ECMWF COPERNICUS REPORT

Copernicus Climate Change Service

Algorithm Theoretical Basis Document

CDR and ICDR Sentinel-3 Land Cover (v2.1.1)

Issued by: UCLouvain/ Pierre Defourny

Date: 22/02/2024

Ref: WP2-FDDP-LC-2021-2022-SENTINEL3-300m-v2.1.1_ATBD_v1.2

Official reference number service contract: 2022/C3S2_312a_Lot5_VITO/SC1

PROGRAMME OF THE EUROPEAN UNION

Copernicus Europe's eyes on Earth

IMPLEMENTED BY ECMWF

Contributors

UCLOUVAIN
Pierre Defourny
Céline Lamarche

BROCKMANN CONSULT GMBH
Carsten Brockmann
Martin Boettcher
Grit Kirches

History of modifications

Version	Date	Description of modification	Chapters / Sections
V1.0	31/12/2022	First version	All
V1.1	19/06/2023	Document amended to account for feedback from independent reviewer and finalized for publication.	All
V1.2	22/02/2024	Update regarding the LC 2021 and 2022 data set	All

List of data sets covered by this document

Deliverable ID	Product title	Product type (CDR, ICDR)	Version number	Delivery date
WP2-FDDP-LC-2021-SENTINEL3-300m-v2.1.1	ICDR Land Cover 2021	ICDR	2.1.1	15/07/2023
WP2-FDDP-LC-2022-SENTINEL3-300m-v2.1.1	ICDR Land Cover 2022	ICDR	2.1.1	30/04/2024

Related documents

Reference ID	Document
RD-1	E.U. Copernicus Climate Change Service (2018) "C3S Product Quality Assurance Document - ICDR Land Cover", E.U. Copernicus Climate Change Service, Document ref.: v1.3.1, D2.2.11-v1.0_PQAD_ICDR_LC_v2.0.7cds_PRODUCTS_v1.0.1 https://datastore.copernicus-climate.eu/documents/satellite-land-cover/D2.2.11-v1.0_PQAD_CDR_LC-CCI_v2.0.7cds_Products_v1.0.1_APPROVED_Ver1.pdf
RD-2	E.U. Copernicus Climate Change Service (2018) " C3S Product User Guide and Specification - ICDR Land Cover", E.U. Copernicus Climate Change Service, Document ref.: V1.3.1, D3.3.11-v1.0_PUGS_ICDR_LC_v2.0.7cds_PRODUCTS_v1.0.1 (2018) https://datastore.copernicus-climate.eu/documents/satellite-land-cover/D3.3.11-v1.0_PUGS_CDR_LC-CCI_v2.0.7cds_Products_v1.0.1_APPROVED_Ver1.pdf
RD-3	E.U. Copernicus Climate Change Service (2021) "C3S Product Quality Assurance Document - ICDR Land Cover", E.U. Copernicus Climate Change Service, Document ref.: V1.1, D5.2.3_PQAD_ICDR_LC_v2.1.x_PRODUCTS_v1.1 (2021) https://datastore.copernicus-climate.eu/documents/satellite-land-cover/D5.2.3_PQAD_ICDR_LC_v2.1.x_PRODUCTS_v1.1.pdf
RD-4	E.U. Copernicus Climate Change Service (2021) "C3S Product User Guide and Specification - ICDR Land Cover", E.U. Copernicus Climate Change Service, Document ref.: V1.1, D5.3.1_PUGS_ICDR_LC_v2.1.x_PRODUCTS_v1.1 (2021) https://datastore.copernicus-climate.eu/documents/satellite-land-cover/D5.3.1_PUGS_ICDR_LC_v2.1.x_PRODUCTS_v1.1.pdf

Acronyms

Acronym	Definition
AATSR	Advanced Along-Track Scanning Radiometer
ADS	Annotation Data Set
AMF	Airmass Factor
AOD	Aerosol Optical Depth
AOD550	AOD at 550nm
ARTDECO	Atmospheric Radiative Transfer Database for Earth Climate Observation
ATBD	Algorithm Theoretical Basis Document
AVHRR	Advanced Very High-Resolution Radiometer
BBDR	Broadband Directional Reflectance
BRDF	Bi-Directional Reflectance Distribution Function
C3S	Copernicus Climate Change Service
C3S BA	C3S Burned Area C3S Fire-BA

Acronym	Definition
C3S LC	C3S Land Cover
CCI	Climate Change Initiative
CCI-LC	Climate Change Initiative Land Cover
CDR	Climate Data Record
CDS	Climate Data Store
CEOS	Committee on Earth Observation Satellites
CEOS-WGCV	CEOS Working Group on Calibration and Validation
CMC	Climate Modelling Community
CMIP	Coupled Model Intercomparison Project
CMUG	Climate Modelling User Group
CRS	Coordinate Reference System
CTH	Cloud top height
CWV	Water Vapor Column Content
DARD	Data Access Requirement Document
DJF	Design Justification File
DOM	Dark Object Method
DPM	Detailed Processing Model
DUE	Data User Element
EC	European Commission
ECV	Essential Climate Variable
ELEV	Elevation
Envisat	Environmental Satellite
EO	Earth Observation
ERS	European Remote Sensing Satellite
ERA Interim	Global atmospheric reanalysis from 1979
ESA	European Space Agency
ET	Evapotranspiration
EU	European Union
fAPAR	Fraction-Absorbed Photosynthetically Active Radiation
FOV	Field Of View
FR	Full Resolution
Gamma-RS	Gamma Remote Sensing
GCOS	Global Climate Observing System
GCS	Global Coordinate System
GDAL	Geospatial Data Abstraction Library
GeoTIFF	Georeferenced Tagged Image File Format
GFED	Global Fire Emissions Database
GHSL	Global Human Settlement Layer
GIMMS	Global Inventory Monitoring and Modelling System
GIS	Geographic Information System
GLC2000	Global Land Cover 2000 Project

Acronym	Definition
GlobAlbedo	ESA Data User Element (DUE) project
GlobCover	ESA Data User Element (DUE) Project
GMM	Global Monitoring Mode
GRASS	Geographic Resources Analysis Support System
GUF	Global Urban Footprint
HYGEOS	HYGEOS (company)
ICDR	Intermediate Climate Data Record
IMM	Image Mode Medium
IPCC	Intergovernmental Panel on Climate Change
ISIN	Integerised, sinusoidal grid
ISODATA	Iterative Self-Organising Data Analysis Technique
ISSI	International Space Science Institute
JD	Julian Day
KM	Kilometre
L0	Level 0
L1B	Level 1B
L2	Level 2
L3	Level 3
LAI	Leaf Area Index
Landsat	Land remote sensing Satellite
LC	Land Cover
LC-CCI	ESA Climate Change Initiative Land Cover Project
LCC	Land Cover Change
LCCS	Land Cover Classification System
LS	Land Surface
LUT	Look-Up Table
M	Metre
MC	Mean Compositing
MERIS	Medium Resolution Imaging Spectrometer
ML	Maximum Likelihood
MODIS	Moderate Resolution Imaging Spectroradiometer
MOMO	Matrix-Operator-Model
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NDWI	Normalized Difference Water Index
NetCDF	Network Common Data Format
NIR	Near InfraRed
NLCD	National Land Cover Database
NN	Neuron Net
OLCI	Ocean and Land Colour Instrument
OZO	Ozone Column Content

Acronym	Definition
PDF	Probability Density Function
PDGS	Payload Data Ground Segment
PFT	Plant Functional Types
PQAD	Product Quality Assurance Document
PROBA	Project for On-Board Autonomy
PROBA-V	Project for On-Board Autonomy, with the V standing for Vegetation instrument
PSD	Product Specification Document
PUG	Product User Guide
PUGS	Product User Guide and Specification
PVASR	Product Validation and Algorithm Specification Report
PVIR	Product Validation and Intercomparison Report
PVP	Product Validation Plan
QA	Quality Assurance
RRA	Relative Azimuth Angle
RR	Reduced Resolution
SAFE	Standard Archive Format for Europe
SAR	Synthetic Aperture Radar
SDR	Surface directional reflectance
SLSTR	Sea and Land Surface Temperature Radiometer
SMAC	Simplified Method for Atmospheric Correction
SM_FLAgs	PROBA-V Status Map
SPOT	Satellite Pour l'Observation de la Terre
SPOT-VGT	SPOT- Vegetation
SR	Surface Reflectance
SRTM	Shuttle Radar Topography Mission
SWBD	SRTM Water Body Database
SWIR	Short Wave Infrared
SYN	Synergy
SZA	Solar Zenith Angle
TOA	Top of Atmosphere
TOC	Top of Canopy
UCLouvain	Université catholique de Louvain
UN	United Nations
UNFCCC	United Nations Framework Convention on Climate Change
UR	User Requirements
URD	User Requirements Document
USGS	U.S. Geological Survey
VISCAL	VISible CALibration system
VNIR	Visible and Near-Infrared
VZA	Viewing Zenith Angle
W	Weight

Acronym	Definition
WB	Water Body
WBP	Water Body Product
WGS84	World Geodetic System 84
WSM	Wide Swath Mode

General definitions

Satellite Data Processing Levels

Description of data processing levels ranging from Level 0 to Level 4 has been cited from the following National Aeronautics and Space Administration (NASA) Earth Observation Data website: https://www.earthdata.nasa.gov/engage/open-data-services-and-software/data-information-policy/data-levels

- **Level 0** (L0) data are reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artifacts (e.g., synchronization frames, communications headers, duplicate data) removed. (In most cases, NASA's EOS Data and Operations System [EDOS] provides these data to the DAACs as production data sets for processing by the Science Data Processing Segment [SDPS] or by one of the SIPS to produce higher-level products.)
- **Level 1A** (L1A) data are reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (e.g., platform ephemeris) computed and appended but not applied to L0 data.
- **Level 1B** (L1B) data are L1A data that have been processed to sensor units (not all instruments have L1B source data).
- **Level 1C** (L1C) data are L1B data that include new variables to describe the spectra. These variables allow the user to identify which L1C channels have been copied directly from the L1B and which have been synthesized from L1B and why.
- **Level 2** (L2) data are derived geophysical variables at the same resolution and location as L1 source data.
- **Level 2A** (L2A) data contains information derived from the geolocated sensor data, such as ground elevation, highest and lowest surface return elevations, energy quantile heights ("relative height" metrics), and other waveform-derived metrics describing the intercepted surface.
- **Level 2B** (L2B) data are L2A data that have been processed to sensor units (not all instruments will have a L2B equivalent).
- **Level 3** (L3) are variables mapped on uniform space-time grid scales, usually with some completeness and consistency.
- **Level 3A** (L3A) data are generally periodic summaries (weekly, ten-day, monthly) of L2 products.
- **Level 4** data are model outputs or results from analyses of lower-level data (e.g., variables derived from multiple measurements).

Pre-processing

Data pre-processing in the case of Land cover processing refers to the preparation of the data for the classification, which includes pixel identification and atmospheric correction and temporal resampling. The algorithms used are adapted to the requirements of the subsequent application.

Classification
Image classification in remote sensing is defined as the process of categorising all pixels in satellite data/images to obtain a given set of labels (Lillesand & Keifer, 1994).

Land cover
Land Cover (LC) and land cover change (LCC) are becoming increasingly related to the climate modelling effort. Land cover change is a pressing environmental issue acting as both a cause and a consequence of climate change. The importance of these issues requires continuous monitoring systems and the most accurate data. The Copernicus Climate Change Service provides Intermediate Climate Data Records for many Essential Climate Variables, including LC. The global C3S LC maps 2016 – 2022 are/will be consistent with the existing European Space Agency Climate Change Initiative global annual LC maps from 1992 – 2015.

Land cover is the observed bio-physical cover on the Earth's surface (Townshend et al., 2008). It is not to be confounded with land use. Land use characterizes the arrangements, socio-economic activities and inputs people are undertaking on a certain land cover type.

The proposed land cover ontology assumes that the land cover is organized along a continuum of temporal and spatial scales and that each land cover type is defined by a characteristic scale, i.e. by the typical spatial extent and period over which its physical traits are observed (Miller, 1994). This twofold assumption requires introducing the time dimension in the land cover characterization to allow distinguishing between the stable and the dynamic components of the land surface.

The stable component, named "land cover", refers to the set of land surface features which remain stable over time and thus define the land cover independently of any sources of temporary or natural variability. Conversely, the dynamic component is directly related to this temporary or natural variability that can induce some variation in land surface features over time but without changing the land cover in its essence. This dynamic component is referred to as "land cover conditions" (Lamarche et al., 2013).

LC Change
In this context, 'LC change' (LCC) is therefore considered as a permanent modification of the LC – and not of its conditions – in comparison with a baseline status.

LC Classes
A LC class refers to a LC category described by a stable ensemble of land surface features forming a LC class (e.g., forest, cropland). Land surface features consist of landscape elementary units (e.g., a house, a tree, a water body, etc.) described by:

(1) the type of the observed features, such as tree, shrub, herbaceous vegetation, moss/lichen vegetation, terrestrial or aquatic vegetation, inland water, built-up areas, permanent snow/ice, etc;

(2) the structure of the observed features, like the vegetation height, vegetation cover, building density, etc;
(3) the nature of the observed features, such as the level of artificiality or some species information (e.g., C3/C4 distinction);
(4) the homogeneity of the observed features at the level of observation, leading to pure or mosaic classes.

The LC classes are well-defined and described using the UN/FAO Land Cover Classification System (LCCS) (di Gregorio and Jansen, 2005).

Current pixel state

The current_pixel_state in the Land cover map product is the classification of the aggregated pixels regarding the following classes:

- invalid – no observation or all observations are not valid
- clear_land – at least one valid observation over clear land
- clear_water – at least one valid observation over clear water and no valid observations over clear land and clear snow/ice areas
- clear_snow_ice – at least one valid observation over a clear snow/ice area and no valid observations over clear land
- cloud – all valid observations are covered by clouds
- cloud_shadow – at least one valid cloud shadow observation and no valid observation over clear snow/ice area, clear land and clear_water

Table of Contents

Scope of the document

This document is the Algorithm Theoretical Basis Document (ATBD) for the products Land Cover 2021 and 2022, version v2.1.1, produced by Université catholique de Louvain and Brockmann Consult GmbH from the L1b data of SENTINEL-3 Ocean and Land Colour (S3-OLCI) Instrument and SENTINEL-3 Sea and Land Surface Temperature Radiometer (S3-SLSTR) (2021 &2022). It describes the algorithms used to generate the products Land Cover 2021 and 2022, including the scientific justification for the algorithms selected to derive the products, an outline of the proposed approach and a listing of the assumptions and limitations of the algorithm.

After this overview, the document is divided into four sections that are briefly described below:

- Section 1 presents the instruments used for the products Land Cover 2021 and 2022 in the C3S Land Cover (C3S LC),
- Section 2 describes the input and auxiliary data for the processing,
- Section 3 presents the processing chain modules in detail,
- Section 4 describes the products Land Cover 2021 and 2022.

Executive summary

The Copernicus Climate Change Service (C3S) provides Intermediate Climate Data Records (ICDRs) for many Essential Climate Variables (ECVs), among which is land cover (LC). The C3S 2021 and 2022 v2.1.1 global LC maps, consistent with the existing European Space Agency (ESA) Climate Change Initiative (CCI) global annual LC maps from 1992 – 2015 and the C3S global annual LC maps from 2016 – 2020, are generated from the SENTINEL-3 OLCI and SLSTR imagery at 300m spatial resolution. The typology of the map is defined based on the hierarchical United Nations (UN) – Land Cover Classification System (LCCS), with a global level of 22 classes and a detailed, regional level.

A new theoretical LC concept, pre-processing and classification chains were initially developed in the LC component of the ESA CCI to generate a series of annual LC maps from 1992 to 2015. Within C3S, these processing chains were then adapted to generate annually and in an ongoing operational manner the LC maps 2016–2022 from PROBA-V (2016 – 2019), from SENTINEL-3 Ocean and Land Colour Instrument (S3-OLCI) (2020) and from SENTINEL-3 Ocean and Land Colour (S3-OLCI) Instrument and SENTINEL-3 Sea and Land Surface Temperature Radiometer (S3-SLSTR) (2021 &2022).

The PROBA-V instrument provided Earth Observation (EO) acquisitions in four spectral bands at 300 m and 1 km with a temporal revisit of two days. The S3-OLCI instrument inherits and outperforms ENVISAT's MERIS specifications and provides data continuity in terms of sensor specifications between MERIS and OLCI. It acquires global EO data over land every three days in 21 bands at a full resolution of 300 m. The SLSTR instrument provides radiometric measurements expressed in Top-Of-Atmosphere (TOA) brightness temperatures for thermal IR and fire channels and expressed in TOA

radiances for the visible / SWIR channels. The Level 1b SENTINEL-3 OLCI and SLSTR data were cloud screened and atmospherically corrected and like the Level 3 PROBA-V Top-of-Canopy daily aggregated into seasonal composites using the mean compositing strategy (Vancutsem et al., 2007).

Building on the ESA CCI LC heritage, annual classifications for 2016 – 2022 were generated at full (300 m) and low (1.2 km) resolutions, extending the annual map series for land cover change analysis. The 1 km spatial resolution served for LCC detection, later remapped in more detail using the 300 m data. These processing steps generate a series of C3S global annual LC maps for the period 2016 – 2022, based on PROBA-V and Sentinel-3-OLCI and Sentinel-3 SLSTR, consistent with the existing CCI global annual LC maps from 1992 – 2015. It is delivered with metadata, product documentation, and validation reports.

This version of the ATBD focuses on the processing for the 2021 and 2022 ICDR LC products. Its main specifications are summarised in Table 0-1.

Table 0-1: Summary of the C3S LC products

Product	Coverage		Resolution		Sensor	Projection	Format
	Spatial	Temporal	Spatial	Temporal			
Annual LC maps	Global	2021	0.002778°	1-year	SENTINEL-3 OLCI / SLSTR	Plate-Carrée	NetCDF[1]
Annual LC maps	Global	2022	0.002778°	1-year	SENTINEL-3 OLCI / SLSTR	Plate-Carrée	NetCDF

[1] *NetCDF, 2010*

1 Instruments

1.1 SENTINEL-3 OLCI

1.1.1 Instrument Payload

The following information has been cited from the European Space Agency´s pages regarding this instrument: https://sentinel.esa.int/web/sentinel/missions/sentinel-3/instrument-payload/olci

“The SENTINEL-3 Ocean and Land Colour Instrument (OLCI) is based on the opto-mechanical and imaging design of Envisat's MERIS instrument (Figure 1-1). The main characteristics of the OLCI are:

- swath width: 1,270 km
- push-broom imaging spectrometer with five cameras, mitigation of sun-glint contamination by tilting cameras in westerly direction
- spatial sampling: 300 m @ SSP
- spectrum: 21 bands [0.4-1.02] µm
- radiometric accuracy: 2% abs, 0.1% rel
- launch mass: 153 kg
- size: 1.3 m^3
- design lifetime: 7.5 years.”

Figure 1-1: SENTINEL-3 OLCI Instrument (Credit: ESA)

1.1.2 Resolutions

The following information has been cited from the European Space Agency´s pages regarding this instrument: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-olci/resolutions

“The OLCI instrument measures reflected solar radiation from the Earth's surface and clouds simultaneously in 21 spectral bands.

OLCI products are available at two spatial resolutions:

- Full Resolution (FR) at approximately 300 m
- Reduced Resolution (RR) at approximately 1.2 km.

OLCI data is acquired in Full Resolution over Land and Ocean, but Level1B[2]processing can generate products in FR or RR resolutions, or both. Level 2 processing ingests either FR or RR Level 1b products and generates Level 2 products at the same resolution."

Spatial Resolution (https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-olci/resolutions/spatial)

"Full Resolution (FR)

- To simplify OLCI operations, maximise instrument autonomy and offer maximum flexibility for data processing and re-analysis, OLCI always operates (both land and ocean) in FR mode.
- For the nominal orbit, at sub-satellite point, OLCI Full Resolution is approximately 300 m on ground.

Reduced Resolution (RR)

- From Level-1B, OLCI products are provided in a sub-sampled version, referred to as RR.
- RR is obtained by averaging the signal of a grid of 16 FR pixels (4 Along Track x 4 Across Track).
- For the nominal orbit, at sub-satellite point, OLCI Reduced Resolution is approximately 1.2 km on ground."

Radiometric Resolution - 21 bands in VIS/SWIR (https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-olci/resolutions/radiometric)

"OLCI observation is performed simultaneously in 21 spectral bands, listed in table 1-1 below, ranging from the visible to the near-infrared (400 nm to 1 020 nm). Each of these bands is programmable in position and width."

Table 1-1: Band characteristics of the SENTINEL-3 OLC, in cyan MERIS heritage

Band	λ centre(nm)	Width (nm)	Function
Oa01	400	15	Aerosol correction, improved water constituent retrieval
Oa02	412.5	10	Yellow substance and detrital pigments (turbidity)
Oa03	442.5	10	Chlorophyll absorption maximum, biogeochemistry, vegetation
Oa04	490	10	High Chlorophyll,
Oa05	510	10	Chlorophyll, sediment, turbidity, red tide
Oa06	560	10	Chlorophyll reference (Chlorophyll minimum)
Oa07	620	10	Sediment loading
Oa08	665	10	Chlorophyll (2nd Chlorophyll absorption maximum), sediment, yellow substance/vegetation
Oa09	673.75	7.5	For improved fluorescence retrieval and to better account for smile together with the bands 665 and 680 nm
Oa10	681.25	7.5	Chlorophyll fluorescence peak, red edge
Oa11	708.75	10	Chlorophyll fluorescence baseline, red edge transition
Oa12	753.75	7.5	O2 absorption/clouds, vegetation
Oa13	761.25	2.5	O2 absorption band/aerosol correction.

[2] *Level-1B (L1B) data are Level-1A (L1A) data that have been processed to sensor units (not all instruments have L1B source data). (L1A data are reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (e.g., platform ephemeris)) - https://www.earthdata.nasa.gov/engage/open-data-services-and-software/data-information-policy/data-levels*

Band	λ centre(nm)	Width (nm)	Function
Oa14	764.375	3.75	Atmospheric correction
Oa15	767.5	2.5	O2A used for cloud top pressure, fluorescence over land
Oa16	778.75	15	Atmospheric correction./aerosol correction
Oa17	865	20	Atmospheric correction/aerosol correction, clouds, pixel co-registration
Oa18	885	10	Water vapour absorption reference band. Common reference band with SLSTR instrument. Vegetation monitoring
Oa19	900	10	Water vapour absorption/vegetation monitoring (maximum reflectance)
Oa20	940	20	Water vapour absorption, Atmospheric correction/aerosol correction
Oa21	1 020	40	Atmospheric correction/aerosol correction

1.1.3 Data Products

The following information has been cited from the European Space Agency´s pages regarding this instrument: https://sentinel.esa.int/web/sentinel/missions/sentinel-3/data-products/olci

"There are different data products associated with the three levels of processing of OLCI:

- **Level-0** is the reconstructed and time-sorted Instrument Source Packet (ISP) at full space-time resolution. All communications artefacts (e.g. synchronisation frames, communications headers and duplicate data) and invalid packets are removed. OLCI data is always sensed in full resolution mode (300 m resolution).
- **Level-1** includes Top-Of-Atmosphere (TOA) radiometric measurements, radiometrically corrected, calibrated and spectrally characterised. It is quality controlled, ortho-geolocated (latitude and longitude coordinates, altitude) and annotated with satellite position and pointing, landmarks and preliminary pixel classification (e.g. land/water/cloud masks). Products are generated in FR (300 m) and in RR (1.2 km) for the whole globe with the same coverage.
- **Level-2** products consist of geophysical quantities derived from the processing of measurement data provided in the Level-1 product. Level-2 products specifically for marine and land application domains are generated separately by the SENTINEL-3 Payload Data Ground Segment (PDGS), with each containing the parameter relevant for the specific field of application. Level-2 atmospheric information relevant for both application domains, such as water vapour, is reported in both data streams.

OLCI Level-1 and 2 data products are available to the general public.

The OLCI files are collected into a Standard Archive Format for Europe (SAFE[3]) container. Level-1 and 2 products are encapsulated in free-standing NetCDF 4 product files.

The timeframe for delivery of products is dependent on the specific application:

- Near Real-Time (NRT) products are delivered to the users less than 3 hours after acquisition of the data by the sensor

[3] *SENTINEL-SAFE is based on the XML Formatted Data Units (XFDU) developed by the Consultative Committee for Space - Data Systems (CCSDS).https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-3-olci/data-formats*

- Non-Time Critical (NTC) products are delivered not later than 1 month (commitment) after acquisition or from long-term archives. Typically, the product should be available within 24 or 48 hours (but this is not guaranteed)."

