Climate Change Initiative Extension (CCI+) Phase 1
New Essential Climate Variables (NEW ECVS)
High Resolution Land Cover ECV (HR_LandCover_cci)

Algorithm Theoretical Basis Document

(ATBD)

Prepared by:

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

Changelog

Issue	Changes	Date
1.0	First version.	02/07/2019
1.1	Revision according to CCI_HRLC_Ph1_KO+9_RID_ESA.xlsx	25/09/2019
2.0	Document entirely updated according to current status of the processing chain. Final legend is included.	03/01/2020
3.0	Global revision of the document updated with processing chain with its state at the end of the second year of activity.	23/11/2020
3.1	Updated version according to CCI_HRLC_Ph1_AR2_RID-ESA.xlsx	04/12/2020
4.0	Updated figures in Section 2. Section 3 has been updated with the pre-processing improvements. Section 6 has been updated, SVM description has been moved here, the use of GLCM and DEM as feature is reported, the inclusion of training points from the available land cover agreement and the division of areas in ecoregions has been included. Section 7.2 has been updated with the descriptions of the urban and water detectors. Section 8 now describes the cascade model, the harmonization modules (spatial and temporal) and the fusion post processing. Section 9.2 has been added with the description of the PCC mask application in the CD production. Section 9.4 has been updated with the steps of the multitemporal change detection processing chain. Section 8.5.2 has been updated with the description of the temporal consistency check. Section 10 has been added with the description of the post processing. The whole document includes updated references.	31/10/2022

Detailed Change Record

Issue	RID	Description of discrepancy	Sections	Change
1.1	FR-01	The ATBD should report a detailed description of the algorithms and methodologies (reported in the technical proposal) that should be used to achieve the objective of the project. We understand that RR#1 activities will provide better indications on which algorithm candidates on classification, but a more detailed description of the listed methods is needed.	7,8,9	Sections are integrated with more detailed information and mathematical insights.
1.1	FR-02	Why as Global Product to use as reference the unique map described is CORINE LC? CLC is not global.	7.1.1 (removed)	Mention to CORINE LC product as global product has been removed.
1.1	FR-03	Why training the S1 data using as reference the 300m CCI-LC maps? We are going to lose the HR of S1 data, or am I wrong? Please add some reference document using this technique	7	Further clarification has been added.
3.1	FR-01	From the documentation it seems that you will apply Sen2Cor to S2 L1C data, but this will not be the case because you will download S2 L2A, therefore the AC will be	2,3,7	The atmospheric correction has been removed from the block scheme, and the input

		applied only on Landsat data if you will download L1. (e.g. in Figure 2 AC box should be removed for S2 data).		products to the processing chain have been specified.

Contents

1 Introduction

1.1 Executive summary

Algorithm development is specifically aimed at addressing the technical requirements as provided by the outcome of Task 1 of the project. Best performing algorithms have been selected among proposed candidates through an internal benchmarking-testing iteration by the Earth Observing Science (EOS) team, as documented in PVASR. The final processing chain as developed in the three years of project activity is presented in this version of the document.

1.2 Purpose and scope

The Algorithm Theoretical Basis Document (ATBD) details the algorithms included in the processing chain needed to produce the land cover products as presented in the PSD [AD3]. It is intended to provide information for the understanding of the processing chain as a whole. This ATBD version is mature, it encapsulates all the updates and modifications that were introduced to address the feedbacks coming from the validation and intercomparison activities during the first benchmarking and testing stages of the project. This version of the document integrates in the whole project workflow and it identifies as main blocks of computation the following ones:

- Optical pre-processing.
- SAR pre-processing.
- Multi-sensor geolocation.
- Optical data classification.
- SAR data classification.
- Decision fusion.
- Multitemporal change detection and trend analysis.

1.3 Applicable documents

Ref.   Title, Issue/Rev, Date, ID

[AD1]   CCI HR Technical Proposal, v1.1, 16/03/2018

[AD2]   CCI Extension (CCI+) Phase 1 – New ECVs – Statement of Work, v1.3, 22/08/2017, ESA-CCI-PRGM-EOPS-SW-17-0032.

[AD3]   CCI_HRLC_Ph1-D1.2_PSD, latest version

[AD4]   CCI_HRLC_Ph1-D1.1_URD, latest version

[AD5]   CCI_HRLC_Ph1-D2.1_PVASR, latest version

1.4 Acronyms and abbreviations

6S	Second Simulation of a Satellite Signal in the Solar Spectrum
AC	Atmospheric correction
AMI	Active Microwave Instrument
AOT	Aerosol Optical Thickness
ASAR	Advanced Synthetic Aperture Radar
ATBD	Algorithm Theoretical Basis Document
BEAST	A Bayesian Estimator of Abrupt change, Seasonal change, and Trend
BFAST	Breaks For Additive Seasonal and Trend
BOA	Bottom of Atmosphere
BoW	Bag of visual Words

CCI	Climate Change Initiative
CD	Change Detection
CFMask	C Version of Function Of Mask
CMA	Climate Modeling Grid - Aerosol
CMG	Climate Modeling Grid
CNN	Deep Convolutional Neural Network
CVA	Change Vector Analysis
DARD	Data Access Requirement Document
DDV	Dark Dense Vegetation
DEM	Digital Elevation Model
DM	Dissimilarity Measure
DSM	Digital Surface Model
DTW	Dynamic Time Warping
ECV	Essential Climate Variables
ERS	European Remote Sensing
ETM	Enhanced Thematic Mapper
ETM+	Enhanced Thematic Mapper Plus
FC	Fully Connected
FS	Feature Space
GCOS	Global Climate Observing System
GMM	Gaussian Mixture Model
GSFC	Goddard Space Flight Center
HLS	Harmonized Landsat/Sentinel-2
HR	High Resolution
IFK	Improved Fisher Kernel
INT	Integer
IRMAD	Iteratively-Reweighted Multivariate Alteration Detection
L-5/7/8	Landsat-5/7/8
LandTrendr	Landsat-based detection of Trends in Disturbance and Recovery
LaSRC	Landsat Surface Reflectance Code
LC	Land Cover
LCC	Land Cover Change
LEDAPS	Landsat Ecosystem Disturbance Adaptive Processing System
LLC	Locality-constrained linear coding
LOP	Linear Opinion Pool
LPF	Low Pass Filter
LSTM	Long Short Term Memory
LTS	Landsat Time Series
LUT	Lookup Table
MDDTW	Multi-Dimension DTW
MEaSUREs	Making Earth Science Data Records for Use in Research Environments

MF-DTW	Multi-Feature DTW
MGRS	Military Grid Reference System
MLP-NN	Multi-Layer Perceptron Neural Network
MMU	Minimum Mapping Unit
MODIS	Moderate Resolution Imaging Spectroradiometer
MR	Medium Resolution
MSS	Multispectral Scanner
NA	Not Applicable
NASA	National Aeronautics and Space Administration
NCEP	National Centers for Environmental Prediction
NDI	Normalized Difference Index
NDVI	Normalized Difference Vegetation Index
NIR	Near infrared
NSPI	Neighbourhood Similar Pixel Interpolator
OA	Overall Accuracy
OLI	Operational Land Imager
OMI	Ozone Monitoring Instrument,
PCA	Principal Component Analysis
PSD	Product Specification Document
QA	Quality Assessment
RBF	Radial Basis Function
RD	Range Doppler
REFEREE	Learning a transferable change Rule From a recurrent neural network for change detection
RF	Random Forest
RNN	Recurrent Neural Network
S-1/2	Sentinel-1/2
S2AC	Sentinel-2 Atmospheric Correction
SAR	Synthetic Aperture Radar
SIFT	Scale Invariant Feature Transform
SITS	Satellite Image Time Series
SLC	Scan-line corrector
SM	Similarity Measure
SoW	Statement of Work
SR	Surface Reflectance
SRTM	Shuttle Radar Topography Mission
SSFA	Supervised Slow Feature Analysis
ST	Similarity Trend
STWR	Spatially and temporally weighted regression
SVM	Support Vector Machine
SWIR	Short-wave infrared
TIMESAT	Time Series of Satellite data

TIRS	Thermal Infrared Sensor
TM	Thematic Mapper
TOA	Top of Atmosphere
TOMS	Total Ozone Mapping Spectrometer
TS	Time Series
UEXT	Urban EXTraction
UTM	Universal Transverse of Mercator
VHR	Very High Resolution
VLAD	Vector of locally aggregated descriptors
WGS84	World Geodetic System 1984
XML	Extensible Markup Language

2   Processing chain overview

The CCI HRLC project will deliver to the climate community regional land cover (LC) and land cover change (LCC) products over three areas in Africa Sahel band, Amazon and Siberia URD [AD4]. LC maps will be provided at 10m resolution for year 2019 (the so-called Static Map) and at 30m resolution for the historical record of LC and LCC from 1990 on, every five years. The high-resolution classification legend as agreed by the Consortium is listed in URD [AD4]. The processing chain, outlined in Figure 1 and Figure 2, is novel and it does not rely on already existing land cover products.
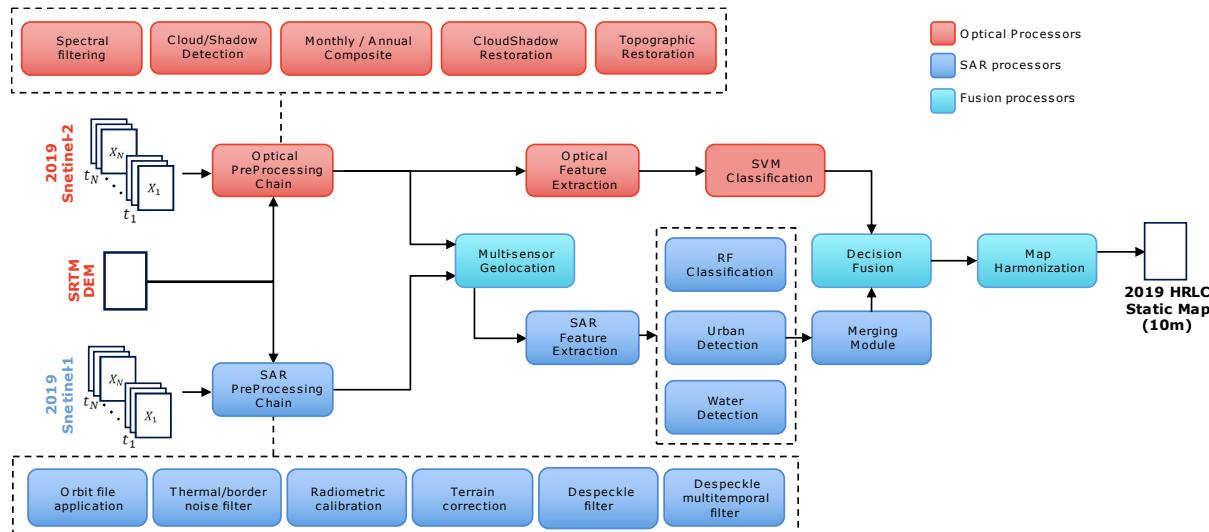
Figure 1. Block-based representation of the processing chain for the production of static HRLC maps.
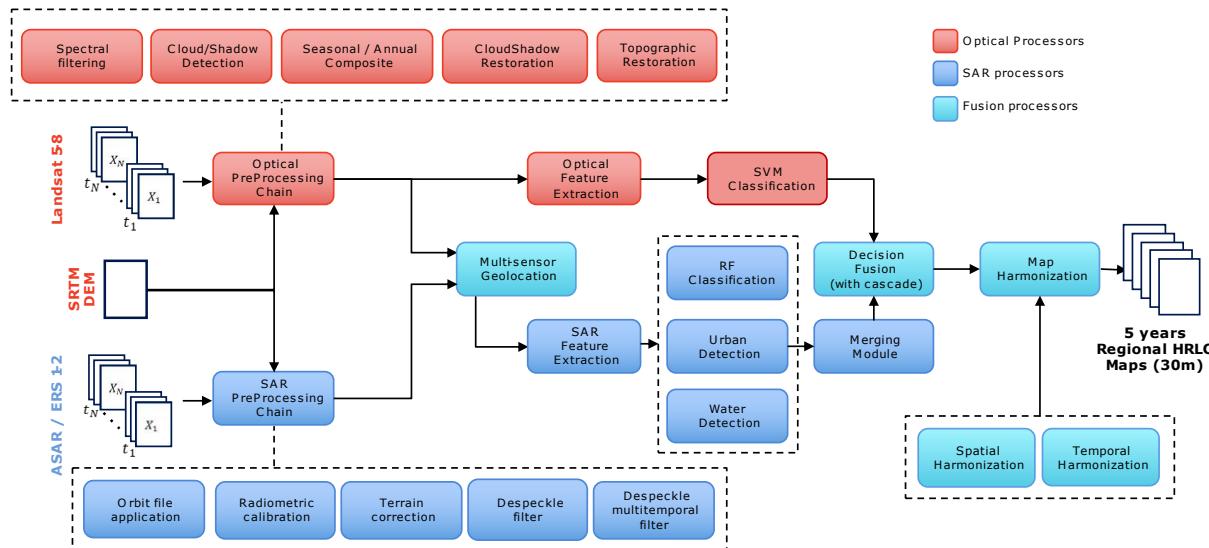
Figure 2. Block-based representation of the processing chain for the production of historical HRLC maps.

The high-level workflow of the processing chain is presented in Figure 1 and Figure 2. Optical multispectral imagery is the main source of data as input for the classification. The optical processing chain is consistent with the possibility to work mainly with images at 10/30m resolution and generating an output at 10/30m, based on multitemporal multispectral data from S-2 and L-8 in the recent years and legacy Landsat-5/7/8 data in the past. The SAR processing chain will be implemented mainly for S-1 in the recent years, and ERS and ASAR data sets in the past (whenever and wherever HR mode data are available). Microwave data sets are useful for classes where SAR has proven to be accurate at medium resolution, such as water bodies and coastal lines, and the option to use SAR for urban areas is considered as well. The products obtained by the optical and the SAR processing chains

will be then integrated in the data fusion module in order to produce the final HRLC products. This design choice of fusion at the decision level makes it possible to develop advanced and ad hoc processing approaches for optical, SAR, and multisensor data, while keeping the system modular and scalable. The output products will be then analyzed in the multitemporal change detection and trend analysis block for identifying different change components to be used for the historical time series HRLC products every 5 years.

Table 1. Final high resolution HR Land Cover classification legend defined during the HRLC project activity.

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

3 Optical pre-processing

Pre-processing operations are intended to correct for sensor- and platform-specific radiometric and geometric distortions of data and harmonization. Radiometric corrections may be necessary due to variations in scene illumination and viewing geometry, atmospheric conditions, and sensor noise and response. Each of these will vary depending on the specific sensor and platform used to acquire the data and the conditions during data acquisition. Cloud coverage is a systematic issue related to optical imagery and it requires specific processing aimed at precisely locating cloud and shadow pixels, with possible restoring steps to recover spectral information over occluded pixel locations. All the steps needed to prepare optical images for classification, see Figure 3, are detailed in the following sections.
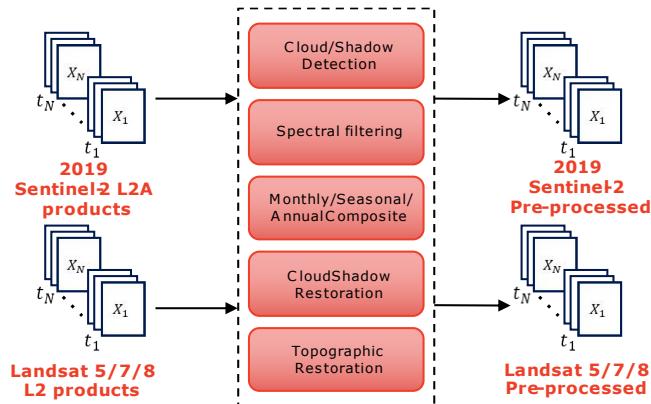
Figure 3. Optical pre-processing chain.

The input data to the processing chain are the atmospherically corrected Sentinel 2 data (i.e., L2A products) and atmospherically corrected Landsat 5-7-8 images (i.e., L2 products). Although the data used in the processing chain are already atmospherically corrected, in the following subsections we described the algorithms used to generate such products as well as the cloud and shadows masks.

3.1 Cloud and cloud shadow detection

3.1.1 Sentinel-2 – sen2cor

The sen2cor processor allows calculation of Bottom of Atmosphere (BOA) reflectance from Top of Atmosphere (TOA) reflectance images available in Level-1C products. Sentinel-2 atmospheric correction (S2AC) is based on an algorithm proposed in [1]. The method performs atmospheric correction based on the LIBRADTRAN radiative transfer model presented in [2].

The model is run once to generate a large LUT of sensor-specific functions (required for the AC: path radiance, direct and diffuse transmittances, direct and diffuse solar fluxes, and spherical albedo) that accounts for a wide variety of atmospheric conditions, solar geometries and ground elevations. This database is generated with a high spectral resolution (0.6 nm) and then resampled to S-2 spectral responses. This LUT is used as a simplified model (running faster than the full model) to invert the radiative transfer equation and to calculate BOA reflectance. All gaseous and aerosol properties of the atmosphere are either derived by the algorithm itself or fixed to an a priori value.

S2AC employs Lambert's reflectance law. Topographic effects can be corrected during the surface retrieval process using an accurate Digital Elevation Model (DEM). S2AC accounts for and assumes a constant viewing angle per tile (sub-scene). The solar zenith and azimuth angles can either be treated as constant per tile or can be specified for the tile corners with a subsequent bilinear interpolation across the scene.

3.1.2 Landsat 5/7/8 – LEDAPS, LaSRC

Landsat-4/5 TM and Landsat-7 ETM+ Surface Reflectance are generated using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm, a specialized software originally developed through a National Aeronautics and Space Administration (NASA) Making Earth System Data Records for Use in Research Environments (MEaSUREs) grant by NASA Goddard Space Flight Center (GSFC) and the University of Maryland [3]. The software applies Moderate Resolution Imaging Spectroradiometer (MODIS) atmospheric correction routines to Level-1 data products. Water vapor, ozone, geopotential height, aerosol optical thickness, and digital elevation are input with Landsat data to the Second Simulation of a Satellite Signal in the Solar Spectrum (6S) radiative transfer models to generate TOA reflectance, surface reflectance, TOA brightness temperature, and masks for clouds, cloud shadows, adjacent clouds, land, and water. Landsat 8 OLI Surface Reflectance are generated using the Landsat Surface Reflectance Code (LaSRC) [4], which makes use of the coastal aerosol band to perform aerosol inversion tests, uses auxiliary climate data from MODIS, and a unique radiative transfer

model. LaSRC hardcodes the view zenith angle to "0", and the solar zenith and view zenith angles are used for calculations as part of the atmospheric correction.

While both the LEDAPS and LaSRC algorithms produce similar SR products, the inputs and methods to do so differ. The table below illustrates both of them.

Table 2. Differences between Landsat-4/5/7 and Landsat-8 surface reflectance algorithms.

Parameter	Landsat-4/5/7 (LEDAPS)	Landsat-8 (LaSRC)
Global Coverage	Yes	Yes
TOA Reflectance	Visible (Bands 1–5,7)	Visible (Bands 1–7, 9 OLI)
TOA Brightness Temperature	Thermal (Band 6)	Thermal (Bands 10 & 11 TIRS)
SR	Visible (Bands 1-5, Band 7)	Visible (Bandsat 1-7) (OLI only)
Thermal bands used in Surface Reflectance processing	Yes (Brightness temperature Band 6 is used in cloud estimation)	No
Radiative transfer model	6S	Internal algorithm
Thermal correction level	TOA only	TOA only
Thermal band units	Kelvin	Kelvin
Pressure	NCEP Grid	Surface pressure is calculated internally based on the elevation
Water vapor	NCEP Grid	MODIS CMA
Air temperature	NCEP Grid	MODIS CMA
DEM	GTOPO5	GTOPO5
Ozone	OMI/TOMS	MODIS CMG Coarse resolution ozone
AOT	Correlation between chlorophyll absorption and bound water absorption of scene	MODIS CMA
Sun angle	Scene center from input metadata	Scene center from input metadata
View zenith angle	From input metadata	Hard-coded to "0"
Undesirable zenith angle correction	SR not processed when solar zenith angle > 76 degrees	SR not processed when solar zenith angle > 76 degrees
Pan band processed	No	No
XML metadata	Yes	Yes
Top of Atmosphere Brightness Temperature calculated	Yes (Band 6 TM/ETM+)	Yes (Band 10 & 11 TIRS)
Cloud mask	CFMask	CFMask

Data format	INT16	INT16
Fill values	-9999	-9999
QA bands	Cloud	Cloud
	Adjacent cloud	Adjacent cloud
	Cloud shadow	Cloud shadow
	DDV	Aerosols
	Fill	Cirrus
	Land water	Aerosol In
	Snow	
	Atmospheric opacity	

Identification of clouds, cloud shadows in optical images is necessary. The well-known program named Fmask has been used to accomplish these tasks for use with images from Landsat-5/7/8 [5].

3.1.3   Sentinel-2 – sen2cor improvement

Cloud and cloud shadow detection are based on cloud and shadow mask provided with the Sen2Cor (for Sentinel-2) and Fmask (for Landsat). The OA of cloud and shadow masks provided by the Sen2Cor (84%) is on average lower than the one provided by Fmask (90%) [6]. Therefore, the Sen2cor masks should be further enhanced to achieve the required accuracy. To this end, we adopt two strategies, one for cloud detection and one for cloud shadow detection and removal.

For cloud detection, we compute the cloudless background image for each season [7]:

$$x_j^B = quantile_{0.25}\{x_{j,Blue}^1, x_{j,Blue}^2, \dots, x_{j,Blue}^S\}$$

The difference between the blue bands of each image from the TS and the background image is computed. The pixels in the difference image are then clustered into 3 clusters. To understand which from the obtained clusters belong to cloud cover, the mean of each cluster is compared with the blue band mean of the cloudy pixels overall image. Finally, we merge of the obtained cloud mask with the original Sen2cor mask. Note that this strategy is performed only for tiles with a sufficiently large cloud cover, in order to properly model the clusters. Figure 4 shows the flowchart of the considered strategy.
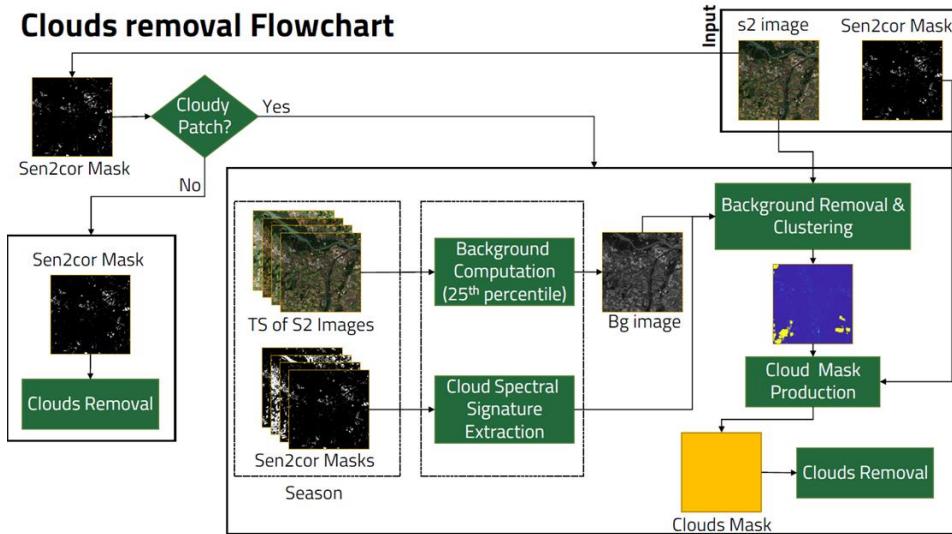
Figure 4. Flowchart of the sen2cor cloud mask improvement

In order to detect and remove cloud shadows, the cloud shadow index (CSI) can be used, which is based on the physical reflective characteristic of cloud shadow [8]. The CSI index is computed by combining information provided by the NIR and SWIR bands:

$$CSI \ = \frac{1}{2} * (B_{NIR} + B_{SWIR})$$

To avoid confusion between shadows and water bodies, as they both have very similar spectral signatures associated with their low reflectance, an additional condition including shorter wavelengths, i.e., the blue band reflectance, should also be analyzed. Thus, the cloud shadow is identified in areas where the following conditions are fulfilled:

$$CSI < \min(CSI) + \ t_1 * (\text{mean(CSI)} - \min(CSI))$$

$$B_{Blue} < \min(B_{Blue}) + \ t_2 * (\text{mean}(B_{Blue}) - \min(CSI))$$

Coefficients were fine-tuned: t1 =1/2 and t2=1/4. Note that this approach is performed only for tiles where there is enough cloud cover and the cloud cover has on average a large reflectance. Figure 5 shows the flowchart of the considered strategy.
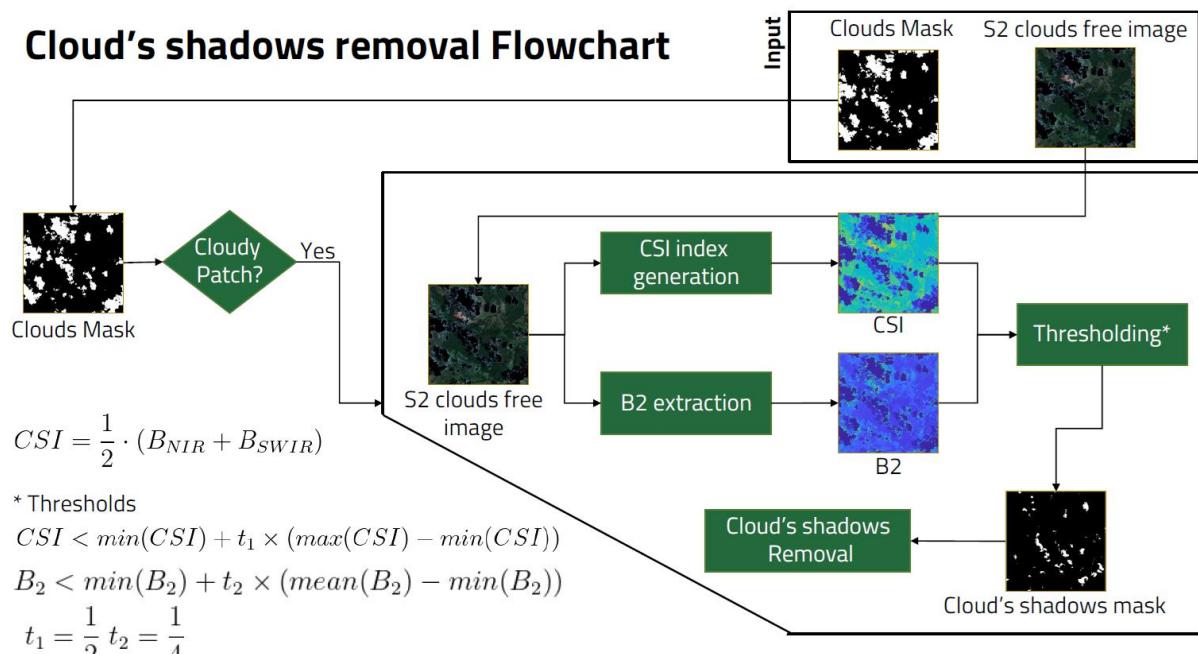
Figure 5. Flowchart of the sen2cor cloud shadow mask improvement and removal

3.2 Spectral Filtering

The spectral filtering aims to detect and remove the outlier present in the optical images. To this end, in this step we discard the reflectance values higher than the 0.999 quantile and lower than the 0.001 quantile of each spectral band. All the images considered in the experiments have cloud coverage less than 40%. In order to mitigate any possible effect of clouds and shadow present on the image, they have been detected by using Sen2cor mask and discarded from the quantitative evaluation.

3.3 Composite Generation

When working at large scale, it is necessary to harmonize the times-series of images acquired over different tiles which are characterized by different lengths and are acquired at different times. This is mainly due to the irregular cloud coverage (which hampers the use of some images of the time-series) and the different orbit acquisitions (different temporal sampling). To solve this problem, in the pre-processing step we generate monthly, seasonal and annual composites. This condition allows us to mitigate cloud occlusions problem and minimize the processing resources. To this end, we consider a statistic-based approach that computes the median value for each pixel. This approach is able to generate consistent results at large scale in an automatic way by sharply reducing the spatial noise. Let $\{X^1, X^2, ..., X^Q\}$ be the considered time-series which includes the optical images acquired over a month, a season or the whole year (i.e., according to the sensor and the

	Ref	CCI_HRLC_Ph1-ATBD	
esa	Issue	Date	Page
	4.rev.0	31/10/2022	16

high resolution
land cover
cci

considered study area). The multitemporal pattern associated to the jth pixel of $\{X^1, X^2, \ldots, X^Q\}$ can be defined as $[\mathbf{x}_j^1, \mathbf{x}_j^2, \ldots, \mathbf{x}_j^Q]$, where $\mathbf{x}_j^1 = [x_{j,1}^1, x_{j,2}^1, \ldots, x_{j,B}^1]$ represents the B spectral values of the jth pixel in the first image of the time-series. The jth pixel of the composite x_j^{Com} is generated by computing the band-wise median of the cloud-free multispectral pixels of the images present in $\{X^1, X^2, \ldots, X^Q\}$ as follows:

$$x_{j,1}^{Com} = Med\{x_{j,1}^1, x_{j,1}^2, \ldots, x_{j,1}^Q\}$$
$$x_{j,2}^{Com} = Med\{x_{j,2}^1, x_{j,2}^2, \ldots, x_{j,2}^Q\}$$
$$\vdots$$
$$x_{j,B}^{Com} = Med\{x_{j,B}^1, x_{j,B}^2, \ldots, x_{j,B}^Q\}$$

It is worth noting that cloud, shadow and snow mask pixels are ignored during median computation. Table 3 summarizes the kind of composite generated per study area according to different optical sensor. Due to the increased revisit time of Sentinel 2 (5 days) with respect to Landsat (16 days), denser time-series are available for 2019 that can be used to generate monthly composites. In the case of Sentinel data over Amazonia and Africa, we computed 12 monthly composites. Due to dense cloud coverage over some regions, each monthly composite is computed using the considered and the following month (e.g., January and February are used for January composite generation, the only exception is December, where only December images are considered). This conservative choice allows us to sharply reduce the probability of having cloudy pixels in the time-series. For Sentinel data over Siberia, we generate yearly composite due to heavy cloud and the snow coverage problem which hampered the use of images acquired for most of the year. Hence, the Siberian yearly composite is computed as the median of data acquired in July and August.

In the case of Landsat data, we similarly consider yearly composite for Siberia, which is computed as the median of data acquired between April and September. Finally, for Landsat data over Amazonia and Africa we compute four seasonal composites considering the optical data acquired in the following months: (i) January – March, (ii) April – June, (iii) July- September, and (iv) October – December.

Table 3. Composites generated for the different study areas according to the availability of cloud free optical images.

