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

Product Validation and Algorithm Selection Report
(PVASR)

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



UNIVERSITÀ
DEGLI STUDI
DI TRENTO



UCLouvain



CREAF

e-geos
AN ASI / TELESPAZIO COMPANY

planetek
italia

GeoVille

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	1	

Changelog

Issue	Changes	Date
1.0	First version, describing activities of first year.	29/10/2019
2.0	Updated with new Section 4, describing activities of second year.	23/11/2020
2.1	Updated version according to CCI_HRLC_Ph1_AR2_RID-ESA.xlsx.	04/12/2020
3.0	Updated with new Section 5, describing activities of third year.	02/11/2022
3.1	Updated with Section 5.4 reporting on quantitative analysis of change percentage. Updated with Section 5.2 on slum detection assessment in the HRLC10 map.	13/01/2023

Detailed Change Record

Issue	RID	Description of discrepancy	Sections	Change
2.1	FR-01	In the table reporting the new legend is not present the 'No Data' class (0 value).	General	Updated Figure 53.
	FR-02	Check the tables and figures number in the text and update them. Several of them are wrong.	4	Fixed references to many figures and tables in Section 4.1
	FR-03	Check the text and update the reference source when message " Error! Reference source not found. " appears.	4	Fixed for all instances of the error.
	FR-04	No Figures number and caption are present. Please add them.	4.3 onwards	Added numbering and captions figures.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	2	

Contents

1	Introduction.....	4
1.1	Executive summary.....	4
1.2	Purpose and scope	4
1.3	Applicable documents	4
1.4	Reference documents.....	4
1.5	Acronyms and abbreviations	4
2	Selection procedure.....	6
2.1	Criteria	6
2.2	Evaluation	7
3	Classification algorithms and procedures (year 1).....	8
3.1	Optical data processing	8
3.1.1	Satellite images	9
3.1.2	Method/algorithm/technique.....	9
3.1.3	Qualitative evaluation	12
3.1.4	Final decision.....	17
3.2	SAR data processing	18
3.2.1	Satellite images	18
3.2.2	Method/algorithm/technique.....	18
3.2.3	Qualitative evaluation	25
3.2.4	Final decision.....	27
3.3	Multi-sensor Optical and SAR Data Fusion	27
3.3.1	Multisensor Geolocation Methods	27
3.3.2	Quantitative evaluation	29
3.3.3	Final decision.....	42
3.3.4	Decision Fusion Methods	42
3.3.5	Qualitative evaluation	45
3.3.6	Final decision.....	48
4	Classification algorithms and procedures (year 2).....	49
4.1	Optical data processing	49
4.1.1	Satellite images	49
4.1.2	Method/algorithm/technique.....	51
4.1.3	Qualitative evaluation	52
4.1.4	Final decision.....	58
4.2	SAR data processing	58
4.2.1	Satellite images	58
4.2.2	Method/algorithm/technique.....	59
4.2.3	Qualitative evaluation	66
4.2.4	Final decision.....	70

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	3	

4.3	Multi-Sensor Optical-SAR Data Fusion	71
4.3.1	Multisensor Geolocation Methods	71
4.3.2	Automatic tiling for large-scale registration.....	74
4.3.3	Quantitative evaluation using COBYLA and the unconstrained Powell's algorithm	75
4.3.4	Quantitative evaluation of Powell's algorithm with barrier functions and automatic tiling	89
4.3.5	Final decision.....	96
4.3.6	Decision Fusion Methods	97
4.3.7	First experimental analysis on the Amazon round robin area	100
4.3.8	Experimental analysis on all round robin areas	105
4.3.9	Final decision.....	117
5	Algorithms and procedures (year 3).....	117
5.1	Optical data processing	117
5.1.1	Satellite images	117
5.1.2	Method/algorithm/technique.....	119
5.1.3	Qualitative evaluation	119
5.1.4	Final decision.....	124
5.2	SAR data processing	124
5.2.1	Satellite images	124
5.2.2	Method/algorithm/technique.....	124
5.2.3	Qualitative evaluation	125
5.2.4	Final decision.....	148
5.3	Multi-Sensor Optical-SAR Data Fusion and Decision Fusion	148
5.3.1	Experimental analysis on the static and historical area	148
5.3.2	Results of the spatial harmonization.....	173
5.3.3	Results of the temporal harmonization	174
5.4	Change Detection	177
5.4.1	Change Detection data processing.....	177
5.5	Post processing.....	188
5.5.1	Results of post processing	188
6	References.....	189

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	4	

1 Introduction

1.1 Executive summary

For the success of the project, key is a well-driven and structured selection of best performers among candidate algorithms for some blocks of the whole processing chain aimed at generating the HR LC products. With the end of the second year of activities, several comparative tests and performance analysis tasks have been carried out by the EOS team. The outcome of these comparative activities is presented and indications on best performing algorithms/techniques are provided in Sec. 4. In the period, the final version of the processing chain has been executed and we document its performance improvements for both static and historical LC and LCC products.

1.2 Purpose and scope

The Product Validation and Algorithm Selection Report (PVASR) provides detailed information about the comparative tasks performed for assessing best performing algorithms and techniques to be included in the classification blocks within the overall processing chain.

The PVASR document is living, being updated at every project cycle (on annual basis) based on the output of the round-robins and internal benchmarking activities. In its current version, activities carried on in the second cycle are presented. Comparison and benchmarking activities have been devoted towards classification of optical and SAR imagery and decision fusion, giving emphasis especially on:

1. Testing classifiers and evaluating their performance in terms of accuracy, computational efficiency and predisposition to model/code modification to meet requirements and implementation needs.
2. Testing approaches for building reliable training datasets out of already existing products, being them sub-optimal in terms of spatial resolution (coarse to medium) and legend detail (incomplete if compared to HR LC products legend as detailed in ATBD).
3. Evaluating sets of multitemporal features provided as input for the classifiers.
4. Testing multisensory decision fusion methods against their accuracy and computational time.

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] Data Standards Requirements for CCI Data Producers, v2.0, 17/09/2018, CCI-PRGM-EOPS-TN-13-0009
- [AD4] CCI_HRLC_Ph1-D1.1_URD, latest version
- [AD5] CCI_HRLC_Ph1-D2.2_ATBD, latest version
- [AD6] CCI_HRLC_Ph1-D4.1_PVIR, latest version
- [AD7] CCI_HRLC_Ph1-D2.4_ADP, latest version

1.4 Reference documents

Ref. Title, Issue/Rev, Date, ID

- [RD1] The Global Climate Observing System: Implementation Needs, 01/10/2016, GCOS-200

1.5 Acronyms and abbreviations

ANN	Artificial Neural Network
ATBD	Algorithm Theoretical Basis Document
CC	Cross Correlation

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	5	

CCI	Climate Change Initiative
COBYLA	Constrained Optimization BY Linear Approximation
ESA	European Space Agency
FFT	Fast Fourier Transform
GCOS	Global Climate Observing System
GLC	Global Land Cover
GLCNMO	Global Land Cover by National Mapping Organizations
GRD	Ground Range Detected
GT	Ground Truth
HR	High Resolution
HPF	High Pass Filter
ICM	Iterated Conditional Mode
IWS	Interferometric Wide Swath
ML	Maximum Likelihood
LC	Land Cover
LCCS	Land Cover Classification System
LOGP	Logarithmic Opinion Pool
LOP	Linear Opinion Pool
LPF	Low Pass Filter
MI	Mutual Information
MMSE	Minimum Mean Square Error
MRF	Markov Random Fields
PDF	Probability Density Function
RBF	Radial Basis Function
RMSE	Root Mean Square Error
RF	Random Forest
S1	Sentinel 1
S2	Sentinel 2
SAR	Synthetic Aperture Radar
SNAP	Sentinel's Application Platform
SFFS	Sequential Forward Floating Selection
SoW	Statement of Work
SVM	Support Vector Machine
TS	Time Series
URD	User Requirements Document

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	6	

2 Selection procedure

The overall procedure for the selection of best performing algorithms and methods is performed according to a three-step procedure. The algorithms presented in the Technical Proposal [AD1] and ATBD [AD5] are considered for the comparisons together with a set of proposed solutions for each task such as generating training samples and building multitemporal features. The evaluation-selection procedure is devised in such a way that the selected algorithms/techniques are the most suitable to satisfy project requirements.

The three steps of the procedure are the following:

- **Step 1: Qualitative pre-screening of algorithms**

A pre-screening of the algorithms and methods from a State-of-the-art pool of competitors is carried out in order to identify the most relevant methodologies with respect to the project objectives. This preliminary analysis is driven by the selection criteria described in Section 2.1. In this first step, a high-level **qualitative evaluation** of these criteria is conducted in order to identify techniques that clearly cannot reach a satisfactory ranking on several categories of parameters. These techniques are discarded and not considered in the next steps. Algorithms and methods that passed the pre-screening are reported in the Technical Proposal [AD1] and more in detail in the ATBD [AD5]. In this report only the methods that passed the pre-screening are considered explicitly.

- **Step 2: Quantitative evaluation of algorithms**

Algorithms that pass the pre-screening in step 1 are analyzed in greater detail with a **quantitative evaluation**. This analysis is based on different parameters, ranging from a scientific and technical analysis to possible impacts on the application and users. For each investigated item (algorithm, method, technique, etc.) details on the quantitative evaluation of the comparison activities can be found in a dedicated section of this document.

- **Step 3: Final decision**

According to the analysis carried out for each individual comparison task, a **final decision** is taken according to the best performer and its relevance with respect to project objectives. Final decision is reported.

It is worth noting that the pre-processing algorithms are not included in the evaluation and ranking procedure because we expect to import in the project basic pre-processing chains already developed for both multispectral and SAR data.

2.1 Criteria

In this section the criteria adopted for evaluating the relevance of methods and algorithms with respect to project requirements are listed. Up to seven categories of parameters are considered divided in different issues.

1. **Scientific Background and Technical Soundness** – The scientific validity of the algorithms and of the methodologies on which the algorithms are based is considered as an important parameter. The rationale is that selected algorithms should be based on a solid theoretical background that guarantees the accuracy of its results also at an operational level. The guidelines for rating are as follows:
 - The methodology is solid;
 - The methodology is technical convincing;
 - The methodology is at the state-of-the-art;
 - The methodology is published in high quality journals;
 - The methodology is included in several other scientific publications or project technical reports.
2. **Robustness and Generality** – In order to obtain a reasonable estimation for the robustness and generality of the investigated algorithms, different parameters are considered, such as:
 - The method is suitable to be used with different kinds of images (e.g., Sentinel-2, Landsat, SAR, etc.);

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	7	

- The method shows high performance on different images (Sentinel, Landsat, etc.) and over the three test areas as described in URD [AD4];
 - There are software implementations or examples for the implementation available;
 - The algorithm can be used in combination with other methodologies.
3. **Novelty** – An appropriate candidate algorithm should have been published or reported for the first time relatively recently in the literature. It is not required that algorithms are completely innovative; the novelty may consist in both combining well established methodologies or applying well-known techniques in a novel way. As a main guideline, a tested method should be already applied in literature to solve existing problems.
4. **Operational Requirements** – The expected operational requirements (in terms of computational complexity, time effort, cost, etc.) for the final implementation of an algorithm/technique are evaluated. Although no actual constraints are fixed on the algorithm computational complexity, the most optimized implementations available in literature are preferred. Other crucial aspects are:
- The algorithm is prone to architectural modifications;
 - The processing time scaling is likely to be linear with image size;
 - The hardware and disk-storage requirements are appropriate.
- Algorithm/method consistency with project requirements is also extremely relevant, following guidelines from GCOS [RD1] and SoW [AD2]:
- Algorithms and methodologies must be effective for high resolution images (e.g., optical data at 10-30m).
 - Documented accuracy must be within the boundaries imposed by GCOS (see [RD1]) and as reported in SoW [AD2].
5. **Accuracy** – An algorithm is positively evaluated if able to provide a high absolute accuracy in all test areas, especially keeping into account the different climatological conditions and possibly different data availability conditions. Accordingly, the following guidelines are used for evaluating accuracy characteristics:
- Accuracy/uncertainty to be in line with GCOS [RD1] requirements as reported in SoW [AD2].
 - The algorithm matches the end-user (climatologist and other users from the community) requirements;
 - For unsupervised tasks the accuracy should not depend on the availability/quality of prior information.
 - For supervised tasks the accuracy should be robust to the availability/quality of prior information.
6. **Level of Automation** – From an operational point of view, it is mandatory that an algorithm runs in a completely automatic way. Algorithms requiring any amount of manual work, strong interaction with the final users are negatively evaluated.
7. **Specific End-users Requirements** – From an operational point of view, capability of an algorithm to satisfy and meet possible end-user requirements is another important parameter of evaluation. The main guidelines for driving this ranking are:
- The algorithm is robust to the use in several climatological regions;
 - The algorithm can be reasonably included in an operational procedure.

2.2 Evaluation

The evaluation procedure of each comparative task aimed at deciding on a specific algorithm/technique is carried out by considering all criteria listed before. To each reported activity, a thorough discussion is given regarding how these criteria are weighted in the overall evaluation, which aspects are given strong emphasis and which ones are considered less relevant. The evaluation activity provides answers about best performing algorithms/techniques that are included in the processing chain of the current version of HR LC products.

3 Classification algorithms and procedures (year 1)

Global climate change as well as the protection and management of natural sources have become central topics for many scientific initiatives in the Earth surface dynamics. Many works investigated the effectiveness of optical and radar data for both local and global scale thematic characterization for Land Cover (LC) analyses [1], [2]. The SAR application in remote sensing has been investigated in several studies and a substantial potential for LC monitoring has been proven [3], [4]. Synthetic Aperture Radar (SAR) data are utilized especially when weather conditions are not suitable for acquiring optical data, because their quality does not depend on weather conditions. Contrary to optical satellites, SAR makes it possible to continually collect data despite of light and weather conditions, providing "cloud-free" images because the cloud-penetrating capability of C-band signal [5]. On the other hand, multispectral optical images provide wide spectral information (ranging from visible to infrared wavelengths) from which detailed information of land properties can be retrieved. Moreover, the inclusion of optical time series provides detailed information about LC dynamic. Time series of multispectral images have proven capacity to characterize trends and environmental phenomena and are widely used for LC classification [6], [7]. Within this section a processing chain for the fully production of LC maps using High Resolution (HR) optical and SAR images is presented.

3.1 Optical data processing

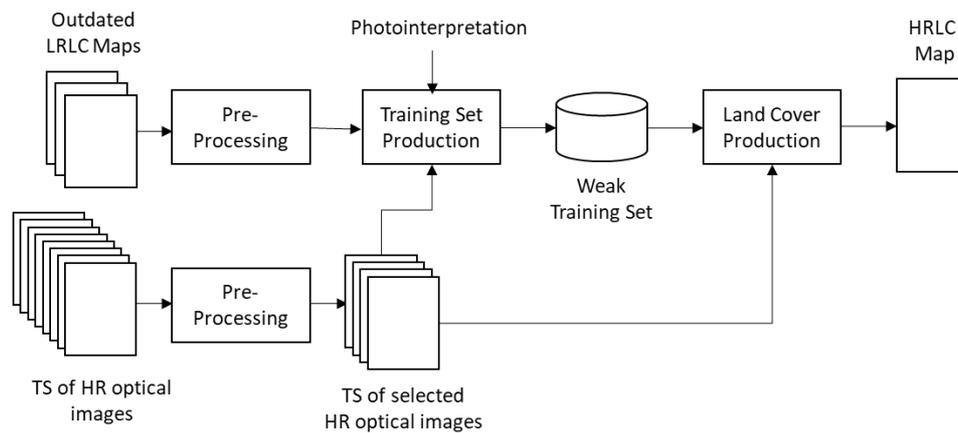


Figure 1. Optical data processing chain for the prototype production of the HR LC map obtained by classifying the time series of Sentinel 2 data.

Figure 1 depicts the optical data processing chain for the prototype production of the HR LC map obtained by classifying Sentinel 2 time series. The images are first pre-processed in order to perform the atmospheric correction and detect the clouds. Then, the best time series of images used to generate the HR static LC map is automatically detected. Due to the missed availability of training data, a training set production step is performed to extract the labeled data necessary to train the supervised classification system. Existing thematic product available on the considered study area are used to create database of weak training samples. The thematic products available are characterized by medium/coarse spatial resolution (e.g., 100m, 300m and 1 km), much coarser than the desired geometrical detail (10 m). The maps are analysed and processed in an unsupervised way to detect and extract the most reliable samples which are included in the weak training set. Moreover, few samples are added by photointerpretation to integrate the missing information on classes which require HR labelled pixels (e.g., building or roads).

Finally, the classification of the time series of Sentinel 2 images is performed to produce the HR LC map for the considered study areas.

3.1.1 Satellite images

The considered study areas are located into the following four thematic regions (according to tiling grid of Sentinel-2 products):

- Amazonia area for 21KUQ Sentinel-2 tile;
- Amazonia area for 21KXT Sentinel-2 tile;
- Siberia area for 42WXS Sentinel 2 tile;
- Africa area for 37PCP Sentinel-2 tile.

2018 Sentinel-2 images downloaded through the Copernicus Open Access Hub (<https://scihub.copernicus.eu/dhus/#/home>) are used over the four CCI-HR LC areas for the year 2018. The complete list of images is given in Table 1.

Table 1. List of Sentinel-2 data

Area	Satellite	# Products	Date list (2018y)
Amazonia – 21KUQ	S2	8	2018-04-17; 2018-08-10; 2018-05-12; 2018-08-30; 2018-05-22; 2018-10-09; 2018-07-16; 2018-12-08;
Amazonia – 21KXT	S2	9	2018-02-23; 2018-06-23; 2018-09-11 2018-04-29; 2018-07-18; 2018-05-09; 2018-08-12; 2018-05-29; 2018-08-27;
Siberia – 42WXS	S2	7	2018-06-06; 2018-08-27; 2018-07-06; 2018-09-01; 2018-07-21; 2018-11-08 2018-08-22;
Africa – 37PCP	S2	9	2018-02-06; 2018-09-24; 2018-12-23 2018-02-16; 2018-10-24; 2018-03-03; 2018-11-13; 2018-04-12; 2018-12-18;

3.1.2 Method/algorithm/technique

Within this Section, several methods are presented and compared as candidate approaches to be develop and implemented in the optical image processing chain according to Figure 2.

3.1.2.1 Optical data pre-processing

This step has the purpose of generating a time series of images able to characterize the HR LC classes. First, the atmospheric correction is performed by using the specific tools provided by ESA, the Sen2Cor processor for Level 2A product generation in the Sentinel-2 toolbox [8]. The main goal is to exploit the high revisit time of Sentinel 2 data to select almost cloud free images available on the considered study area. This condition allows for an accurate temporal characterization of the HR LC classes. By analysing the dense time series of images atmospherically corrected, the method automatically retrieves the images having low cloud coverage according to the available cloud and shadow masks. This is done by using the cloud masks generated by Sen2cor [8]. Small cloud gaps are filled according to [9].

The peculiar multi-resolution property of Sentinel 2 images involves four spectral bands acquired at a spatial resolution of 10 m, six spectral bands acquired at a spatial resolution of 20 m and three spectral bands acquired at a spatial resolution of 60 m. Because the 60 m spectral bands are mainly dedicated to atmospheric corrections and cloud screening [10], only the 10 and 20 m bands are used to produce the HR LC maps. A nearest neighbour interpolation technique is used to match the spatial resolution of the 20 m bands to the 10 m ones for the entire tie series. The nearest neighbour interpolation technique has the drawback of generating smoothed images, thus losing in sharpness with respect to more sophisticated interpolation technique such as High Pass Filter (HPF). However, no new values are calculated by interpolation. This condition allows us to keep the original spectral information recorded by the sensor. Finally, we perform a spectral outlier detection and removal by discarding the pixels having values higher than the 0.999 quantile and lower than the 0.001 quantile of the spectral band. A radiometric normalization is eventually applied to the interpolated images so that each spectral band is rescaled between zero and one.

3.1.2.2 Training set preparation

Due to missing training data, existing thematic products available at global scale are considered to produce the training set. To extract samples able to represent the LC classes in the legend [AD4], three thematic products are considered: (1) the 2015 ESA CCI LC map available at 300m spatial resolution [11], (2) the 2015 Copernicus Global Land Cover (GLC) map produced at 100m spatial resolution [12], and (3) the Global Land Cover by National Mapping Organizations (GLCNMO) produced at 1 km spatial resolution [13].

First, we rescale the existing LC map at 10 m spatial resolution and then to convert the considered legend into the required one. To this end, we refer to the Land Cover Classification System (LCCS) [12], which is the standard common LC language for translating and comparing existing legends. Table 2 presents the translation of the considered thematic products into the HRLC legend.

Table 2. Training Set Production: the translation of the considered coarse thematic products into the desired map legend is reported. Bare rocks, built-up areas and bare soil classes are inserted via photointerpretation.

CCI-HRLC	ESA CCI LC 2015	CGLC	GLCNMO	Photo Interpretation
 Evergreen, broadleaf		3, 9		
 Evergreen needleleaf		1,7		
 Deciduous broadleaf		4,10		
 Deciduous needle leaf		2,8		
 Shrubland		13		
 Permanent cropland		18,14		
 Annual summer cropland			12	
 Grassland	130			
 Lichens and mosses		16		
 Permanent water bodies		21		
 Permanent snow and ice		20		
 Beaches dunes and sands			17	
 Bare soils		16		x
 Bare rock				x
 Built-up areas				x

A weak training set production is performed by selecting from the available thematic maps those samples having the highest probability of belonging to areas correctly associated to their label. Many difficulties arise when exploiting existing thematic maps generated with RS data characterized by properties different from Sentinel 2 images. Due to the coarse spatial resolution, the label assigned to mixed pixel is propagated to the pure pixels of Sentinel 2 images. Moreover, the considered maps are outdated and thus, they are not completely reliable. To address all these issues, we perform an automatic and unsupervised analysis which extracts from the existing thematic maps a weak, but reliable training set. First, a random stratified sampling is performed by using the LC

classes as strata. Five training sets are generated via bootstrap statistical method (e.g., without replacement) and used to train an ensemble of statistically independent classifiers. This condition allows us to generate an intermediate thematic product obtained at 10 m spatial resolution by classifying the time series of Sentinel 2 images. Only the areas where the ensemble of classifiers agree are kept. This condition allows us to increase the probability of selecting reliable samples to produce the final weak training database. Finally, few samples were added by photointerpretation for the classes that required a geometrical detail higher than the resolution of the considered thematic products (e.g., building, roads, small bare soil areas).

Figure 2 shows a qualitative comparison over a portion of the study area located in tile 21KUQ (Amazonia) between: (a) the coarse thematic product obtained by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (b) the intermediate product produced by the ensemble of five classifiers, (c) the weak training samples selected, and (d) a true colour composition of the Sentinel 2 image acquired on 17th April 2018. The qualitative example demonstrates the importance of generating the intermediate product at 10 m spatial resolution to sharply increase the probability of selecting samples correctly associated to their labels with respect to the ones that can be directly selected from the coarse thematic product.

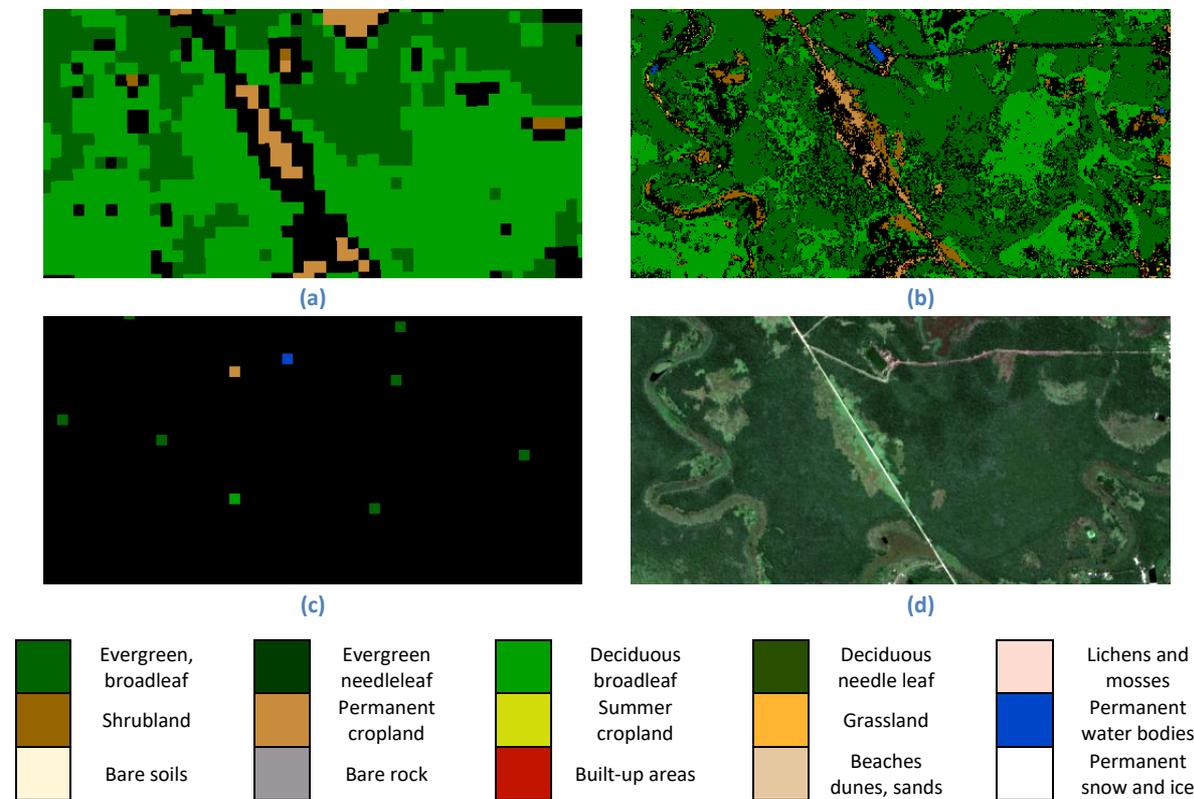


Figure 2. Visual comparison of the: (a) coarse thematic product generated by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (b) intermediate HR LC product produced by the ensemble of 5 classifiers, (c) extracted weak training samples, and (d) true color composition of the Sentinel 2 image acquired on 17 April 2018.

3.1.2.3 Classification

Automatic classification is a crucial processing step to produce accurate LC maps. The selected classification algorithm must achieve the best trade-off between classification accuracy and computational burden due to the need of processing a huge amount of data.

By analysing the recent literature, the team identified several successful core approaches to the classification. Some of them are now very consolidated, such as the Support Vector Machine (SVM) classifier [14] given the almost unanimous consensus obtained on its effectiveness to generate HR LC maps. SVM classifiers are based on kernel methods that have been extensively employed for the classification of RS data. In the considered implementation, we exploit SVM with Gaussian Radial Basis Function (RBF) kernels because of its capability of

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	12	

dealing with noisy samples in a robust way and to produce sparse solutions. A feature selection step is performed to detect the feature subspace where the LC classes are more discriminable. In the considered implementation, a Sequential Forward Floating Selection (SFFS) method based on the Jeffreys-Matusita distance as separability criterion is used [15]. The optimal kernel parameters (i.e., the regularization parameter C and the spread of the kernel γ) are selected by a 3-fold cross-validation. The LC maps generated by the SVM classifier are compared with the ones obtained by using other classification algorithms widely employed to generate global LC maps: Random Forest (RF) [16] and Maximum Likelihood (ML) [17]. RF is a commonly used classifier for LC classification due to its capability of being robust to label noise, yielding high classification accuracy with a low computational complexity. In the considered implementation we follow the parameter setting suggested by [18]: namely the number of trees to build equal to 200; and the number of input features randomly selected by each node equal to the square root of total number of features (i.e., the total number of spectral bands of the time series of Sentinel 2 images).

ML is one of the most common basic parametric classifiers based on statistical approach. ML is based on the statistical representation of the class distribution, thus achieving good accuracy for data with normal distribution and often poor quality for data with non-normal distribution. Although ML are easy to understand and interpret, they require a large training set with preferably only pure training samples. The last classifier used to generate HR LC maps is the Artificial Neural Network (ANN). Neural networks manage well with large feature space and generally obtain high classification accuracy. However, they require a large diversity of training set and are computationally expensive. Like other non-parametric methods, they are often a good choice for large LC applications where the data distribution is unknown. In the considered implementation, we exploit a simple feedforward neural network having one hidden layer characterized by ten neurons. The team is also evaluating the possibility of using sophisticated deep learning technique such as Long Short Term Memory classifier [19] to extensively exploit the spectral information provided by the long time series of Sentinel 2 images (see ADP v1.0). However, ground reference data are mandatory for the proper training of deep learning architectures.

The weak training sets automatically generated for the considered study areas are used to train all the above-mentioned classification algorithms in order to compare their classification performances. Due to the missed availability of ground reference data, a qualitative analysis is carried out to determine the classifier able to provide the best classification results.

3.1.3 Qualitative evaluation

In the following examples of qualitative analysis performed in the different study areas are reported. Figure 3 reports a comparison of classification results obtained in Amazonia (tile 21KUQ). The low geometrical resolution of the coarse thematic map (Figure 3b) is sharply improved in the HR classification maps. By comparing the LC classes in the ESRI HR optical images used to perform the qualitative evaluation results (Figure 3a), the best LC map is obtained by using the SVM (Figure 3f). It is the only classifier that correctly extracts the build-up areas. The worst result is obtained by the ML (Figure 3c), which misclassifies evergreen broadleaf as permanent cropland.

Figure 4 presents examples of classification results obtained in Amazonia (tile 21KXT). In this case, there is small correspondence between the coarse thematic product (Figure 4b) and the LC in the scene (Figure 4a). However, the classification results accurately retrieve the geometrical detail of the scene, thus recovering the river, lakes and road. All the classifiers correctly detect build up areas, river and lakes and obtain similar classification map. To identify the best classifier ground truth data are needed.

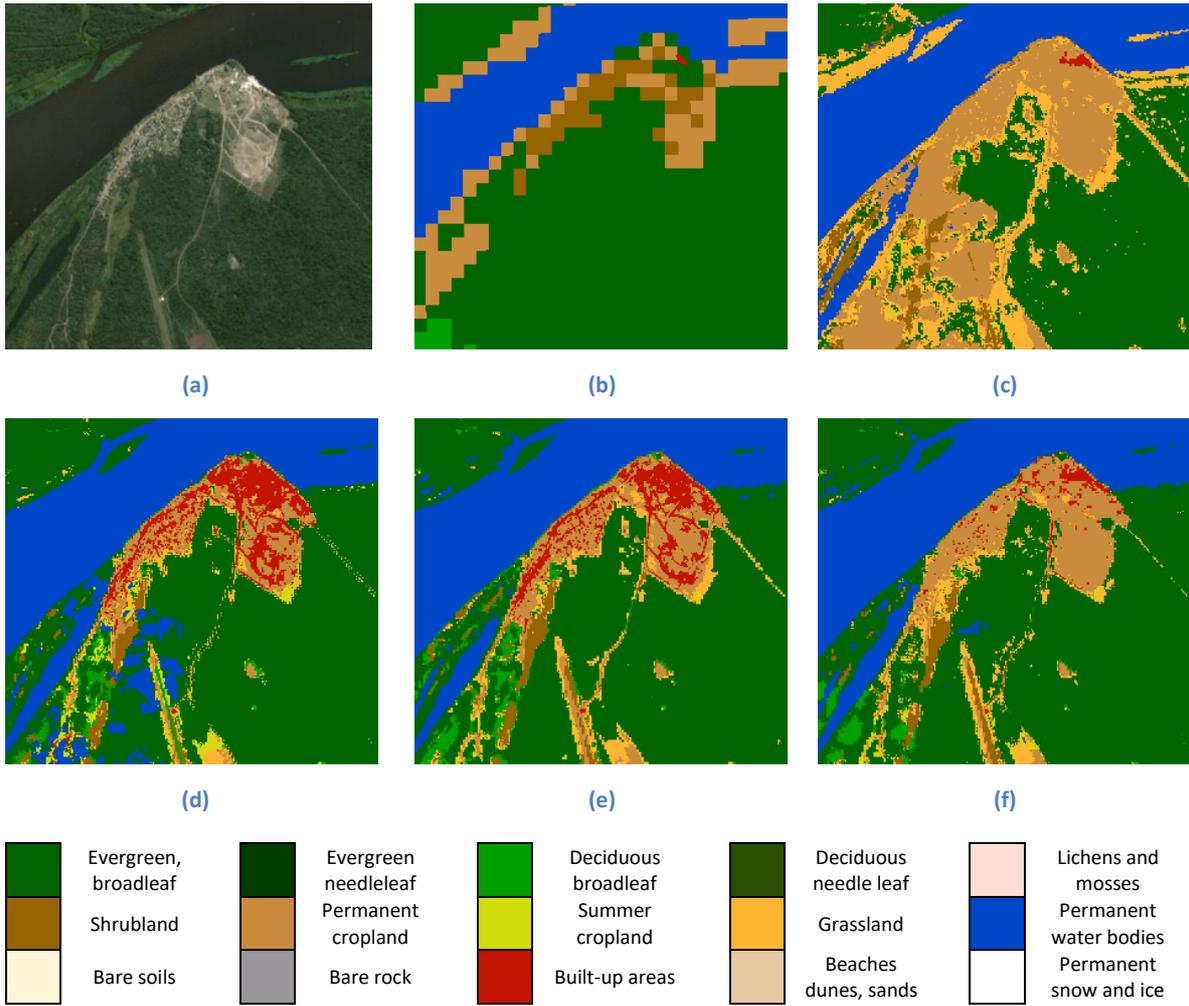


Figure 3. Visual comparison of the: (a) HR optical image used to evaluate the results obtained; (b) coarse thematic product generated by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (c) LC map obtained by using ML, (d) LC map obtained by using ANN, (e) LC map obtained by using RF, and (f) LC map obtained by using the SVM. The study area is in Amazonia (Tile 21KUQ).