1.2 SENTINEL-3 SLSTR

1.2.1 Instrument Payload

The following information has been cited from the European Space Agency´s pages regarding this instrument: https://sentinel.esa.int/web/sentinel/missions/sentinel-3/instrument-payload/slstr

"Following ENVISAT's Advanced Along-Track Scanning Radiometer (AATSR) instrument, the SLSTR instrument is a conical scanning imaging radiometer employing the along track scanning dual view technique.

The main characteristics of the SLSTR are:

- swath width: dual view scan, 1,420 km (nadir) / 750 km (backwards)
- spatial sampling: 500 m (VIS, SWIR), 1 km (MWIR, TIR)
- spectrum: nine bands [0.55-12] µm
- noise equivalent dT: 50 mK (TIR) at 270 K
- launch mass: 90 kg
- size: 2.116 m3
- design lifetime: 7.5 years."

Figure 1-2: SENTINEL-3 SLSTR Instrument (Credit: Selex-Galileo & Jena-Optronik)

1.2.2 Resolutions

The following information has been cited from the European Space Agency´s pages regarding this instrument: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-slstr/resolutions

"The SLSTR products (Level-1B and Level-2 measurement, annotation and auxiliary data sets) are generated separately for the two instrument views (nadir and oblique) at a spatial resolution that depends on which channel was used:

- 500 m resolution for solar reflectance bands (S1-S6)
- 1 km resolution for thermal infrared bands (S7-S9 and F1-F2)"

Spatial Resolution – 500m VNIR/SWIR (https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-slstr/resolutions/spatial/500m)

"The visible and ShortWave Infra-Red (SWIR) channels S1 - S6 are collected and stored at 500 m resolution.

The SLSTR detectors are located in a Focal Plane Assembly (FPA). Figure 1-3 shows the relative orientations and nominal sizes of the SLSTR nadir view detector Instantaneous Fields Of View (IFOVs) at the sub-satellite track position. The X-axis is the nadir scan direction, the Y-axis is the direction of flight and O is the principal ray. The rightmost four detector elements (blue with a solid outline) are implemented in all six channels, while the remainder (green with a dashed outline) are only present in channels S4 - S6."

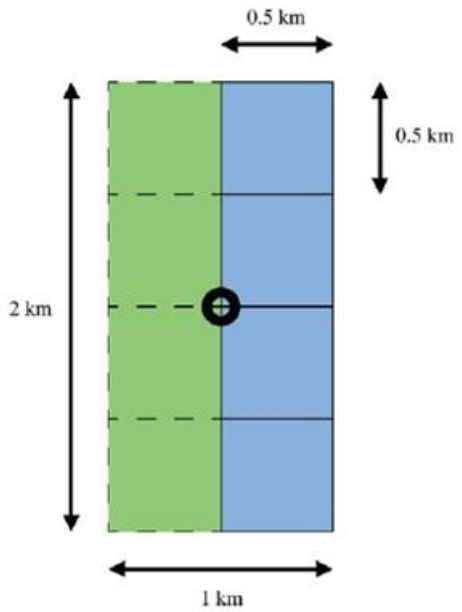

Figure 1-3: SLSTR Detector Configuration 500 m Spatial Resolution Channels S1 - S6 (VIS/NIR to SWIR)- (https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-slstr/resolutions/spatial/500m)

Spatial Resolution – 1 km TIR (https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-slstr/resolutions/spatial/1km)

"Thermal infra-red channels S7 - S9, F1 and F2 are collected and stored at 1 km resolution. Each channel has two detectors, aligned along track at the sub-satellite point.

As shown in Figure 1-4, the leftmost detector elements (yellow) are common to channels S7 - S9 and F2, while the remaining detector elements (red) are used in channel F1 only."

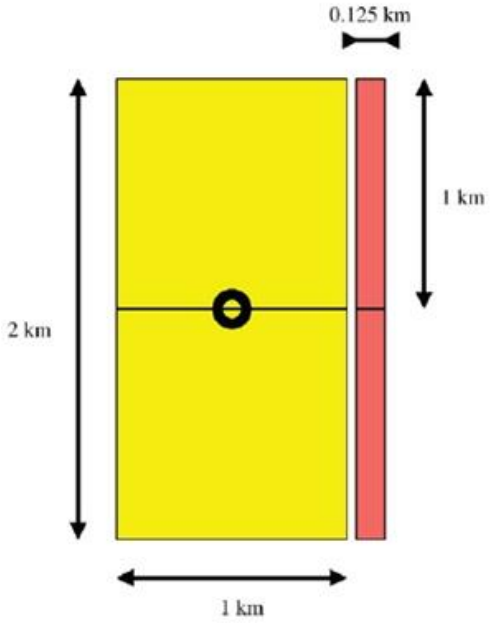

Figure 1-4: SLSTR Detector Configuration 1 km Spatial Resolution Channels S7 - S9, F1 and F2 (MIR to TIR)- (https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-slstr/resolutions/spatial/1km)

1.2.3 Data Products

The following information has been extracted from the European Space Agency´s pages regarding this instrument: https://sentinel.esa.int/web/sentinel/missions/sentinel-3/data-products/slstr

Similar to the OLCI products, there are different data products associated with the three levels of processing of SLSTR: Level-0 , Level-1 and Level-2. The SLSTR Level-1 and 2 data products are available to the general public also. The SLSTR files are also collected into a SAFE[4] container. Level-1 and 2 products are encapsulated in free-standing NetCDF 4 product files. The SLSTR Level-1 and 2 data products are provided as NRT products and/or NTC products are available.

[4] *SENTINEL-SAFE is based on the XML Formatted Data Units (XFDU) developed by the Consultative Committee for Space Data Systems (CCSDS) -.*

2 Input and auxiliary data

2.1 Input data

2.1.1 SENTINEL-3 OLCI L1B data

The following information has been cited from the European Space Agency´s pages regarding this instrument: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-olci/product-types/level-1b .

"From the three processing modes of the OLCI Level-1 processor (Earth Observation (EO), Radiometric Calibration (RC) and Spectral Calibration (SC)) and the two available Resolutions in EO (Full and Reduced Resolution), four different Level-1B products can be obtained:

- *OL_1_EFR*, output during EO processing mode for Full Resolution
- *OL_1_ERR*, output during EO processing mode for Reduced Resolution
- *OL_1_RAC*, output during Radiometric Calibration mode
- *OL_1_SPC*, output during Spectral Calibration mode

NOTE: *OL_1_RAC* and *OL_1_SPC are* internal PDGS products, and not disseminated to SENTINEL-3 users.

The Level-1B products in EO processing mode contain calibrated, ortho-geolocated and spatially re-gridded Top-Of-Atmosphere (TOA) radiances for the 21 OLCI spectral bands. The associated error estimates are also contained in the measurement data files. In Full Resolution products (i.e. at native instrument spatial resolution), these parameters are provided for each re-gridded pixel on the product image and for each removed pixel. In Reduced Resolution products (i.e. at a resolution four times coarser), the parameters are only provided on the product grid.

In addition to measurement data sets, annotation data sets provide:

- time stamps for each line of the product grid
- geolocation information for each pixel (and for each removed pixel in case of *OL_1_EFR*)
- quality flags, concerning surface or cloud identification, invalid or cosmetically filled pixels
- meteorological variables for each tie-point (defined on a specific grid: every 16 pixels for RR and every 64 pixels for FR products)
- geographical information and angles associated with each tie-point
- instrument features and settings needed in further processing such as detector index or OLCI channels, central wavelength and bandwidths."

2.1.2 SENTINEL-3 SLSTR L1B data

The following information has been cited from the European Space Agency´s pages regarding this instrument: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-3-slstr/product-types/level-1b .

“The SLSTR Level-1B product gathers, for each view and for each channel, the full-resolution geolocated radiometric measurements.

For thermal IR and fire channels (labelled as S7 to S9 for TIR channels and F1, F2 for fire channels), the radiometric measurements are expressed in Top-Of-Atmosphere (TOA) brightness temperatures. In the case of visible / SWIR channels (labelled as S1 to S6), these measurements are expressed in TOA radiances.

The radiometric measurements are indexed according to the across track and along track direction:

- on a 1 km grid for brightness temperature for channels S7-S9 and F2
- on a 1km grid that is applicable to the F1 channel
- on a 0.5 km grid for radiances. In this case, two stripes are distinguished: stripe A for S1-S6 and stripe B for S4-S6.

There are essentially eight 'grids' due to the TIR 1km infrared 'i' grid, the 'f' grid specific to F1, and the two different 0.5km grids 'a', 'b'. Each of these is represented on the nadir and oblique views, making a total of eight.

The SLSTR Level-1B product also contains the following annotation data sets (ADSs):

- Quality - quality annotations per channel, such as estimates of detector noise measured at the VISible CALibration system (VISCAL) or black body calibration sources
- Flags - flags concerning cloud identification, surface classification and scan and flip mirror testing provided per grid
- Indices - arrays of indices to retrieve the position of each pixel in the instrument measurement frame provided per grid
- Cartesian - a file providing the Cartesian coordinates per grid
- Geodetic - geodetic coordinates associated with the radiometric measurements on each grid and to the tie-point grid defined for SLSTR
- Time - the timings of the pixels provided per grid for the nadir view only
- Met - meteorological annotations, provided once per product”

2.2 Auxiliary data

2.2.1 Auxiliary data applied in the pre-processing chain

Currently, no auxiliary data have been used in the pre-processing chain.

2.2.2 The reference land cover layer

A critical auxiliary data set in the classification chain is a global reference LC database. This database consists of a set of existing global, regional and local LC maps merged after preparatory transformations and according to specific merging rules. It is used as a training and labelling layer in the supervised and unsupervised classification modules, respectively. The reference LC layer and data sets from the following sections are described in detail in the [ESA_CCI_LC_ATBD, 2017].

2.2.3 The stratification layer

Before running the pre-processing and classification modules, the world is split into 22 strata from an ecological and remote sensing point of view (Defourny et al., 2006) (*Figure 2-1*). The stratification objectives are twofold: (1) reducing the land surface reflectance variability in the data set in order to improve the classification efficiency and (2) allowing a regional tuning of the classification parameters to consider regional characteristics such as the vegetation seasonality, cloud coverage, etc. The generation of the seasonal composites and the classification chain runs independently for each stratum with specific and fine-tuned parameters. This strategy provides the high level of generality required for land cover processing on a global scale, while allowing a high level of specificity and flexibility to achieve a certain level of accuracy in mapping regional landscapes.

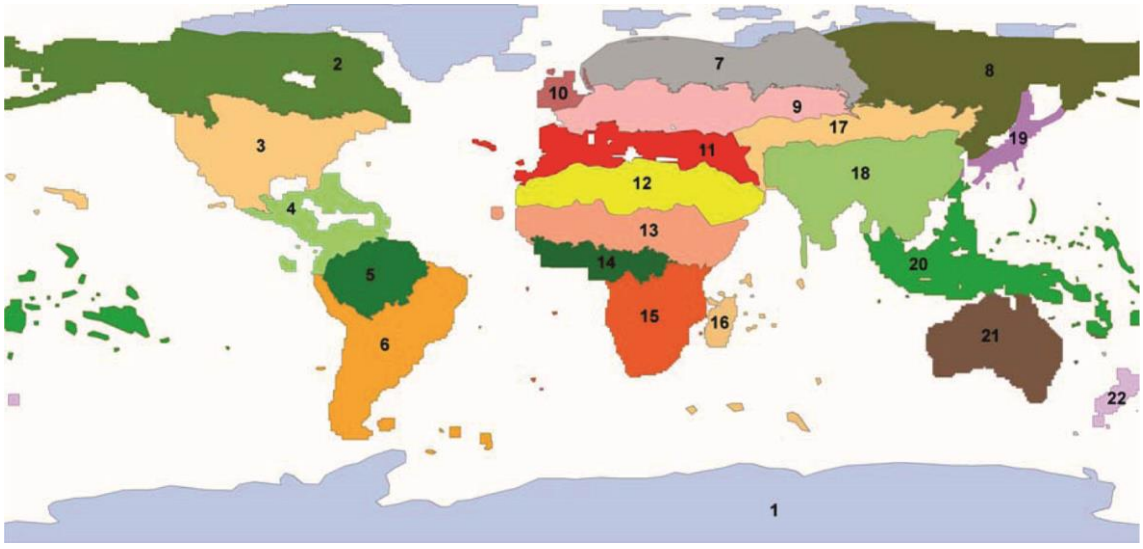

Figure 2-1: Stratification layer, which consists of the 22 strata (Defourny et al., 2006).

2.2.4 The baseline land cover map

The baseline global land cover (LC) map was created by classifying the complete Envisat MERIS data archive from 2003 to 2012, with a spatial resolution of 300 m. The reference land cover layer was used as a source of information for training and labelling during the classification process. The baseline global LC map describes the stable part of the LC as 22 LCCS classes. The baseline LC map is updated where and when changes have been detected in the change detection module.

2.2.5 Existing annual LC maps series

The spatio-temporal consistency should be ensured between each newly produced LC map and the existing LC map series.

3 Algorithms

3.1 The land cover pre-processing chain

The land cover (LC) pre-processing chain requires surface reflectance composites as input products. The complete automated pre-processing chain, developed during the Fire-BA CCI and LC-CCI projects (ESA_CCI_LC_ATBD, 2017 & ESA_CCI_LC_ATBD, 2017a), has been selected for the pre-processing of the SENTINEL-3 Level 1B OLCI and SLSTR data. The pre-processing performs the following operations depending on the used sensors: radiometric correction, geometric correction, pixel identification, atmospheric correction with aerosol retrieval as well as compositing and mosaicking. In the case of SENTINEL-3 Level 1B OLCI and SLSTR data, the pre-processing chain of MERIS data has been adapted and used. The surface reflectance composite includes the non-absorbing SENTINEL-3 OLCI bands (1 … 12, 16 … 18), see also Table 1-1, and the SLSTR bands 1-6 at nadir and from stripe A. The pre-processing chain is shown in Figure 3-1 (data that are not used and modules that are not applied are indicated with the dashed structure) and the different applied modules are exhaustively detailed in the following.

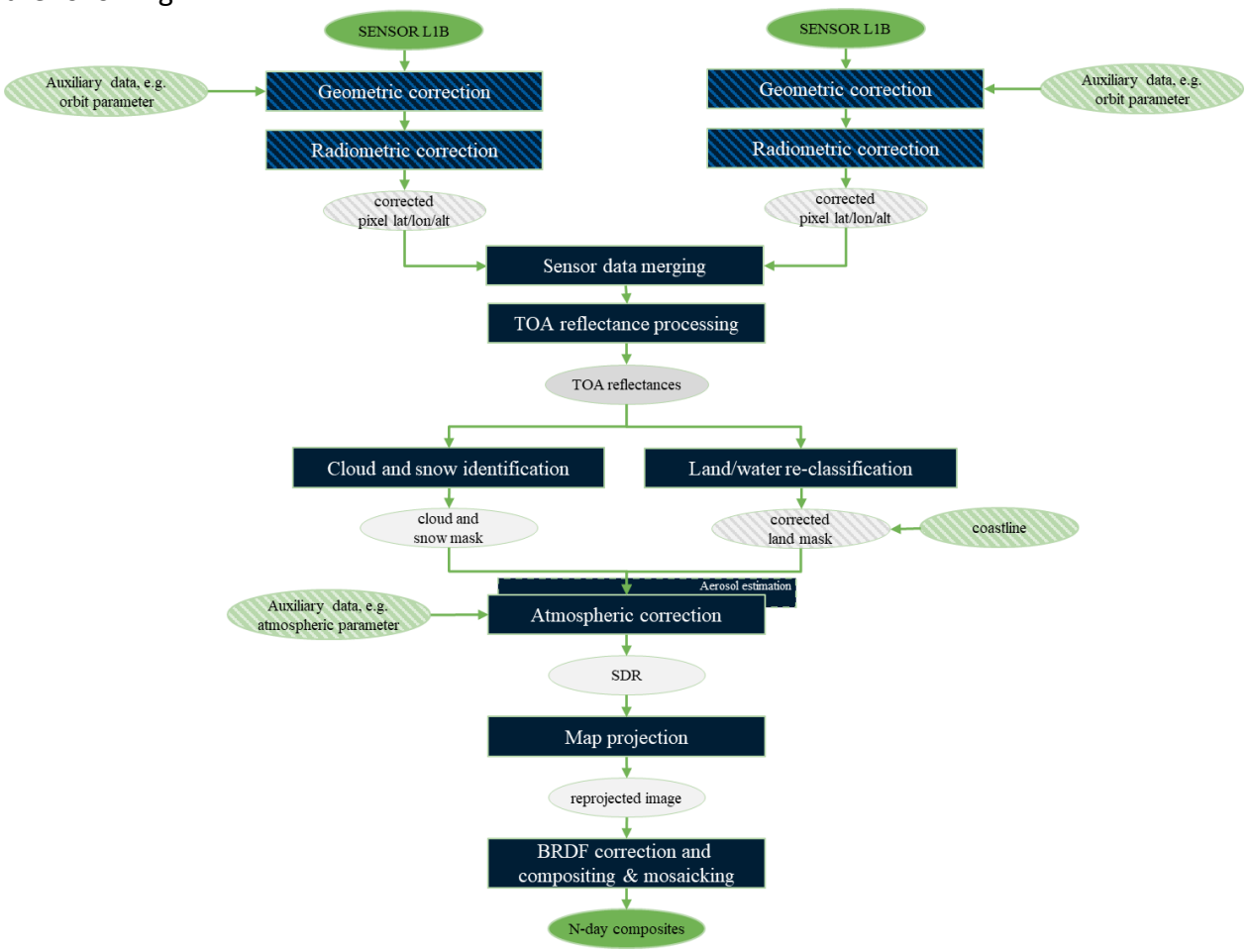

Figure 3-1: Schematic representation of the adapted Land Cover CCI and Fire-BA CCI pre-processing chain including input data (pre-processing chain based on the GlobAlbedo chain, GlobAlbedo_ATBDv4.12, 2013)[5]

[5] *Note: not all modules, processing steps and auxiliary data (indicated by structure) have been applied/used in the SENTINEL-3 OLCI & SLSTR LC processing chain.*

3.1.1 Geometric correction

This module is not applied to the SENTINEL-3 OLCI L1b or SENTINEL-3 SLSTR L1b data, as there is sufficient geolocation accuracy in the products.

3.1.2 Radiometric correction

This module is also not applied to the SENTINEL-3 OLCI L1b or SENTINEL-3 SLSTR L1b data, as there is sufficient radiometric accuracy in the products.

3.1.3 Sensor data merging

The processing steps to merge the SENTINEL-3 OLCI L1b and SENTINEL-3 SLSTR L1b data are: (i) band sub-setting of SENTINEL-3 SLSTR L1b data, (ii) applying the pixel geocoding for both products, and (iii) the collocation of SENTINEL-3 SLSTR L1b data on the SENTINEL-3 OLCI L1b reference grid as a master using nearest neighbour resampling. The reference grid is defined by the acquisition geometry of OLCI channel 17 to be consistent with the official SENTINEL3 SYN products.

The merged SENTINEL-3 OLCI/SLSTR product contains the TOA radiance data for the 6 solar reflective SENTINEl-3 SLSTR bands at nadir and stripes A and for all SENTINEl-3 OLCI bands. Furthermore, the information regarding the original solar/view geometry of the SENTINEl-3 OLCI and SLSTR data and the SENTINEl-3 SLSTR corresponding and related flag bands are also included.

3.1.4 TOA reflectance processing

Some modules of the processing chain require the apparent spectral directional reflectance instead of the spectral radiance as input data. The apparent spectral directional reflectance $\rho^*(\lambda, \theta s, \theta v)$ in the waveband λ of the coupled surface atmosphere system is related to the spectral radiance $L_{sat}(\lambda, \theta s, \theta v)$ measured by a satellite at the TOA by:

$$\rho^*(\lambda, \theta_s, \theta_v) = \frac{\pi \cdot L_{sat}(\lambda, \theta_s, \theta_v)}{\mu_s \cdot E_{sun}(\lambda, \theta_s, \theta_v)}$$

eq. 3-1

where $E_{sun}(\lambda, \theta_s, \theta_v)$ is the extraterrestrial irradiance at the time of the measurement. The view and solar angles are denoted by θ_v respectively θ_s, while μ_s denotes the cosine of solar zenith. The extra-terrestrial irradiance, as well as the corresponding sun and viewing geometry are part of SENTINEL-3 L1b OLCI and SENTINEL-3 L1b SLSTR products and are also part of the merged SENTINEL-3 OLCI/SLSTR products.

3.1.5 Pixel Identification: Land water delineation and cloud screening

The Multi Sensor Pixel Identification approach (which is also applied in the ESA GlobAlbedo, ESA LC-CCI and ESA Fire-BA CCI project and is explained, e.g., in the GlobAlbedo ATBD (GlobAlbedo_ATBDv4.12, 2013) classifies each pixel to be processed according to a series of pixel categories, which include cloud and cloud shadow, clear-land, clear-water and clear-snow/ice. Cloudy

pixels are not processed in LC processing, while land, water and snow pixels must be distinguished because of the particular processing steps associated with each surface type. In particular, water pixels must be separated from land surfaces even in the case of continental water bodies, as these are flagged in the final pre-processed product. Snow and snow-free surfaces will also be considered separately in the final pre-processed product.

The term "Pixel identification" refers to a classification of a measurement made by a space borne radiometer, to identify properties of the measurement which are influencing further algorithmic processing steps. Most importantly is the classification of a measurement as being made over cloud, a clear sky land surface or a clear sky ocean surface. The term "pixel" is often used for such a measurement to express it being part of a spatially oriented collection of many measurements, which all are geo-located and which form, as a whole, an image of the Earth below the satellite.

While the information regarding whether a pixel is made over water or land can be taken from a static map with suitable spatial resolution, the cloud coverage is spatially and temporally highly variable and needs to be derived from the measurement itself. Also, the snow coverage can be spatially and temporally variable and is therefore also derived from the measurements itself. After determining whether a pixel is cloudy or clear, in the clear sky case the land-water information can be refined using the measurement. This is particularly necessary in the coastal zone where the actual land-water boundary is changing due to tides, when the pixel size is small enough to resolve this difference. Also, maps are not always correct and therefore a radiometric refinement is advisable.

The pixel identification (IdePix) is a framework for the identification of properties of the measurement for different sensors, e.g. MERIS, SPOT-VGT, AVHRR, PROBA-V and SENTINEL-3 OLCI and SLSTR data considered (Figure 3-2). The uniqueness consists of a certain set of features, which are calculated for each instrument and then combined in order to calculate a set of pixel classification attributes. The implementation of how and which features are calculated is instrument specific. This approach has the advantage of being easily extendable to other instruments, and it is also applied in the GlobAlbedo, LC-CCI and Fire CCI BA projects.

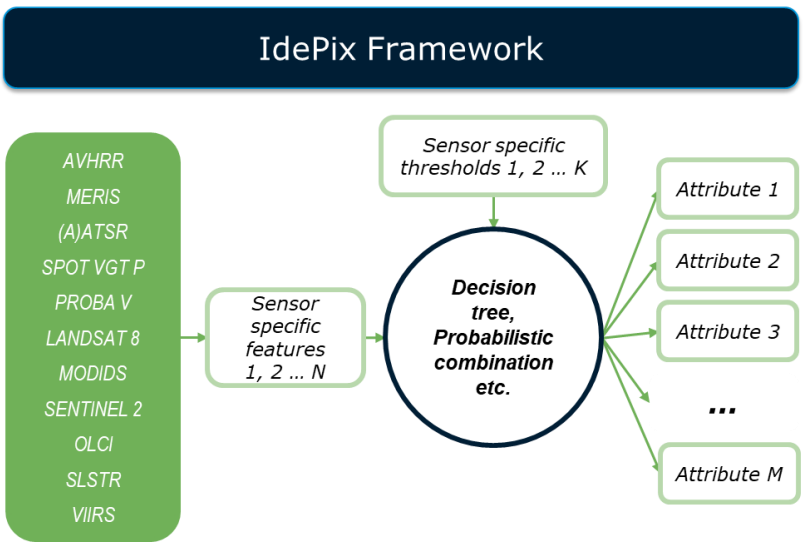

Figure 3-2: Pixel identification Framework (IdePix) including the integrated sensors

In the case of the merged SENTINEL-3 OLCI/SLSTR data the SENTINEL-3 OLCI IdePix is applied and complemented by additional tests based on the SENTINEL-3 SLSTR bands and cloud flags.

SENTINEL-3 OLCI IdePix

The SENTINEL-3 OLCI IdePix is fully described in the corresponding ATBD (Wevers et al., 2022), therefore only a brief description is provided here. In the current SENTINEL-3 OLCI IdePix only the cloud value feature has been calculated. The cloud value feature is predicted by a neural network (NN) model trained with a back-propagation learning algorithm by using the PixBox[6] data. The training data set was split into two based on the surface properties, namely land and water and then these two data sets have been used separately for the training. The PixBox data are described in the PVASR (ESA_CCI_LC_PVASRv2.0, 2012). The cloud value feature is calculated through the NN by using the SENTINEL-3 OLCI TOA reflectance spectrum as input. The cloud value feature has a continuous range from 0 through 4. The user defined thresholds are used for the determination of the current pixel status as clear land, clear water, clear snow/ice and cloud. For example, all pixel with a cloud value feature value between 0.0 and 1.1 are identified as clear snow/ice trough the NN.