Area	Sentinel 2	Landsat 5/7/8
Siberia	Yearly (July - August)	Yearly (April – September)
Amazonia	12 Monthly Composites	4 Seasonal Composites
Africa	12 Monthly Composites	4 Seasonal Composites

3.4   Cloud and cloud shadow restoration

3.4.1   Landsat-7 SLC-off

The scan-line corrector (SLC) of the Landsat-7 Enhanced Thematic Mapper Plus (ETM+) sensor failed in 2003, resulting in about 22% of the pixels per scene not being scanned. The SLC failure has seriously limited the scientific applications of ETM+ data. In particular, this problem affects the considered composite strategy when the available acquisitions are scarce and come mainly or only from Landsat-7 (e.g., Africa 2005 and 2010). In order to avoid the considered composite strategy and the classifier to be affected by this problem (i.e., striping in the composites and in the land-cover maps), a gap-filling strategy based on interpolation has been used to fill in the values of the missing pixels. While accurate spatial information is not retrieved, the considered composite strategy is able to partially retrieve it by exploiting the multitemporal acquisitions. Even though the spatial detail might be reduced, this strategy results in improved spectral uniformity and consistency across pixels in the composite. This improved the performance of the classifier, which uses the spectral bands as primary features.

	Ref	CCI_HRLC_Ph1-ATBD	high resolution
esa	Issue	Date	Page	land cover cci
	4.rev.0	31/10/2022	17	

3.4.2 Cloud restoration

Cloud restoration is an important step in the optical images pre-processing part. Although we are considering the composites instead of original time-series of images, missing information due to poor atmospheric conditions (e.g., thick clouds and related shadows) or defective sensors may be present in the composites. In the literature, a large effort has been devoted to solve this problem. However, to properly recover missing information, sophisticated and usually computationally intensive techniques should be used, increasing significantly the computational complexity of the pre-processing part. Instead of considering computationally demanding approaches, a simple and effective linear temporal gap filling was employed. In this method the missing information are restored as the average of the spectral values acquired in the previous and the following images in the time-series. If clouds are present in the first or last image in time-series, the second or the one before last image are considered, respectively.

3.4.3 Topographical Shadow Reconstruction

Another problem associated with low reflectance values is related to topographical shadow. Here, similarly to the shadow detection, we identified topographic shadow by using thresholding of the CSI index. Again, to avoid confusion between shadow and water bodies, we consider also the topographical slope information. To restore topographic shadow, we rely on the assumption that the signal registered in the shadow areas is weak but can be exploited for shadow reconstruction. The shadow region can be retrieved according to the following equation:

$$y_j = S_f/S_s(x_j - M_s) + M_f$$

Where S_s is the standard deviation of the shadow, M_s is the mean value in the shadow, S_f is the standard deviation in the shadow-free surrounding area, and M_f is the mean values in the shadow-free surrounding area. To harmonize the obtained shadows restoration results and the surrounding spectral signatures of the original data, we used the inpainting technique based on the fast marching method presented in [9].

4 SAR pre-processing

For the production of Static Map at 10m resolution for 2019 Sentinel-1 data were adopted, while the mapping LC and LCC back to 1990 at 30m resolution requires the use of Sentinel-1, ERS 1/2 and ENVISAT ASAR acquisitions.

We considered Sentinel-1 data acquired in Interferometric Wide swath (IW) mode and Ground Range Detected (GRD) type, which derive from an application of a proper multi-looking and ground range projection based on an Earth ellipsoid model. The datasets are in High resolution (HR) and provide images with a native range by azimuth resolution 20×22 meters and pixel spacing equals to 10x10 m. Over land surfaces, the orbital period of each satellite is about 12 days. Consequently, acquisitions have been available since 2015 for time-periods of 6 or 12 days depending on the study region. About the other radar data used for LC and LCC products, we assumed the SAR Level 1 Precision Image Product (SAR_IMP_1P) of ERS 1/2 satellites and ASAR IM Precision L1 (ASA_IMP_1P) product given by ENVISAT platform.

The SAR_IMP_1P product is a multi-look (speckle-reduced), ground range image acquired in Image Mode. This product type has already been submitted to early preprocessing operations so as to make it very suitable in many remote sensing applications [10]. ASA_IMP_1P is a multi-look, ground range, digital Precision Image generated from Level 0 data collected when the instrument was in Image Mode (7 possible swaths HH or VV polarisation). The product includes slant range to ground range correction. The processing uses up to date (at time of processing) auxiliary parameters corrected for antenna elevation gain, and range spreading loss. Engineering corrections and relative calibration are applied to compensate for well-understood sources of system variability. This product provides a continuation of the ERS-SAR product [11].

By an accurate analysis of data availability into the past, we noted that SAR data do not present a good spatial and temporal coverage in all the years of the period of interest, which goes from 2015 to 1990. The major issue is related to the tile coverage, since is not uniform, and may generate misclassification effects. For this reason, the adopted strategy is based on the use of just a single season per year to produce the historical land cover

	Ref	CCI_HRLC_Ph1-ATBD	high resolution
esa	Issue	Date	Page	land cover cci
	4.rev.0	31/10/2022	18	

maps every five years from 2015 backwards. By this way, it is possible to achieve results for a larger set of tiles and considering a much more homogeneous data set.

The following table shows the final situation of SAR dataset distribution concerning the three areas, i.e., Amazonia, Africa, and Siberia, dedicated to the historical maps' production.

Table 4. SAR data availability for the historical production

Area	Year	Date range	Season	SAR historical product	# images
Amazonia	2005	01.01 - 03.31	Winter	ENVISAT_ASA.IMP.1P	466
Amazonia	2000	01.01 - 03.31	Winter	ERS_SAR.IMP.1P	396
Amazonia	1995	04.01 - 06.30	Spring	ERS_SAR.IMP.1P	421
Africa	2010	01.01 - 03.31	Winter	ENVISAT_ASA.IMP.1P	274
Africa	2005	07.01 - 09.30	Summer	ERS_SAR.IMP.1P	350
Africa	2000	07.01 - 09.30	Summer	ERS_SAR.IMP.1P	350
Africa	1995	04.01 - 06.30	Spring	ERS_SAR.IMP.1P	323
Siberia	2010	07.01 - 09.30	Summer	ERS_SAR.IMP.1P	895
Siberia	2005	07.01 - 09.30	Summer	ENVISAT_ASA.IMP.1P	315
Siberia	1995	07.01 - 09.30	Summer	ERS_SAR.IMP.1P	548

Looking at Table 4, we see that the SAR data distribution does not cover all the years of the historical period. As mentioned before, it is due to the poor availability of datasets in the past, which leads to the lack of any contribution by SAR to the land cover maps for 2015, 2010 and 1990 in Amazonia, 2015 and 1990 in Africa, and 2015, 2000 and 1990 in Siberia.

For processing and analysing SAR data, several codes have been developed in Python programming language, which were then deployed by means of dockers, i.e., general automated applications that can be launched in every OS.
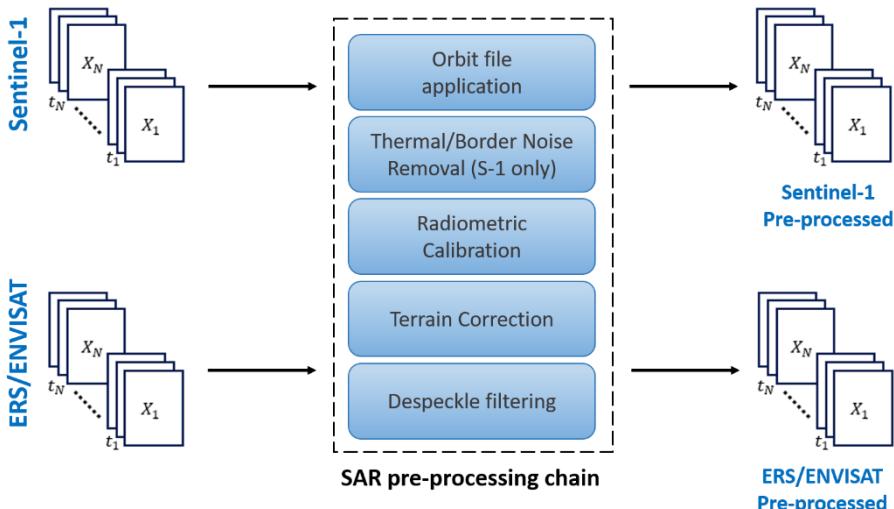

Figure 6. Block scheme of SAR pre-processing chain.

Before applying any classification algorithm on radar images, a preliminary pre-processing phase is required, and it consists in the following basic steps:

- Orbit File application
- Thermal Noise removal (for Sentinel-1 only);
- Border Noise removal (for Sentinel-1 only);
- Radiometric calibration of data;
- Geometric Terrain correction;
- Despeckle filtering.

4.1 Orbit File application

The first processing step is to apply the orbit files in SAR products to provide accurate satellite position and velocity information. Usually, the satellite flying orbit track is being detected by many sensors, like mounted gyroscope, GPS, and also ground observations. To calculate out the precise orbit data takes time, so the precise orbit data are not included in many SAR satellite data bundles. SNAP is trying to get more precise orbit data to help to improve the geocoding and other SAR processing results.

4.2 Thermal Noise removal

In general, an SAR image product contains not only the wanted signal but also the unwanted noise that is superposed within the same pixel. Among the various noise components in SAR systems, the thermal noise is an additive noise, which is processed with the same processing gains applied to the true signal.

In general, an SAR image product contains not only the wanted signal but also the unwanted noise that is superposed within the same pixel. Among the various noise components in SAR systems, the thermal noise is an additive noise, which is processed with the same processing gains applied to the true signal.

Sentinel-1 image intensity is disturbed by additive thermal noise, particularly in the cross-polarization channel [12]. Thermal noise removal reduces noise effects in the inter-sub-swath texture, in particular, normalizing the backscatter signal within the entire Sentinel-1 scene and resulting in reduced discontinuities between sub-swaths for scenes in multi-swath acquisition modes. The thermal noise removal operator available in SNAP for Sentinel-1 data can also re-introduce the noise signal that could have been removed during level-1 product generation, and update product annotations to allow for re-application of the correction. Sentinel-1 level-1 products provide a noise *Look-Up Table* (LUT), provided in linear power, for each measurement data set and used to derive calibrated noise profiles matching the calibrated GRD data [13].

4.3 Border Noise removal

While generating level-1 products, it is necessary to correct the sampling start time in order to compensate for the change of the Earth's curvature. At the same time, azimuth and range compression leads to radiometric artefacts at the image borders. The border noise removal algorithm [14],available as an operator in SNAP, was designed in order to remove low intensity noise and invalid data on scene edges.

4.4 Radiometric calibration

Radar images are firstly calibrated with respect to their intrinsic sensor and signal acquisition properties, for expressing the echoes of distributed target (e.g. grass, dirt, etc.) in terms of the radar backscattering coefficient. In other words, the VV and VH intensities are expressed in terms of sigma naught. Generally, this operation was performed during the generation of a SAR product, but for the land cover map generation is not recommended to use raw data because of the inconsistency of the uncalibrated signal. The radiometric calibration is therefore needed since the grey-value of SAR imagery must be adjusted respect the backscattering signals of the objects present into the scene.

4.5 Geometric terrain correction

Due to the active nature of the system, every SAR image is acquired in slant looking geometry. If the ground is elevated because of hills and valleys, the time of the signal to travel to the Earth surface and back to the sensor is distorted, causing geometric shifts in the image (foreshortening, layover and shadow). These can only be corrected if a model representing the topography under the image is known. In particular, the Range Doppler (RD) Terrain Correction is applied, and it shifts all pixels to their correct locations according to ancillary data Shuttle Radar Topography Mission (SRTM) 3 arc sec (i.e. around 20 m of resolution) DEM as input. RD Terrain Correction increases the location accuracy of your image. The first two steps of pre-processing phase have been conducted using ESA Sentinel-1 toolbox implemented in the official Sentinel Application Platform software provided by ESA (for more detailed information, ones should refer to the proper Wiki for Developer Documentation to [13].

4.6 Despeckle filtering

The SAR images are inherently affected by speckle that is a "noise like" signal due to the coherent nature of the electromagnetic scattering [15]. Even though speckle carries itself information about the illuminated area, it degrades the appearance of images and affects the performance of scene analysis tasks carried out by computer programs (e.g., segmentation and classification). To mitigate this problem several suitable filtering methodologies have been developed for reducing the disturbance significantly and preserve at the same time all the relevant scene features, such as radiometric and textural information. The speckle in SAR is a *multiplicative* effect, i.e. it is in direct proportion to the local grey level in any area. Speckle filtering is needed to suppress the noise in order to allow better interpretation and backscatter analysis. However, it is essential mentioning that the speckle filter not only suppress the noise, but also remove observations that are not affected by noise and contain valuable land surface information (i.e. soil moisture, biomass and flood extent). The process of removal of speckle in SAR image is very essential for the analyst to interpret. A filter should remove speckle without sacrificing image structures.

There are various speckle removal methods. Speckle removal is necessary for quantitative, analysis but there exists a tradeoff between speckle removal and resolution. Speckle Suppression can be done using various techniques. The first technique is Lee filter, known for being one of the first approach designed for suppressing speckle effect [16]. Second technique is time series-based processing. Proper developed docker containers provide both classical Lee method and a better suitable and advanced de-speckle filter (called *multitemporal de-speckle filter)* that exploits a SAR time series. Multitemporal denoising methods take advantage of the increasing availability of SAR image time-series to solve the spatial denoising problems, for the benefit of a better spatial resolution preservation.

4.6.1 Lee filter

The Lee filter is an adaptive filter, and reportedly to be the first model-based filter dedicated to speckle noise suppression [17]. It is also derived from the Minimum Mean-Square Error (MMSE) algorithm that converts the multiplicative model into an additive one, thereby reducing the problem of dealing with speckle to a known tractable case (more details are reported in [18]). In Lee filter, the statistical distribution of the values of the pixels within the moving kernel is utilized to estimate the value of the pixel of interest. This assumes that the mean and variance of the pixel of interest are equal to the local mean and local variance of all pixels within the user-selected moving kernel. The resulting grey level value Y for the smoothed pixel is:

$$Y = I_c W + I_m (1 - W),$$

where:

- $W = \left(1 - \frac{C_u{}^2}{C_i{}^2}\right);$
- $C_u = \sqrt{1/ENL};$

- $C_i = S/I_m$

- I_c is the central pixel of filter kernel;

- I_m value is the mean between all pixels falling within kernel;

- S is the standard deviation of all pixels falling within kernel;

W is the weighting function that measures the estimated noise variation coefficient C_u over the image variation coefficient C_i. The number of looks parameter ENL is the Equivalent Number of Looks of the radar image, which is used to estimate the noise variance and control the amount of smoothing applied to the image by the filter. The user may experimentally adjust the ENL value to control the effect of the filter. A small ENL value leads to more smoothing while a large ENL preserves more image features.

Several works [19], [20] have proven, with quantitative assessments, that a good tradeoff between speckle suppression, details and textures preservation is achieved with 5x5 or 7x7 moving kernel size. Moreover, the Lee filter is reportedly superior in its ability to preserve prominent edges, linear features, point target, and texture information, by minimizing either the mean square error or the weighted least square estimation.

4.6.2   Multi-look anti-speckle filter

Multi-look processing is known to improve SAR image quality because it reduces the speckle noise that produces the "salt and pepper" appearance of image. Either in range or azimuth direction or in both the directions, subsequent lines are averaged to get a better image. Averaging the different looks offers a trade-off in terms of resolution, and an image with reduced resolution but less grainy appearance can be formed.

(a)   (b)   (c)
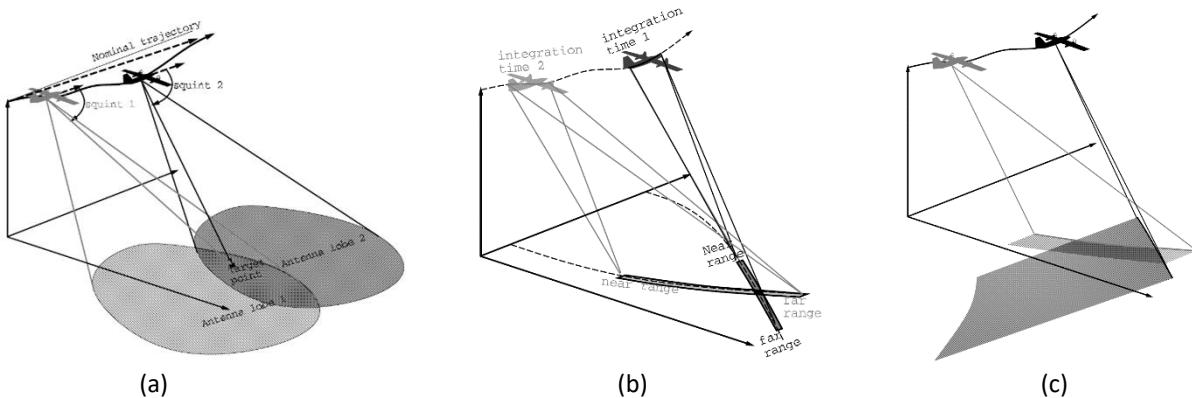
Figure 7. Principle of multi-look processing (a), acquiring a point on the ground from separated integration intervals (synthetic antennae) (b) and corresponding single-look images with range axis oriented along different squint angles (c).

Due to low directivity of the radar antenna, one given point on the ground can be illuminated by radar during a much longer time than integration time. For example, in Figure 7 (a) we have a point illuminated from the time it appears under the 1st squint (on the forward side of the antenna lobe) to the time it appears under the 2nd squint (on the rear side of the antenna lobe). SAR images can be computed for a same point with different integration intervals, as well as shown in Figure 7 (b) which depicts two image rows containing the target point for each integration interval. Clearly, due to different direction of observation, the "range axis" in the two images, also called "single-look", do not match as displayed in Figure 7 (c). The accurate matching of the single-look images requires a good geometrical model, especially if the platform trajectory is far from linear [21]. As it is well known, multi-look averaging has the advantage of reducing the speckle of the diffuse reflections on rough surfaces. One of the speckle property is that grain locations are independent for single-look images calculated from disjoint integration time intervals, thus averaging several single-look images (after co-registration in the same coordinate system) will provide a smoother multi-look image. This result on "real" rough textures is empirical. In fact, it is related to the statistical characteristics of the texture. Theoretically, a texture with fractal surface (fortunately, unlike any real surface is) would remain grainy whatever the number of looks combined.

The simplest approach to despeckling is to average the intensity over several pixels within a window centered on a specific pixel. This is tantamount to assuming that the information at each pixel carried by the *Radar Cross Section* (RCS) (or *Backscattering Coefficient* σ) is constant over the filter window. If this assumption is incorrect, the method is fundamentally flawed. The joint probability that all N pixels have this mean value is given by

$$P(\sigma|I_1, I_2, \ldots, I_N) \propto \prod_{j=1}^{N} P(I_j|\sigma) \propto \prod_{j=1}^{N} \left(\frac{L}{\sigma}\right)^L \frac{I_j^{L-1}}{\Gamma(L)} \exp\left[-\frac{LI_j}{\sigma}\right]$$

where I_j is the jth observed intensity. $P(\sigma)$ is the likelihood function which describes the effect of speckle during imaging for L -look SAR. Assuming N independent pixels, the *Maximum Likelihood Estimate* (MLE) of σ is then given by

$$\sigma_{ML} = \underline{I}$$

which is the average intensity over all the pixels in the window, corresponding to multi-looking. Note that if this is applied to a single pixel the MLE is equal to the intensity of that pixel. Only if features of interest within the scene occur on a large enough scale can multi-look images be effective. An adaptive approach that matches the size of a window to the scale of objects of interest would provide optimum despeckling at the required resolution. Where fluctuations were consistent with a constant RCS the window size can be increased to improve despeckling as far as possible [15]. Although multi-look causes a degradation in the image resolution, it greatly improves the interpretability of the SAR image. Also the effect of speckle tends to weaken for very high-resolution systems, since the number of elemental scatters within a resolution cell decreases.

4.6.3   Multitemporal despekling filter

The proposed approach is a ratio-based multitemporal denoising framework based on the use of a ratio image composed of a noisy image and the temporal mean of the stack. This ratio image is easier to denoise than a single image thanks to its improved stationarity. Besides, temporally stable thin structures are well preserved thanks to the multi-temporal mean [22]. Because of the improved spatial stationarity of the ratio images, denoising these ratio images with a speckle-reduction method is more effective than denoising images from the original multi-temporal stack. The amount of data that is jointly processed is also reduced compared to other methods through the use of the 'super-image' that sums up the temporal stack in order to fully exploit the significant information of the multi-temporal stack.

The method consists in three steps that are grouped into the following list and represented in Figure 8:

1.   Super image;
2.   Denoising of the ratio image;

	Ref	CCI_HRLC_Ph1-ATBD	
esa	Issue	Date	Page
	4.rev.0	31/10/2022	23

**high resolution
land cover**
cci

3. Computation of the final image through the multiplication between denoised ratio and super image.
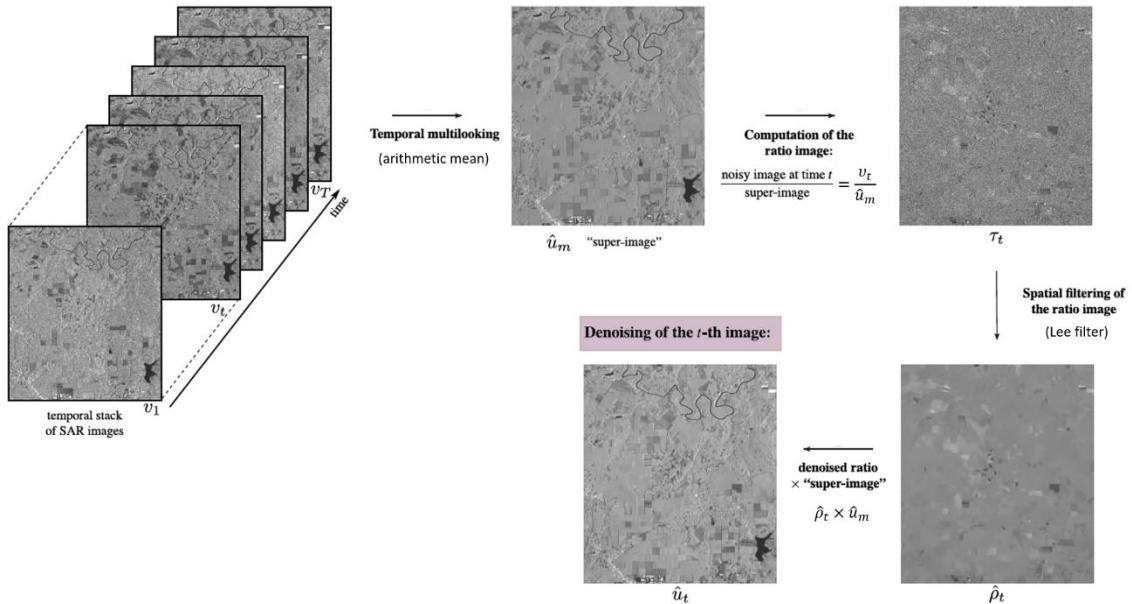
Figure 8. Summary of multitemporal despeckling method applied on SAR time series.

The temporal averaging (also called temporal multi-looking) of SAR time series generates an image with reduced speckle and a preserved spatial resolution, that has been identified as "super-image".

Given T spatially registered and radiometrically calibrated intensity SAR images $\{v_1, v_2, \ldots, v_t, \ldots, v_T\}$, the super image is given by a simple average that offers good properties in terms of modelling the statistics [23]. Hence, the arithmetic mean is calculated at pixel p by:

$$\hat{u}_m(p) = \frac{1}{T}\sum_{t=1}^{T} v_t(p) \qquad t \in [1, T].$$

After temporal averaging the second step consists in using the super-image to form the ratio image τ_t between the image v_t at time t and the super image $\hat{u}_m$, at each spatial location p:

$$\tau_t(p) = \frac{v_t(p)}{\hat{u}_m(p)}$$

It contains the residual speckle noise between the two images, and the radiometric shifts when changes occur. When the length of the time series increases and in the absence of change, the super image $\hat{u}_m$ converges to u_t, the reflectivity of the scene (the signal of interest). The ratio image τ_t then tends to pure speckle (i.e., a collection of independent identically distributed random variables with unitary mean and the same number of looks as the original image). In contrast, when changes occur in the time series, these changes impact the super image which then differs from the reflectivity u_t of the image at time t. Processing the ratio image τ_t is necessary to correctly recover the reflectivity u_t. Anyway, the ratio image still needs of speckle reduction methodologies since both the noisy image v_t and the super-image $\hat{u}_m$ suffer from speckle (although speckle in the super-image is strongly reduced). The use of this additional spatial filtering step to form the ratio image seems beneficial in terms of restoration quality: the obtained image is smoother.

Finally, in the latter step the filtered image is recovered by multiplying the denoised ratio image with the original super image $\hat{u}_m$. The estimated image $\hat{u}_t$ at location p is given by:

$$\hat{u}_t(p) = \hat{u}_m(p) \cdot \hat{\rho}_t(p)$$

Based on the processing of SAR stack corrupted by speckle noise, the approach has showed the potential to better preserve structures in multi-temporal SAR images while efficiently removing speckle. A classic application of this approach has been well reported in Figure 7.

4.7 Multiscale data merging

In case we do not have enough IW S-1 data to cover the area of interest, it may become necessary to merge data at different spatial resolution. This is achieved by a multi-scale SAR merging following [24]. We selected the Discrete Wavelet Transform and Histogram Matching framework (DWT/HM) because among all other filters, the DWT is the most common way for dealing the multiscale signal representation at pixel-level in simply and effective manner, due its ease implementation and the low computational cost [25].

Firstly, let us model the vector $X = [X_1, X_2, \dots, X_k]$ that represents a multiscale SAR dataset, with k satellite imagery that having different resolution levels. In particular, the elements are arranged in ascending order in terms of resolution level, where the data with subscript 1 has finest resolution while the k-th element denotes the product with coarsest resolution.

4.7.1 Discrete Wavelet Transform and Histogram Matching framework (DWT/HM)

Generally, the wavelet transform decomposes a signal into a set of basis so-called wavelets. The wavelet representation provides a way for analyzing signals in both time and frequency domains. This makes it ideal for representing non-stationary signals, to which most real-world signals belong. The DWT transforms a discrete time signal to a discrete wavelet representation [26]. This procedure carries out a lossy compression, since components of signal that are known to be redundant, are discarded. The classical DWT is implemented by considering two filters: low-pass (LPF) and high-pass (HPF) filters. The DWT method is implemented also in bi-dimensional (2D) case. In fact, in image processing, the image X_m, with $m = 1,2,..,k$, is filtered by means a high-pass and a low-pass filter combination. After the filtering, the outputs are all downscaled by a factor of two. In figure 2, a simple diagram that report the basic architecture of the DWT procedure is shown.
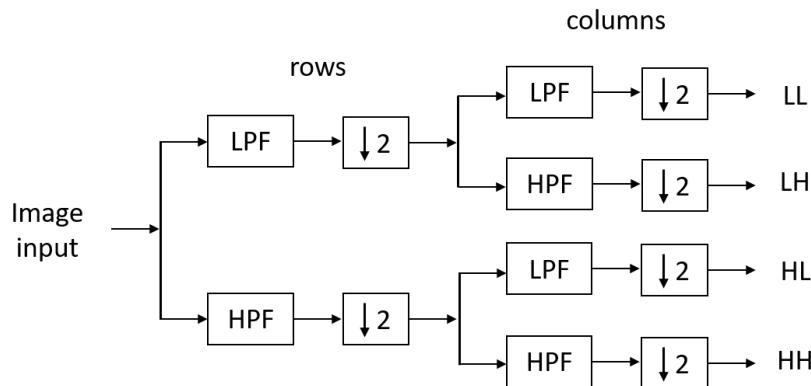
Figure 9. Block scheme of 2D DWT algorithm: the 1-Level 2D analysis DWT image decomposition process.

The original image is decomposed into four sub-band images, it deals with row and column directions separately. First, the HPF and the LPF are exploited for each row data, and then are down-sampled by two to get high- and low-frequency components of the row. Next, the high- and the low-pass filters are applied again for each high- and low-frequency components of the column, and down-sampled by two. By way of the above processing, the four sub-band images are generated: $HH_{(m+1)}$, $HL_{(m+1)}$, $LH_{(m+1)}$, and $LL_{(m+1)}$, with a resolution level equals to $(m + 1)$ due to the down-sampling (note that we used the round brackets for emphasize that we were passed from m to (m+1) resolution by applying the DWT approach). Each sub-band image has its own feature, such as the low-frequency information is preserved in the $LL_{(m+1)}$-band (named *context image* also) and the high-frequency information is almost preserved in the $HH_{(m+1)}$-, $HL_{(m+1)}$-, and $LH_{(m+1)}$-bands.

The $LL_{(m+1)}$-subband image can be further decomposed in the same way (in recursive manner) for the second level sub-band image. By using 2D DWT, an image can be decomposed into any level sub-band images, as shown in **Figure 10**.

	Ref	CCI_HRLC_Ph1-ATBD		high resolution
	Issue	Date	Page	land cover
	4.rev.0	31/10/2022	25	cci
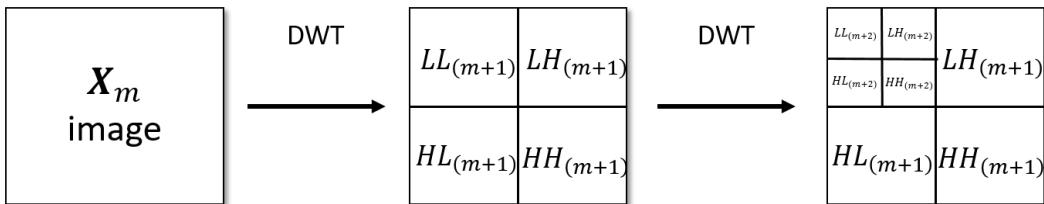
Figure 10. Diagrams of DWT image decomposition: the 2-Level 2D analysis DWT subband.

For carrying out the data fusion between two images having different resolution levels, X_i and X_j, with $i < j$, for example, we could be use the DWT in addition to the standard Histogram Matching (HM) method. The HM is a transformation used for generating an image that is harmonized from a statistical point of view, since its probability distribution function matches a specified histogram [27].

In fact, the DWT is recursively applied on the finer image X_i until achieving the desired level j, and the $LL_{(j)}$ subimage is hence calculated. Then, the HM is applied on the coarser image X_j in order to obtain the HM image version, X_j^{HM}(calculated respect the target image $LL_{(j)}$) for resampling it on a common (much finer) grid. The next step provides to substitute $LL_{(j)}$ with the derived image given by the mean $LL_{(j)}$ and X_j^{HM}, i.e. $(LL_{(j)} + X_j^{HM})/2$. Finally, we derive the data fusion result at $i - th$ resolution level by applying the inverse DWT, i.e. the reconstruction process, opposite to the decomposition one, is formed by synthesis filters and up-samplers [28] going back until the finer scale.