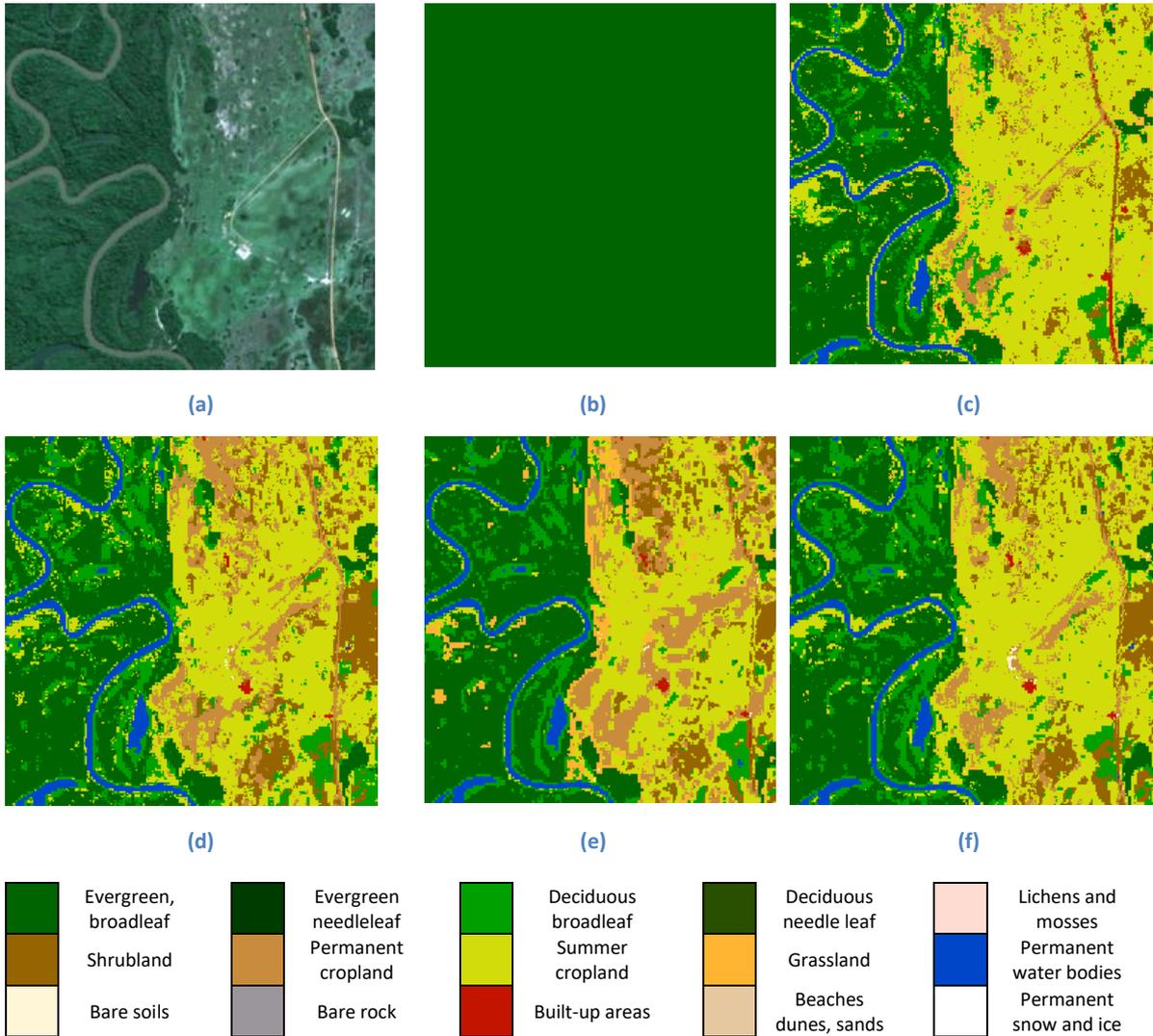


Figure 4. Visual comparison of the: (a) Sentinel 2 image acquired on the 9th May 2018; (b) coarse thematic product generated by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (c) LC map obtained by using ML, (d) LC map obtained by using ANN, (e) LC map obtained by using RF, and (f) LC map obtained by using SVM. The study area is in Amazonia (Tile T1KXT).

Figure 5 represents the classification products obtained by the considered classifiers in Africa (tile 37PCP). An evident correspondence between coarse thematic products and HR LC maps can be noticed. The classification map generated by the ML classifier is the only one where water and shrubland are confused (Figure 5c). One can notice that the build-up areas are accurately classified in all the HR maps.

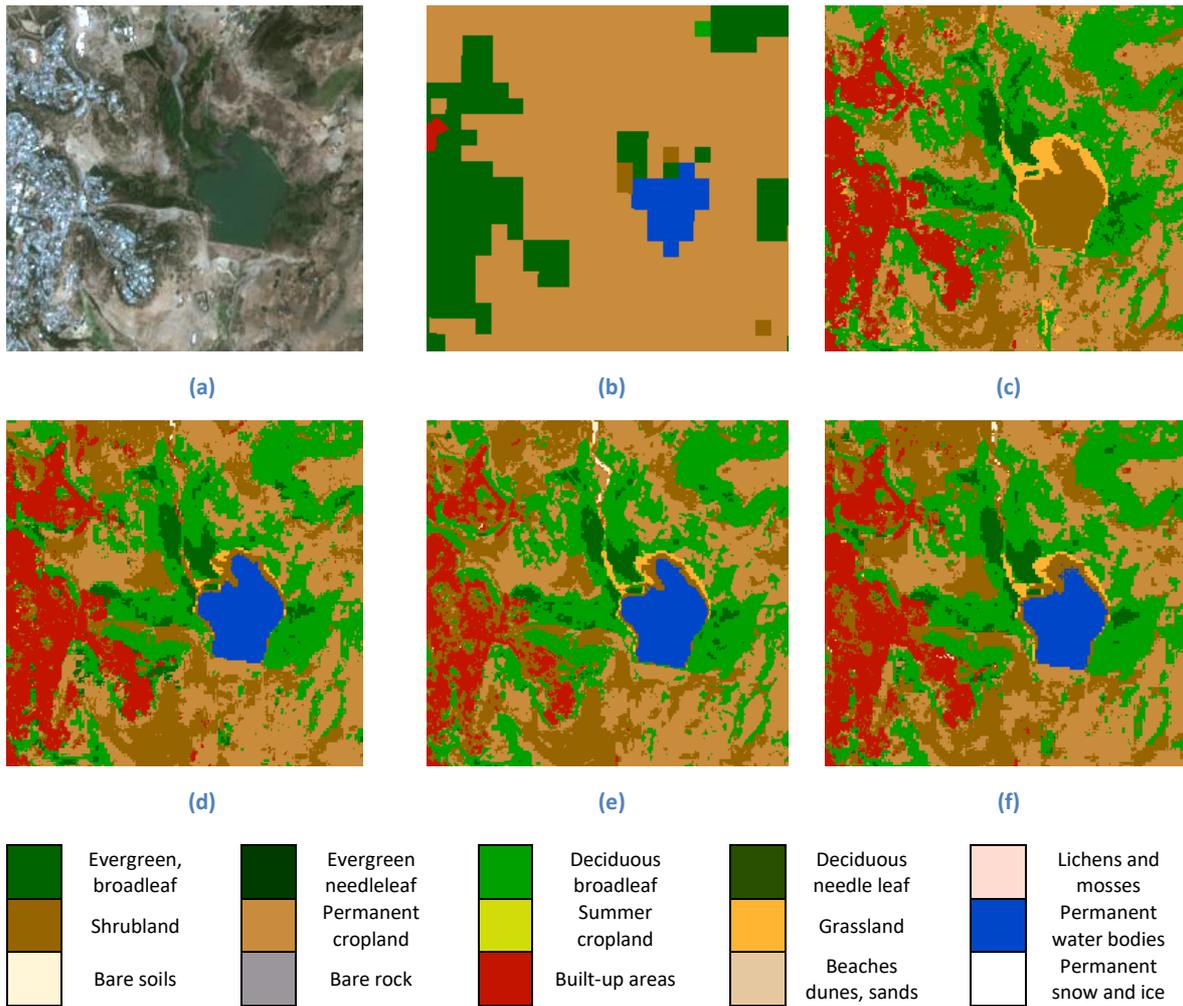


Figure 5. Visual comparison of the: (a) Sentinel 2 optical image acquired on the 3rd March 2018; (b) coarse thematic product generated by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (c) LC map obtained by using ML, (d) LC map obtained by using ANN, (e) LC map obtained by using RF, and (f) LC map obtained by using SVM. The study area is located in Africa (Tile 37PCP).

Figure 6 reports a comparison between the SVM classifier (Figure 6b and Figure 6e) and the African Prototype produced by ESA CCI (Figure 6c and Figure 6f) on two areas located in Africa (tile 37PCP). The African Prototype legend is converted into the HRLC one. From a qualitative analysis, one can notice that SVM correctly detects the build-up areas, the shrubland and the permanent cropland areas. As expected, the geometrical detail of the HRLC map is better than the one provided by the African Prototype.

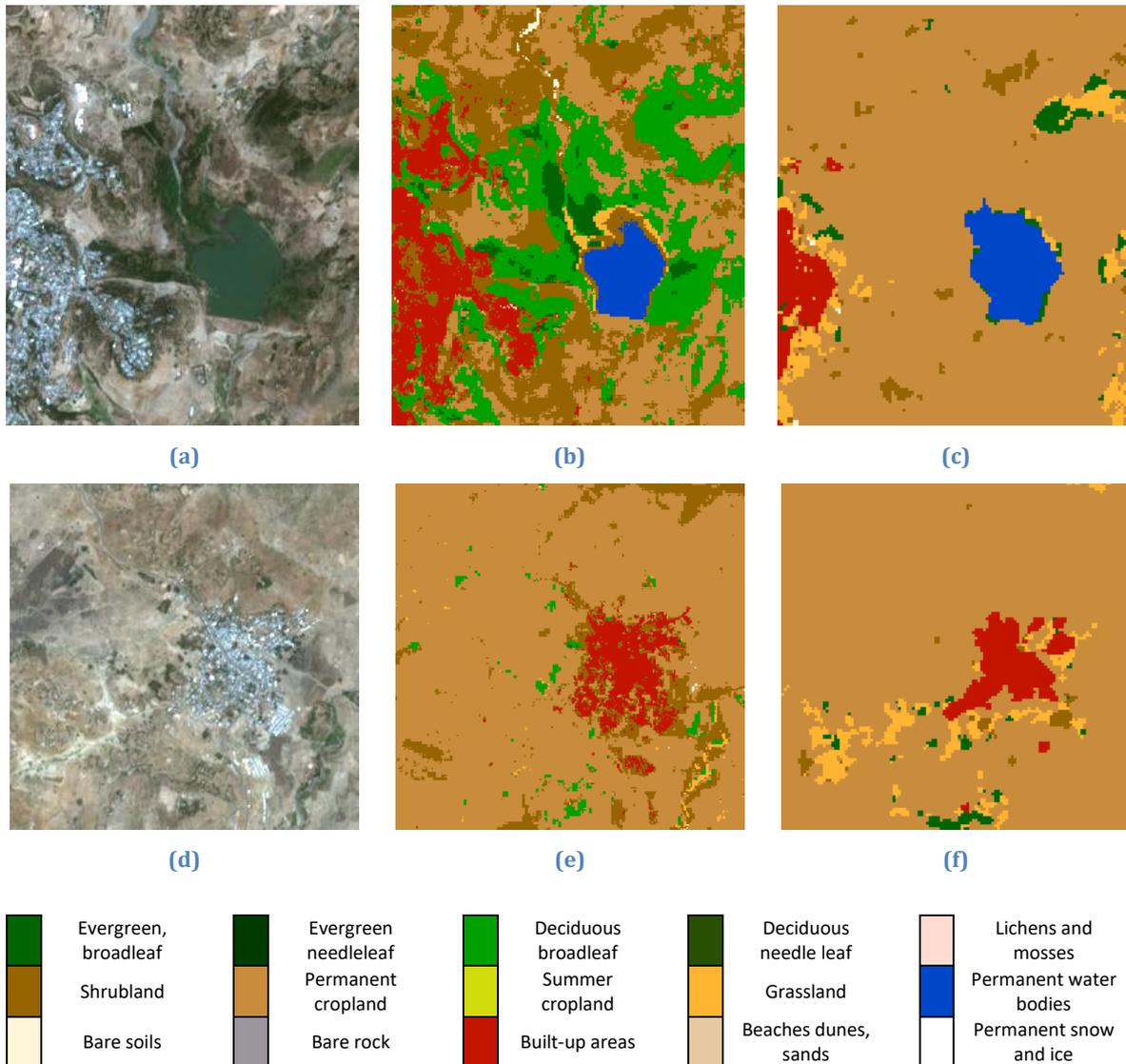


Figure 6. Visual comparison of the: (a)(d) true color composition of Sentinel 2 image acquired on the 3rd March 2018; (b)(e) LC maps obtained by using SVM, (c)(f) African prototype LC maps produced by ESA CCI converted to the HRLC map legend. The study area is located in Africa (Tile 37PCP).

Figure 7 represents the classification products obtained in Siberia (tile 42WXS). Like the other tiles, there is a correspondence between the coarse thematic products and the classification results. However, the HR maps sharply improve the geometrical detail in the scene thus highlighting the presence of lichens and mosses. All the classifiers correctly detect build up areas and lakes. From the visual interpretation point of view, Siberia is the most challenging area as expected. This is because the area is mostly covered with permanent water, shrubland, lichens and mosses which are difficult to be evaluated without reference data.

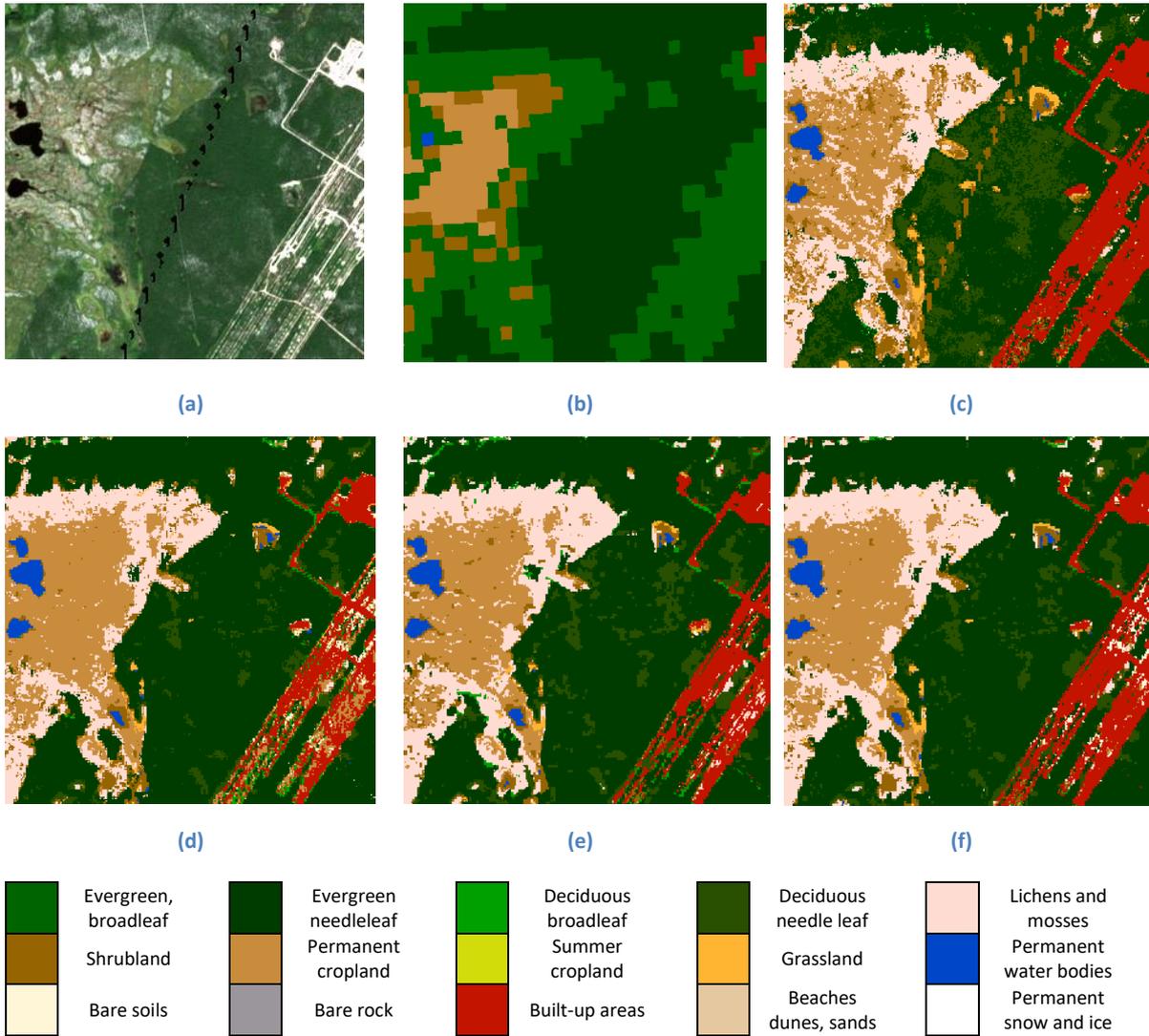


Figure 7. Visual comparison of the: (a) Sentinel 2 optical image acquired on the 21st July 2018; (b) coarse thematic product generated by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (c) LC map obtained by using ML, (d) LC map obtained by using ANN, (e) the LC maps obtained by using RF, and (f) LC map obtained by using SVM. The study area is in Siberia (Tile 42WXS).

3.1.4 Final decision

According to the qualitative analysis the best LC maps are achieved by SVM and RF classifiers. However, SVM better distinguishes between build up areas and bare soil. This is due to the capability of SVM of performing well with small training dataset. Thus as training samples of build-up areas are added to the weak training set by photointerpretation. ML presents the poorest results, by making many classification mistakes. This is mainly due to the fact that the considered HR LC classes do not follow the normal distribution, thus leading to poor classification performances. Although these preliminary results are encouraging, ground reference data are needed to perform a quantitative evaluation and to achieve better classification results.

3.2 SAR data processing

According to what is reported in the ATBD document [AD5], Figure 8 reports the general work flow for the kernel of the processing chain.

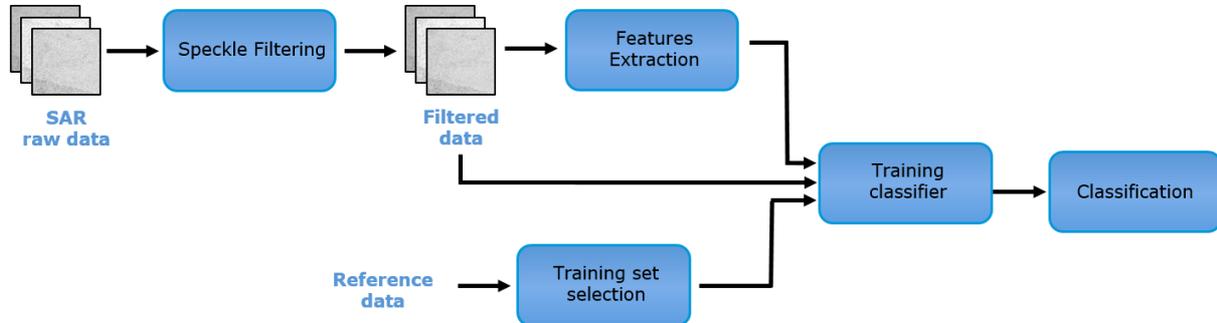


Figure 8. Block diagram for SAR data of the land cover map production procedure.

The reference data is split into a training subset. The image data is properly pre-processed in order to ensure spatial and temporal homogeneity and then image features for the classification are extracted. After that, classifier training and image classification are performed.

3.2.1 Satellite images

The Round Robin has been voted in analysing the following three thematic regions according to tiling grid of Sentinel-2 Level-1C products:

- Amazon area for 21KUQ S-2 tile;
- Amazon area for 21KXT S-2 tile;
- Siberia area for 42WXS S-2 tile.

Sentinel-1 is the only source of radar images used in Round Robin. Level-1 data processed into Ground Range Detected (GRD) products, acquired in Interferometric Wide Swath (IWS) and available through the Copernicus Open Access Hub (<https://scihub.copernicus.eu/dhus/#/home>) were used over the three Round Robin areas for the year 2018. The Level-1 GRD data contained the detected amplitude (phase information is lost) and are multi-looked to reduce the impact of speckle at a cost of reducing spatial resolution. The products are projected to ground range using Earth ellipsoid model, generating images with approximately square resolution pixels and square pixel spacing. The complete list of images is given in Table 3.

The growing availability of “free and open” satellite imagery at 10-30 m spatial resolution encourages the development of innovative and advanced methodologies in the context of the climate change initiative.

3.2.2 Method/algorithm/technique

Within this Section, several methods are presented and compared as candidate approaches to be developed and implemented in the SAR image processing chain in Figure 8.

3.2.2.1 SAR data pre-processing

The first step consists in a preprocessing task aiming to provide a time series devoted to the land cover map production. The Level-1 products of Sentinel-1 GRD data are properly calibrated and terrain corrected before any other processing. The preprocessing is accomplished by Sentinel’s Application Platform (SNAP) software provided by ESA

3.2.2.2 Training set preparation

The training set extraction refers to procedure presented in section 8.2 of ATBD document [AD5]. Consistent and accurate training data that cover a large area is not available and reference data originate from existing databases have been used for classifier trained. Two data sources have been taken into account:

- ESA CCI-LC 2015 300m [20];
- GLCNMO 1km [13];

These products were therefore combined in order to ensure a unique reference data set. Then, random sampling is applied for extracting a consistent set of training samples, as described in ATBD [AD5].

A set of classes of interest were defined, namely: evergreen broadleaf tree, evergreen needleleaf tree, deciduous broadleaf tree, deciduous needleleaf tree, shrubland, permanent cropland, annual summer cropland, grassland, lichens and mosses, permanent water bodies, permanent snow and ice, beaches, dunes and sands, bare soils, bare rock and built-up areas (Figure 9).

In this document it is proved that the use of these combined data sources allows obtaining a classification maps that is largely independent from the random set of points used for training, hence effectively implementing an automatic supervised approach.

Table 3. List of Sentinel-1 data

Area	Section	Satellite	Band	# Products	Date list (2018y)
Amazonia – 21KUQ	Upper	S1B	VH and VV	25	03-10 03-22 04-03 04-15 04-27 05-09 05-21 06-02 06-14 06-26 07-08 07-20 08-01 08-13 08-25 09-06 09-18 09-30 10-12 10-24 11-05 11-17 11-29 12-11 12-23
	Lower	S1B	VH and VV	25	02-26 03-10 03-22 04-03 04-15 04-27 05-09 05-21 06-02 06-14 06-26 07-08 07-20 08-13 08-25 09-06 09-18 09-30 10-12 10-24 11-05 11-17 11-29 12-11 12-23
Amazonia – 21KXT	Upper	S1B	VH and VV	27	01-16 03-05 03-17 03-29 04-10 04-22 05-04 05-16 05-28 06-09 06-21 07-03 07-15 07-27 08-08 08-20 09-01 09-13 09-25 10-07 10-19 10-31 11-12 11-24 12-06 12-18 12-30
	Lower	S1B	VH and VV	26	01-04 03-05 03-17 03-29 04-10 04-22 05-04 05-16 05-28 06-09 06-21 07-03 07-15 07-27 08-08 08-20 09-01 09-13 09-25 10-07 10-19 10-31 11-12 11-24 12-06 12-18
Siberia – 42WXS	-	S1B	VH and VV	20	01-08 03-09 03-21 04-14 05-08 05-20 06-13 06-25 07-07 07-19 07-31 08-12 08-24 09-05 09-17 09-29 10-11 11-04 11-16 12-22

Value	Label	Color
0	No data	Black
1	Evergreen broadleaf tree	Dark Green
2	Evergreen needleleaf tree	Medium Green
3	Deciduous broadleaf tree	Bright Green
4	Deciduous needleleaf tree	Light Green
5	Shrubland	Brown
6	Permanent cropland	Light Brown
7	Annual summer cropland	Yellow
8	Grassland	Orange
9	Lichens and mosses	Pink
10	Permanent water bodies	Blue
11	Permanent snow and ice	White
12	Beaches dunes and sands	Tan
13	Bare soils	Light Yellow
14	Bare rock	Grey
15	Built-up areas	Red

Figure 9. Land cover classes for Sentinel-1 data classification

3.2.2.3 Speckle filtering

For speckle reduction, two approaches have been compared. The first one is the Lee filter, one of the well-known filters for despeckling and enhancing SAR images. It uses the minimum mean square error (MMSE) filtering criterion as explained in [21]. The second approach is the multitemporal despeckling method developed by Zhao et al. in [22]. This filter is based on the calculation of a super-image exploiting the spatial and temporal information of a SAR time series. Both methods aim to enhance the quality of image by means an effective speckle reduction and spatial resolution preservation, but they have different impacts on classification performance.

3.2.2.4 Feature extraction

Feature extraction methods encompass characteristics and texture, structural and graph descriptors. 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 [23]. The feature extraction could be carried out considering both single and double bands analysis.

In single band case, for analyzing and exploring the spatial information contained in satellite images, a set of filters that operate especially in spatial domain have been assumed. The rationale for selecting these algorithms is especially due to their velocity of the execution and versatility. Although they might not be the most accurate one, 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.

- *Lee filter* is an adaptive filter based on minimum mean-square error (MMSE) that converts the multiplicative model into an additive one, thereby reducing the problem of dealing with speckle to a known tractable case.
- *Minimum (maximum) filter* is a non-linear filter that locates 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.

In dual polarimetry analysis case, polarimetric information of Sentinel-1 data were extracted using intensities of the VH and VV channels, and several composite images given by:

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

In this document, possible combinations of these feature sets (and subset of) are compared and discussed, in order to identify a (sub-)set of features for the proper detection of the classes of interest.

3.2.2.5 Classification

The global mapping systems using high resolution imagery could implement rule or advanced approaches based on the definition of the classes. For classifying satellite images, two supervised classifiers have been considered, the *Random Forest* (RF) and *Support Vector Machine* (SVM). They are superior to unsupervised methods and more robust [16]. Some erroneous reference data (e.g., slightly outdated ones) are acceptable in training [24].

The whole classification chain has been investigated on Sentinel-1 time-series data assuming different scenarios. Both RF and SVM classifier have been applied. The experiments presented in this documents are carried on two sites of continental Amazonian region, see Figure 10, which amounts to about 20.000 Km².

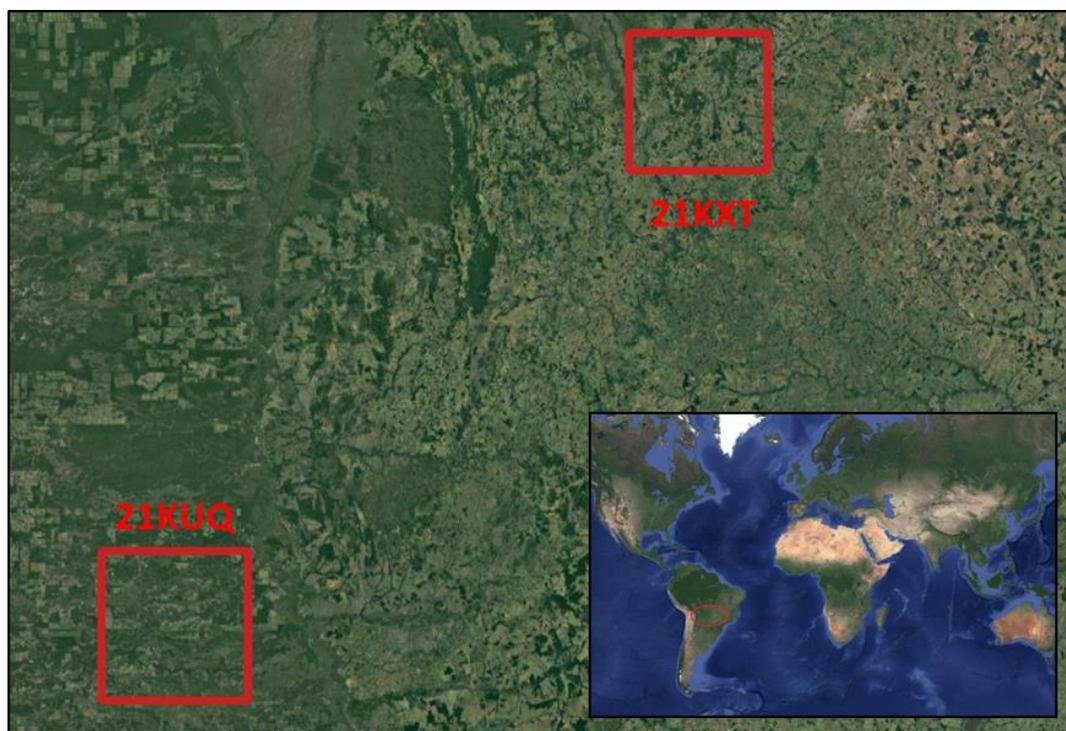


Figure 10. Study areas (highlighted in red) referring to 21KUQ and 21KXT Sentinel-2 granules.

The RF was applied for classification performance evaluation in the following cases:

- Single image, single band analysis (VH band);
- Single image, double band analysis (VV and VH);

- Multitemporal analysis, on the basis of a Sentinel-1 time series.

The choice of parameters for the RF classifier is not very sensitive for this kind of problem [25], and one-hundred trees have been used.

In single band analysis, the VH channel of a single image has been considered, and the speckle noise was reduced applying multitemporal despeckling approach. The classification was carried on features aimed to identify textural and spatial information of the scene (described in Section 3.2.2.4). For double band analysis, also VV channel has been taken account. The features given by VV and VH combination (*mean*, *ratio* and *difference*) have been extracted and used in classification.