Additional test using the SENTINEL-3 SLSTR bands and cloud flags and their combination

An additional test has been applied to avoid erroneous classification trough the NN, e.g. classification of bright surfaces as clouds.

Bright surface vs cloud test (Ishida and Nakajima, 2009): The test is based on the ratio of the SENTINEL-3 SLSTR Band 3 (867nm) and Band 5 (1.64µm) TOA reflectance.

$$isSlstrBright35Ratio = 0.82 \leq \left(\frac{\rho_{SLSTR_3}}{\rho_{SLSTR_5}}\right) \&\& \left(\frac{\rho_{SLSTR_3}}{\rho_{SLSTR_5}}\right) \leq 0.92 \;\&\&\; isLand$$
$$= \begin{cases} true & then\ desert\ or\ bright\ surface \\ false & then\ cloud \end{cases}$$

where ρ_{SLSTR_3}, ρ_{SLSTR_5} are the TOA reflectance in the SLSTR bands 3 and 5 and *isLand* is the boolean value regarding the land water mask (true=land, false = water)

eq. 3-2

Snow test (Krijger et al.,2011): The test is based on the ratio of the SENTINEL-3 SLSTR Band 3 (867nm) and Band 2 (1.64µm) TOA reflectance as well as on the ratio of the SENTINEL-3 SLSTR Band 1 (867nm) and Band 5 (1.64µm) TOA reflectance.

$$W43Ratio = \left(\frac{\rho_{SLSTR_3}}{0.795\ \rho_{SLSTR_2}}\right)$$

[6] *PixBox is a proprietary software developed by Brockmann Consult GmbH to conduct expert pixels collections.*

$$W15Ratio = \left(\frac{\rho_{SLSTR_1}}{\rho_{SLSTR_5}}\right)$$

$$isSlstrSnowIce = W43Ratio \geq 0.77 + \left(\frac{1}{W15Ratio - 0.08}\right) = \begin{cases} true\ then\ snow \\ false\ then\ not\ snow \end{cases}$$

where ρ_{SLSTR_1}, ρ_{SLSTR_2}, ρ_{SLSTR_3}, ρ_{SLSTR_5} are the TOA reflectance in the SLSTR bands 1, 2, 3 and 5

eq. 3-3

Cloud test based on the SENTINEL-3 SLSTR cloud flag:

$$isSlstrCloud = \text{isSlstrCloudAnGrossCloud} \,||\, \text{isSlstrCloudAn137Thresh}$$

where isSlstrCloudAnGrossCloud and isSlstrCloudAn137Thresh are boolean values of the two internal SLSTR cloud screening test – SLSTR GROSS CLOUD and the SLSTR 137 THRESHOLD TEST.

eq. 3-4

Combination of SENTINEL-3 OLCI and SLSTR cloud flags:

$$isSynCloud = cloudSureNN \,||\, isSlstrCloud$$

where *cloudSureNN* is the boolean values regarding the cloud detection via NN and the cloud sure class.

eq. 3-5

The results of the additional tests have to be considered for the final determination of the pixel status as clear land, clear water, clear snow/ice and cloud

Finalisation of the IdePix Flags:

$$IDEPIX\_CLOUD\_AMBIGUOUS = cloudAmbiguousNN \,\&\&\, !\, isSlstrBright35Ratio \,\&\&\, !\, isSLSTRSnowIce$$
$$IDEPIX\_CLOUD\_SURE = isSynCloud \,\&\&\, !\, isSlstrBright35Ratio \,\&\&\, !\, isSLSTRSnowIce$$
$$IDEPIX\_(SNOW\_ICE) = SnowIceNN \,||\, isSlstrSnowIce$$

where *cloudAmbigousNN* is the boolean values regarding the cloud detection via NN and the cloud ambigous class and *SnowIceNN* is the boolean values regarding the snow/ice detection via NN.

eq. 3-6

Mountain, Cloud Shadow and Cloud Edge detection

A more comprehensive representation of clouds is provided by two additional pixel properties, 'cloud edge' and 'cloud shadow'. These two properties are also kept in the pixel identification attributes/flags and can be considered as a kind of "post-processing" in the cloud detection.

To detect dark cloud shadow pixels -whose spectra are polluted by the shadow- the position of the shadow is determined by projecting the 'cloud pixels' onto the ground using the sun position, the pixel's altitude on Earth, and the cloud height estimated from cloud top pressure (Figure 3-3).

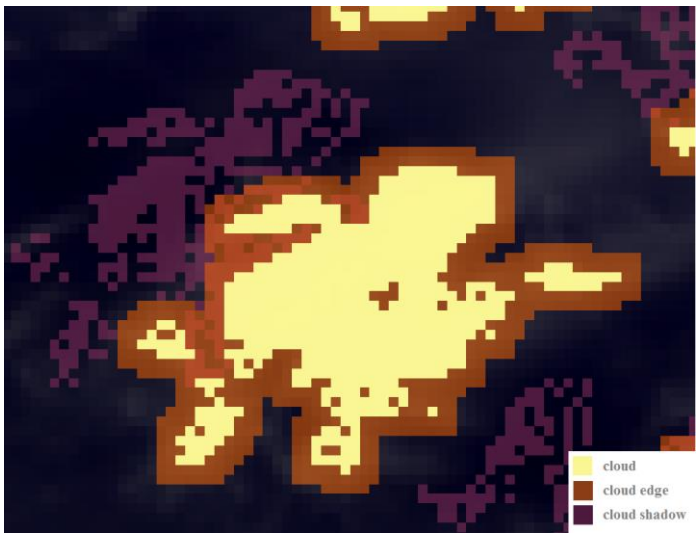

Figure 3-3: Example of cloud, cloud edges and cloud shadow

Cloud edge pixels are in principle regarded as neighbour pixels of a 'cloud' as identified before in the cloud detection as part of the pixel identification. The width of this edge (in number of pixels) can be set by the user. Briefly, the algorithm to identify cloud edge pixels works as follows:

- use a 2x2 square with reference pixel in upper left
- move this square row-by-row over the given tile
- if reference pixel was not cloud, don't do anything
- if reference pixel is a cloud
  - if 2x2 square only has cloud pixels, then set cloud buffer of two pixels in both x and y direction of reference pixel;
  - if 2x2 square also has non-cloudy pixels, do the same but with cloud buffer of only 1.

The IdePix for SENTINEL-3 OLCI also includes a mountain and cloud shadow detection. The mountain shadow algorithm uses the Hillshading technique, to create a hypothetical illumination of a surface by computing slope, aspect and orientation (e.g., Schuckmann 2020, ESRI 2016). A pixel is flagged as mountain shadow if the following condition is met:

$$cos\beta = \cos(\theta_s)cos(slope) + \sin(\theta_s)sin(slope)cos\big(\phi_s - (aspect + orientation)\big)$$
$$\text{Mountain shadow condition: } cos\beta < mntShadowExtent - 1$$

where *mntShadowExten*t is the mountain shadow extent, *slope*, *aspect* and *orientation* are slope, aspect and orientation and have been calculated as described in the corresponding ATBD (Wevers et al., 2022).

eq. 3-7

Mountain shadow extent is set by default to 0.9, the parameter accepts values between 0 and 1, corresponding to no shadow extension at 0 and maximum shadow extension at 1.

The cloud shadow algorithm contains two main parts:

- Step 1: Cloud top height (CTH) estimation
  - CTH: Use O_2 absorption (OLCI) to estimate CTH,
- Step 2: cloud shadow area identification
  - Intersection of line-of-sight with cloud
  - Identify connected cloud shadow areas

Figure 3-4 and Figure 3-5 show the result of the cloud shadow detection for the SENTINEL-3 OLCI data, the clouds are marked in yellow, and the cloud shadows are marked in purple. The quality of detection of the cloud shadows is influenced by the quality of the cloud screening and the quality of the cloud top height estimation. In both images, most of the clouds and cloud shadows are correctly detected. Only a few clouds and the resulting shadows are not identified.

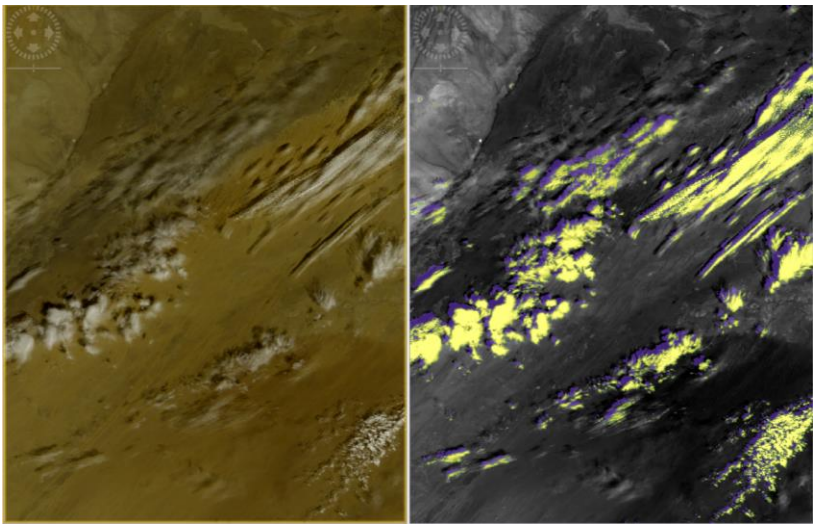

Figure 3-4: Cloud shadow detection for a negative viewing azimuth [7](Eastern edge of swath) – left side RGB image, right side: OLCI band 17 with cloud and cloud shadow mask - clouds are marked in yellow, cloud shadows are marked in purple (SENTINEL-3 OLCI L1b product 2018/10/29, 10:37:44)

[7] *Negative viewing azimuth angle- azimuth angles are expressed to the local North and the so called compass azimuth can be calculated by the difference of 360° and the viewing azimuth with the range of [-180°,180°]*

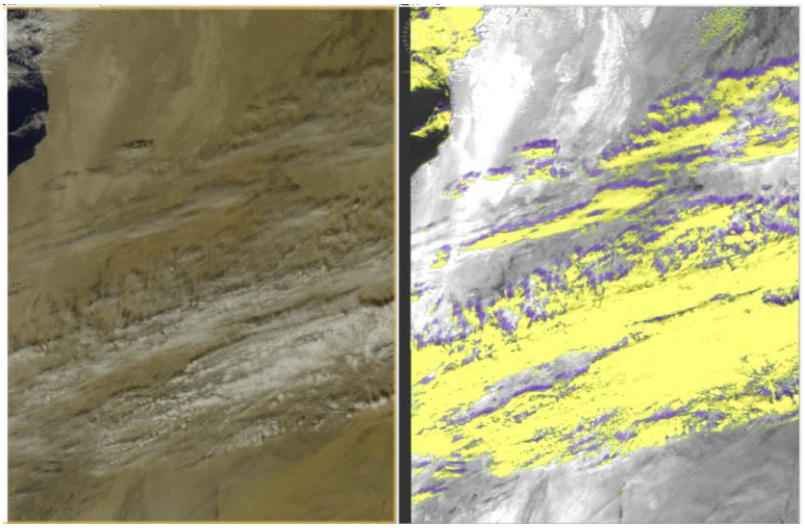

Figure 3-5: Cloud shadow detection for a positive viewing azimuth[8] (Western edge of swath) - – left side RGB image, right side: OLCI band 17 with cloud and cloud shadow mask clouds are marked in yellow, cloud shadows are marked in purple (SENTINEL-3 OLCI L1b product 2018/10/29, 10:37:44)

3.1.6 Atmospheric correction

The atmospheric correction includes the correction for the absorbing and scattering effects of atmospheric gases, in particular ozone, oxygen and water vapour, of the scattering of air molecules (Rayleigh scattering) and the correction of absorption and scattering due to aerosol particles. All components except aerosols can be rather easily corrected. Ozone content can be taken from other satellites, available from met services, and oxygen and water vapour content can be taken from sensor measurements thanks to dedicated spectral bands or can be taken also from meteorological data sets, e.g. ERA Interim (Simmons et al. 2006 & Dee et al., 2011) or ERA5[9] data set (Hersbach et al., 2020). However, aerosols are spatially and temporally highly variable and do not have a distinct spectral absorption feature, so that they cannot be measured easily from merged SENTINEL-3 OLCI/SLSTR measurements.

As described previously the adapted pre-processing chain of MERIS data has been applied to the merged SENTINEL-3 OLCI/SLSTR data. The atmospheric correction including aerosol retrieval is fully documented in GlobAlbedo ATBD (GlobAlbedo_ATBDv4.12, 2013) and here we give a short summary of points relevant to atmospheric correction.

The aerosol correction typically consists of two parts: the calculation of the aerosol properties and secondly the actual correction of the reflectance (after Rayleigh and gas absorption correction). This can be a sequential process where the reflectance correction takes an aerosol optical depth and its spectral dependency as input, or a one-step approach where aerosol properties and reflectance corrections are one implicit step. An approach for retrieval of aerosol properties for MERIS has been

[8] *Positive viewing azimuth angle - azimuth angles are expressed to the local North and the so called compass azimuth can be calculated by the difference of 360° and the viewing azimuth with the range of [-180°,180°]*

[9] *https://confluence.ecmwf.int/display/CKB/ERA5*

developed by the University of Swansea for the GlobAlbedo project (GlobAlbedo_ATBDv4.12, 2013) and is applied to SENTINEL-3 OLCI heritage bands only. It is based on an inverse modelling approach, working on a larger area of 8 x 8 km to estimate the aerosol properties, which is then used to correct every single pixel. This method provides significantly more aerosol retrievals than the LARS method (Santer and Ramon, 2011), which is the baseline for the GlobCover aerosols. The iterative retrieval of atmospheric aerosol requires a fast approximation of atmospheric radiative transfer, to relate top of atmosphere (TOA) to surface reflectance for varying aerosol loading.

To retrieve estimates of aerosol properties from measured satellite radiances, the inverse problem is to be solved to separate the atmospheric and surface scattering contributions to the observed signal. This normally requires some assumptions to be made on the land surface brightness. Within the proposed framework, these assumptions are expressed as constraints defined by error of fit to a parameterised model describing the surface angular or spectral reflectance. For the single view instruments, they apply constraints based on the dark object method (DOM[10]).

The aerosol retrieval includes the calculation of the aerosol optical depth (AOD[11]). Estimates of aerosol extinction are needed for the conversion from top-of-atmosphere measurements to surface reflectance. Aerosol optical depth and aerosol model plus an estimate of the uncertainty in AOD are derived by the Aerosol Retrieval processor from every data set to be processed. It must be noted that AOD and aerosol model are assumed to sufficiently account for the variability in the atmospheric conditions to calculate these terms, while water vapour and ozone column contents are needed in addition to aerosol parameters to retrieve the most accurate Lambertian equivalent reflectance[12]. All other atmospheric constituents are just set to climatology values in the algorithm. This selection is justified by the relatively higher impact of aerosol extinction in the spectral channels of the OLCI instrument, particularly in the visible spectrum.

The problem of surface reflectance and aerosol retrieval can essentially be formulated as one of multivariate optimisation subject to multiple constraints. Figure 3-6 illustrates the retrieval framework followed here. The two-stage optimization process is employed as follows:

(1) Given a set of satellite TOA reflectance and an initial guess of atmospheric profile, the corresponding set of surface reflectance is estimated
(2) Application of the observed set to the estimated set of reflectance results in an error metric, where a lower value of the metric corresponds to a set of surface reflectance (and hence atmospheric profile) that is more realistic

Step (1) is repeated with a refined atmospheric profile until convergence at an optimal solution.

The algorithm components are therefore: (i) design of an efficient and accurate scheme for deriving surface reflectance for known atmospheric profile, and (ii) formulation of constraints on the land

[10] *Fully described in the GlobAlbedo ATBD (GlobAlbedo_ATBDv4.12, 2013)*

[11] https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-aerosol-properties?tab=overview

[12] *Tilstra, L. G., Tuinder, O. N. E., Wang, P., and Stammes, P.: Directionally dependent Lambertian-equivalent reflectivity (DLER) of the Earth's surface measured by the GOME-2 satellite instruments, Atmos. Meas. Tech., 14, 4219–4238, https://doi.org/10.5194/amt-14-4219-2021, 2021.*

surface reflectance suitable for the spectral sampling of the instruments used. The method is applied to estimate aerosol at a coarser spatial resolution (8x8 km) than the underlying surface reflectance, and a subsequent bilinear interpolation step is used to obtain per-pixel values.

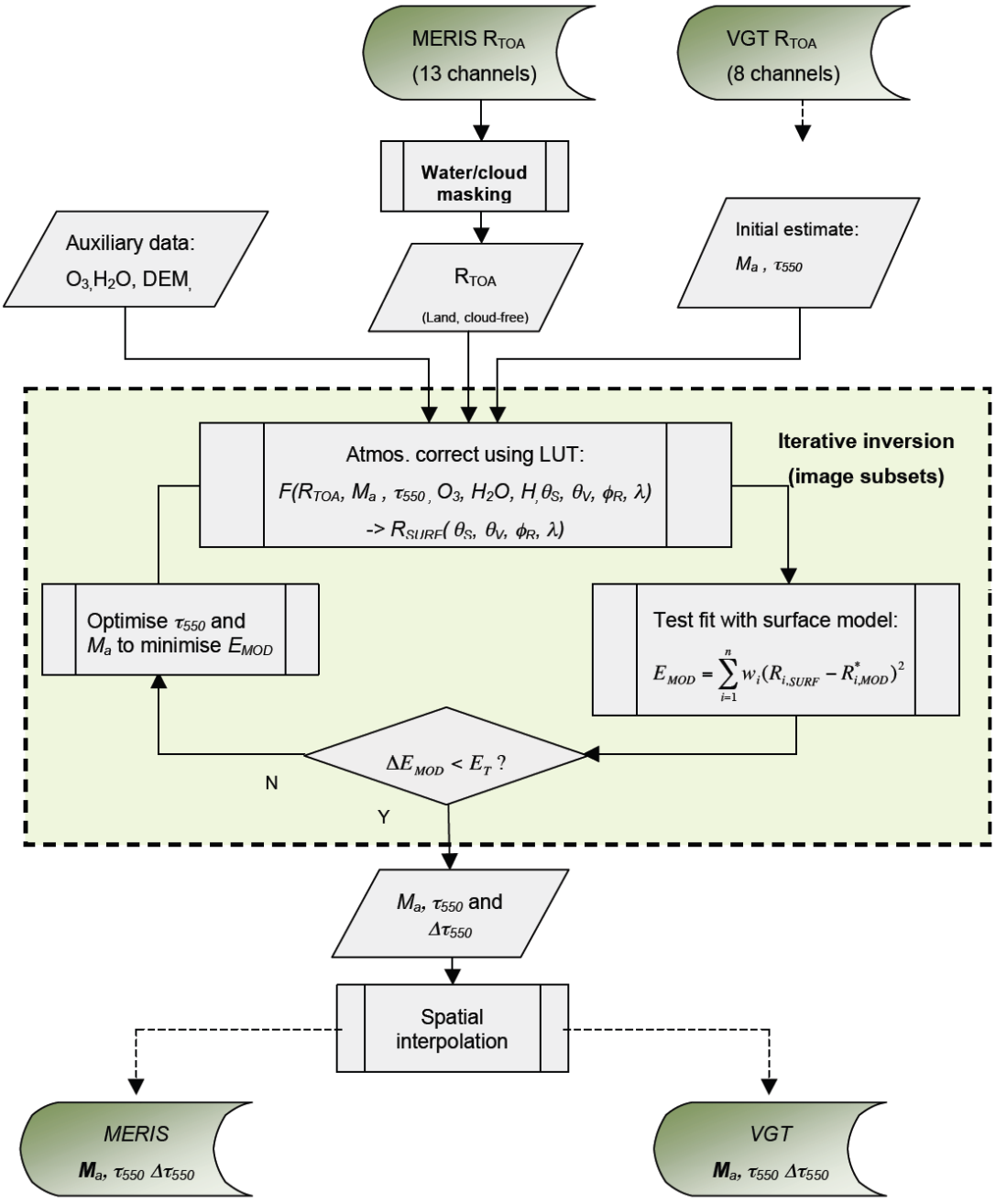

Figure 3-6: Scheme of algorithm for retrieval of aerosol properties, graphic adopted from [GlobAlbedo_ATBDv4.12, 2013]

The GlobAlbedo surface directional reflectance SDR retrieval method was developed by the Institute for Space Sciences, Free University of Berlin, fully documented in GlobCover Design Justification File (GlobCover_DJFv3.8, 2008) and an improved version documented in GlobAlbedo ATBD

(GlobAlbedo_ATBDv4.12, 2013). The algorithm was designed to produce SDR (including uncertainties) from TOA radiance data. The algorithm requires the inputs from the pixel classification and atmospheric condition parameters, e.g. aerosol properties. As mentioned in the previous paragraph the aerosol properties for the merged SENTINEL-3 OLCI/SLSTR products have been retrieved by the aerosol retrieval developed by the University Swansea for the GlobAlbedo project. The TOA reflectance or apparent spectral directional reflectance R_λ^* in the waveband λ of the coupled surface-atmosphere system is related to the radiance L_λ measured by a satellite at the TOA by:

$$R_\lambda^*(\Omega_s, \Omega_v) = \frac{\pi L_\lambda(\Omega_s, \Omega_v)}{\mu_s E_{s,\lambda}}$$

eq. 3-8

where $E_{s,\lambda}$ is the extra-terrestrial irradiance at the time of the measurement. The solar and view vectors are denoted by Ω_s and Ω_v respectively, while μ_s denotes the cosine of solar zenith.

A uniform, Lambertian surface is normally assumed as a basis for the modelling of the atmosphere-surface radiative transfer for operational atmospheric correction algorithms of single-view instruments. Under that assumption, the relationship between top of atmosphere reflectance R_λ^* and the surface directional reflectance R_λ can be approximated by:

$$R_\lambda^*(\Omega_s, \Omega_v) = R_{atm,\lambda}(\Omega_s, \Omega_v) + \gamma_\lambda(\mu_s)\gamma_\lambda(\mu_v)\frac{R_\lambda}{1-\bar{\bar{\rho}}_\lambda R_\lambda}$$

eq. 3-9

Where $R_{atm,\lambda}$ means the atmospheric scattering term (TOA reflectance for zero surface reflectance), γ_λ denotes atmospheric transmission for either sensor to ground or ground to sensor for waveband λ, and $\bar{\bar{\rho}}_\lambda$ denotes atmospheric bi-hemispherical albedo with respect to the surface. The solar and view vectors are denoted by Ω_s and Ω_v respectively, while μ_s and μ_v denote the cosines of solar and view zenith. Spectral directional reflectance R_λ is derived from TOA reflectance by means of the analytical inversion of the previous equation (eq. 3-9) and the corresponding equation is:

$$R_\lambda(\Omega_s, \Omega_v) = \frac{R_\lambda^*(\Omega_s, \Omega_v) - R_{atm,\lambda}(\Omega_s, \Omega_v)}{\gamma_\lambda(\mu_s)\gamma_\lambda(\mu_v) + \bar{\bar{\rho}}_\lambda\left[R_\lambda^*(\Omega_s, \Omega_v) - R_{atm,\lambda}(\Omega_s, \Omega_v)\right]}$$

eq. 3-10

As discussed in GlobAlbedo ATBD/BBDR (GlobAlbedo_ATBD/BBDRv1.0, 2010) and in GlobAlbedo ATBD (GlobAlbedo_ATBDv4.12, 2013), the Lambertian equivalent reflectance R_λ derived from eq. 3-10, which is taken as SDR in the GlobAlbedo processing chain, represents a smoothed version of the surface bi-directional reflectance distribution function (BRDF), with errors up to 15% for turbid atmospheres (Hu et al., 1999).

R_λ^* is calculated from TOA radiance with eq. 3-8 and first corrected from gaseous absorption with a look-up table for a set of input of the geometric airmass factor (AMF), the water vapor column content (CWV) and the ozone column content (OZO). AMF and CWV are considered pixel-wise, while a mean value representative of the imaged area is used for OZO. The atmospheric parameters $R_{atm,\lambda}$, γ_λ and $\bar{\bar{\rho}}_\lambda$ are provided by a look-up table per pixel, with the input view zenith angle (VZA), sun zenith

angle (SZA), relative azimuth angle (RAA), elevation (ELEV) and AOD at 550nm (AOD550) also provided on a per-pixel basis. No correction of adjacency effects (Mekler and Kaufman, 1982; Kaufman, 1989) has been found to be necessary.
The look-up tables intended for aerosol retrieval have been compiled with the Matrix-Operator-Model (MOMO) radiative transfer code (Fischer and Grassl, 1984; Fell and Fischer, 2001). MOMO is a widely accepted radiative transfer code which provides all the features required for atmospheric radiative transfer simulations. The look-up tables intended for gas correction and SDR retrieval are now calculated with radiative transfer package ARTDECO[13] (Compiègne et al., 2013) by HYGEOS instead of the 6S (Second Simulation of the Satellite Signal in the Solar Spectrum) radiative transfer code (Vermote et al., 1997a and 1997b).