The method can be implemented for the whole multiscale dataset X, starting from two images with coarsest resolution. The DWT procedure is hence iteratively repeated by using the fused result and the image with the finest resolution (among all those still unused for the fusion) as input. The output is a unique final fused image.

As a backup, in the procedure has been implemented a second method for multi-scale SAR merging, the Multiscale Kalman Filter (MKF) can be considered. MKF is a pyramidal approach where the spatial resolution is assumed as an independent variable as the time. As described in [29], the MKF algorithm can be applied following two different modes respect the DWT one, since the fusion data with different scales might be carrying out starting both from finer resolution data to coarser resolution (upward step).

5   Multi-sensor geolocation

Given the outputs of the optical and SAR pre-processing chains, a further pre-processing stage, prior to their joint use for land cover mapping, is generally necessary to make them spatially aligned. In general terms, the process of aligning different sets of image data and of referencing them into a common coordinate system (Figure 11) is named image registration. Input data for registration may be multiple photographs, data from different sensors, times, or viewpoints [30]. One image is taken as the "reference image", and all other images are registered to the reference image and are called "sensed (or input) images". Besides remote sensing, it is used in computer vision, medical imaging, military automatic target recognition, etc. Registration is necessary in order to be able to compare or integrate the data corresponding to the same scene but obtained from different measurements. Here, the focus is put on multi-sensor geolocation, which corresponds to the case where image registration is applied to data gathered by different sensors, namely optical and SAR sensors in the CCI+ HRLC pipeline.

Figure 11: Image registration example with aerial photos.

By definition, multi-sensor geolocation enables the integration of complementary information from different sensors. A registration method is broadly composed of different elements, i.e.: (i) the geometric transformation used to warp the input image; (ii) the similarity measure used to compare the reference and input images during the registration process; and (iii) the optimization strategy used to minimize or maximize the similarity measure, depending on the semantic of the metric.

Figure 12: Building blocks of multi-sensor geolocation.

The following subsections cover each one of such aspects, focusing on the choices related to the processing steps of the multi-sensor geolocation block in the CCI+ HRLC pipeline. Hence, Section 5.1 describes all the geometric transformations utilized within multi-sensor geolocation. Section 5.2 details the similarity measures, while Section 5.3 deals with the minimization strategies. Finally, Section 5.4 introduces the possibilities of using deep learning methods for geolocation purposes.

5.1 Geometric Transformations

Image registration assumes a consistent geometric transformation between the sensed and reference images. Suppose that the sensed (or input) image $In(x, y)$ is defined over an (x, y) coordinate system, while the reference image $Ref(X, Y)$ is defined over an (X, Y) coordinate system. The goal of image registration is to find the transformation $T: (X, Y) \mapsto (x, y)$ that modifies the input image so as to be referenced in the same coordinate system as the reference image:

$$Ref(X, Y) \simeq In\{T(X, Y)\}$$

Within the CCI+ HRLC pipeline, the focus is put on global transformations, i.e., transformations operating on the entire image or on an image patch of non-negligible size. A rather general case is represented by the affine transformation. Affine transformations are identified by a vector of six parameters, i.e. translation over the x axis T_x, translation over the y axis T_y, rotation angle θ, scale factor on the x axis s_x, scale factor on the y axis s_y, and shear angle ϕ_{sh}. Particular cases of affine transformations are represented by rotation-scale-translation (RST) transformations (similarity transformations), where the shear angle is zero ($\phi_{sh} = 0$) and the scale factor is equal in the two dimensions ($s_x = s_y = s$); rigid transformations, a particular case of similarity transformation where there is no impact on the scale ($s = 1$); and shift transformations, characterized by a simple translation of the image ($\theta = 0$) [31].

In more details, the transformation $T:(X,Y) \mapsto (x,y)$ can be formulated as in the following equations, for each of the aforementioned cases, starting from the simpler shift transformation and moving to the more complex affine transformation. It is worth noting that, with the following convention, all the rotations are intended to be counter-clockwise.

- Shift transformations

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -T_x \\ 0 & 1 & -T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

- Rigid transformations

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -T_x \\ 0 & 1 & -T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

- Similarity transformations (RST)

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -T_x \\ 0 & 1 & -T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

- Affine transformation

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -T_x \\ 0 & 1 & -T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin(\theta+\phi_{sh}) & 0 \\ \sin\theta & \cos(\theta+\phi_{sh}) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Together with a given geometric transformation, to complete the mapping between the reference and the input image $Ref(X,Y) \simeq In\{T(X,Y)\}$, it is also necessary to define a resampling strategy [31]. In the case of the CCI+ HRLC processing chain, the chosen resampling strategy is the nearest neighbour (NN) interpolation. More in detail, considering again the transformation $T:(X,Y) \mapsto (x,y)$, the value of the output pixel (X,Y) is chosen equal to that of the input pixel (x',y') whose location is closest to the reverse sampled position (x,y) (whose components are generally non-integer). The advantage of nearest neighbour resampling is that the output image only contains intensity values present in the original image.

5.2 Similarity Measures

Image registration is aimed at aligning two images, the input and the reference. The reference image is fixed, and the input image is transformed to match the reference image. The matching strategies may be feature-based (e.g., speeded-up robust features (SURF) [32], Harris corner detection [33], maximally stable extremal regions (MSER) [34], etc.), area-based (cross-correlation, information theoretic measures [35], etc.), or hybrid. Within the CCI+ HRLC pipeline we focus on area-based methods and in particular on mutual information [36], [37], [38] and possibly cross correlation. Additional details on such strategy are reported in Section 5.2.3.

5.2.1 Area-based Methods

Area-based strategies [39], [40] rely on similarity and information-theoretic measures. In general, area-based methods are computationally heavier than the feature-based strategies because of the necessity to compute the similarity measure taking into consideration the whole image or generally large image regions. Nevertheless, the accuracy achievable by such techniques is generally higher than that achieved by feature-based methods [39].

As anticipated above, within the HRLC pipeline, two similarity measures are taken into consideration. On one hand, mutual information, an information-theoretic measure based on comparing local intensity distributions rather than individual pixel values, is particularly suited for multi-sensor geolocation where the images to be registered have different statistics and acquisition geometries. The main drawback is that, even though it is more robust and less sensitive to noise than the correlation-based measures, statistical distributions are heavier to be estimated on large-scale imagery, which may contribute to the computation time of the overall HRLC pipeline in

a non-negligible way. On the other hand, similarity measures like the cross correlation are faster to compute but less suited for multi-sensor data, as they are based on the pixel-wise comparisons of intensity values.

5.2.2   Feature-based Methods

Feature-based methods are generally faster but less accurate than area-based methods, and the accuracy of the registration result depends on the accuracy of the feature extraction method that is being used. There exist different strategies for the extraction of informative features. In particular, feature-point registration algorithms [39] extract a set of distinctive and highly informative individual points from both images, and then find the geometric transformation that matches them. Feature points are named in different ways, including control points, tie-points, and landmarks.

Well-known approaches in this area are those based on scale-invariant feature transforms (SIFT) [41], speeded-up robust features (SURF) [32], maximally stable extremal regions (MSER) [34], and Harris point detectors [33]. Other features of interest may be curvilinear and could be extracted by using edge detection algorithms [39], generalized Hough transforms [42], or stochastic geometry (e.g., marked point processes) methods [43].

5.2.3   The CCI+ HRLC strategy

Within the CCI+ HRLC processing chain, where the reference and the input images are the optical and the SAR images, respectively, the choice is the use of area-based methods based on the estimation of the mutual information between the two images.

Another common possibility is the use of cross-correlation as the similarity measure; however, such option is particularly critical in the multi-sensor case of the CCI+ HRLC chain. The computation of the cross-correlation, especially using the strategy based on the fast Fourier transform (FFT) [44], is usually faster and hence more convenient in an iterative process like image registration. However, the sharply different statistics of the optical and SAR images, together with the different acquisition geometries, prevent the use of cross-correlation within the CCI+ HRLC pipeline. Nevertheless, the use of cross-correlation and the fast computation through FFT will be dealt with in Section 5.4, where the focus will be put on the possibility of using generative adversarial networks (GANs) to perform domain adaptation as a pre-processing step of registration.

With respect to mutual information $\text{MI}(Ref, In)$ between the reference and the input images, let $In(\cdot)$ and $Ref(\cdot)$ indicate the input and reference images (which are both assumed composed of $M \times N$ pixels), respectively. Let also $p_{Ref,In}$ be their joint distribution, and p_{Ref} and p_{In} be their marginal distributions. The mutual information is thus computed according to:

$$\text{MI}(Ref, In) = \sum_r \sum_i p_{Ref,In}(r,i) \log \frac{p_{Ref,In}(r,i)}{p_{Ref}(r)\, p_{In}(i)}$$

There are different methods to compute such quantity. Within the CCI+ HRLC the mutual information is estimated by approximating the probability distributions through the normalized histograms. Another option, which is computationally heavier, is to estimate such distributions using kernel-based methods like Parzen window density estimation [45]. Due to the large scale of the project and the iterative optimization process, using heavy estimators is not recommended because of the registration process requiring multiple sequential estimations.

5.3   Optimization Strategies

As anticipated in the introduction to this chapter, the registration task is viewed as the combination of the following sub-processes [46]:

1. Selecting a transformation model and a resampling strategy.
2. Selecting a similarity metric to decide if a transformed input image closely matches the reference image.
3. Selecting a search strategy, which is used to match the images based on maximizing or minimizing the similarity metric.

	Ref	CCI_HRLC_Ph1-ATBD	
esa	Issue	Date	Page
	4.rev.0	31/10/2022	29

high resolution
land cover
cci

We already discussed points 1 and 2 in the Sections above; here the focus is put on the optimization strategy that has been chosen for the CCI+ HRLC multi-sensor geolocation step. The optimization strategies that are considered in the pipeline are the unconstrained Powell's algorithm and constrained optimization by linear approximation (COBYLA) method. On one hand, the unconstrained Powell's algorithm uses Powell's formulation of an approximate conjugate direction method. The objective function does not need to be differentiable, and no derivatives are required (differently from the standard conjugate gradient algorithm). The method minimizes the function using a line search along a set of search vectors [47]. Moreover, the line search is done by the Golden-section and Brent's methods [48].

On the other hand, COBYLA addresses constrained optimization by a linear approximation. It works by iteratively approximating the actual constrained optimization problem with linear programming problems. At each iteration, the resulting linear programming problem is solved to obtain a candidate for the optimal solution. The candidate solution is evaluated using the original objective and constraint functions, yielding a new data point in the optimization space. This information is used to improve the approximating linear programming problem used for the next iteration of the algorithm. When no improvement is possible, the step size is reduced, refining the search. When the step size becomes sufficiently small, the algorithm stops [49]. COBYLA allows the user to choose a starting search radius. The tuning of such parameter allows the registration process to explore regions of the search space that are far away from the initial point.

Indeed, the standard Powell's algorithm is unconstrained and, therefore, may perform its search on a large portion of the parameter space. On one hand, this is a benefit because it allows the method to perform a broad exploration of the parameter space in search of an appropriate transformation to match the input optical and SAR images. On the other hand, it is also one of the biggest downsides of the method from the viewpoint of reaching convergence in a limited timeframe. Indeed, the input optical and SAR images are natively cropped on the same pixel grid based on their georeferencing information. Given this model assumption, it is not hard to determine bounded intervals on the transformation parameters. Limiting the search to the multidimensional box determined by these ranges is not restrictive from the viewpoint of registration accuracy and makes for a significant reduction in computational burden. A modification to the standard Powell's minimization method integrates a set of barrier functions to cope with the unconstrained nature of the original optimization algorithm and combine it with the aforementioned box constraint. This way it is possible to restrict the search space to a subspace of feasible transformations (i.e., based on the size and spatial resolution of the input images). This constrained formulation of Powell's algorithm with barrier functions is integrated in the HRLC pipeline of the first production.

5.4 Multi-sensor Geolocation using Deep Learning Architectures

Another approach that is taken into consideration within the CCI+ HRLC processing chain is the use of deep learning architectures [47] for multi-sensor geolocation. Deep learning solutions for the registration of multi-sensor data is becoming of great interest for the remote sensing community.

In the context of the CCI+ HRLC processing chain, a deep learning solution is being investigated. Such strategy uses auto encoders [47] and adversarial networks [48] with the goal of developing a domain adaptation [49] method and transform optical images into SAR-like data or vice versa. With such a domain adaptation, the application of the aforementioned area-based techniques is further favoured because the optical and SAR data are brought together in a common homogeneous domain in which they are more directly comparable [50]. The adversarial network considered here is based on the interconnection of convolutional neural networks (CNNs), which have been proven highly effective in the application to the semantic segmentation of remote sensing images for land cover mapping purposes [50].

In particular, as anticipated before, the cross-correlation similarity measure, together with the fast computation through FFT, may take the place of the heavier-to-estimate mutual information. Therefore, in the following section, the details of such computation through the fast Fourier transform are presented.

5.4.1 Cross-correlation via Fast Fourier Transform

Let again $In(\cdot)$ and $Ref(\cdot)$ indicate the input and reference images (which are both assumed composed of $M \times N$ pixels). Their cross correlation can be computed according to:

$$CC(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} In(m, n) \, Ref(m - x, n - y)$$

There exists a formulation of such quantity computed using the fast Fourier transform [44]. Such process takes advantage of the relation between the convolution operation in the spatial or time domain and the product operation in the frequency domain. Let $\mathcal{F}(\cdot)$ denote the Fourier transform operator and let f and g be two signals defined in the spatial or time domain. It is straightforward to write the cross-correlation in terms of a convolution operator, which allows taking benefit of the computational efficiency of the FFT [44] and derive the cross-correlation by combining transformation, products, and inverse transformations (up to introducing the appropriate zero padding):

$$\mathcal{F}(f * g) = \mathcal{F}(f) \cdot \mathcal{F}(g) \quad \rightarrow \quad f * g = \mathcal{F}^{-1}\big(\mathcal{F}(f) \cdot \mathcal{F}(g)\big)$$

In more details, to compute the cross-correlation between two images it is necessary to: (i) compute the FFT of each image (up to zero padding) to pass from the spatial domain to the frequency domain; (ii) compute the complex conjugate of one of the two resulting signals in the frequency domain because of the mirroring operation performed during convolution and not during correlation; (iii) multiply the images in the frequency domain; and (iv) compute the inverse FFT transform of the product to obtain the cross-correlation of the two images in the spatial domain. The flowchart of such computation is provided in Figure 13.
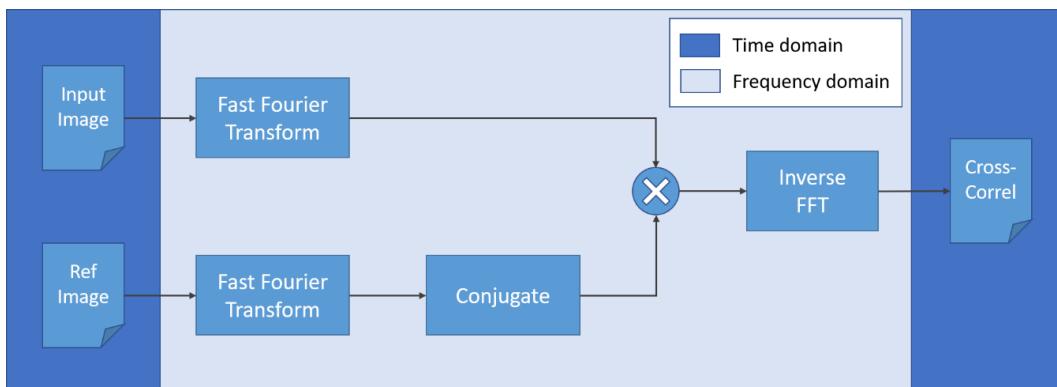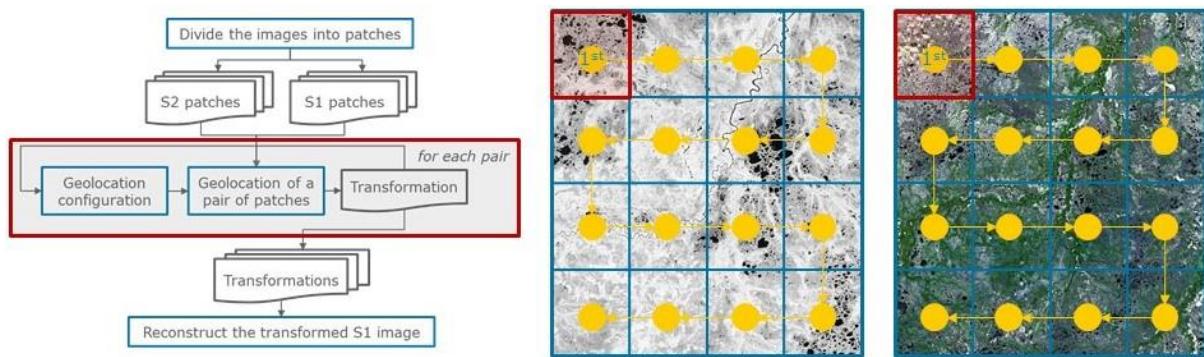
Figure 13: Computation of the cross-correlation via FFT.

5.5 Automatic tiling for large-scale registration

The HRLC pipeline operates geometrically on the pixel grid of the Sentinel-2 granules. Accordingly, the multisensor geolocation module also works with respect to this lattice. Given the size of the Sentinel-2 granules, taken as reference images in the multisensor geolocation module, a single global transformation may not be sufficient to carefully address local distortions. In fact, the images acquired by the two satellites may present distortions which significantly differ locally. This is especially true in the cases in which a single S2 granule lays over a border among different S1 images. Furthermore, the application of the aforementioned area-based approach to the entire granule area may be computationally heavy.

In order to address these problems, a dedicated automatic tiling algorithm has been developed in the project: both the reference (output of the optical pre-processing chain) and input (output of the SAR pre-processing chain) images are divided into patches. Each of this patches is registered separately using a global transformation and the final image is reconstructed starting from the patches. On one hand, considering one patch at a time reduces the computational complexity of the registration. Furthermore, the resulting overall transformation is non-global across the whole granule and allows to better account for misaligned details. On the other hand, the

developed algorithm also takes into account the need for favouring consistency across the patches and to prevent border artefacts. While for the first patch the algorithm is initialized with the identity transformation T_0 consisting of: (i) no translation along neither one of the two axis; (ii) no rotation; and (iii) unitary scaling (i.e. $T_0 = [T_x = 0, T_y = 0, \theta = 0, s = 1]$), the result $\hat{T}_1$ of the geolocation for this first patch is passed to the next one. That way, the next patch is initialized with $T_1 = \hat{T}_1$. The same is done for each subsequent patch and transformation value is propagated so as to initialize the next iteration of the algorithm.

This allows for faster convergence, giving the algorithm a reliable initialization for each patch after the first one. Moreover, it generally leads to more precise results, considering that local distortions among adjacent patches can be assumed to be minor and the initialization allows the minimization algorithm to start from a good solution and quickly exploring the closer ones in the search space [51]. Dedicated procedures are also integrated in the algorithm to detect anomalies in this iterative process and prevent them for affecting the registration output.

6 Optical imagery classification

From analysis of the recent literature related to large-scale land cover mapping problems the following crucial aspects must be considered in order to achieve efficient and robust classification of optical high-resolution images [52]:

- automation for efficiency and timeliness;
- spatial continuity of the maps;
- temporal coherence between updates of the product;
- reproducibility of the results;
- support of changes of nomenclature without changing the system.

The CCI HRLC project addresses each of these points. To maximize the outcome, the following strategies must be implemented and assessed in an operational context of land cover map production at the large-scale:

- all available images having cloud cover lower than 40 % acquired during the reference period are used;
- the procedure is fully automatic without a need for manual operations;
- the processing chain is implemented using a massively parallel work-flow which achieves a reduced computation time allowing timely map production and data reprocessing for ensuring continuity across reference years in the case of updating the product specification.

Project activity for optical imagery classification is oriented in two directions: on the one side, internal benchmarking activities were entirely devoted to the selection of the best performing classification algorithms; on the other side, a selection of reference/ancillary data to empower and enrich the feature extraction and training processes.
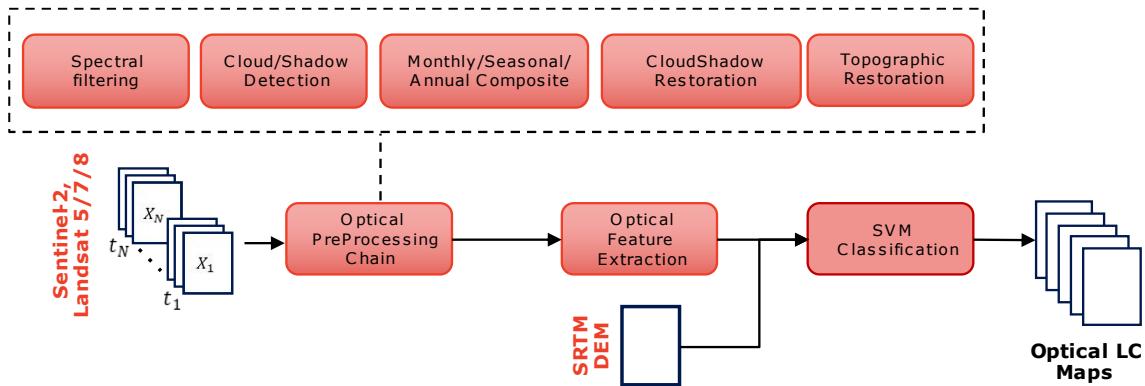
Figure 14. Optical data processing chain for the prototype production of both the static and the historical HR LC maps obtained by classifying the time series of high resolution optical data.

Figure 14 depicts the optical data processing chain for the production of both the static and the historical HR LC maps obtained by classifying the time series of Sentinel 2 and Landsat data. The images are first pre-processed in order to perform the atmospheric correction and detect the clouds. Then, the composites are generated and the possible missing data (caused by cloud and shadow coverage) are restored. Due to the missed availability of training data, a lot of effort has been devoted to the preparation of photo-interpretation activity carried out to define the training sets. In order to generate a representative and informative training set, a stratified random sampling strategy was defined to define to the prior probabilities of the land cover classes, computed according to the 2015 Copernicus Global Land Cover map. Details are given in the following.

6.1 Classification

To generate the high-resolution LC maps, the SVM classifier is trained on the time-series of optical composites generated in the pre-processing step. To accurately represent the spatial information together with the temporal one, an optical feature extraction step is performed to extract textural features from the first composite. Moreover, the altitude of the land cover is added as a feature, such that to better characterize the land cover typical of specific altitude levels. Hence, while temporal and spectral features are good in representing the seasonality of the classes, the aim of the textural and altitude features extraction is to provide to the classifier information about spatial context of the samples which can provide better classes discrimination.

In order to perform the supervised training of the classifier, a lot of effort has been devoted generating a training set by photo-interpretation. The labelled samples associated with time-series of composites are extracted from the training database. The training is performed once for each year in each considered area independently (Amazonia, Sahel and Siberia).

6.1.1 Support Vector Machine

As a classifier, the Support Vector Machine (SVM) is one of the most effective methods in pattern and texture classification to the land cover mapping [53]. Its fundamental idea is that the feature of input space is mapped into a high-dimensional feature space through nonlinear transformation. The nonlinear transformation is implemented by defining proper kernel function. SVM has two important features. Firstly, the upper bound on the generalization error does not depend on the dimension of the space. Secondly, the error bound is minimized by maximizing the margin, that is, the minimal distance between the hyperplane and the closest data points [54], [55]. SVMs are particularly appealing in remote sensing field due to their ability to successfully handle small training datasets, often producing higher classification accuracy than traditional methods, as well as to be the best algorithm when classes are separable [55]. In contrast, for larger dataset, it requires a large amount of time to process.

SVM implements a classification strategy that exploits a margin-based "geometrical" criterion rather than a purely "statistical" criterion. In other words, SVM does not require an estimation of the statistical distributions of classes to carry out the classification task. Instead, the classification model exploits the concept of margin maximization. The main properties that make SVM particularly attractive in the considered application are the following:

- their intrinsic effectiveness with respect to traditional classifiers thanks to the structural risk minimization principle, which results in high classification accuracies and very good generalization capabilities;

- the possibility to exploit the kernel trick to solve non-linear separable classification problems by projecting the data into a high dimensional feature space and separating the data with a simple linear function;

- the convexity of the objective function used in the learning of the classifier, which results in the possibility to solve the learning process according to linearly constrained quadratic programming (QP) characterized from a unique solution (i.e., the system cannot fall into sub-optimal solutions associated with local minima);

- the possibility of representing the convex optimization problem in a dual formulation, where only non-zero Lagrange multipliers are necessary for defining the separation hyperplane (which is a very important advantage in the case of large datasets). This is related to property of sparseness of the solution.

Let us assume that a training set is given $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i = (x_i^1, \ldots, x_i^j)$ is the i-th primitive feature and $Y = \{y_i\}_{i=1}^{N}$ is the corresponding set of labels. Accordingly, let us assume that $y_i \in \{+1; -1\}$ is the binary label of the sample x_i. The goal of the binary SVM is to divide the d-dimensional feature space in two subspaces, one for each class, through a separating hyperplane $H: y = \langle w \cdot x \rangle + b = 0$. The final decision rule used to find the membership of a test sample is based on the sign of the discrimination function $f(x) = \langle w \cdot x \rangle + b$ associated to the hyperplane. Therefore, a generic sample x will be labelled according to the following rule: $y = sgn\, f(x)$.

The training of an SVM consists in finding the position of the hyperplane H, estimating the values of the vector w and the scalar b, according to the solution of an optimization problem. From a geometrical point of view, w is a vector perpendicular to the hyperplane H and thus defines its orientation. The distance of the H to the origin is $b/\|\mathbf{w}\|$, while the distance of a sample x to the hyperplane is $f(\mathbf{x})/\|\mathbf{w}\|$. Let us define the functional margin $F = \min\{y_i f(\mathbf{x}_i)\}$, $i = 1, \ldots, N$ and the geometric margin $G = F/\|\mathbf{w}\|$. The geometric margin represents the minimum Euclidean distance between the available training samples and the hyperplane.

In the case of a linearly separable problems, the learning of an SVM can be performed with the maximal margin algorithm, which consists in finding the hyperplane H that maximizes the geometric margin G. However, the maximum margin-training algorithm cannot be used in case the available training samples are not linearly separable because of noisy samples and outliers. In these cases, the soft margin algorithm is used in order to handle nonlinear separable data. This is done by defining the so-called slack variables as:

$$\xi\left[(x_i, y_i), (w, b)\right] = \xi_i = \max\left[0, 1 - y_i(\langle w \cdot x_i \rangle + b)\right]$$

Slack variables allow one to control the penalty associated with misclassified samples. In this way the learning algorithm is robust to both noise and outliers present in the training set, thus resulting in high generalization capability. The optimization problem can be formulated as follows:

$$\begin{cases} \min_{\mathbf{w}, b} \left\{ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi_i \right\} \\ y_i(\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) \geq 1 - \xi_i, \ \xi_i \geq 0, \ \forall i = 1, \ldots, N \end{cases}$$

	Ref	CCI_HRLC_Ph1-ATBD	
esa	Issue	Date	Page
	4.rev.0	31/10/2022	34

high resolution
land cover
cci

where $C \geq 0$ is the regularization parameter that allows one to control the penalty associated to errors (if $C = \infty$ we come back to the maximal margin algorithm), and thus to control the trade-off between the number of allowed mislabelled training samples and the width of the margin. If the value of C is too small, many errors are permitted and the resulting discriminant function will poorly fit with the data; on the opposite, if C is too large, the classifier may overfit the data instances, thus resulting in low generalization ability. A precise definition of the value of the C parameter is crucial for the accuracy that can be obtained in the classification step and should be derived through an accurate model selection phase. Similarly to the case of the maximal margin algorithm, the optimization problem can be rewritten in an equivalent dual form:

$$\begin{cases} \max_{\alpha} : \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \right\} \\ \sum_{i=1}^{N} y_i \alpha_i = 0, \ 0 \leq \alpha_i \leq C, \ 1 \leq i \leq N \end{cases}$$

Because of the constraint introduced by the multipliers $\{\alpha_i\}_{i=1}^{N}$ that for the soft margin algorithm are bounded by the parameter C, the problem is also known as box constrained problem. The Karush–Kuhn–Tucker (KKT) complementarity conditions provide useful information about the structure of the solution. They state that the optimal solution should satisfy:

$$\begin{cases} \alpha_i \left[y_i (\langle \mathbf{w} \cdot \mathbf{x}_i \rangle + b) - 1 + \xi_i \right] = 0, & i = 1, ..., l \\ \xi_i (\alpha_i - C) = 0, & i = 1, ..., l \end{cases}$$

Varying the values of the multipliers $\{\alpha_i\}_{i=1}^{N}$ three cases can be distinguished:

$$\begin{cases} \alpha_i = 0 & \Rightarrow y_i f(\mathbf{x}_i) \geq 1 \\ 0 < \alpha_i < C_i & \Rightarrow y_i f(\mathbf{x}_i) = 1 \\ \alpha_i = C_i & \Rightarrow y_i f(\mathbf{x}_i) \leq 1 \end{cases}$$

The support vectors with multiplier $\alpha_i = C$ are called bound support vectors (BSV) and are associated to slack variables $\xi_i \geq 0$; the ones with $0 < \alpha_i < C_i$ are called non-bound support vectors (NBSV) and lie on the margin hyperplane H_1 or H_2 ($y_i f(\mathbf{x}_i) = 1$).
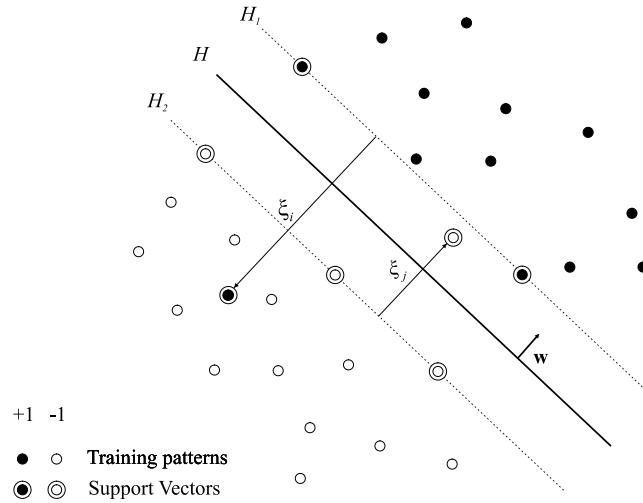
Figure 15: Qualitative example of a separating hyperplane in the case of a non-linear separable classification problem.