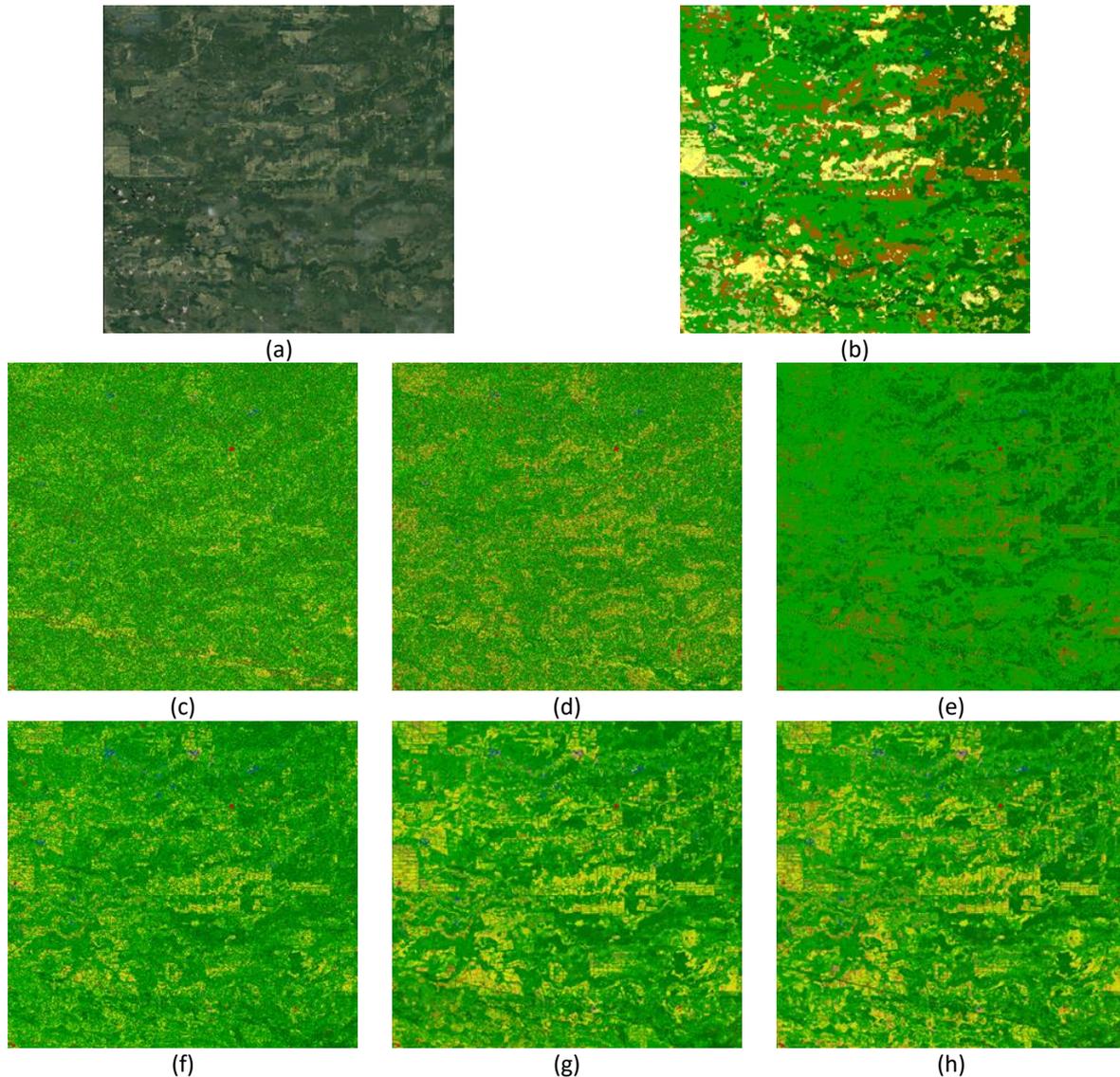


Figure 11. 21KUQ Amazonian tile: the figure shows a comparison among the satellite image (a); ESA CCI LC 2015 reference data (b), and classification maps using single image-single band (c), single image-double band (d), multitemporal sequence (e), multitemporal sequence with majority voting (f), seasonal multitemporal sequence with three images per season (g) and seasonal multitemporal sequence with five images per season (h).

To carry out the multitemporal analysis, a time series of Sentinel-1 VH images has been assumed. All images have been divided according to annual seasons: *winter*, *spring*, *summer* and *autumn*. For each season, a multitemporal despeckling filter has been computed and subsequently applied. Hence, three filtered images have been randomly chosen for each season (with a total of twelve images) and applied in input to RF classifier. To try and improve the performances of classification, more spatial and textural information have been added, by means of a seasonal mean image. Two scenarios have been evaluated. In the first case the average of three images per

season has been selected, whereas five images have been considered in the second one. The spatial features (Lee, Min, Max, Max Min, Mean and Median) of each mean image were then extracted, leading to a total of twenty-eight features for classification. Still in the multitemporal analysis framework, also the majority voting algorithm has been evaluated as potential rule to be deployed in classification chain. This approach consists in applying the whole classification chain for each season, in order to have four classification maps. Hence the majority voting is carried out pixel-per-pixel over the maps for accomplishing a single final product. All classification results for 21KUQ and 21KXT Amazonian granules are reported in Figure 11 and Figure 12, respectively.

For further comparison, SVM classifier was also applied to carry out a multitemporal analysis (based on mean image of five images per season). Results are reported in Figure 13 and Figure 14.

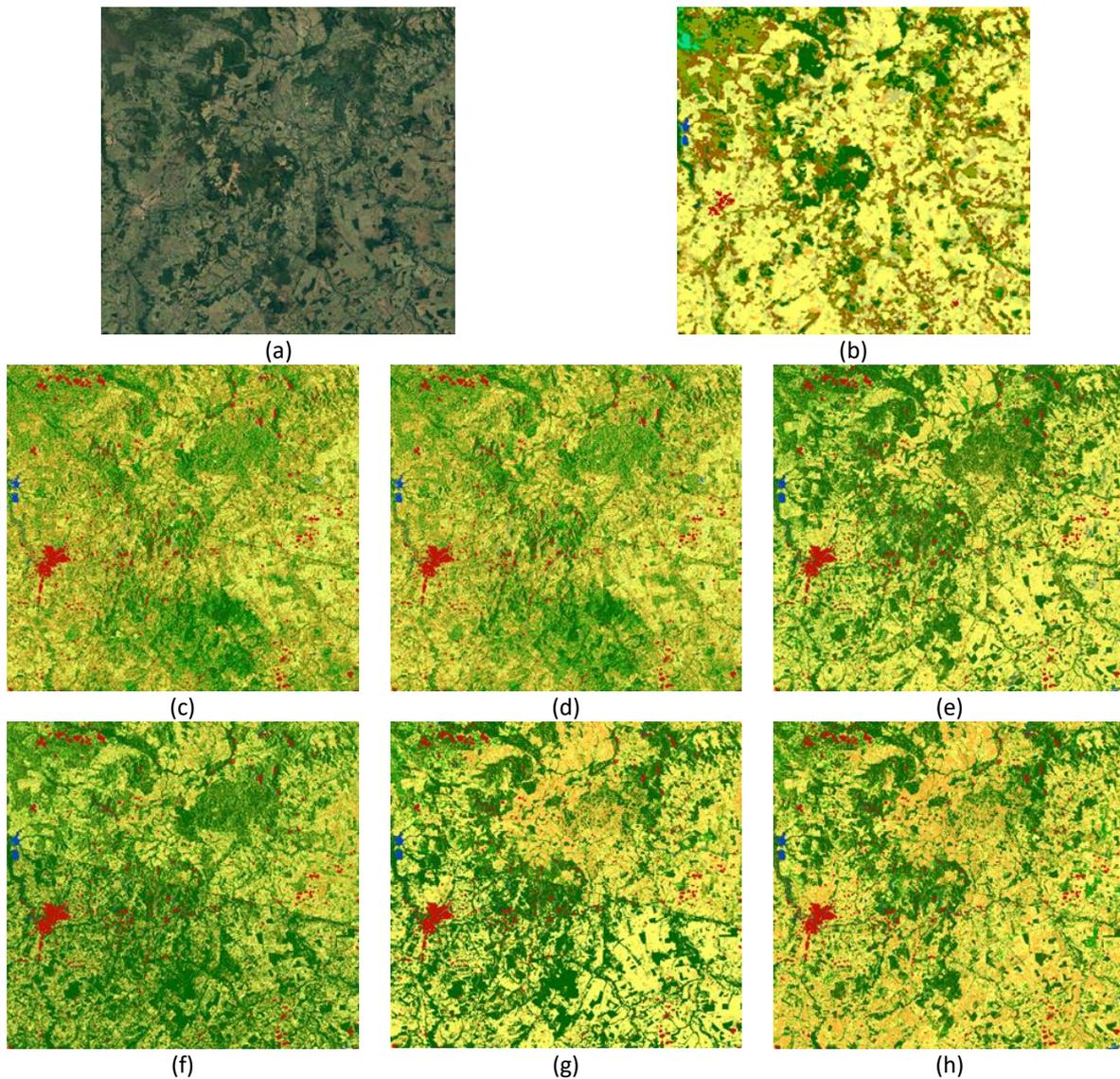


Figure 12. 21KXT Amazonian tile: the figure shows a comparison among the satellite image (a); ESA CCI LC 2015 reference data (b), and classification maps using single image-single band (c), single image-double band (d), multitemporal sequence (e), multitemporal sequence with majority voting (f), seasonal multitemporal sequence with three images per season (g) and seasonal multitemporal sequence with five images per season (h).

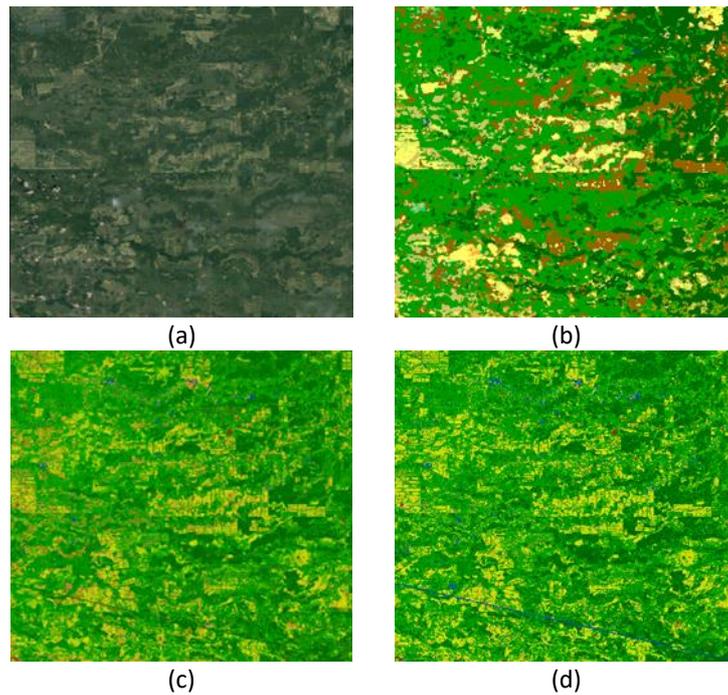


Figure 13. Comparison of RF and SVM classification for the 21KUQ Amazonian tile: the satellite image (a), ESA CCI LC 2015 reference data (b), RF map using a seasonal multitemporal sequence (five image per season) (c) and SVM map with the same input (d).

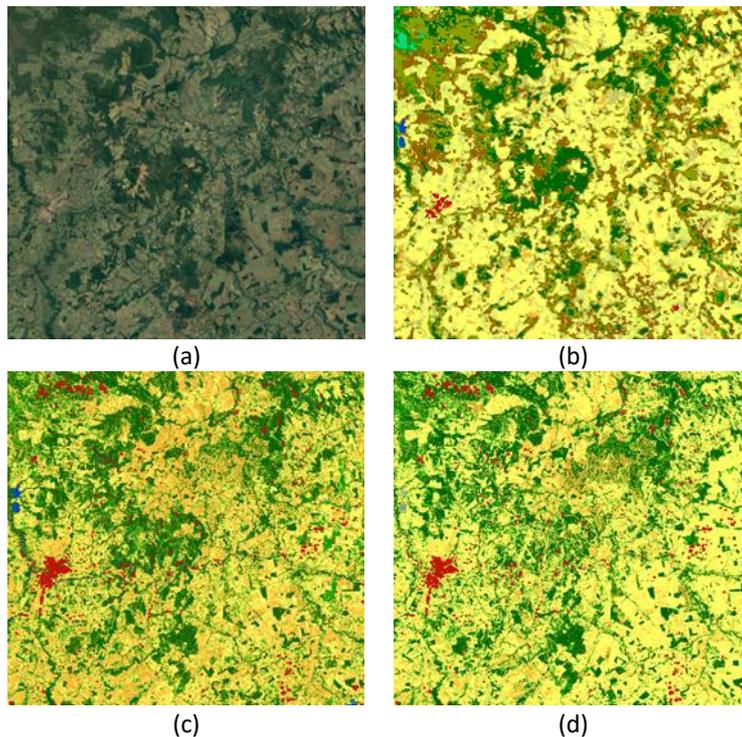


Figure 14. Comparison of RF and SVM classification in 21KXT Amazonian tile: the satellite image (a), ESA CCI LC 2015 reference data (b), RF map using a seasonal multitemporal sequence (five image per season) (c) and SVM map with the same input (d).

The whole processing chain has also been investigated in terms of despeckling filters. In fact, two experimental environments have been considered and the classical Lee speckle filter has been assumed both for single- and multi-image analysis. In first case, speckle noise was reduced by Lee filtering applied on single S-1 image. In the other one, five images of each season have been separately processed with Lee filter and their average has been

calculated. Features have been properly extracted as described in previous analysis and RF classification has applied in both times.

To inquire into the robustness and effective of classification with respect to reference data, several training set have been randomly extracted and used in the processing. Once applied multitemporal despeckling over the whole SAR dataset, a multitemporal analysis have been carried out by mean RF classification. Resulting maps are displayed in Figure 15 and Figure 16.

The urban class (highlighted in red in reported classification maps) has been separately extracted with the developed UEXT algorithm (download is available at tlclab.unipv.it/downloads/1/20171207/UEXT-files.zip), well described and validated by Lisini *et al.* [26].

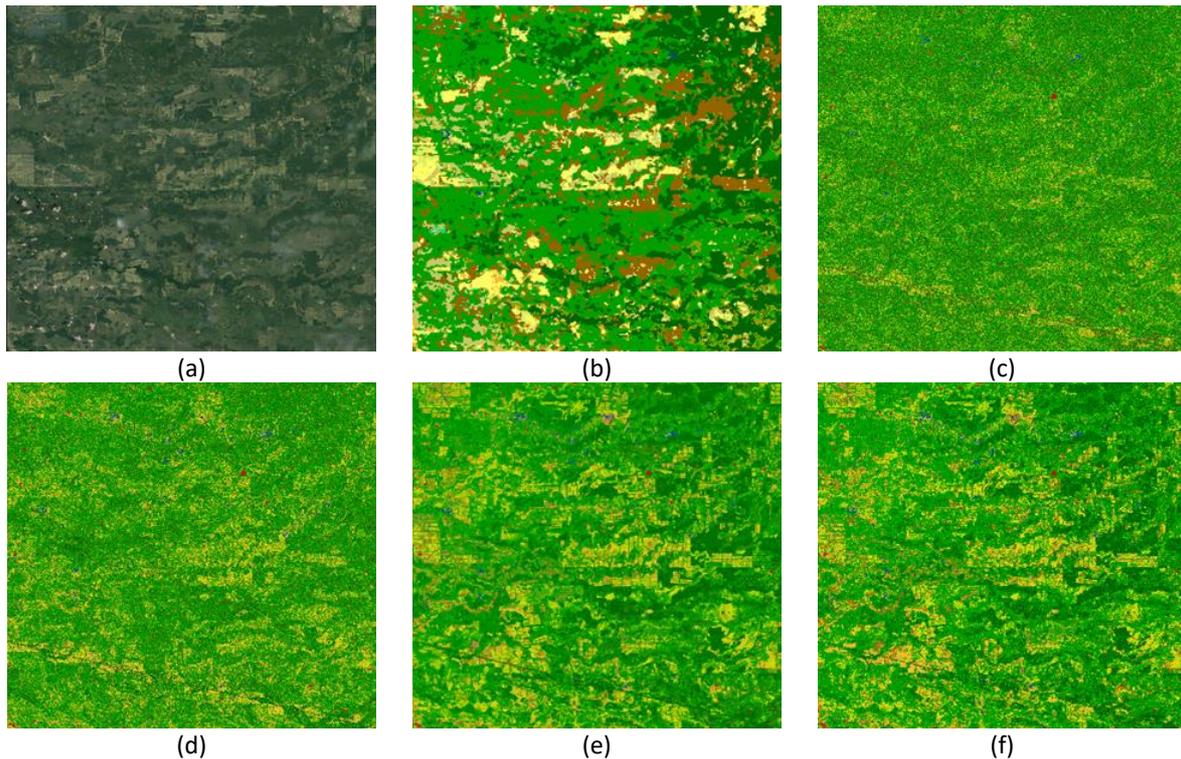


Figure 15. Lee filter applied in RF classification for the 21KUQ Amazonian tile: the satellite image (a), ESA CCI LC 2015 reference data (b), single image classification with multitemporal despeckling (c), single image classification with Lee despeckling (d), multitemporal classification with multitemporal filtering (e) and multitemporal classification with Lee filtering (f).

3.2.3 Qualitative evaluation

By looking at the maps reported above, the different performance levels of the considered approaches can be appreciated. By comparing classification results in Figure 8 with the satellite image (a) and referenced data (b), it can be seen that the severe classification noise is gradually mitigated with multi-image analysis. The single-image case (Figure 11 (c)) shows poor performance in terms of class recognition, as well as attained in double band analysis (Figure 11(d)), since images appear very noisy with a lot of classes hard to be visually distinguishable. The multitemporal approach produces instead interesting results, with nice classification capabilities, confirming the state-of-the-art level of this methodology. In Figure 11 (h) several croplands and vegetative areas are very well extracted proving the capability of the approach to produce a more reliable map. Moving to the classification maps reported in Figure 12, a similar behaviour can be appreciated, up to the map in Figure 12 (h).

In Figure 13 and Figure 14, a comparison between Random Forest and Support Vector Machine performances is reported for the 21KUQ and 21KXT tiles, respectively. In both cases, RF maps (Figure 13 (c) and Figure 14 (c)) show a greater capability in class recognition than SVM ones. The performance can be explained due to the nature of RF in being inherently multiclass whereas Support Vector Machines need workarounds to deal with

classification tasks involving multiple classes. Moreover, RF works well with a mixture of numerical and categorical features. In other words, with Random Forest one can use inhomogeneous data without any refined pre-processing.

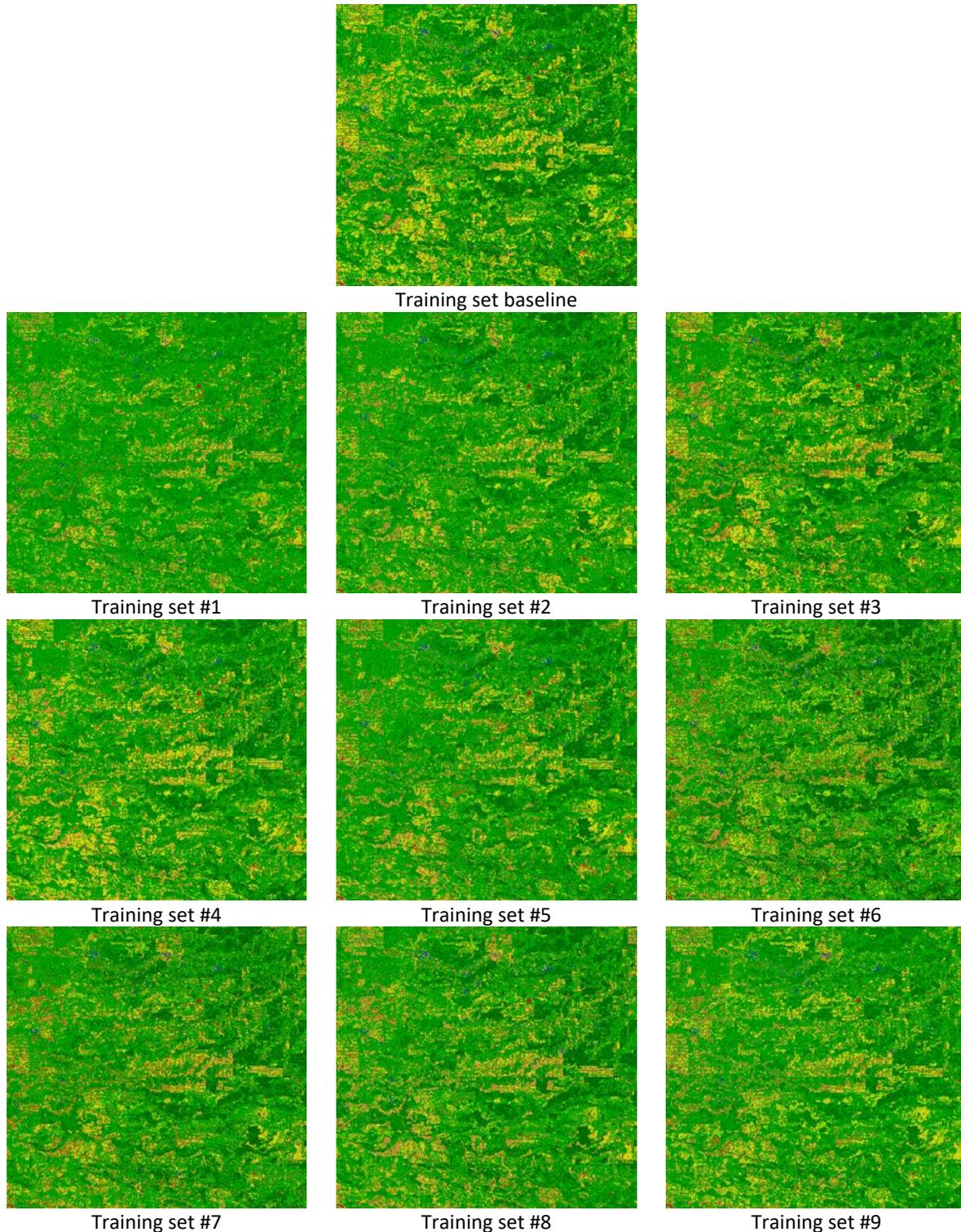


Figure 16. Random Forest maps for to the 21KUQ Amazonian tile using randomly selected different training sets.

Figure 15 can be used to visually assess the effectiveness of speckle reduction when using different filters, in this case the Lee and the multitemporal approaches. The Lee filter (Figure 15 (f)) performs a linear combination of the observed intensity and the local average intensity value within the fixed window (use of the local statistics). By applying the filter, a mild smoothing effect is achieved, and the noise reduction does not appear visually

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	27	

strong. In comparison, the multitemporal filter, applied to a temporal series of Sentinel-1 images (Figure 15(e)), gives very good results: the noise is well reduced and image sharpness is remarkable. Multitemporal despeckling better preserves spatial structures in multitemporal SAR images while effectively removing speckle.

Finally, the last test (Figure 16), aim at showing the robustness of the classification approach with respect to several randomly extracted training sets. According to the maps, although the maps are obtained using different training sets, they visually show similar performance.

3.2.4 Final decision

According to the previous subsections and the comparisons, in despeckling, the improvement given by multitemporal classification appears evident. The method delivers better performance than the Lee adaptive filter, in addition to the capability to be versatile and scalable by jointly processing multiple images, i.e. different imaging sequences.

Moving to classification, RF is state-of-the-art in remote sensing image processing, and yields classification accuracies as high as SVM, but with a much lower computational complexity. It is also more stable with respect to the choice of parameters [25], and easily provides the probability of belonging to a class (necessary input to the data fusion step). This makes it an excellent candidate for operational processing chains.

Having that said, it has to be underlined that this comparison has been performed only qualitatively, and a quantitative analysis should be provided for a better evaluation of the quality.

3.3 Multi-sensor Optical and SAR Data Fusion

3.3.1 Multisensor Geolocation Methods

In the context of multi-sensor geolocation, different image registration methods and strategies have been designed and validated on the available dataset, following up on the methodological analysis conducted in the deliverables of the previous milestones.

The multi-sensor geolocation process is composed of different elements (Figure 17), 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 17. General block diagram of the geolocation process

According to the CCI proposal, we have developed and validated area-based image registration methods with a combination of different geometric transformations, similarity measures, and optimization strategies.

3.3.1.1 Geometric Transformations

With respect to the geometric transformations, we have considered:

- Translation 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

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

- Rotation-Scale-Translation (RST) transformations

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

Here, (x, y) and (X, Y) indicate the axes of the coordinate frames in the reference and input images, respectively, T_x and T_y are translation parameters, θ is a rotation angle, and s is a scale parameter [27].

3.3.1.2 Similarity Measures

With respect to the similarity measures, we have considered:

- Cross-correlation

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

- Mutual Information

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

Here, $In(\cdot)$ and $Ref(\cdot)$ indicate the input and reference images (which are both assumed composed of $M \times N$ pixels), respectively, $p_{Ref, In}$ is their joint probability density function (PDF), p_{Ref} and p_{In} are their marginal PDFs, $CC(\cdot)$ is their cross-correlation evaluated on a given pixel location, $MI(\cdot)$ is their mutual information [27], [28], [29].

It is worth noting that the cross correlation is computed through the fast Fourier transform (FFT) algorithm. 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:

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

where $\mathcal{F}(\cdot)$ denotes the Fourier transform operator and f and g are 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 from the computational efficiency of the FFT [30].

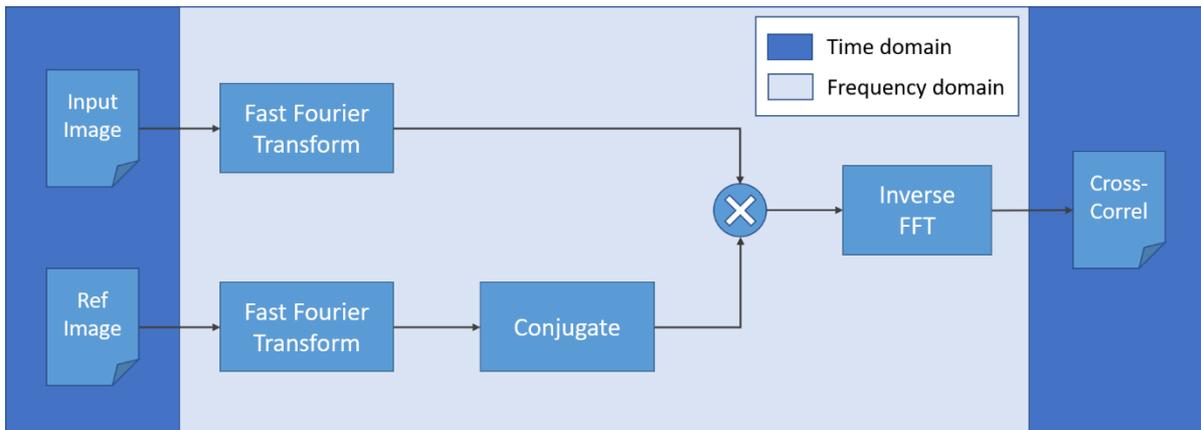


Figure 18. Flowchart of the cross-correlation computation between two images

In order to compute the cross-correlation between two images it is necessary to: (i) compute the FFT of each image 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

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	29	

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 shown in Figure 18.

3.3.1.3 Optimization Strategies

With respect to the optimization strategies, we have considered:

- Powell's algorithm, for unconstrained minimization. It 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 bi-directional search along a set of search vectors [31]. Moreover, the bi-directional line search is done by Golden-section search and Brent's method [32].
- Constrained optimization by linear approximation (COBYLA), for constrained minimization. It 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 [33].

3.3.2 Quantitative evaluation

Several experiments have been carried out with respect to multi-sensor geolocation. First, experiments with synthetic data generated from single-sensor measurements have been performed. Second, the aforementioned registration methods have been applied to multi-sensor S1-S2 data. The imagery associated with the first round robin were used for experiments

In particular, the dataset used for the experiments is composed of two S1 SAR tiles and an S2 optical tile. The single-sensor synthetic datasets have been built by extracting an area from either an S1 or an S2 tile and by transforming it according to predefined transformations (translation, rigid, RST, etc.). This strategy allows to quantitatively determine the registration accuracy obtained by registering the resulting couples.

The multi-sensor dataset has been built by extracting an area from the result of stacking the S1 and S2 tiles. This way, it is possible to test the registration methods in a fully real-world scenario, although analytical and quantitative results are not available of course, because the "true" ideal matching of the optical and SAR data is undefined. In this case, registration accuracy can be qualitatively appreciated using false color representations and checkerboard visualizations.

Finally, the synthetic multi-sensor dataset has been built by registering the S1 and S2 tiles using accurate georeferencing information, extracting an area from the stacking of the two georeferenced tiles, and applying a well-known transformation to one of the two. As for the first scenario, this strategy allows to quantitatively evaluate the accuracy of the proposed methods in a multi-sensor scenario as well.

The experiments that have been carried out are the following:

- Finding the optimal translation using cross-correlation as metric;
- Finding the optimal translation using both cross-correlation and mutual information as metrics;
- Finding the optimal rigid transformation using both cross-correlation and mutual information as metrics;
- Finding the optimal RST transformation using mutual information as a metric.

In the case of the synthetic dataset, where the ground truth (GT) transformation is known, the root mean square error (RMSE) between the real and the resulting transformation is used as a measure of registration accuracy. For details on the RMSE computation refer to [34] and [29]. Additionally, the computer used for the experiments is a Windows system equipped with a quad-core Intel i7 processor with a working frequency of 3.60GHz and 24GB of RAM.

Before analyzing the results in more detail, it is worth anticipating that the usage of cross-correlation as a similarity metric is effective and satisfactory only in the case of the single-sensor dataset. In the multi-sensor

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	30	

cases, the different nature of the images prevented the usage of such a similarity metric, because of the overall low values achieved by the inter-sensor correlations. On the contrary, the mutual information measure resulted to be more robust in this case.

Therefore, cross-correlation has been experimentally validated only in the cases of translation and rigid transformations. The results have been proved satisfactory only in the single-sensor scenarios, with poor performance in the multi-sensor cases. Hence, multi-sensor experiments with more challenging scenarios (higher-order transformations and multi-step registration) have been focused on the usage of the more promising mutual information similarity metric.

Moreover, another interesting conclusion is the comparison of the two optimization methods used in the experiments. As it will be shown, Powell's algorithm performs well in case of transformations where the input image and the reference image are not "very distant," i.e., when the optimal solution is in the neighborhood of the starting point. Conversely, the COBYLA algorithm allows the user to choose the starting search radius. The tuning of such parameter allows the registration process to explore regions of the search space that a simple conjugate-gradient method would never reach.

3.3.2.1 Finding the optimal translation using cross-correlation as metric

This type of registration performed well only in synthetic cases where the translation transformation was manually applied. Experiments were carried out with respect to optical-optical, SAR-SAR, and optical-SAR matching. Note that from now on, the convention used to represent the parameters of the geometric transformation is the following: Transformation = [translation on the x axis, translation on the y axis, rotation (deg.), scale factor].

3.3.2.1.1 Synthetic Optical-to-optical matching

Elapsed Time: 0.636 seconds

True Transformation: [-76, 52, 0, 1]

Resulting Transformation: [-76.0, 52.0, 0.0, 1.0]

Registration RMSE: 0.0 pixels

Figure 19 shows: (i) the reference optical image (left panel); (ii) the input optical image to be registered (central panel); and (iii) the resulting cross-correlation with the corresponding point of maximum (right panel).

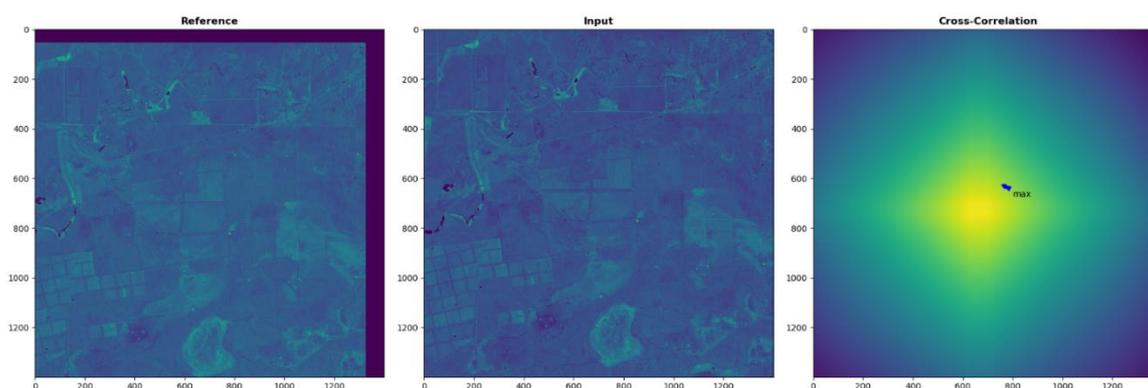


Figure 19. Synthetic optical-to-optical matching: reference optical image (left), input optical image to be registered (center), and cross-correlation (right).

Figure 20 shows a false color composition of the images before and after registration. The composition uses the green and the magenta colors for the input and the reference images (left) and for the transformed input and reference images (right), respectively.