The atmospheric corrected merged SENTINEL-3 OLCI/SLSTR product includes the 6 solar reflective SENTINEL-3 SLSTR bands at nadir and stripes A and for all OLCI bands except the spectral bands affected by oxygen (band 13, 14, and 15) and water vapor absorption (band 19 and 20).

Tracking of uncertainties is one of the specific requirements for the C3S-LC processing. The purpose in this part of the processing is to generate a reliable estimation of the errors in the SDR products including the error covariance between the different spectral ranges. Different error sources are considered: instrumental noise, atmosphere and directional reflectance effects. The uncertainty of the SDR values can be propagated as one of the source uncertainties in the LC processing. Alternatively, it can be used in a Monte-Carlo approach, including ensemble building, to determine the output parameter uncertainty in the LC processing.

In the current pre-processing chain, the uncertainty calculation of the surface reflectance follows the approach from North and Heckel (ESA_S3_SYN_ATBD, 2010) and also applied for the calculation of the SENTINEL-3 SYN L2 products (ESA_S3_SYN_ATBD, 2010). The following paragraphs are cited parts of this document:

Error in AOD (ESA_S3_SYN_ATBD, 2010)

"The error in retrieving optical thickness of the aerosols is estimated from the error metric E_{mod} and the curvature of the error surface near the minimum. The optimization procedure determines the minimum E_{min} of $E_{mod}\ (\tau)$, where τ represents the aerosol optical thickness defined at a reference wavelength (550nm). The value τ_{min}, where $E_{mod}\ (\tau_{min}) = E_{min}$, is the resulting aerosol optical thickness. The curvature term (a) of a parabolic fit to the error metric $E_{mod}\ (\tau)$ is calculated near the minimum. This curvature term allows calculation of the uncertainty of the retrieved aerosol optical thickness τ_{min}. This uncertainty can be expressed as:

$$\Delta\tau = k\sqrt{\frac{E_{min}}{a}}$$

eq. 3-11

[13] *https://www.icare.univ-lille.fr/artdeco/*

The curvature term (a) of the error surface gives a measure of the sensitivity of the location of the minimum to error in model fit. The term k is estimated to be 1.58"

Surface reflectance error (ESA_S3_SYN_ATBD, 2010)

"The error in surface reflectance at each waveband is calculated as the sum of errors due to the error in AOD estimation Δ_τ, sensor noise Δ_{sens}, and estimated error in the radiative transfer model Δ_{RT}. For each channel, the uncertainty is given by

$$\Delta R_{surf} = (\Delta_\tau^2 + \Delta_{sens}^2 + \Delta_{RT}^2)^{0.5}$$

eq. 3-12

The uncertainty in surface reflectance due to aerosol error is

$$\Delta_\tau = \frac{\delta R_{surf}}{\delta \tau} \Delta\tau$$

Where $\Delta\tau$ is obtained from (12). An estimate for $\frac{\delta R_{surf}}{\delta \tau}$ at each wavelength is given by linear fit of the variation of R with τ near τ_{min}, using surface reflectance at the same two points as used to find the parabola parameter a. We estimate error in surface reflectance Δ_{sens} due to instrument noise in the TOA measurement Δ'_{sens} using:

$$\Delta_{sens} = \frac{\delta R_{surf}}{\delta \tau} \Delta'_{sens} \approx \frac{\Delta'_{sens}}{T_{O_3}\, T(\theta_s) T(\theta_v)}$$

eq. 3-13

Estimate of the channel-dependent instrument noise Δ'_{sens} should include the combined effects of quantisation and calibration error. The error in the radiative transfer model Δ_{RT} includes the net effect of numerical approximation of atmospheric radiative transfer variation and composition (column ozone, water vapour and aerosol model) from reality. A value of 0.005 should be used as default."

3.1.7 Level3-Processing

The processing steps to retrieve the L3 surface reflectance products from an input set of single satellite observations (i.e., SDR and pixel classification data as derived in the previous sections) are (i) the reprojection of the input products onto a Plate Carrée grid, (ii) the aggregation of the single satellite observations for given binning cells (tiles), and (iii) the mosaicking of the binning cells to a (usually global) Level 3 product with the aggregation results. This processing flow is illustrated in Figure 3-7. The Level 3 product will in return serve as input for the classification algorithms to derive a final landcover map (ESA_CCI_LC_DPM, 2013).

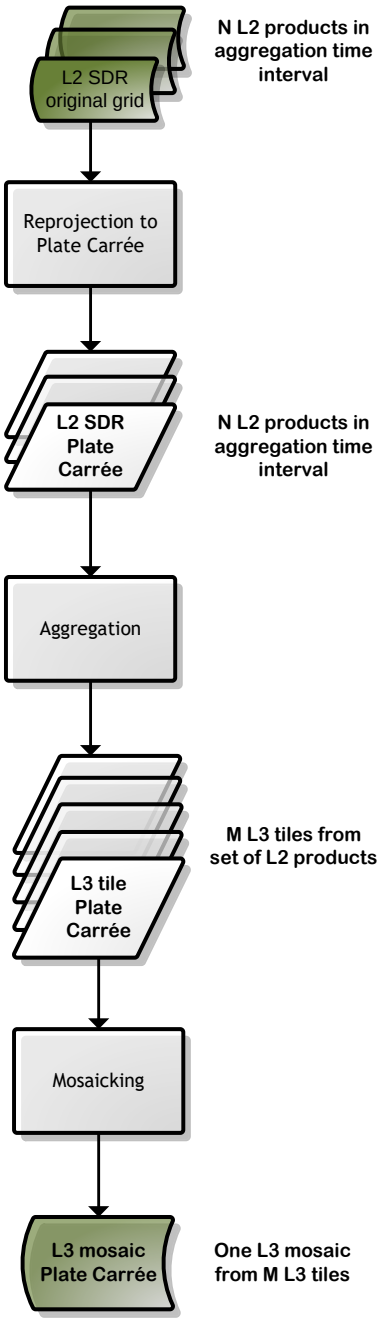

Figure 3-7: Overall logical flow of the CCI-LC L3 processing as the final step of the pre-processing part (from (ESA_CCI_LC_DPM, 2013))

3.1.7.1 Plate-Carrée projection

The Plate-Carrée projection or geographic projection or equirectangular projection, is a very simple map projection that has been in use since the earliest days of spherical cartography. The name is from the French for "flat and square". It is a special case of the equidistant cylindrical projection in which the horizontal coordinate is the longitude and the vertical coordinate is the latitude.

The spherical earth can only be mapped onto a developable surface by allowing distortion, so certain geometric properties on the sphere are not preserved. The Plate-Carrée projection is a cylindrical projection, but, unlike the Mercator projection, the entire sphere, including the poles, can be represented on a finite sized map. The projection is not a conformal map so angles are not preserved.

Because of the distortions introduced by this projection, it has little use in navigation or cadastral mapping and finds its main use in thematic mapping. It has also become a de-facto standard for computer applications that process global maps, such as Celestia[14], because a given co-ordinate is very easily identifiable in an image file.

The following equations (eq. 3-14) describe the mapping of geographic coordinates in terms of latitude φ and longitude λ onto the x and y coordinates of a point on the map. From its latitude φ and longitude λ (with φ_0 and λ_0 being the latitude and longitude in the centre of map) and k being an appropriate scale factor at the equator:

$$x = k(\lambda - \lambda_0)$$
$$y = k(\varphi - \varphi_0)$$

eq. 3-14

For the LC products, the geoid WGS84[15] is used.

3.1.7.2 Temporal sample aggregation and spatial resampling to 1-day surface reflectance composites

The temporal sample aggregations are, for a given pixel in a given aggregation grid cell, a selection of the best observation over all reprojected input data – but separated regarding SENTINEL-3A and SENTINEL-3B, providing best pixel values of SDR and SDR uncertainties, and a status flag. Table 3-1 summarises the types of final status flag and the following pseudo-code describes the temporal sample aggregation. It is worth noting that the selection process of the best pixels does not include a BRDF correction part.

Table 3-1: Final status flag - CURRENT_PIXEL_STATE

FLAG NAME	DESCRIPTION
CLEAR_LAND	pixel was classified as land
CLEAR_WATER	pixel was classified as water
CLEAR_SNOW_ICE	pixel was classified as snow/ice
CLOUD	pixel was classified as cloud or pixel was classified as 'temporal' cloud (a posteriori cloud check within the aggregation scheme)
CLOUD_SHADOW	pixel was classified as cloud shadow

[14] *https://celestia.space/*

[15] *https://earth-info.nga.mil/index.php?dir=wgs84&action=wgs84*

FLAG NAME	DESCRIPTION
INVALID	pixel was classified as invalid or cosmetic[16]

- pixel classification after aggregation (X= LAND, CLOUD etc.)

$$STATUS\_X$$

- surfaces reflectances for N wavelengths of a given sensor after aggregation

$$for\ j = 1 \dots N_\lambda\ and\ N_{obs} \coloneqq\ N_{obs,STATUS_X}\ \ S_{agg} \coloneqq SDR_{MAX(NDVI(j;N_{obs}))}$$

- NDVI[17] after aggregation

$$for\ j = 1 \dots N_\lambda\ and\ N_{obs} \coloneqq\ N_{obs,STATUS_X}\ \ NDVI \coloneqq NDVI_{MAX(NDVI(j;N_{obs}))}$$

- uncertainty of surfaces reflectances for N wavelengths of a given sensor after aggregation

$$for\ j = 1 \dots N_\lambda\ and\ N_{obs} \coloneqq\ N_{obs,STATUS_X}\ \ \triangle S_{agg} \coloneqq \triangle SDR_{MAX(NDVI(j;N_{obs}))}$$

eq. 3-15

The 1-day surface reflectance composites are delivered in Plate Carrée projection, therefore a spatial resampling including super-sampling is applied to SDR data. Because of the difficulty to aggregate categorical data like the status, the selection approach of the contributing pixel part is the same as for the temporal selection process.

3.1.7.3 Temporal sample aggregation for the seasonal surface reflectance products

For the temporal sample aggregation for the SENTINEL-3 OLCI/SLSTR merged seasonal surface reflectance (SR) products based on the SENTINEL-3A and SENTINEL-3B 1-day surface reflectance composites the following approach has been applied:

- mean composite based on all observations which satisfy conditions regarding the Normalised Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI)

The 1-day SR products are the basis for the seasonal composites. The selection of observations for the calculation of the mean values is based on two different selection criteria, which are defined as follows:

[16] *A pixel is classified as invalid, if the pixel observation is marked as erroneous by the L1b flag or the pixel values cannot be processed by the pre-processing chain, e.g. the sun observation geometry is outside the valid range for the aerosol retrieval.*

[17] *Uncertainty has not been considered in the generation of NDVI values; no uncertainty value for NDVI is provided.*

$$if\ (status = LAND\ \ or\ \ CLOUD\ SHADOW)$$

$$\{$$

$$if\ \ (max_{NDVI} - \sigma_{NDVI}\ - \varepsilon \leq NDVI\ and\ \ NDVI\ \leq\ max_{NDVI} + \sigma_{NDVI} + \varepsilon)$$

$$then\ select\ observation\ for\ further\ processing$$

$$\}\ with\ \ NDVI = \frac{\rho_{NIR}\ \ - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$$

$$if\ (status = WATER\ or\ SNOW)$$

$$\{$$

$$if\ \ (max_{NDWI} - \sigma_{NDWI}\ - \varepsilon\ \leq\ \ NDWI\ \ and\ NDWI\ \leq max_{NDWI} + \sigma_{NDWI} + \varepsilon\ )$$

$$then\ select\ observation\ for\ further\ processing$$

$$\}\ with\ \ NDWI = \frac{\rho_{NIR} - \rho_{SWIR}}{\rho_{NIR} + \rho_{SWIR}}$$

where status is the CURRENT_PIXEL_STATE, NDVI is the Normalized Difference Vegetation Index, NDWI is the Normalized Difference Water Index, max_{xx} is the maximum value of XX, e.g. XX= NDVI, σ_{XX} is the standard deviation of XX, e.g. XX= NDVI and ε is a user defined threshold.

eq. 3-16

The temporal sample aggregation sums up all observations, which have fulfilled the conditions, to the intermediate variables and saved different status counts for a pixel. Figure 3-8 and the following pseudo-code (eq. 3-17) describe and illustrate the temporal sample aggregation (ESA_CCI_LC_DPM, 2013].

pixel classification status after aggregation and filtering (X=LAND, CLOUD, etc.)

$$STATUS\_X$$

surface reflectances for wavelengths of given sensor

$$for\ \ j = 1\ \ \ldots N_{\lambda,filter}\ \ \ and\ \ N_{\lambda,filter} := N_{obs,STATUS\_X,filter}$$

$$S_{agg}(j,i) = \overline{SDR_j}(i) := \frac{1}{N_{obs}(i)} \sum_{m=1}^{N_{obs}(i)} SDR_j(i,m)$$

NDVI

$$S_{agg}(N_{\lambda+1},i) = \overline{NDVI}\ (i) := \frac{1}{N_{obs}(i)} \sum_{m=1}^{N_{obs}(i)} NDVI(i,m)$$

where STATUS_X is the CURRENT_PIXEL_STATE after aggregation and filtering and N are the number of observations corresponding to the STATUS_X.

eq. 3-17

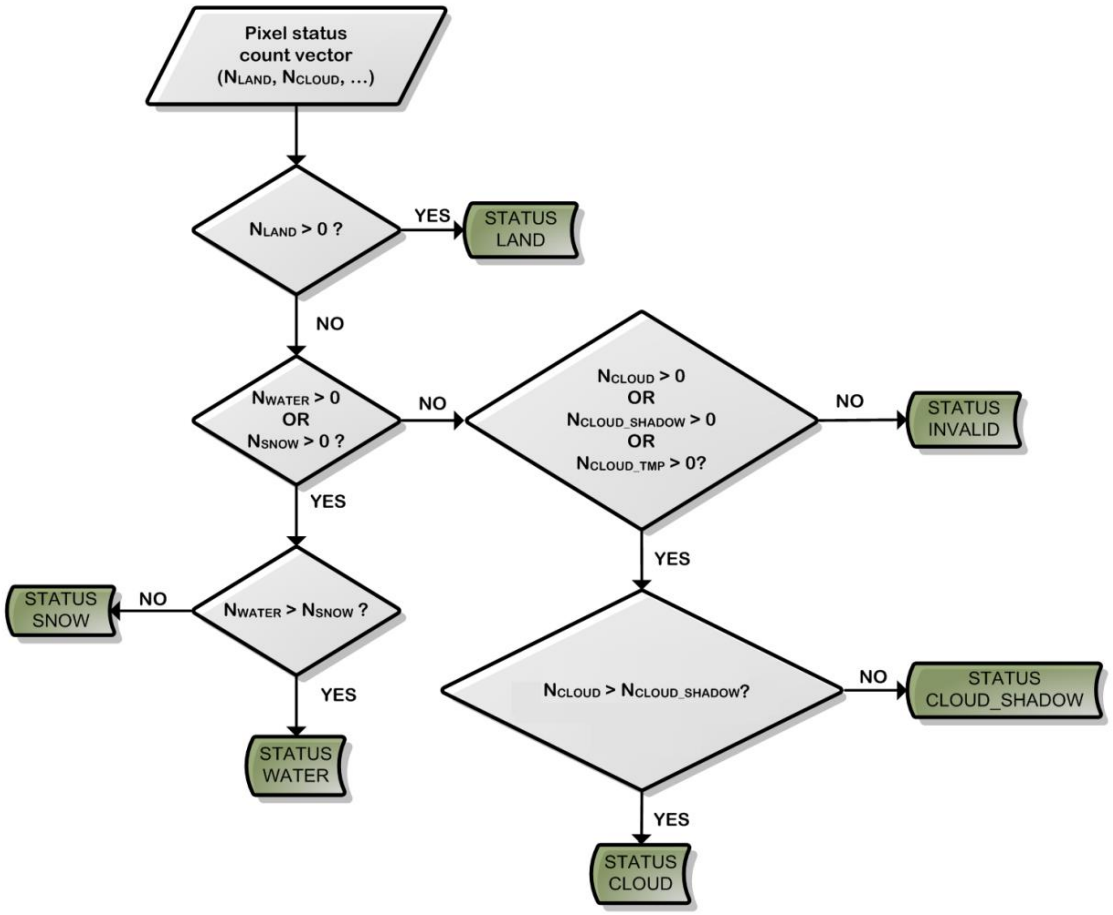

Figure 3-8: Logical flow of the pixel classification after aggregation. N_X (with STATUS_X from the pixel identification) represents the number of single observations of a pixel Pi with status X, the green boxes show the final status (taken from (ESA_CCI_LC_DPM, 2013)).

Per year, 16 different seasonal composites have been processed. The processed seasons with the corresponding start and end date using as an example the year 2020 are presented in Table 3-2.

Table 3-2: Processed seasons using as an example the year 2020 (PxxW- Period of XX Weeks, e.g.P13W means period of 13 weeks)

No.	SEASON	START	END
1	2019-12-03-P13W	03.12.2019	04.03.2020
2	2019-12-03-P17W	03.12.2019	01.04.2020
3	2020-01-01-P22W	01.01.2020	03.06.2020
4	2020-01-01-P52W	01.01.2020	31.12.2020
5	2020-04-02-P13W	02.04.2020	01.07.2020
6	2020-04-16-P7W	16.04.2020	03.06.2020
7	2020-04-16-P13W	16.04.2020	15.07.2020
8	2020-04-16-P17W	16.04.2020	12.08.2020

No.	SEASON	START	END
9	2020-06-04-P13W	04.06.2020	02.09.2020
10	2020-06-18-P15W	18.06.2020	30.09.2020
11	2020-07-16-P7W	16.07.2020	02.09.2020
12	2020-07-30-P13W	30.07.2020	28.10.2020
13	2020-08-13-P13W	13.08.2020	11.11.2020
14	2020-09-03-P13W	03.09.2020	02.12.2020
15	2020-10-01-P13W	01.10.2020	31.12.2020
16	2020-12-03-P13W	03.12.2020	04.03.2021

3.1.7.4 Generation of mosaics

As a final step, all grid cells (tiles) which were part of one or more input products for the aggregation and which were processed as described above are merged into a single global Level 3 result product. Table 3-3 summarises the bands, which are included in the final L3 pre-processing products for merged SENTINEL-3 OLCI/SLSTR data.

The atmospheric corrected merged SENTINEL-3 OLCI/SLSTR product includes the 6 solar reflective SENTINEL-3 SLSTR bands at nadir and stripes A and for all OLCI bands except the spectral bands affected by oxygen (band 13, 14, and 15) and water vapor absorption (band 19 and 20).

No BRDF correction module is currently applied in the pre-processing chain.

Table 3-3: Variables used in Level 3 mosaicking for SENTINEL-3 OLCI

Parameter	Description
SR of band x_{sensor}; $x_{OLCI/SLSTR}$ = 1 … 12, 16 … 18 (OLCI) and 1 … 6 (SLSTR)	surface reflectance of band x_{sensor}; $x_{OLCI/SLSTR}$ = 1 … 12, 16 … 18 (OLCI) and 1 … 6 (SLSTR)
ε_{SR} of band x_{sensor}; $x_{OLCI/SLSTR}$ = 1 … 12, 16 … 18 (OLCI) and 1 … 6 (SLSTR)	uncertainties of the surface reflectance of band x_{sensor}; $x_{OLCI/SLSTR}$ = 1 … 12, 16 … 18 (OLCI) and 1 … 6 (SLSTR)
vegetation_index	Normalised Difference Vegetation Index $NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$
status	current pixel status • invalid • clear land • clear snow/ice • clear water • cloud • cloud shadow
obs_count	number of valid observations over pixel
status_counts of status y; y=land, snow/ice, water, cloud, etc.	number of observations over pixel covered by the corresponding status: invalid, clear land, clear snow/ice, clear water, cloud, cloud shadow

3.1.8 Validation and intercomparison of surface reflectance composites

Validation is needed in order to assess the effectiveness and quality of the different algorithms. Several mathematical and other empirical methods of validation are available. The validation of an algorithm is fairly complex and the validation procedure proposed for the pre-processing algorithms is to run through an iterative process with the following major points:

- Testing the algorithms on (real) remote sensing data for the use cases
- Comparison between independent in situ data and outputs created by the algorithms if possible

The validation and intercomparison plan is provided in the Product Quality Assurance Document (PQAD) (RD-1, RD-3).

3.1.9 Uncertainty

The uncertainty is calculated and included in the 1-day surface reflectance composites as described in the previous sections. A low uncertainty associated with surface reflectance indicates that the values are spread around the mean tightly and are close to the true value if no unknown systematic errors are existing and known correction are applied. At this stage of the project, the uncertainty is not propagated to the seasonal composites. But, the quality of each global multispectral seasonal SR composite is described, on a per-pixel basis, by a set of flags and values:

- number of observations, for each pixel;
- current status of surface associated with the surface reflectance in the aggregation period, for each pixel;
- number of observations associated with the selected status, for each pixel;

This could enable the users to evaluate the pre-processing products for their own applications or to integrate the information about uncertainties in the processing chains.

3.2 The CCI land cover processing chain

The classification chain developed in the ESA CCI LC project to generate the 1992 - 2015 global LC map series is applied in the C3S framework to produce the LC maps from 2016 to 2022 (Figure 3-9). The chain is organised into two main processes: (i) the generation of a baseline global LC map using the entire archive of the Envisat Medium Resolution Imaging Spectrometer (MERIS) data and (ii) the generation of global annual LC maps from this baseline product.

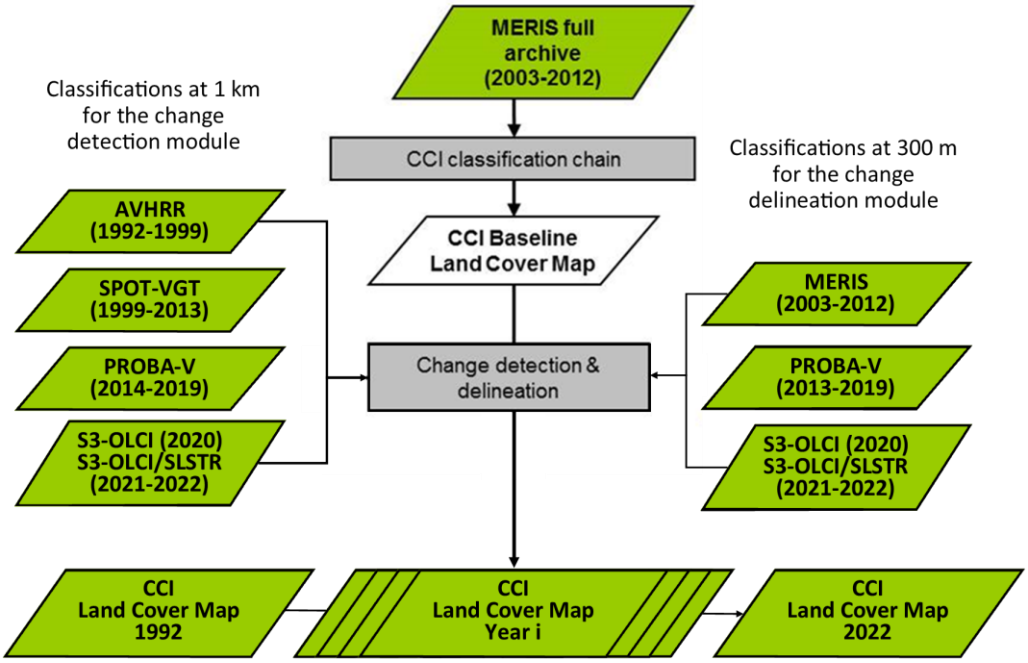

Figure 3-9: Schematic representation of the CCI LC classification chain made of 2 main processes to generate global annual LC maps using the entire archives of Envisat MERIS, AVHRR time series between 1992 to 1999, SPOT-VGT time series between 1999 and 2013, PROBA-V data for 2014 – 2019[18] ,SENTINEL-3 OLCI for 2020 and SENTINEL-3 OLCI & SLSTR for 2021 –2022. Updated from [ESA_CCI_LC_ATBD, 2017a].

The CCI LC baseline classification module aims at generating a global LC map representing the LC remaining stable over time (CCI Baseline Landcover Map – Figure 3-9). Then, for each year of the map series, two global LC classifications are produced at 1 km and 300 m using the change detection and delineation modules, respectively. The CCI baseline LC map is then updated with the LCC detected at 1 km, then refined at 300 m from 2004 onwards to generate the LC map of the year of interest.

