASR [AD6].

4.3 Uncertainty products description

Many steps of the processing chain (e.g., pre-processing, geolocation, etc.) involve well-known best-performing algorithms that come with uncertainty models associated with them. For instance, the classification task is able to output probabilistic posteriors that can be managed at the fusion level to infer uncertainty score pixel-wise. Both uncertainties of input data sets and processing model-related ones must be considered, including error propagation dynamics. The CCI HRLC project provides the climate community with uncertainty measures for each pixel in the HRLC10 products. As the first indicator of the uncertainty, together with the output LC map (corresponding to the most probable class on each individual pixel), also the second most probable one is considered on a pixelwise basis. Moreover, the posterior probability values for these best and second best classes are also provided together with the corresponding classification maps. This allows to greatly reduce the data size for the uncertainty while keeping the information that has been deemed more relevant by the Climate Group. In fact, that formulation shows among which classes the major uncertainty is and the relevance of the possible doubt (higher when the posterior for these two classes are similar and lower as their difference increases). In addition, the optical input quality index is included in the uncertainty output, as the optical data is the main source of information and its quality directly affects the classifier performance. Therefore, the output for uncertainty includes:

- Classification maps corresponding to best and second-best thematic classes (CL01 and CL02 as product id in the file naming);
- Posterior probabilities corresponding to best and second-best thematic classes (PS01 and PS02 as product id in the file naming).
- Input quality index corresponding to the optical input quality, related to the number and temporal distribution of the optical acquisitions used in the composite generation step (IQIX as product id in the file naming).

See ATBD [AD7] and E3UB [AD8] for further details.

4.4 Format, metadata and projection

Product specifications are compliant with the last update of CCI Data Standards Requirements [AD3]. In particular, this ensures complete interoperability of HRLC10 products with other products delivered in the framework of CCI. The metadata are stored in the NetCDF/GeoTIFF as specified in Section 3.3 and the format as specified in Section 3.2. Table 6 presents detailed information of pixel data available in each different product of HRLC10.

Table 6. Raster pixel values description for each product of HRLC10.

product type	product id	Description	NetCDF DataType	GeoTIFF DataType	Values Range
MAP	CL01	First class as returned by the classification algorithm according to posterior probability ranking. Class values are those described in Figure 2.	NC_BYTE	Byte	0 - 150
UNCERT	CL02	Second class as returned by the classification algorithm according to posterior probability ranking. Class values are those described in Figure 2.	-	Byte	0 - 150
UNCERT	PS01	Posterior probability associated to the first ranked class. Original probability values in the range [0,1] are mapped to 0-100 intgere values.	-	Byte	0 - 100
UNCERT	PS02	Posterior probability associated to the second ranked class. Original probability values in the range [0,1] are mapped to 0-100 intgere values.	-	Byte	0 - 100
UNCERT	IQIX	Optical imagery input quality evaluated before classification by considering the number of valid images at pixel level used for classification. Ranges from 0–low quality to 3–high quality.	-	Byte	0 - 3

The reference projection of the products are based on the original projection coming from Sentinel-2 (UTM local projections) data in order to allow the overlay with original Sentinel-2 images without warping. The final delivery is done in tiles (as specified in the naming) following the Sentinel 2 MGRS tiling scheme.

The mosaicked products of the full study areas are delivered in WGS84 Lat-Lon fulfilling the requirements coming from Climate modelers community.

4.4.1 Naming detailed structure

The files for each product type are named according to the CCI Data Standards:

ESACCI-<CCI Project>-<Processing Level>-<Data Type>-<Product String>[-<Additional Segregator>] <IndicativeDate>[<Indicative Time>]-fv<File version>.tif
<CCI Project> **ESACCI- HRLC** <Processing Level> **L4**

<Data Type>

MAP - UNCERT

<Product String>

CL01 – CL02 – PS01 – PS02 - IQIX

<Additional Segregator Tile>

Area code A01=Africa, A02=Amazon, A03=Siberia (3 chars) concatenated with the tile code following Sentinel-2 tiling system based on MGRS (6 chars).

<Additional Segregator Resolution>

10m

<Indicative Date>

The identifying date for this data set is YYYY, where YYYY is the four digit year. As prefix, the indication of the period is given as specified in Data Standard document with ISO notation.

<File Version>

File version number in the form n{1,}[.n{1,}]. This is 1 or more digits followed by optional . and another 1 or more digits.

Examples:

- ESACCI-HRLC-L4-MAP-CL01-A01MOSAIC-10m-P1Y-2019-fv01.0.tif
- ESACCI-HRLC-L4-MAP-CL01-A01T32NPF-10m-P1Y-2019-fv01.0.tif

4.5 Independent validation

All three HRLC10 products have subcontinental extent, which at the HR scale of 10m is measurable in the order Gpixels. This is better quantified in number of tiles (each of size 100 km x 100 km as in the MGRS scheme) and pixels in Table 7. To address the complexity of a proper validation on the large size of the data, two independent validation methodologies have been used: i) large extent validation on the basis of labeled samples extracted from the agreement of available official HRLC products; and ii) validation on a reduced set of samples extracted by photo-interpretation. The inter-comparison activity consisted of providing a very significant number (in the order of billions) of "reliable" pixels obtained through systematic intersection and analysis of several available HR LC products. The photo-interpretation activity focused on providing a much smaller number (in the order of thousands) of labeled samples obtained through manual interpretation by experts. Advantage in the former case is the huge number of samples considered, the disadvantage is related to the possibility of coherent misclassification of all the land cover maps used for producing the agreement map. The main advantage of the latter is that it allows to analyze seasonal classes (not present in the products used for inter-comparison); the disadvantage is the very small number of labels that results in high uncertainty on the considered area.

Table 7. Extent of the three HRLC10 products and number of benchmarking and validation samples used for reporting products accuracies.

	Product extent		Intercomparison agreement	Photo-interpretation
Area	**N. of tiles**	**N. of pixels**	**N. of pixels**	**N. of pixels**
Amazon	568	~53 Gpixels	~35 Gpixels	857
Africa	768	~72 Gpixels	~36 Gpixels	748

Siberia	392	~33 Gpixels	~21 Gpixels	549
TOTAL	**1728**	**~158 Gpixels**	**~92 Gpixels**	**2154**

Given the importance of all classes for climatological modelling, in the following tables the statistics per class are provided.