An important improvement to the above-described methods consists in considering nonlinear discriminant functions for separating the two information classes. This can be obtained by transforming the input data into a high dimension (Hilbert) feature space $\Phi(\mathbf{x}) \in \Re^{d'}$ ($d' > d$) where the transformed samples can be better separated by a hyperplane (Figure 16). The main problem is to explicitly choose and calculate the function $\Phi(\mathbf{x}) \in \Re^{d'}$ for each training samples. Given that the input points in dual formulation appear in the form of inner products, we can do this mapping in an implicit way by exploiting the so-called kernel trick. Kernel methods provide an elegant and effective way of dealing with this problem by replacing the inner product in the input space with a kernel function such that:

$$K(x_i, x_j) = \langle (\phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)) \rangle \qquad i, j = 1, ..., N \tag{7}$$

implicitly calculating the inner product in the transformed space. The soft margin algorithm for nonlinear function can be represented by the following optimization problem:

$$\begin{cases} \max_\alpha \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \right\} \\ \sum_{i=1}^{N} y_i \alpha_i = 0, \ 0 \le \alpha_i \le C \text{ and } 1 \le i \le N \end{cases} \tag{8}$$

And the discrimination function becomes:

$$f(\mathbf{x}) = \sum_{i \in SV} y_i \alpha_i^* k(\mathbf{x}_i \cdot \mathbf{x}) + b \tag{9}$$

The condition for a function to be a valid kernel is given by the Mercer's theorem. The most widely used non-linear kernel functions are the following:

- homogeneous polynomial function: $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^p, \quad p \in Z$

- inhomogeneous polynomial function: $k(\mathbf{x}_i, \mathbf{x}_j) = (c + (\mathbf{x}_i \cdot \mathbf{x}_j))^p, \quad p \in Z, c \ge 0$

- Gaussian function: $k(\mathbf{x}_i, \mathbf{x}_j) = e^{\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}}, \qquad \sigma \in \Re$

	Ref	CCI_HRLC_Ph1-ATBD		high resolution
esa	Issue	Date	Page	**land cover** cci
	4.rev.0	31/10/2022	36	
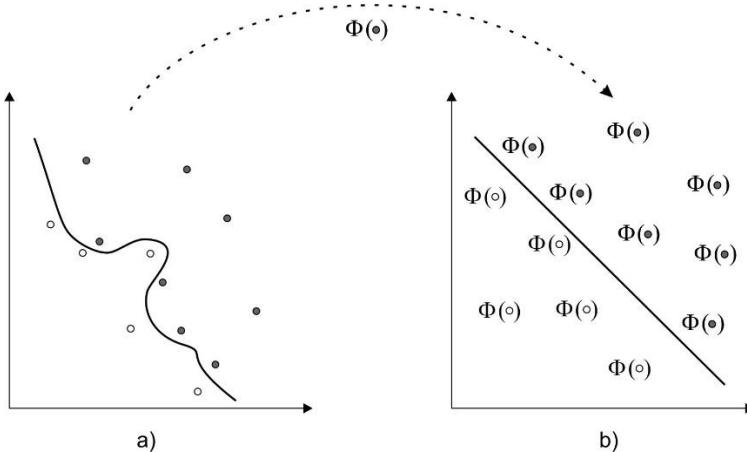
Figure 16: Transformation of the input samples by means of a kernel function into a high dimension feature space: a) Input feature space; b) kernel induced high dimensional feature space.

From an operational perspective, a possible implementation would use the Gaussian kernel since linear and polynomial kernels are less time consuming but provide in general less accuracy. The Sigma σ parameter is a positive parameter whose behavior regulates the fitting property: if its value increases the model gets overfits, while decreasing the model underfits. In our implementation, the default value for gamma is initially set equals to 1 over the number of features [56], optimal choice can be made in proper training stage.

6.1.2   Textural features extraction

Texture allows the accurate characterization of the contextual information of a pixel in the image. In the literature, it can be found that the use of textural information can significantly improve the classification results. Hence, such features can be more distinctive than spectral features for some land cover classes. Instead of considering complex spatial features, such as shape and size, which required the unsupervised segmentation of the image, we considered other textural feature extractors. First, the Gray-Level Co-Occurrence Matrix (GLCM) is computed. Then the GLCM is used to extract the following statistical measures, used as features:

- Dissimilarity;
- Correlation;
- Contrast;
- Homogeneity;
- Energy;
- Angular Second Moment (ASM).

6.1.3   Photo-interpreted training sets generation

Operational land cover map production over large areas cannot rely on field campaigns because huge amounts of costly data have to be collected, most importantly jeopardising the timeliness of the land cover map. In order to generate the training set used to perform the supervised classification of the considered study areas, a lot of effort has been devoted to photo-interpretation activities. Hence, even though existing thematic products represent a valid source of information, ground reference data are needed to model complex classes which require reliable samples that cannot be extracted from the outdated coarse thematic products. Although extremely complex and time consuming the reference data allows the production of high quality training set which matches the definition of the legend and corresponds to the exact same time frame (see Figure 17 and Figure 18).

To properly generate the training set, which is representative of the considered area, the team first estimated the prior probabilities of the classes by considering the information provided by the 2015 Copernicus Global Land Cover (CGLC) map. Then, the samples to be labeled, where selected according to the stratified random sampling strategy. The label of each sample was defined by photo-interpretation of both Sentinel 2 data and SPOT images in the RR areas. For areas, where POT images were not available, we exploited the public very high-resolution Google and ESRI images (i.e., 50 cm). The labels assigned to samples are those of the HRLC classes (see Table 1) which have been defined by applying the schematic approach from LCCS system (see Figure 19) and adapting to what is actually observable with remote sensing data. In particular, the data were pixel-wise labeled, thus we avoided the strong positive correlation between samples units, which is the case for polygon-wise labeling.

Figure 17. Training Set Production conducted via photo-interpretation.
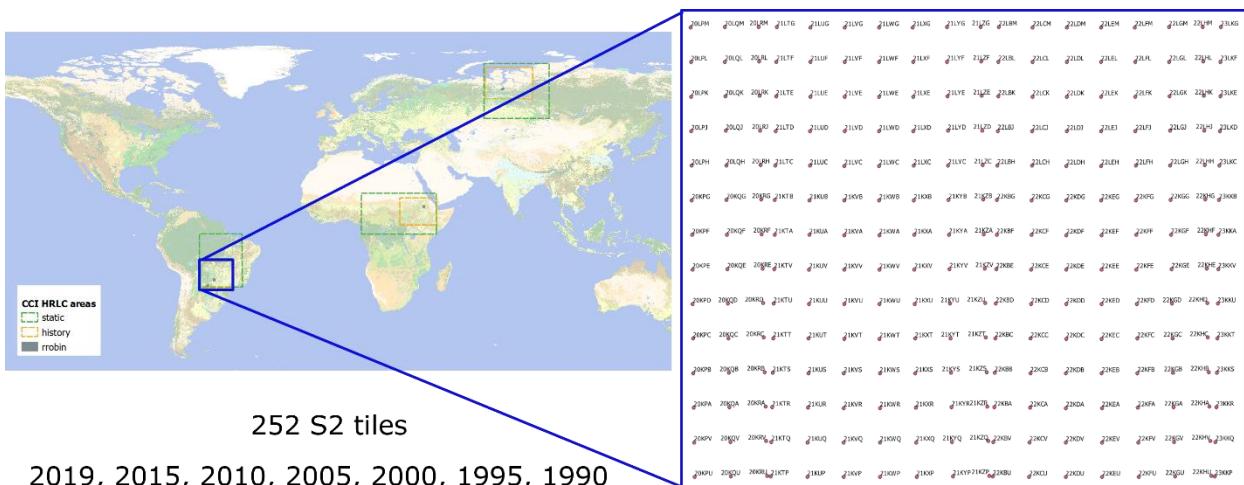
Figure 18. Example of number of tiles to be covered by photointerpretation in Amazonia.

Although the photo-interpretation represents a valid solution for generating the training set, the legend scheme reported in Figure 19 presents some discrepancy with a set of classes which can be discriminated by the considered remote sensing data. Figure 20 reports an example of ambiguity between grassland (permanent herbaceous cover) and cropland (a herbaceous cover that is harvested). In the reported example it is not clear

which is the correct class considering the set of rules defined in the classification scheme. Hence, if bare soil is present during the year, it should be classified as cropland. However, by taking the context of cattle raising in this region, you might guess this is grassland. Another problem is reported in Figure 21, where the pixels belonging to the road are classified as bare soil since they are not asphalted. Finally, differently from the medium resolution no mixed classes are present in the legend (e.g., Mosaic herbaceous cover (>50%) / and shrub (<50%)). Although we are working at 10 m spatial resolution, the detection of shrubs in the Sentinel 2 images is challenging (see Figure 22). The identification of deciduous and evergreen shrubland is even more challenging.
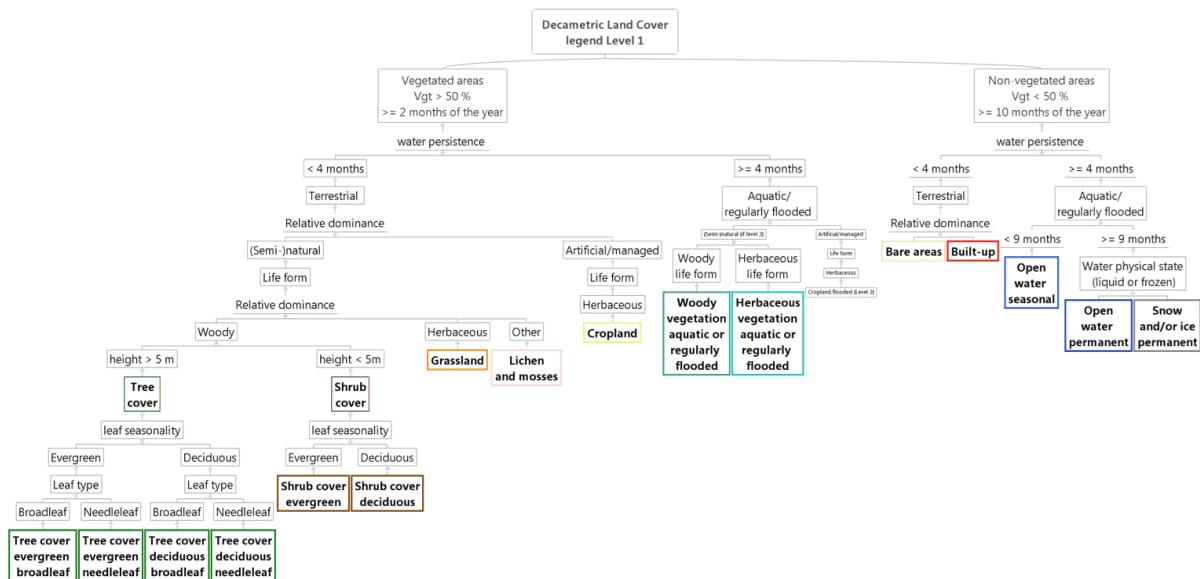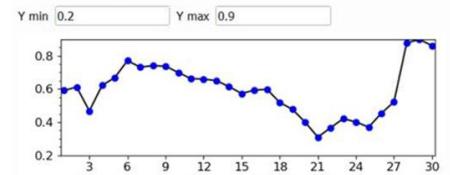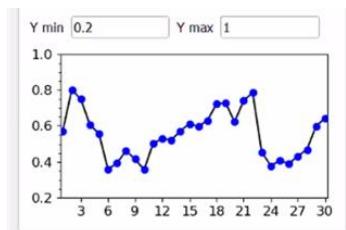
Figure 19. The classification scheme of the training-set production.

NDVI trend

NDVI trend

Figure 20. Example of ambiguity between the classes: (i) grassland (permanent herbaceous cover), and (ii) cropland (a herbaceous cover that is harvested). If bare soil is present during the year, it should be classified as cropland. By taking the context of cattle raising in this region, you might guess this is grassland.

Figure 21. In Amazonia, the roads far away the city are not asphalted, and they are very similar to bare areas (ambiguous definition from the RS view point).
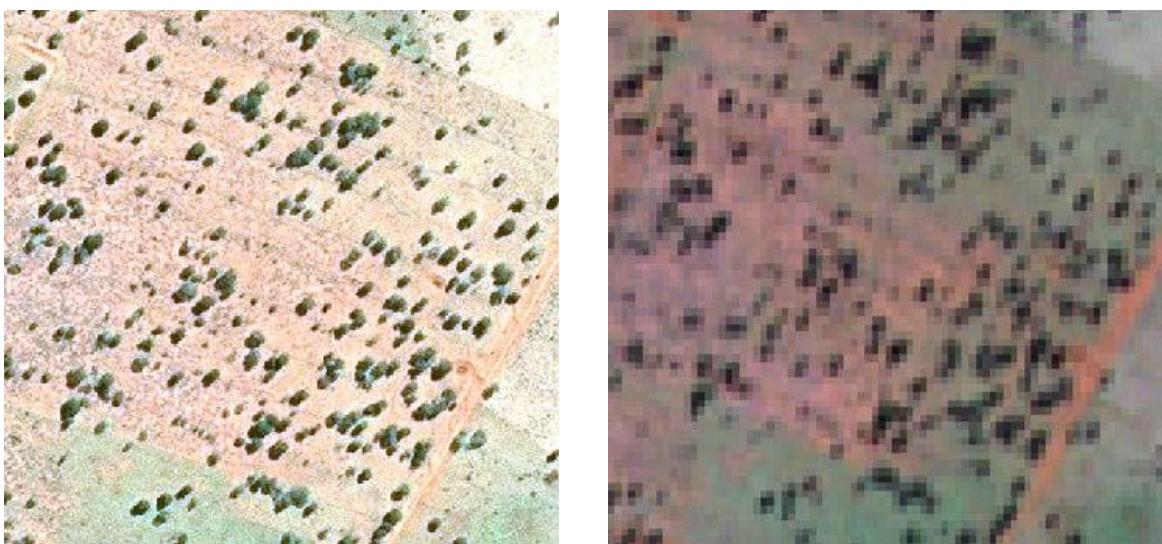
Figure 22. Differently from the medium resolution no mixed classes are present in the legend (e.g., Mosaic herbaceous cover (>50%) / and shrub (<50%). Although we are working at 10 m spatial resolution, the detection of shrubs in the Sentinel 2 images is challenging. The identification of deciduous and evergreen shrubland is even more challenging.

In the case of the historical training set photo-interpretation activity, and at the same time changing the resolution of the available images from 10 to 30 meters, the team has identified following challenges:

- Less high resolution images are available;
- Landsat 7 images are corrupted;
- spatial resolution of 30m hampers extraction of training points as the spectral information is often mixed. Moreover, the NDVI and NDWI trends (crucial to differentiate some very similar classes e.g. grassland vs cropland) are unclear and difficult to interpret.

Taking into account all of the above mentioned points, the team has decided to update the training set extracted in 2019. This means to confirm the label assigned to a sample in 2019 or otherwise to eliminate the sample from the training set. Thus, the training set produced in the past have smaller number of samples compared to the

one used to classify the static map. Figure 23 shows a qualitative example of the data used to perform the photointerpretation in the past.
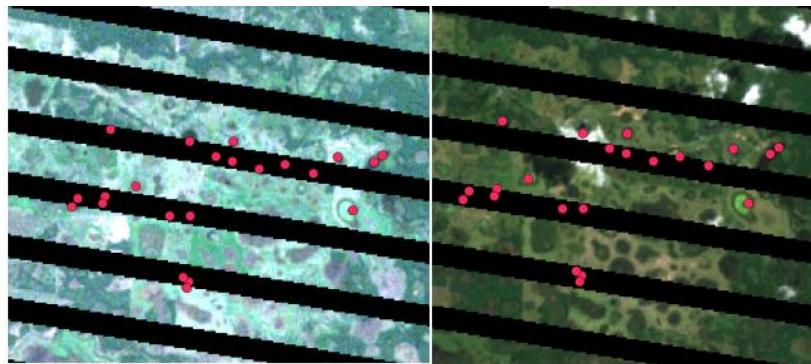
Figure 23. Many difficulties going back in the past for the photo-interpretation process: (i) less images are available; (ii) Landsat 7 Corrupted; (iii) NDVI and NDWI trend not clear; (iv) the spatial resolution of 30m.

6.1.4   Final static training sets generation

While complex classes require reliable samples that cannot be extracted from the outdated coarse thematic products, existing thematic products represent a valid source of information for the other classes, allowing to significantly expand the training set and properly represent the whole areas to map. For this reason, only for the static map production, we integrated the training sets delivered through photointerpretation with samples extracted from the agreement of land cover products available. Oversampling of the complex classes was performed to keep the training set prior distribution of the land-cover classes constant. Moreover, the increased amount of training labels unlocked the possibility of exploiting the specific properties of the local land cover. This can be done by considering the global ecoregions, which are areas of water or land that contain characteristic assemblages of natural communities and species. By training a classifier for each ecoregion, we can exploit the fact that inside an ecoregion the probability of encountering different vegetation species (which may be mapped in the same class) and communities remains relatively constant. This feature is important in land-cover mapping with remote sensing images, as it allows to mitigate the intra-class variance, a well-known issue in remote sensing.

Therefore, we combined the photo-interpreted training sets with samples extracted from the agreement of land cover products available, and then divided each area in smaller areas defined by taking into account the ecoregions. This was done at tile level and by aggregating ecoregions to avoid excessive fragmentation of training set. Figure 24 shows as examples the photo-interpreted training set and the final training set of Amazon static, respectively. Figure 25 shows the final division into ecoregions of the three mapped areas. Note that the ecoregion training sets are slightly larger and overlapping with each other to guarantee consistent predictions of the land cover on the ecoregion borders.

(a)   (b)
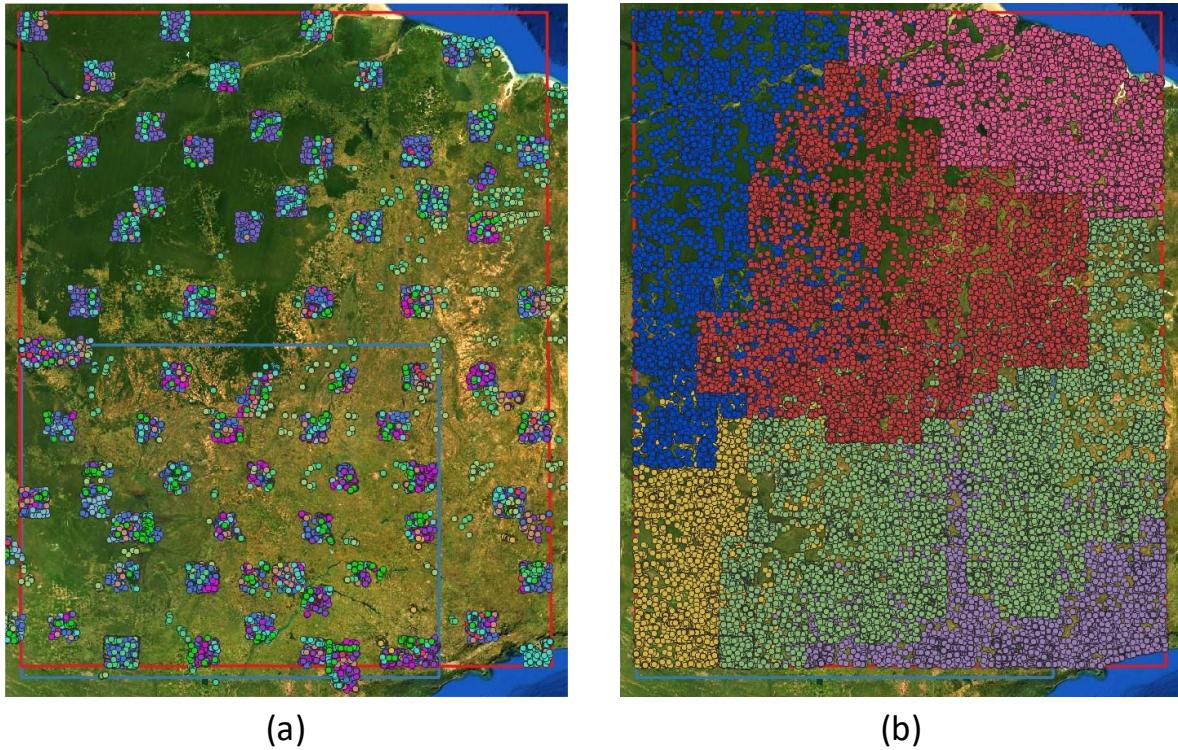
Figure 24. Amazon static: (a) photo-interpreted training set, (b) final training sets divided by ecoregions.

Siberia	
Ecoregion	**# tiles**
10553	61
10576	19
10586_1	67
10586_2	66
10595	31
10596	36
10645	71
10648	44

Africa	
Ecoregion	**# tiles**
10108	85
10110	59
10122	213
10124	33
10700	232
10132	129
10176	29

Amazon	
Ecoregion	**# tiles**
10384_10349	118
10352_10381_10353	50
10380_10354	117
10394_10438	133
10400_10439	63
10312_10316	87
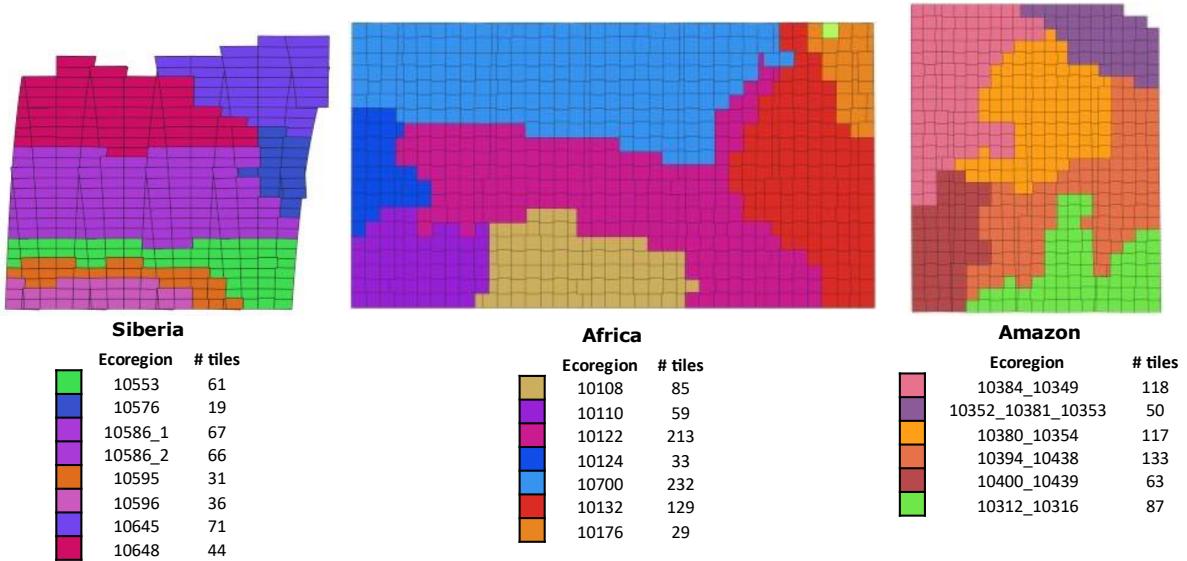
Figure 25. Final division into ecoregion for the three mapped areas.

7   SAR imagery classification

7.1   Feature extraction

To carry out the land cover classification using Sentinel-1 dual-Pol data sets based on the defined classes, reported in Table 1, the feature extraction will be based on the polarimetric information of data [57], [58].

To improve the ability of classifier to recognize and discriminate the different environment textures and morphological structures (e.g. urban areas, agricultural crops, forests, etc.), the amplitude of VH and VV channels and their combinations have been assumed.

	Ref	CCI_HRLC_Ph1-ATBD	high resolution
eSa	Issue	Date	Page	land cover cci
	4.rev.0	31/10/2022	42	

Although the S1 data are not fully polarized, we can exploit the polarimetric information arising from the intensities of the VH and VV channels by means analysis on single channel (by choosing VH or VH) or on their combination (their mean or ratio, for instance). These features contain essential polarimetric information provided by the dual-Pol data since the polarimetry combination distinguishes specular scattering from diffuse scattering.

7.1.1   Texture analysis on single polarization

To analyze and exploring the spatial information contained in a single S1 image (VH or VV), a docker application has been developed in order to provide a set of filters that operate especially in spatial domain. The rationale for selecting these algorithms is the velocity of the execution. Although they might not be the most accurate ones, the possibility to apply them quickly to the SAR images in a large stack in a reasonable amount of time is an invaluable asset for wide area processing. The implemented techniques are summarized in the following list:

- *Mean filter* is one of the most widely used low-pass filters (LPF). It substitutes the pixel value with the average of all the values in the local neighborhood (filter kernel).

- *Median filter*, a non-adaptive filter and replaces each pixel value with the median of the pixel values in the local neighborhood.

- *Minimum (maximum) filter* is a non-linear filter that is located the darkest (brightest) point in an image. It is based on median filter since it is defined as his 0th (100th) percentile, i.e. by considering the minimum (maximum) of all pixels within a local region of an image.

- *Max-Min filter*, blurs the image by replacing each pixel with the difference of the highest pixel and the lowest pixel (with respect to intensity) within the specified window-size.

7.1.1.1   Mean filter

The Mean filter is a low-pass filter (LPF) and represents the simplest and easiest method of smoothing images, in addition to being very easy to implement. Mean filtering is usually thought of as a convolution filter. Like other convolutions it is based around a kernel, which represents the shape and size of the neighborhood to be sampled when calculating the mean. The idea of mean filtering is simply to replace each pixel value in an image with the mean ("average") of values belonging to neighborhood, including itself. Then, the filter window will be moved pixel-by-pixel until to scanner the whole image.

It does not remove the speckle from the image but averages it into one. In fact, the noise becomes less apparent, but the image looks "softened". Theoretically, dark and bright speckle pixels within the filter window can cancel each other out. The probability of such situations increases with the size of the filter window, 7×7 or 9x9 for example. However, it produces image blur, loss of details and, eventually, loss of spatial resolution, giving an image with less noise but less high frequency detail. For this reasons, 3x3 or 5x5 size filter are recommended. Note that the mean filtering is not suitable in case of pulse and spike noise since the shot noise pixel values are often very different from the surrounding values, they tend to significantly distort the pixel average calculated by the mean filter. The median filter is successful at removing pulse and spike noise while retaining step and ramp functions [59].

7.1.1.2   Median filter

The median filter is normally used to reduce noise in an image, somewhat like the mean filter. However, it often does a better job than the mean filter of preserving useful detail in the image. Like the mean filter, the median filter considers each pixel in the image in turn and looks at its nearby neighbors to decide whether it is representative of its surroundings or not. Instead of simply replacing the pixel value with the mean of neighboring pixel values, it replaces it with the median of those values. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into numerical order and then replacing the pixel being considered with the middle pixel value.

By calculating the median value of a neighborhood rather than the mean filter, the median filter has two main advantages over the mean filter:

- The median is a more robust average than the mean and so a single very unrepresentative pixel in a neighborhood will not affect the median value significantly.

- Since the median value must be the value of one of the pixels in the neighborhood, the median filter does not create new unrealistic pixel values when the filter straddles an edge. For this reason, the median filter is much better at preserving sharp edges than the mean filter.

Hence, the median filter is edge preserving [60] although it may lead to the removal (or suppression) of small (also linear) objects from the image, exactly in the same way as it removes (or suppresses) speckle noise.

Applying a 3×3 median filter produces a noise reduction at the expense of a slight degradation in image quality. If we smooth the noisy image with a larger median filter, e.g. 7×7, all the noisy pixels disappear, but the image looks a bit "blotchy".

A good solution is to use 3×3 or 5x5 median filter [61] and passing it over the image more times in order to remove all the noise with less loss of detail, alternatively.

The mean and median filters meet with only limited success when applied to SAR data. One reason for this is the multiplicative nature of speckle noise, which relates the amount of noise to the signal intensity. The other reason is that they are not adaptive filters in the sense that they do not account for the speckle properties of the image. Adaptive filters, such as the Lee filter, assume that the mean and variance of the pixel of interest are equal to the local mean and variance of all pixels within the user-selected moving window.

7.1.1.3 Minimum and maximum filters

Minimum and maximum filters, also known as erosion and dilation filters, respectively, are morphological filters that work by considering a neighborhood (running window) around each pixel. The running window is an image area around a current pixel with a defined radius. For example, if we specify the radius = 1, the running window will be a 3x3 square around the target pixel, which is the smallest box size. The maximum and minimum filters are shift-invariant. Whereas the minimum filter replaces the central pixel with the darkest one in the running window, the maximum filter replaces it with the lightest one. In other words, the minimum filter extends object boundaries, whereas the maximum filter erodes shapes on the image. The odd size of the neighborhood considered for each pixel. Also in this case, the recommended size are 5x5 or 7x7 in order not to incur in issues have been addressed previously, see Section 7.1.1.1.

The docker offers also the possibility for each user to choose the kernel filter size adapted to its needs, but the default dimension is 9x9 because the implemented filter has shown satisfactory results both in terms of computational complexity and the quality of output image, due its ability in details preservation, edges definition.

7.1.1.4 Max-Min filter

The output image is given by the difference between dilation and erosion filters (described in previous section 7.1.1.3).

Hiring X as input image, the max-min filtered image is given taking into the account to the following simple expression:

$$Y = X_{max} - X_{min}$$

where Y is the resulting gray level image, whereas X_{max} and X_{min} are the maximum and minimum filtered version of input image X, respectively. The Max-Min filter blurs the image by replacing each pixel with the difference of the highest pixel and the lowest pixel (with respect to intensity) within a specified window-size (for example, the grayscale 3x3 or 5x5 pixel neighborhoods). To preserve much more spatial details and texture structures, we have set up window size to 9 by default, also according to the evaluations explained above.

7.1.2 Texture analysis on dual- polarization

SAR polarimetry allows for the retrieve of shape, orientation, and dielectric property information of scatterers [62],[63]. Since there are multiple polarimetric channels, it provides more information than single-pol SAR data. However, the richness of polarimetry is achieved by sacrificing the spatial resolution. To balance the trade-off, instead of a fully polarized SAR, Sentinel-1 mission provides partially polarized SAR data, known as dual-Pol data, with the VV and VH channels. To extract the polarimetric information of Sentinel-1 data, we used the signal acquired from VH and VV channels, and several composite images given by:

- *Ratio*, VV/VH
- *Sum*, $VH + VV$
- *Mean*, $(VH + VV)/2$
- *Difference*, $VV - VH$

These four features contain essential polarimetric information provided by the dual-Pol data. This polarimetry combination is able to distinguish specular scattering from diffuse scattering [64]. For the purpose of classification, these features are highly beneficial to differ classes with different surface roughnesses, such as water, plant, building, and soil. The aim is basically exploit the dual-polarization capability of S1 for providing as many ground surface information as possible [65].

7.1.3 Texture analysis by statistics

To increase the feature space it is also possible to add texture features by applying the *Grey Level Co-occurrence* matrix (GLCM), in order to retrieve second order textures [66]–[68].

This operation is done before applying the speckle filtering, since the despeckling destroys most of the image texture. For example, the classification accuracy related to perennial agroforestry land cover can be improved by using less correlated GLCM texture measures: Contrast, Entropy, Correlation, and Variance. The GLCM texture can be measured using a 5×5 moving window, one-pixel displacement, for example. In [69] it is shown that the GLCM texture measures are appropriated to discriminate vegetation types, and less sensitive to no vegetation cover. Is shown that the more informative variables are the VH Variance and Correlation of SAR images acquired in a dry season and, and VV Contrast of images in a wet season.