The entire archives of Envisat MERIS (2003-2012), AVHRR (1992-1999), SPOT-VGT (1999-2013), PROBA-V (2014 – 2019), SENTINEL-3 OLCI (2020) and SENTINEL-3 OLCI & SLSTR (2021 –2022) were used to generate the global ESA CCI/C3S LC map series 1992-2022. The following sections describe the generic methods of LC classifications, LCC detection, and delineation applied to the sensors mentioned above.

[18] *The left hand-side concerns data at 1 km. As SPOT_VGT was available up to 2013, we could use PROBA-V starting 2014. On the right hand side, MERIS ends in 2012 so that we had to use PROBA-V partially in 2013 to have less data gaps for the Land cover change delineation at 300m.*

3.2.1 Land cover classification modules

3.2.1.1 Logical model

The classification process transforms the Level 3 (L3) seasonal surface reflectance composites at 1 km and 300 m produced by the pre-processing step (section 3.1) into global LC maps at 1 km and 300 m. The 1 km maps are used in the change detection module (Figure 3-9, left-hand side) while the 300 m LC maps are used in the delineation module (Figure 3-9, right-hand side). AVHRR data are available at 1 km only. The 300-m S3-OLCI data were resampled to 1 km using an average resampling.

The LC baseline and annual global LC products representative of each year were obtained through two parallel classifications, namely a machine learning and an unsupervised classification. The outputs from the machine learning and unsupervised classifications were then merged to generate the baseline LC map (see section 3.2.2). The three significant processing steps to generate the baseline (noted as Step 1, Step2 and Step 3 in Figure 3-10) rely on the reference LC layer and the stratification layer inherited from the CCI LC project (section 2). Step 1 and 2 outputs directly feed the LCC module to generate the global annual LC maps (see section 3.2.3).

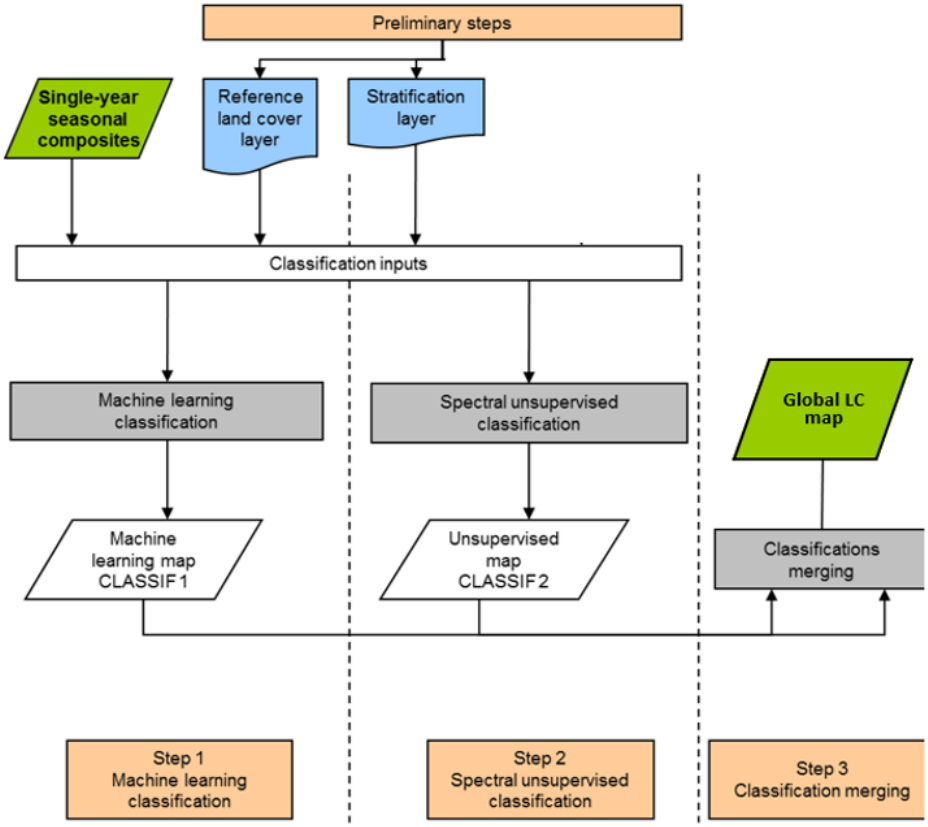

Figure 3-10: Schematic representation of the classification process developed to generate the 300-m baseline LC product (steps 1 – 3) using the entire archives of Envisat MERIS. Step 1 and 2 outputs are used to generate the global LC products at 1 km and 300 m. Source data vary according to the year being processed: AVHRR time series between 1992 to 1999, SPOT-VGT time series between 1999 and 2013, PROBA-V data for 2014 – 2019, SENTINEL-3 OLCI for 2020, SENTINEL-3 OLCI & SLSTR 2021–2022.

In the first and second steps of the classification, machine learning and unsupervised classification algorithms are run using the spectral properties of seasonal composites as input, resulting in two different maps, "CLASSIF1" and "CLASSIF2". In the third step, these two maps are merged to produce a global LC map.

The different algorithmic stages are presented in detail below. For each stage, the algorithm rationale is presented, inputs and outputs are identified, algorithms parameters and pseudo-codes are provided. In general, classification chain parameters related to each stratum are stored in Look-Up Tables (LUT), which are included throughout the document.

3.2.1.2 Stage 1: machine learning spectral classification

The machine learning classification is a supervised classification approach. Any supervised classification procedure relies on the following steps:

- Gathering training data (i.e. spectral signatures for each LC class of interest) as representative of the "real-world" as possible;
- Determining the type of the classifier to use – the classifier corresponding to the function which (i) analyses the training data and (ii) predicts the output class of any input pixel by generalising the training data to "unseen" situations;
- Defining the classification parameters to optimise the algorithm's performance.

Two refinements have been introduced to improve the collection of the training data set compared to traditional approaches. Firstly, the training data is defined locally, taking into consideration that the spectral signature of a specific land cover class may not be consistent across large areas. It is important to acknowledge that at larger scales each land cover label may correspond to multiple spectral signatures. For example, different spectral signatures may exist for crops such as maize and wheat, even though they will be both labelled as crops. To address this variability, a locally-adjusted training data set collection and classification approach has been developed.

In this approach, each stratum from the stratification layer (section 2.2.3) to be classified is divided into smaller moving windows that represent the classification areas. The size of these classification areas varies depending on the specific stratum. The regions in which the training data set is gathered are referred to as search areas. These search areas are 240*240 km² regions centered on each classification area. By decoupling the classification areas from the search areas, the approach considers any imperfections in the auxiliary reference data set used to collect training samples. For example, if a particular land cover class is not adequately represented within a classification area, it can still be included in the training data thanks to the larger search extent. Additionally, the overlap of the search areas ensures a smooth transition of the training data, avoiding artefacts at the boundaries between two classification areas. Figure 3-11 provides illustrations and demonstrates the relationships between the strata, classification areas, and search areas.

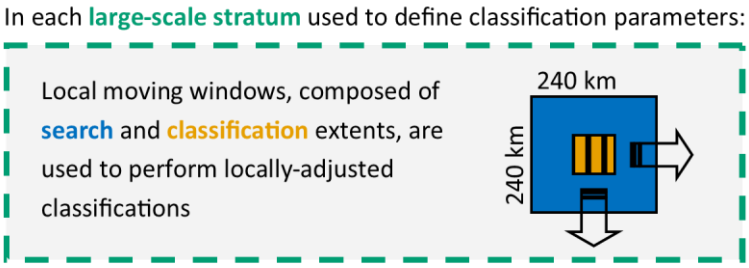

Figure 3-11: Illustration of the three spatial units (stratum (dashed green rectangle), classification (orange central square with vertical stripes) and search areas (blue larger square)) of the developed locally-adjusted supervised classification procedure.

It should be clearly stated that these classification and search areas are entirely distinct from the strata defined in the stratification phase. Strata allow the classification parameters such as the spectral bands and compositing periods to be adjusted, while classification and search areas allow training data for a local classification to be selected. The preparation of such training data is achieved in the ***first preliminary step*** of the supervised spectral classification procedure (see section 3.2.1.2.1). The second training data set refinement consists of combining the representation of each LC class through spectral signatures with a priori information about the occurrence probability of each LC class. The a priori information is extracted from an auxiliary reference data set at the spatial scale of the search areas. Indeed, while the local classification strategy allows accounting for local specificities, it also makes the algorithm more sensitive to the spatial inconsistency of the training data set. The occurrence probabilities computation at an intermediate scale is performed in a ***second preliminary step*** of the supervised spectral classification procedure (see section 3.2.1.2.2).

As for the machine learning classification algorithm itself, it relies on the classical maximum likelihood principle. The algorithm and associated parameters are described in section 3.2.1.2.3, along with the entire procedure.

3.2.1.2.1 Preliminary step 1 – Training data set preparation

The training data set provides representative spectral signatures for each LC class of interest for the 240*240 km^2 search areas centred on the classification areas. It is derived from the reference LC database (section 2.2.2). This reference database consists of global, regional, and local reference LC maps selected as the most accurate ones for a given region, with the highest spatial resolution and an LCCS-compatible legend. Its spatial resolution is 20 m.

The workflow of this preliminary step for training data set preparation is provided in Figure 3-12. It is organised in three sub-steps:

- The first step consists of aggregating and re-sampling the reference LC layer from 20 m to 300 m. In order to aggregate the data, a set of decision rules are used that associate proportions of the 20-m LC labels to reference LC labels in the resulting 300 m x 300 m pixel;
- The second step consists of applying a morphological filter, which is based on the erosion principle, to the aggregated 300m reference LC layer in order to ensure the training data set is as "pure" as possible (i.e. not contaminated by "border effects"). LUT1 contains the list of IDs of the LC classes on which the morphological filter is applied;

- The third step consists of extracting, for each eroded LC class, representative spectral signatures at the search area level.

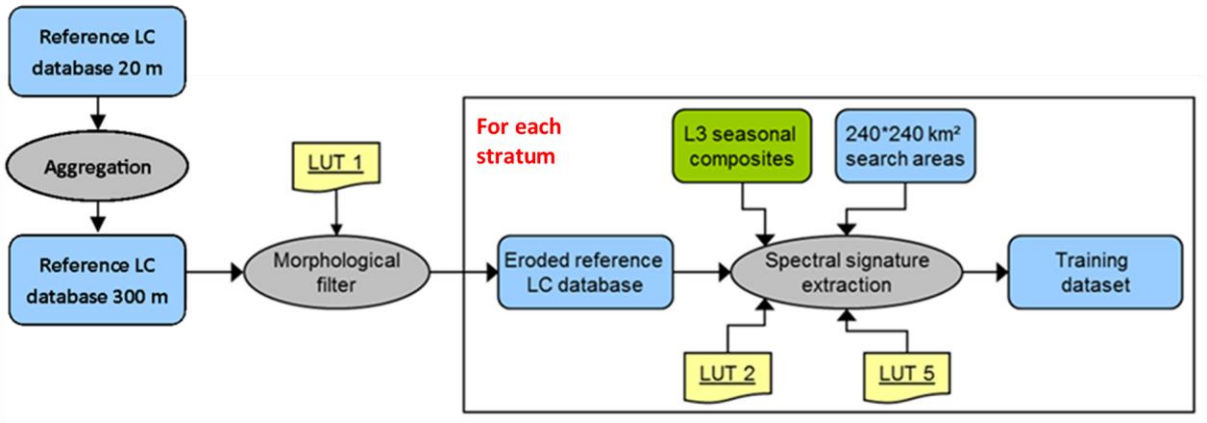

Figure 3-12: Activity diagram illustrating the training data set preparation: (1) the reference LC database is aggregated to 300 m, (2) the reference LC database is eroded via a morphological filter and (3) spectral signatures are extracted for each eroded class.

The aggregation step (first step) is the process of resampling the reference LC database from 20 m to 300 m. It relies on pre-defined rules that associate proportions of the 20-m LC labels to reference LC labels in the resulting 300 m x 300 m pixel. Those rules aim at mimicking the proportions of LC types that form the definition of the global LC map itself. The underlying hypothesis of this aggregation step is that each 20-m pixel is labelled according to the LCCS legend with minimum ambiguity between similar classes.

The morphological filter (second step) is based on the erosion principle, i.e. removing pixels along the boundaries of each LC class. However, eroding each class with the same number of pixels would erase the thin classes from the reference. To address this issue, the erosion process was designed to preserve at least one pixel from each group of adjacent pixels belonging to the same class. The morphological filter selectively retains only the pixels that have the highest number of neighbouring pixels from the same class. A fixed neighbourhood size of 3*3 pixels was used consistently across all strata. The filter operates in two passes. In the first pass, the number of pixels belonging to the same class as the central pixel is counted (as illustrated in Figure 3-13 B). In the second pass, the label of the central pixel is erased if it has fewer neighbours from the same class than another pixel in the neighbourhood (as demonstrated in Figure 3-13 D). This approach ensures that isolated pixels in the reference layer are retained if there is no larger group of pixels from the same class in the neighbourhood. This is in contrast to a standard erosion filter, which would remove such isolated pixels (as shown in Figure 3-13 C).

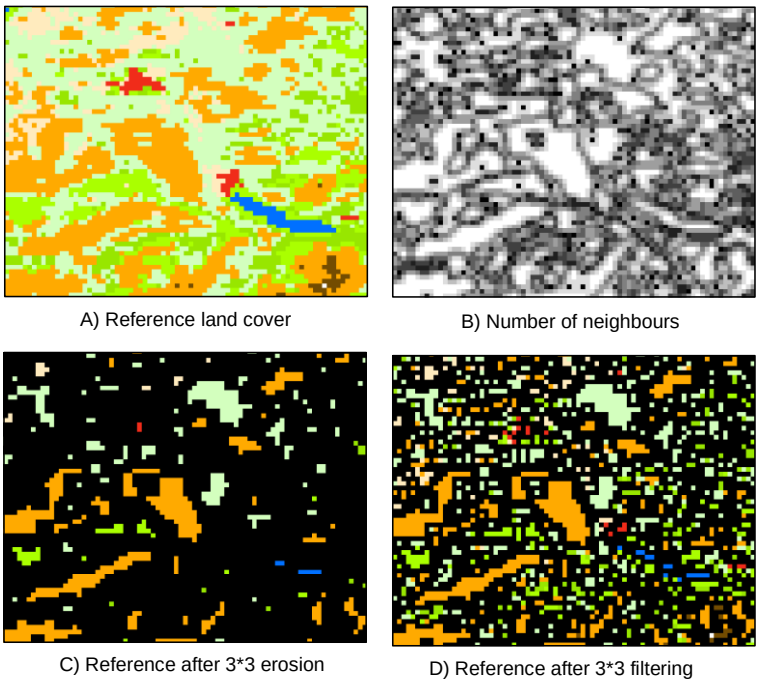

Figure 3-13: Result comparison between two filtering techniques: morphological erosion (D) and majority neighbors filtering (C), both applied on a 3 by 3 pixel window of the reference land cover (LC) layer aggregated at 300 meters (A). The morphological erosion approach (D) is designed to selectively retain pixels that have the highest number of neighboring pixels from the same class (B). This enables the preservation of pixels belonging to LC classes that are sparsely represented in the data set. Zoom on a rural landscape.

This composite filter results in a new raster file that contains a new eroded reference LC layer. It is then used to extract spectral signatures for each class and build a training data set.

This spectral signature extraction (third step) is operated at the 240*240 km^2 search area level while running with the generic parameters of the machine learning algorithm, defined at the stratum level. LUT 5 encompasses the relevant parameters for each stratum, including the L3 seasonal composites and spectral bands required for extracting spectral signatures. Within each search area, the algorithm uses the specified seasonal composites and spectral channels to extract reflectance values for each class. These reflectance values are then used to define representative spectral distributions, which form the training data set.

The LUT 5 contains the following fields:

- **NB_ST**, which lists the numbers (IDs) ranging between 1 and 22 and corresponding to each stratum;
- **Startdate_Season(i)**, specifying the exact date which marks the start of the seasonal composite i (i ranging from 1 to 3 according to the stratum);
- **Enddate_Season(i)**, specifying the exact date which marks the end of the seasonal composite i (i ranging from 1 to 3 according to the stratum);
- **Channels**, spectral bands to be used as input for the machine learning classification algorithm.

The input and output data associated with applying the aggregation of the reference LC database, the morphological filter and the generation of the training data set (through the extraction of a spectral signature by class) are described in Table 3-4,
Table 3-5 and
Table 3-6, respectively.

Table 3-4: Input and output data of the 1st preliminary step of the supervised spectral classification for the aggregation of the 20-m reference to 300 m.

Data	Description	(in, out, in-out)	Physical unit	Range
Ref20_LC	Global reference LC database at 20 m, where each pixel is associated with a LC class of the LCCS CCI-legend through an ID	IN	None	[0 … 255]
LUT 1	Look-up table describing the ID of each LC class	IN	/	/
Ref300_LC	Global reference LC database aggregated at 300 m, where each pixel is associated with a LC class of the LCCS CCI-legend through an ID	OUT	None	[0 … 255]

Table 3-5: Input and output data of the 1st preliminary step of the supervised spectral classification, for the application of the morphological filter.

Data	Description	(in, out, in-out)	Physical unit	Range
Ref300_LC	Global reference LC database at 300 m, where each pixel is associated with a LC class of the LCCS CCI-legend through an ID	IN	None	[0 … 255]
LUT 1	Look-up table describing the ID of each LC class	IN	/	/
Ref_LC_Training	Reference LC database, processed with the morphological filter. Each pixel is associated with a LC class of the LCCS CCI-legend or with a no data value through an ID	OUT	None	[0 … 255]

Table 3-6: Input and output data of the 1st preliminary step of the supervised spectral classification, for the spectral signature extraction

Data	Description	(in, out, in-out)	Physical unit	Range
L3_StartDate_EndDate_SR_<n> n = 1,..,15 (MERIS)	Global raster of surface reflectance in each band of the seasonal composites GeoTiff format	IN	None	[0 … 1]
Ref_LC_Training	Reference LC database, processed with the morphological filter. Each pixel is associated with a LC class of the LCCS CCI-legend or with a no data value through an ID	IN	None	[0 … 255]
GRID_240	Raster grid for each stratum defining the 240*240 km^2 search areas localisation	IN	None	[0 … ~60000]
LUT 2	Look-Up-Table describing the location of the 22 strata	IN	/	/
LUT 5	Look-Up-Table describing the parameters of the machine learning algorithm	IN	/	/
ROI_<NB_LAB>_<GRID>_StartDate _EndDate_SR - NB_LAB representing each class of the LCCS legend - GRID representing the search area - StartDate and EndDate defining the source seasonal composite	The pure training data set for each class ("NB_LAB") and each search area ("GRID"), consisting of representative reflectance values distributions in specific seasonal composites (defined by "StartDate" and "EndDate")	OUT	None	[0 … 1]

Table 3-7 provides the parameters needed to run the morphological filter. The parameters needed to extract the spectral signatures correspond to the generic ones defined for the machine learning algorithm and are thus contained in the LUT 5.

Table 3-7: Parameter needed in the 1st preliminary step of the supervised spectral classification and the application of the morphological filter.

Parameters	Description	(in, out, in-out)	Format	Range
k	Size of the window inside which the filter is applied	IN	Short	[0 … 255]

3.2.1.2.2 Preliminary step 2 – Computation of land cover class occurrence probabilities at the stratum level

As already mentioned in section 3.2.1.2, the machine learning classification is locally adjusted: the algorithm is run within classification areas, using a training data set collected within 240*240 km^2 search areas. In order to minimise the sensitivity of the algorithm to the possible spatial inconsistency of the training data set, the algorithm also uses a priori occurrence probabilities for each LC class defined at the spatial scale of the search areas.

The a priori occurrence probability of each LC class is computed inside each search area through a comparison with the reference LC database. These occurrences will serve as a priori information for the machine learning algorithm and complement the classical training data set made of spectral signatures.

Table 3-8 presents the input and output data associated with the step of LC class occurrence computation.

Table 3-8: Input and output data of the 2nd preliminary step of the supervised spectral classification for the LC class occurrence computation

Data	Description	(in, out, in-out)	Physical unit	Range
REF_LC	Reference LC layer where each pixel is associated with a LC class through an ID	IN	None	[0 … 255]
GRID_240	Raster grid for each stratum defining the 240*240 km2 search areas localisation	IN	None	[0 … ~60000]
LUT 1	Look-Up-Table describing the CCI LCCS LC legend	IN	/	/
OCC_<NB_LAB>_<GRID> - NB_LAB representing each class of the LCCS legend - GRID representing the search area	The pure training data set for each class ("NB_LAB") and each search area ("GRID"), consisting of representative reflectance value distributions in specific seasonal composites (defined by "StartDate" and "EndDate")	OUT	None	[0 … 1]

3.2.1.2.3 Machine learning spectral classification

This section describes the machine learning algorithm and the classifier which predicts the output class of all pixels.

The algorithm makes use of the most common classifier, which is the Maximum Likelihood (ML) and relies on a statistical approach. The ML classifier assumes that a multivariate normal distribution can describe each spectral class. Therefore, the ML algorithm takes advantage of both the mean vectors and the multivariate spreads of each class and can also identify elongated classes. More precisely, the Probability Density Function (PDF) of each class is estimated under a Gaussian assumption. It is relatively rapid and straightforward, but it assumes that there is only one population per class (i.e. a

unique spectral signature by LC class). This assumption seems realistic since the algorithm is applied at a local scale below 240 by 240 km².

ML classification is a statistical decision criterion to assist in classifying overlapping signatures; pixels are assigned to the class of highest probability.

As already mentioned, the supervised algorithm is run with parameters defined at a larger scale while being applied at a more local scale. On the one hand, the input EO data (i.e. seasonal composites in a definite number of spectral channels) are defined at the ***stratum*** level. On the other hand, the algorithm is processed via moving ***classification areas*** using (i) training data valid within 240*240 km^2 ***search areas*** centred on the classification areas and (ii) a priori LC class occurrence probabilities defined at the same scale as the ***search areas***.

Classification areas inside which the machine learning algorithm is run are not the same for all strata. This information is included in the LUT 6 (Table 3-9), which includes the following fields:

- **NB_ST**, which indicates the number of the stratum;
- **CLASSIF_AREA,** specifying the size (in km*km) of the areas where the ML algorithm is running

Table 3-9: Parameters associated with the classification areas of the machine learning algorithm (contained in LUT 6)

NB_ST	Classif_area (km * km)	NB_ST	Classif_area (km * km)
1	120 *120	12	60 *60
2	120 *120	13	60 *60
3	120 *120	14	120 *120
4	120 *120	15	60 *60
5	120 *120	16	120 *120
6	60 *60	17	120 *120
7	120 *120	18	120 *120
8	120 *120	19	120 *120
9	120 *120	20	120 *120
10	120 *120	21	120 *120
11	60 *60	22	120 *120

The principle of this machine learning classification strategy is illustrated in Figure 3-14.

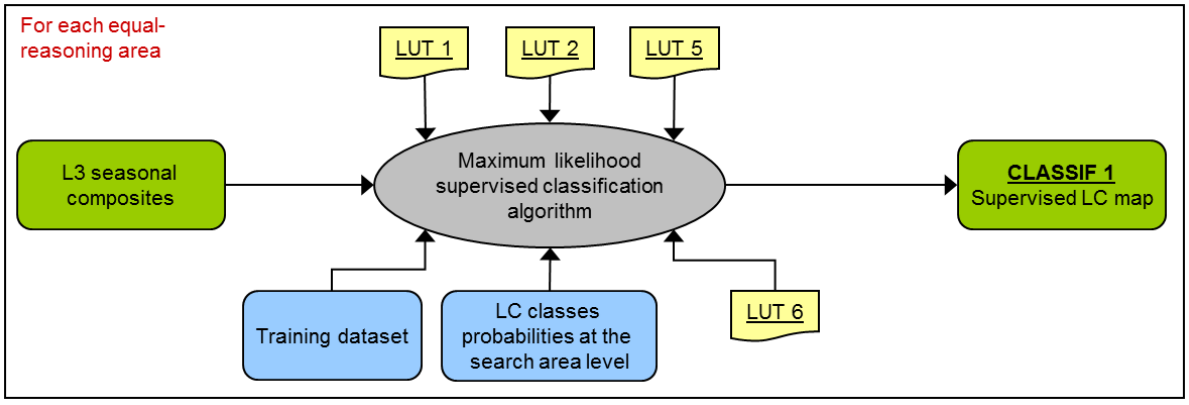

Figure 3-14: Activity diagram illustrating the Gaussian Maximum Likelihood supervised classification strategy developed in the CCI-LC project

The machine learning classification algorithm is applied on L3 seasonal composites independently for each stratum (NB_ST).

The principle of the maximum likelihood classification relies on the Bayes theorem. In the context of the supervised classification of remote sensing data, the likelihood of a LC class for a given set of spectral values can be derived from the a priori probability and the distribution of the spectral values for this class. For each pixel, a label is assigned to the class with the highest probability. It is worth noting that the highest probability is stored together with the label to be used as a quality flag.