4.5.1 Validation by inter-comparison with existing products

As a result of the class aggregation process needed to intersect the existing HRLC products for benchmarking, a total of eight classes are finally obtained for the map of agreement: Bareland, Built-up, Cropland, Forest, Grassland, Shrubland, Water, Wetland. More details on this and on how HRLC10 classes are linked to these aggregated classes can be found in PVP [AD9], Section 6.1.3.

In the following, for each of the three HRLC10 products, number of benchmarking pixels and accuracy metrics obtained by comparison of the products with their corresponding maps of agreement are given. Reported accuracy metrics User and Producer accuracies for the benchmarking case are notated with a "b" subscript to distinguish them from those reported in the validation section: UA_b, PA_b.

4.5.2 Validation by photo-interpretation

Validation samples were randomly selected at the locations of the TREEs dataset used in the CCI MRLC and GlobCover projects. The sample original labels served as a first indication to the photo-interpreters who could inherit from the MR project hints about the regional LC interpretation of the landscape. Samples are then re-labeled with a HRLC class according to the process described in the PVP [AD9]. Not all the located samples have been considered reliable, a first set of 316, 638, and 560 systematic random samples for Amazonia, Africa and Siberia were considered certain and part of a homogeneous landscape. Initially, the low sample representation in the first set for some of the classes resulted in large confidence intervals around the precision estimates. To increase such precision, a sample numerosity augmentation process was carried out using posterior estimates of the class proportions derived from the HRLC10 maps themselves. As a result, 638, 560 and 347 further samples for Amazon, Africa and Siberia respectively, were included in the validation sample set. A final selection of 857, 748 and 549 samples, from the complete set formed by the first set and the augmented set, was done to perform accuracy estimations. The ratio of retained certain and homogeneous samples to the total number of samples available in the pool formed by the first set and the augmented set is 74%, 78% and 83% for Amazon, Africa, and Siberia respectively. The above-mentioned process of selecting certain and homogeneous samples implies that the accuracy estimates presented hereafter are marginal with respect to a homogeneous and uniform sampling of the classified area.

In order to provide useful information for a wide range of HRLC10 users, accuracy results are presented class wise for both the original HRLC legend and also with respect to a further class aggregation where "Shrub" and "Tree" classes are grouped into a "Woody vegetation" class and where the "Woody aquatic vegetation" and "Grassland aquatic vegetation" are grouped into an "Aquatic or regularly flooded vegetation" class. The first grouping is justified by the current absence of shrub PFT in the ORCHIDEE scheme, thus useful for climate users. The last grouping is to highlight the ability of the maps to detect the wetness of a landscape rather than the vegetation type. Reported accuracy metrics User and Producer accuracies for validation are notated as: UA, PA, and they are calculated via area-based methodology, as described in PVP [AD9].

4.5.3 Amazon product

In Amazon seven existing HRLC products were used to derive the map of agreement: FROM GLC, GL30, MapBiomas, GHS BU S1, WSF, FNF and GSW seasonality. Three of them with generic LC classes, two focused only on the Built-up class, one focused mainly on the Forest class but also includes Water, and one focused on Water.

In Table 8 below the accuracy metrics of the HRLC10 product classes as compared with the map of agreement is reported.

Table 8. Amazon area: accuracy metrics for the HRLC10 product as compared to the map of agreement. It includes the total number of benchmarking pixels used for each class.

Label	UA_b	PA_b	# of samples
Forest	97%	99%	22 075 650 573
Shrubland	92%	38%	3 296 020 597
Grassland	68%	92%	4 605 613 119
Cropland	93%	93%	3 978 385 040
Wetland	29%	40%	56 213 614
Bareland	3%	74%	2 263 169
Built-up	65%	95%	50 210 662
Water	100%	96%	1 361 349 442
TOTAL			35 425 706 216

Accuracy figures obtained by comparison of the HRLC10 product with photo-interpreted validation samples are given in Table 9 for both the case of the grouped legend and the case of the native HRLC legend.

Table 9. Amazon area: class wise validation accuracy metrics for the HRLC10 product with the grouped and native legend as compared with validation samples. Accuracy figures are reported with 95% confidence interval for error margins.

Aggregated Legend			Native Legend			
Label	**UA (%)**	**PA (%)**	**Label**	**UA (%)**	**PA (%)**	**# of samples**
Woody Vegetation	94.5 ± 2.4	82.6 ± 3.3	Tree cover evergreen broadleaf	77.6 ± 5.3	89.0 ± 2.9	241
			Tree cover deciduous broadleaf	50.9 ± 13.3	43.9 ± 10.6	60
			Shrub cover evergreen	26.8 ± 13.7	4.5 ± 2.4	82
			Shrub cover deciduous	6.7 ± 9.0	4.0 ± 5.5	29
Grassland	61.3 ± 7.4	90.2 ± 3.9	Grassland	61.3 ± 7.4	91.6 ± 3.3	132
Croplands	81.4 ± 7.6	75.7 ± 9.6	Croplands	81.4 ± 7.6	76.1 ± 9.6	100
Vegetation aquatic or regularly flooded	50.0 ± 12.9	10.7 ± 5.3	Woody vegetation aquatic or regularly flooded	42.4 ± 17.1	3.1 ± 2.2	25
			Grassland vegetation aquatic or regularly flooded	28.0 ± 18.0	10.6 ± 9.4	22
Bare areas	66.7 ± 17.2	56.8 ± 39.2	Bare areas	66.7 ± 17.2	56.8 ± 39.2	24
Built-up	85.7 ± 10.8	100 ± 0.0	Built-up	85.7 ± 10.8	100 ± 0.0	36
Open Water seasonal	8.3 ± 11.4	50.8 ± 59.6	Open Water seasonal	8.3 ± 11.4	73.4 ± 46.5	3
Open Water permanent	97.1 ± 3.9	90.8 ± 2.4	Open Water permanent	97.1 ± 3.9	88.9 ± 2.9	103

4.5.4 Africa product

In Africa seven existing HRLC products were used to derive the map of agreement: FROM GLC, GL30, CCI Prototype, GHS BU S1, WSF, FNF and GSW seasonality. Three of them with generic LC classes, two focused only on the Built-up class, one focused mainly on the Forest class but also includes Water, and one focused on Water. In Table 10 below the accuracy metrics of the HRLC10 product classes as compared with the map of agreement is reported.

Table 10. Africa area: class wise accuracy metrics for the HRLC10 product as compared to the map of agreement. It includes the total number of benchmarking pixels used for each class.