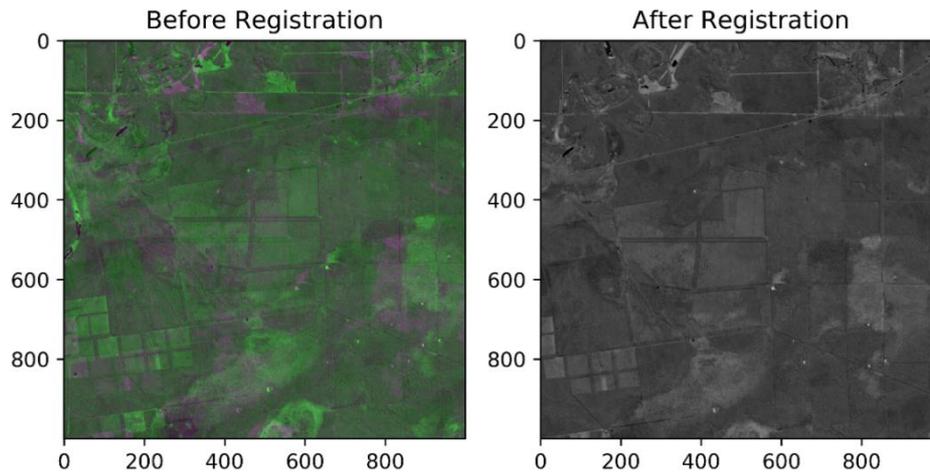


Figure 20. Synthetic optical-to-optical matching: false color composition of the images before (left) and after registration (right) based on translation and cross-correlation metric.

3.3.2.1.2 Synthetic SAR-to-SAR matching

Elapsed Time: 0.680 seconds

True Transformation: [-76, 52, 0, 1]

Resulting Transformation: [-76.0, 52.0, 0.0, 1.0]

Registration RMSE: 0.0 pixels

As in previous case, Figure 21 shows the reference SAR image, the input SAR image to be registered, and the resulting cross-correlation with its own point of maximum.

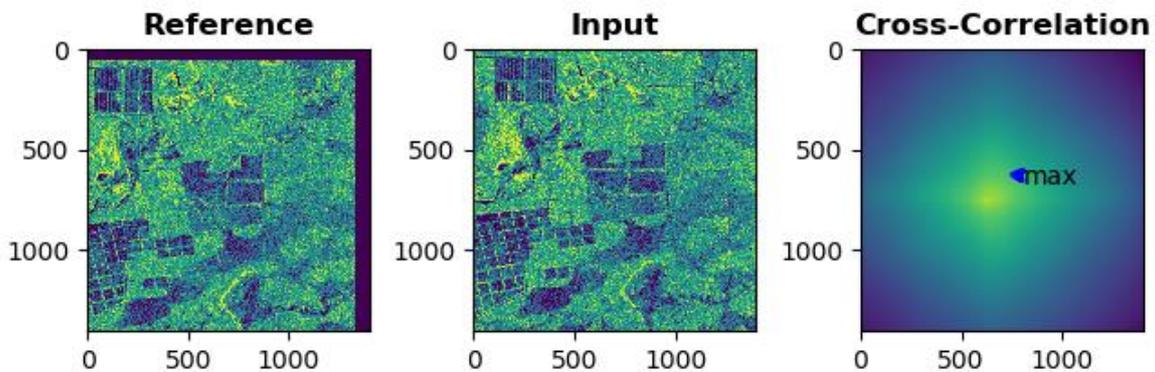


Figure 21. Synthetic SAR-to-SAR matching: reference SAR image (left), input SAR image to be registered (center) and cross-correlation (right).

Figure 22 shows a false color composition of the images before and after registration. The same composition as in the Optical-to-Optical case is used.

In these simple cases of optical-to-optical and SAR-to-SAR registration with only a shift transformation, the cross-correlation proves indeed a valid metric.

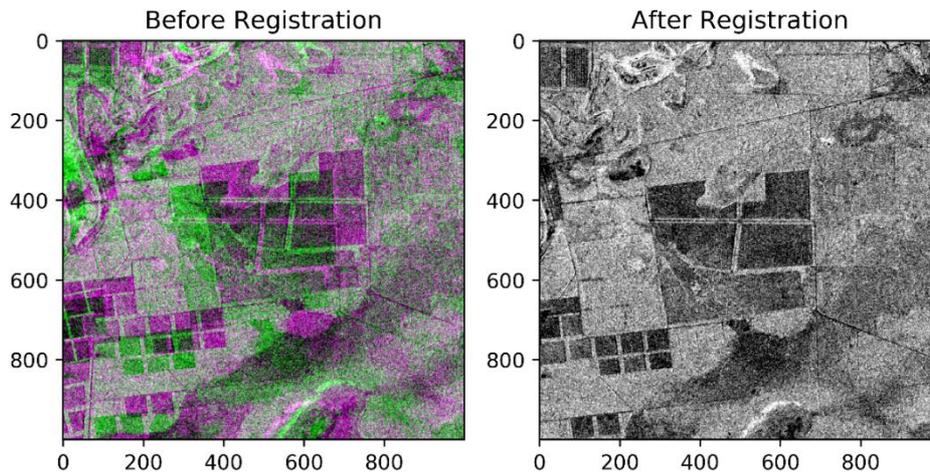


Figure 22. Synthetic SAR-to-SAR matching: false color composition of the images before (left) and after registration (right) based on translation and cross-correlation metric.

3.3.2.1.3 Synthetic Optical-to-SAR matching

Elapsed Time: 0.616 seconds

True Transformation: [-76, 52, 0, 1]

Resulting Transformation: [-0.0, 52.0, 0.0, 1.0]

Registration RMSE: 76.0 pixels

Figure 23 shows the reference optical image, the input SAR image to be registered, and the resulting cross-correlation with its point of maximum.

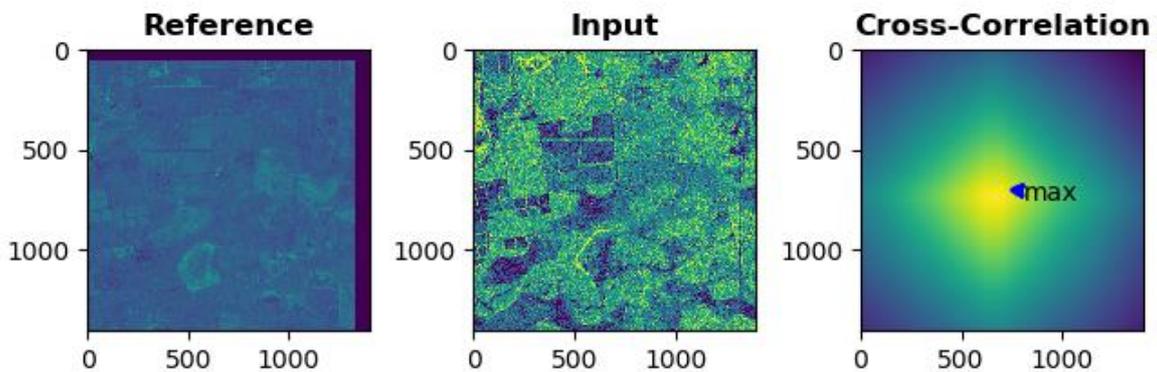


Figure 23. Synthetic Optical-to-SAR matching: reference optical image (left), input SAR image to be registered (center) and cross-correlation (right).

In this case, Figure 24 shows a checkerboard representation of the images before and after registration. On the left, alternate rectangles show either the input or the reference images. On the right, alternate rectangles show either the transformed input or the reference images. A qualitative assessment of the registration accuracy can be achieved by looking at the borders between each rectangle and at the continuity of linear image features across those borders.

Such a result points out the non-satisfactory performance of using cross-correlation as a measure of similarity in the multi-sensor case.

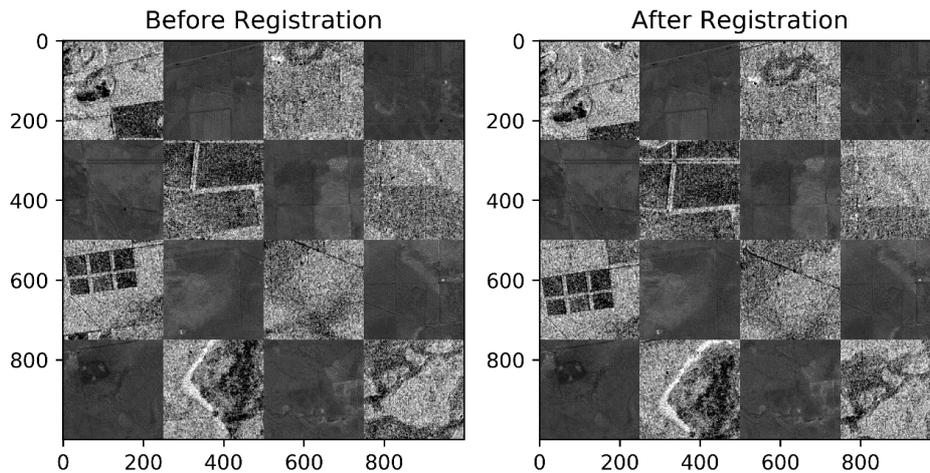


Figure 24. Synthetic Optical-to-SAR matching: checkerboard of the optical and SAR images before (left) and after registration (right) based on translation and cross-correlation.

3.3.2.2 Finding the optimal translation using both cross-correlation and mutual information as metrics

The difference with respect to the previous case is that, in these experiments, the optimal translation is found by maximizing either the cross-correlation or the mutual information. Again, experiments have been carried out with respect to both multi-sensor and single-sensor synthetic scenarios. For the sake of brevity, we report here the results obtained for the SAR-to-SAR and for the Optical-to-SAR registration using the COBYLA optimization method. The experiments with the optical-to-optical and with the Powell optimization methods showed comparable results. It is worth noting that COBYLA usually outperforms Powell's method when the solution is far from initialization (for a complete analysis refer to the experiments with the RST transformation below).

3.3.2.2.1 Synthetic SAR-to-SAR matching

The results obtained by using cross-correlation or mutual information are similar in this synthetic case. What it is worth noting is the required time needed for convergence, which is longer in the mutual information case.

Cross-correlation

This experiment is like the one reported above, but here the cross-correlation is maximized using COBYLA and not by computing the whole cross-correlation function through the FFT. Figure 25 shows the result using the false-color composite defined for the previous single-sensor experiments.

Elapsed Time: 21.172 seconds

True Transformation: [-76, 52, 0, 1]

Resulting Transformation: [-75.9, 51.9, 0.0, 1.0]

Registration RMSE: 5.65 e-4 pixels

Mutual Information

Elapsed Time: 42.221 seconds

True Transformation: [-76, 52, 0, 1]

Resulting Transformation: [-76.0, 51.9, 0.0, 1.0]

Registration RMSE: 5.33 e-4 pixels

The false color composition in Figure 26 points out the accuracy of the registration result obtained. As anticipated, the accuracy when using either cross-correlation or mutual information is similar. The difference is the computation time, which is longer in the mutual information case, thus suggesting a preference for the cross-correlation metric in the single-sensor case.

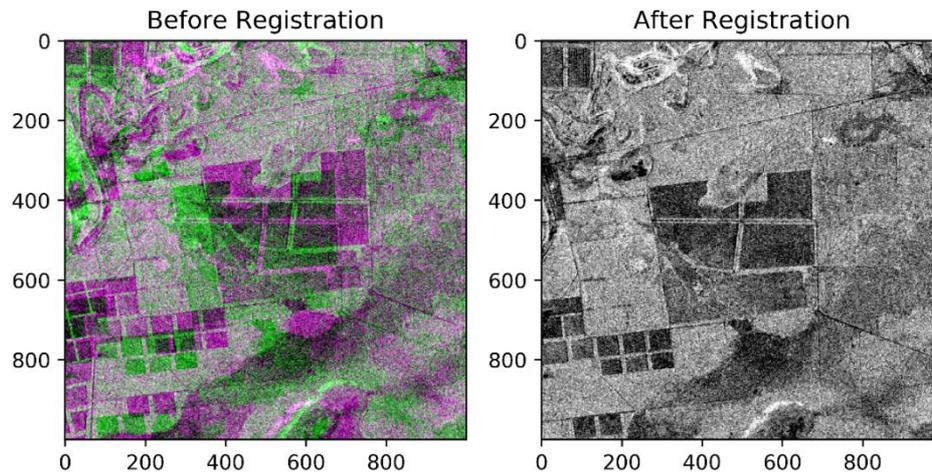


Figure 25. Synthetic SAR-to-SAR matching: false color composition of the images before (left) and after registration (right) based on translation, COBYLA, and cross-correlation metric.

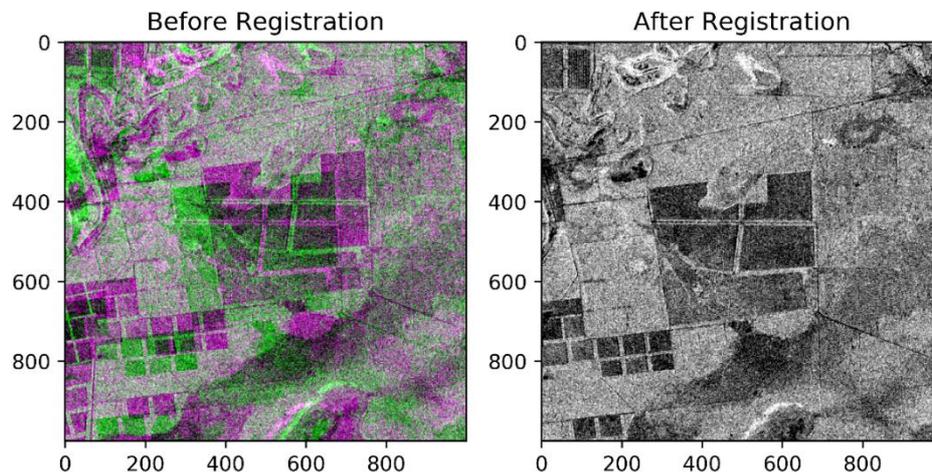


Figure 26. Synthetic SAR-to-SAR matching: false color composition of the images before (left) and after registration (right) based on translation, COBYLA, and mutual information metric.

3.3.2.2.2 Synthetic Optical-to-SAR matching

This experiment, in combination with the one considering cross-correlation through FFT in the optical-to-SAR matching case, highlights the importance of using mutual information as a similarity measure in the multi-sensor scenario [35], [36].

Mutual Information

Elapsed Time: 7.59 seconds

True Transformation: [-76, 52, 0, 1]

Resulting Transformation: [-81.5, 49.6, 0.0, 1.0]

Registration RMSE: 5.99 pixels

The checkerboard representation of the images before and after registration shown in Figure 27 highlights that, even in the multi-sensor case, the mutual information is effective for determining the optimal translation.

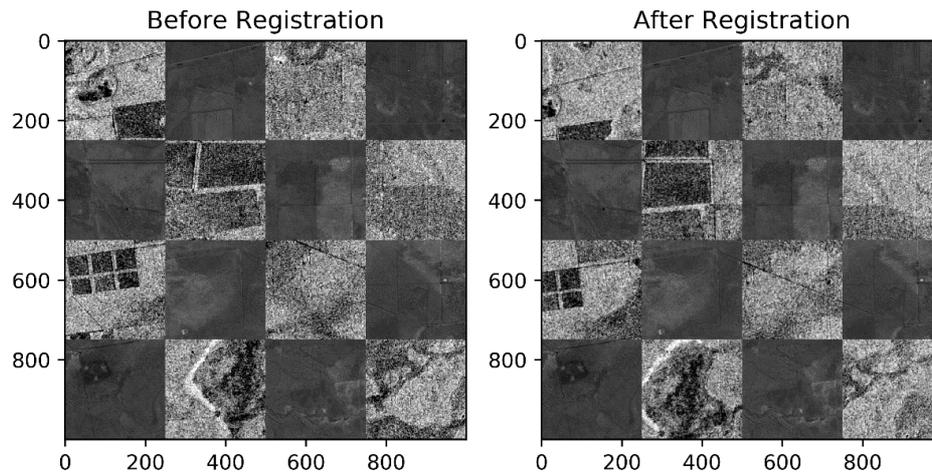


Figure 27. Synthetic Optical-to-SAR matching: checkerboard of the images before (left) and after registration (right) based on translation, COBYLA, and mutual information metric.

3.3.2.3 Finding the optimal Rigid transformation using both cross-correlation and mutual information as metrics

As in the case above, in these experiments, the optimal transformation is found by maximizing either the cross-correlation or the mutual information using COBYLA. The difference is in the geometric transformation, since here the model is considered as a rigid transformation (i.e., translation + rotation).

For the sake of brevity, the experiments reported here consider: (i) using cross-correlation with synthetic single-sensor data; (ii) using mutual information with real multi-sensor data. It is worth noting that, according to its poor results in the multi-sensor case, cross-correlation will not be considered for the further experiments in this document. Conversely, experiment (ii) shows the satisfactory accuracy achievable in the synthetic case through the mutual information, but also confirms that the rigid transformation is often a too restrictive model for the application to real S1-S2 data.

3.3.2.3.1 Synthetic Optical-to-Optical Matching

Cross-correlation

Elapsed Time: 37.39 seconds

True Transformation: [-45, 26, 1.1, 1]

Resulting Transformation: [-44.39, 25.34, 1.14, 1.0]

Registration RMSE: 1.48 pixels

As expected, and as pointed by the following false color composition (Figure 28), an accurate registration was obtained in this case of rigid transformation of a single-sensor image pair.

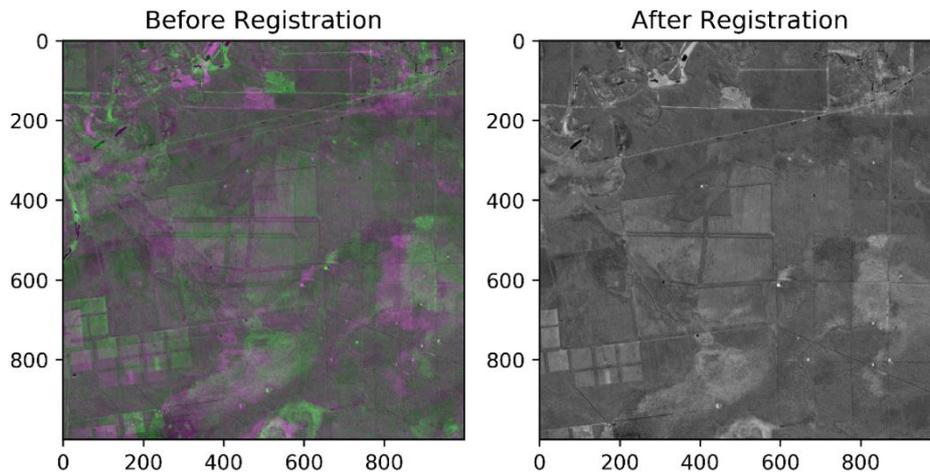


Figure 28. Synthetic Optical-to-Optical matching: false color composition of the images before (left) and after registration (right) based on rigid transformation, COBYLA, and cross-correlation metric.

3.3.2.3.2 Real Optical-to-SAR matching

Mutual information

Elapsed Time: 13.52 seconds

Resulting Transformation: [66.44, -44.43, -0.75, 1.0]

Figure 29 shows a checkerboard representation of the images before and after registration. Well-registered areas are highlighted in green, while areas that still suffer from unprecise registration are highlighted in red.

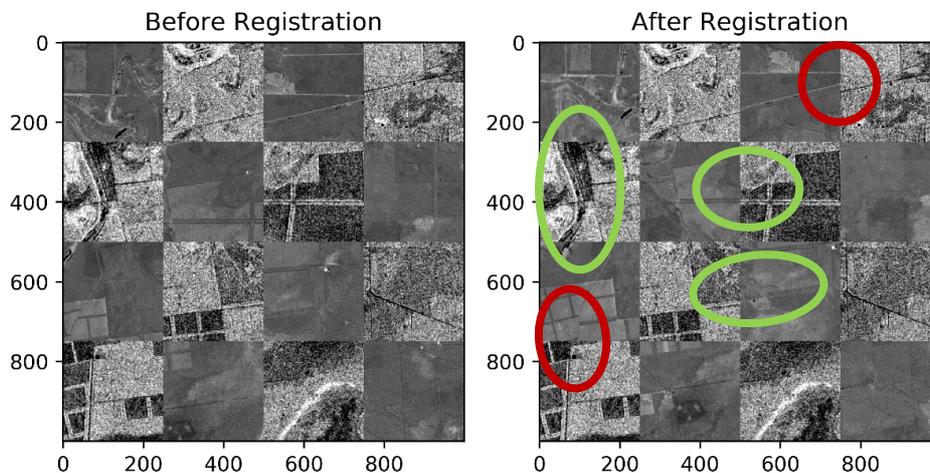


Figure 29. Real Optical-to-SAR matching: checkerboard of the images before (left) and after registration (right) based on rigid transformation, COBYLA, and mutual information metric.

As anticipated above, on one hand, the registration through the mutual information improved as compared to the initial georeferencing. A particularly effective alignment can be noted in the green circles. Nevertheless, the resulting rigid transformation is not globally accurate, as it can be noted by inspecting the crop fields in the lower left corner of the image, together with the other areas highlighted in red.

3.3.2.4 Finding the optimal RST transformation using mutual information as a metric

The following experiments consider the more general case of rotation-scale-translation transformations. The experiments reported in the following sections do not consider anymore the synthetic single-sensor cases to focus on the HRLC multi-sensor scenario.

Also in this case, the performance is evaluated both quantitatively, with synthetic data where the true transformation is known, and visually-qualitatively in the case of the real dataset. The following experiments are also aimed at testing the registration method on different areas of the available Amazonian tile. Three separate areas (1000x1000 pixels each) have been registered using the RST transformation and the COBYLA and Powell's optimization strategies.

As anticipated in the "Quantitative Evaluation" section, it is worth noting here that the Powell's algorithm performs well in the cases where the solution of the registration process falls in a neighborhood of the starting point. Conversely, the COBYLA method, with the initial search radius parameter, allows to better explore the search space and usually grants better convergence properties. This is shown especially in Areas from 2 to 4 in the following experiments, where the solution is significantly far away from the initialization and Powell algorithm fails to converge.

3.3.2.4.1 Synthetic Optical-to-SAR Matching

Elapsed Time: 23.81 seconds

True Transformation: [-45, 42, 2.1, 0.98]

Resulting Transformation: [-35.26, 43.02, 1.82, 0.97]

Registration RMSE: 5.16 pixels

The checkerboard representation of the images before and after registration is shown in Figure 30. For all the experiments within the synthetic optical-to-SAR matching, the COBYLA optimization method has been used.

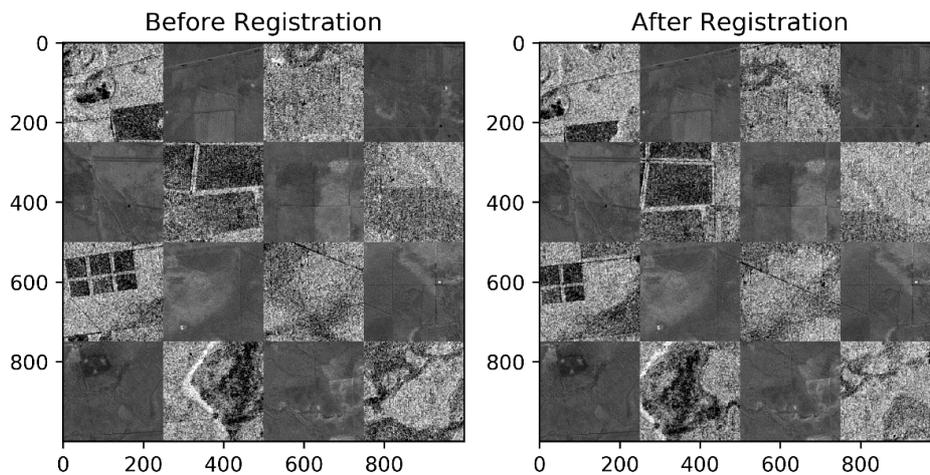


Figure 30. Synthetic Optical-to-SAR matching: checkerboard of the images before (left) and after registration (right) based on RST transformation, COBYLA, and mutual information metric.

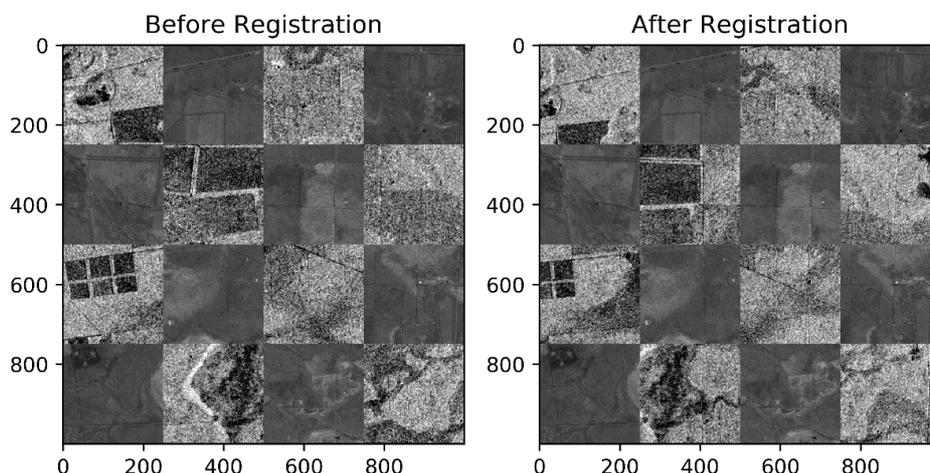


Figure 31. Synthetic Optical-to-SAR matching (scaling factor of 0.98): checkerboard of the images before (left) and after registration (right) based on RST transformation, COBYLA, and mutual information metric.

This result is comparable in accuracy with the case of rigid transformation. Nevertheless, it is worth noting that for the registration process to converge to a qualitative satisfactory result, the scaling factor should not be too large (or small). An experimental example is reported here, where a scaling factor of 0.9 prevented the registration to succeed. During the experiments, all the other parameters were kept constant and the scaling factor was reduced from 0.98 to 0.90. With a scale factor reduced from 0.98 to 0.95 (Figure 31), the rigid transformation still converged to a satisfactory result.

Scaling even further to 0.92 still granted acceptable results (Figure 32).

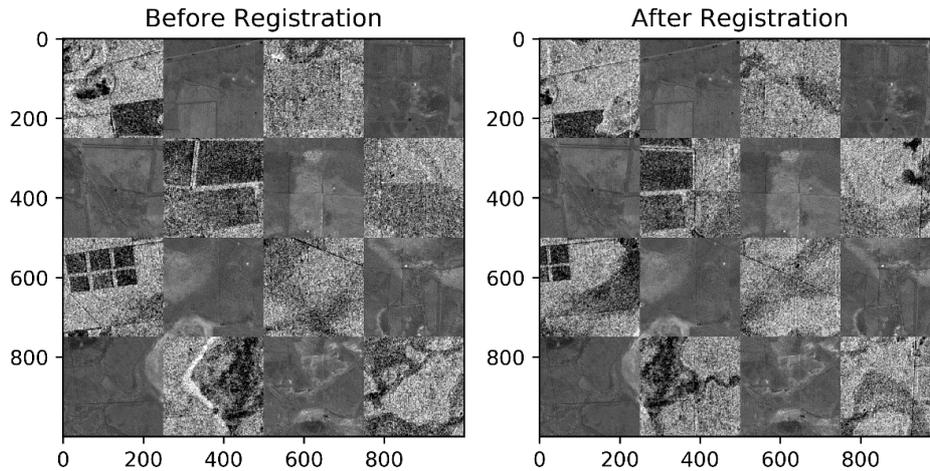


Figure 32. Synthetic Optical-to-SAR matching (scaling factor of 0.92): checkerboard of the images before (left) and after registration (right) based on RST transformation, COBYLA, and mutual information metric.

However, further tests failed to converge effectively. An example is reported in Figure 33, where the scaling factor was chosen equal to 0.9 and the other parameters were kept constant.

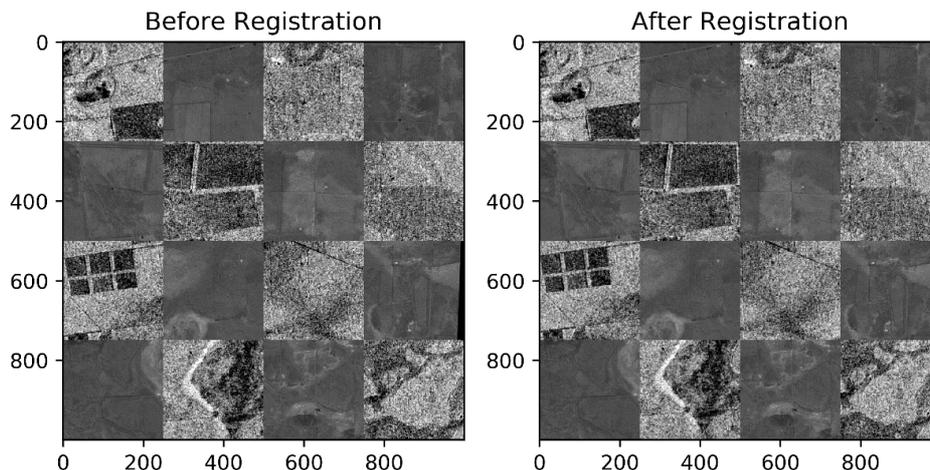


Figure 33. Synthetic Optical-to-SAR matching (scaling factor of 0.90): checkerboard of the images before (left) and after registration (right) based on RST transformation, COBYLA, and mutual information metric.

This analysis is important especially with respect to the COBYLA minimization method. Indeed, such technique allows defining constraints in the search space, thus preventing the scale parameter to assume excessive values. This consideration is also consistent with the project, as the available images are not expected to be characterized by a large-scale difference. Indeed, as the input S1 and S2 data are first georeferenced on pixel grids associated with the same nominal pixel spacing of 10 m, the possible scaling factor is indeed expected to take values in a narrow neighborhood of unity.

3.3.2.4.2 Real Optical-to-SAR Matching: Area 1

Elapsed Time: 13.24 seconds

Resulting Transformation: [76.40, -44.31, -0.17, 0.99]

The checkerboard in Figure 34 points out that this result is comparable with the one obtained in the case of rigid transformation (i.e., only translation and rotation) in the application to the real multi-sensor dataset as well. The scale factor does not have great importance in the registration process of this dataset, as it is suggested by the convergence value of 0.99 and in accordance with the aforementioned comments on the expected behavior of this parameter. COBYLA is used for minimization both for Area 1 and for all the following experiments considering other areas.

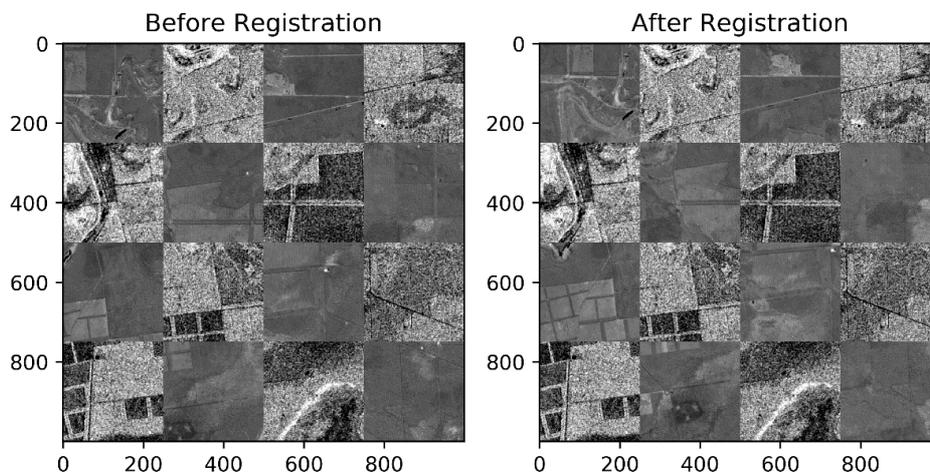


Figure 34. Real Optical-to-SAR matching – Area 1: checkerboard of the images before (left) and after registration (right).