As an output, the machine learning classification algorithm creates, for each stratum (NB_ST), a raster file (CLASSIF1) where each pixel is associated with a LC class through the NB_LAB identifier. It also produces an auxiliary layer providing the classification probability associated with each pixel (CL1_PROB). This probability informs about the confidence in the classification output.

Local classification could lead to tiling artefacts because the classifiers change for each classification area. However, this is mitigated by using a search area larger than the classification one and a continuous bilinear interpolation of the a priori probabilities. While the classifier's training based on spectral signature is locally based, the a priori brings some consistency, and the classifier's parameters are seamlessly changing. The size of the classification area was set by tuning the method to remove tiling artefact with optimal processing time. In addition, the tiling artefacts can also be avoided by shifting the starting point of the classification areas year after year.

The Gaussian ML approach underlying the supervised classification algorithm assumes that a Gaussian distribution can describe each spectral class. It assumes that classes in the input data have a Gaussian distribution and that signatures were well selected; this is not always a safe assumption.

Input and output data associated with this machine learning algorithm are described in Table 3-10.

Table 3-10: Input and output data of the step 1a of the classification chain, i.e. the spectral machine learning classification

Data	Description	(in, out, in-out)	Physical unit	Range
L3_StartDate_EndDate_SR_<n> n = 1,..,15 (MERIS)	Global raster of surface reflectance in each band of the seasonal composites GeoTiff format	IN	None	[0 … 1]
ROI_<NB_LAB>_<GRID>_StartDate _EndDate_SR - NB_LAB representing each class of the LCCS legend - GRID representing the search area - StartDate and EndDate defining the source seasonal composite	The pure training data set for each class (“NB_LAB”) and each search area (“GRID”), consisting of representative reflectance value distributions in specific seasonal composites (defined by “StartDate” and “EndDate”)	IN	None	[0 … 1]
OCC_<NB_LAB>_<GRID> - NB_LAB representing each class of the LCCS legend - GRID representing the search area	The pure training data set for each class (“NB_LAB”) and each search area (“GRID”), consisting of representative reflectance value distributions in specific seasonal composites (defined by “StartDate” and “EndDate”)	IN	None	[0 … 1]
LUT 1	Look-up table describing the ID of each LC class	IN	/	/
LUT 2	Look-Up-Table describing the location of the 22 strata	IN	/	/
LUT 4	Look-Up-Table indicating which strata require the generation of multi-year seasonal composites	IN	/	/
LUT 5	Look-Up-Table describing the parameters of the machine learning algorithm	IN	/	/
LUT 6	Look-Up-Table providing the classification areas for each stratum	IN	/	/
CLASSIF1	Land cover map resulting from the supervised classification algorithm, where each pixel is associated with a LC class through an ID	OUT	None	[0…255]
Code_CLASSIF1	Classification probability associated with the label selected for each pixel	OUT	None	[0… 1]

The classifier can be tuned to adjust the size of the processing window, the size of the search window and the weight (W) of the a priori information for the spectral information.

No specific equations need to be implemented.

3.2.1.3 Stage 2: Detailed processing scheme of the unsupervised spectral classification

The use of an unsupervised classification algorithm allows a high degree of automation to be achieved while reducing the processing time. These advantages were successfully demonstrated in the GlobCover (ESA Data User Element (DUE) project) project [GlobCover 2005 and Arino et al., 2008].

Unsupervised image classification is based solely on image statistics, without the availability of training data or other a priori knowledge of the area. The unsupervised algorithm relies on the clustering principle, which assigns pixels with similar reflectance values into clusters. The basic premise is that reflectance values within a given LC class should be close together in the measurement space, whereas pixels belonging to different LC classes should be comparatively well-separated (Lillesand et al., 2000).

The unsupervised classification algorithm used in this project relies on the Iterative Self-Organizing Data Analysis Technique (ISODATA) clustering technique, which represents each cluster by a single mean vector [GlobCover 2005 and Arino et al., 2008]. More detail about this algorithm (and about all the classification procedure) is given hereafter in section 3.2.1.3.1.

Directly linked with the unsupervised algorithm is the labelling procedure, which aims at transforming the spectral clusters into LC classes. The unsupervised algorithm indeed generates spectrally separable clusters, for which the LC label is not known. The next step thus compares clusters with a LC reference data set. This procedure is detailed in section 3.2.1.3.2.

3.2.1.3.1 Unsupervised ISODATA algorithm

This section describes the unsupervised classification algorithm used in the project, relying on the ISODATA clustering technique. The ISODATA algorithm is an iterative optimization clustering procedure, also called the migrating means technique. It is based upon estimating some reasonable assignment of the pixel vectors into candidate clusters and then moving them from one cluster to another in such a way that the Sum of Squared Error (SSE) is progressively reduced.

The following set of basic steps implements the algorithm:

- First, the procedure starts by randomly selecting C points in the multidimensional input data space that will serve as candidate cluster centres: $\mathbf{m}_i$, i = 1 … C. These centers are represented as circles in Figure 3-15;
- Second, each pixel in the image (or segment of image) to classify is assigned to the nearest candidate cluster. These pixels are represented as blue stars in Figure 3-15. This assignment is based on the minimization of the Euclidean distance function between that pixel and the candidate cluster centres $\mathbf{m}_i$;
- Third, after each iteration, the new set of means that result from the grouping produced in the previous step are computed;
- Fourth, the entire process is repeated. At each iteration, a new mean is calculated per cluster based on the actual spectral locations of the pixels. Then, these new means are used for defining clusters in the next iteration. The process continues until there is little change between iterations, i.e. if the normalised percentage of pixels whose assignments are unchanged since the last iteration reaches a convergence threshold, or until the maximum number of iterations is reached.

The ISODATA principle is illustrated in Figure 3-15.

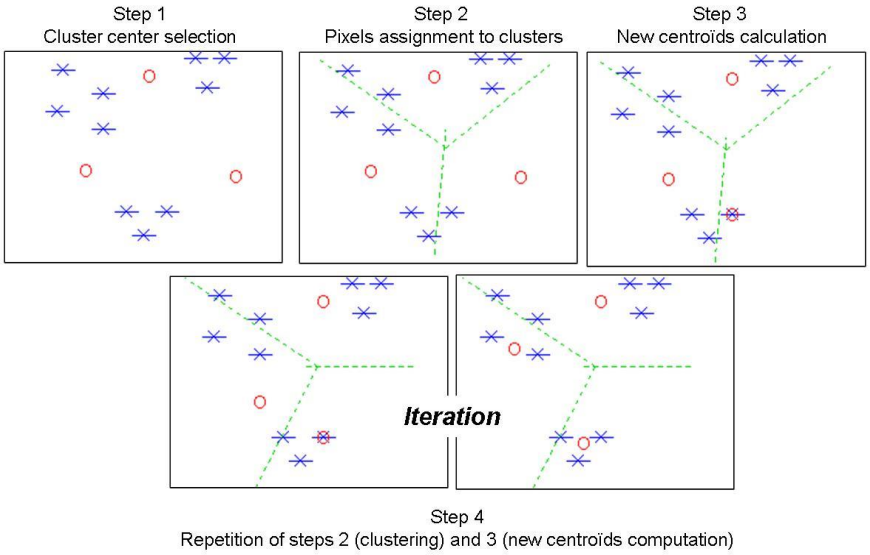

Figure 3-15: Principle of the ISODATA clustering technique. Stars represent the pixels to be classified, while circles represent the centers of the clusters.

The ISODATA algorithm is applied to L3 seasonal composites independently for each stratum (NB_ST). For each stratum, specific compositing periods and spectral channels were selected. These parameters are included in the LUT 7 (Table 3-11). As an output, the ISODATA algorithm creates, for each stratum (NB_ST), an output raster file (L4_<NB_ST>_Clusters_Spectral) where each pixel is associated with a spectrally homogeneous but unlabelled cluster (NB_Cluster).

The LUT 7 contains the following fields:

- **NB_ST**, which lists the numbers (IDs) ranging between 1 and 22 and corresponding to each stratum;
- **Startdate_Season(i)**, specifying the exact date which marks the start of the seasonal composite i (i ranging from 1 to 3 according to the stratum);
- **Enddate_Season(i)**, specifying the exact date which marks the end of the seasonal composite i (i ranging from 1 to 3 according to the stratum);
- **Channels**, specifying the spectral channels to use as input for the unsupervised classification algorithm;
- **N**, indicating the maximum number of clusters to generate;
- **NB_PIX**, indicating the minimum number of pixels in a cluster.

The ISODATA clustering technique assumes that each LC class can be well-represented by a single mean vector.

Input and output data associated with this unsupervised classification process are described in Table 3-11.

Table 3-11: Input and output data of the spectral unsupervised (ISODATA) classification algorithm

Data	Description	(in, out, in-out)	Physical unit	Range
L3_StartDate_EndDate_SR_<n> n = 1, ...,15 (MERIS)	Global raster of surface reflectance in each band of the seasonal composites GeoTiff format	IN	None	[0 … 1]
LUT 2	Look-Up-Table describing the location of the 22 strata	IN	/	/
LUT 4	Look-Up-Table indicating which strata require the generation of multi-year seasonal composites	IN	/	/
LUT 7	Look-Up-Table describing the parameters of the unsupervised algorithm	IN	/	/
L4_<NB_ST>_Clusters_Spectral	Raster at the stratum level resulting from the unsupervised classification algorithm where each pixel is associated with a cluster ID (NB_Cluster)	OUT	None	[0 … 255]

Two parameters are associated with the ISODATA process, described in Table 3-12.

Table 3-12: Parameter of the ISODATA algorithm

Parameters	Description	(in, out, in-out)	Format	Range
NB_IT	Number of iterations to be performed	IN	Short	[0 … 255]
T	Percentage of pixels remaining unchanged between iterations	IN	Short	[0… 100]

3.2.1.3.2 Automatic reference-based labelling

The ISODATA algorithm has interpreted, for each stratum, L3 seasonal composites into a set of clusters being spectrally homogeneous but unidentified by a LC label. Transforming these clusters into LC classes (identified by a number NB_LAB and a name LABEL) is the objective of this labelling step. The labelling is achieved by comparing the cluster raster file (L4_<NB_ST>_CLUSTERS_Spectral) and the auxiliary LC reference database.

The reference LC layer (see section 2.2.3) (REF_LC) consists of global, regional and local reference LC maps selected as the most accurate ones available for a given region, with the highest spatial resolution and a legend compatible with CCI LC. It is associated with the LC legend, i.e. with the NB_LAB, LABEL and colour codes.

A superposition of the cluster raster (L4_<NB_ST>_CLUSTERS) with the reference layer (REF_LC) is carried out. For each cluster, a histogram of class frequency is computed (Figure 3-16). The most

represented classes inside the cluster are identified and ranked using (i) the number of pixels they cover (NB_Pix1, …, NB_Pixi) and (ii) their label (NB_LAB1, …, NB_LABi).

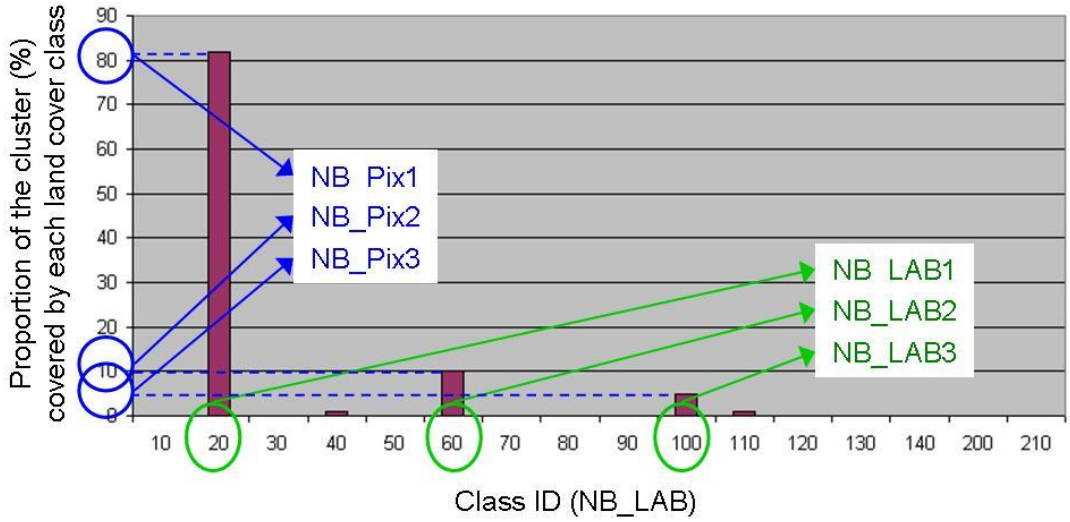

Figure 3-16: To assign a land cover label to each cluster, the histogram frequency of land cover labels derived from the reference data is interpreted. In this example, three class IDs representing three specific land cover labels (NB_LAB 20, 60, and 100) have been observed with high frequencies within the cluster. These frequencies are approximately 80%, 10%, and 5% respectively, based on the total number of pixels (NB_Pix) in the cluster. A set of predefined decision rules are used to assign a final land cover label to each cluster based on these frequencies.

The class frequency histogram is then interpreted according to a set of decision rules defined a priori. As a result, a unique LC class (i.e. a unique class number NB_LAB and name LABEL) is associated with each cluster (NB_Cluster), and an output raster file (CLASSIF_2) is generated. The histogram interpretation process also associates each cluster with an ambiguity code (CODE) that characterizes the ambiguity of the interpretation, and thus the reliability of the associated LC label. The decision rules are indicated in Table 3-14.

Input and output data associated with the automated labelling procedure are described in Table 3-13.

Table 3-13: Input and output data of step 2 of the classification chain, which is the spectral unsupervised classification process, for the automated labelling procedure

Data	Description	(in, out, in-out)	Physical unit	Range
L4_<NB_ST>_Clusters_Spectral	Raster at the stratum level resulting from the unsupervised classification algorithm where each pixel is associated with a cluster ID (NB_Cluster)	IN	None	[0 … 1024]
REF_LC	Reference LC layer where each pixel is associated with a LC class through an ID	IN	None	[0 … 255]
LUT 1	Look-Up-Table describing the CCI LCCS LC legend	IN	/	/
L4_<NB_ST>_Histo	Text file (one for each stratum) containing for each cluster (NB_Cluster) the LC class frequency (with indices NB_Pixi and NB_LABi)	INOUT	None	[0 … 100]

Data	Description	(in, out, in-out)	Physical unit	Range
CLASSIF 2	Land cover map resulting from the unsupervised classification approach (ISODATA algorithm + labelling process), where each pixel is associated with a LC class through an ID	OUT	None	[0 … 255]
Code_Classif2	Ambiguity code that characterizes the ambiguity of the classification process, thus reflecting, at the pixel level, the reliability of the CLASSIF_1B	OUT	None	0 … 10

No parameters are needed to process this aggregation. However, a critical input is the set of pre-defined labelling rules used to interpret the histograms. They are provided in Table 3-14.

Table 3-14: Decision rules defined to label clusters resulting from the ISODATA unsupervised algorithm in the spectral classification step

```
Decision rules are:
// for clusters with proportions of the first LC class (NB_Pix1_GL) above 60 %

• if (NB_Pix1_GL > 85%), do:
     CODE = 1
     * if (NB_Pix1 > 60%), do:
          NB_LAB(CLASSIF_1B) = NB_LAB1
     * else, do:
          NB_LAB(CLASSIF_1B) = NB_LAB1_GL
• if (70% < NB_Pix1_GL ≤ 80%) do:
     CODE = 2
     * if (NB_LAB1_GL = 210) and (159 < NB_LAB2_GL < 189), do:
          NB_LAB(CLASSIF_1B) = NB_LAB2_GL
     * if (NB_LAB1_GL = 200) and (NB_LAB2_GL = 190) and (NB_Pix1_GL > 10%), do:
          NB_LAB(CLASSIF_1B) = NB_LAB2_GL
     * if (NB_Pix1 > 60%), do:
          NB_LAB(CLASSIF_1B) = NB_LAB1
     * else, do:
          NB_LAB(CLASSIF_1B) = NB_LAB1_GL
• if (60% < NB_Pix1_GL ≤ 70%) do:
     CODE = 3
     * if (NB_LAB1_GL = 210), do:
          NB_LAB(CLASSIF_1B) = NB_LAB2_GL
     * if (NB_LAB1_GL = 200) and (NB_LAB2_GL = 190) and (NB_Pix1_GL > 10%), do:
          NB_LAB(CLASSIF_1B) = NB_LAB2_GL
     * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and [(NB_LAB2_GL = 130)
     or (99 < NB_LAB2_GL < 119)], do:
          NB_LAB(CLASSIF_1B) = 100
     * if (NB_LAB1_GL = 130) and (49 < NB_LAB2_GL < 129), do:
          NB_LAB(CLASSIF_1B) = 110
     * if (9 < NB_LAB1_GL < 29) and [(49 < NB_LAB2_GL < 99) or (119 < NB_LAB2_GL
     < 139)], do:
          NB_LAB(CLASSIF_1B) = 30
     * if [(49 < NB_LAB1_GL < 99) or (119 < NB_LAB1_GL < 139)] and (9 <
     NB_LAB2_GL < 29), do:
          NB_LAB(CLASSIF_1B) = 40
     * else, do:
          ▪ if (NB_Pix1 > 60%), do: NB_LAB(CLASSIF_1B) = NB_LAB1
          ▪ else, do: NB_LAB(CLASSIF_1B) = NB_LAB1_GL
```

```
// for clusters with proportions of the first LC class (NB_Pix1_GL) below 60 %,
consider also the second cluster (NB_Pix2_GL)

• if (40% < NB_Pix1_GL ≤ 60%) and (NB_Pix2_GL > 20%) do:
    CODE = 4
    * if (NB_LAB1_GL = 210), do:
          NB_LAB(CLASSIF_1B) = NB_LAB2_GL
    * if (NB_LAB1_GL = 200) and (NB_LAB2_GL = 190) and (NB_Pix1_GL > 10%), do:
          NB_LAB(CLASSIF_1B) = NB_LAB2_GL
    * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and [(NB_LAB2_GL = 130)
    or (99 < NB_LAB2_GL < 119)], do:
          NB_LAB(CLASSIF_1B) = 100
    * if (NB_LAB1_GL = 130) and (49 < NB_LAB2_GL < 129), do:
          NB_LAB(CLASSIF_1B) = 110
    * if (9 < NB_LAB1_GL < 29) and [(49 < NB_LAB2_GL < 99) or (119 < NB_LAB2_GL
    < 139)], do:
          NB_LAB(CLASSIF_1B) = 30
    * if [(49 < NB_LAB1_GL < 99) or (119 < NB_LAB1_GL < 139)] and (9 <
    NB_LAB2_GL < 29), do:
          NB_LAB(CLASSIF_1B) = 40
    * else, do:
          ▪ if (NB_Pix1 > 60%), do: NB_LAB(CLASSIF_1B) = NB_LAB1
          ▪ else, do: NB_LAB(CLASSIF_1B) = NB_LAB1_GL

• if (40% < NB_Pix1_GL ≤ 60%) and (NB_Pix2_GL ≤ 20%) and (NB_Pix1_GL+NB_Pix2_GL >
  50%), do:
    CODE = 5
    * if (NB_LAB1_GL = 210), do:
          NB_LAB(CLASSIF_1B) = NB_LAB2_GL
    * if (NB_LAB1_GL = 200) and (NB_LAB2_GL = 190) and (NB_Pix1_GL > 10%), do:
          NB_LAB(CLASSIF_1B) = NB_LAB2_GL
    * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and [(NB_LAB2_GL = 130)
    or (99 < NB_LAB2_GL < 119)], do:
          NB_LAB(CLASSIF_1B) = 100
    * if (NB_LAB1_GL = 130) and (49 < NB_LAB2_GL < 129), do:
          NB_LAB(CLASSIF_1B) = 110
    * if (9 < NB_LAB1_GL < 29) and [(49 < NB_LAB2_GL < 99) or (119 < NB_LAB2_GL
    < 139)], do:
          NB_LAB(CLASSIF_1B) = 30
    * if [(49 < NB_LAB1_GL < 99) or (119 < NB_LAB1_GL < 139)] and (9 <
    NB_LAB2_GL < 29), do:
          NB_LAB(CLASSIF_1B) = 40
    * else, do:
          ▪ if (NB_Pix1 > 60%), do: NB_LAB(CLASSIF_1B) = NB_LAB1
          ▪ else, do: NB_LAB(CLASSIF_1B) = NB_LAB1_GL

• if (40% < NB_Pix1_GL ≤ 60%) and (NB_Pix2_GL ≤ 20%) and (NB_Pix1_GL+NB_Pix2_GL ≤
  50%), do:
    CODE = 6
    * if (NB_LAB1_GL = 210), do:
          NB_LAB(CLASSIF_1B) = NB_LAB2_GL
    * if (NB_LAB1_GL = 200) and (NB_LAB2_GL = 190) and (NB_Pix1_GL > 10%), do:
          NB_LAB(CLASSIF_1B) = NB_LAB2_GL
    * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and (S_100110120 > 20%),
    do:
          ▪ if (NB_Pix1_GL > S_100110120), do: NB_LAB(CLASSIF_1B) = 100
          ▪ if (NB_Pix1_GL ≤ S_100110120), do: NB_LAB(CLASSIF_1B) = 110
    * if (NB_LAB1_GL = 130) and (S_50to120 > 20%), do:
          ▪ if (NB_Pix1_GL > S_50to120), do: NB_LAB(CLASSIF_1B) = 110
          ▪ if (NB_Pix1_GL ≤ S_50to120), do: NB_LAB(CLASSIF_1B) = 100
    * if (9 < NB_LAB1_GL < 29) and (S_50to90_120130 > 20%), do:
          ▪ if (NB_Pix1_GL > S_50to90_120130), do: NB_LAB(CLASSIF_1B) = 30
```

```
            ▪ if (NB_Pix1_GL ≤ S_50to90_120130), do: NB_LAB(CLASSIF_1B) = 40
      * if [(49 < NB_LAB1_GL < 99) or (119 < NB_LAB1_GL < 139)]  and (S_1020 >
      20%), do:
            ▪ if (NB_Pix1_GL > S_1020), do: NB_LAB(CLASSIF_1B) = 40
            ▪ if (NB_Pix1_GL ≤ S_1020), do: NB_LAB(CLASSIF_1B) = 30
      * if (9 < NB_LAB1_GL < 29) and (S_3040 > NB_Pix1_GL), do:
            NB_LAB(CLASSIF_1B) = 40
      * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and (S_100110 >
      NB_Pix1_GL), do:
            NB_LAB(CLASSIF_1B) = 100
      * if (NB_LAB1_GL = 130) and (S_100110 > NB_Pix1_GL), do:
            NB_LAB(CLASSIF_1B) = 110
      * else, do:
            ▪if (NB_Pix1 > 60%), do: NB_LAB(CLASSIF_1B) = NB_LAB1
            ▪else, do: NB_LAB(CLASSIF_1B) = NB_LAB1_GL

// for clusters with proportions of the first LC class (NB_Pix1_GL) below 40 %,
consider the second cluster (NB_Pix2_GL) and a combination of both

• if (NB_Pix1_GL ≤ 40) and (NB_Pix2_GL > 20%) and (NB_Pix1_GL+NB_Pix2_GL > 50%),
  do:
      CODE = 7
      * if (NB_LAB1_GL = 210), do:
            NB_LAB(CLASSIF_1B) = NB_LAB2_GL
      * if (NB_LAB1_GL = 200) and (NB_LAB2_GL = 190) and (NB_Pix1_GL > 10%), do:
            NB_LAB(CLASSIF_1B) = NB_LAB2_GL
      * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and [(NB_LAB2_GL = 130)
      or (99 < NB_LAB2_GL < 119)], do:
            NB_LAB(CLASSIF_1B) = 100
      * if (NB_LAB1_GL = 130) and (49 < NB_LAB2_GL < 129), do:
            NB_LAB(CLASSIF_1B) = 110
      * if (9 < NB_LAB1_GL < 29) and [(49 < NB_LAB2_GL < 99) or (119 < NB_LAB2_GL
      < 139)], do:
            NB_LAB(CLASSIF_1B) = 30
      * if [(49 < NB_LAB1_GL < 99) or (119 < NB_LAB1_GL < 139)] and (9 <
      NB_LAB2_GL < 29), do:
            NB_LAB(CLASSIF_1B) = 40
      * else, do:
            ▪if (NB_Pix1 > 60%), do: NB_LAB(CLASSIF_1B) = NB_LAB1
            ▪else, do: NB_LAB(CLASSIF_1B) = NB_LAB1_GL

• if (NB_Pix1_GL ≤ 40) and (NB_Pix2_GL > 20%) and (NB_Pix1_GL+NB_Pix2_GL ≤ 50%),
  do:
      CODE = 8
      * if (NB_LAB1_GL = 210), do:
            NB_LAB(CLASSIF_1B) = NB_LAB2_GL
      * if (NB_LAB1_GL = 200) and (NB_LAB2_GL = 190) and (NB_Pix1_GL > 10%), do:
            NB_LAB(CLASSIF_1B) = NB_LAB2_GL
      * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and (S_100110120 > 20%),
      do:
            ▪ if (NB_Pix1_GL > S_100110120), do: NB_LAB(CLASSIF_1B) = 100
            ▪ if (NB_Pix1_GL ≤ S_100110120), do: NB_LAB(CLASSIF_1B) = 110
      * if (NB_LAB1_GL = 130) and (S_50to120 > 20%), do:
            ▪ if (NB_Pix1_GL > S_50to120), do: NB_LAB(CLASSIF_1B) = 110
            ▪ if (NB_Pix1_GL ≤ S_50to120), do: NB_LAB(CLASSIF_1B) = 100
      * if (9 < NB_LAB1_GL < 29) and (S_50to90_120130 > 20%), do:
            ▪ if (NB_Pix1_GL > S_50to90_120130), do: NB_LAB(CLASSIF_1B) = 30
            ▪ if (NB_Pix1_GL ≤ S_50to90_120130), do: NB_LAB(CLASSIF_1B) = 40
      * if [(49 < NB_LAB1_GL < 99) or (119 < NB_LAB1_GL < 139)] and (S_1020 >
      20%), do:
            ▪ if (NB_Pix1_GL > S_1020), do: NB_LAB(CLASSIF_1B) = 40
            ▪ if (NB_Pix1_GL ≤ S_1020), do: NB_LAB(CLASSIF_1B) = 30
      * if (9 < NB_LAB1_GL < 29) and (S_3040 > NB_Pix1_GL), do:
```

```
            NB_LAB(CLASSIF_1B) = 40
        * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and (S_100110 >
        NB_Pix1_GL), do:
            NB_LAB(CLASSIF_1B) = 100
        * if (NB_LAB1_GL = 130) and (S_100110 > NB_Pix1_GL), do:
            NB_LAB(CLASSIF_1B) = 110
        * else, do:
            ▪ if (NB_Pix1 > 60%), do: NB_LAB(CLASSIF_1B) = NB_LAB1
            ▪ else, do: NB_LAB(CLASSIF_1B) = NB_LAB1_GL
• if (NB_Pix1_GL ≤ 40) and (NB_Pix2_GL ≤ 20%) and (NB_Pix1_GL+NB_Pix2_GL > 50%),
  do:
        CODE = 9
        * if (NB_LAB1_GL = 210), do:
            NB_LAB(CLASSIF_1B) = NB_LAB2_GL
        * if (NB_LAB1_GL = 200) and (NB_LAB2_GL = 190) and (NB_Pix1_GL > 10%), do:
            NB_LAB(CLASSIF_1B) = NB_LAB2_GL
        * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and [(NB_LAB2_GL = 130)
        or (99 < NB_LAB2_GL < 119)], do:
            NB_LAB(CLASSIF_1B) = 100
        * if (NB_LAB1_GL = 130) and (49 < NB_LAB2_GL < 129), do:
            NB_LAB(CLASSIF_1B) = 110
        * if (9 < NB_LAB1_GL < 29) and [(49 < NB_LAB2_GL < 99) or (119 < NB_LAB2_GL
        < 139)], do:
            NB_LAB(CLASSIF_1B) = 30
        * if [(49 < NB_LAB1_GL < 99) or (119 < NB_LAB1_GL < 139)] and (9 <
        NB_LAB2_GL < 29), do:
            NB_LAB(CLASSIF_1B) = 40
        * else, do:
            ▪ if (NB_Pix1 > 60%), do: NB_LAB(CLASSIF_1B) = NB_LAB1
            ▪ else, do: NB_LAB(CLASSIF_1B) = NB_LAB1_GL
• if (NB_Pix1_GL ≤ 40) and (NB_Pix2_GL ≤ 20%) and (NB_Pix1_GL+NB_Pix2_GL ≤ 50%),
  do:
        CODE = 10
        * if (NB_LAB1_GL = 210), do:
            NB_LAB(CLASSIF_1B) = NB_LAB2_GL
        * if (NB_LAB1_GL = 200) and (NB_LAB2_GL = 190) and (NB_Pix1_GL > 10%), do:
            NB_LAB(CLASSIF_1B) = NB_LAB2_GL
        * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and (S_100110120 > 20%),
        do:
            ▪ if (NB_Pix1_GL > S_100110120), do: NB_LAB(CLASSIF_1B) = 100
            ▪ if (NB_Pix1_GL ≤ S_100110120), do: NB_LAB(CLASSIF_1B) = 110
        * if (NB_LAB1_GL = 130) and (S_50to120 > 20%), do:
            ▪ if (NB_Pix1_GL > S_50to120), do: NB_LAB(CLASSIF_1B) = 110
            ▪ if (NB_Pix1_GL ≤ S_50to120), do: NB_LAB(CLASSIF_1B) = 100
        * if (9 < NB_LAB1_GL < 29) and (S_50to90_120130 > 20%), do:
            ▪ if (NB_Pix1_GL > S_50to90_120130), do: NB_LAB(CLASSIF_1B) = 30
            ▪ if (NB_Pix1_GL ≤ S_50to90_120130), do: NB_LAB(CLASSIF_1B) = 40
        * if [(49 < NB_LAB1_GL < 99) or (119 < NB_LAB1_GL < 139)] and (S_1020 >
        20%), do:
            ▪ if (NB_Pix1_GL > S_1020), do: NB_LAB(CLASSIF_1B) = 40
            ▪ if (NB_Pix1_GL ≤ S_1020), do: NB_LAB(CLASSIF_1B) = 30
        * if (9 < NB_LAB1_GL < 29) and (S_3040 > NB_Pix1_GL), do:
            NB_LAB(CLASSIF_1B) = 40
        * if [(49 < NB_LAB1_GL < 99) or (NB_LAB1_GL = 120)] and (S_100110 >
        NB_Pix1_GL), do:
            NB_LAB(CLASSIF_1B) = 100
        * if (NB_LAB1_GL = 130) and (S_100110 > NB_Pix1_GL), do:
            NB_LAB(CLASSIF_1B) = 110
        * else, do:
            ▪ if (NB_Pix1 > 60%), do: NB_LAB(CLASSIF_1B) = NB_LAB1
            ▪ else, do: NB_LAB(CLASSIF_1B) = NB_LAB1_GL
```