Label	UA_b	PA_b	# of samples
Forest	100%	99%	12 149 855 046
Shrubland	73%	26%	1 416 129 826
Grassland	67%	84%	4 193 468 040
Cropland	92%	85%	3 710 989 939
Wetland	12%	71%	4 843 131
Bareland	97%	98%	12 252 823 426
Built-up	42%	93%	7 232 581
Water	100%	100%	1 910 729 003
TOTAL			35 646 070 992

Accuracy figures obtained by comparison of the HRLC10 product with photo-interpreted validation samples are given in Table 11 for both the case of the grouped legend and the case of the native HRLC legend.

Table 11. Africa area: class wise validation accuracy metrics for the HRLC10 product with the grouped and native legend as compared with validation samples. Accuracy figures are reported with 95% confidence interval for error margins.

Aggregated Legend			Native Legend			
Label	**UA (%)**	**PA (%)**	**Label**	**UA (%)**	**PA (%)**	**# of samples**
Woody Vegetation	84.2 ± 5.1	78.2 ± 4.7	Tree cover evergreen broadleaf	96.2 ± 4.3	92.7 ± 4.9	88
			Tree cover deciduous broadleaf	43.2 ± 14.9	40.9 ± 12.3	43
			Shrub cover evergreen	56.5 ± 20.8	0.2 ± 0.0	32
			Shrub cover deciduous	56.1 ± 12.9	60.1 ± 10.8	56
Grassland	44.8 ± 7.3	70.8 ± 7.3	Grassland	44.8 ± 7.3	73.2 ± 7.3	126
Croplands	63.8 ± 11.4	66.8 ± 9.8	Croplands	63.8 ± 11.4	67.7 ± 9.8	68
Vegetation aquatic or regularly flooded	58.7 ± 14.3	7.7 ± 4.1	Grassland vegetation aquatic or regularly flooded	58.7 ± 14.3	7.6 ± 4.1	42
Bare areas	94.9 ± 4.9	70.9 ± 5.9	Bare areas	94.9 ± 4.9	70.9 ± 5.9	118
Built-up	88.9 ± 10.4	15.9 ± 13.3	Built-up	88.9 ± 10.4	15.9 ± 13.3	36
Open Water seasonal	10.0 ± 8.4	1.0 ± 1.2	Open Water seasonal	10.0 ± 8.4	1.0 ± 1.2	19
Open Water permanent	86.7 ± 7.3	77.9 ± 16.9	Open Water permanent	86.7 ± 7.3	77.9 ± 16.9	120

4.5.5 Siberia product

In Siberia seven existing HRLC products were used to derive the map of agreement: FROM GLC, GL30, ESA DUE GlobPermafrost, GHS BU S1, WSF, FNF and GSW seasonality. Three of them with generic LC classes, two focused only on the Built-up class, one focused mainly on the Forest class but also includes Water, and one focused on Water. In Table 12 below the accuracy metrics of the HRLC10 product classes as compared with the map of agreement is reported.

Table 12. Siberia area: class wise accuracy metrics for the HRLC10 product as compared to the map of agreement. It includes the total number of benchmarking pixels used for each class.

Label	UA_b	PA_b	# of samples
Forest	98%	98%	10 634 989 050
Shrubland	0%	11%	1 503 113
Grassland	87%	81%	3 377 418 725
Cropland	89%	90%	2 496 718 168
Wetland	43%	56%	347 911 760
Bareland	58%	83%	46 129 964
Built-up	27%	94%	5 087 773
Water	98%	99%	3 962 130 715
Perm. snow and/or ice	0%	0%	357 212
TOTAL			20 872 246 480

In Siberia, the distinction between the leaf seasonality of evergreen and deciduous shrubs could not be photo interpreted with certainty. The two types of shrubs were merged into a single "Shrub Cover" class for validation. Accuracy figures obtained by comparison of the HRLC10 product with photo-intrerpreted validation samples are given in Table 13 for both the case of the grouped legend and the case of the native HRLC legend.

Table 13. Siberia area: class wise validation accuracy metrics for the HRLC10 product with the grouped and native legend as compared with validation samples. Accuracy figures are reported with 95% confidence interval for error margins.

Aggregated Legend			Native Legend			
Label	**UA (%)**	**PA (%)**	**Label**	**UA (%)**	**PA (%)**	**# of samples**
Woody Vegetation	88.7 ± 4.7	76 ± 4.5	Tree cover evergreen needleleaf	68.5 ± 10.8	72.2 ± 8.2	72
			Tree cover deciduous broadleaf	54.4 ± 12.0	76.0 ± 11.6	48
			Tree cover deciduous needleleaf	42.9 ± 26.9	23.6 ± 13.3	27
			Shrub cover*	59.1 ± 21.0	13.8 ± 5.3	77
Grassland	55.6 ± 9.8	69.3 ± 7.1	Grassland	55.6 ± 9.8	70.5 ± 6.9	110
Croplands	81.2 ± 9.6	72.7 ± 12.7	Croplands	81.2 ± 9.6	72.8 ± 12.7	61
Vegetation aquatic or regularly flooded	20.6 ± 9.6	51.8 ± 20.4	Woody vegetation aquatic or regularly flooded	2.9 ± 5.7	18.8 ± 34.7	4
			Grassland vegetation aquatic or regularly flooded	15.2 ± 12.3	23.5 ± 17.4	19

Lichens and mosses	48.5 ± 17.2	22.2 ± 12.5	Lichens and mosses	48.5 ± 17.2	22.2 ± 12.5	29
Bare areas	34.2 ± 15.3	12.9 ± 8.0	Bare areas	34.2 ± 15.3	12.9 ± 8.0	26
Built-up	85.7 ± 15.3	13.2 ± 10.4	Built-up	85.7 ± 15.3	13.2 ± 10.4	25
Open Water	95.9 ± 5.7	93.1 ± 7.3	Open Water	95.9 ± 5.7	93.2 ± 7.3	51

* Both Shrub cover types (i.e., deciduous and evergreen) are mapped, however the validation did not consider the seasonality of the Shrub cover.