3.3.2.4.3 Real Optical-to-SAR Matching: Area 2

Elapsed Time: 13.39 seconds

Resulting Transformation: [77.52, -62.89, 0.16, 1.03]

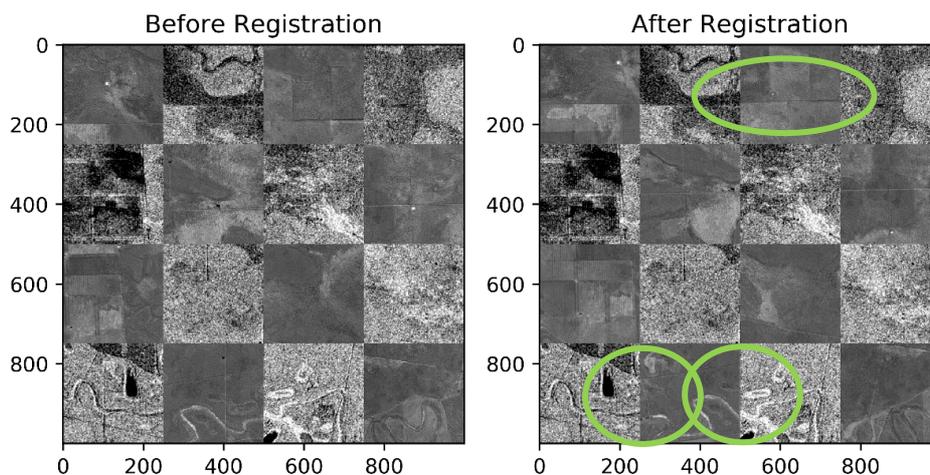


Figure 35. Real Optical-to-SAR matching – Area 2: checkerboard of the images before (left) and after registration (right).

The same comments made for Area 1 hold with regard to Area 2 as well. In particular, the green ellipses in the checkerboard (see Figure 35) emphasize regions where accurate registration of linear image features is especially evident.

3.3.2.4.4 Real Optical-to-SAR Matching: Area 3

In this area, the spatial difference between the reference and input images is large, especially in the translation along the x axis. COBYLA was able to reach an effective registration result, though, while Powell's method failed to converge due to the large distance between the solution and the initialization point. The green and red ellipses superimposed to the following two sets of checkerboards emphasize this significant difference in performance.

COBYLA

Elapsed Time: 15.12 seconds

Resulting Transformation: [222.11, -3.23, 0.48, 1.00]

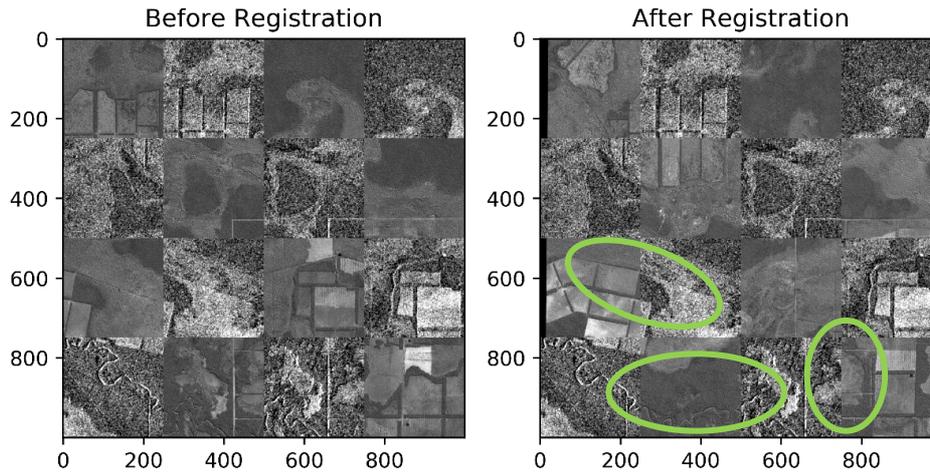


Figure 36 Real Optical-to-SAR matching – Area 3: checkerboard of the images before (left) and after registration (right) using COBYLA.

POWELL

Elapsed Time: 46.29 seconds

Resulting Transformation: [28.07, 16.23, 0.06, 1.00]

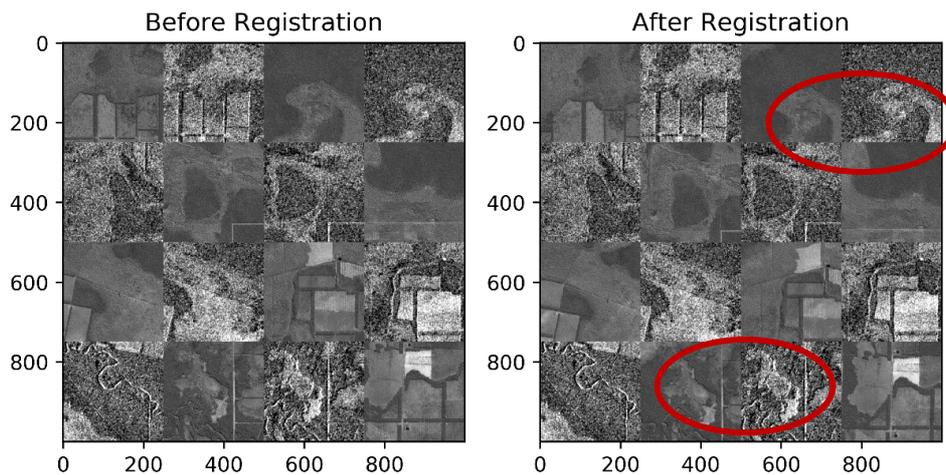


Figure 37 Real Optical-to-SAR matching – Area 3: checkerboard of the images before (left) and after registration (right) using Powell's algorithm.

3.3.2.4.5 Real Optical-to-SAR Matching: Area 4

As for Area 3, also Area 4 is characterized by a large spatial difference, which is even larger than the previous case. Nevertheless, COBYLA succeed to find a good transformation fitting the input and the reference images, and outperformed Powell's method.

Another experimental advantage of COBYLA is a lower time needed for convergence, most probably due in this case to the fact that Powell failed to find a good matching. However, the short convergence time suggests the possibility of tuning online the "initial search radius" parameter. Indeed, it should be possible to implement a new version of such registration procedure that integrates a grid search on that parameter. The resulting similarity metrics could be compared to decide which value of such parameter is able to grant the best registration result for the dataset at hand.

COBYLA

Elapsed Time: 13.33 seconds

Resulting Transformation: [270.68, -55.25, 0.51, 1.02]

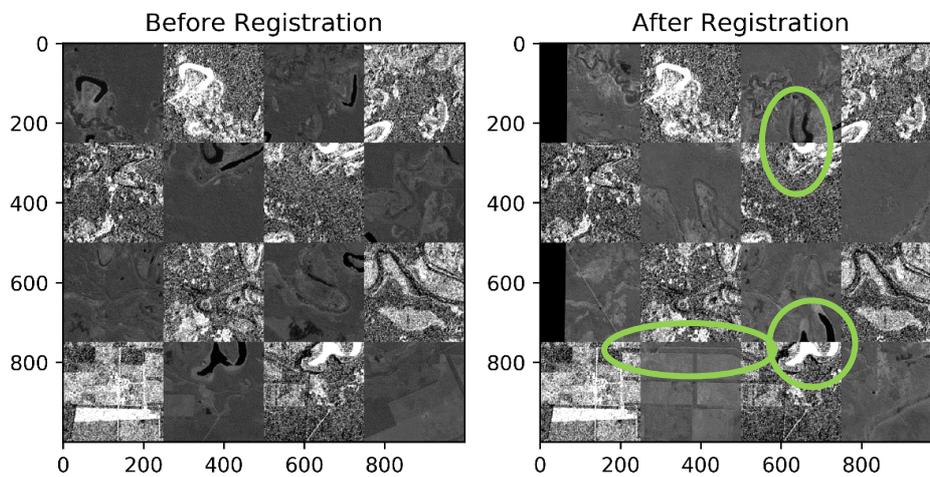


Figure 38 Real Optical-to-SAR matching – Area 4: checkerboard of the images before (left) and after registration (right) using COBYLA.

POWELL

Elapsed Time: 38.93 seconds

Resulting Transformation: [3.63, -1.58, 0.01, 1.00]

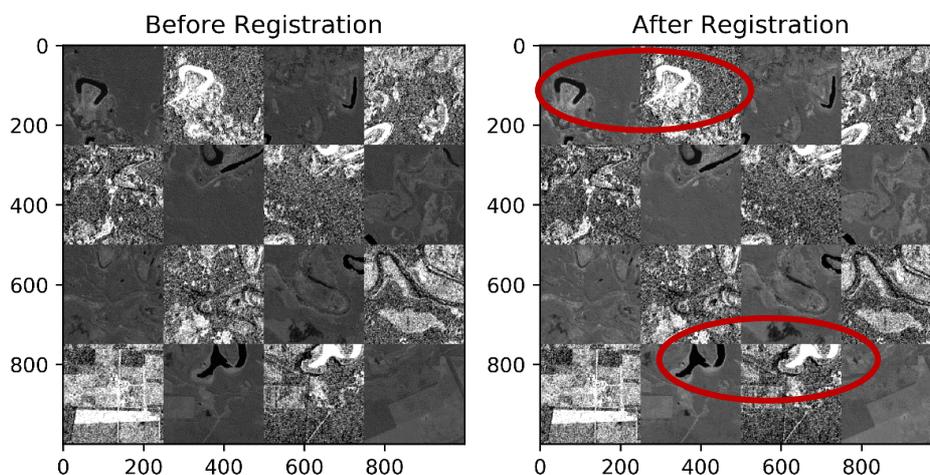


Figure 39 Real Optical-to-SAR matching – Area 4: checkerboard of the images before (left) and after registration (right) using Powell's algorithm.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	42	

3.3.3 Final decision

The experimental analysis conducted with respect to multi-sensor geolocation and the Round Robin dataset pointed out a list of guidelines helpful to decide which registration strategy to pursue and to identify where to focus the future work on the topic. The following list enumerates such consideration and gives some insight in each of the points:

- The choice of the similarity metric. Both cross-correlation and mutual information have been experimented with. The results pointed out that cross-correlation is valuable in the single-sensor case, while mutual information should be preferred in the multi-sensor scenarios. As a result, within the CCI+ HRLC product, mutual information should be prioritized. Different implementations and approximations of such metric will be considered (e.g., Parzen window estimation, Parzen window applied to a random selection of sample to reduce the computation time, etc.). Moreover, other metrics will be studied in the future Round Robin.
- The choice of the geometric transformation. The experiments pointed out that, at least with respect to the available dataset, a simple translation transformation is not enough to find a good match between the input and reference images. However, rigid and RST transformations performed equally, as the available dataset was not affected by a consistent scale factor between the input couple. The last two transformations will therefore be considered in the future, with the exploration of additional higher-order transformations (e.g., affine transformation) in the future Round Robin.
- The choice of the optimization strategy. Both Powell's method and COBYLA performed effectively when the spatial difference between the reference and input images is not large. Conversely, COBYLA outperformed Powell's method in the other case, due to the possibility to control the starting search radius. Future work will focus more on this last minimization method, with details on the possibility of integrating a strategy for the optimization of the starting search radius in the registration process. Other more sophisticated minimization methods (e.g., global search methods like genetic algorithms and simulated annealing) will be considered within the next Round Robin. In this case, the focus will be in analyzing the time versus accuracy tradeoff.

3.3.4 Decision Fusion Methods

Decision fusion methodologies are used in the HRLC project in order to combine the posterior probabilities coming from the disjoint classification of SAR and optical images. Within the first year of CCI+ HRLC, different families of decision fusion methods have been implemented and experimentally compared, including consensus-theoretic methods and fusion strategies based on Markovian modelling (i.e., Markov random fields), and combined with a class-specific matching rule. Experimental comparisons have been focused on the Amazon round robin area, thanks to the availability of both the optical-based and the SAR-based classification outputs on this area.

Specifically, the posterior probabilities coming from the optical and SAR processing chains are quantized and coded into unsigned integers using 8 bits per class and per pixel. This choice is aimed at minimizing memory requirements without any expected loss in appreciable precision. As discussed in the deliverables of the previous milestones, the sets of classes that can be accurately discriminated by using optical and SAR data exclusively generally do not coincide. SAR data are expected to well discriminate especially built-up classes and water bodies. Accordingly, SAR and optical classification algorithms work on different sets of classes (see also the previous sections of the present document). Decision fusion methodologies are aimed at fusing posterior probabilities related to the common classes. Hence, a class-specific combination rule has been devised to take this into account. Specifically, with respect to the HRLC project, for what concerns the Amazon round robin areas, the posterior probabilities coming from optical-based classification relate to classes 1, 3, 5, 6, 7, 8, 10, 15 (Figure 40).

Value	Label	Color
0	No data	Black
1	Evergreen broadleaf tree	Green
2	Evergreen needleleaf tree	Dark Green
3	Deciduous broadleaf tree	Bright Green
4	Deciduous needleleaf tree	Olive Green
5	Shrubland	Brown
6	Permanent cropland	Tan
7	Annual summer cropland	Yellow
8	Grassland	Orange
9	Lichens and mosses	Pink
10	Permanent water bodies	Blue
11	Permanent snow and ice	White
12	Beaches dunes and sands	Light Brown
13	Bare soils	Yellow
14	Bare rock	Grey
15	Built-up areas	Red

Figure 40 Class legend for the classification experiments with optical, SAR, and multi-sensor data

Therefore the optical set of classes is $\Omega_o = \{1,3,5,6,7,8,10,15\}$. SAR classification is performed on 7 classes: 1, 3, 5, 6, 7, 8, 10 in the previous table. The SAR class set is therefore, in the Amazonian case, a strict subset of the optical one ($\Omega_s = \{1,3,5,6,7,8,10\} \subset \Omega_o$). For that reason, optical posterior probabilities are divided into two subsets:

- Ω_s , including the 7 common classes (1,3, 5, 6, 7, 8, 10).
- $\bar{\Omega}_s = \Omega_o - \Omega_s$, including class 15 only.

The next sections focus on the consensus (non-contextual) and Markovian (contextual) approaches. The class-specific combination is explained together with the former.

3.3.4.1 Consensus Theory and class-specific combination

Consensus theory [37],[38] involves general procedures with the goal of combining multiple probability distributions to summarize their estimates in a non-contextual manner. Under the assumption that both the SAR and the optical 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 most common consensus theory methods are linear opinion pool (LOP) and logarithmic opinion pool (LOGP) [37],[38]. Both these methods were implemented and tested within the first year of CCI+ HRLC in combination with the aforementioned class-specific rule.

Let $x = [O, S]$ be the input data vector on a generic pixel, resulting from the stacking of optical (O) and SAR (S) individual feature vectors, and let ω_j be the j th information class ($j = 1,2, \dots, M$). The LOP functional can be expressed as:

$$C(\omega_j|x, \Omega_s) = \alpha P(\omega_j|O, \Omega_s) + \beta P(\omega_j|S, \Omega_s)$$

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	44	

where $P(\omega_j|O, \Omega_s)$ is the optical posterior probability of ω_j conditioned to the common subset of classes Ω_s and $P(\omega_j|S, \Omega_s)$ is the SAR posterior probability conditioned to the same subset Ω_s . In the Amazon RR experiments, $P(\omega_j|S, \Omega_s) = P(\omega_j|S)$ because Ω_s includes all classes in the SAR legend, while $(\omega_j|O, \Omega_s)$ can be derived from the posteriors $P(\omega_j|O)$ associated with the classes in Ω_o through straightforward probability manipulations. α and β are optical and SAR source-specific weights, respectively, and control the relative influence of the two sources.

According to the same definitions, the logarithmic opinion pool (LOGP) functional can be defined as:

$$\mathcal{L}(\omega_j|x, \Omega_s) = \alpha \log P(\omega_j|O, \Omega_s) + \beta \log P(\omega_j|S, \Omega_s)$$

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, then by definition $\mathcal{L}(\omega_j|x, \Omega_s) = 0$. This dramatic behaviour is a drawback when the single-source predictions are very inaccurate and can be generated also 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 $P_{\mathcal{F}}(\omega_j|x, \Omega_s)$. In the case of LOP, $P_{\mathcal{F}}(\omega_j|x, \Omega_s)$ is computed from $\mathcal{C}(\omega_j|x, \Omega_s)$ by just re-normalizing so that the sum over all $\omega_j \in \Omega_s$ 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:

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

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_s) can be defined according to the total probability theorem:

$$\begin{aligned} P_{\mathcal{F}}(\omega_j|x) &= P(\omega_j|x, \Omega_s)P(\Omega_s|x) + P(\omega_j|x, \bar{\Omega}_s)P(\bar{\Omega}_s|x) = \\ &= P_{\mathcal{F}}(\omega_j|x, \Omega_s)[1 - P(\bar{\Omega}_s|O)] + P(\omega_j|O, \bar{\Omega}_s)P(\bar{\Omega}_s|O), \end{aligned}$$

where the quantities associated with $\bar{\Omega}_s$ are conditioned only on the optical observations O .

It is worth noting that, while this class-specific rule is consistent with the peculiarities of the optical and SAR classification results on the Amazon round robin data set, this combination rule is applicable to all cases in which one of the two sources discriminates among a larger set of classes than the other source – as expected in the fusion of optical and SAR data.

3.3.4.2 Markov Random Fields

Markov random fields (MRFs) can include contextual information in the form of class interactions between neighbouring pixels. An MRF is determined by an energy function, composed of two main terms: one characterizing class likelihood at the pixel level (depending on per-class scores obtained by the optical and SAR classification methods), and another promoting label smoothness in a local neighbourhood. This means that the model encourages two neighbouring pixels to be labelled with the same class and that the fusion processing stage allows incorporating spatial-contextual information as well.

Define the regular pixel lattice as I , and let y_i be the class label of the i -th pixel ($y_i \in \Omega_o, i \in I$). The MRF consider y_i as a 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 [39]. In this case, ∂i was chosen to be made of the four pixels adjacent to the i -th pixel (four-connected).

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

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

where $X = \{x_i\}_{i \in I}$ is the random field of all optical and SAR observations ($x_i = [O_i, S_i]$ on each pixel $i \in I$), $P_{\mathcal{F}}(y_i|x_i)$ is the pixelwise fusion output described in the previous section, and the spatial energy contribution has been modelled using the Potts model [39].

In the application of MRF-based methods to decision fusion, special focus has been devoted to the minimization of the energy function U with respect to the random field Y of the class labels. For that reason, the iterated conditional mode (ICM) algorithm has been applied, since it represents an efficient trade-off between accuracy and computational burden [40]. Moreover, a lot of effort was spent to make the implementation of ICM as fast as possible. This was achieved by making use of convolution-like operations only, which have a very fast implementation in the Python environment. That way it was possible to speed-up the computations, at a cost of a slight increase in memory occupation, and make MRF feasible to be applied to large scale dataset as the one of HRLC.

3.3.5 Qualitative evaluation

The aforementioned approaches to multi-sensor fusion were experimentally evaluated on the case study of the Amazonian S2 tile 21kuq and of the S1 images that spatially overlap with it. The area of spatial intersection between these S1 and S2 sources was obviously considered for the fusion experiments. Analogous results with other input data will be progressively generated as soon as the corresponding optical and SAR inputs become available.

Figure 41 collects a detail of the classification maps obtained by the optical and SAR processing chains (the color legend is reported at the beginning of the current section).

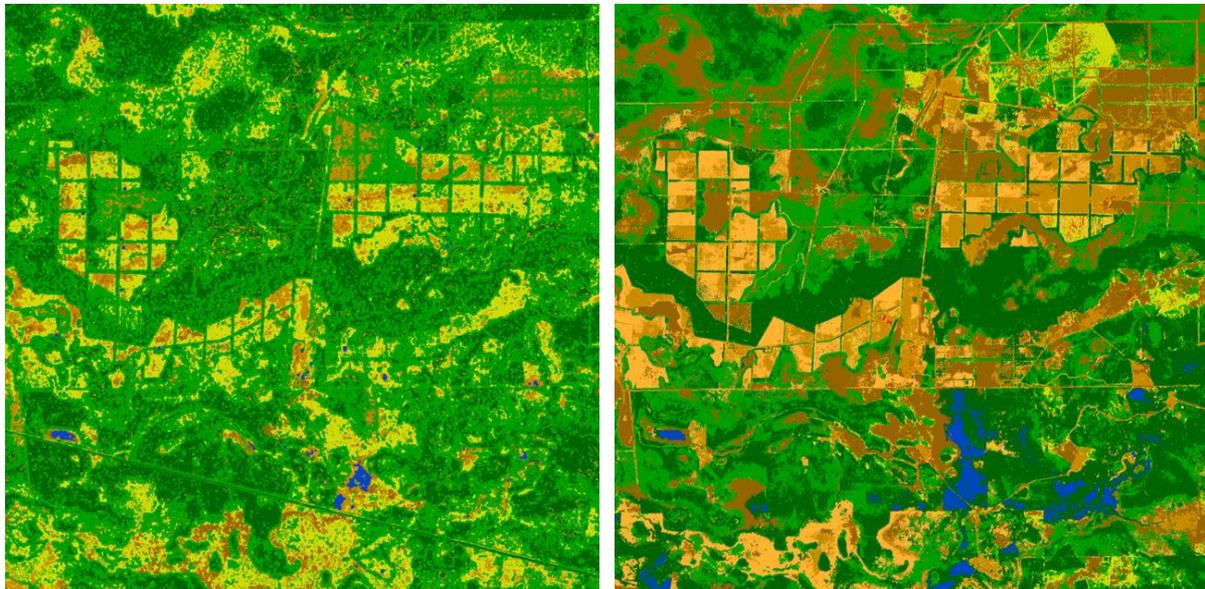


Figure 41 Details of the classification maps obtained from the classification of optical (left) and SAR data (right).

As expected, the two results differ in spatial regularity (smoother in the optical-based case than in the SAR-based case, because of the influence of residual speckle), class legend (see above) and in the labels assigned by the classification of the two sources to several areas. While optical data are expected to be fundamental in discriminating the considered land cover classes, the following figure provides an example of a case in which residual impact of cloud cover may affect the optical-based map but obviously not the SAR-based map.



Figure 42 Details of the classification maps obtained from the classification of optical (top left) and SAR data (top right), indicating the impact of a residual cloud on the former. A true-color composite of the optical image is also shown (bottom).

3.3.5.1 Results - Consensus Theory

The results of applying linear opinion pool to the posterior probabilities associated with the aforementioned maps is shown in Figure 43.

In particular, LOP was applied by giving slightly larger weight to the optical source than to the SAR source ($\alpha = 0.6$ and $\beta = 0.4$), consistently with the expected reliability of the corresponding land cover classification output. Indeed, linear opinion pool made minor changes as compared to the optical-based classification map. It is however the fastest method among those described above, requiring around 30 seconds on a 8200x8200 tile on a standard desktop machine (no GPU). In particular, it does not compensate for the impact of the aforementioned residual cloud on the classification map. Indeed, the stronger influence of the optical-based result than of the SAR-based result on the fusion map is consistent with both the expected contribution of each type of sensor to class discrimination and of the different behaviour of the corresponding posteriors. The posterior probabilities obtained from the optical source on a given pixel most often indicate the most probable class membership with higher confidence as compared to the posteriors obtained from the SAR source.

Accordingly, they more strongly affect the pixelwise decision fusion outcome. The results of logarithmic opinion pool are shown in Figure 44.

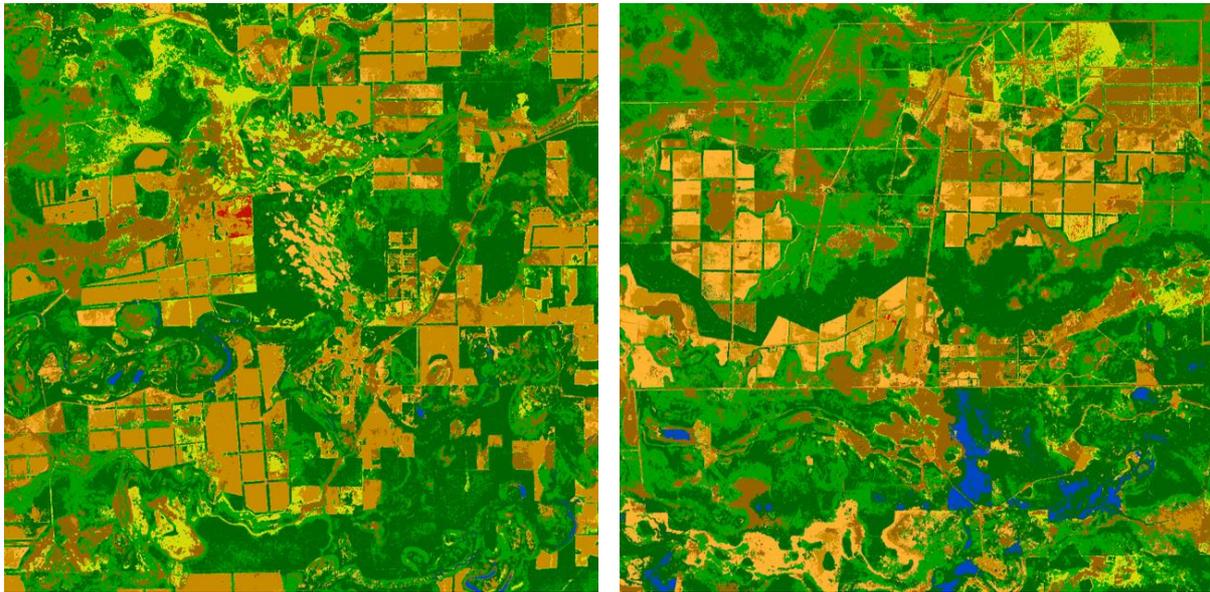


Figure 43 Details of the classification map obtained by LOP in the same areas as in Figure 41 (left) and Figure 42 (right; cloudy area on the right)

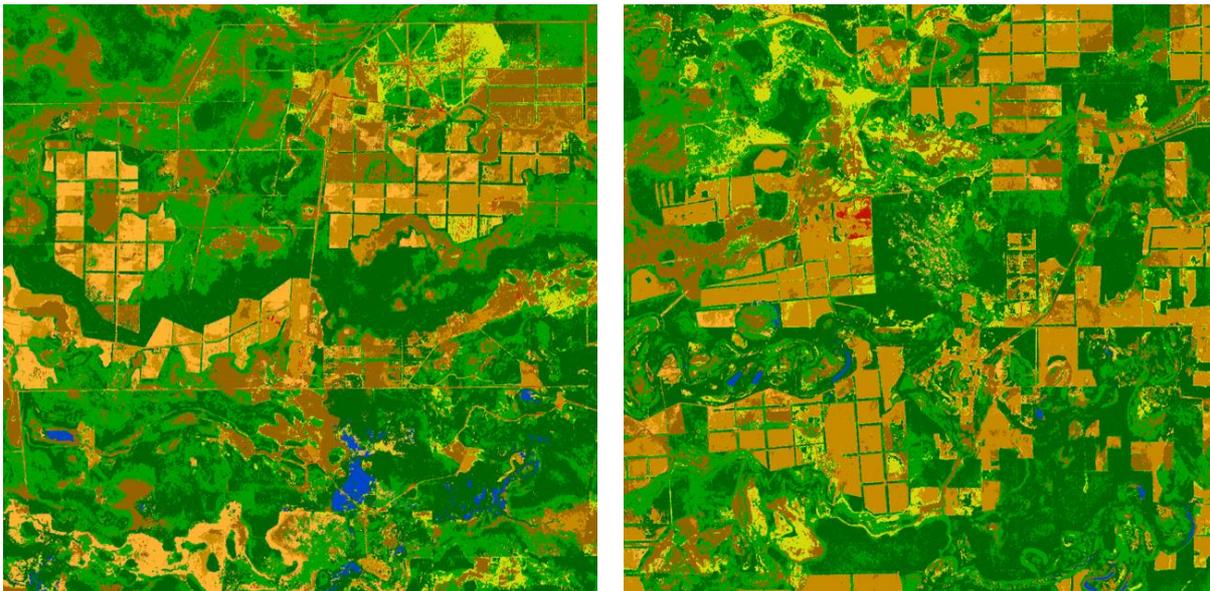


Figure 44 Details of the classification map obtained by LOGP in the same areas as in Figure 41 (left) and Figure 42 (right; cloudy area on the right)

In the case of the logarithmic opinion pool, while the overall classification output is still significantly influenced by the optical source, it is worth noting that the erroneously labelled residual cloud is mostly corrected thanks to the SAR source, thus confirming the relevance of fusing the two separate classification results. LOGP required around 50 seconds to run on the 8200x8200 tile. Hyper-parameter setting was addressed for LOGP analogously to the case of LOP.

In both cases of LOP and LOGP, the spatial regularity of the output classification map was similar to that of the optical-based result. On one hand, this suggests that the two fusion strategies are not negatively affected by the impact of residual speckle on the SAR-based result. On the other hand, further smoothness is obviously not achieved by LOP or LOGP, because they are fully non-contextual approaches.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	48	

3.3.5.2 Results – Markov Random Fields

While the previous methods make no use of contextual information, Markov random fields model this type of information explicitly. Here, consistently with the results discussed in the previous subsection, the adopted MRF used a unary energy term derived according to LOGP, while the contextual term depending on a 4-connected neighbourhood was given by the Potts model. A unary based on LOP is not presented for brevity as it could benefit less than LOGP from the fusion of the two sources, as argued in the previous subsection. The classification map obtained by applying ICM to this Markovian energy is shown in Figure 45.

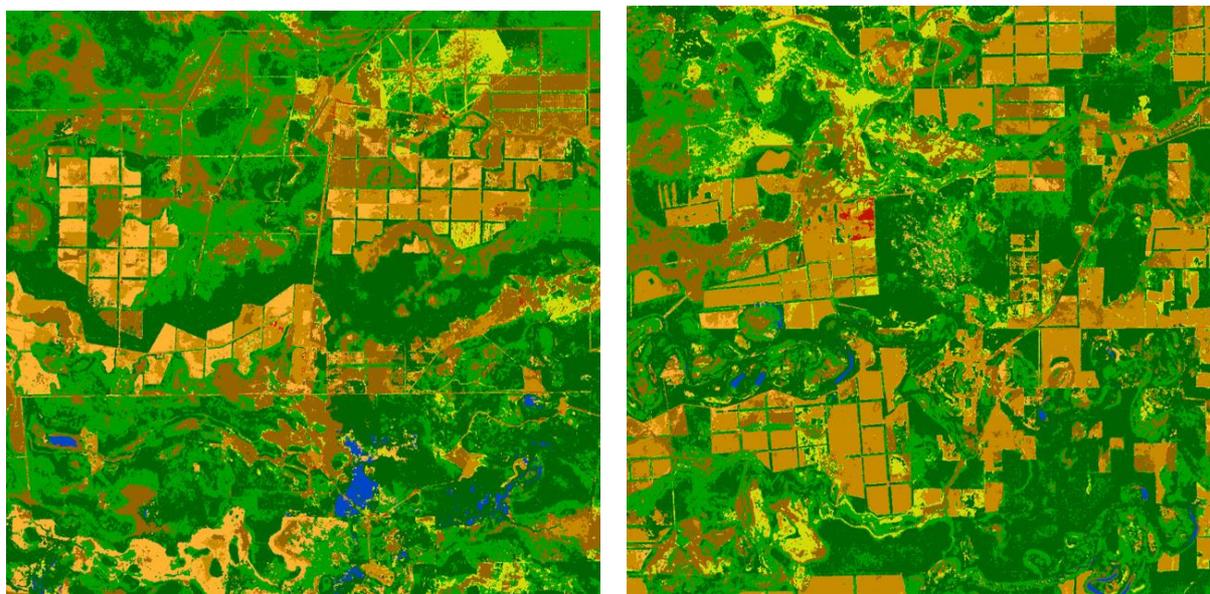


Figure 45. Details of the classification map obtained by MRF-ICM in the same areas as in Figure 41 (left) and Figure 42 (right; cloudy area on the right).

MRF required around 7 minutes to run on the 8200x8200 tile. It is possible to appreciate the spatial regularization achieved with MRF as compared to the previous classification map. On one hand, the comments made in the previous section with regard to LOGP and its relation to the optical-based and SAR-based individual outputs hold in this case as well. In particular, the impact of the residual cloud is significantly mitigated in the MRF result, too. On the other hand, the Markovian output is overall smoother than the aforementioned pixel-wise results. In particular, these maps were obtained by setting the spatial parameter γ to 0,8. Stronger spatial regularization can as well be achieved increasing the value of γ . That way, a spatially even smoother result can be achieved, although at the cost of a minor degradation of small-scale spatial details. On one hand, this trade-off can be tuned by applying automatic hyper-parameter optimization methods that determine an optimal value for γ through mean-square-error or Bayesian approaches [5], [6]. Moreover, alternate spatial models, including contrast sensitive conditional random fields, mitigate the possible degradation in small-scale details. On the other hand, the opportunity to explicitly tune the trade-off between spatial smoothness and detail provides further flexibility in the generation of an output HRLC product that meets the requirements of the climate community.

3.3.6 Final decision

Both LOP and LOGP are fast since they are one-shot non-iterative pixel-wise methods. However, data fusion performed that way does not benefit from contextual information, which can be crucial in order to remove noise and favour the spatial regularity of the output map. Between LOP and LOGP, the experiments suggest a stronger potential of the latter in taking benefit from both input sources.