Instead, [70] highlights the importance of VH image that is the best band for differentiating agricultural land from other land cover types. The major differences in vegetation, their vertical structure, are captured in co-polarized (VV) band.

Another way for land-cover classification is to use multi-temporal SAR data (i.e. SAR data time series) analysis and extract features by considering the temporal variation of backscattering coefficients and information from interferometric data processing. The work [71] exploits the combination of the average backscattering coefficient and temporal variability. The average backscattering coefficient permits to classify water and urban areas, since they present very low and high signatures, respectively. The temporal variability, which is a main feature in multitemporal analysis, can be used to distinguish cultivated areas and water from the forest and urban classes.

The behavior of VH and VV backscatter signal is different over vegetated areas. Over vegetation land covers, there is much volume scattering of the radar signal. And volume scattering tends to cause a depolarization of the return signals, which then corresponds to a high backscatter in cross-polarization (VH or HV) bands. Thus, VH bands show a higher sensitivity to vegetation.

For the purposes of classification, these features are highly beneficial to diversify classes with different surface roughness, such as water, plants, buildings, and soil. In this manner, the classification maps may achieve high classification accuracy values. Specifically, the feature extraction step is preliminary to the classification step in the sense that only specific features for peculiar classes may be extracted and used each time. In addition, according to the technical literature, we also identified several works describing most performing classification methods able to classify different classes (water, urban areas, snow, for example) with a proper combination of features set. A preliminary list is reported in the following Table 5.

Table 5. Preliminary list of SAR features for subsets of classes.

Class	Feature(s)	Reference
Urban	Occurrence range, DEM slope	[G. Lisini, A. Salentinig, P. Du, P. Gamba, "SAR-based urban extents extraction: from ENVISAT to Sentinel-1", IEEE J. of Selected Topics in Applied Earth Observation and Remote Sensing, doi: 10.1109/JSTARS.2017.2782180, vol. 11, no. 8, pp. 2683-2691, Aug. 2018.]
Water	Average backscatter, the minimum backscatter of a time series and standard deviation of the backscatter	[Santoro, Maurizio, and Urs Wegmüller. "Multi-temporal SAR metrics applied to map water bodies." 2012 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2012.
Snow	σ^0VV band; backscattering ratio	[Tsai, Ya-Lun S., et al. "Wet and Dry Snow Detection Using Sentinel-1 SAR Data for Mountainous Areas with a Machine Learning Technique." Remote Sensing 11.8 (2019): 895.]
Crop	Occurrence variance; co-occurrence contrast	[Fontanelli, Giacomo, et al. "Agricultural crop mapping using optical and SAR multi-temporal seasonal data: A case study in Lombardy region, Italy." 2014 IEEE Geoscience and Remote Sensing Symposium. IEEE, 2014.]
Deciduous vegetation	Temporal signature	[Rüetschi, Marius, Michael Schaepman, and David Small. "Using multitemporal Sentinel-1 C-band backscatter to monitor phenology and classify deciduous and coniferous forests in northern Switzerland." *Remote Sensing* 10.1 (2017): 55.]
Evergreen vegetation	VV and VH channels	[Abdikan, Saygin, et al. "Land cover mapping using sentinel-1 SAR data." The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 41 (2016): 757]
Soil	VV and VH channels	[Hu, Jingliang, Pedram Ghamisi, and Xiao Zhu. "Feature Extraction and Selection of Sentinel-1 Dual-Pol Data for Global-Scale Local Climate Zone Classification." ISPRS International Journal of Geo-Information 7.9 (2018): 379.]

At this point, features are computed according to the following steps:

a) Initially, the SAR time series is properly pre-processed by means of the methods and filters previously introduced in Section 4.

b) Then, all the de-speckled images in one year are first divided according to the season and then merged into one image per season by means of a temporal average. This step is performed as a thread-off between the need to keep multitemporal information and the one to reduce the computational load of the classification procedure.

c) Finally, the features useful for the extraction of the classes reported in the table above are computed for the final multitemporal sequence.

7.2 Classification

The classification procedure implemented in this work is based on a hierarchical extraction of specific classes followed by a general classification applied to the rest of the scene. Specifically:

- First, some of the classes that are recognizable using a specific subset of features are extracted from the data by means of unsupervised techniques. This is currently performed for the urban and water classes.
- Then, supervised classifiers, namely Random Forest (RF) and Support Vector Machines with Radial Basis Function (RBF) kernel, are applied to the set of features.

For the latter step, suitable training data are necessary, and two methodologies have been adopted to assess the performance of the classification chain. To avoid the unbearable cost of a manual extraction of high-

resolution samples a procedure able to extract samples for many of the desired classes from existing maps has been identified. This procedure must be complemented for specific areas and classes by more performing sample extraction methods (manual selection, for example), but it helps to reduce the cost of that procedure to a level which is manageable in the context of a global mapping methodology.

The following sections are dedicated to present the extractors developed for the urban and water classification, and to describe the ways followed to generate the training sets.

7.2.1   Description of the Urban EXTraction (EXT) algorithm

The main idea of The UEXT algorithm [24] is to use artificial structures, such as buildings, that show up as very bright points in multitemporally averaged and despeckled sets of SAR images (possibly reduced to single SAR images in case the temporal intervals of interest do not include more than one data set). The artificial structures corresponding to the largest values of the normalized backscattered power are selected as "seed pixels", and a flooding algorithm is subsequently and iteratively applied to their neighbourhood, until a lower threshold value is reached. Eventually, a series of post-processing steps, also considering the Digital Elevation Model of the area to avoid false positives due to mountainous slopes, are applied. The approach in [24] includes several intermediate steps that have been simplified in the approach implemented into the project to reduce the computational cost. Specifically, instead of the above-mentioned iterative flooding procedure, a single watershed technique has been introduced. The seed pixels are used as "markers" of the watershed algorithms, while the saliency map of the same algorithm is obtained by applying an occurrence data range filter to the SAR data sets. The size of the data range windows has been selected to keep as much as possible the spatial resolution of the data, and it is just a 3x3 pixel window.
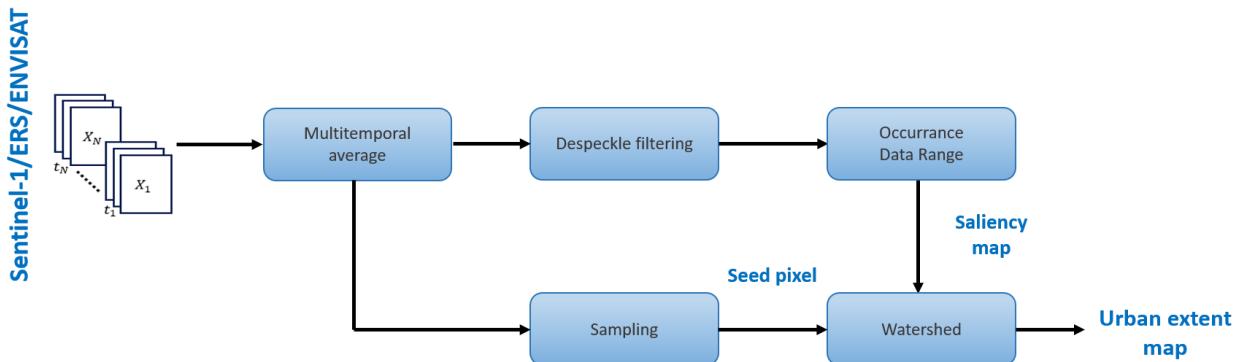
Figure 26. Graphical representation of the workflows for the UEXT method.

This method is visually described by the workflow in the Figure 26. The first step is the above-mentioned temporal average over the data sets in the temporal range of interest. This operation exploits the fact that urban areas respond coherently along the temporal axis, and thus a multitemporal average still results in bright backscattering pixels inside human settlements. This is not true for vegetated areas, whose backscattered values changes during seasons, hence their average will result in a lower value. The resulting image is then used twice: to identify seed pixel by hard thresholding the image, and to extract a saliency map by applying a data-range filter. This filter highlights urban areas, and thanks to the previous low-pass filter, the map results in homogeneous areas. The final step is the watershed algorithm, where the identified seeds would grow within the saliency map, thus producing the final urban extent map.

7.2.2   Description of the water extraction algorithm

The water extraction algorithm applied in this work has been recently presented in [72] and is summarized in the following figure. First, a multitemporal denoising step is performed on the SAR stack. Then, a set of statistical and temporal features are computed to train a k-Nearest Neighbor (k-NN) clustering algorithm. Finally, a few

post-processing steps (such as morphological operations and steep slope removing) are applied to the classified result to improve the final water surface mapping accuracy.
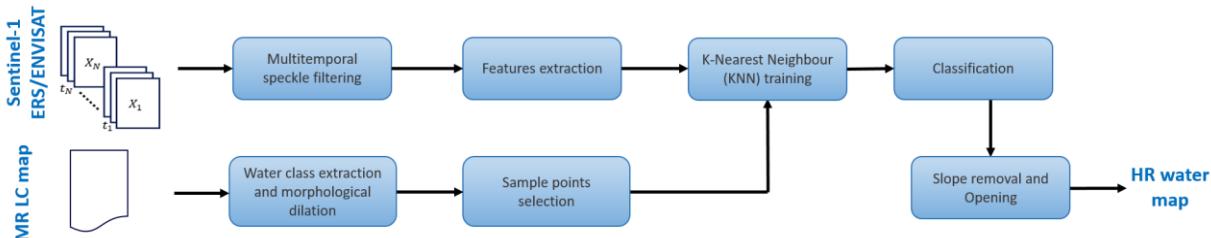
Figure 27. Overview of the complete procedure designed to extract both temporary and permanent water bodies.

To extract water bodies, a few features are computed: composites obtained by averaging in time all the available SAR images (if necessary, within more temporal windows) and their temporal mean, minimum, maximum, and variance. Then, starting from the 300 m map by the ESA CCI project, a mask representing water areas is extracted, resampled at 10 m spatial resolution, and dilated with a 3 x 3 kernel window. In each tile 10.000 pixels that fall either on water or non-water regions are randomly selected. These training points are clustered, and the extracted model is used to classify all points in the tile. The k-NN algorithm is performed with $k=4$, as inside the initial mask there may be water, vegetation, bare and impervious soil. Finally, once the clusters are extracted, the same number of points is randomly extracted from each cluster and compared with the water layer in the ESA-CCI map. The cluster with more points belonging to that layer is recognized as the "water cluster".

This result optionally undergoes a post-processing step using a Digital Surface Model (DSM) is performed. The slope on a per-pixel basis is used to filter out radar shadows in hilly and mountainous regions. Lastly, an "opening" morphological operation is applied to the slope-filtered result to remove foreground pixels. Contrary to a simple erosion, which may be too disrupting, the opening operation is defined by an erosion followed by a dilation, which ensures to remove only isolated false positives.

7.2.3   Training set generation from medium resolution maps

To automatically carry out a classification based on a training set extracted from the medium resolution products, we start from the assumption to classify in high resolution only pure classes that were recognized in medium resolution maps.

Specifically, the medium resolution maps that were considered are:

- *ESA CCI-LC 2018 (300m)*: The annual ESA CCI LC maps cover a period of 24 years from 1992 to 2018 at a spatial resolution of 300m. These maps describe the Earth's terrestrial surface in 37 original LC classes based on the United Nations Land Cover Classification System (UN-LCCS) [73]. The product that covers the 2015 year have been assumed as baseline.
- *GLCNMO (1km)*: The Global Land Cover by National Mapping Organizations (GLCNMO) is geospatial information in raster format which classifies the status of land cover of the whole globe into 20 classes at a spatial resolution of 1 km [74]–[76]. The classification is based on LCCS developed by Food and Agriculture Organization of the United Nations (FAO).

The proposed strategy aims to exploit the information associated only to those MR classes that present a good correlation with the high-resolution legend, excluding for instance the MR mixed-classes of ESA CCI-LC product reported in Table 4.

Table 6. Mixed classes list of the ESA CCI-LC 2015 (330m) product

Values	ESA CCI-LC 2015 (300m) labels
30	Mosaic cropland (>50%)/natural vegetation (tree, shrub, herbaceous cover) (<50%)

40	Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%)/cropland (<50%)
100	Mosaic tree and shrub (>50%)/herbaceous cover (<50%)
110	Mosaic herbaceous cover (>50%) tree and shrub (<50%)
180	Shrub and herbaceous cover, flooded, fresh/ saline/brakish water
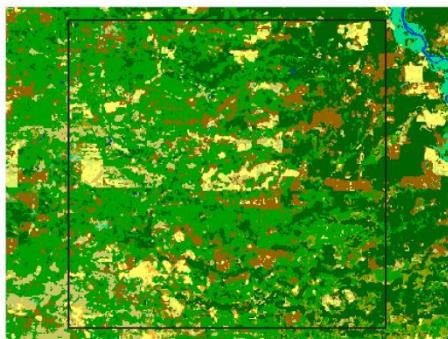
A comparison of the above-mentioned medium resolution map and the desired classes for HR mapping in this project has eventually brought to the results summarized in the following table:

Table 7. List of several medium resolution classes dealing to a training set for a high resolution classes subset
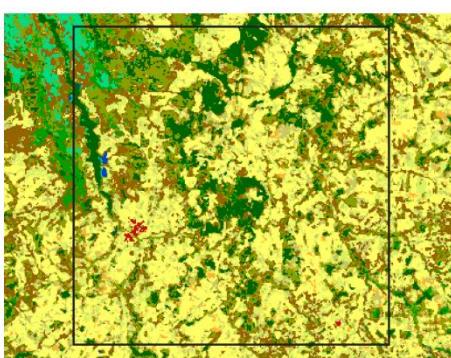
CCI-HR LC classes	ESA CCI LC 2018 (300m) values	GLCNMO (1km) values
Tree cover evergreen broadleaf	50	
Tree cover evergreen needleleaf	70, 71, 72	
Tree cover deciduous broadleaf	60, 61, 62	
Tree cover deciduous needleleaf	80, 81, 82	
Shrub cover evergreen	121	
Shrub cover deciduous	122	
Grassland	130	
Cropland		11,13
Vegetation acquatic or regularly flood	160,170,180	
Lichens and mosses	140	
Bare areas	200,201,202	
Sands		17
Rocks		16
Built-up areas	Urban extraction methodology [22]	
Open water seasonal Open water permanent	210	
Snow and/or ice	220	

In the second and third columns of Table 7, several values of ESA and GLCNMO legends, respectively, are selected in a way to provide a redundant yet meaningful set of training points for the corresponding high-resolution classes reported in the first column. For an easier reading of Table 7, the legends of ESA CCI-LC 2015 and GLCNMO maps have been shown in Figure 28 and Figure 29, respectively.

Legend of the global CCI-LC maps, based on LCCS

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

Figure 28. The CCI-LC MR maps referred to Amazonian tile 21KUQ (q) Amazonian tile 21KXT (b) classified according to the legend of the global CCI-LC maps (c).

Value	Class Name
1	Broadleaf Evergreen Forest
2	Broadleaf Deciduous Forest
3	Needleleaf Evergreen Forest
4	Needleleaf Deciduous Forest
5	Mixed Forest
6	Tree Open
7	Shrub
8	Herbaceous
9	Herbaceous with Sparse Tree/Shrub
10	Sparse vegetation
11	Cropland
12	Paddy field
13	Cropland / Other Vegetation Mosaic
14	Mangrove
15	Wetland
16	Bare area,consolidated(gravel,rock)
17	Bare area,unconsolidated (sand)
18	Urban
19	Snow / Ice
20	Water bodies

Figure 29. Legend associated to GLCNMO medium resolution map.

	Ref	CCI_HRLC_Ph1-ATBD	high resolution
	Issue	Date	Page	land cover
	4.rev.0	31/10/2022	50	cci

The final step for training point selection is performed using random sampling. This is performed by first selecting the points belonging (for each HR class) to the corresponding classes in the MR maps into binary maps. To avoid inaccuracies and collect more reliable samples, a morphological erosion step is applied to this binary map, and only its "internal" area is considered. Then, random sampling is applied. The procedure is repeated for each class, and a consistent set of training samples is extracted.

This approach does not reduce the resolution of the final HR map, which is obtained considering the original resolution of S-1 data. Moreover, by selecting only classes that are not mixed, it allows to obtain reasonably good training samples at a very limited cost. Of course, these samples are as accurate as (in average) the maps from which they are taken, and this is the reason why robust classifiers, such as RF and SVM has been selected.

7.2.4 Training set extraction from HR satellite images by visual interpretation

Reference data for large-scale land cover map are commonly acquired by visual interpretation of multiple remotely sensed data. Visual interpretation of high-resolution imagery is a valid alternative to a ground truth collection that would supply such very high-quality data, but populating a global dataset with a sufficiently large sample of field measurements is extremely costly. The proposed method consists in adopting a stratified strategy that aims in interpreting the land cover classes well-defined into the CCI HRLC legend taking into the account their characterization. The hierarchical approach is a simple but powerful methodology and consists of a top-down analysis organized as a clearly defined sequence of tasks grouped in levels, also called branches.
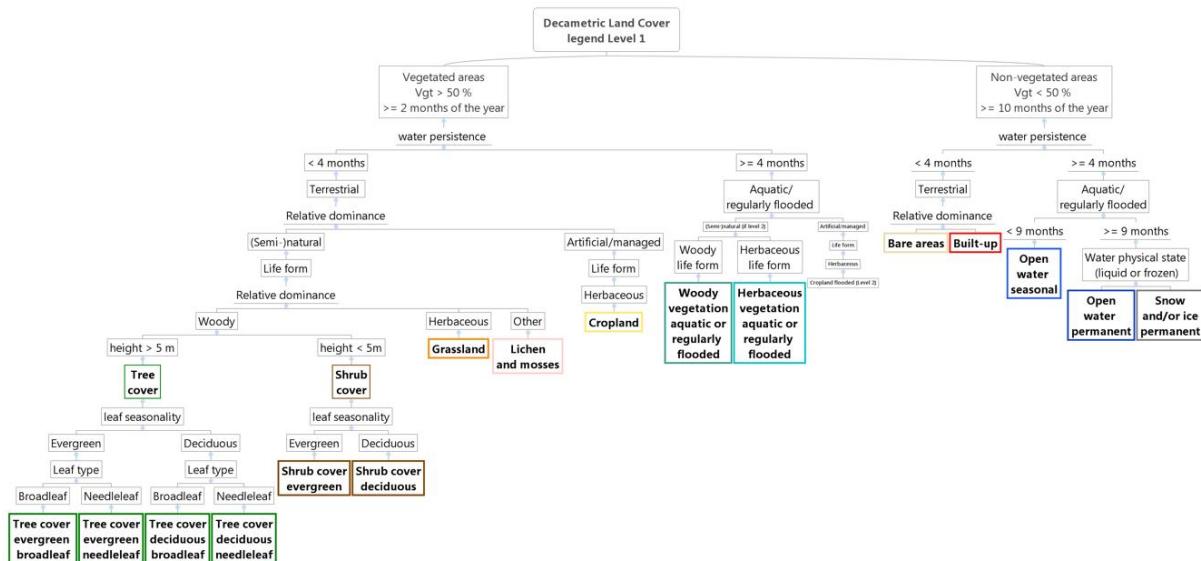

Figure 24. Decision tree of hierarchical approach based on 1st Level of CCI HRLC Legend

Each level solves a fundamental decision problem where systematic methods can be applied for the synthesis of further subsystems. All this can be depicted as a simple decision tree, as shown in Figure 24, where clear rules and exhaustive definition of land cover classes are needed to reach completeness of labelling training points via thresholds. With reference to Figure 24, a decision tree with four levels has been designed for identifying the Level 1 of CCI HRLC legend. Firstly it needs to distinguish in which percentage the area is covered or not by vegetation, in order to differentiate between vegetated and non-vegetated area. A similar reasoning is then replicated in each subsequent branch of decision tree, step-by-step, to verify relative dominance.

All the rules are explained below:

- **1st Level:**
  - *Vegetated area:* presence of vegetation >= 50% and for >= 2 months a year;
  - *Non-vegetated area:* presence of vegetation < 50 % and for >= 10 months a year.
- **2nd Level:**
  - *Terrestrial*: water persistence < 4 months;

- o *Aquatic/regularly flooded*: water persistence >= 4 months:
  - ▪ Seasonal: presence of water of water >= 4 months and < 9 months a year;
  - ▪ Permanent: presence of water of water >= 9 months a year.
- **3rd Level:**
  - o *(Semi-)natural*: natural vegetated areas are defined as areas where the vegetative cover is in balance with the abiotic and biotic forces of its biotope. Semi-natural vegetation is defined as vegetation not planted by humans but influenced by human actions.
  - o *Artificial/managed*: natural vegetation has been removed or modified and replaced by other types of vegetative cover of anthropogenic origin [93].
- **4th Level:**
  - o *Tree*: A tree is a woody, perennial plant with a simple and well-defined stem, bearing a more or less defined crown and a minimum height of 5 m;
  - o *Shrub*: A shrub is a woody perennial plant with persistent woody stems and without any defined main stem, being less than 5 m tall.
  - o *Grassland*: Herbaceous plants without persistent stem or shoots above ground and lacking definite firm structure
  - o *Croplands*: Mainly herbaceous plants are sowed/planted and harvestable at least once within the 12 months after the sowing/planting date. Herbaceous plants are defined as plants without persistent stem or shoots above ground and lacking definite firm structure. Cropland includes rain fed crops, irrigated crops, aquatic crops and annual pastures. It is an adaptation of the Joint Experiment for Crop Assessment and Monitoring (JECAM) cropland definition.
  - o *Lichens and Mosses*: Mosses are a group of photo-autotrophic land plants without true leaves, stems or roots. Lichens are composite organisms formed from the symbiotic association of fungi and algae.

To cover the HRLC classes, it needs also to define the leaf seasonality, for distinguish between evergreen (never entirely without green foliage) and deciduous (leafless for a certain period during the year), and the leaf type (needled or broad).

For instance, following the tree in Figure 24, a shrub cover deciduous class is a vegetated area where water persists for less than 4 months, that is (semi-)natural, where the shrub cover dominates other life forms and where the shrub cover has deciduous leaf seasonality. A vegetated area where the shrub cover is semi-natural and flooded for more than 4 months a year is a woody vegetation aquatic or regularly flooded.

There are challenges associated with the visual extraction of points as an abstraction of the nature of the land surface at a given point in space and time, and suitable sampling models have to be properly designed. About the sampling plan, the training points are typically extracted from sample sites with size 10x10m, composed of smaller sized square elements, with approximate 3x3m size. According the stratified approach and the land cover types, the interpreter can be assign to training point the class which predominates into the sample site of interest.

Figure 25. Example of sample site (highlighted in red) with 10x10m size. Working in hierarchy and taking into the account the scenario surrounding the site, the sample can be labelled as grassland.

Sample sites are extracted in random order, so to interpret different land cover types over the course of the samples labelling task.

A useful tool used to localize and analysis HR training points in a set of samples site is Collect Earth. Collect Earth is a free and open source software for land monitoring developed by the Food and Agriculture Organization of the United Nations (FAO). Built on Google desktop and cloud computing technologies, Collect Earth facilitates access to multiple freely available archives of satellite imagery, including archives with very high spatial resolution imagery (Google Earth) and those with very high temporal resolution imagery (e.g., Google Earth Engine, Google Earth Engine Code Editor) [94]. Collect Earth offers accesso to geo-synchronized visualization and use of imagery of varying spatial and temporal resolutions, including Sentinel 2, Landsat and MODIS imagery within Google Earth and Google Earth Engine. These imagery provide near-daily multispectral imaging of the Earth's land surface at resolutions ranging from 20 to 250m. Their frequent coverage provides a higher probability of observing the surface without interference from clouds, thus allowing the construction of global datasets in which nearly all points on the Earth's land surface have been imaged. Once the point of interest to be labelled was properly identified on the basis of an accurate photointerpretation of site sample, collect Earth enables the point confirmation by the NDVI analysis in Google Earth Engine of its corresponding optical time series (Sentinel-2, MODIS or Landsat).

	Ref	CCI_HRLC_Ph1-ATBD	high resolution
	Issue	**Date**	**Page**	land cover
	4.rev.0	31/10/2022	53	cci
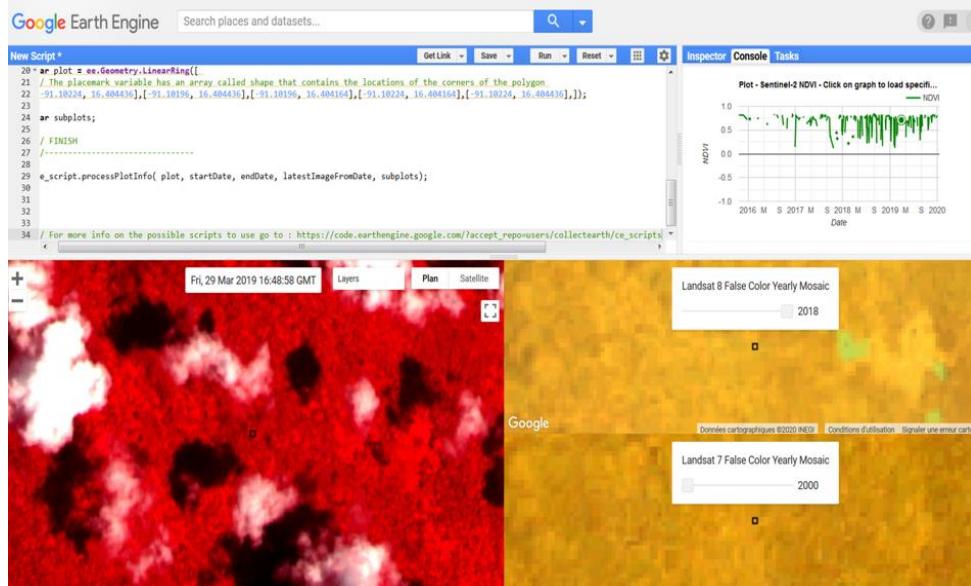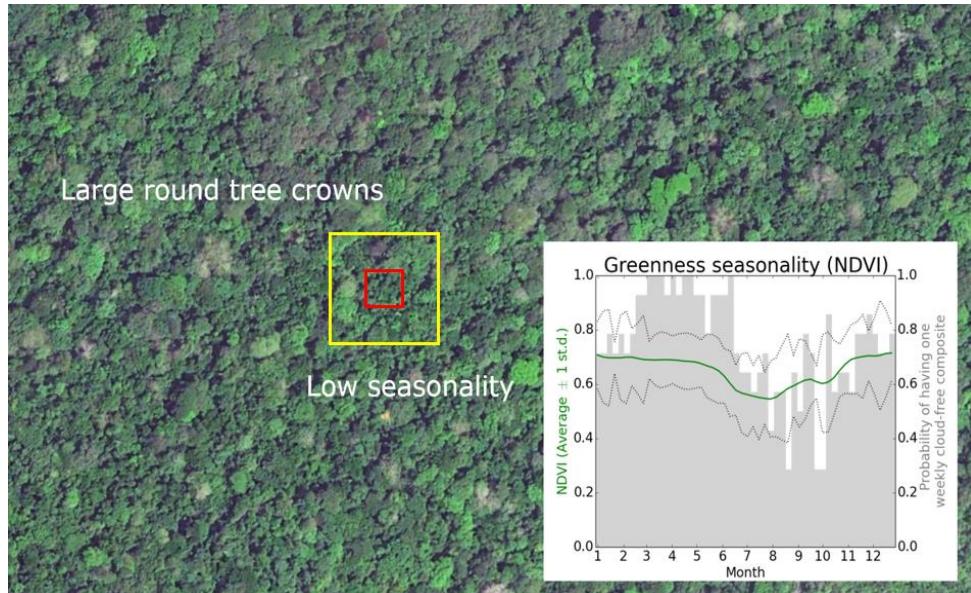
Figure 26. An example of Collect earth application. On the top, the visual interpretation of a point localised in a sample site belonging to the tree cover evergreen broadleaf HR land cover class. Its confirmation is realised according the corresponding NDVI analysis (on the bottom).

In Figure 26 an example of Collect Earth application for labelling a training point is shown. The considered sample site belongs to a tree cover broadleaf area, easily recognized by visual inspection within the image on the top because the crowns have a large round form. The point is also classified as evergreen by the NDVI seasonality analysis in Google Earth Engine (shown on the lower image), since its profile is constant over the year, indicating a low seasonality.

	Ref	CCI_HRLC_Ph1-ATBD	high resolution
⦿esa	Issue	Date	Page	land cover
	4.rev.0	31/10/2022	54	cci
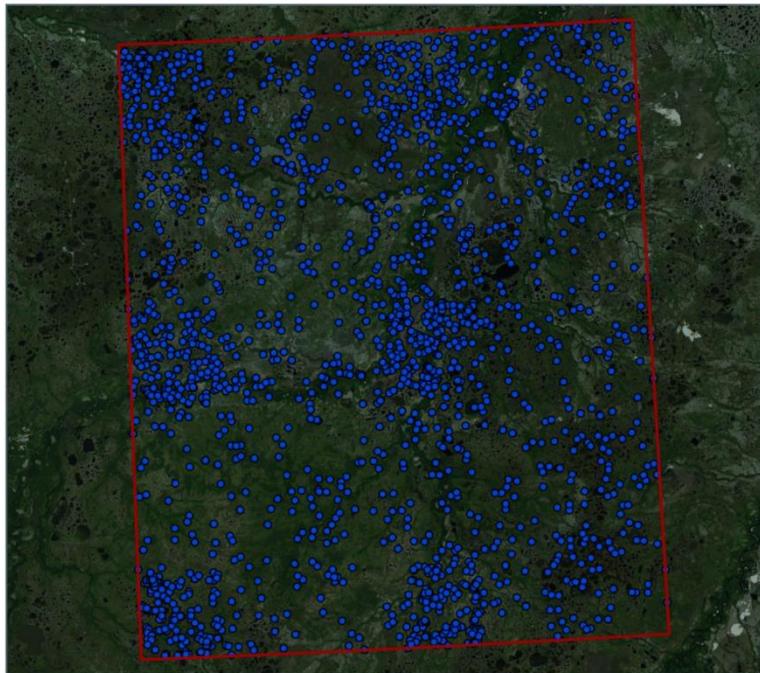
Figure 27. HR training points collected by visual interpretation and stratified approach. The area of interests is the benchmark tile 42WXS, belonging to Siberia area.

8 Decision fusion

Data fusion methodologies, and specifically the sub-class of decision fusion, allow making a common decision in case of multiple actors and opinions. Within the CCI+ HRLC pipeline, decision fusion combines the posterior probabilities associated with the outputs of the single classifiers that are applied to optical and SAR data separately. Therefore, multiple decisions are combined into a final result by taking into account the level of uncertainty associated with each source. This uncertainty is expressed precisely by the probabilistic characterization provided on a pixelwise basis by the aforementioned posteriors.

The sets of classes that can be accurately discriminated by using optical and SAR data separately do not coincide in general. While optical data are generally expected to be useful to the discrimination of all considered land cover classes, SAR data are expected to well discriminate especially built-up classes and water bodies. Accordingly, SAR and optical classification algorithms generally work on different, although obviously non-disjoint, sets of classes. Decision fusion methodologies are aimed at fusing posterior probabilities related to the classes in common across the two sets. Hence, a class-specific combination rule has been devised to take this into account and, correspondingly, integrate this fusion result on the common classes with the results obtained using only optical or SAR data for the remaining classes.