5 HRLC30: High Resolution Land Cover Map at 30m products

5.1 Product description

The CCI HRLC project provides the climate community consistent land-cover maps at the resolution of 30m (HRLC30) on three regional areas (subsets of the sub-continental regions of the HRLC10 products, see Figure 1) every five years: 1990, 1995, 2000, 2005, 2010, 2015, 2019. The 5-year period is due to the necessity of data availability for historical optical and SAR data. The main sources of information for historical mapping are Landsat-5/7/8 for optical data, and ERS-1, ERS-2, ENVISAT-ASAR for SAR imagery. The legend of the classification map is presented in Section 3.4. The HRLC30 product is available in two formats: i) GeoTIFF tiles (as specified in the naming) following the Sentinel 2 MGRS tiling scheme, and ii) NetCDF mosaics. The tile products use as Coordinate Reference System (CRS) the Universal Transverse Mercator (UTM) map projection system based on the World Geodetic System 84 (WGS84) reference ellipsoid. The mosaic products use as CRS a geographic coordinate system (i.e., latitude and longitude) based on the WGS84 reference ellipsoid. Note that UTM provides a constant distance relationship anywhere on the map. Instead, in angular coordinate systems like latitude and longitude, the distance covered by a degree of longitude differs as you move towards the poles and only equals the distance covered by a degree of latitude at the equator.

The HRLC30 product is illustrated in Figure 6, where the full mosaics over the three selected regional areas are illustrated, and in Figure 7, where some sample tiles (MGRS scheme, 100km x 100km) are presented in a key area of Paraguay that shows and incremental growth of cropping activity from 1990 to 2019.

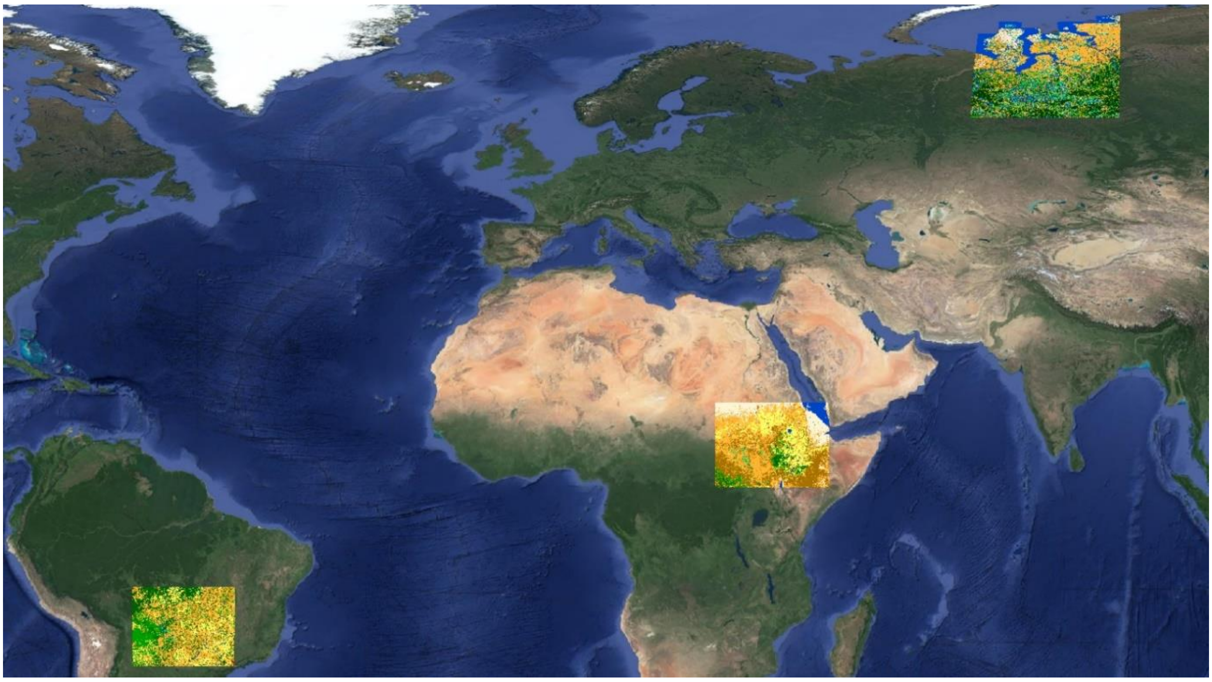
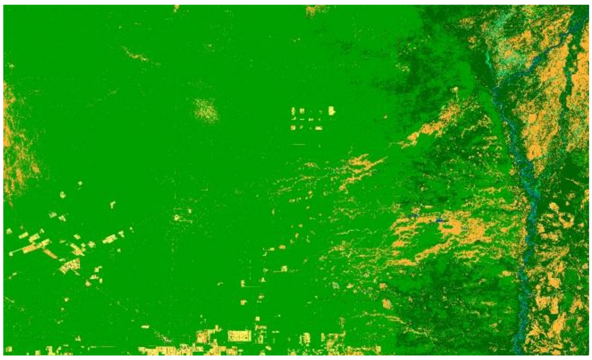
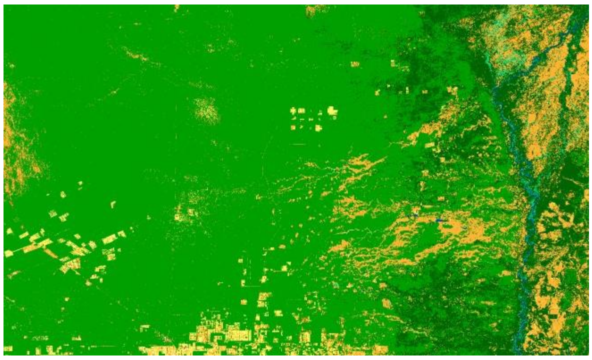


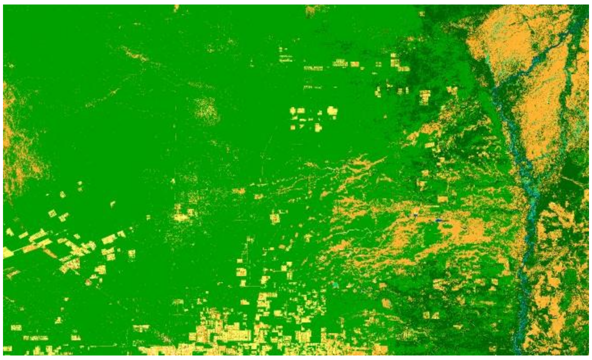
Figure 6. HRLC30 maps (year 2019) on the three sub-continental areas within Amazon, Africa and Siberia.