Markov Random fields can also capture local spatial information, thus providing more spatial regularization at a cost of a comparatively higher, yet still quite short, computational time.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	49	

Overall, LOGP and MRF-ICM are identified here as appropriate solutions for HRLC, in combination with the aforementioned class-specific combination strategy. In this framework, further development may regard the automatic optimization of the hyper-parameters of these methods (weights of the various information sources), the integration of edge-preserving / contrast-sensitive spatial terms in the Markovian approach, and the evaluation of the potential of alternate energy minimization algorithms based on graph theory. The last item is possibly promising from the viewpoint of classification accuracy, although at the cost of significantly longer computation time. This time-vs-accuracy trade-off will be addressed carefully.

4 Classification algorithms and procedures (year 2)

4.1 Optical data processing

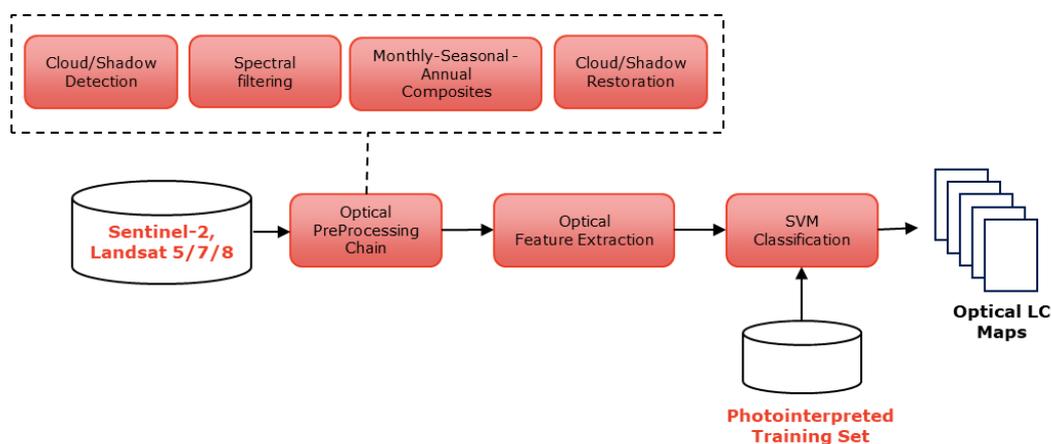


Figure 46. Optical data processing chain for the prototype production of the HR LC map obtained by classifying the time series of Sentinel 2 data.

Figure 46 depicts the optical data processing chain for the prototype production of the HR LC maps obtained by classifying the time series of Sentinel 2 L2A data for the static map of 2019 and the time series of Landsat L2 data for the historical maps of 1990, 1995, 2000, 2005, 2010 and 2015. The images are first pre-processed in order to remove spectral outliers and to detect clouds and shadows. Then, monthly, seasonal and annual composites are generated according to the availability of cloud free images. Finally, the missing data (caused by cloud and shadow coverage) are restored. Due to the missed availability of training data, the team has defined the pixel-based photo-interpretation activity to ensure the production of a representative and informative training set. To provide spatial information to the classifier the textural features (standard deviation and gradient) have been computed. Finally, we performed the classification by using SVM on the times series of the considered composites. The SVM results have been compared with state of the art deep based LSTM classifier.

4.1.1 Satellite images

The considered study areas are located into the following four thematic regions (according to tiling grid of Sentinel-2 products):

- Amazonia area for 21KUQ Sentinel-2 tile;
- Amazonia area for 21KXT Sentinel-2 tile;
- Siberia area for 42WXS Sentinel 2 tile;
- Africa area for 37PCP Sentinel-2 tile.

2019 Sentinel-2 L2A images downloaded through the Copernicus Open Access Hub (<https://scihub.copernicus.eu/dhus/#/home>) are used over the four CCI-HR LC areas for the year 2019. The complete list of images is given in Table 4.

Table 4. List of Sentinel-2 data

Area	Satellite	# Products	Date list (2019y)
Amazonia – 21KUQ	S2	41	2019-01-17;2019-01-22;2019-02-01 2019-02-06;2019-02-11;2019-02-16 2019-02-21;2019-03-03;2019-03-08 2019-03-23;2019-03-28;2019-04-07 2019-04-07;2019-04-12;2019-04-17 2019-05-07;2019-05-17;2019-05-27 2019-06-06;2019-06-26;2019-07-06 2019-07-16;2019-07-21;2019-07-31 2019-08-05;2019-08-15;2019-08-20 2019-08-30;2019-09-14;2019-09-19 2019-09-24;2019-09-29;2019-10-09 2019-10-24;2019-10-29;2019-11-18 2019-11-23;2019-11-28;2019-12-18 2019-12-23;2019-12-28
Amazonia – 21KXT	S2	53	2019-01-14;2019-02-03;2019-02-08 2019-02-18;2019-02-23;2019-03-05 2019-03-30;2019-03-30;2019-04-09 2019-04-19;2019-04-24;2019-04-24 2019-04-29;2019-05-09;2019-05-09 2019-05-19;2019-05-29;2019-06-08 2019-06-13;2019-06-13;2019-06-18 2019-06-23;2019-06-23;2019-06-28 2019-07-08;2019-07-08;2019-07-13 2019-07-23;2019-07-28;2019-08-07 2019-08-12;2019-08-12;2019-08-17 2019-08-22;2019-08-22;2019-09-06 2019-09-16;2019-09-16;2019-09-26 2019-10-01;2019-10-11;2019-10-11 2019-10-26;2019-10-31;2019-11-05 2019-11-20;2019-11-20;2019-11-30 2019-12-15;2019-12-20;2019-12-20 2019-12-25;2019-12-30
Siberia – 42WXS	S2	11	2019-06-23;2019-07-03 2019-07-13;2019-07-15;2019-07-16 2019-07-18;2019-07-23;2019-07-25 2019-08-10;2019-08-15;2019-08-22
Africa – 37PCP	S2	56	2019-01-02;2019-01-12;2019-01-17;

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	51	

			2019-01-22;2019-01-27;2019-02-01; 2019-02-06;2019-02-11;2019-02-16; 2019-02-21;2019-02-26;2019-03-03; 2019-03-08;2019-03-13;2019-03-18 2019-03-23;2019-03-28;2019-04-02 2019-04-07;2019-04-12;2019-04-17 2019-04-27;2019-05-02;2019-05-07 2019-05-12;2019-05-17;2019-05-27 2019-06-11;2019-06-21;2019-06-26 2019-07-06;2019-07-16;2019-08-15 2019-08-20;2019-08-25;2019-08-30 2019-09-04;2019-09-19;2019-09-24 2019-09-29;2019-10-04;2019-10-09 2019-10-14;2019-10-24;2019-10-29 2019-11-03;2019-11-13;2019-11-18 2019-11-23;2019-11-28;2019-12-03 2019-12-08;2019-12-13;2019-12-18 2019-12-23;2019-12-28
--	--	--	--

4.1.2 Method/algorithm/technique

Within this Section, several methods are presented and compared as candidate approaches to be developed and implemented in the optical image processing chain according to Figure 46.

4.1.2.1 Optical data pre-processing

This step has the purpose of generating a time series of composites able to characterize the HR LC classes. We consider all the Sentinel-2 L2A images with cloud coverage lower than 40%. First, cloud and shadow masks are detected by using the Sen2Cor provided by ESA [8], which provides cloud and shadow masks. According to the amount of cloud-free images available, the team decided to generate monthly composites for Amazonia and Africa and annual composite for Siberia. This condition allows us to mitigate cloud occlusions problem and minimize the processing resources. In the case of a small cloud or shadow gaps they were filled with linear temporal gap filling approach (see ATBD 3.0)

The peculiar multi-resolution property of Sentinel 2 images involves four spectral bands acquired at a spatial resolution of 10 m, six spectral bands acquired at a spatial resolution of 20 m and three spectral bands acquired at a spatial resolution of 60 m. Because the 60 m spectral bands are mainly dedicated to atmospheric corrections and cloud screening [10], only the 10 and 20 m bands are used to produce the HR LC maps. A nearest neighbour interpolation technique is used to match the spatial resolution of the 20 m bands to the 10 m ones for the entire time series. The nearest neighbour interpolation technique has the drawback of generating smoothed images, thus losing in sharpness with respect to more sophisticated interpolation technique such as High Pass Filter (HPF). However, no new values are calculated by interpolation. This condition allows us to keep the original spectral information recorded by the sensor. Finally, we perform a spectral outlier detection and removal by discarding the pixels having values higher than the 0.999 quantile and lower than the 0.001 quantile of the spectral band. A radiometric normalization is eventually applied to the interpolated images so that each spectral band is rescaled between zero and one.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	52	

4.1.2.2 Classification

Automatic classification is a crucial processing step to produce accurate LC maps. The selected classification algorithm must achieve the best trade-off between classification accuracy and computational burden due to the need of processing a huge amount of data.

By analysing the recent literature, the team identified two successful core approaches to the classification. By comparing the results obtained in the RR tiles, the Support Vector Machine (SVM) classifier [14] proved its effectiveness to generate HR LC maps. SVM classifiers are based on kernel methods that have been extensively employed for the classification of RS data. In the considered implementation, we exploit SVM with Gaussian Radial Basis Function (RBF) kernels because of its capability of dealing with noisy samples in a robust way and to produce sparse solutions. A feature selection step is performed to detect the feature subspace where the LC classes are more discriminable. In the considered implementation, a Sequential Forward Floating Selection (SFFS) method based on the Jeffreys-Matusita distance as separability criterion is used [15]. The optimal kernel parameters (i.e., the regularization parameter C and the spread of the kernel γ) are selected by a 3-fold cross-validation. The LC maps generated by the SVM classifier are compared with the state of the art deep based Long Short Term Memory architecture (LSTM) [19]. LSTM is a special type of RNN, which has been proposed to overcome the problem of the vanishing gradient on long data sequences. In particular, by using a forget gate's vector, LSTM has a better control of the gradients values at each time step, thus avoiding the gradient vanishing in a long TS. LSTM is able to catch the temporal dependency of sequential data, since the network is using feedback loops connected to their past decisions. LSTM classifier is very promising for classification of dense time series of images. Moreover, it has proven to be able to handle the presence of clouds [16],[17], [18]. Nevertheless, as a typical deep based architecture LSTM needs a much bigger training set for its proper training. Thus, apart from photo-interpreted labelled samples, we have included additional samples generated by the weak training set analysis (see ATBD 3.0). Finally, since LSTM is performing well on long time-series, we have considered two LSTM models: first trained on the whole time-series of images (reported in the Table 4) and second trained on the time series of monthly composites. Considering SVM, the photo-interpreted training sets generated for all the considered study areas were used to train the model. A qualitative analysis was carried out to determine the classifier able to provide the best classification results.

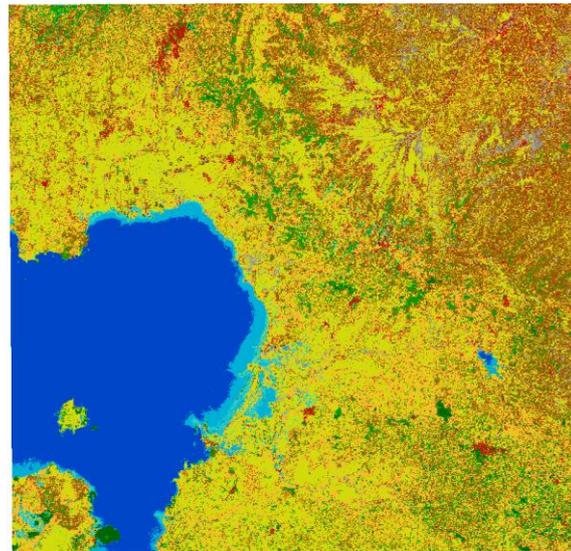
4.1.3 Qualitative evaluation

In the following examples of qualitative analysis performed in the different study areas are reported.

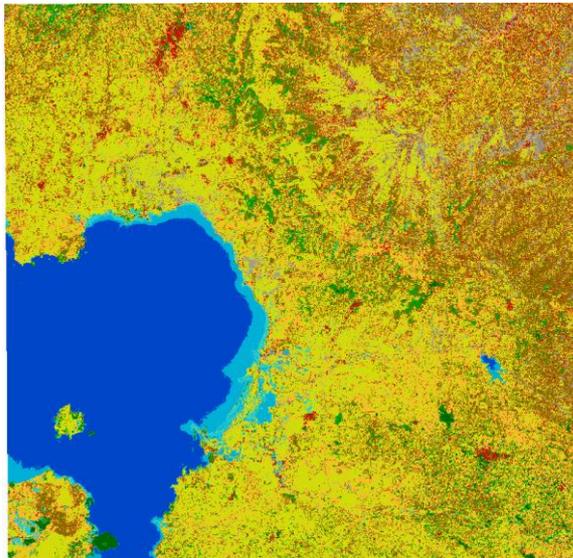
Figure 47 reports a comparison of classification results obtained in Africa (tile 37PCP) in macro scale. It can be noticed that all three classifiers provide similar results. This is particularly evident for the two HRLC maps generated by the LSTM: on the whole time series (see Figure 47 b) and on the time series of monthly composites (see Figure 47 c). Although the SVM map (see Figure 47 d) is quite similar to the LSTM maps, there can be noticed a significant difference in recognition of shrub cover evergreen class. Moreover, there is also a disagreement between seasonal vs permanent water. Overall, classifiers accurately retrieve the geometrical detail of the scene, thus recovering the river, lakes and cities. To identify the best classifier, we compared classifiers at the areas of disagreement.



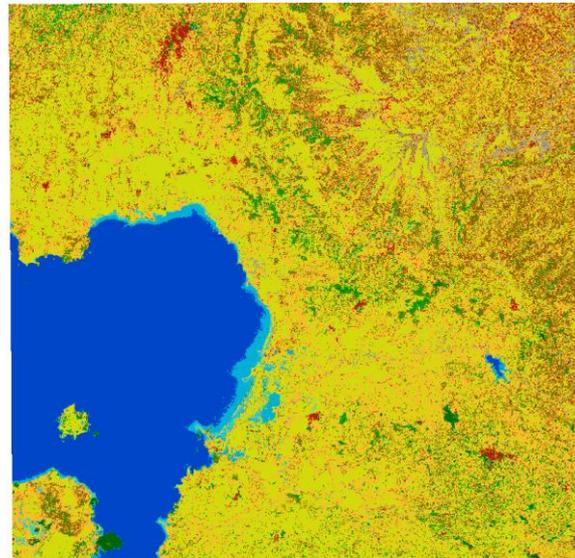
(a)



(b)



(c)



(d)

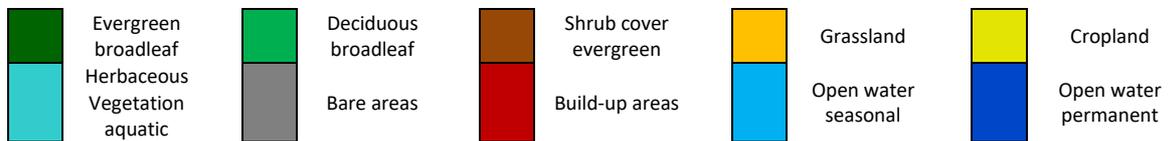


Figure 47. Visual comparison of the: (a) HR optical image used to evaluate the results obtained; (b) LC map obtained by using LSTM on the time series, (c) LC map obtained by using LSTM on composites, (d) LC map obtained by using SVM. The study area is in Africa (Tile 37PCP).

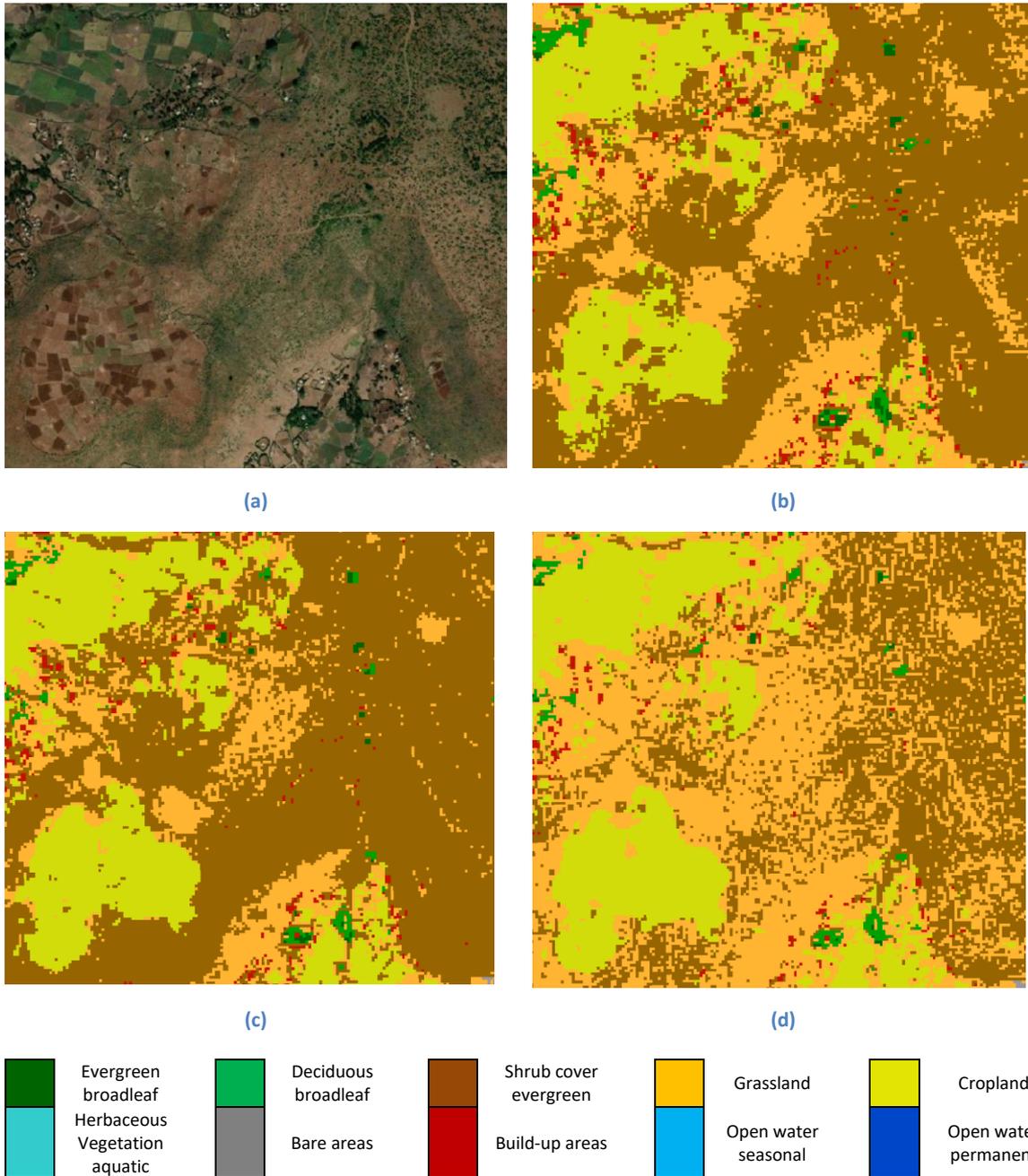


Figure 48. Visual comparison of the: (a) HR optical image used to evaluate the results obtained; (b) LC map obtained by using LSTM on the time series, (c) LC map obtained by using LSTM on composites, (d) LC map obtained by using SVM. The study area is in Africa (Tile 37PCP).

Figure 48 represents the classification products obtained in Africa (tile 37PCP). It can be clearly noticed that the SVM accurately detects shrub cover evergreen. Having a look to the both LSTM results, we can notice quite significant overestimation of that class. Moreover, SVM seems to also provide the best results for the cropland, while LSTM classifier with the monthly composites seems to provide better result than the one using full time-series.

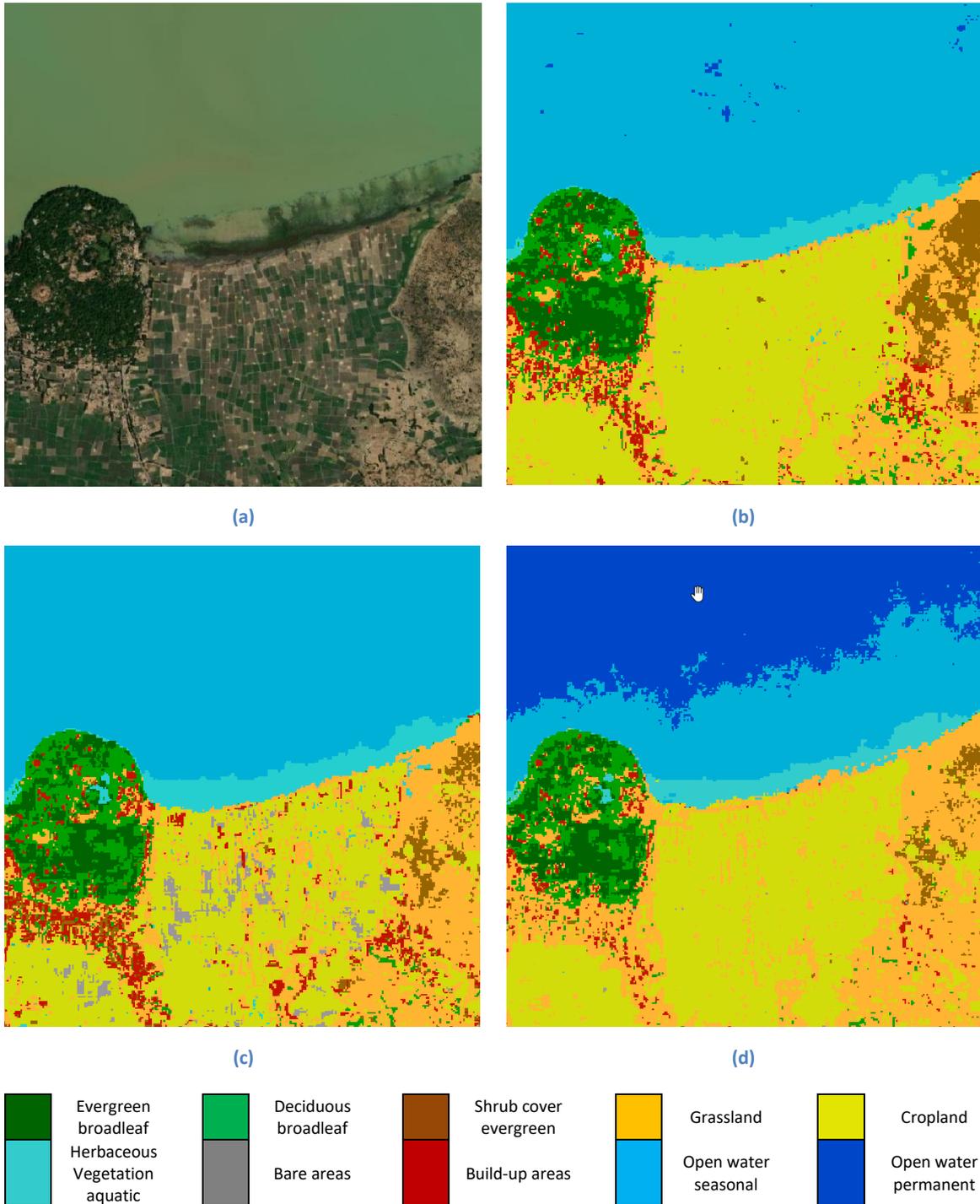


Figure 49. Visual comparison of the: (a) HR optical image used to evaluate the results obtained; (b) LC map obtained by using LSTM on the time series, (c) LC map obtained by using LSTM on composites, (d) LC map obtained by using SVM. The study area is in Africa (Tile 37PCP).

Figure 49 reports the classification products obtained in Africa (tile 37PCP). It can be noticed that SVM is the only classifier that correctly recognizes the permanent water class. Moreover, one can also notice that LSTM with all the time-series obtains better results for cropland, when the LSTM with seasonal composites seems to confuse cropland with bare areas and urban.

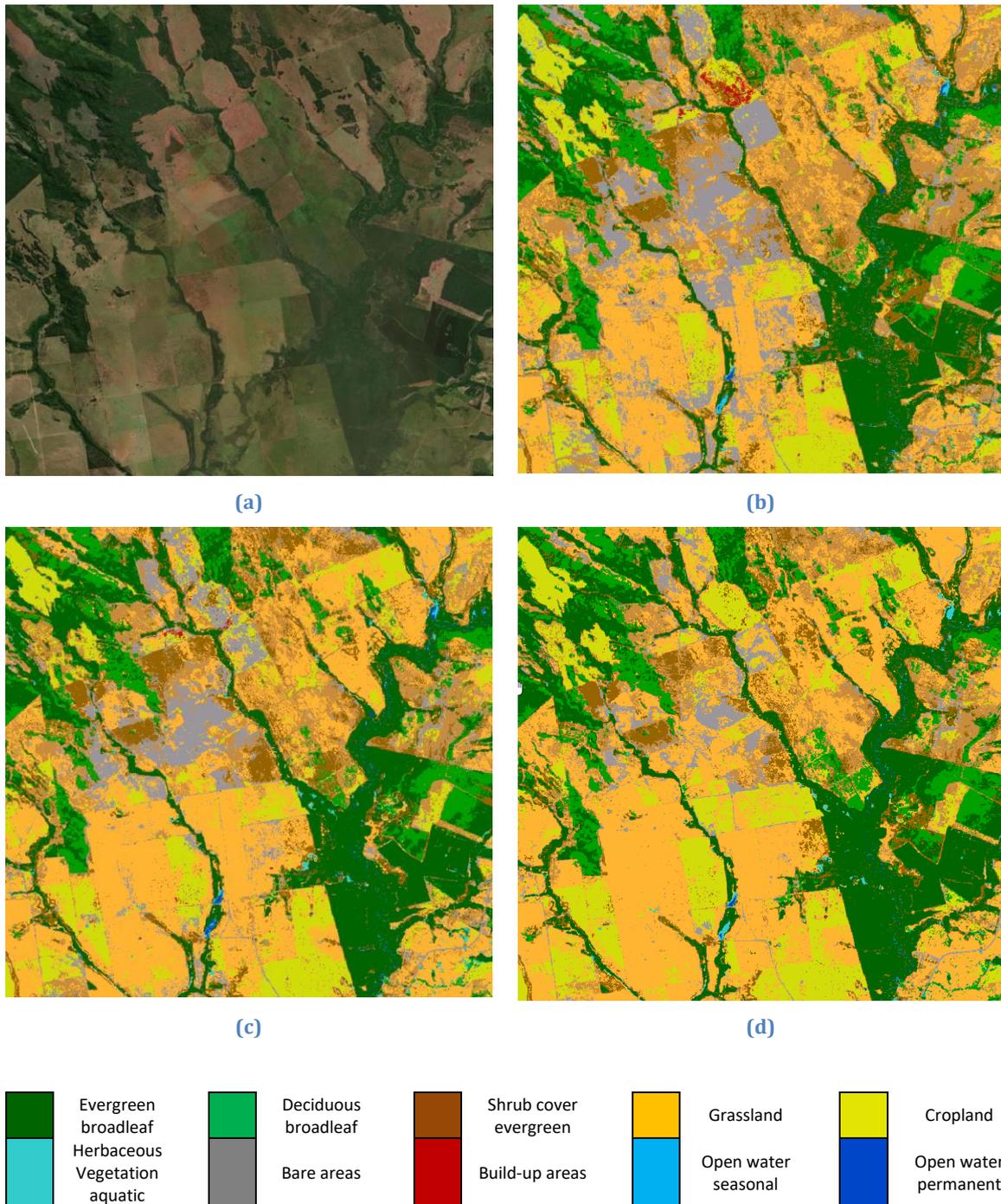


Figure 50. Visual comparison of the: (a) HR optical image used to evaluate the results obtained; (b) LC map obtained by using LSTM on the time series, (c) LC map obtained by using LSTM on composites, (d) LC map obtained by using SVM. The study area is in Amazonia (Tile 21KXT).

Figure 50 represents the classification products obtained in Amazonia (tile 21KXT). Similarly to the Africa HRLC map, SVM seems to provide the most accurate classification map. On the one hand in all three maps there seems to be overestimation of the bare class. However, classes characterized by peculiar geometrical details, such as roads, are very well mapped. Nevertheless, the bare areas overestimation is the least visible on the SVM map.



(a)



(b)



(c)

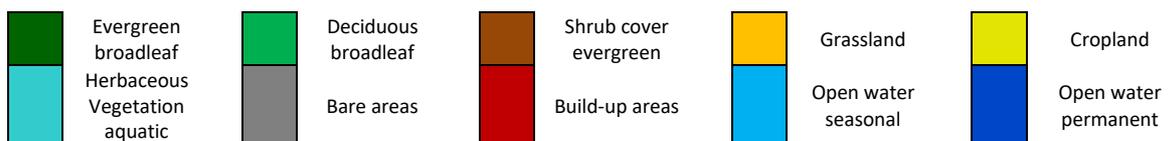


Figure 51. Visual comparison of the: (a) HR optical image used to evaluate the results obtained; (b) LC map obtained by using LSTM on the time series, (c) LC map obtained by using SVM. The study area is in Siberia (Tile 42WXS).

Figure 51 represents the classification products obtained in Siberia (tile 42WXS). Since for Siberia, we are considering only one yearly composite, the comparison with LSTM can be carried out only with the LSTM trained on a full time series of images. Looking at the results, it seems that LSTM is providing better result for permanent water classification. However, it is significantly overestimating the lichens class. The SVM correctly classifies the build-up areas. Also, in the case of urban class SVM seems to provide better results, while LSTM is misinterpreting it as shrub.

4.1.4 Final decision

According to the qualitative analysis, the best LC maps are achieved by SVM. LSTM trained with full time-series and with monthly composites provides poorer accuracy than SVM. This can be attributed to the too small size of the training set for a deep based architecture and the high heterogeneity of the considered landscape. On the other hand, it is well known that SVM is able to provide reasonable accuracy even if a small training dataset is available. It can be also concluded that there is not a big difference between results provided by LSTM trained with full time-series and with monthly composites. Overall, our conclusions are in line with the report provided by the validation team (see PVIR [AD6]).

4.2 SAR data processing

The main novelty introduced during the second year of activity is the use of a high-resolution training data set, much more performant than the one extracted from thematic products at a global scale. The method applied for the extraction of these new training points is based on a stratified approach and the photo-interpretation of very high-resolution imagery, as explained in the ATBD document [AD5]. Once several local sample sites are localized, the training points are firstly labelled via photo-interpretation following a hierarchical decision tree. The seasonality then will be confirmed by means of the analysis of the NDVI profile. The updated SAR processing chain for the production of static and dynamic HRLC products is represented in Figure 52.

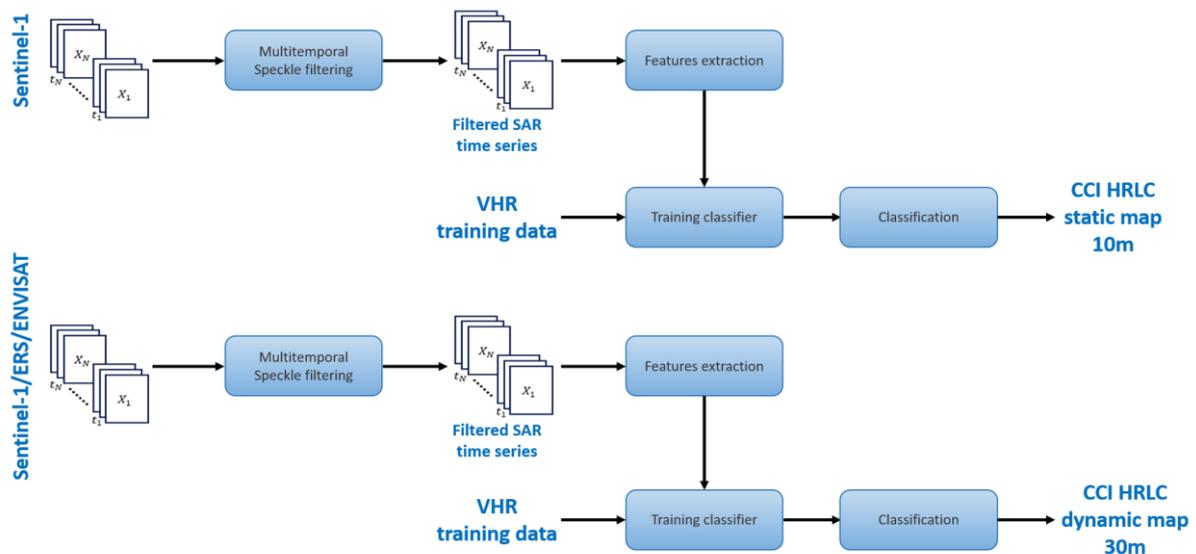


Figure 52. SAR processing chain for the production of static and dynamic HRLC products

4.2.1 Satellite images

The tests have been carried out on the four benchmark areas of the Round Robin, devoted to the analysis of the following three regional areas tiles refer to the tiling grid of Sentinel-2 Level-1C products:

- Amazons: tiles 21KUQ S-2 and 21KXT S-2;
- Siberia: tile 42WXS S-2;
- Sahel: tile 37PCP S-2.