Specifically, the whole class legend Ω is divided into three disjoint subsets of thematic classes: Ω_O, the set of classes that are distinguished only by using optical data ("optical-exclusive"); Ω_S, the set of classes that are distinguished only by using SAR data ("SAR-exclusive"); and Ω_C, the set of classes which are discriminated by the classifiers operating with both data modalities (common classes). Accordingly, $\Omega = \Omega_O \cup \Omega_S \cup \Omega_C$. While the optical classifier works on the set of classes $\Omega_O \cup \Omega_C$, the SAR classifier outputs posterior probabilities for the set of classes $\Omega_S \cup \Omega_C$.

As a trade-off between computational complexity and expected accuracy, in the context of the CCI+ HRLC processing chain the following families of decision fusion methods are developed: (i) weighted voting and consensus-theoretic methods, and (ii) fusion strategies based on Markovian modelling (i.e., Markov and

	Ref	CCI_HRLC_Ph1-ATBD		high resolution
	Issue	Date	Page	land cover
	4.rev.0	31/10/2022	55	cci

conditional random fields). Both families are combined with class-specific combination rules that take into account the aforementioned rationale. Details can be found in the following subsections.

8.1 Consensus Theory and Class-Specific Combination Rule

Consensus theory [78], [79] involves general procedures with the goal of combining multiple probability distributions to summarize their estimates. The problem can be formulated as the combination of different opinions. This is represented as the fusion of posterior probabilities coming from different classifiers, each associated with a particular data source.

Under the assumption that all the classifiers can be made into generating Bayesian outputs and that, accordingly, their predictions are endowed with a probabilistic characterization, i.e., pixelwise posteriors are available, the goal is to produce a single probability distribution that summarizes their estimates. The study of such combination procedures is called consensus theory.

A first well-known consensus rule is the linear opinion pool (LOP). Focusing on the specific case of optical and SAR classifiers as sources generating the posterior probabilities and keeping in mind that the two classifiers generally work on different sets of classes, let $\underline{x} = [\underline{O}, \underline{S}]$ be the input data vector on a generic pixel, resulting from the stacking of optical ($\underline{O}$) and SAR ($\underline{S}$) individual feature vectors, and let ω_j be the jth information class ($\omega_j \in \Omega$). The LOP functional can be expressed as:

$$\mathcal{C}(\omega_j | \underline{x}, \Omega_C) = \alpha_j \, P(\omega_j | \underline{O}, \Omega_C) + \beta_j \, P(\omega_j | \underline{S}, \Omega_C),$$

where $P(\omega_j | \underline{O}, \Omega_C)$ is the optical posterior probability of ω_j conditioned to the common subset of classes Ω_C and $P(\omega_j | \underline{S}, \Omega_C)$ is the SAR posterior probability conditioned to the same subset Ω_C. α_j and β_j are optical and SAR source-specific weights, respectively, and control the relative influence of the two sources on the fusion output corresponding to class ω_j. We note that the pixelwise outputs of the optical-based and SAR-based classification chains are $P(\omega_j | \underline{O}, \Omega_O \cup \Omega_C)$ and $P(\omega_j | \underline{S}, \Omega_S \cup \Omega_C)$, respectively, i.e., the pixelwise posteriors associated with the corresponding sets of classes. Deriving $P(\omega_j | \underline{O}, \Omega_C)$ and $P(\omega_j | \underline{S}, \Omega_C)$ (as well as $P(\omega_j | \underline{O}, \Omega_O)$ and $P(\omega_j | \underline{S}, \Omega_S)$) is straightforward.

LOP has several good properties: it is simple, it yields a probabilistic formulation, and the weights α_j and β_j can reflect the relative expertise of the optical and SAR classifiers, respectively. The opportunity to use different values for these weight parameters in relation to different classes also allows reflecting possible prior information on the expected capability of optical and SAR data to discriminate each class. This can be interpreted indirectly as a weighted voting scheme that incorporates class-specific rules. Moreover, if the data sources have absolutely continuous probability distributions, LOP may be related to an absolutely continuous distribution [79]. LOP also assumes that all the experts (classifiers) observe the input vector $\underline{x}$. Therefore, LOP can be viewed as a weighted average of the probability distributions from the experts that results in a combined probability distribution. Yet, LOP is a simple method and, besides the aforementioned advantages, has also several weaknesses [80]. An example is a possible "dictatorship effect" when Bayes' theorem is applied (i.e., a specific data source dominates in making a decision). Moreover, not deriving from the joint probabilities using Bayes' rule, it is also not externally Bayesian (does not obey Bayes' rule).

Another well-known and usually effective consensus rule, the logarithmic opinion pool (LOGP), has been proposed to overcome some of the problems of LOP. In the optical-SAR case addressed here, the LOGP functional can be defined as:

$$\mathcal{L}(\omega_j | \underline{x}, \Omega_C) = \alpha_j \log P(\omega_j | \underline{O}, \Omega_C) + \beta_j \log P(\omega_j | \underline{S}, \Omega_C)$$

LOGP differs from the linear version in that it is usually unimodal and less dispersed. Zeros are considered vetoes: if any of the two sources assigns a zero posterior (i.e. $P(\omega_j | \underline{O}, \Omega_C) = 0$ or $P(\omega_j | \underline{S}, \Omega_C) = 0$), then by definition $\mathcal{L}(\omega_j | \underline{x}, \Omega_C) = 0$. This dramatic behaviour is a drawback when the single-source predictions are very inaccurate and can be generated even by roundoff error. In order to prevent this, all posterior values are increased by the machine epsilon (the minimum number that can possibly be represented given a certain data type).

$\mathcal{C}(\cdot)$ and $\mathcal{L}(\cdot)$ provide probabilistic fusion results associated with the classes in common between the two single-sensor outputs, although they generally do not take values in the interval [0, 1]. Either can be mapped to proper posteriors by suitably transforming to a probabilistic output, which represents a fused posterior probability $P_\mathcal{F}(\omega_j|\underline{x}, \Omega_C)$. In the case of LOP, $P_\mathcal{F}(\omega_j|\underline{x}, \Omega_C)$ is computed from $\mathcal{C}(\omega_j|\underline{x}, \Omega_C)$ by just re-normalizing so that the sum over all $\omega_j \in \Omega_C$ is unity. In the case of LOGP, the following softmax operator is appropriate to take into account the logarithmic relation between the $\mathcal{L}(\cdot)$ functional and the original probabilities:

$$P_\mathcal{F}(\omega_j|\underline{x}, \Omega_C) = \frac{\exp \mathcal{L}(\omega_j|\underline{x}, \Omega_C)}{\sum_{\omega_k \in \Omega_c} \exp \mathcal{L}(\omega_k|\underline{x}, \Omega_C)}$$

This probabilistic fusion output $P_\mathcal{F}(\cdot)$ covers the subset of classes in common across the two single-sensor classifications. To extend it to the whole set of classes, the posterior probability (unconditional with respect to Ω_C) can be defined according to the total probability theorem:

$$P_\mathcal{F}(\omega_j|\underline{x}) = P(\omega_j|\underline{x}, \Omega_C)P(\Omega_C|\underline{x}) + P(\omega_j|\underline{x}, \Omega_O)P(\Omega_O|\underline{x}) + P(\omega_j|\underline{x}, \Omega_S)P(\Omega_S|\underline{x}) =$$
$$= P_\mathcal{F}(\omega_j|\underline{x}, \Omega_C)P(\Omega_C|\underline{x}) + P(\omega_j|\underline{O}, \Omega_O)P(\Omega_O|\underline{x}) + P(\omega_j|\underline{S}, \Omega_S)P(\Omega_S|\underline{x}),$$

where the aforementioned probabilistic fusion result $P_\mathcal{F}(\omega_j|\underline{x}, \Omega_C)$ is used for the common classes, whereas the optical-based and SAR-based posteriors $P(\omega_j|\underline{O}, \Omega_O)$ and $P(\omega_j|\underline{S}, \Omega_S)$ are used for the two sets of exclusive classes. The aggregated posteriors of the three subsets of thematic classes Ω_O, Ω_S and Ω_C are modelled as follows:

$$P(\Omega_O|\underline{x}) = \lambda P(\Omega_O|\underline{O}, \Omega_O \cup \Omega_C), \qquad P(\Omega_S|\underline{x}) = (1-\lambda)P(\Omega_S|\underline{S}, \Omega_S \cup \Omega_C),$$
$$P(\Omega_C|\underline{x}) = \lambda P(\Omega_C|\underline{O}, \Omega_O \cup \Omega_C) + (1-\lambda)P(\Omega_C|\underline{S}, \Omega_S \cup \Omega_C),$$

where $0 \le \lambda \le 1$. This choice makes sure that the resulting terms correctly sum to unity (for all $\lambda \in [0,1]$), combines the optical- and SAR-specific probabilistic outputs using a LOP-like formulation on the common classes, and expresses the items associated with the exclusive classes as functions of the output of one of the two single-sensor processing chains. To determine an appropriate value for λ, we note that, in the limit case $\Omega_S = \emptyset$ (i.e., if the set of classes discriminated using SAR is a subset of the set of classes discriminated using optical data), $\lambda = 1$ is a desired choice. Vice versa, in the limit case $\Omega_O = \emptyset$, a desired value is $\lambda = 0$. A suitable weight that covers both limit cases is:

$$\lambda = \frac{P(\Omega_O)}{P(\Omega_O) + P(\Omega_S)},$$

where the prior probabilities $P(\Omega_O)$ and $P(\Omega_S)$ can be estimated on the training set. Therefore:

$$P_\mathcal{F}(\omega_j|\underline{x}) = P_\mathcal{F}(\omega_j|\underline{x}, \Omega_C)[\lambda P(\Omega_C|\underline{O}, \Omega_O \cup \Omega_C) + (1-\lambda)P(\Omega_C|\underline{S}, \Omega_S \cup \Omega_C)]$$
$$+ P(\omega_j|\underline{O}, \Omega_O)\lambda P(\Omega_O|\underline{O}, \Omega_O \cup \Omega_C) + P(\omega_j|\underline{S}, \Omega_S)(1-\lambda)P(\Omega_S|\underline{S}, \Omega_S \cup \Omega_C).$$

This combination rule is applicable to all cases, independently on the set of classes with which the two classifiers works. It is worth noting that, in the fusion of optical and SAR data, a frequent scenario is that one of the two sources discriminates among a larger set of classes than the other source. In particular, SAR-based classifiers typically work on a set of classes which is a proper subset of the set of classes considered by optical classifiers. In this case, we have $\Omega_S = \emptyset$ and then $\Omega_S \cup \Omega_C = \Omega_C, \Omega_O \cup \Omega_C = \Omega$, and $\lambda = 1$. Therefore the previous formulation simplifies as follows:

$$P_\mathcal{F}(\omega_j|\underline{x}) = P_\mathcal{F}(\omega_j|\underline{x}, \Omega_C)P(\Omega_C|\underline{O}, \Omega) + P(\omega_j|\underline{O}, \Omega_O)P(\Omega_O|\underline{O}, \Omega),$$

where it is possible to remove the conditioning on the whole set of classes:

$$P_\mathcal{F}(\omega_j|\underline{x}) = P_\mathcal{F}(\omega_j|\underline{x}, \Omega_C)P(\Omega_C|\underline{O}) + P(\omega_j|O, \Omega_O)P(\Omega_O|\underline{O}).$$

Within the HRLC pipeline, special focus is given to the definition of the weights α and β. Several approaches are being explored. The first is the use of uniform weights, which formalizes the case in which the decision maker has no knowledge on which source is more reliable. On one hand, this is straightforward; on the other hand, it does not benefit from the aforementioned properties of optical and SAR data in terms of the capability to

discriminate the various classes. More accurately, it is possible to assign the weights proportionally to a score that is set according to the "goodness" of each source, where a higher score indicates a better (i.e., more reliable) source. This scoring may be accomplished by assessing the accuracy of the land-cover predictions coming from the optical and the SAR sources. Another solution that is considered is to compute the weights through linear or nonlinear optimization [81] [78]. In particular, the method in [17] which is based on the expectation-maximization (EM) algorithm, can be incorporated into the HRLC pipeline. It regards a LOGP-type model in the framework of unsupervised change detection and will be generalized here to the case of supervised land-cover classification. Finally, case-specific tuning can be performed according to possible intermediate feedbacks from the Validation Team and the Climate Group about the quality of the resulting HRLC map. In the formulation of the method integrated in the first production, this last option has been selected to align as much as possible the output product to the requirements of the climate community.

8.2 Markov Random Fields

Markov random fields (MRFs) are probabilistic graphical models able to include contextual information in the form of class interactions between neighbouring pixels. An MRF is determined by an energy function, whose minimization with respect to the labels is equivalent to the application of a maximum a-posteriori criterion [82]. Considering an MRF model in which only up to pairwise clique potentials are non-zero (comparing items one couple of nodes at a time), this energy is composed of two main terms: one characterizing class likelihood at the pixel level (depending on per-class scores obtained from any method able to estimate posterior or class-conditional probability density functions), and another promoting label smoothness in a local neighbourhood [82]. This means that the model encourages two neighbouring pixels to be labelled with the same class.

Let Ω be again the set of thematic classes. Define the regular pixel lattice as I, and let y_i be the class label of the i-th pixel ($y_i \in \Omega, i \in I$). The MRF considers y_i as sample of the random field $Y = \{y_i\}_{i \in I}$ of class labels, which is discrete-valued. A neighbourhood system $\{\partial i\}_{i \in I}$, which provides each i-th pixel with a set $\partial i \subset I$ of neighbouring pixels, is defined [83]. It is possible to choose different kinds of adjacency systems: the ones that have being used the most include the first- and second-order connectivity [83]. In the former, ∂i is made of the four pixels adjacent to the i-th pixel (four-connected) while in the latter the eight pixels surrounding it are considered.

Considering the aforementioned frequently used family of the MRF models in which only up to pairwise clique potentials are non-zero, the energy can be written as:

$$U(Y|X) = -\sum_{i \in I} \alpha \log P(y_i | \underline{x}_i) - \gamma \sum_{\substack{i \in I \\ j \in \partial i}} \delta(y_i, y_j).$$

where α and γ are positive weights and $\delta(\cdot)$ is the Kronecker impulse. In the multi-sensor case, a different unary term is added for each sensor, so that it is possible to fuse the different posterior probabilities while enforcing contextual relationships. The overall equation is:

$$U(Y|X) = -\sum_{i \in I} \sum_{s=1}^{S} \alpha_s \log P(y_i | \underline{x}_{is}) - \gamma \sum_{\substack{i \in I \\ j \in \partial i}} \delta(y_i, y_j),$$

where the notation x_{is} indicates the dependence of image data on both the pixel location i and the sensor s ($s = 1,2, \dots, S$), S is the number of sensors, and $\{\alpha_s\}_{s=1}^{S}$ is a set of positive weights.

Within the HRLC pipeline, in order to ensure consistency with the aforementioned pixelwise formulation and inspired by the similarity between the unary term and LOGP, the MRF approach is applied to the posterior probabilities resulting from the pixelwise fusion of the outputs of the optical and SAR classifiers. Therefore, in our specific setting, the overall equation becomes:

$$U(Y|X) = -\sum_{i\in I} \alpha \log P_{\mathcal{F}}(y_i|\underline{x}_i) - \gamma \sum_{\substack{i\in I \\ j\in\partial i}} \delta(y_i, y_j),$$

As compared to the previous fusion approaches, the strategy based on MRFs incorporates spatial information, an important contribution in the application to high-resolution remote sensing imagery, which is intrinsic in the HRLC project. The weights α and γ that tune the tradeoff among the various contributions to the energy function U are optimized by extending to the MRF fusion formulation the approaches described in Section **Error! R eference source not found.** with regard to the consensus formulation. Also in this respect, intermediate feedback from the Validation Team and the Climate Group about the desired characteristics of the output HRLC map in terms of smoothness, removal of salt-and-pepper classification noise, and detail preservation have been taken into account.

In the application of MRF-based methods to decision fusion, special focus is devoted to the minimization of the energy function U with respect to the random field Y of the class labels [84]. First, as an efficient tradeoff between accuracy and computational burden, the iterated conditional mode (ICM) algorithm is adopted. It ensures short execution times, yet, it converges to a local minimum of the energy, which can be possibly suboptimal [82]. We shall investigate, either methodologically or experimentally, the opportunity to make use of global (or near-global) energy minimization methods based on graph-theoretic concepts (namely, graph cuts [83] and belief propagation methods [82]). On one hand, they ensure convergence to minima with stronger optimality properties than ICM. On the other hand, their computational burden is significantly higher and needs to be properly evaluated according to the data size involved in the HRLC project.

8.3 Cascade multitemporal model

Multitemporal models are used to propagate information towards years. This is especially important in the case of historical maps since, in these cases, the availability of data has lot of variability. This may cause inconsistencies in the classification, due to sparse acquisitions taken in different months of the year.

With the use of a multitemporal cascade model, it is possible to enforce consistency in the maps while preserving the actual changes of interest for the study. According to the rationale of the HRLC project to generate a static product in 2019 and historical products going backward to 1990, a cascade multitemporal approach is used. The scheme of the cascade approach is shown in Figure 8.1:
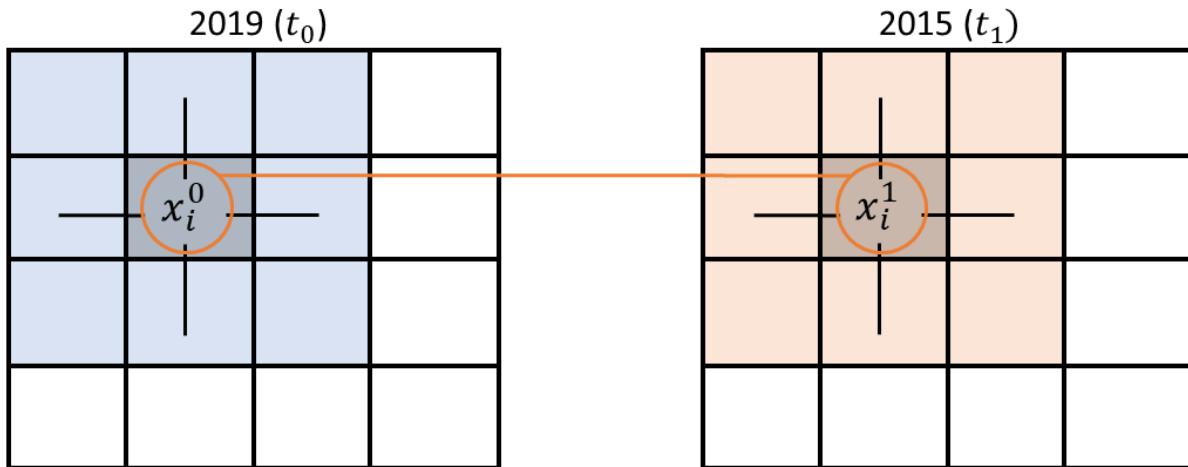

Figure 8.1: Temporal dependence in cascade model (example for the pair 2015-2019).

Keeping the same notations used in previous section, let the entire feature vector for pixel i at times t_0 and t_1 be, respectively x_i^0 and x_i^1. In the same way, the corresponding labels are $\{y_i^t\}_{t\in\{0,1\}}$.

Then, the fused posteriors at time t_0 (the ones to propagate) are $P_F(y_i^0|x_i^0)$, while the ones at time t_1 are $P_F(y_i^1|x_i^1)$. It is worth noting that $P_F(y_i^0|x_i^0)$ coincides with the final posteriors probabilities obtained at time t_0, as described in section 8.1.

The final fused posterior probability obtained by the cascade method is then given by[85]:

$$P_F(y_i^1|x_i^1, x_i^0) \propto \frac{P_F(y_i^1|x_i^1)}{P(y_i^1)} \sum_{y_i^0} \frac{P_F(y_i^0|x_i^0)}{P(y_i^0)} P(y_i^0, y_i^1),$$

where $P(y_i^1)$ and $P(y_i^0)$ are the prior probabilities corresponding to t_1 and t_0, respectively, which are often omitted considering that spatial MRF prior is already in the model. Then $P(y_i^0, y_i^1)$ is the joint probability matrix (JPM) representing the chances of having a temporal transition among certain classes.

This matrix is different for each area of study and derived according to the indications of the Climate group. It is worth recalling that the adopted cascade approach is a rigorous probabilistic Bayesian strategy to incorporate temporal dependence information in the classification of a time series. Its integration in the pipeline for the historical product is aimed at minimizing inconsistencies across the different years. At the same time, its probabilistic Bayesian formulation prevents the risk of erroneously "censoring" specific temporal transitions (like a deterministic rule-based approach could do). In particular, the use of a dedicated JPMs on each one of the three study areas of the project is consistent with the fact that diverse temporal transitions are expected in Sahel, Siberia, and the Amazon. In this respect, the use of this prior information about the JPM plays a role similar to the use of prior information about the land cover classes, as represented by their training samples.

The obtained $P_F(y_i^1|x_i^1, x_i^0)$ is then integrated into an MRF formulation at time $t = t_1$, having:

$$U(Y^1|X^1, X^0) = -\sum_{i \in I} \left[-log \, P_F(y_i^1|x_i^1, x_i^0) - \gamma \sum_{\substack{i \in I \\ j \in \partial i}} \delta\left(y_i^1, y_j^1\right) \right],$$

8.4   Spatial Harmonization

The spatial harmonization module is in charge of ensuring consistency across the granule borders in the final fusion product. In general, it may happen that, due to the different characteristics of data, and more generally due to possible issues on data availability, neighbouring granules may result in land cover maps with slightly different characteristics. These differences may generally impact the output product in terms of edge artifacts at the interface between the two granules. Therefore, in order to prevent such possible artifacts in the final mosaic, a spatial harmonization module is run on such neighbouring granules. The overlapping parts among the two granules are indeed fused using an ad hoc linear opinion pool that incorporates space-varying weights to ensure seamless spatial fusion (see 8.2).
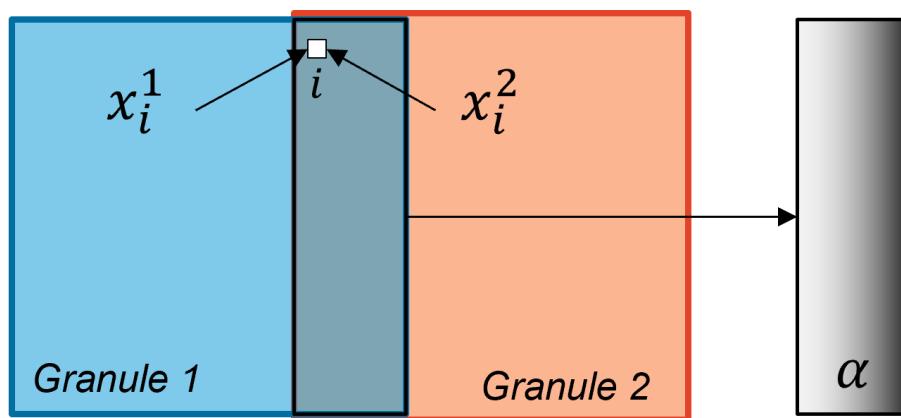

Figure 8.30: Overlapping granules and space-varying weights (brighter values are closer to 1, while darker values are closer to 0)

	Ref	CCI_HRLC_Ph1-ATBD	high resolution
	Issue	Date	Page	land cover
	4.rev.0	31/10/2022	60	cci

Let x_i^1 be the value of the i-th pixel in the lattice of the first granule, let x_i^2 be the value of the corresponding pixel in the lattice of the second granule, and let α be a set of weights that spatially vary with a constant gradient over the intersecting area, with $\alpha_i \in [0,1]$ be the weight corresponding to the previously considered i-th pixel. The spatial harmonization module computes the harmonized posterior probability $p(y_i|x_i^1, x_i^2)$ according to:

$$P(y_i|x_i^1, x_i^2) = \alpha_i\, P(y_i|x_i^1) + (1 - \alpha_i)\, P(y_i|x_i^2),$$

where $P(y_i|x_i^G)$ is the posterior probability resulting from the decision fusion module applied to the granule G, with $G = \{1,2\}$. More specifically, the two input posteriors are meant as the output of the MRF stage. The spatial harmonization process, which has been described here with regard to a pair of neighbouring granules, is then applied across all four borders of each granule (i.e., up, down, left, and right).

8.5   Temporal Harmonization

8.5.1   Post classification comparison module

The PCC generator module is responsible for computing the post classification comparison (PCC) map that is an input to the change detection module.

The change detection module analyses the whole time series of raw data to determine whether a change occurred in a given time window and, in the positive case, in which year it occurred. To minimize its computational burden, the processing is performed only on those pixels that are marked as changed by the optical, SAR, and fusion processing chains, i.e., the pixels that exhibit different land cover labels in the LC maps obtained five years apart from each other. The PCC generator processor is indeed responsible for providing the change detection module the indication about such pixels to process.

Let Y^t and Y^T be the land cover maps generated by the decision fusion module, let t and T be the corresponding years, with $t < T$, and let Y_i^t and Y_i^T be the labels assigned to the i-th pixel in the two maps. In case Y^t and Y^T are defined on the same pixel lattice (i.e., they are generated with the same spatial resolution), the computation of the PCC map is straightforward and only requires comparing the two maps on a pixel-by-pixel basis:

$$PCC_i = \begin{cases} 1 & if \quad Y_i^t \neq Y_i^T \\ 0 & if \quad Y_i^t = Y_i^T \end{cases}$$

When the two land-cover maps are defined on two different pixel lattices (i.e., one map corresponds to the year 2019 at 10m resolution and the other corresponds to the year 2015 at 30m resolution), a different processing scheme is adopted (see 8.3). Please note that the PCC map is always defined on the coarser lattice, as the change detection module works at the resolution of 30m.
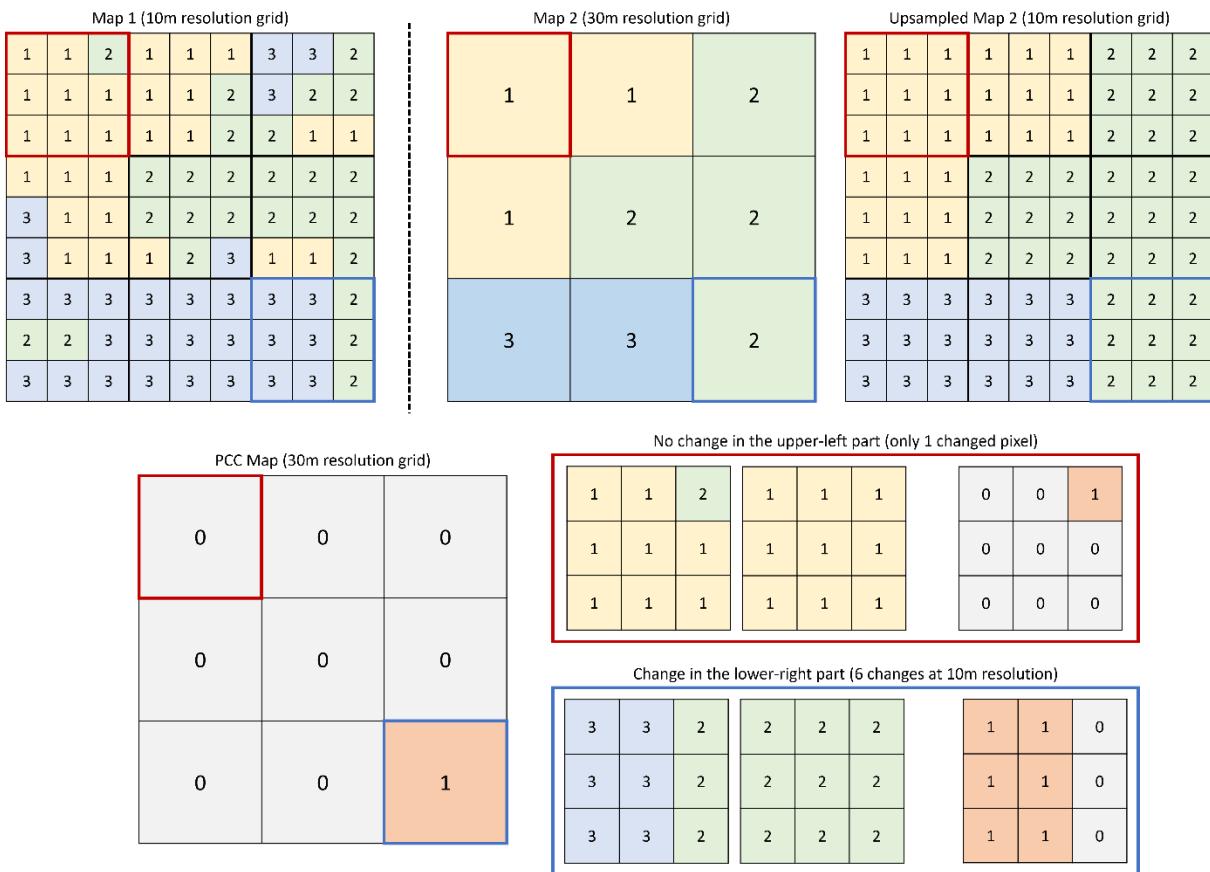
Figure 8.3: Flowchart of the PCC generation when two different resolutions are considered.

First, the coarser-resolution image is upsampled to the finer-resolution grid. Due to the integer resolution ratio, there exists a 3x3 window on the finer lattice that corresponds to a single pixel in the coarser lattice. The upsampling is done by replicating each pixel at 30m resolution onto the corresponding window of the 10m lattice. Then, the PCC map is computed by comparing each 3x3 window: a change label is assigned when there are more than 4 pixels (out of the 9 pixels in the window) with different values.

8.5.2 Temporal harmonization module

The temporal harmonization module is deployed in the historical processing chain and is responsible for the post-processing of the fused land cover maps and the successive update based on the results of the change detection module. In practice, the module acts as a feedback loop between the change detection module and the fusion module.