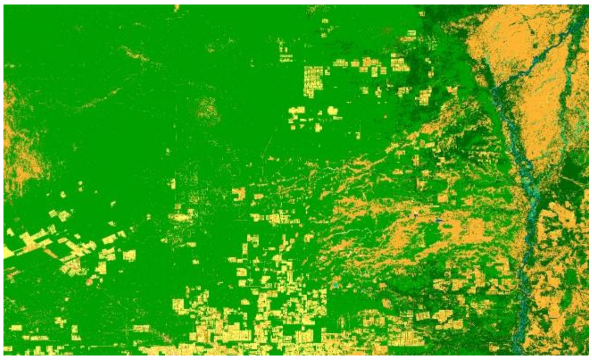
(a) 1990



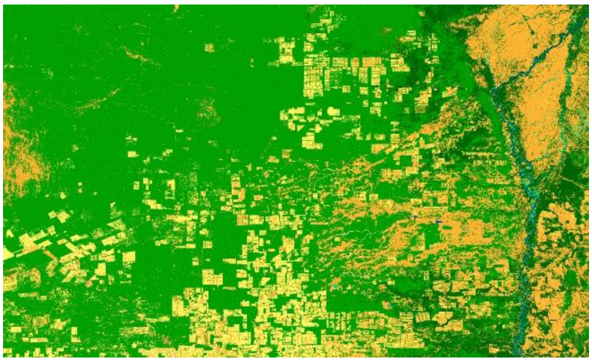
(b) 1995



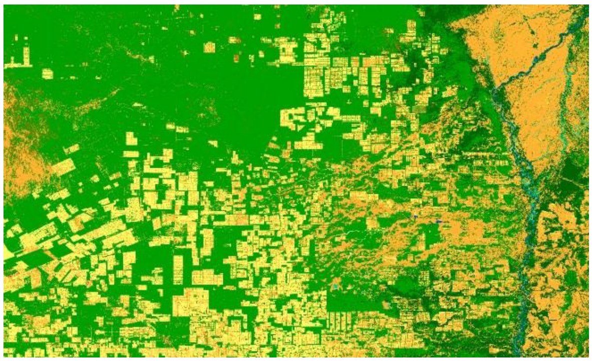
(c) 2000



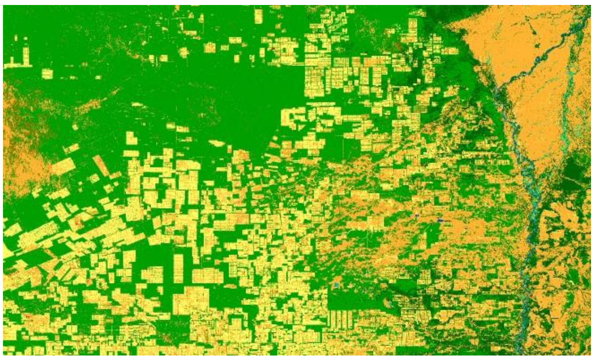
(d) 2005



(e) 2010



(f) 2015



(g) 2019

Figure 7. Illustration of a sequence of the HRLC30 land cover maps for years 1990, 1995, 2000, 2005, 2010, 2015 and 2019 for a large area in Paraguay.

5.2 Processing chain

The high-level workflow of the processing chain is presented in Figure 8. Optical multispectral imagery is the main source of data as input for the classification. The optical processing chain is consistent with the possibility to work mainly with images at 10/30m resolution and generating an output at 10/30m, based on multitemporal multispectral data from Landsat-8 in the recent years and legacy Landsat-5/7/8 data in the past. The SAR processing chain is implemented for ERS and ASAR data sets (whenever and wherever HR mode data are available). The products obtained by the optical and the SAR processing chains are then integrated in the data fusion module in order to produce the final HRLC30 products. This step also includes a multitemporal cascade model that exploits a Bayesian approach to ensure temporal consistency with the HRLC10 product. The output products are then analyzed in the multitemporal change detection and trend analysis block for identifying different change components to be used for the historical time series HRLC30 products every 5 years.

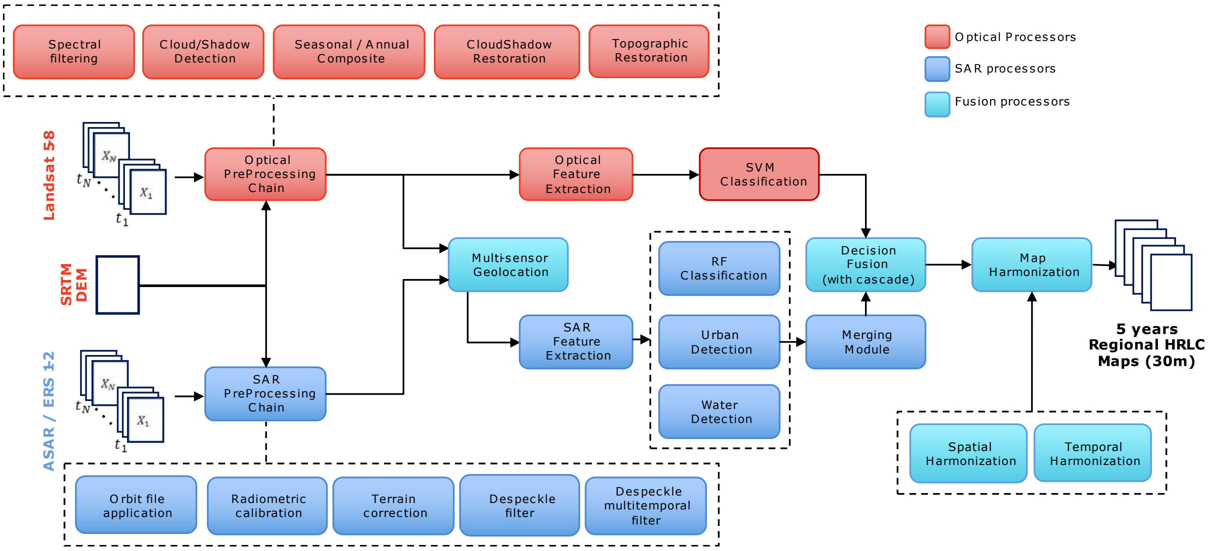


Figure 8. Block-based representation of the processing chain for the production of historical HRLC30 maps.

The processing chain is described in greater detail in ATBD [AD7] and PVASR [AD6].