3.2.2 Stage 3: Baseline LC generation

At this stage of the classification chain, two global LC maps have been produced:

- the map called CLASSIF1_MY, resulting from the machine learning classification approach detailed in section 3.2.1.2;
- the map called CLASSIF2_MY, resulting from the unsupervised classification approach detailed in section 3.2.1.3.

They are merged to generate the baseline LC map.

Input and output data associated with this merging process are described in Table 3-15.

Table 3-15: Input and output data for stage 3 of the classification chain, i.e. the merging of LC maps obtained by the supervised and unsupervised classification approaches

Data	Description	(in, out, in-out)	Physical unit	Range
CLASSIF1_MY	Land cover map resulting from the supervised classification algorithm, where each pixel is associated with a LC class through an ID	IN	None	[0 … 255]
Code_CLASSIF1_MY	Classification probability associated with the label selected for each pixel	IN	None	[0 …1]
CLASSIF2_MY	Land cover map resulting from the unsupervised classification approach (ISODATA algorithm + labelling process), where each pixel is associated with a LC class through an ID)	IN	None	[0 … 255]
Code_CLASSIF2_MY	Ambiguity code that characterizes the ambiguity of the classification process, thus reflecting, at the pixel level, the reliability of the classification	IN	None	[0 … 10]
CLASSIF3	Land cover map resulting from the spectral classification approach (both supervised and unsupervised), where each pixel is associated with a LC class through an ID	OUT	None	[0 … 255]
Code_CLASSIF3	The quality flag that characterizes, at the pixel level, the reliability of the classification	OUT	None	[0 … 100]
Source_CLASSIF3	Flag that indicates, at the pixel level, if the LC label is derived from the supervised (CLASSIF1) or unsupervised (CLASSIF2) algorithm	OUT	None	[1, 2]

No parameters are required to merge the supervised and unsupervised maps. However, a critical input of this step is the set of decision rules that guides the merging. These rules are mainly intended to be applied to wetland and mosaic classes. They are provided in Table 3-16.

Table 3-16: Decision rules defined to merge the supervised and unsupervised LC maps

```
Decision rules are:
• if CLASSIF1_MY = (160, 170, 160 or 190), do:
      CLASSIF3 = CLASSIF1_MY
      Code_CLASSIF3 = Code_CLASSIF1_MY
      Source_CLASSIF3 = 1
• if CLASSIF2_MY = (30 or 40) and CLASSIF1_MY = (10 or 20), do:
      CLASSIF3 = CLASSIF1_MY
      Code_CLASSIF3 = Code_CLASSIF1_MY
      Source_CLASSIF3 = 1
• if CLASSIF2_MY = (100 or 110) and CLASSIF1_MY = (50, 60, 70, 80 or 90), do:
      CLASSIF3 = CLASSIF1_MY
      Code_CLASSIF3 = Code_CLASSIF1_MY
      Source_CLASSIF3 = 1
```

3.2.3 Stage 4: Generation of the global annual land cover maps

The machine learning and unsupervised LC classification map series serve as input for the LCC module. The global annual LC map for a given year results from the update of the CCI baseline LC map with the change information, as illustrated in Figure 3-17. The process is organized into three significant steps: (i) the annual change detection at 1 km, (ii) the change delineation at 300 m and (iii) the baseline update. These processes are detailed in the following sections.

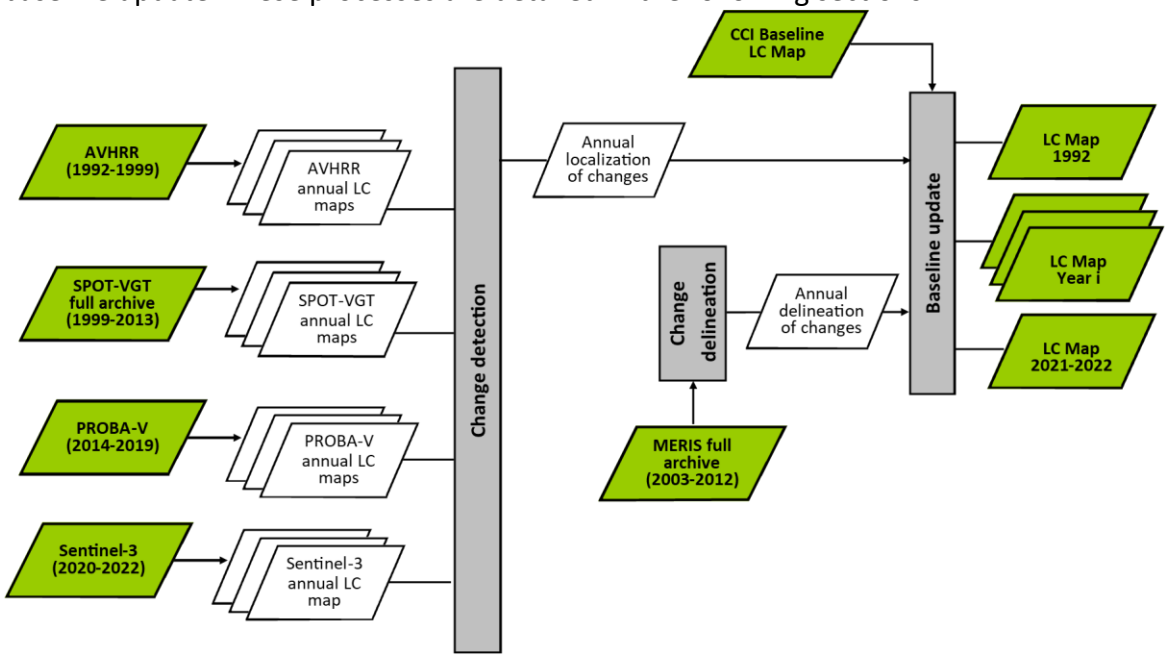

Figure 3-17: Schematic representation of the methodology developed to derive global annual LC maps from the baseline global LC map and the LCC, detected and delineated at 1 km and 300 m, respectively.

3.2.3.1 Annual change detection at 1 km

The first step of the change module consists of mapping the dynamics of the land surface by analysing, on a per-pixel basis, annual time series of 1-km global unsupervised classifications (see section 3.2.1.3) derived from AVHRR (1992–1999), SPOT-VGT (1999–2013), PROBA-V (2014 – 2019), 300-m SENTINEL-3 OLCI (2020) and 300-m SENTINEL-3 OLCI & SLSTR (2021-2022) averaged at 1 km.

Analysing the sequence of global LC classifications over time allows the capturing of the dominant LC transitions. Nevertheless, to avoid false change detections due to the inter-annual variability in classifications, each change has to be confirmed in at least two successive years in the classification time series, except for deforestation/reforestation transitions. Thus, the algorithm is applied on a time window of approximately 14 years (W in Table 3-20) to allow a robust yet flexible change detection when the LC map time series is extended. As each LC map of a given year is released nine months after the last acquisition of the year, six months extra from the following year are being used for LCC confirmation. For example, the LC map $Year_i$ is being produced considering a $Year_{i-13}$ – $Year_{i+0.5}$ window, where "i" represents a given year.

More than one LCC can be detected in the world's most dynamic regions during the period. Most pixels are associated with 0, 1, 2 or 3 LCCs, knowing that most LCCs need to last at least two years to be detected. Water loss, forest gain and loss are recorded without using the confirmation procedure presented above.

The change detection method allows (i) the year of change and (ii) the type of change observed to be determined. These are both inputs to the change delineation step described in 3.2.3.2.
The change is detected on a pixel basis between groups of classes that are different from the LC legend. The classes considered here are cropland, forest, grassland, wetland, settlement and other (shrubland, sparse vegetation, bare area, water). This grouping corresponds to the main Intergovernmental Panel on Climate Change (IPCC) land categories [IPCC 2006], which was a requirement expressed by the climate users [ESA_CCI_LC_URD, 2015 and ESA_CCI_LC_PSD, 2015]. The correspondence between these groups of classes is defined in Table 3-17. It is used to generate the input "IPCC_Grouping".

Table 3-17: Correspondence between the IPCC classes and the LCCS classes of the LC maps legend.

Classes considered for the change detection (IPCC)	**LCCS LC classes (class ID | text description)**	
Cropland	10, 11, 12	Rainfed cropland
	20	Irrigated cropland
	30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)
	40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (< 50%)
Forest	50	Tree cover, broadleaved, evergreen, closed to open (>15%)
	60, 61, 62	Tree cover, broadleaved, deciduous, closed to open (> 15%)
	70, 71, 72	Tree cover, needle-leaved, evergreen, closed to open (> 15%)
	80, 81, 82	Tree cover, needle-leaved, deciduous, closed to open (> 15%)
	90	Tree cover, mixed leaf type (broadleaved and needle-leaved)
	100	Mosaic tree and shrub (>50%) / herbaceous cover (< 50%)
	160	Tree cover, flooded, fresh or brackish water
	170	Tree cover, flooded, saline water
Grassland	110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)
	130	Grassland
Wetland	180	Shrub or herbaceous cover, flooded, fresh-saline or brackish water
Settlement	190	Urban areas

Classes considered for the change detection (IPCC)		LCCS LC classes (class ID | text description)	
Other	Shrubland	120, 121, 122	Shrubland
	Sparse vegetation	140	Lichens and mosses
		150, 152, 153	Sparse vegetation (tree, shrub, herbaceous cover)
	Bare area	200, 201, 202	Bare areas
	Water	210	Water

The change detection method is applied over twelve types of changes (Table 3-18). Note that changes in urban areas are derived in a separate module (see section 3.2.3.3).

Table 3-18: Descriptions for each type of change, ID and band number associated with the output of the change detection at 1km

Change Type	ID	Band number	Description
Forest Loss	101	1	Changes from forest classes to crops, shrub, grasses, sparse vegetation, wetlands and bare areas
Forest Gain	102	2	Changes from crops, shrub, grasses, sparse vegetation, wetlands and bare areas to forest classes
Cropland Loss	103	3	Changes from crops classes to grasses, sparse vegetation and bare areas
Cropland Gain	104	4	Changes from grasses, sparse vegetation and bare areas to crop classes
Grassland Loss	105	5	Changes from grasses to sparse vegetation and bare areas, and from sparse vegetation to bare areas
Grassland Gain	106	6	Changes from sparse vegetation and bare areas to grasses, and from bare areas to sparse vegetation
Shrubland Loss	107	7	Changes from shrub to crops, grasses, sparse vegetation and bare areas
Shrubland Gain	108	8	Changes from crops, grasses, sparse vegetation and bare areas to shrub
Wetland Loss	109	9	Changes from wetland to crops, grasses, shrub, sparse vegetation and bare areas
Wetland Gain	110	10	Changes from crops, grasses, shrub, sparse vegetation and bare areas to wetland
Water Loss	111	11	Changes from water to crops, forests, grasses, shrub, sparse vegetation, wetland and bare areas
Water Gain	112	12	Changes from crops, forests, grasses, shrub, sparse vegetation, wetland and bare areas to water

The change detection method is guided by decision rules that analyse the suite of annual classifications and determine if there has been an LCC during this period. These decision rules follow a general structure but are also fine-tuned per stratum and type of change (Table 3-19). Those specific decision rules concern, for example, the number of years for which change should be confirmed over the period and the IPCC classes observed as a result of the change (see Table 3-17).

Table 3-19: Decision rules applied in the change detection algorithm at 1km for each stratum by type of change

Change Type	Specific and generic decision rules per stratum
Forest Loss	Specific decision rules for strata 1 – 6, 14, 16, 20
	Specific decision rules for strata 7 – 13, 15, 17 – 19, 22
	Specific decision rules for stratum 21
Forest Gain	Specific decision rules for strata 2 – 12, 14 – 20, 21
	Specific decision rules for stratum 13
	Specific decision rules for stratum 1
	Specific decision rules for stratum 21
Cropland Loss	Generic for all strata
Cropland Gain	Specific decision rules for strata 1 – 11, 14, 16, 17, 19 – 21
	Specific decision rules for stratum 18
	Specific decision rules for strata 12, 13, 15, 22
Grassland Loss	Generic for all strata
Grassland Gain	Generic for all strata
Shrubland Loss	Specific decision rules from strata 1 – 20 and 22
	Specific decision rules for stratum 21
Shrubland Gain	Generic for all strata
Wetland Loss	Generic for all strata
Wetland Gain	Generic for all strata
Water Loss	Generic for all strata
Water Gain	Generic for all strata

The generic decision rules for the change detection at 1 km over the 14-year window follow a series of three conditional statements for each type of change which aim at:

- Identifying if a potential change exists between two different groups of IPCC classes by sequentially scanning the series of annual LC classifications;
- Identifying if, among the potential changes, some are confirmed over time by observing the same over several years;
- Identifying the year of change as the year of the first LC class diverging from the original group of IPCC classes.

The standard pre-processing chain applied over the various sensors provides stable surface reflectance from one sensor to another. The annual LC classifications reached a high level of accuracy, allowing a comparison of LC labels over a time series

Input and output data associated with the change detection process are described in Table 3-20.

Table 3-20: Input and output data for the change detection at the stratum level

Data	Description	(in, out, in-out)	Physical unit	Range
IPCC_Grouping_1_14_<NB_ST>_<W>	Matrix for each stratum (NB_ST) converting the LCCS legend to the IPCC legend based on Table 3-17 on each time window (W)	IN	None	[1…10]
M_1_14_<NB_ST>_<W>(i)	Matrix at the stratum level comprising the annual SPOT-VGT and PROBA-V classifications reshaped in vectors for each stratum (NB_ST) and each time window (W). i = pixel	IN	None	[0…220]
ChangeYear_<NB_ST>_<W> (band)	Multi-band raster at the stratum level of the year of change by type of change (1 band by type of change) for each time window (W).	IN-OUT	None	[2004… 2022]
ChangeType_<NB_ST>_<W> (band)	Multi-band raster at the stratum level of the type of change for each time window (W). Each code corresponds to the change type ID in Table 3-18.	OUT	None	[101… 114]

3.2.3.2 Change delineation at 300 m

The change information extracted from the 1 km time series was enhanced thanks to the higher spatial resolution of MERIS, PROBA-V and S3-OLCI annually classified between 2004 and 2022. Before that, LCC is described at 1 km only. The supervised machine learning algorithm, described in section 3.2.1.2, was used for the LCC delineation.

3.2.3.2.1 Multi-annual change-based composite

The rationale of the multi-annual change-based composite is the same as the rationale behind the compositing process: increasing the number of observations for a given seasonal stage will reduce the signal's noise. However, multi-annual compositing assumes that no LCC occurred inside the pixel during the compositing period, which could make changed pixels undetectable. The multi-annual change-based compositing mitigates this issue by splitting the compositing period at the date of change in areas where the change is likely to occur.

There are two assumptions for this algorithm. First, it assumes that the LC is stable between two dates of change. Second, it relies on stable radiometric corrections from year to year. The main limitation of the algorithm is the potential absence of data during the compositing period.

The methodology to derive multi-annual change-based composites are inherited from the CCI LC project. In addition to the 1 km change date as presented in section 3.2.3.1, a buffer layer is also created as described in section 3.2.3.4.

3.2.3.3 Change detection for the urban class

The methodology described in section 3.2.3.1 did not show satisfactory results to classify the urban class both spatially and temporarily. Therefore, a new algorithm was dedicated to urban change detection for the 300 m era (i.e. from 2003 onwards). Before 2003, the 1 km spatial resolution brought by AVHRR and SPOT-VGT did not allow for robust urban change detections. Between 1992 and 1999 and between 2000 and 2003, the urban footprint corresponds to the Global Human Settlement Layer (GHSL) 1990 and GHSL 2000. The urban footprint is stable inside and expands between each of these two periods.

From 2003 onwards, the urban change detection relied on a series of annual machine learning spectral classifications of MERIS 300 m from 2004 and 2011 and PROBA-V 300 m from 2012 and 2016. A filling procedure aims to overcome yearly data gaps. Two constraints ensured spatio-temporal consistency in the change detection: (i) the urban footprint is only allowed to expand over time; (ii) the minimum and maximum urban footprints are constrained by the Landsat-based Joint Research Centre Global Human Settlement Layer (GHSL) [Pesaresi et al., 2016] resampled to 300 m for both 2000 and 2014. In addition, the Global Urban Footprint (GUF) [Esch et al., 2017] was fused to the GHSL 2014 to compensate for urban area omissions.

With no auxiliary urban product like the GHSL and GUF available for 2016 onwards at the time of generating the LC maps 2016–2022, a strict trimming procedure and maximum likelihood classification were used to detect urban areas on the seasonal composites at 300 m and the urban class of the previous LC map as the training data set.

Assumptions are the following:

- urban footprints expand over time, with no destruction;
- the GHSL data sets for years 2000 and 2014 represent the urban state correctly;
- the error implied by using the GHSL of 2014 in 2015 is minimal at the global scale.

Limitations concern:

- the lack of change detections between 1992 and 1999 and between 2000 and 2003;
- a temporal shift in the urban change detection could be present in highly cloud-covered areas resulting from the gap-filling procedure.

Input and output data associated with the change detection for the urban class are described in

Table 3-21.

Table 3-21: Input and output data for the change delineation at 300 m.

Data	Description	(in, out, in-out)	Physical unit	Range
SUP_CLASSIF_300m_<sensor>_<year> where: <sensor> is “MERIS” or “PROBA-V” <year> is in range [2003, 2015]	Land cover map resulting from the supervised classification algorithm, where each pixel is associated with a LC class through an ID	IN	None	[0 … 255]
GHSL_<year> where <year> is 2000 OR 2014	Landsat-based binary layer with urban presence/absence for year <year> at ~38 m spatial resolution	IN	None	[0 ... 190]
GUF_2014	The TerraSAR-X/TanDEM-X binary layer of built-up areas using images from 2011 and 2012, complemented by images from 2013 and 2014 to fill gaps (~12 m, near the equator)	IN	None	[0 ... 190]
GHSL_300m_<year> where <year> is 2000 OR 2014	Binary layer with urban presence/absence for year <year> at 300 m spatial resolution	IN-OUT	None	[0 ... 190]
Urban_Mask_<Year> where <year> is range [2003, 2015]	binary layers with urban area presence/absence for year <year>	OUT	None	[0 ... 190]

3.2.3.4 Baseline update

From 2004 onwards, the 1-km LCC was further delineated at 300 m with observations from MERIS, PROBA-V and S3-OLCI. In the 1992– 2003 period, LCC is detected at 1 km only. LCC information is now used to derive annual LC maps from the baseline LC map.
Under the 1-km change detection, annual classifications are compared to the baseline. If the classification indicates a change during a specific year, the baseline label can be modified from this specific year if it respects certain rules.