First, after applying the cascade multitemporal module, a check is performed on the produced land-cover maps for assessing proper temporal consistency. The cascade multitemporal module is able to model the probability of land-cover transitions between two separated years and uses this information to properly regularize the predicted class posterior probabilities of the historical classifiers by exploiting a robust Bayesian formulation. However, it does not model land-cover trends over more than two years as it is allowed to see the posteriors of only two dates at time. For this reason, some wrong land-cover changes might remain in the maps. In many cases, a priori information can be used to detect and remove them. For example, for some classes is not reasonable that their presence in the time series is allowed to oscillate, such as the urban areas. Hence, expert rules have been adopted to detect those unrealistic trends and then to correct them with the expected trend. For example, in the case of urban areas, if we have built-up in the whole time series but not in 2005, then the land cover of 2005 is expected to be built-up as well. It is worth noting that the historical time series of land-cover maps is at the resolution of 30m. In the case of 2019, the static map at the resolution of 10m is first

downsampled to the target resolution of 30m. Due to the resolution ratio with the historical maps, there is always a 3x3 window at 10m resolution that corresponds to a single pixel in the 30m resolution lattice. The label assigned to the 30m pixel corresponds to the most frequent label in such a 3x3 window. In the case of a tie, the label is chosen randomly among such most frequent labels. There are cases where the temporal consistency check finds errors and inconsistency in the 2019 map with respect to the whole historical time series, which are thus corrected. Hence, the output 30m resolution version of 2019 is saved and delivered as part of the historical products.
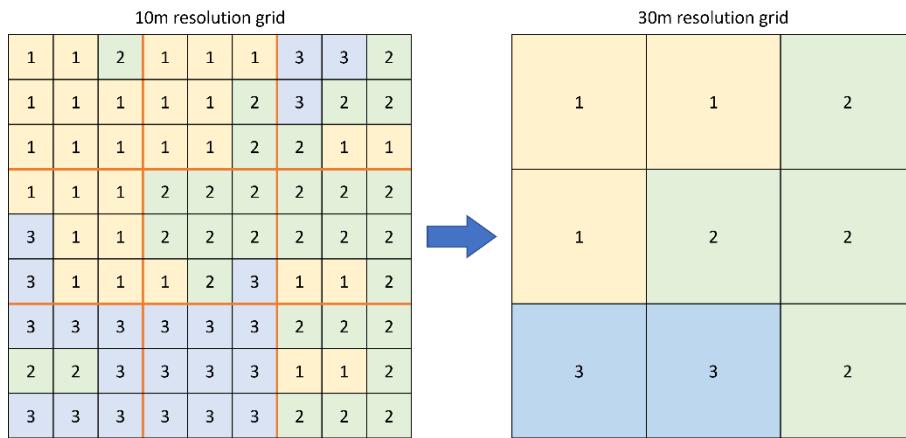

Figure 31. Practical example of the downsampling of the high-resolution LC map.

Then, the change detection module is run iteratively starting from 2015-2019 to 1990-1995. After each iteration the change detection output is used to update the least recent land-cover map. The rationale is that the temporal consistency check outputs the land cover maps of all the considered years. Then, the change detection module analyses the pixels that, according to the fusion module, have changed within a given time frame. For such an analysis, the change detection module uses the entire time series of data for a given pixel. There are cases in which this thorough analysis reports that no change is detected in the considered time window. Therefore, the temporal harmonization module uses this information to ensure consistency with the thorough analysis done by the change detection module. Figure 32 details the processing scheme of the temporal harmonization module. The image only considers the best-class product. Nevertheless, the output of the CCI HRLC processing chain consists of the best-class map, the second best-class map, and the two posterior probabilities corresponding to the aforementioned maps. Concerning the second best-class map, it is always kept unchanged and, for those pixels in which the change is reverted, it inherits the best-class label that was chosen by the fusion module and then reverted. Finally, the output uncertainty measures mirror those coming from the optical, SAR, and fusion processing chains.
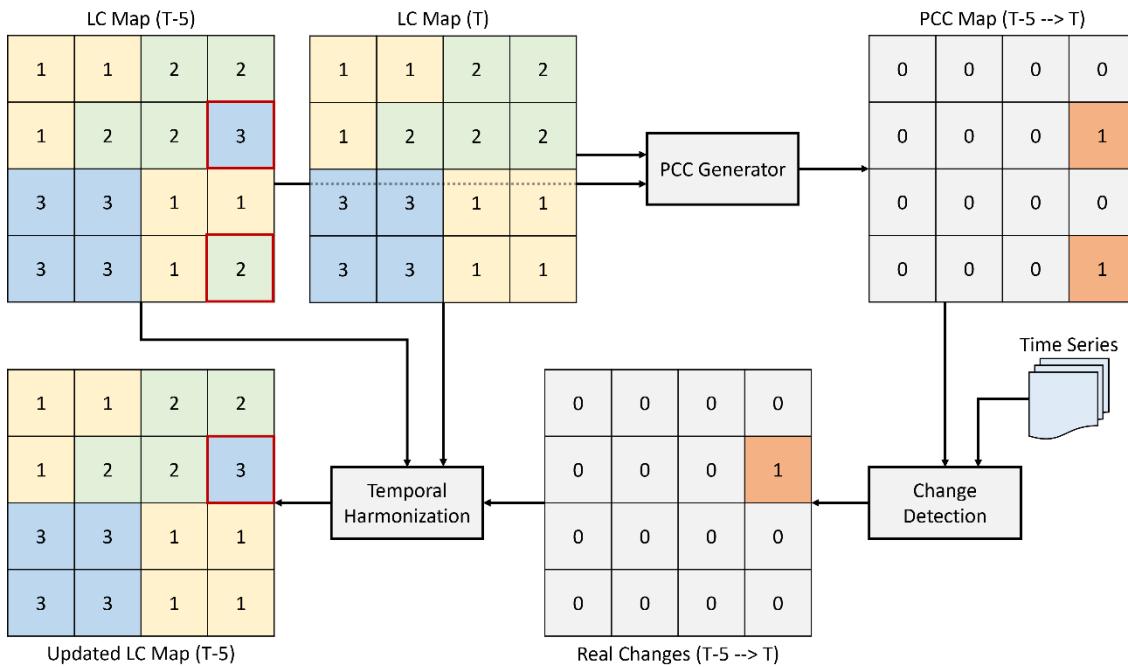
Figure 32. Flowchart of the temporal harmonization processing scheme

8.6 Deep Learning Solution

As an alternative to the aforementioned approaches to decision fusion, the multisensor fusion stage of the HRLC processing chain can also benefit from deep learning architectures. In this case, multi-sensor classification and fusion are dealt with by a deep convolutional neural network [47], [84], [85], [86] rather than by the specific aforementioned formulations. This is promising from the viewpoint of classification performance as confirmed by the accuracy gain observed in several recent international contests, in which deep learning solutions have overcome previous methods (e.g., recent IEEE GRSS Data Fusion Contests [87], [88] or ISPRS 2D Semantic Labeling competitions [89], [90]). On the other hand, the implementation, training, and computational complexity of the deep formulation will be significantly higher than those involved by the previous, more traditional, approaches.

In the specific case of the decision fusion block of the HRLC processing chain, an effective deep learning formulation would be based on the aforementioned CNN, autoencoder, and adversarial components that have been mentioned in previous sections. Adversarial networks are especially promising in this case thanks to their domain adaptation capabilities and to the opportunity to use them to map optical and SAR products into a homogeneous domain [49] (see also Section 5).

9 Multitemporal change detection and trend analysis

In accordance with the SoW [AD2] and as per the lessons learned from the CCI MRLC, the scheme shown in Figure 33 is used for the generation of HRLC change products. In particular, the multitemporal change detection (CD) and trend analysis processing chain, assumes to have the entire optical data time series from 1990-2019 already pre-processed (i.e., radiometrically corrected and co-registered). Additional to this information, this processing chain requires as input the five years regional HRLC maps (30m). As output from the processing chain, there will be the change information at 30 m in yearly time scale.
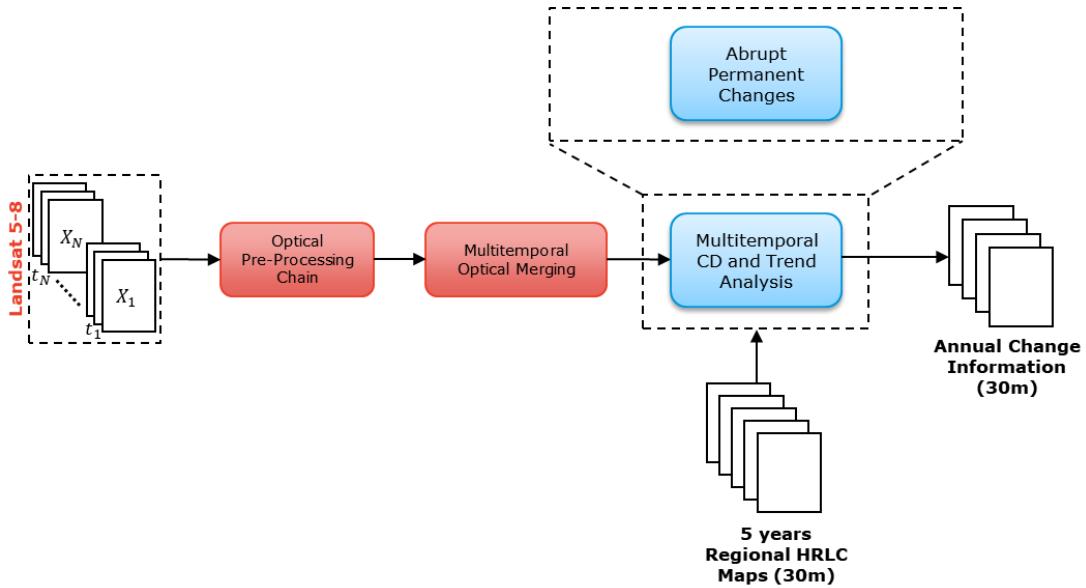
Figure 33. Block-based representation of the processing chain for the multitemporal change detection and trend analysis.

LC Changes can be divided into three classes [86]: (1) seasonal changes, impacting plant phenology or proportional cover of LC types with different plant phenology; (2) gradual changes such as inter-annual climate variability (e.g., trends in mean NDVI) or gradual change inland management or land degradation; and (3) abrupt (or permanent) changes, caused by disturbances such as deforestation, urbanization, floods, and fires.

The CCI HRLC change products will be developed with an emphasis on quantification of abrupt/permanent changes since climate change tends to be more abrupt than gradual. The analysis is performed over the products derived from the multitemporal optical merging step, plus the HRLC static and five years regional maps.

9.1 Timeline analysis and Cascade Paradigm

Backward timeline analysis considering every five years is defined in order to be in-line with land cover maps production and faster in the processing chain (see Figure 34). The analysis will be performed in a top-down time scale direction and abrupt/permanent changes occurring at longer time scales will be identified in an unsupervised way.
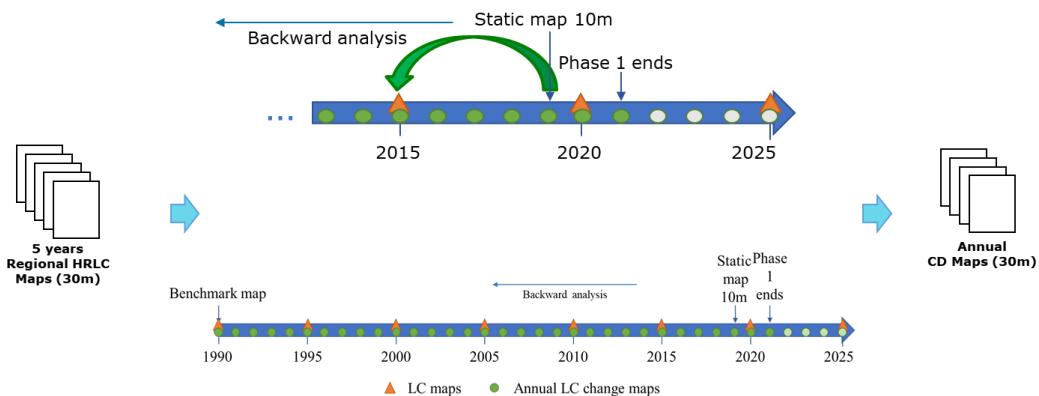

Figure 34. Timeline analysis in the processing chain.

The processing chain is based on a cascade paradigm (Figure 35) and the CD algorithm will use the LC maps every five years to produce the yearly change maps.
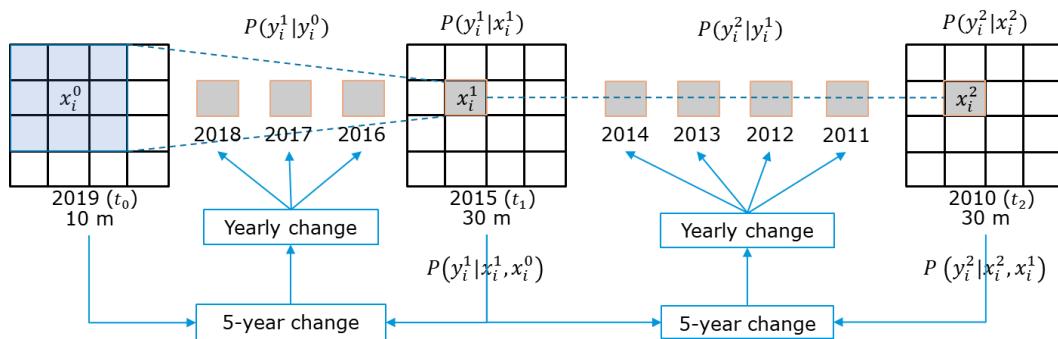
Figure 35. cascade paradigm of the processing chain.

9.2   Post Classification Comparison Map

Considering the LC maps produced every five years, it is possible to generate the Post Classification Comparison (PCC) maps. These maps have been produced in order to: 1) align the changes that have occurred during five years derived from the LC maps to the changes detected in multitemporal change detection processing chain, and 2) reduce the computational burden. Here we should take into account that the CD processing chain works in pixel level in a yearly basis from 1990 to 2019 considering three subcontinental area in Amazonia, Africa and Siberia [AD4]. The changes in different areas have been categorized to what really matters for climate modelling as it is visualized in Figure 36.

	Amazonia	Africa	Siberia
Land disturbances	– deforestation, land clearing by fires or cutting – crop plantation – afforestation – floodings linked to heavy precipitation or dam construction	– drought events – deforestation – afforestation – crop plantation – floodings linked to heavy precipitation or dam construction	– landslides linked to heavy precipitation on thawed permafrost areas – floodings linked to extreme precipitation events or dam construction – longer freezing periods during cold years – fires/storm/tree cuttings leading to land clearing – infrastructure developments for gas exploitation in the arctic zone

Figure 36. High priority transitions for different areas.

As a result, the changed pixels derived from PCC map have been divided into two categories, high priority changes and low priority changes. The pixels that have highlighted as the high priority changes will be further analyzed in the multi-temporal CD and trend analysis step.

9.3   Abrupt/permanent change and trend detection

A limited number of methods have been developed in the literature that allow the analysis of long time series (with 16 days acquisitions) and can be considered as scalable to the spatial resolutions of the available sensors in this project. Possible adaptation/combination is foreseen, given the fact that most of state-of-the-art methods: (1) have been developed for medium and/or low spatial resolution applications; (2) make use of a single spectral value per each evaluated year; and (3) focus on single LC only (e.g., forest and/or vegetation). In order to map the abrupt/permanent changes, Breaks For Additive Seasonal and Trend (BFAST) [87] to be considered that is a generic CD approach for time series, involving the detection and characterization of change. BFAST integrates the iterative decomposition of time series into trend, seasonal and noise components with methods for detecting changes, without the need to select a reference period, set a threshold, or define a change trajectory. In other words, using BFAST methodology will allow us to: i) detect multiple abrupt/permanent changes in the seasonal

and trend components of the time series, ii) characterize the gradual and abrupt ecosystem changes by deriving the time, magnitude, and direction of change within the trend component of the time series; and iii) generate color-coded maps where different colours represent the year in which a given change has occurred. The main limitation of this method is that it has been developed for MODIS data and tested mainly in NDVI index, and a few vegetation indices, and in particular for forest change detection. Adaptation to both HR data and other spectral information is thus required.

Figure 37 shows the general block scheme followed in this case, where features are first extracted from optical TS. As second step, feature TS are regularized to compensate for further errors like cloud/shadow from pre-processing steps. As third step, the adapted BFAST algorithm will be used to generate the color-coded change map. As the output, the method uses the multi-feature regular SITS to provide the information of: i) pixels with abrupt changes in long term SITS across regional extents ii) the time in which a change has happened, iii) the probability of a given change.
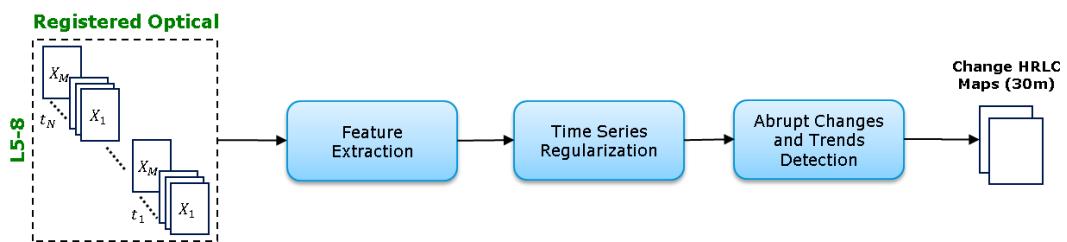

Figure 37. General block scheme multitemporal change detection processing chain.

9.4 Multitemporal change detection processing chain

Let $SITS = \{X_1, X_2, \ldots, X_N\}$ be a pre-processed satellite image time series acquired over the same geographical area in the period $[t_1, t_N]$. Assume the SITS have non-uniform time sampling, and each image has a total number of P pixels. Given an image $X_n \in SITS$, each pixel value represents the surface reflectance value in a given spatial position and a temporal instant t_n. Let $B = \{b_1, b_2, \ldots, b_K\}$ be the set of bands that compose the images and K the total number of bands.

In details (Figure 38), the input of the processing chain is pre-processed SITS that is employed in feature extraction together with feature reduction to distinguish the spectral trends of different sets of LC changes. Then the time series reconstruction is developed to generate a continues and dense SITS. Here, the cloud/shadow mask and the PCC mask have imposed to the algorithm to remove cloudy pixels and select high priority changed pixels, respectively. Finally, the abrupt change detection by means of BFAST methodology is implemented to detect the year in which the change has happened. A feature fusion step is considered before BFAST methodology to fuse all the available features by using a modified multifeature Change Vector Analysis (CVA) based method.
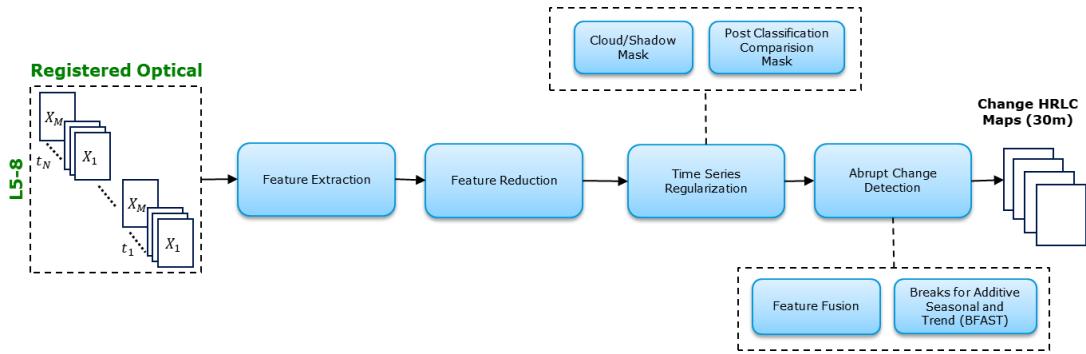
Figure 38. Detailed block-based representation of the multitemporal change detection processing chain.

The SITS carries information about the ground response in different spectrum. The extraction of the information from the multi-band SITS needs the employment of algorithms to learn the suitable feature space for the problem-solving. The main character that involves in the feature space analysis is the discrimination capability of different LC classes. As a result, the employment of just a single feature is not enough. Multiple features are required, and the identification of the most proper is an essential task.

9.4.1 Feature Extraction and Feature Reduction

The first stage is to determine a suitable Feature Space (FS), which is one of the most important factors in distinguishing the spectral trends of various sets of LC changes. The combination of the spectral bands acquired by the sensor provides suitable information to analyse the LC behaviour for SITS. Thus, all possible couples of the available sensor bands are considered to compute a set of Normalized Difference Indices ($NDI_f, f = (1, ..., F)$) of different bands as follow:

$$NDI_f = \frac{b_i - b_j}{b + b_j}, f = (1, ..., F) \tag{1}$$

This stage transforms the B-dimensional FS into a F-dimensional FS (equation (2)). where b_i and b_j belong to B, the set of bands available in a sensor, and i and $j \in [1, 2, ..., K]$.

$$F = \frac{1}{2}(K - 1) \times K \tag{2}$$

Depending on the number of bands (K), the number of resulting NDIs is equal to F. The obtained NDI indices are in the [-1,1] interval, they are more stable to noisy changes, they are able to highlight the interaction between singular bands and mitigate the undesirable oscillations of the spectral bands.

It has been shown in literature that combinations of the different spectral bands can result in redundant information [88] . Equation (1), when applied to numerous spectral bands of satellite images, yields a huge number of NDIs, the majority of which are uninformative since they are either unrelated or redundant to the class concept. Additionally, developing a CD methodology with a large number of features is computationally demanding. Principal Component Analysis (PCA) can be used to select the most relevant features based on the statistical association of different bands in order to execute band selection. However, PCA can only detect linear correlations between data features. Kernel Principal Component Analysis (KPCA) [89] expands PCA to include non-linear patterns in the data. As a result, a feature reduction method based on the KPCA is used to keep most informative and reliable features.

9.4.2 Time series reconstruction.

At this stage, the temporal signature is a raw signal characterized by non-equally distributed temporal sampling and non-continuous trend, also affected by noisy oscillation not corrected in the pre-processing step. The state-of-the-art mainly compares vegetative profiles between inner class temporal signatures. The behaviours are modelled, taking into account vegetation cycles and cycling harmonics models. The usage of those strategies

does not fit the case of multiple class trends and fails in the presence of abrupt changes or number of cycles different from the pre-established values. The development of an ad-hoc non-parametric strategy to reconstruct the vegetation-based temporal signature is needed. Moreover, the choice of the appropriate interpolation approach is critical in this case since it has a direct impact on the CD accuracy. To this end, preliminary symmetrical and consistently sampled NDI-SITS must be created, which is often accomplished by filling in the gaps with a linear combination of nearby values.

To produce reliable and continuous time series for the non-vegetation profiles the strategy is based on two steps: i) for each pixel in the image extract the NDI-SITS, ii) perform NDI data-SITS augmentation by upper envelope and dropout strategy (a piecewise cubic interpolation is used here) [90].

Further details on the augmentation by upper envelope strategy are illustrated as follows:

- Define a NDI-SITS set (NDI_{tr}), corresponding to a year (365 days), plus the two previous and two later months of data;
- For each NDI_{tr}, select the samples that are above a given threshold (defined by trial and error as NDI = 0.4). This threshold identifies when a given $SITS_p$ experiences a significant variability over time;
- Calculate the local maxima (as the points with zero first derivative and negative second derivative) of the selected samples and withdraw the remaining ones (from NDI_{tr}). This leads to the upper envelope of the data;
- Use the samples below the threshold and the local maxima from previous step for data imputation by means of a Piecewise Cubic Hermite Interpolating Polynomial (PCHIP). The selection of PCHIP over other interpolation methods is justified by its characteristic to preserve the shape of the data and respect monotonicity. The combination of these samples is defined as the upper-envelope set;
- Subtract the imputed data from NDI_{tr}. Reinsert the withdrawn samples with a difference greater than zero to the upper-envelope set. This step allows to better follow the shape of the original data;
- Impute the updated upper-envelope set by means of PCHIP;
- Remove the two previous and two later months from NDI_{tr}.

The definition of NDI_{tr} allows to better model the beginning and the end of the SITS, thus smoothing discontinuities and possible errors in LCCD analysis.

In the case of complex land cover classes like vegetation type (i.e., grass, shrubs, forest and crops) that show strong variabilities over space and time due to intrinsic seasonality and the large amount of species around the world, a third step is added that performs adaptive non-parametric regression of NDI-SITS by considering a General Regression Neural Network (GRNN) by taking inspiration from [91] [92](see Figure 39). The non-parametric regression is used and adapted to produce continuous and regularly sampled temporal signatures for vegetation pixels. To do so, four steps are followed: (1) Computation of Normalized Difference Indices (NDI), (2) uniform sampling interpolation, (3) low pass filtering and; (4) non-parametric regression through a Multi-Layer Perceptron Neural Network (MLP-NN). First, the spectral temporal signatures are combined, generating NDI arrays (FS). The combination of the source signals in the K bands produce an increased number of features. The NDI temporal signatures are then interpolated, considering the density and the shape maintenance requirement. A low pass filter reduces the intensity of high-frequency oscillations not usual in the LC temporal signatures, achieving a smoother behaviour. Last, a non-parametric regression captures the temporal signatures trend reducing the profile complexity and arithmetic dependency.
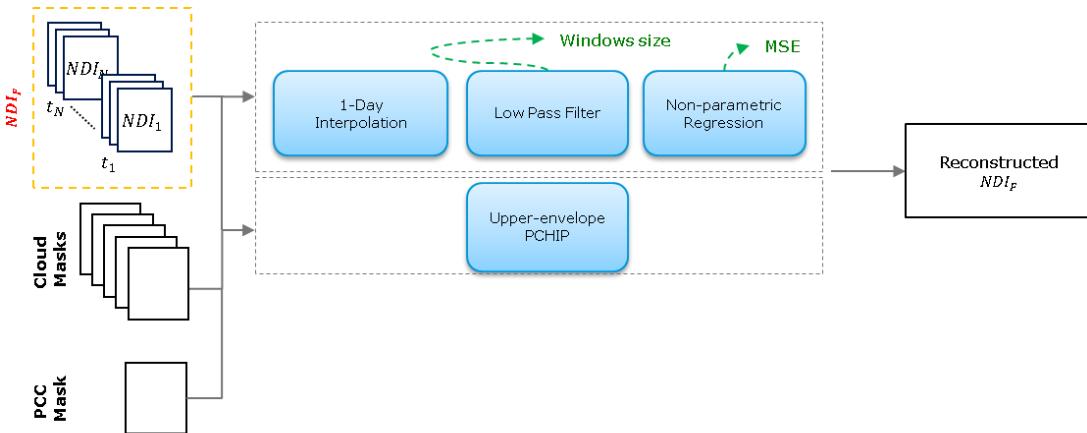
Figure 39. Block-based representation of the time series reconstruction.

The cloud/shadow mask for each image in SITS is imposed to exclude cloudy and cloud shadow pixels [93]. Moreover, the PCC map is considered to identify the pixels that have changed during the five years processing time. As a result, the high priority changed pixels have been selected to detect the year of change from them only and limit the computational burden. The low and high priority changes have been defined by climate users considering the climate LC transition matrix. These two filtering processes have a significant impact on improving the CD map reliability while reducing the overall processing chain computing load [AD5].

9.4.3 Abrupt Change Detection

A binary CD method is implemented to effectively discriminate changed and unchanged pixels. The approach is based on a generic CD approach for time series, involving the detection and characterization of Breaks For Additive Seasonal and Trend (BFAST) [87] . BFAST detects multiple abrupt changes in the seasonal and trend components of the time series and characterizes gradual and abrupt ecosystem change by deriving the time, magnitude, and direction of change within the trend component of the time series. However, BFAST was developed using MODIS data and has been tested mostly on NDVI and a few other vegetation indices, focusing on forest change detection. In this study the BFAST is employed for non-vegetated environments, with a set of features that accurately capture the properties of different classes. The pixel-wise abrupt CD based on BFAST imposes heavy computations to the system. Thus, in order to meliorate the computational burden, a feature fusion strategy is considered to fuse informative features derived from feature selection step. The process is based on feature magnitude calculation and is performed by considering data of two adjacent years in SITS. If data were not available for two adjacent years in the pixel level, the algorithm considers the next year to produce the feature magnitude and change information becomes biannual. Let $\left\{NDI_1^{m,1}, NDI_1^{m,2}, \dots, NDI_f^{m,n}\right\}$ and $\{NDI_1^{m+1,1}, NDI_1^{m+1,2}, \dots, NDI_f^{m+1,n}\}$ be the sets of NDIs of f features for the mth and $(m+1)$th year of SITS, respectively. Let n correspond to the total number of days for each year. $\{NDI_1^{(m,m+1),1}, NDI_1^{(m,m+1),2}, \dots, NDI_f^{(m,m+1),n}\}$ is computed by subtracting NDI_f^m and NDI_f^{m+1} for each feature. Finally, a hyper magnitude $NDI_F^{m,m+1}$ is calculated following the popular technique CVA [94]:

$$NDI_F^{m,m+1} = \sqrt{\sum_1^f (NDI_f^{m,m+1})^2} \tag{3}$$

The output of this process will be used as the input for BFAST. The final product is a four-channel image, one is related to the year in which a change has occurred, the second provides information on the probability of a certain change occurring, the third one shows how reliable is the reported year, considering the time span

between the years in which the feature magnitude is calculated and the last channel is the PCC map giving the information of the high and low priority pixels.

The analysis will be fully unsupervised and the CD accuracy is strongly reliant on the SITS fitting quality, as well as the characteristics and speed of change over time. The method allows the user to know if there has been some disturbance/change, but not the type of disturbance/change. Therefore, the user can derive the information of the LC of interest from the five year regional HRLC maps.
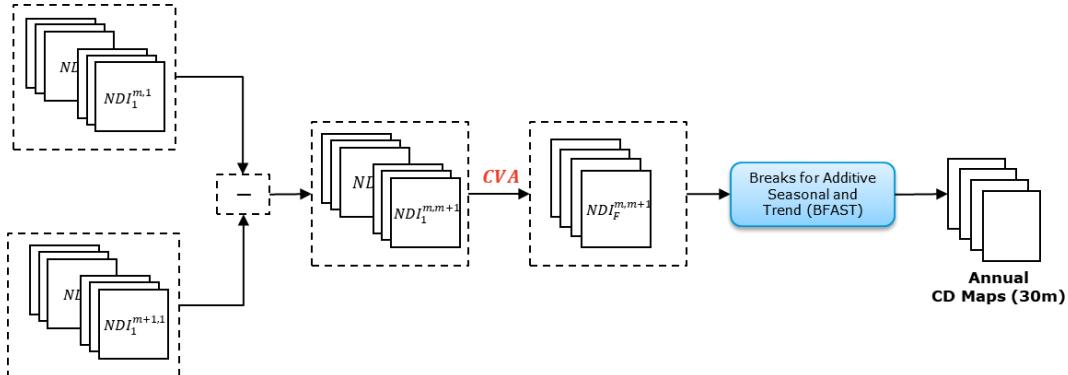

Figure 40. Block-based representation of the abrupt change detection.