5.3 Uncertainty products description

Many steps of the processing chain (e.g., pre-processing, geolocation, etc.) involve well-known best-performing algorithms that come with uncertainty models associated to them. For instance, the classification task is able to output probabilistic posteriors that can be managed at the fusion level to infer uncertainty score pixel-wise. Both uncertainties of input data sets and processing model-related ones must be considered, including error propagation dynamics. The CCI HRLC project provides the climate community with uncertainty measures for each pixel in the HRLC10 products. As the first indicator of the uncertainty, together with the output LC map (corresponding to the most probable class on each individual pixel), also the second most probable one is considered on a pixelwise basis. Moreover, the posterior probability values for these best and second best classes are also provided together with the corresponding classification maps. This allows to greatly reduce the data size for the uncertainty while keeping the information that has been deemed more relevant by the Climate Group. In fact, that formulation shows among which classes the major uncertainty is and the relevance of the possible doubt (higher when the posterior for these two classes are similar and lower as their difference increases). In addition, the optical input quality index is included in the uncertainty output, as the optical data is the main source of information, and its quality directly affects the classifier performance. Therefore, the output for uncertainty includes:

- Classification maps corresponding to best and second-best thematic classes (CL01 and CL02 as product id in the file naming);

- Posterior probabilities corresponding to best and second-best thematic classes (PS01 and PS02 as product id in the file naming).
- Input quality index corresponding to the optical input quality, related to the number and temporal distribution of the optical acquisitions used in the composite generation step (IQIX as product id in the file naming).

See ATBD [AD7] and E3UB [AD8] for further details.

5.4 Format, metadata and projection

Product specifications are compliant with last update of CCI Data Standards Requirements [AD3]. In particular, this ensures complete interoperability of HRLC30 products with other products delivered in the framework of CCI. The metadata are stored in the NetCDF/GeoTIFF as specified in Section 3.3 and the format as specified in Section 3.2. Table 14 presents detailed information of pixel data available in each different product of HRLC30.

Table 14. Raster pixel values description for each product of HRLC30.

product type	product id	Description	NetCDF DataType	GeoTIFF DataType	Values Range
MAP	CL01	First class as returned by the classification algorithm according to posterior probability ranking. Class values are those described in Figure 2.	NC_BYTE	Byte	0 - 150
UNCERT	CL02	Second class as returned by the classification algorithm according to posterior probability ranking. Class values are those described in Figure 2.	-	Byte	0 - 150
UNCERT	PS01	Posterior probability associated to the first ranked class. Original probability values in the range [0,1] are mapped to 0-100 intgere values.	-	Byte	0 - 100
UNCERT	PS02	Posterior probability associated to the second ranked class. Original probability values in the range [0,1] are mapped to 0-100 intgere values.	-	Byte	0 - 100
UNCERT	IQIX	Optical imagery input quality evaluated before classification by considering the number of valid images at pixel level used for classification. Ranges from 0–low quality to 3–high quality.	-	Byte	0 - 3

The reference projection of the products is based on the projection coming from Sentinel-2 (UTM local projections) data in order to allow the overlay with other HRLC products without warping. The final delivery is done in tiles (as specified in the naming) following the Sentinel-2 MGRS tiling scheme. The mosaicked products of the full study areas are delivered in WGS84 Lat-Lon fullfilling the requirements coming from Climate modelers community.

5.4.1 Naming detailed structure

The files for each product type are named according to the CCI Data Standards:

ESACCI-<CCI Project>-<Processing Level>-<Data Type>-<Product String>[-<Additional Segregator>] <IndicativeDate>[<Indicative Time>]-fv<File version>.tif
<CCI Project> **ESACCI- HRLC** <Processing Level> **L4** <Data Type> **MAP - UNCERT** <Product String> **CL01 – CL02 – PS01 – PS02 - IQIX** <Additional Segregator Tile> **Area code A01=Africa, A02=Amazon, A03=Siberia (3 chars) concatenated with the tile code following Sentinel-2 tiling system based on MGRS (6 chars).** <Additional Segregator Resolution> **30m** <Indicative Date> **The identifying date for this data set is YYYY, where YYYY is the four digit year. As prefix, the indication of the period is given as specified in Data Standard document with ISO notation.** <File Version> **File version number in the form n{1,}[.n{1,}]. This is 1 or more digits followed by optional . and another 1 or more digits.**

Examples:

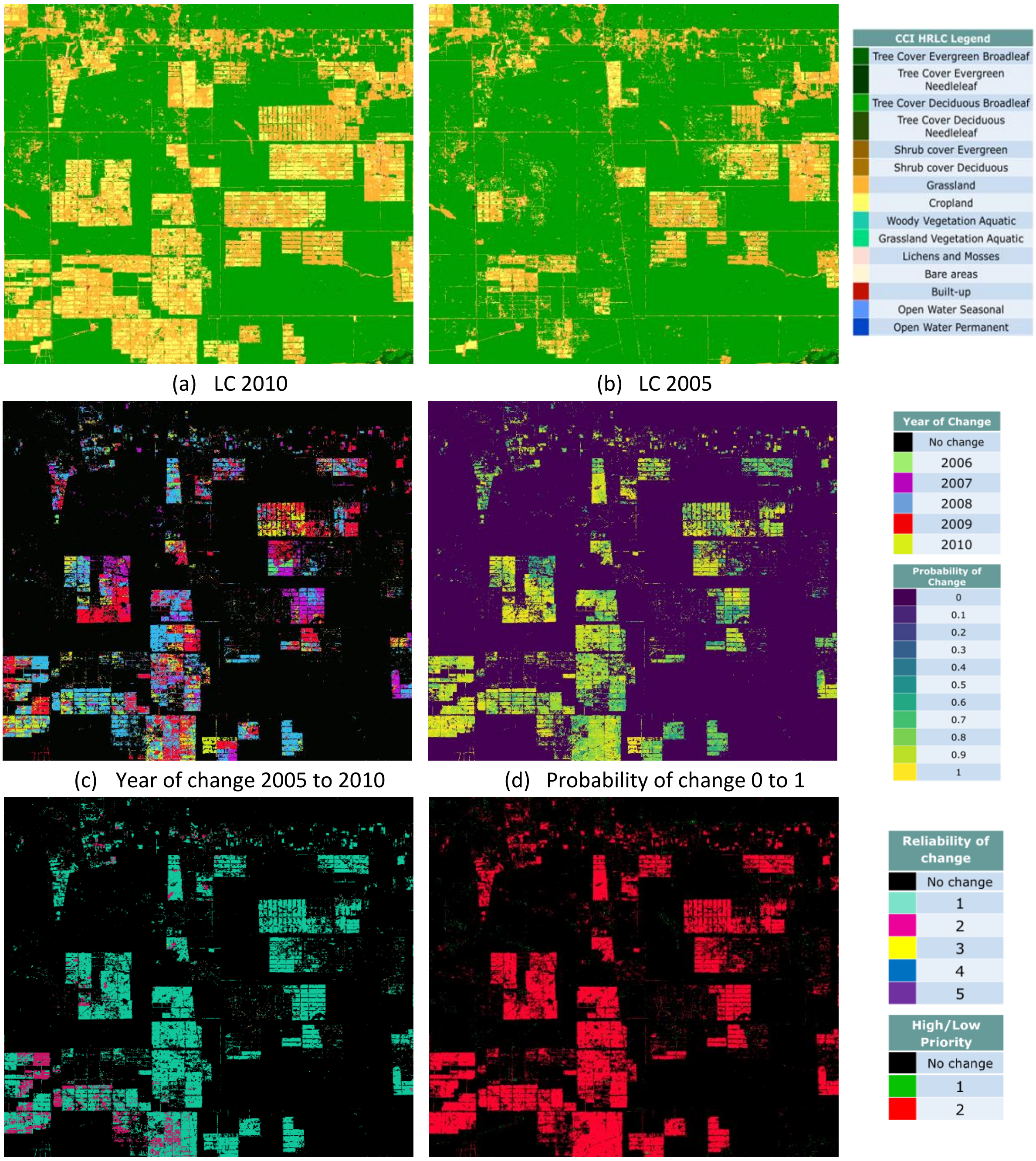
- ESACCI-HRLC-L4-MAP-CL01-A03MOSAIC-30m-P5Y-2010-fv01.0.tif
- ESACCI-HRLC-L4-MAP-CL01-A02T20LPJ-30m-P5Y-1995-fv01.0.tif