High resolution (10m) Sentinel-1 IWS Level-1 GRD products have been processed over the four benchmark areas to test the static map production for 2019. The Level-1 GRD data contains the detected amplitude and are multi-looked to reduce the impact of speckle at a cost of reducing spatial resolution. The products are projected to ground range using the Earth ellipsoid model, generating images with approximately square resolution pixels and square pixel spacing. Table 4 provides a list of S-1 images that have been used for the local experiments.

4.2.2 Method/algorithm/technique

Within this Section, several methods are presented and compared as candidate approaches to be developed and implemented in the SAR image processing chain depicted in Figure 52.

4.2.2.1 SAR data pre-processing

The first step consists in a preprocessing task aiming to provide a time series devoted to the land cover map production. The production of static map of 2019 requires the Level-1 products of Sentinel-1 GRD data that first are properly calibrated and terrain corrected before any other processing. About the dynamic product with 30m spatial resolution, the data in 2015 are S1 data, while ERS and ENVISAT data will be considered for 2010, 2005, 2000, 1995 and 1990. The typical processing sequence applied to any kind of SAR data entails multi-looking, speckle filtering, orthorectification, radiometric calibration, and terrain correction. Multi-looking averages over range and azimuth cells to improve radiometric resolution and produce near-square pixels. Adaptive filters use local statistics to filter the data and so reduce image speckle and, in some cases, preserve or enhance edges and other features. As with optical data, SAR data are orthorectified and radiometrically calibrated to produce suitable images for comparison. The best available DEM is used to correct for spatial distortions in the range (across-track) and azimuth (along-track) directions. The process converts the pixel data from slant to ground range geometry and in a defined cartographic (map coordinate) system. During radiometric calibration, standard radar equations are used to correct pixel data for systematic errors and brightness variations due to terrain. An additional terrain illumination correction step is applied to correct for geometric and radiometric distortions present in images collected over steep terrain. These distortions mask the useful backscatter related to land cover or geophysical features and need to be corrected for effective land cover mapping using SAR data. Finally, linearly-scaled data are converted into decibels (dB) and properly resampled to match the spatial resolution of 30m. All the preprocessing task is easily accomplished with the Sentinel's Application Platform (SNAP) software provided by ESA.

4.2.2.2 Training set preparation

The training set extraction is performed according to the hierarchical approach presented in section 8.2 of the ATBD document [AD5]. The main purpose is to achieve a consistent and accurate training data set thanks to the photointerpretation of very high resolution images. Figure 53 shows the updated list of land cover classes of the final HR legend. See in Figure 54 an example of training set extraction for four selected tiles.

Table 5. List of Sentinel-1 data on the four benchmark areas

Area	Section	Satellite	Band	# Products	Date list (2019y)
Amazon – 21KUQ	Upper	S1B	VH	31	01-04 01-16 01-28 02-09 02-21 03-05 03-17 03-29 04-10 04-22 05-04 05-16 05-28 06-09 06-21 07-03 07-15 07-27 08-08 08-20 09-01 09-13 09-25 10-07 10-19 10-31 11-12 11-24 12-06 12-18 12-30
	Lower	S1B	VH	31	01-04 01-16 01-28 02-09 02-21 03-05 03-17 03-29 04-10 04-22 05-04 05-16 05-28 06-09 06-21 07-03 07-15 07-27 08-08 08-20 09-01 09-13 09-25 10-07 10-19 10-31 11-12 11-24 12-06 12-18 12-30

Amazon – 21KXT	Upper	S1B	VH	20	01-23 02-04 02-16 03-12 03-24 04-05 04-17 04-29 08-15 08-27 09-08 09-20 10-02 10-14 10-26 11-07 11-19 12-01 12-13 12-25
	Lower	S1B	VH	20	01-23 02-04 02-16 03-12 03-24 04-05 04-17 04-29 08-15 08-27 09-08 09-20 10-02 10-14 10-26 11-07 11-19 12-01 12-13 12-25
Siberia – 42WXS	-	S1B	VH	20	01-03 01-15 01-27 03-04 03-16 03-28 04-21 05-03 05-15 06-08 06-20 07-14 07-26 08-19 09-12 09-24 10-18 11-11 12-05 12-17
Sahel – 37PCP	-	S1B	VH	16	01-11 01-23 02-16 03-24 04-05 05-11 06-04 06-16 07-10 07-22 08-27 09-20 10-14 11-07 12-01 12-25

HRLC CLASSES											
CODE	DN	1 st LEVEL	CODE	DN	2 nd LEVEL	CODE	DN	3 rd LEVEL	CODE	DN	4 th LEVEL
--	0	No data									
10	1	Tree cover evergreen broadleaf									
20	2	Tree cover evergreen needleleaf									
30	3	Tree cover deciduous broadleaf									
40	4	Tree cover deciduous needleleaf									
50	5	Shrub cover evergreen	51	17	Broadleaf						
			52	18	Needleleaf						
60	6	Shrub cover deciduous	61	19	Broadleaf						
			62	20	Needleleaf						
70	7	Grasslands	71	21	Natural						
			72	22	Managed						
80	8	Croplands	81	23	Winter	811	24	Rainfed			
						812	25	Irrigated	8121	26	Sparkling
									8122	27	Flooding
			82	28	Summer	821	29	Rainfed			
						822	30	Irrigated	8221	31	Sparkling
									8222	32	Flooding
			83	33	Multicropping	831	34	Rainfed			
						832	35	Irrigated	8321	36	Sparkling
									8322	37	Flooding
90	9	Woody vegetation aquatic or regularly flooded									
100	10	Grassland vegetation aquatic or regularly flooded									
110	11	Lichens and Mosses									
120	12	Bare areas	121	38	Unconsolidated	1211	39	Sands			
						1212	40	Bare soils			
			122	41	Consolidated						
130	13	Built-up	131	42	Buildings						
			132	43	Artificial Roads						
140	14	Open Water seasonal									
150	15	Open Water permanent									
160	16	Permanent snow and/or ice	161	44	Snow						
			162	45	Ice						

Figure 53. List of classes of the High Resolution Land Cover (HRLC) legend

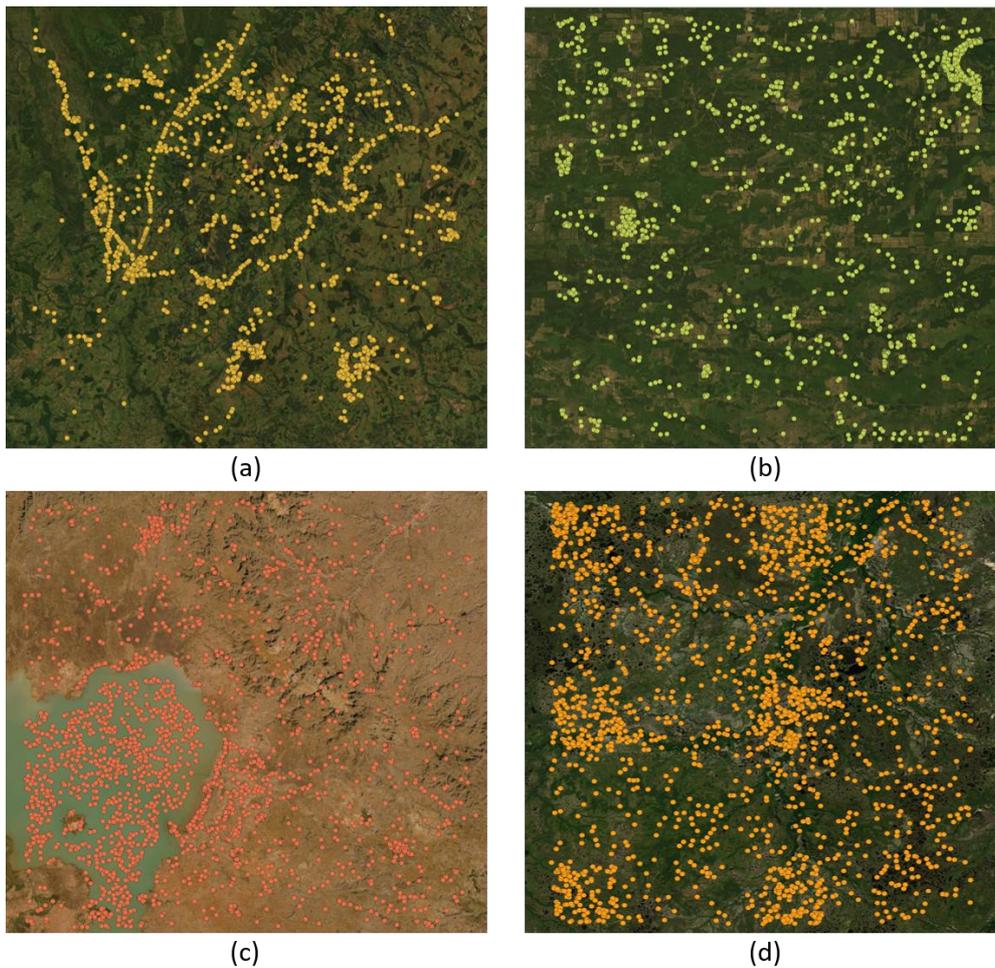


Figure 54. Training data acquired via hierarchical decision tree for Amazon 21KXT (a), Amazon 21KUQ (b), Sahel 37PCP (c) and Siberia 42 WXS (d) tiles.

Speckle filtering

According to the tests performed during the first year, the multi temporal approach to speckle filtering [22] is much more recommended for speckle effective removal in a time series data set. This filter is based on the calculation of a super-image exploiting the spatial and temporal information of the whole SAR time series. The filter aims to enhance the quality of each image in the series by means of an effective speckle reduction and spatial resolution preservation.

4.2.2.3 Feature extraction

Feature extraction methods encompass characteristics and texture, structural and graph descriptors. To improve the ability of a classifier to recognize and discriminate the different environment textures and morphological structures (e.g. urban areas, agricultural crops, forests, etc.), the amplitude of VH channels have been considered [23].

To analyze and explore the spatial information contained in satellite images, a set of filters operating in the spatial domain have been considered. The rationale for selecting these algorithms is their execution velocity and versatility. Although they might not be the most accurate ones, the possibility to apply them quickly to 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 as follows:

- *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 that replaces each pixel value with the median of the pixel values in the local neighborhood.
- *Lee filter*, an adaptive filter based on minimum mean-square error (MMSE) that converts the multiplicative model into an additive one, thereby reducing the problem of dealing with speckle to a known tractable case.
- *Minimum (maximum) filter*, a non-linear filter that locates 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.
- *Maxmin filter* a non-linear filter that computes the difference between the maximum filtered version of the image and the minimum one

4.2.2.4 Classification

Two supervised classifiers have been considered, i.e., *Random Forest* (RF) and *Support Vector Machine* (SVM). They are superior and more robust than unsupervised methods [16]. Some erroneous reference data (e.g., slightly outdated ones) are acceptable in training [24].

The whole classification chain has been investigated on Sentinel-1 time-series data assuming different scenarios. The experiments presented in this section have been carried on the four Round Robin benchmark areas and visually photointerpreted reference data have been used. The RF and SVM were applied to evaluate their classification performance in multitemporal analysis of SAR data. The choice of parameters for the RF classifier is not very sensitive to this kind of problem [25], and one-hundred trees have been used. According to the results in Section 3.2, the multitemporal classification was the most performant analysis in terms of land cover recognition. In the multitemporal analysis framework, a time series of Sentinel-1 VH images has been considered. All images have been divided according to annual seasons: *winter*, *spring*, *summer* and *autumn*. For each season, a multitemporal despeckling filter has been applied and the super image provided as output. To test the performances of the classification chain, more spatial and textural information have been added. Spatial features (Lee, Min, Max, Max Min, Mean and Median) have been extracted on each seasonal super image, leading to a classification based on twenty-eight features in total. All classification results and the comparison between RF and SVM classifier are depicted in Figure 55 to Figure 58.

Table 6. Training Set produced on MR products: the translation of the considered coarse thematic products into the desired map legend is reported. (*) The built-up class has been extracted with the UEXT algorithm, described in Lisini et al. [26]

	CCI-HRLC (1st Level)	ESA CCI LC 2018	GLCNMO
	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	120	
	Grasslands	130	
	Croplands		11, 12, 13
	Lichens and Mosses	140	
	Bare areas	200, 201, 202	
	Built-up(*)		
	Open water	210	
	Permanent Snow and/or ice	220	

The first classification experiments have been carried out and compared considering both photointerpreted training points and reference data from existing medium resolution GLCM, including the ESA CCI land cover map of 2018 with 300m resolution. About reference data based on GLCM, the products were selected as the best candidates considering: i) good spatial resolution, ii) detailed hierarchical legend that includes many classes in common with the CCI HRLC legend, and iii) relatively recent temporal coverage. The existing legends of GLCM have been compared and then translated into the desired HRLC legend. A weak training set has been produced by selecting from the available thematic maps those samples having the highest probability of belonging to areas correctly associated to their label.

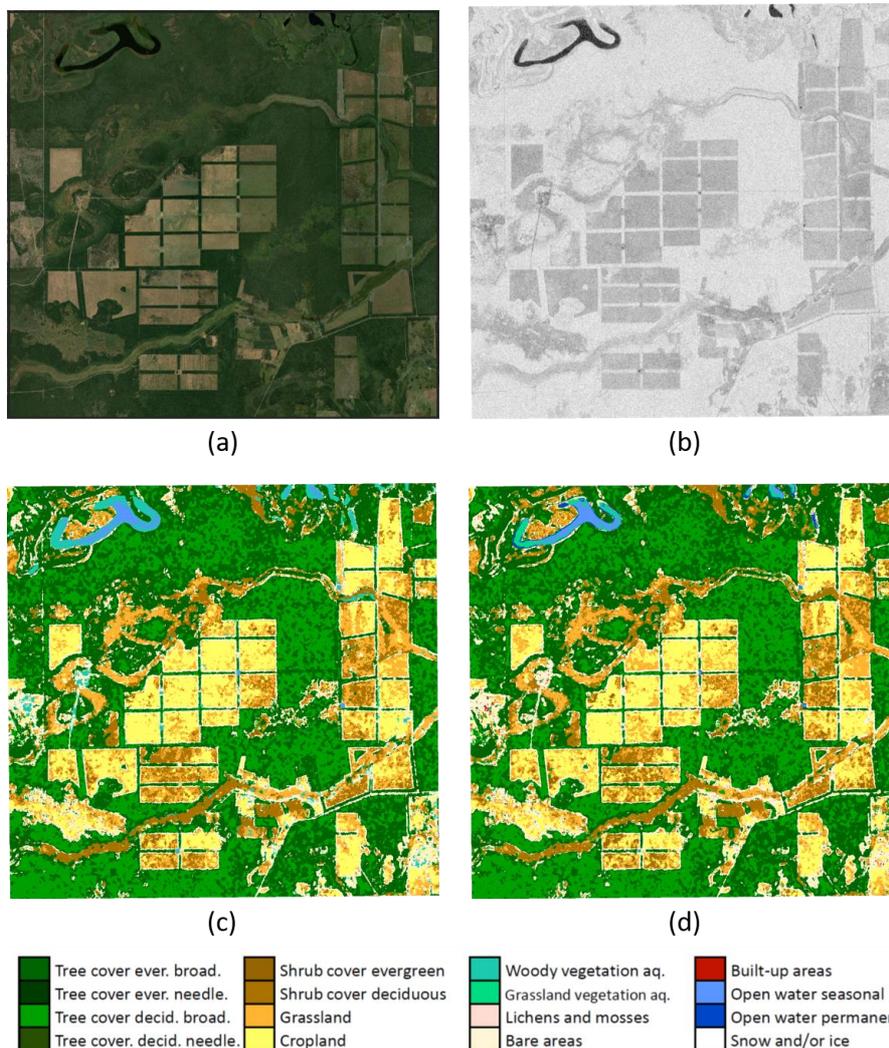
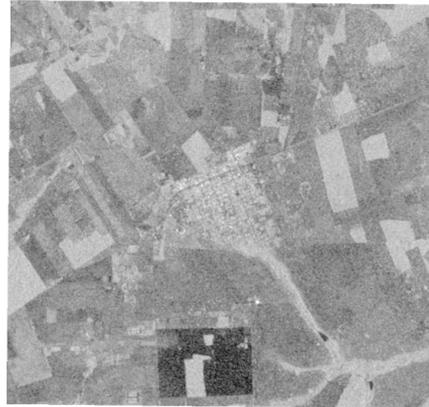


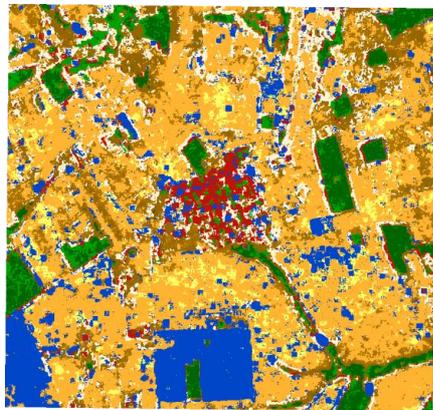
Figure 55. Comparison between RF and SVM classification for of the 21KUQ Amazonian tile: ESRI satellite image (a), SAR seasonal super image (b), SVM static map of 2019 generated by using the seasonal multitemporal sequence (c) and RF static map of 2019 generated by using the same input (d).



(a)



(b)



(c)



(d)

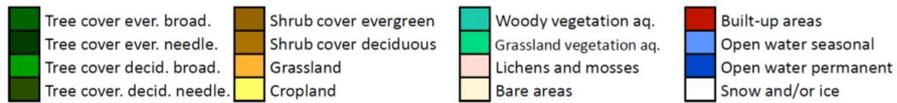
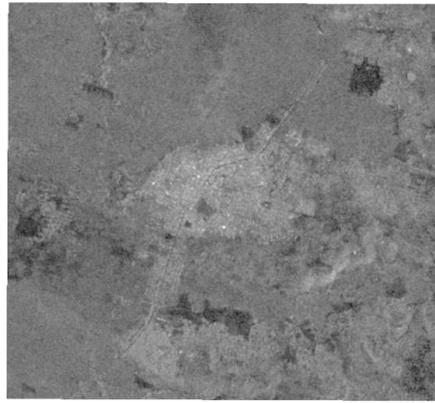


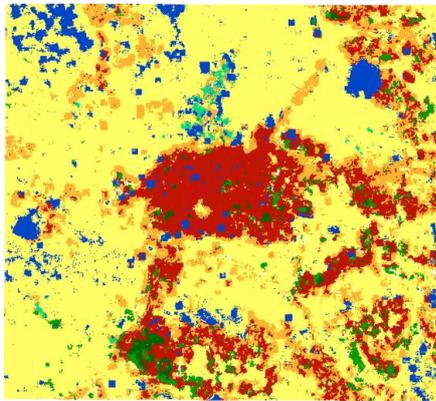
Figure 56. Comparison between RF and SVM classification for an AOI of 21KXT Amazonian tile: ESRI satellite image (a), SAR seasonal super image (b), SVM static map of 2019 generated by using the seasonal multitemporal sequence (c) and RF static map of 2019 generated by using the same input (d).



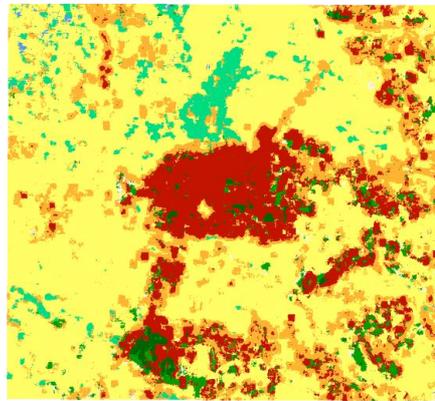
(a)



(b)



(c)



(d)

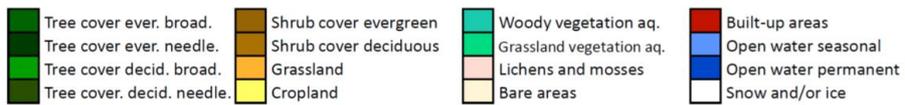


Figure 57. Comparison between RF and SVM classification for an AOI of the 37PCP Saharan tile: ESRI satellite image (a), SAR seasonal super image (b), SVM static map of 2019 generated by using the seasonal multitemporal sequence (c) and RF static map of 2019 generated by using the same input (d).

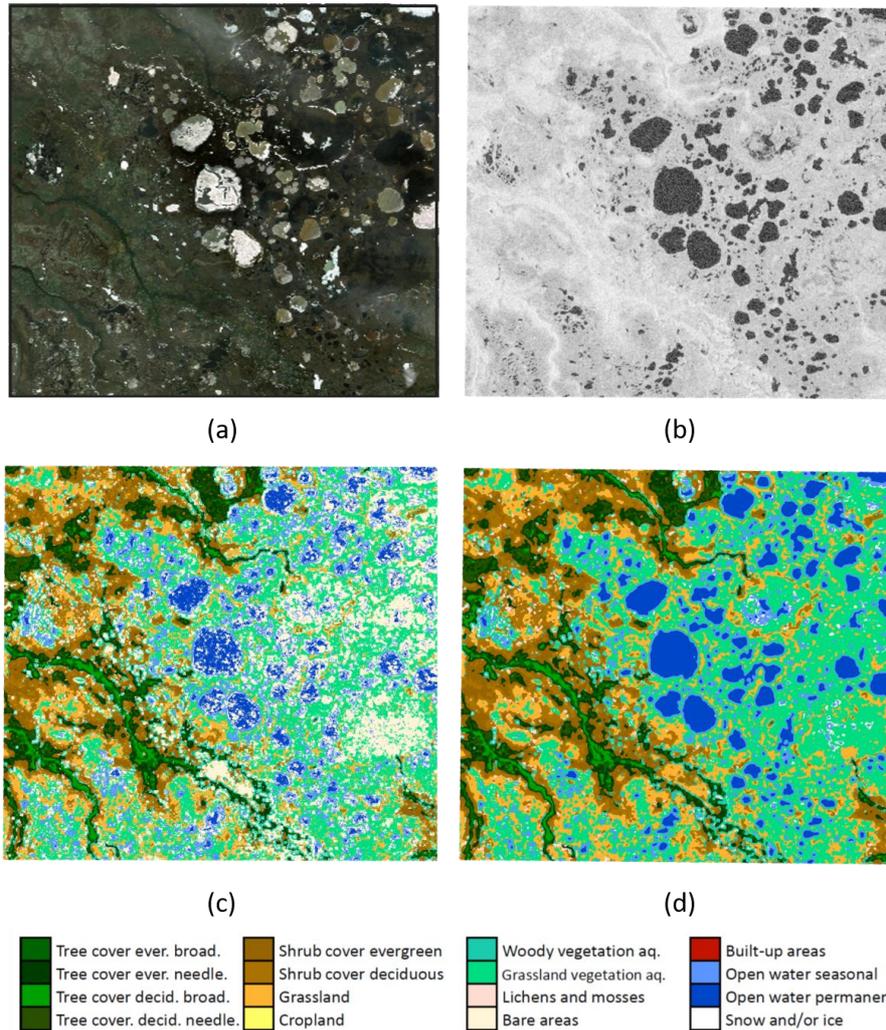


Figure 58. Comparison between RF and SVM classification for an AOI of the 42WXS Siberian tile: ESRI satellite image (a), SAR seasonal super image (b), SVM static map of 2019 generated by using the seasonal multitemporal sequence (c) and RF static map of 2019 generated by using the same input (d).

4.2.3 Qualitative evaluation

By looking at the maps reported above, the comparison between Random Forest and Support Vector Machine are reported, and the different performance levels of the two classifiers can be appreciated (see Figure 55 to Figure 58). The multitemporal approach produces interesting results, with nice classification capabilities, confirming the state-of-the-art level of this methodology.

Moving to the static products reported in Figure 55(c) and Figure 55 (d), several croplands and vegetative areas are very well extracted, showing the capability of the approach to produce a reliable map. By a first visual assessment, both in Amazonia and Sahel, SAR maps seem to contain several thematic errors, especially a huge overestimation of the built-up class. Moreover, SAR SVM prototypes for Amazon 21KXT and Sahel, in Figure 56(c) and Figure 57(c) respectively, contain a strong contamination of the water class. About the Siberia area, taking into the account the SVM map reported in Figure 58(c), the recognition of the water class is not well performant and many outliers of bare soil class can be noticed. In all the scenarios, RFs (Figure 55(d), Figure 56(d), Figure 57(d) and Figure 58(d)) show a greater capability in class recognition than SVMs. The performance can be explained due to the nature of RF in being inherently multiclass whereas Support Vector Machines need workarounds to deal with classification tasks involving multiple classes. Moreover, RF works well with a mixture

of numerical and categorical features. In other words, with Random Forest one can use inhomogeneous data without any refined pre-processing.

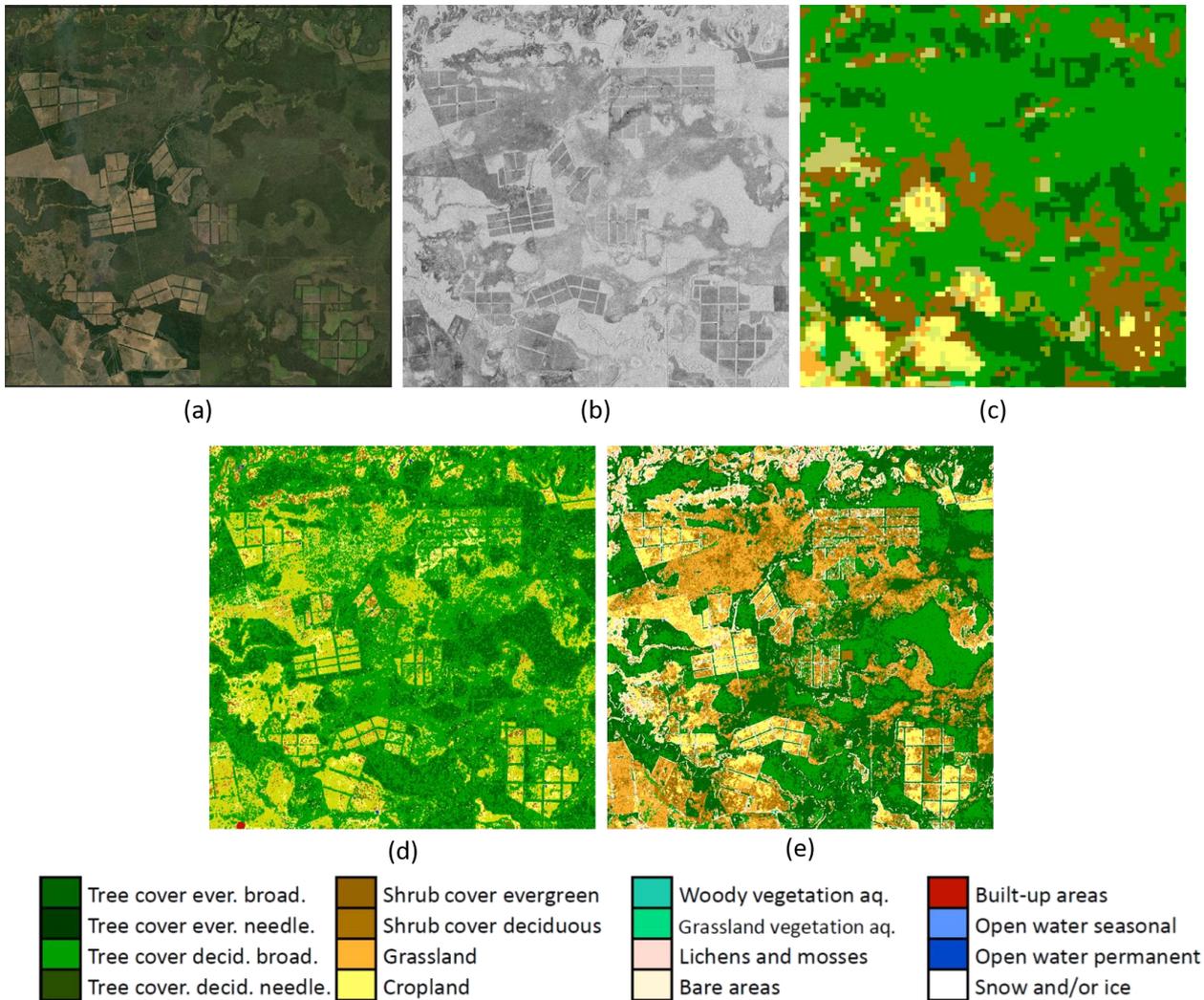


Figure 59. Classification performance comparison for different reference data on an AOI of 21KUQ Amazonian tile: ESRI satellite image (a), SAR seasonal super image (b) ESA CCI LC MR map of 2018 (c), RF classification based on MR reference data (d) and RF classification based on photointerpreted reference data (e).