The detected changes from step 3.2.3.1 are cleaned with the combination of a connection algorithm, followed by an erosion algorithm. It removes isolated pixels more susceptible to be false change detections. The remaining changes are extended with a buffer (to avoid artefacts coming from the 1km to 300m change of resolution). Inside this buffer, the rules are more restrictive.

The following sections describe the baseline update generically for both changes at 1-km and 300 m spatial resolution.

Input and output data associated with the change detection process are described in Table 3-22.

Table 3-22: Input and output data for the change delineation and baseline update

Data	Description	(in, out, in-out)	Physical unit	Range
CCI_Baseline_LC_Map	Land cover map resulting from the editing of the step 3 output (CLASSIF 3) through a posteriori addition of external data set, where each pixel is associated with a LC class through an ID	IN	None	[0 … 255]
Buffer_ChangeYr_1km_ <NB_ST>	Multi-band raster of the year of LCC (value 0 indicating no change), available at the stratum level. Several bands are used if several changes occur inside a given pixel. The MERIS & PROBA-V changes are merged with the SPOT-VGT changes. A majority buffer of 5x5 km is applied around the pixels of step 3.2.3.1 after applying a connection and an erosion algorithm. The buffer is tagged to discern originals from added. GeoTiff format.	IN	None	[1992 … 2015]
Buffer_ChangeTy_1km_ <NB_ST>	Multi-band raster of the thirteen types of change (value 0 indicating no change) available at the stratum level. The comments are the same as for ChangeYr GeoTiff format.	IN	None	[0 … 210]
Classif_<YEAR>_AVHRR	Raster files where each pixel is associated with a LC class, identical as in step 3.2.3.1	IN	None	[0 … 255]
Classif_<YEAR>_SPOT-VGT	Raster files where each pixel is associated with a LC class, identical as in step 3.2.3.1.	IN	None	[0 … 255]
Classif_<YEAR>_MERIS	Raster file where each pixel is associated with a LC class is generated in step raster files where each pixel is associated with a LC class, identical to those in step 3.2.3.2.	IN	None	[0 … 255]
Classif_<YEAR>_PROBA-V	Raster files where each pixel is associated with a LC class	IN	None	[0 … 255]
Urban_Mask_<YEAR>	Raster files where a value of 0 or 190 (urban class) is set annually	IN	None	0 or 190
LC_Map_<year> where <year> can vary from 1992 to 2015	Land cover map representative of year <year>, derived from the Baseline map with the outputs of the change detection, where each pixel is associated with a LC class through an ID	OUT	None	[0 … 210]

The algorithm is driven by a set of decision rules used to modify the baseline map and derive the global annual LC maps. Those rules rely on labels before the change (BEF_CLASS), change type (CHGT_TYPE) and the label of the LCC fate (FATE_CLASS). The links between those labels are given in Table 3-23.

Table 3-23: Decision rules to convert the LC baseline to annual LC maps using the change layers

chgt_type	BEF_class	FATE _class
Forest Loss	50 to 100, 160 or 170	10 to 40, 110, 120, 130, 150, 180 or 200

chgt_type	BEF_class	FATE _class
Forest gain	10 to 40, 110, 120, 130, 150, 180 or 200	50 to 100, 160 or 170
Cropland loss	10 to 40	110, 130, 150 or 200
Cropland gain	110, 130, 150 or 200	10 to 40
Grassland loss	110, 130 150	150, 200 200
Grassland gain	150 200	110, 130 150
Shrubland loss	120	10 to 40, 110, 130, 150, 200
Shrubland gain	10 to 40, 110, 130, 150, 200	120
Wetland loss	180	10 to 40, 110, 120, 130, 150, 200
Wetland gain	10 to 40, 110, 120, 130, 150, 200	180
Water loss	210	10 to 40, 50 to 100, 160 or 170, 110, 120, 130, 150, 180, 200
Water gain	10 to 40, 50 to 100, 160 or 170, 110, 120, 130, 150, 180, 200	210
Urban gain	10 to 40, 50 to 100, 160 or 170, 110, 120, 130, 150, 180, 200	190

Table 3-24: Parameters relating to change detection

Parameters	Description	(in, out, in-out)	Format	Range
Change_threshold	Percentage of minimum valid classes to validate a period between changes or first/last periods	IN-OUT	float	[0..1]
Change_threshold_buffer	Identical to change_threshold, applied only under the buffer	IN-OUT	float	[0..1]

4 Output data

The following section presents the output products of the processing chain.

4.1 Seasonal composites of 2021 & 2022

As described in sec. 3.1.7.3, for each LC Map, a total of 16 different seasonal composites[19] have been processed. The processed seasons with the corresponding start and end date for the year 2021 and 2022 are presented in Table 4-1 & Table 4-2 and the content in Table 4-3.

Table 4-1: Processed seasons for the year 2021

No.	SEASON	START	END
1	2020-12-03-P13W	03.12.2020	04.03.2021
2	2020-12-03-P17W	03.12.2020	01.04.2021
3	2021-01-01-P22W	01.01.2021	03.06.2021
4	2021-01-01-P52W	01.01.2021	31.12.2021
5	2021-04-02-P13W	02.04.2021	01.07.2021
6	2021-04-16-P7W	16.04.2021	03.06.2021
7	2021-04-16-P13W	16.04.2021	15.07.2021
8	2021-04-16-P17W	16.04.2021	12.08.2021
9	2021-06-04-P13W	04.06.2021	02.09.2021
10	2021-06-18-P15W	18.06.2021	30.09.2021
11	2021-07-16-P7W	16.07.2021	02.09.2021
12	2021-07-30-P13W	30.07.2021	28.10.2021
13	2021-08-13-P13W	13.08.2021	11.11.2021
14	2021-09-03-P13W	03.09.2021	02.12.2021
15	2021-10-01-P13W	01.10.2021	31.12.2021
16	2021-12-03-P13W	03.12.2021	04.03.2021

Table 4-2: Processed seasons for the year 2022

No.	SEASON	START	END
1	2021-12-03-P13W	03.12.2021	04.03.2022
2	2021-12-03-P17W	03.12.2021	01.04.2022
3	2022-01-01-P22W	01.01.2022	03.06.2022
4	2022-01-01-P52W	01.01.2022	31.12.2022
5	2022-04-02-P13W	02.04.2022	01.07.2022
6	2022-04-16-P7W	16.04.2022	03.06.2022
7	2022-04-16-P13W	16.04.2022	15.07.2022
8	2022-04-16-P17W	16.04.2022	12.08.2022
9	2022-06-04-P13W	04.06.2022	02.09.2022
10	2022-06-18-P15W	18.06.2022	30.09.2022
11	2022-07-16-P7W	16.07.2022	02.09.2022

[19] *Note: These are intermediate products and will be not available for users and are not described in the PUGS*

No.	SEASON	START	END
12	2022-07-30-P13W	30.07.2022	28.10.2022
13	2022-08-13-P13W	13.08.2022	11.11.2022
14	2022-09-03-P13W	03.09.2022	02.12.2022
15	2022-10-01-P13W	01.10.2022	31.12.2022
16	2022-12-03-P13W	03.12.2022	04.03.2022

Table 4-3: Bands in seasonal composites of the merged SENTINEL-3 OLCI/SLSTR data

Bands	Description
SR of band x_{sensor}; $x_{OLCI/SLSTR}$ = 1 … 12, 16 … 18 (OLCI) and 1 … 6 (SLSTR)	surface reflectance of band x_{sensor}; $x_{OLCI/SLSTR}$ = 1 … 12, 16 … 18 (OLCI) and 1 … 6 (SLSTR)
ε_{SR} of band x_{sensor}; $x_{OLCI/SLSTR}$ = 1 … 12, 16 … 18 (OLCI) and 1 … 6 (SLSTR)	uncertainties of the surface reflectance of band x_{sensor}; $x_{OLCI/SLSTR}$ = 1 … 12, 16 … 18 (OLCI) and 1 … 6 (SLSTR)
vegetation_index	Normalised Difference Vegetation Index $NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$
status	current pixel status
obs_count	number of valid observations over pixel
status_counts of status y; y=land, snow/ice, water, cloud, etc.	number of observations over pixel covered by • invalid • clear land • clear snow/ice • clear water • cloud • cloud shadow

4.2 LC map

The C3S LC project delivers global LC maps at 300 m spatial resolution from 2016 onwards, consistent with the CCI annual global LC maps from 1992 – 2015. The Coordinate Reference System used for the global LC database is a geographic coordinate system (GCS) based on the World Geodetic System 84 (WGS84) reference ellipsoid. The validation and intercomparison plan is provided in the PQAD [RD-1 & RD-3].

Figure 4-1 presents the LC map from the year 2022 at the global scale.

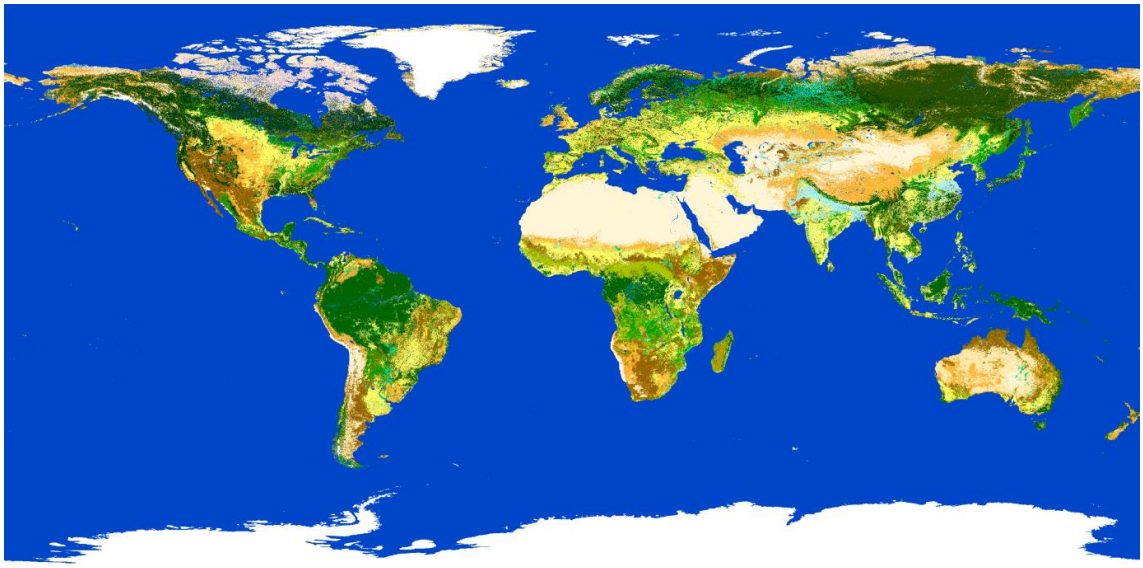

Figure 4-1: The most recent map from the LC map series from the year 2022, at 300 m spatial resolution. The associated legend is shown in table 4-4.

The methods and characteristics of the 2016 LC map are consistent with the process used to create the CCI-LC maps. The following sections will describe the legend of the LC maps, the processing chain including the classification and change detection modules, and finally, the format of the maps and their four quality flags, valid for the entire time series.

4.3 Legend

The typology was defined using the Land Cover Classification System (LCCS) developed by the United Nations (UN) Food and Agriculture Organization (FAO), with the view to be as much as possible compatible with the Global Land Cover 2000 Project (GLC2000), GlobCover 2005 and 2009 products [Defourny et al., 2009]. In addition, the UN-LCCS was found to be quite compatible with the Plant Functional Types (PFTs) used in climate models [ESA_CCI_LC_URD, 2011].

The UN-LCCS defines LC classes using a set of classifiers. The system was designed as a hierarchical classification, which allows adjustment of the thematic detail of the legend to the amount of information available to describe each LC class, whilst following a standardized classification approach.

As the LC maps are designed to be globally consistent, their legend is determined by the level of available information, which makes sense at the scale of the entire world. The “level 1” legend – also called “global” legend – presented in Table 4-4 meets this requirement. This legend has 22 classes, and each class is associated with a unique code (i.e. class codes of 10, 20, 30).

The LC maps are also described by a more detailed legend, called “level 2” or “regional”. This level 2 legend uses more accurate and regional information – where available – to define more LCCS classifiers and reach a higher level of detail in the legend. Therefore, this regional legend has more classes, which can be found in the Product User Guide and Specification (PUGS) [RD-2 & RD-4]. The

regional classes are associated with a set of discrete values (i.e. class codes such as 11, 12). They are not present all over the world since they were not adequately discriminated at a global scale.

Table 4-4: Level 1 (or global) legend of the LC maps, based on the UN-LCCS.

Value	Label	Color	RGB
0	No Data		0, 0, 0
10	Cropland, rainfed		255, 255, 100
20	Cropland, irrigated or post-flooding		170, 240, 240
30	Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%)		220, 240, 100
40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%)		200, 200, 100
50	Tree cover, broadleaved, evergreen, closed to open (>15%)		0, 100, 0
60	Tree cover, broadleaved, deciduous, closed to open (>15%)		0, 160, 0
70	Tree cover, needle-leaved, evergreen, closed to open (>15%)		0, 60, 0
80	Tree cover, needle-leaved, deciduous, closed to open (>15%)		40, 80, 0
90	Tree cover, mixed leaf type (broadleaved and needle-leaved)		120, 130, 0
100	Mosaic tree and shrub (>50%) / herbaceous cover (<50%)		140, 160, 0
110	Mosaic herbaceous cover (>50%) / tree and shrub (<50%)		190, 150, 0
120	Shrubland		150, 100, 0
130	Grassland		255, 180, 50
140	Lichens and mosses		255, 220, 210
150	Sparse vegetation (tree, shrub, herbaceous cover) (<15%)		255, 235, 175
160	Tree cover, flooded, fresh or brackish water		0, 120, 90
170	Tree cover, flooded, saline water		0, 150, 120
180	Shrub or herbaceous cover, flooded, fresh/saline/brackish water		0, 220, 130
190	Urban areas		195, 20, 0
200	Bare areas		255, 245, 215
210	Water bodies		0, 70, 200
220	Permanent snow and ice		255, 255, 255

Among these LC classes, four were primarily identified thanks to external data sets: the “tree cover, flooded, saline water” (class value 170) class which is based on the global mangrove atlas [UNEP-WCMC, 2002], the “urban areas” (class value 190) which relies both on the Global Human Settlement Layer [Pesaresi et al., 2016] and on the Global Urban Footprint [Esch et al., 2017], the “water bodies” (class value 210) which have been inherited from the CCI global map of open water bodies [Lamarche et al., 2017] and the “permanent snow and ice” (class value 220) which comes from the Randolph Glaciers Inventory [Pfeffer et al., 2014] (to which the CCI-Glaciers project is one of the main contributors).

The format, including naming convention and quality flags, are described in detail in the product user guide (PUGS) [RD-2 & RD-4].

References

Arino et al., 2008: GlobCover the most detailed portrait of Earth, ESA Bulletin, 136, pp 25-31, November 2008.,available online at: http://www.esa.int/esapub/bulletin/bulletin136/bul136d_arino.pdf, accessed on 19/02/2024

Compiègne et al., 2013: The phase matrix truncation impact on polarized radiance, AIP Conference Proceedings; May2013, Vol. 1531 Issue 1, p95

Dee et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137(656), pp. 553–597. doi: 10.1002/qj.828.

Defourny et al., 2009: The first 300 m global land cover map for 2005 using Envisat MERIS time series: a product of the GlobCover system. In: "Proceedings of the 33rd International Symposium on Remote Sensing of Environment", Stresa, Italy, 4–8 May 2009, TS-5-1 (Ref 791)

Defourny, P., Vancutsem, C., Bicheron, P., Brockmann, C., Nino, F., Schouten, L., & Leroy, M. (2006, May). GLOBCOVER: a 300 m global land cover product for 2005 using Envisat MERIS time series. In Proceedings of ISPRS Commission VII Mid-Term Symposium: Remote Sensing: from Pixels to Processes, Enschede (NL) (pp. 8-11).

di Gregorio and Jansen, 2005: Land Cover Classification System. Classification concepts and user manual., FAO. Edited by R. Food & Agriculture Organization of the United Nations. Italy. doi: 10.1017/CBO9781107415324.004.

Dubuisson et al. 2016: ARTDECO, an Atmospheric Radiative Transfer Database for Earth and Climate Observation, International Radiation Symposium, 17–22 April 2016, Auckland (https://www.icare.univ-lille.fr/artdeco/), accessed on 19/02/2024

ECMWF, 2017: C3S Copernicus Procurement: Invitation to Tender Copernicus project: Proposal for Lot 5: Surface albedo, LAI, fAPAR, Land Cover and Fire, 15.06.2017 ECMWF – not public available

ESA_CCI_LC_ATBD, 2017: ESA LAND COVER CCI – ALGORITHM THEORETICAL BASIS DOCUMENT – VERSION 1.2 – CCI-LC-ATBDv2– D2.2 – V1.2 – https://www.esa-landcover-cci.org/webfm_send/139 >ESACCI-LC-Ph2-ATBDv2_PartIII-LCClassification_1.2.pdf. Accessed on 19/02/2024.

ESA_CCI_LC_ATBD, 2017a: Land Cover CCI: Algorithm Theoretical Basis Document: Pre-Processing – Year 2 – v1.0 – 02-05.2015. online]. Available at: https://www.esa-landcover-cci.org/?q=webfm_send/136, accessed on 19/02/2024.

ESA_CCI_LC_DPM, 2013: ESA LAND COVER CCI DPM Phase I. Land Cover Climate Change Initiative Detailed Processing Model, available online at: https://www.esa-landcover-cci.org/?q=webfm_send/77, accessed on 19/02/2024.

ESA_CCI_LC_PSD, 2015: Land Cover CCI: Product Specification Document – Year 1 – v1.2 – 13-04-2015. Available online at: https://www.esa-landcover-cci.org/?q=webfm_send/97, accessed on 19/02/2024.

ESA_CCI_LC_PVASRv2.0 (2012) LC-CCI PVASR Phase I. Land Cover Climate Change Initiative Product Validation Algorithm Selection Report – only available on request

ESA_CCI_LC_URD, 2011: LAND COVER CCI – USER REQUIREMENTS DOCUMENT – VERSION 2.2 – LAND_COVER_CCI_URD_2.2 – V2.2 – 22-02-2011", available online at https://www.esa-landcover-cci.org/?q=webfm_send/46, accessed on 19/02/2024.

ESA_CCI_LC_URD, 2015: ESA, 2015, "Land Cover CCI: User Requirements Document –Year 2 – v1.2 – 16-12-2015". online] available at: https://www.esa-landcover-cci.org/?q=webfm_send/128 , accessed on 19/02/2024.

ESA_S3_SYN_ATBD, 2010: ESA_S3_SYN_ATBD – North, P. and Heckel, A. (2010), "SENTINEL-3 OPTICAL PRODUCTS AND ALGORITHM DEFINITION - SYN Algorithm Theoretical Basis Document", S3-L2-SD-03-S02-ATBD v2.3, 30.06.2010,

https://sentinels.copernicus.eu/documents/247904/349589/SYN_L2-3_ATBD.pdf, accessed on 19/02/2024.

Esch et al., 2017: Breaking New Ground in Mapping Human Settlements from Space – The Global Urban Footprint. ISPRS Journal of Photogrammetry and Remote Sensing 134: 30–42. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2017.10.012 .

ESRI 2016: How hillshade works. https://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/how-hillshade-works.htm

Fell and Fischer, 2001: Numerical simulation of the light field in the atmosphere-ocean system using the matrix-operator method, Journal of Quantitative Spectroscopy & Radiative Transfer, 69, 351-388.

Fischer and Grassl, 1984: Radiative transfer in an atmosphere-ocean system: an azimuthally dependent matrix-operator approach, Applied Optics, 23, 1032-1039.

GlobCover 2005: GlobCover 2005 Project - Validation Report V2.1 (Report editors: Bicheron P. and Leroy, M, End-User Meeting 3 – November 2008

GlobAlbedo_ATBDv4.12 (2013): GlobAlbedo Algorithm Theoretical Basis Document Version 4.12 , v4.12, Available at: http://www.globalbedo.org/docs/GlobAlbedo_Albedo_ATBD_V4.12.pdf, accessed on 19/02/2024.

GlobAlbedo_ATBD/BBDRv1.0 (2010): GlobAlbedo_BBDR_ATBD version 1.1 – only available on request.

GlobCover_DJFv3.8 (2008): GlobCover 2005 project Design Justification File, I3.8 – only available on request.

Hersbach, H. et al., 2020: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146(730), pp. 1999–2049. doi: 10.1002/qj.3803.

Hu et al., 1999: The Interrelationship of Atmospheric Correction of Reflectances and Surface BRDF Retrieval: A Sensitivity Study, IEEE Trans. Geosci. Remote Sensing 37 (2): 724-738

IPCC 2006: IPCC Guidelines for National Greenhouse Gas Inventories, Prepared by the National Greenhouse Gas Inventories Programme, Eggleston H.S., Buendia L., Miwa K., Ngara T. and Tanabe K. (eds). Published: IGES, Japan.

Ishida and Nakajima,2009: Development of an unbiased cloud detection algorithm for a spaceborne multispectral imager, Journal of Geophysical Research, 114(D7), p. D07206. doi: 10.1029/2008JD010710.

Kaufman, 1989: The atmospheric effect on remote sensing and its correction. In Asrar, G., editor, Theory and Applications of Optical Remote Sensing, pages 336–428. Wiley and Sons, New York.

Krijger et al.,2011: Improved identification of clouds and ice/snow-covered surfaces in SCIAMACHY observations, Atmospheric Measurement Techniques, 4(10), pp. 2213–2224. doi: 10.5194/amt-4-2213-2011.

Lamarche et al., 2017: Compilation and Validation of SAR and Optical Data Products for a Complete and Global Map of Inland/Ocean Water Tailored to the Climate Modeling Community. Remote Sensing 9.1 (2017): 36. Web. 18 May 2017.

Lamarche et al., 2013: Characterizing the surface dynamics for Land Cover mapping. ESA Living Planet Symposium (Edinburgh, UK, du 09/09/2013 au 13/09/2013).

Lillesand & Keifer, 1994: Remote Sensing and Image Interpretation

Lillesand et al., 2000: Remote sensing and image interpretation, 4th Edition, New York: John Wiley & Sons, Inc.

Mekler and Kaufman, 1982: Contrast reduction by atmosphere and retrieval of nonuniform surface reflectance. Applied Optics, 21:310–316.

Miller, 1994,: Mapping the diversity of the nature, London, New York: Chapman & Hall

NetCDF, 2010: NetCDF Climate and Forecast (CF) Metadata Conventions. Issue 1.5. Date 25.10.2010.

Pesaresi et al., 2016: Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. Publications Office of the European Union, EUR 27741 EN, 2016. doi: 10.2788/253582.

Pfeffer et al., 2014: The Randolph Glacier Inventory: A Globally Complete Inventory of Glaciers. Journal of Glaciology 60 (221): 537–52. https://doi.org/10.3189/2014JoG13J176.

Santer and Ramon, 2011: MERIS ATBD - Aerosol remote sensing over land, ATBD2.15, v4

Schuckman, 2020: GEOG 480, Exploring Imagery and Elevation Data in GIS Applications, Lesson 8: Terrain Modeling and Analysis, Slope, Aspect, and Hillshade, John A. Dutton e-Education Institute, College of Earth and Mineral Sciences, The Pennsylvania State University. https://www.e-education.psu.edu/geog480/node/490

Simmons et al. 2006: 17713-Era-Interim-New-Ecmwf-Reanalysis-Products-1989-Onwards', ECMWF Newsletter, (110). doi: 10.21957/pocnex23c6.

Townshend et al., 2008: Integrated Global Observations of the Land: an IGOS-P Theme. IGOL Report No. 8, GTOS 54.

UNEP-WCMC, 2002: Global Mangrove Atlas. http://geodata.grid.unep.ch/options.php?selectedID=181&selectedDatasettype=16, accessed: October 15, 2018

Vancutsem et al., 2007: Mean Compositing, an Alternative Strategy for Producing Temporal Syntheses. Concepts and Performance Assessment for SPOT VEGETATION Time Series. International Journal of Remote Sensing 28.22 (2007): 5123–5141. Web. 3 Apr. 2014.

Vermote et al., 1997a:. Atmospheric correction of visible to middle infrared EOS-MODIS data over land surface: Background, operational algorithm and validation, Journal of Geophysical Research, 102:17131–17141.

Vermote et al., 1997b: Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview, IEEE Transactions on Geoscience and Remote Sensing, 35:675–686.

Wevers et al., 2022: IdePix for Sentinel-3 OLCI Algorithm Theoretical Basis Document (1.0). Zenodo. https://doi.org/10.5281/zenodo.6517333

Copernicus Climate Change Service

ECMWF - Robert-Schuman-Platz 3, 53175 Bonn, Germany

Contact: https://support.ecmwf.int/