6 HRLC30: High Resolution Land Cover Change Map at 30m products

6.1 Product description

The CCI HRLC project provides the climate community HRLCC30: a consistent land-cover change detection maps at 30m resolution on the three regional areas of the HRLC30 products (subsets of the sub-continental regions of the HRLC10 products, see Figure 1). The change detection products are obtained through massive processing of Landsat-5/7/8 acquisitions from 1990-2019 for each five years temporal periods (1990-1995, 1995-2000, 2000-2005, 2005-2010, 2010-2015, 2015-2019) and they contain, within each 5-year period, the yearly change information. Considering the availability of the data especially for the earlier years, in some cases it is not possible to report the changes between two consecutive years. As a result, additional information is provided that indicates the distance of the years among which the change was detected. Moreover, following prioritization suitable for Climate analysis (see URD [AD4], CAR [AD13] and ATBD [AD7]), the change information is provided only for high priority transitions, therefore a map of no-change and high/low priority change pixels is provided as well to complement the pixel information.

The HRLCC30 product is available in two formats: i) GeoTIFF tiles (as specified in the naming) following the Sentinel 2 MGRS tiling scheme, and ii) NetCDF mosaics. The tile products use as Coordinate Reference System (CRS) the Universal Transverse Mercator (UTM) map projection system based on the World Geodetic System 84 (WGS84) reference ellipsoid. The mosaic products use as CRS a geographic coordinate system (i.e., latitude and longitude) based on the WGS84 reference ellipsoid. Note that UTM provides a constant distance relationship anywhere on the map. Instead, in angular coordinate systems like latitude and longitude, the distance covered by a degree of longitude differs as you move towards the poles and only equals the distance covered by a degree of latitude at the equator. An example of HRLCC30 ensemble of products is given in Figure 9, where for the 5-year period 2005-2010, The HRLC30 maps, the year of change, the probability of change, the reliability of change and the high/low priority changes have been visualized for a sample area with crop fields, in Amazon.



(a) LC 2010

(b) LC 2005

(c) Year of change 2005 to 2010

(d) Probability of change 0 to 1

(e) Reliability of change 1 to 5

(f) High/low priority (PCC) 1 or 2

Figure 9. The HRLC30 maps and HRLCC30 change detection products in a sample area in Amazon. Reference period 2005 - 2010.

6.2 Processing chain

In Figure 10 the processing chain for the generation of HRLCC30 products is illustrated. The processing chain consists of three main steps: 1) feature extraction/reduction to extract relevant information of the spectral trends of different sets of LC changes and preserve the most informative ones, 2) time series reconstruction to produce a continue, reliable and dense Satellite Image Time Series (SITS) for change detection analysis (here the cloud/shadow mask and the Post Classification Comparison (PCC) mask have been imposed to the algorithm to remove cloudy pixels and select high priority changed pixels), 3) the abrupt change detection to detect the time of the change and probability of change by using Breaks For Additive Seasonal and Trend (BFAST). As an input of the last step, the features extracted from the feature extraction/reduction module are fused in the feature fusion step to reduce the computational burden of the abrupt change detection. The processing chain requires in input the entire optical data time series of Landsat 5/7/8 from 1990-2019 and the five years regional HRLC30 maps. As the output from the processing chain, there will be the change information at 30m in yearly time scale. Additional information on the processing chain is given in ATBD [AD7], including what concerns the selection of the best performing algorithms.

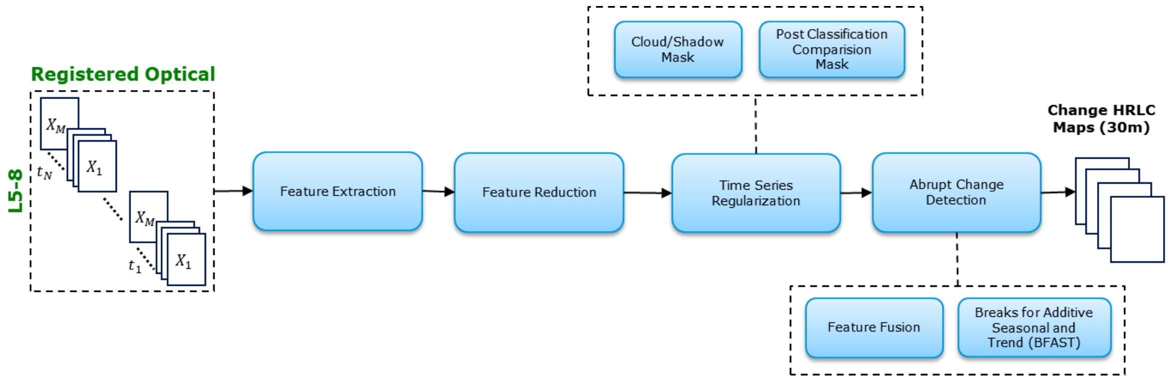


Figure 10. Detailed block-based representation of the multitemporal change detection processing chain.

The processing chain is described in greater detail in ATBD [AD7] and PVASR [AD6].