9.5 A deep learning perspective

The processing chain for the multitemporal CD and trend analysis could be also analysed from a Deep Learning (DL) perspective (see Figure 41). In particular, some works [95]–[99] can be found nowadays in literature that deal with rather longer time scale changes or inter-annual changes. The main problem for deep learning approaches remains the lack of enough training samples to train the algorithms. This problem is even bigger when we talk about CD and long-time series. When training samples are available, the potential in terms of accuracy is quite remarkable. Some examples of works carried out in literature, rather in Landsat like data or long time series, are provided in the next in order to show the potential of DL. Inspiration could be taken from these works in order to be applied on the CCI HRLC with some extra work for training data collection.
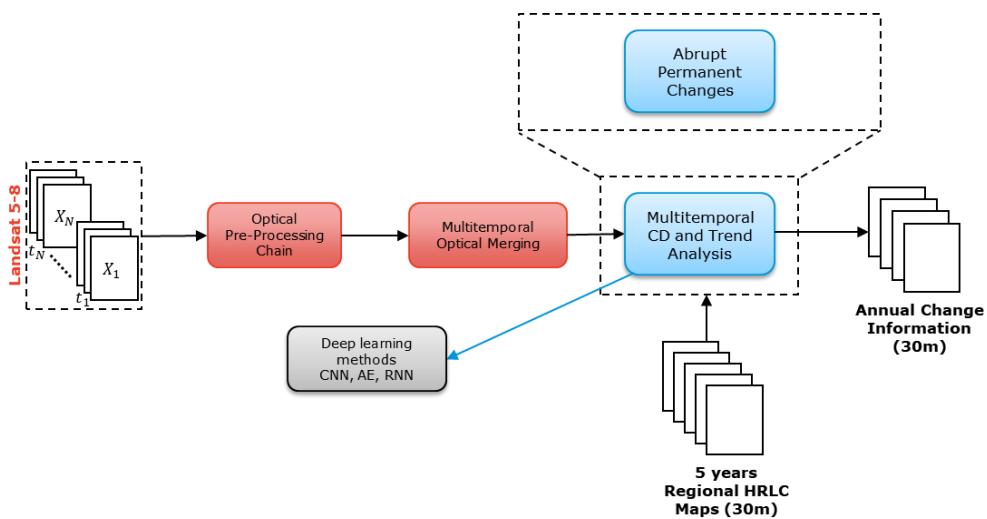

Figure 41. New deep learning block-based representation of the processing chain for the multitemporal change detection and trend analysis.

9.5.1 Learning a Transferable Change Rule from a Recurrent Neural Network (RNN) for Land Cover Change Detection (REFEREE).

The goal of this work is to design an efficient transferable change rule for binary and multi-class CDs. To do so, the method relies on an improved Long Short-Term Memory (LSTM) model, in a RNN learning framework, that acquires and records the change information of long-term sequences of remote sensing data. Experiments were carried out in three different datasets/cities (Taizhou, Kunshan and Yancheng in China), with different types of changes. The results of REFEREE were compared with non-deep learning approaches such as Change Vector Analysis (CVA), Principal Component Analysis (PCA), Iteratively-Reweighted Multivariate Alteration Detection (IRMAD) and Supervised Slow Feature Analysis (SSFA). The results, summarized in Table 8, show the high potential of REFEREE over standard methods with an increase of accuracy over 10-30% for the binary CD case and over 10-25% for the multiple CD case.

Table 8. Kappa coefficient and Overall Accuracy (OA) for the three datasets in (a) binary and (b) multiple change detection cases.

	TaiZhou		KunShan		Yancheng	
	KAPPA	OA	KAPPA	OA	KAPPA	OA
CVA	0.3755	0.6982	0.4011	0.7160	0.7907	0.8722
PCA	0.5413	0.7419	0.633	0.7741	0.8174	0.9025
IRMAD	0.7942	0.9133	0.87	0.9397	0.6973	0.8352
SSFA	0.8229	0.9454	0.9361	0.9763	0.9032	0.9516
REFEREE	**0.9477**	**0.9777**	**0.9573**	**0.9837**	**0.9563**	**0.9828**

(a)

		OA	Kappa	F-score				
				Unchanged	City (C)	Water (C)	Soil (C)	Farmland (C)
Taizhou	REFEREE	0.95	0.8689	0.9788	0.7887	0.8749	0.7524	/
	CNN	0.9235	0.8063	0.9675	0.6679	0.8721	0.5521	/
	SVM	0.8391	0.6758	0.8717	0.5203	0.8326	0.3558	/
	Decision tree	0.7113	0.5221	0.8701	0.6403	0.7496	0.3558	/
Kunshan	REFEREE	0.9587	0.8988	0.9432	0.9735	/	/	0.8750
	CNN	0.9336	0.8413	0.8844	0.9559	/	/	0.8491
	SVM	0.8024	0.6654	0.6830	0.8762	/	/	0.3743
	Decision tree	0.6979	0.4844	0.6642	0.7913	/	/	0.1542

(b)

9.5.2 Forest Change Detection in Incomplete Satellite Images with Deep Neural Network.

The goal of this work is to detect forest cover changes (deforestation and fire) over a period of 29 years (1987-2015). The study area is located in Australia and Landsat images are used. This is the closest example to what we will face in the CCI HRLC project, both in time span and data type. Given the well-known problem of incomplete and contaminated Landsat data, this approach includes the pre-processing steps as well, which are not addressed with deep learning approaches. The CD problem is addressed as a classification problem itself, where features are learnt using a deep neural network in a data-driven fashion. Based on these highly discriminative representations, it is possible to determine forest changes and predict their onset and offset timings. Results are compared to state-of-the-art approaches such Support Vector Machines (SVM), Random Forest (RF), Bag of visual Words (BoW) and Scale Invariant Feature Transform (SIFT). The proposed approach in this paper showed an improvement of about 16-24% for the forest changes (see Table 9) and a mean onset/offset prediction error of 4.9months (an error reduction of five months – see Table 9 and Figure 42).

Table 9. Example of classification/change detection and onset/offset detection. Accuracies are given in percentage, whereas the error units are months.

| Method	Accuracy	Avg. Recall	Onset Error (Mn)	Offset Error (Mn)
SIFT+l-SVM	68.1	57.3	8.7 ± 4.1	15.1 ± 7.5
SIFT+k-SVM	71.3	61.4	8.3 ± 4.1	14.9 ± 7.2
SIFT+RF	69.7	58.8	8.9 ± 4.3	15.9 ± 7.7
BoW+l-SVM	72.6	63.1	7.4 ± 3.6	13.5 ± 6.9
BoW+k-SVM	74.1	64.9	7.1 ± 3.4	12.6 ± 6.8
BoW + RF	71.7	64.0	7.4 ± 3.7	13.8 ± 7.1
This paper	**92.0**	**84.6**	$\mathbf{3.2 \pm 2.3}$	$\mathbf{5.5 \pm 5.5}$
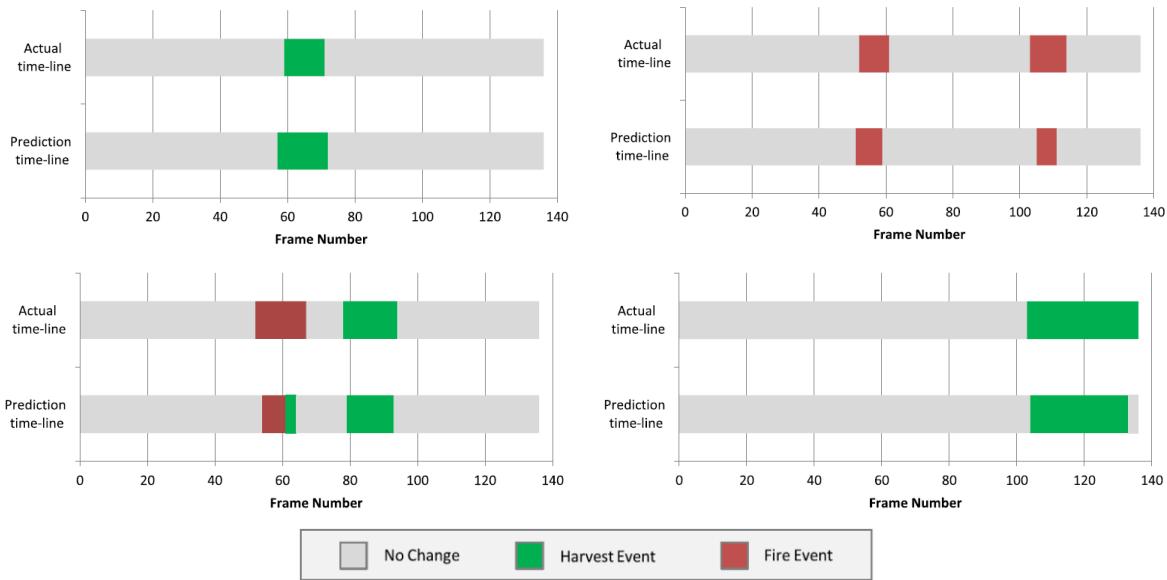
Figure 42. Sample result of the ground truth onset/offset events. In each plot, the top bar shows ground truth, and the bottom bar shows prediction from the proposed approach.

9.5.3 Long-Term Annual Mapping of Four Cities on Different Continents by Applying a Deep Information Learning Method to Landsat Data

The goal of [98] is to detect long-term urban changes by addressing temporal spectral variance and a scarcity of training samples in Landsat images from 1984-2016. Once again, we are in a similar situation to the CCI HRLC project. This time the focus is on urban changes, and not on vegetation LC like, which is indeed complementary to the paper presented in section 10.4.2. The method is applied to Landsat observations over urban areas in four cities in the temperate zone (Beijing, New York, Melbourne, and Munich). The method is trained using observations of Beijing collected in 1999, and then used to map urban areas in all target cities for the entire 1984–2016 period. The method uses two main steps: (1) use of RNN to minimize seasonal urban spectral variance; and (2) introduce an automated transfer strategy to maximize information gain from limited training samples when applied to new target cities in similar climate zones. The method is compared to other state-of-the-art methods (SVM, RF and RNN-LSTM), achieving comparable or even better accuracies (see Table 10). The overall accuracy of single-year urban maps is approximately 96±3% among the four target cities.

	Ref	CCI_HRLC_Ph1-ATBD		high resolution
esa	Issue	Date	Page	**land cover**
	4.rev.0	31/10/2022	73	cci

Table 10. Detection results from state-of-the-art methods and proposed method with OA and run-time.

		SVM-RBF (%)	RF (%)	RNN-LSTM (%)	Proposed Framework (%)
Temporal transfer	**Beijing**	68.63	71.38	76.25	81.87
	New York	69.13	72.75	80.63	82.08
Spatial transfer	**Melbourne**	71.25	67.63	85.88	84.75
	Munich	79.25	78.2	86.87	90.63
Run-Time (min)	**-**	7.53	0.37	0.78	0.82

Given computational burden of deep learning base methodologies, they have been let aside in the context of the project, but could be considered for further activities.

10 Post-processing

The post processing module has been developed to address the residual misregistration in the land-cover map time series and the residual errors of the change detection module. The post processing considers the spatial extent of the detected changes. Those changes that are too small to be realistic are removed and the land cover of the most recent year is used. The priority of the changes is taken into account in order to avoid removing small high priority changes that are still reasonable. Also, the classes involved in the changes are considered regardless of the priority.

The module works by iteratively computing the PCC map of the output land cover maps from the decision fusion modules, starting from 2015-2019 to 1990-1995. The small changes detected in the PCC are analysed in terms of best classes involved in the land cover transition and priority of the change. Generally, for high priority changes only isolated clusters of 1 to 5 changed pixels are considered in the post processing (the cluster size depends on both mapped region and involved classes), apart from cases where they are clearly errors due to residual misregistration. For low-priority changes, we generally consider slightly larger minimum allowed clusters, but it still depends on both the region and the involved classes. For example, changes related to road construction are always kept, even if they are associated to small changes. Concerning the second best-class map, it inherits the best-class label that was chosen by the fusion module for those pixels in which the change is removed. Finally, the output uncertainty measures mirror those coming from the optical, SAR, and fusion processing chains.

11 References

[1] R. Richter and D. Schlapfer, "Atmospheric / Topographic Correction for Satellite Imagery (ATCOR-2/3 UserGuide, Version 9.0.2, March 2016)."

[2] B. Mayer and A. Kylling, "The libRadtran software package for radiative transfer calculations - description and examples of use," *Atmos. Chem. Phys.*, vol. 5, pp. 1855–1877, 2005.

[3] J. G. Masek *et al.*, "A Landsat surface reflectance dataset for North America, 1990-2000," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 68–72, Jan. 2006, doi: 10.1109/LGRS.2005.857030.

[4] E. Vermote, C. Justice, M. Claverie, and B. Franch, "Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product," *Remote Sensing of Environment*, vol. 185, pp. 46–56, Nov. 2016, doi: 10.1016/j.rse.2016.04.008.

[5] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sensing of Environment*, vol. 118, pp. 83–94, Mar. 2012, doi: 10.1016/j.rse.2011.10.028.

[6] L. Baetens, C. Desjardins, and O. Hagolle, "Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure," *Remote Sensing*, vol. 11, no. 4, p. 433, Jan. 2019, doi: 10.3390/rs11040433.

[7] G. Mateo-García, L. Gómez-Chova, J. Amorós-López, J. Muñoz-Marí, and G. Camps-Valls, "Multitemporal Cloud Masking in the Google Earth Engine," *Remote Sensing*, vol. 10, no. 7, p. 1079, Jul. 2018, doi: 10.3390/rs10071079.

[8] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 144, pp. 235–253, Oct. 2018, doi: 10.1016/j.isprsjprs.2018.07.006.

[9] A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method," *Journal of Graphics Tools*, vol. 9, no. 1, pp. 23–34, Jan. 2004, doi: 10.1080/10867651.2004.10487596.

[10] Barstow, "Format Specification for ERS Products within ENVISAT Format".

[11] R. Barstow, "ENVISAT-1 Products Specifications Volume 8: ASAR Products Specifications," vol. 8, p. 179.

[12] J.-W. Park, A. Korosov, M. Babiker, S. Sandven, and J.-S. Won, "Efficient Thermal Noise Removal for Sentinel-1 TOPSAR Cross-Polarization Channel," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, pp. 1–11, Dec. 2017, doi: 10.1109/TGRS.2017.2765248.

[13] "Developer Guide - SNAP - SNAP Wiki." https://senbox.atlassian.net/wiki/spaces/SNAP/pages/8847381/Developer+Guide (accessed Dec. 10, 2019).

[14] F. Filipponi, "Sentinel-1 GRD Preprocessing Workflow," presented at the Proceedings, Jun. 2019, vol. 18, p. 6201. doi: 10.3390/ECRS-3-06201.

[15] C. Oliver and S. Quegan, *Understanding synthetic aperture radar images*. Boston: Artech House, 1998.

[16] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone, "A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 3, pp. 6–35, Sep. 2013, doi: 10.1109/MGRS.2013.2277512.

[17] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone, "A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 3, pp. 6–35, Sep. 2013, doi: 10.1109/MGRS.2013.2277512.

[18] J.-S. Lee, "Speckle analysis and smoothing of synthetic aperture radar images," *Computer Graphics and Image Processing*, vol. 17, no. 1, pp. 24–32, Sep. 1981, doi: 10.1016/S0146-664X(81)80005-6.

[19] P. Kupidura, "Comparison of Filters Dedicated to Speckle Suppression in SAR Images," *ISPAr*, vol. 41B7, pp. 269–276, Jun. 2016, doi: 10.5194/isprs-archives-XLI-B7-269-2016.

[20] F. Qiu, J. Berglund, J. R. Jensen, P. Thakkar, and D. Ren, "Speckle Noise Reduction in SAR Imagery Using a Local Adaptive Median Filter," *GIScience & Remote Sensing*, vol. 41, no. 3, pp. 244–266, Sep. 2004, doi: 10.2747/1548-1603.41.3.244.

[21] H. Cantalloube and C. Nahum, "How to Compute a Multi-Look SAR Image?," Jan. 2000.

[22] W. Zhao, C.-A. Deledalle, L. Denis, H. Maître, J.-M. Nicolas, and F. Tupin, "Ratio-Based Multitemporal SAR Images Denoising: RABASAR," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 6, pp. 3552–3565, Jun. 2019, doi: 10.1109/TGRS.2018.2885683.

[23] G. Quin, B. Pinel-Puysségur, J. Nicolas, and P. Loreaux, "MIMOSA: An Automatic Change Detection Method for SAR Time Series," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 9, pp. 5349–5363, Sep. 2014, doi: 10.1109/TGRS.2013.2288271.

[24] A. Salentinig and P. Gamba, "A General Framework for Urban Area Extraction Exploiting Multiresolution SAR Data Fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 2009–2018, May 2016, doi: 10.1109/JSTARS.2016.2546553.

[25] M. V. Wickerhauser, *Adapted Wavelet Analysis: From Theory to Software*. CRC Press, 1996.

[26] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, Jul. 1989, doi: 10.1109/34.192463.

[27] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4 edition. New York, NY: Pearson, 2017.

[28] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2207–2232, Sep. 1992, doi: 10.1109/78.157221.

[29] G. Simone, F. C. Morabito, and A. Farina, "Radar image fusion by multiscale Kalman filtering," in *Proceedings of the Third International Conference on Information Fusion*, Jul. 2000, vol. 2, p. WED3/10-WED3/17 vol.2. doi: 10.1109/IFIC.2000.859858.

[30] L. G. Brown and L. Gottesfeld, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, Dec. 1992, doi: 10.1145/146370.146374.

[31] A. A. Goshtasby, "Fusion of multi-exposure images," *Image and Vision Computing*, vol. 23, no. 6, pp. 611–618, Jun. 2005, doi: 10.1016/j.imavis.2005.02.004.

[32] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," in *Computer Vision – ECCV 2006*, Berlin, Heidelberg, 2006, pp. 404–417. doi: 10.1007/11744023_32.

[33] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[34] M. Donoser and H. Bischof, "Efficient Maximally Stable Extremal Region (MSER) Tracking," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Jun. 2006, vol. 1, pp. 553–560. doi: 10.1109/CVPR.2006.107.

[35] R. B. Ash, *Information Theory*. Courier Corporation, 1990.

[36] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, Eds., *Image Registration for Remote Sensing*. Cambridge: Cambridge University Press, 2011. doi: 10.1017/CBO9780511777684.

[37] A. A. Cole-Rhodes, K. L. Johnson, J. LeMoigne, and L. Zavorin, "Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1495–1510, 2003, doi: 10.1109/TIP.2003.819237.

[38] D. Solarna, A. Gotelli, J. Le Moigne, G. Moser, and S. B. Serpico, "Crater Detection and Registration of Planetary Images through Marked Point Processes, Multiscale Decomposition, and Region-Based Analysis," *TGRS*, Accepted up to minor modifications.

[39] J. Le Moigne, N. S. Netanyahu, and R. D. Eastman, Eds., *Image Registration for Remote Sensing*. Cambridge: Cambridge University Press, 2011. doi: 10.1017/CBO9780511777684.

[40] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, Oct. 2003, doi: 10.1016/S0262-8856(03)00137-9.

[41] D. G. Lowe, D. G. Lowe, and D. G. Lowe, "Object Recognition from Local Scale-Invariant Features," in *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*, Washington, DC, USA, 1999, pp. 1150–. Accessed: Dec. 11, 2019. [Online]. Available: http://dl.acm.org/citation.cfm?id=850924.851523

[42] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972, doi: 10.1145/361237.361242.

[43] X. Descombes, R. Minlos, and E. Zhizhina, "Object Extraction Using a Stochastic Birth-and-Death Dynamics in Continuum," *Journal of Mathematical Imaging and Vision*, vol. 33, no. 3, pp. 347–359, Mar. 2009, doi: 10.1007/s10851-008-0117-y.

[44] A. B. Carlson and P. Crilly, *Communication Systems*, 5 edizione. Boston: McGraw-Hill Education, 2009.

[45] E. Parzen, "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[46] J. M. Murphy, J. Le Moigne, and D. J. Harding, "Automatic Image Registration of Multimodal Remotely Sensed Data With Global Shearlet Features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1685–1704, Mar. 2016, doi: 10.1109/TGRS.2015.2487457.

[47] M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *Comput J*, vol. 7, no. 2, pp. 155–162, Jan. 1964, doi: 10.1093/comjnl/7.2.155.

[48] R. P. Brent, "An algorithm with guaranteed convergence for finding a zero of a function," *Comput J*, vol. 14, no. 4, pp. 422–425, Jan. 1971, doi: 10.1093/comjnl/14.4.422.

[49] M. J. D. Powell, "A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation," in *Advances in Optimization and Numerical Analysis*, S. Gomez and J.-P. Hennart, Eds. Dordrecht: Springer Netherlands, 1994, pp. 51–67. doi: 10.1007/978-94-015-8330-5_4.

[50] L. Maggiolo, D. Solarna, G. Moser, and S. B. Serpico, "Registration of Multisensor Images through a Conditional Generative Adversarial Network and a Correlation-Type Similarity Measure," *Remote Sensing 2022, Vol. 14, Page 2811*, vol. 14, no. 12, p. 2811, Jun. 2022, doi: 10.3390/RS14122811.

[51] D. Solarna, L. Maggiolo, G. Moser, and S. B. Serpico, "A Tiling-Based Strategy for Large-Scale Multisensor Optical-Sar Image Registration," pp. 127–130, Sep. 2022, doi: 10.1109/IGARSS46834.2022.9884048.

[52] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery," *Remote Sensing*, vol. 7, no. 11, pp. 14680–14707, Nov. 2015, doi: 10.3390/rs71114680.

[53] S. Fukuda and H. Hirosawa, "Support vector machine classification of land cover: application to polarimetric SAR data," in *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217)*, Jul. 2001, vol. 1, pp. 187–189 vol.1. doi: 10.1109/IGARSS.2001.976097.

[54] R. S. Hosseini, I. Entezari, S. Homayouni, M. Motagh, and B. Mansouri, "Classification of polarimetric SAR images using Support Vector Machines," *Canadian Journal of Remote Sensing*, vol. 37, no. 2, pp. 220–233, Nov. 2011, doi: 10.5589/m11-029.

[55] P. Mantero, G. Moser, and S. B. Serpico, "Partially Supervised classification of remote sensing images through SVM-based probability density estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 559–570, Mar. 2005, doi: 10.1109/TGRS.2004.842022.

[56] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.

[57] S. Abdikan, F. B. Sanli, M. Ustuner, and F. Calò, "Land Cover Mapping Using SENTINEL-1 SAR Data," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 41B7, pp. 757–761, Jun. 2016, doi: 10.5194/isprs-archives-XLI-B7-757-2016.

[58] S. Niculescu, H. Talab Ou Ali, and A. Billey, "Random forest classification using Sentinel-1 and Sentinel-2 series for vegetation monitoring in the Pays de Brest (France)," in *Remote Sensing for Agriculture, Ecosystems, and Hydrology XX*, Berlin, Germany, Oct. 2018, p. 6. doi: 10.1117/12.2325546.

[59] "ERDAS Field Guide," Mar. 03, 2016. https://community.hexagongeospatial.com/t5/IMAGINE-Q-A/ERDAS-Field-Guide/ta-p/3179 (accessed Dec. 10, 2019).

[60] W. K. Pratt, *Digital Image Processing: PIKS Scientific Inside*, 4 edition. Hoboken, N.J: Wiley-Interscience, 2007.

[61] J. S. Lee, L. Jurkevich, P. Dewaele, P. Wambacq, and A. Oosterlinck, "Speckle filtering of synthetic aperture radar images: A review," *Remote Sensing Reviews*, vol. 8, no. 4, pp. 313–340, Feb. 1994, doi: 10.1080/02757259409532206.

[62] J. S. Lee and E. Pottier, *Polarimetric Radar Imaging: from Basics to Applications*, 1st ed. FL, USA: CRC Press: Boca Raton, 2009.

[63] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013, doi: 10.1109/MGRS.2013.2248301.

[64] N. Yokoya, "Texture-Guided multisensor superresolution for remotely sensed images," *Remote Sens.*, vol. 9, p. 316, 2017.

[65] H. Jingliang, P. Gamisi, and X. Zhu, "Feature extraction and selection of sentinel-1 dual-pol data for global-scale local climate zone classification," *ISPRS International Journal of Geo-Information*, vol. 7, no. 9, p. 379, 2018.

[66] A. Braun and V. Hochschild, "Combined use of SAR and optical data for environmental assessments around refugee camps in semiarid landscapes," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 7, pp. 777–782, Apr. 2015, doi: 10.5194/isprsarchives-XL-7-W3-777-2015.

[67] P. Du, A. Samat, B. Waske, S. Liu, and Z. Li, "Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 38–53, Jul. 2015, doi: 10.1016/j.isprsjprs.2015.03.002.

[68] M. Wurm, H. Taubenböck, M. Weigand, and A. Schmitt, "Slum mapping in polarimetric SAR data using spatial features," *Remote Sensing of Environment*, vol. 194, pp. 190–204, Jun. 2017, doi: 10.1016/j.rse.2017.03.030.

[69] F. N. Numbisi, F. V. Coillie, and R. D. Wulf, "MULTI-DATE SENTINEL1 SAR IMAGE TEXTURES DISCRIMINATE PERENNIAL AGROFORESTS IN A TROPICAL FOREST-SAVANNAH TRANSITION LANDSCAPE," in *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Sep. 2018, vol. XLII–1, pp. 339–346. doi: https://doi.org/10.5194/isprs-archives-XLII-1-339-2018.

[70] O. Cartus, M. Santoro, C. Schmullius, P. Y. Yong, C. Er-xue, and L. Zeng-yuan, "CREATION OF LARGE AREA FOREST BIOMASS MAPS FOR NORTHEAST CHINA USING ERS-1 / 2 TANDEM COHERENCE," 2007.

[71] N.-W. Park and K.-H. Chi, "Integration of multitemporal/polarization C-band SAR data sets for land-cover classification," *International Journal of Remote Sensing*, vol. 29, no. 16, pp. 4667–4688, Aug. 2008, doi: 10.1080/01431160801947341.

[72] D. Marzi and P. Gamba, "Inland Water Body Mapping Using Multi-temporal Sentinel-1 SAR Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. PP, pp. 1–1, Nov. 2021, doi: 10.1109/JSTARS.2021.3127748.

[73] "Land Cover Classification System - Classification concepts and user manual." http://www.fao.org/3/y7220e/y7220e00.htm (accessed Dec. 10, 2019).

[74] T. Kobayashi *et al.*, "Production of Global Land Cover Data – GLCNMO2013," *Journal of Geography and Geology*, vol. 9, no. 3, p. p1, Jun. 2017, doi: 10.5539/jgg.v9n3p1.

[75] R. Tateishi *et al.*, "Production of global land cover data – GLCNMO," *International Journal of Digital Earth*, vol. 4, no. 1, pp. 22–49, Jan. 2011, doi: 10.1080/17538941003777521.

[76] R. Tateishi, N. Hoan, T. Kobayashi, B. Alsaaideh, G. Tana, and D. Phong, "Production of Global Land Cover Data – GLCNMO2008," *Journal of Geography and Geology*, vol. 6, no. 3, p. p99, Jul. 2014, doi: 10.5539/jgg.v6n3p99.

[77] G. Lisini, A. Salentinig, P. Du, and P. Gamba, "SAR-Based Urban Extents Extraction: From ENVISAT to Sentinel-1," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 8, pp. 2683–2691, Aug. 2018, doi: 10.1109/JSTARS.2017.2782180.

[78] J. A. Benediktsson, "Hybrid consensus theoretic classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 833–843, Jul. 1997, doi: 10.1109/36.602526.

[79] J. A. Benediktsson and P. H. Swain, "Consensus theoretic classification methods," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, no. 4, pp. 688–704, 1992, doi: 10.1109/21.156582.

[80] N. C. Dalkey, "An Impossibility Theorem for Group Probability Functions.," 1972.

[81] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1367–1377, May 1999, doi: 10.1109/36.763301.

[82] S. Z. Li, *Markov random field modeling in image analysis*. Springer, 2009.

[83] Z. Kato and J. Zerubia, "Markov Random Fields in Image Segmentation," *Foundations and Trends in Signal Processing*, vol. 5, no. 1–2, pp. 1–155, 2012, doi: 10.1561/2000000035.

[84] L. Maggiolo, D. Solarna, G. Moser, and S. B. Serpico, "Optical-SAR Decision Fusion with Markov Random Fields for High-Resolution Large-Scale Land Cover Mapping," pp. 5508–5511, Sep. 2022, doi: 10.1109/IGARSS46834.2022.9884751.

[85] Swain, "Bayesian Classification in a Time-Varying Environment".

[86] J. Verbesselt, R. Hyndman, G. Newnham, and D. Culvenor, "Detecting trend and seasonal changes in satellite image time series," *Remote Sensing of Environment*, vol. 114, pp. 106–115, 2010.

[87] J. Verbesselt, R. Hyndman, G. Newnham, e D. Culvenor, "Detecting trend and seasonal changes in satellite image time series," *Remote Sensing of Environment*, vol. 114, pp. 106–115, 2010, doi: 10.1016/j.rse.2009.08.014.

[88] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," *Artificial Intelligence Review*, vol. 53, pp. 907–948, 2020.

[89] L. Zhang, H. Su, and J. Shen, "Hyperspectral Dimensionality Reduction Based on Multiscale Superpixelwise Kernel Principal Component Analysis," *Remote Sensing 11*, vol. 10, p. 1219, 2019.

[90] Y. T. Solano-Correa, K. Meshkini, F. Bovolo, and L. Bruzzone, "A land cover-driven approach for fitting satellite image time series in a change detection context," *Proc. SPIE 11533, Image and Signal Processing for Remote Sensing XXVI*, 2020.

[91] Y. T. Solano-Correa, F. Bovolo, L. Bruzzone, and D. Fernández-Prieto, "A Method for the Analysis of Small Crop Fields in Sentinel-2 Dense Time Series," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 2150–2164, Mar. 2020, doi: 10.1109/TGRS.2019.2953652.

[92] Y. T. Solano-Correa, F. Bovolo, L. Bruzzone, and D. Fernández-Prieto, "Automatic Derivation of Cropland Phenological Parameters by Adaptive Non-Parametric Regression of Sentinel-2 NDVI Time Series," in *IGARSS 2018*, Jul. 2018, pp. 1946–1949. doi: 10.1109/IGARSS.2018.8519264.

[93] Z. Zhu, S. Wang, and C. E. Woodcock, "Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images," *Remote Sensing of Environment*, vol. 159, pp. 269–277, Mar. 2015, doi: 10.1016/j.rse.2014.12.014.

[94] F. Bovolo and L. Bruzzone, "A Theoretical Framework for Unsupervised Change Detection Based on Change Vector Analysis in the Polar Domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, pp. 218–236, 2007.

[95] S. Saha, F. Bovolo, L. Bruzzone, "Unsupervised Deep Change Vector Analysis for Multiple-Change Detection in VHR Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 3677–3693, 2019, doi: 10.1109/TGRS.2018.2886643.

[96] S. H. Khan, X. He, F. Porikli, M. Bennamoun, "Forest Change Detection in Incomplete Satellite Images With Deep Neural Networks," vol. 55, pp. 5407–5423, 2017, doi: 10.1109/TGRS.2017.2707528.

[97] H. Lyu, H. Lu, L. Mou, "Learning a Transferable Change Rule from a Recurrent Neural Network for Land Cover Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 5407–5423, 2017, doi: 10.1109/TGRS.2017.2707528.

[98] H. Lyu et al, "Long-Term Annual Mapping of Four Cities on Different Continents by Applying a Deep Information Learning Method to Landsat Data," *Remote Sensing*, vol. 10, p. 471, 2018, doi: 10.3390/rs10030471.

[99] C. Pelletier, G. I. Webb, F. Petitjean, "Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series," vol. 11, p. 523, 2019, doi: 10.3390/rs11050523.