6.3 Uncertainty products description

The algorithms applied in the change detection processing chain have an impact on the propagation of uncertainty, from both input SITS and the HRLC30 products, to the final change detection products. Change detection and trend analysis are mainly sensitive to LC classification.

The probability in abrupt change detection can be defined as a certainty associated to the change detection processing chain. In detail, BFAST allows the computation of pixelwise certainty measures using the output of the breakpoints function for the trend model that represents the detected breaks in the trend component considering a segment size between potentially detected breaks. The output of the breakpoints confint function for the trend model can be retrieved as the confidence level associated to a change happening within a day from the breakpoint.

The reliability of the change is calculated using the feature magnitude between two consecutive years. If enough data (at least three acquisitions in consecutive months in a year) is not available for consecutive years, the algorithm checks the other years to find the closest year with sufficient acquisitions. As a result, the reliability is defined as a factor of certainty of the detected year of change. See ATBD [AD7] and E3UB [AD8] for further details.

6.4 Format, metadata and projection

Product specifications are compliant with last update of CCI Data Standards Requirements [AD3]. In particular, this ensures complete interoperability of HRLCC30 products with other products delivered in the framework of CCI. The metadata are stored in the NetCDF/GeoTIFF as specified in Section 3.3 and the format as specified in Section 3.2. Table 15 presents detailed information of pixel data available in each different product within HRLCC30. Note that each change detection product is a multi-layered file including 4 bands.

Table 15. Raster pixel values description for each product of HRLCC30. The change detection products are multi-layered files with 4 bands each.

product type	**product id**	*band*	**Description**	**NetCDF DataType**	**GeoTIFF DataType**	**Values Range**
CHANGE	CDET	1	YEAR OF CHANGE: the year of change in the 5-year span of the two years from which the LCC product is calulated.	NC_FLOAT	Float32	1990 - 2019 (NaN)
		2	PROBABILITY OF CHANGE: probaility of the change as returned by the change detection algorithm. Original probability values in the range [0,1] are mapped to 0-100 intgere values.	-	Float32	0 - 100 (NaN)
		3	RELIABILITY: distance between the couple of years for which the change has been calculated	-	Float32	0 - 5 (NaN)
		4	PCC PRIORITY CHANGES: from Post Classification Comparison of HRLC30 maps, the class cover transition of the change in its degree of priority. 0=no change, 1=low priority change, 2=high priority change.	-	Float32	0 - 2

The YEAR OF CHANGE is calculated when enough data is available (at least three acquisitions in consecutive months in a year) for each pixel in consecutive years. If enough data are not available for consecutive years, the algorithm keeps the first year (with enough data) and checks the number of acquisitions in the other years. This procedure (checking the availability of data for the next years) is continued to find a year with enough acquisitions in the whole six years of data. If the data availability is verified, the year of the change and the PROBABILITY OF CHANGE are reported on a multiannual span (instead of yearly). Thus, the value of RELIABILITY shows the distance between the couple of years that the change information is provided and it is a number between 1 – 5 for the changed pixels and 0 for no-change. If every year has not enough data (considering the limitations described in PVASR [AD6]) the change information is missing, reliability and probability are not calculated and NaN is assigned as a value to those parameters.

The year of the change is calculated only for the high priority changes (due to the computational burden) according to the LCC matrices of relevant changes for different areas [AD8]. The PCC PRIORITY CHANGES represents the low and high priority changes using values 1 and 2, respectively and 0 for no-change. The class to which the pixel has changed can be derived from the LC map using the legend values described in Figure 2.

The reference projection of the products is based on the projection coming from Sentinel-2 (UTM local projections) data in order to allow the overlay with other HRLC products without warping. The final delivery is done in tiles (as specified in the naming) following the Sentinel-2 MGRS tiling scheme. The mosaicked products of the full study areas are delivered in WGS84 Lat-Lon fulfilling the requirements coming from Climate modelers community.

6.4.1 Naming detailed structure

The files for each product type are named according to the CCI Data Standards:

ESACCI-<CCI Project>-<Processing Level>-<Data Type>-<Product String>[-<Additional Segregator>] <IndicativeDate>[<Indicative Time>]-fv<File version>.tif
<CCI Project> **ESACCI- HRLC** <Processing Level> **L4** <Data Type> **CHANGE** <Product String> **CDET** <Additional Segregator Tile> **Area code A01=Africa, A02=Amazon, A03=Siberia (3 chars) concatenated with the tile code following Sentinel-2 tiling system based on MGRS (6 chars).** <Additional Segregator Resolution> **30m** <Indicative Date> **The identifying date for this data set is YYYY, where YYYY is the four digit year. As prefix, the indication of the period is given as specified in Data Standard document with ISO notation.** <File Version> **File version number in the form n{1,}[.n{1,}]. This is 1 or more digits followed by optional . and another 1 or more digits.**

Examples:

- ESACCI-HRLC-L4-CHANGE-CDET-A01MOSAIC-30m-P1Y-2000-2005-fv01.0.tif
- ESACCI-HRLC-L4-CHANGE-CDET-A03T42VXM-30m-P1Y-2010-2015-fv01.0.tif

7 Software Tools

All the HRLC products will be available through the CCI Climate Open Data portal. Currently no visualization software is available. The link to the Delivery System will be made available to the Users. The Delivery System is meant to be interfaced by the CCI Toolbox in order to provide easy functionalities for retrieving.

As part of the Software delivered to the users, a reprojection tool is delivered to create mosaics from selected products and translate them from UTM to Lat-Long projection.

7.1 Reprojection tool

A reprojection tool has been developed in order to create a mosaic of the classification map produced in lat/lon coordinates in NetCDF format. Such tool is available here: https://lab.egeos-services.it/gitlab/fronci/cci-hrlc-reprojection-tool

Installation

- Clone the repository, or download the file "reprojectionenv.yml". That's all you need;
- Conda must be installed in your system: https://docs.conda.io/projects/conda/en/latest/user-guide/install/index.html;
- create the environment using the yml file provided in this repository through the command:

```
conda env create -f reprojectionenv.yml
```

- Activate the environment through the command:

```
conda activate reprojectionenv
```

Usage

- Once the environment is activated, launch the python interpreter through the command python and write the following commands:

```
import reprojection_and_mosaic
reprojection_and_mosaic.reproject_and_mosaic()
```

- The script will now ask to insert the path to the folder containing the images to be processed and an output folder.

Output

Once the script successfully finishes, output consists of each input image reprojected to lat/lon coordinates and a mosaic of the files in NetCDF format.