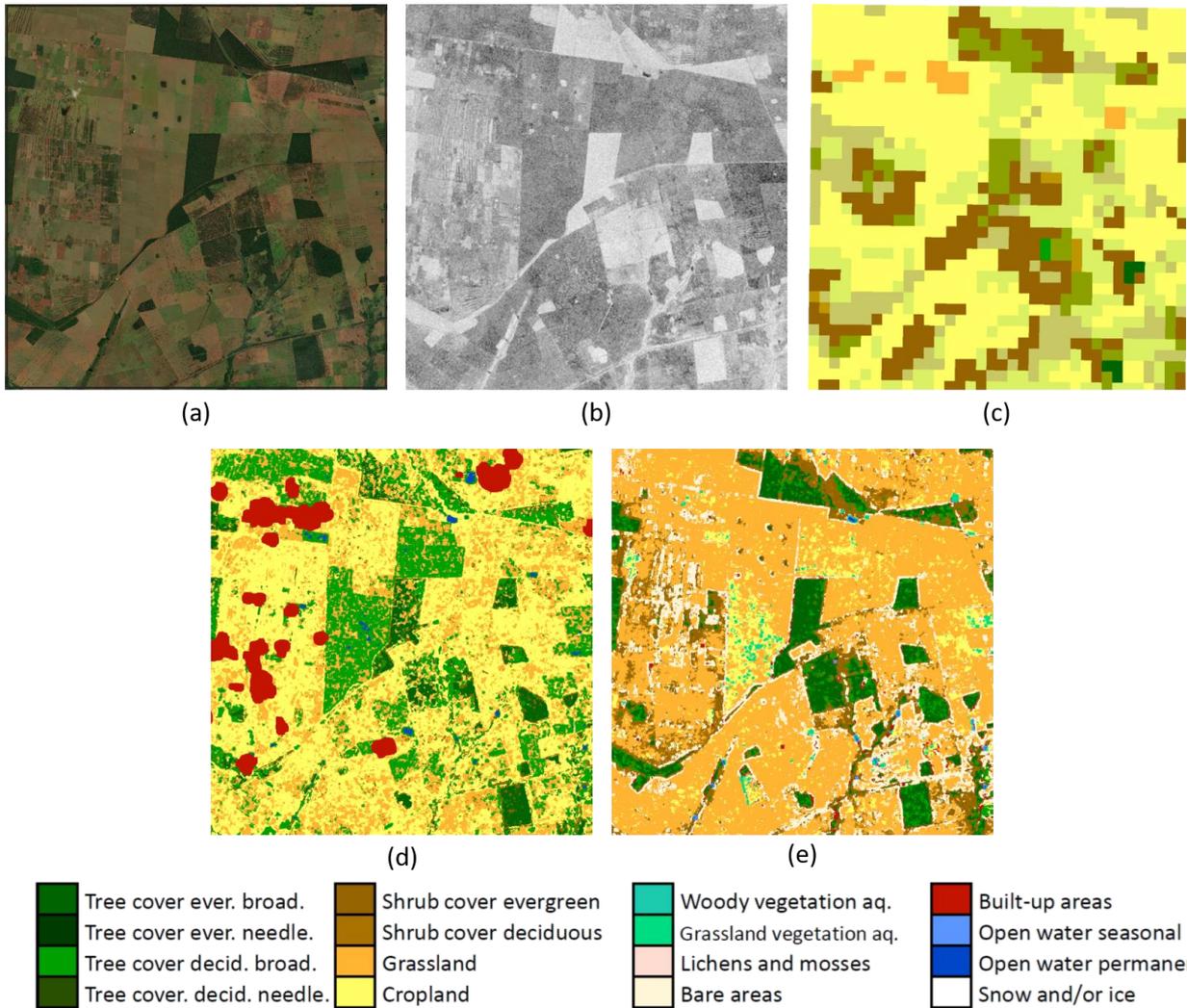
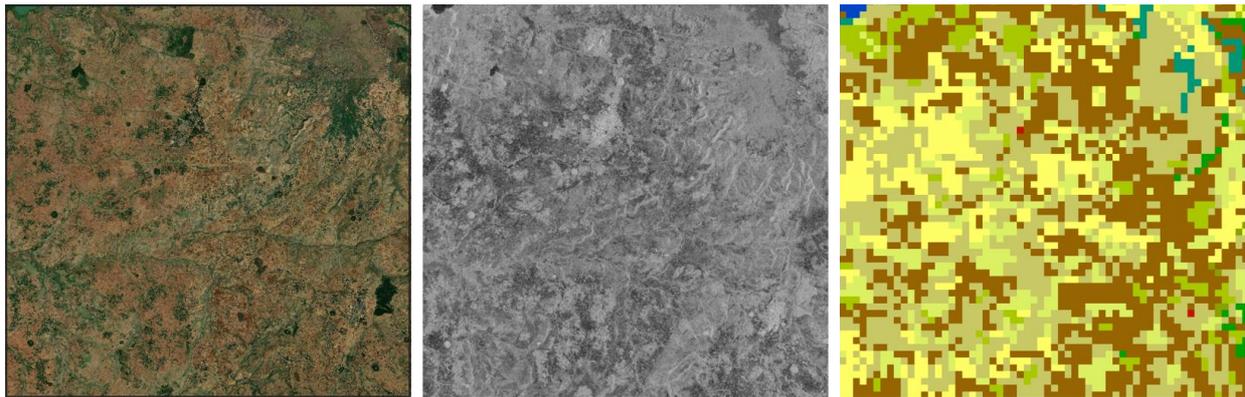


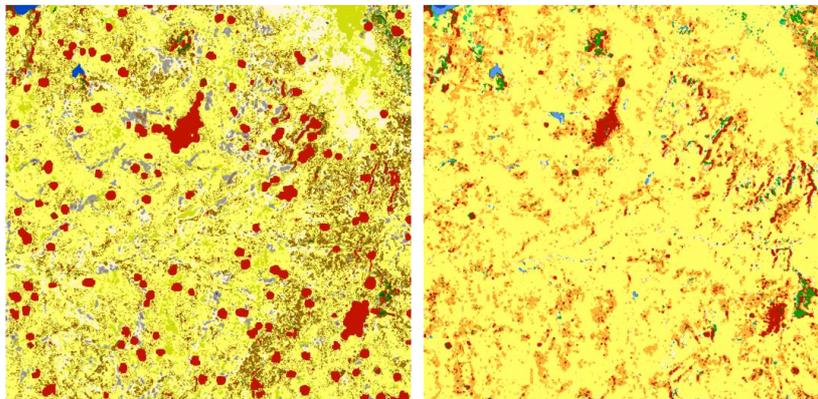
Figure 60. Classification performance comparison for different reference data on an AOI of 21KXT Amazonian tile: ESRI satellite image (a), SAR seasonal super image (b) ESA CCI LC MR map of 2018 (c), RF classification based on MR reference data (d) and RF classification based on photointerpreted reference data (e).



(a)

(b)

(c)



(d)

(e)

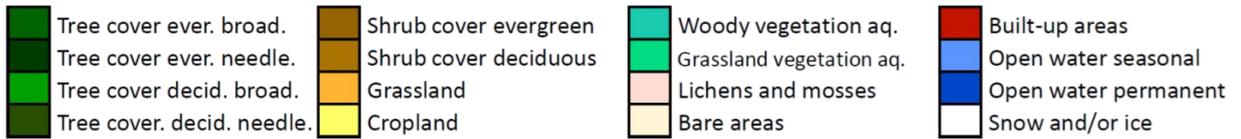


Figure 61. Classification performance comparison for different reference data on an AOI of 37PCP Saharan tile: ESRI satellite image (a), SAR seasonal super image (b) ESA CCI LC MR map of 2018 (c), RF classification based on MR reference data (d) and RF classification based on photointerpreted reference data (e).

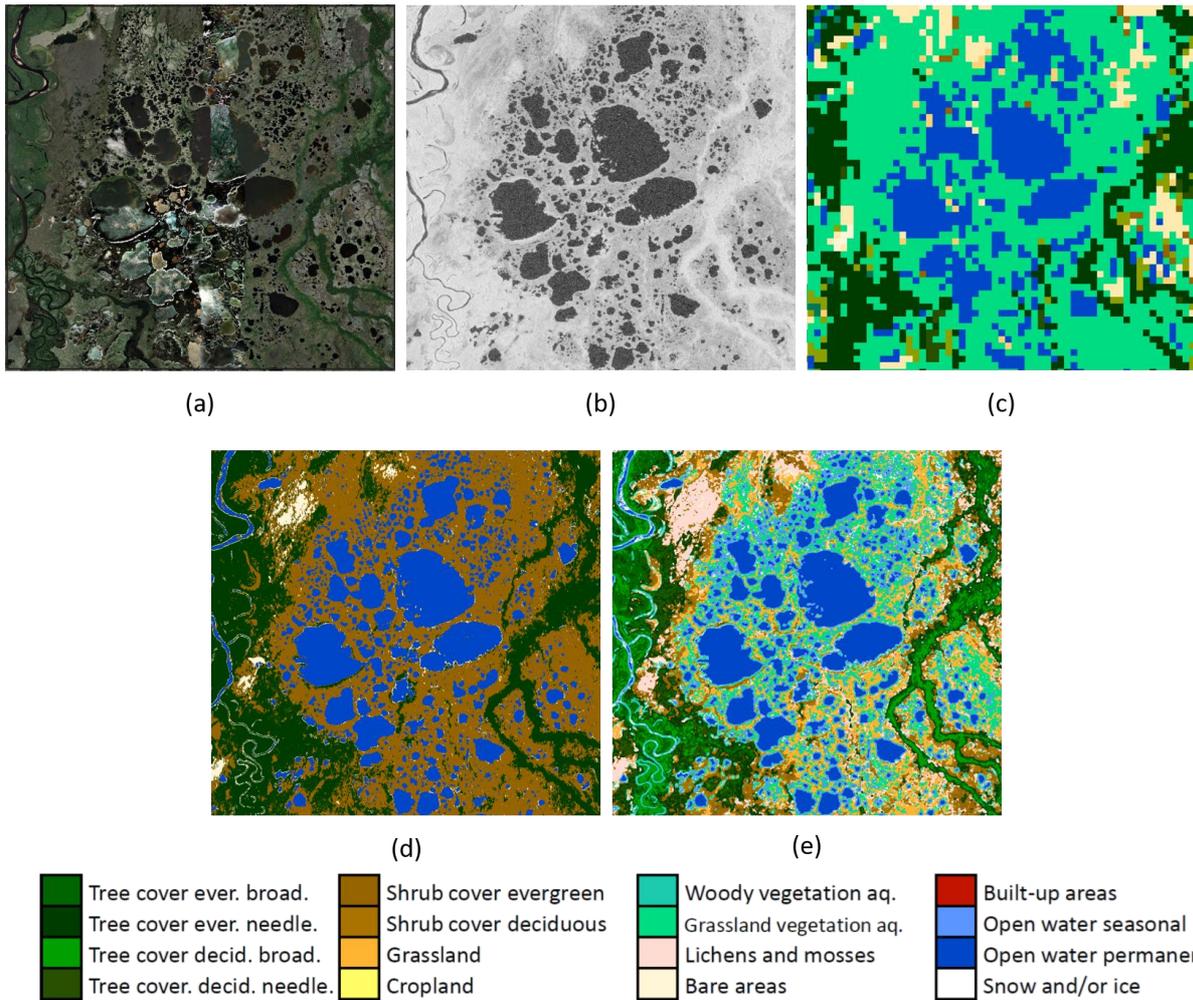


Figure 62. Classification performance comparison for different reference data on an AOI of 42WXS Siberian tile: ESRI satellite image (a), SAR seasonal super image (b) ESA CCI LC MR map of 2018 (c), RF classification based on MR reference data (d) and RF classification based on photointerpreted reference data (e).

Figure 59 to Figure 62 show a qualitative comparison over a portion of each study areas. We can appreciate that the classification based on the reliable training data hierarchically extracted solves many outliers present in the results, where a weak training set was applied. For the Siberia, as shown in Figure 62, the results are very encouraging, since the classified map based on the photointerpreted training points brings to a land cover performance that can be easily appreciated with a comparison with the ESA CCI medium resolution product, but. For a more detailed qualitative and quantitative assessment please refer to PVASR v1.0.

4.2.4 Final decision

According to the previous subsections and the visual assessment, the best LC maps are achieved by using RFs. RF is state-of-the-art in remote sensing image processing, and yields classification accuracies as high as SVM, but with a much lower computational complexity. It is also more stable with respect to the choice of parameters [25], and easily provides the probability of belonging to a class (necessary input to the data fusion step). This makes it an excellent candidate for operational processing chains. Reference data acquired by photointerpretation based on a hierarchical method have been extracted in each study area and better classification performance have been achieved in all test areas.

Having that said, it has to be underlined that this comparison has been performed only qualitatively, and a quantitative analysis is provided in PVIR v1.0 for a better quality evaluation.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	71	

4.3 Multi-Sensor Optical-SAR Data Fusion

4.3.1 Multisensor Geolocation Methods

In the context of multi-sensor geolocation, different image registration methods and strategies have been designed and experimented on the available datasets, following up on the methodological analysis conducted in the deliverables of the previous milestones.

The multi-sensor geolocation process is composed, as depicted in Figure 63, 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 63. Composition of the multisensor geolocation process.

According to the HRLC proposal, we have developed and tested area-based image registration methods with a combination of different geometric transformations, similarity measures, and optimization strategies (see also ATBD-v3).

4.3.1.1 Geometric Transformations

With respect to the geometric transformations, we have considered:

- 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)

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

Here, (x, y) and (X, Y) indicate the axes of the coordinate frames in the reference and input images, respectively, T_x and T_y are translation parameters, θ is a rotation angle, and s is a scale parameter [27].

4.3.1.2 Similarity Measures

With respect to the similarity measures, we have considered:

- Cross-correlation

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

- Mutual Information

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

Here, $In(\cdot)$ and $Ref(\cdot)$ indicate the input and reference images (which are both assumed composed of $M \times N$ pixels), respectively, $p_{Ref, In}$ is their joint probability density function (PDF), p_{Ref} and p_{In} are their marginal PDFs, $CC(\cdot)$ is their cross-correlation evaluated on a given pixel location, $MI(\cdot)$ is their mutual information [27], [28], [29].

It is worth noting that the cross correlation is computed through the fast Fourier transform (FFT) algorithm. 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:

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

where $\mathcal{F}(\cdot)$ denotes the Fourier transform operator and f and g are 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 from the computational efficiency of the FFT [30].

In order to compute the cross-correlation between two images it is necessary to: (i) compute the FFT of each image 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 given in Figure 64.

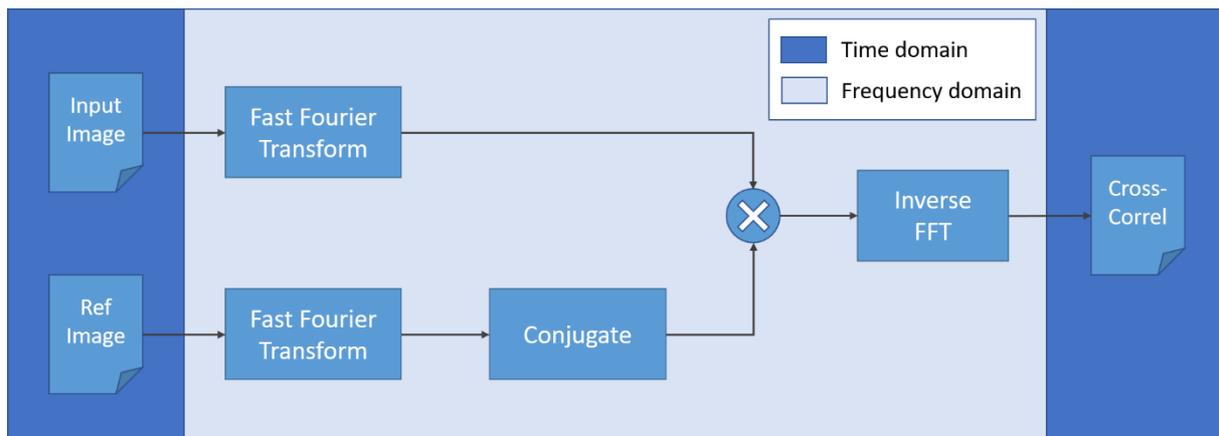


Figure 64. Block diagram of the application of fast Fourier transform (FFT) to compute cross-correlation.

4.3.1.3 Optimization Strategies

With respect to the optimization strategies, we have considered:

- Powell's algorithm, for unconstrained minimization. It uses Powell's formulation of an approximate conjugate direction method. The objective function does not need to be differentiable, and no

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	73	

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 [31]. Moreover, the line search is done by the Golden-section and Brent's methods [32].

- Constrained optimization by linear approximation (COBYLA), for constrained minimization. It 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 [33].
- Constrained reformulation of the Powell's method (Powell's algorithm with barrier functions). 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.

4.3.2 Automatic tiling for large-scale registration

The HRLC pipeline operates geometrically on the pixel grid of the Sentinel-2 granules. Accordingly, the multi-sensor geolocation module also works with respect to this lattice. Given the size of the Sentinel-2 granules, taken as reference images in the multi-sensor 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 these 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. Figure 65 details the tiling scheme of the multi-sensor geolocation procedure.

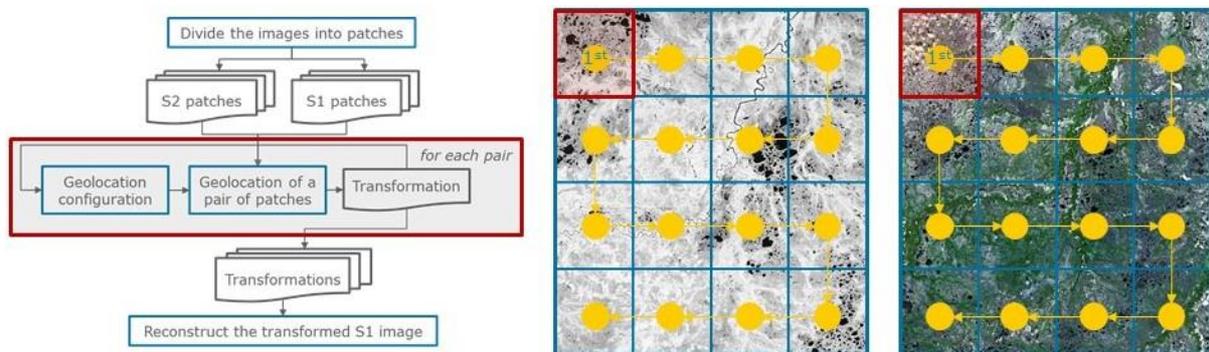


Figure 65. Detail of the tiling scheme for multi-sensor geolocation.

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. Dedicated procedures are also integrated in the algorithm to detect anomalies in this iterative process and prevent them for affecting the registration output.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	75	

4.3.3 Quantitative evaluation using COBYLA and the unconstrained Powell's algorithm

Several experiments have been carried out with respect to multi-sensor geolocation. First, experiments with synthetic data generated from single-sensor measurements have been performed. Second, the aforementioned registration methods have been applied to multi-sensor S1-S2 data. The imagery associated with the first round robin were used for experiments. The focus has been initially on the unconstrained Powell and constrained COBYLA methods.

In particular, the dataset used for the experiments is composed of two S1 SAR tiles and an S2 optical tile. The single-sensor synthetic datasets have been built by extracting an area from either an S1 or an S2 tile and by transforming it according to predefined transformations (translation, rigid, RST, etc.). This strategy allows to quantitatively determine the registration accuracy obtained by registering the resulting couples.

The multi-sensor dataset has been built by extracting an area from the result of stacking the S1 and S2 tiles. This way, it is possible to test the registration methods in a fully real-world scenario, although analytical and quantitative results are not available of course, because the "true" ideal matching of the optical and SAR data is undefined. In this case, registration accuracy can be qualitatively appreciated using false color representations and checkerboard visualizations.

Finally, the synthetic multi-sensor dataset has been built by registering the S1 and S2 tiles using accurate georeferencing information, extracting an area from the stacking of the two georeferenced tiles, and applying a well-known transformation to one of the two. As for the first scenario, this strategy allows to quantitatively evaluate the accuracy of the proposed methods in a multi-sensor scenario as well.

The experiments that have been carried out are the following:

- Finding the optimal translation using cross-correlation as metric;
- Finding the optimal translation using both cross-correlation and mutual information as metrics;
- Finding the optimal rigid transformation using both cross-correlation and mutual information as metrics;
- Finding the optimal RST transformation using mutual information as a metric.

In the case of the synthetic dataset, where the ground truth (GT) transformation is known, the root mean square error (RMSE) between the real and the resulting transformation is used as a measure of registration accuracy. For details on the RMSE computation we refer to [34] and [29]. Additionally, the computer used for the experiments is a Windows system equipped with a quad-core Intel i7 processor with a working frequency of 3.60GHz and 24Gb of RAM.

Before analyzing the results in more detail, it is worth anticipating that the use of cross-correlation as a similarity metric is effective and satisfactory only in the case of the single-sensor dataset. In the multi-sensor cases, the different nature of the images prevented the usage of such a similarity metric, because of the overall low values achieved by the inter-sensor correlations. On the contrary, the mutual information measure resulted to be more robust in this case.

Therefore, cross-correlation has been experimentally validated only in the cases of translation and rigid transformations. The results have been proved satisfactory only in the single-sensor scenarios, with poor performances in the multi-sensor cases. Hence, multi-sensor experiments with more challenging scenarios (higher-order transformations and multi-step registration) have been focused on the use of the more promising mutual information similarity metric.

4.3.3.1 Optimal translation using cross-correlation

This type of registration performed well only in synthetic cases where the translation transformation was manually applied. Experiments were carried out with respect to optical-optical, SAR-SAR, and optical-SAR matching.

Note that from now on, the convention used to represent the parameters of the geometric transformation is the following: Transformation = [translation on the x axis, translation on the y axis, rotation (deg.), scale factor]

4.3.3.1.1 Synthetic Optical-to-optical matching

Elapsed Time:	True Transformation:	Resulting Transformation:	Registration RMSE:
0.636 seconds	[-76, 52, 0, 1]	[-76.0, 52.0, 0.0, 1.0]	0.0 pixels

Figure 66 shows: (i) the reference optical image (left panel); (ii) the input optical image to be registered (central panel); and (iii) the resulting cross-correlation with the corresponding point of maximum (right panel).

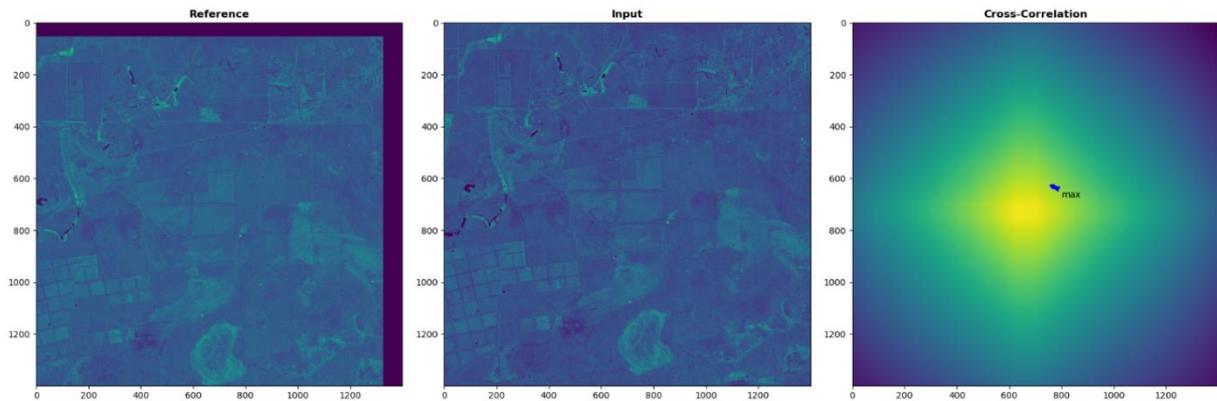


Figure 66. The reference optical image (left panel); the input optical image to be registered (central panel); and the resulting cross-correlation with the corresponding point of maximum (right panel).

Figure 67 shows a false color composition of the images before and after registration. The composition uses the green and the magenta colors for the input and the reference images (left) and for the transformed input and reference images (right), respectively.

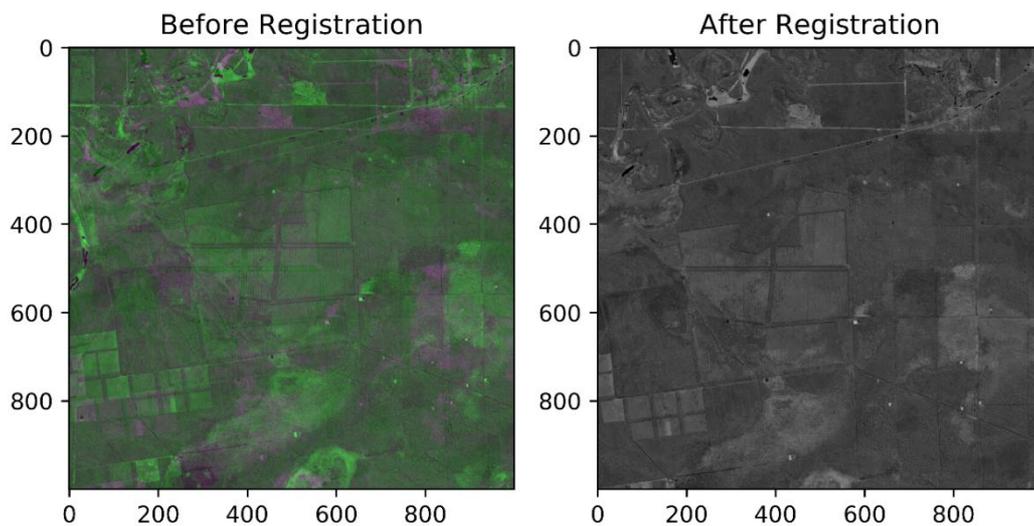


Figure 67. False color composition of the optical images before (left) and after (right) the registration.

4.3.3.1.2 Synthetic SAR-to-SAR matching

Elapsed Time:	True Transformation:	Resulting Transformation:	Registration RMSE:
0.680 seconds	[-76, 52, 0, 1]	[-76.0, 52.0, 0.0, 1.0]	0.0 pixels

As in previous case, Figure 68 shows the reference SAR image, the input SAR image to be registered, and the resulting cross-correlation with its own point of maximum.

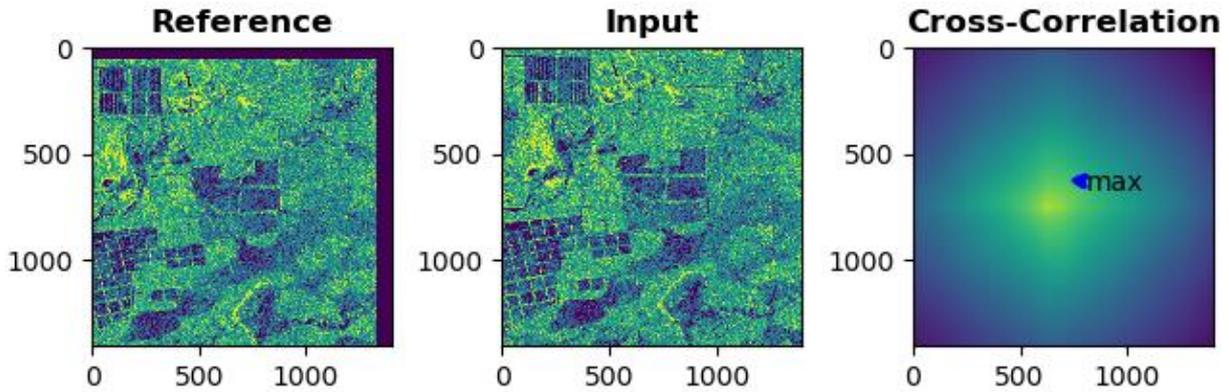


Figure 68. The reference SAR image (left panel); the input SAR image to be registered (central panel); and the resulting cross-correlation with the corresponding point of maximum (right panel).

Figure 69 shows a false color composition of the images before and after registration. The same composition as in the Optical-to-Optical case is used.

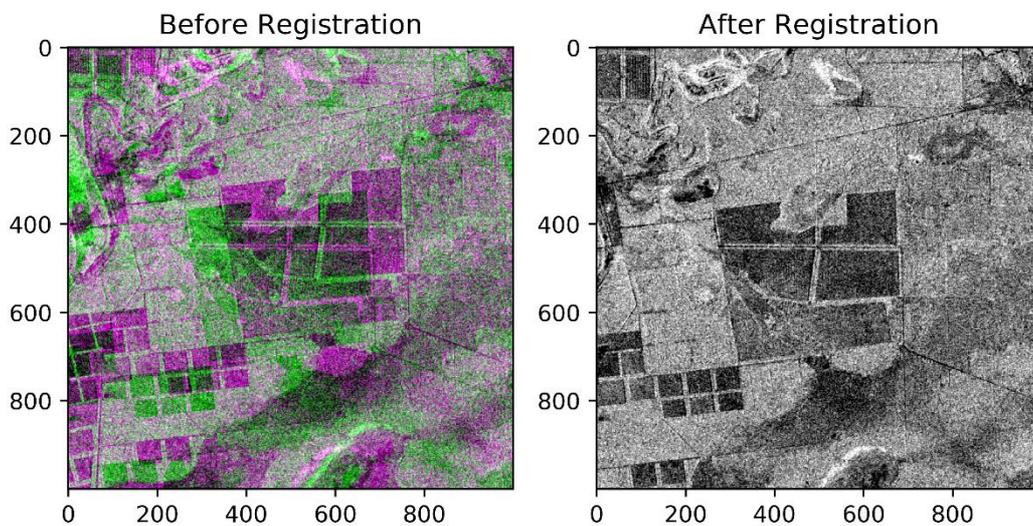


Figure 69. False color composition of the SAR images before (left) and after (right) the registration.

In these simple cases of optical-to-optical and SAR-to-SAR registration with only a shift transformation, the cross-correlation proves indeed a valid metric.

4.3.3.1.3 Synthetic Optical-to-SAR matching

Elapsed Time:	True Transformation:	Resulting Transformation:	Registration RMSE:
0.616 seconds	[-76, 52, 0, 1]	[-0.0, 52.0, 0.0, 1.0]	76.0 pixels

Figure 70 shows the reference optical image, the input SAR image to be registered, and the resulting cross-correlation with its point of maximum.

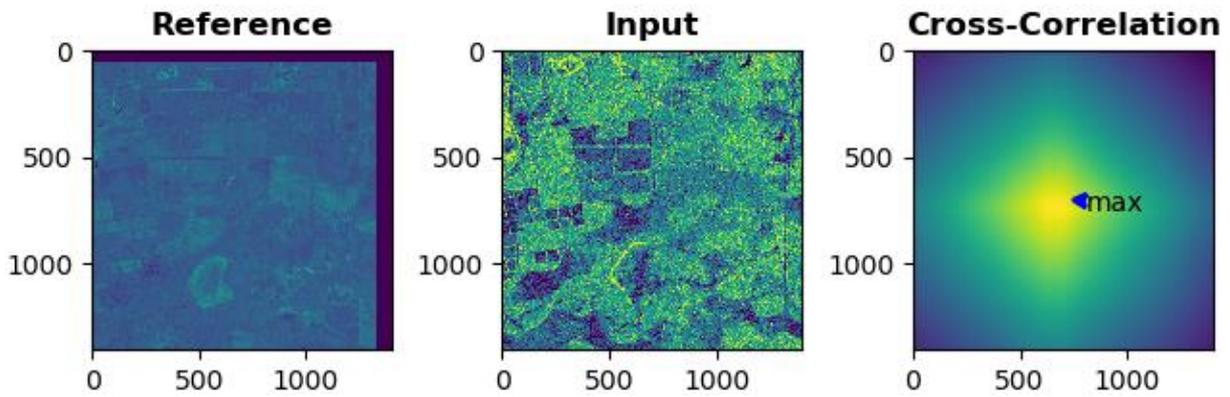


Figure 70. The reference optical image (left panel); the input SAR image to be registered (central panel); and the resulting cross-correlation with the corresponding point of maximum (right panel).

In this case, Figure 71 shows a checkerboard representation of the images before and after registration. On the left, alternate rectangles show either the input or the reference images. On the right, alternate rectangles show either the transformed input or the reference images. A qualitative assessment of the registration accuracy can be achieved by looking at the borders between each rectangle and at the continuity of linear image features across those borders.

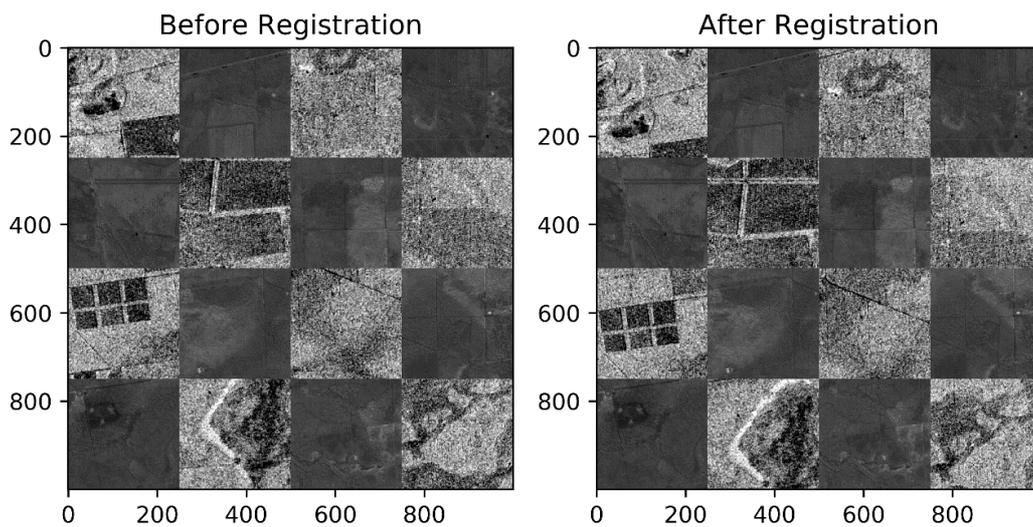


Figure 71. Checkerboard representation of the optical and SAR images before (left) and after (right) registration.

Such a result points out the non-satisfactory performances of using cross-correlation as a measure of similarity in the multi-sensor case.

4.3.3.2 Optimal translation using cross-correlation and mutual information

The difference with respect to the previous case is that, in these experiments, the optimal translation is found by maximizing either the cross-correlation or the mutual information. Again, experiments have been carried out with respect to both multi-sensor and single-sensor synthetic scenarios. For the sake of brevity, we report here the results obtained for the SAR-to-SAR and for the Optical-to-SAR registration using the COBYLA optimization method. The experiments with the optical-to-optical and with the Powell optimization methods showed comparable results. It is worth noting that COBYLA usually outperforms Powell's method when the solution is far from the initialization (for a complete analysis refer to the experiments with the RST transformation below).

4.3.3.2.1 Synthetic SAR-to-SAR matching

The results obtained by using cross-correlation or mutual information are similar in this synthetic case. What it is worth noting is the required time needed for convergence, which is longer in the mutual information case.

Cross-correlation

This experiment is similar to the one reported above, but here the cross-correlation is maximized using COBYLA and not by computing the whole cross-correlation function through the FFT. Figure 72 shows the result using the false-color composite defined for the previous single-sensor experiments.

Elapsed Time:	True Transformation:	Resulting Transformation:	Registration RMSE:
21.172 seconds	[-76, 52, 0, 1]	[-75.9, 51.9, 0.0, 1.0]	5.65 e-4 pixels

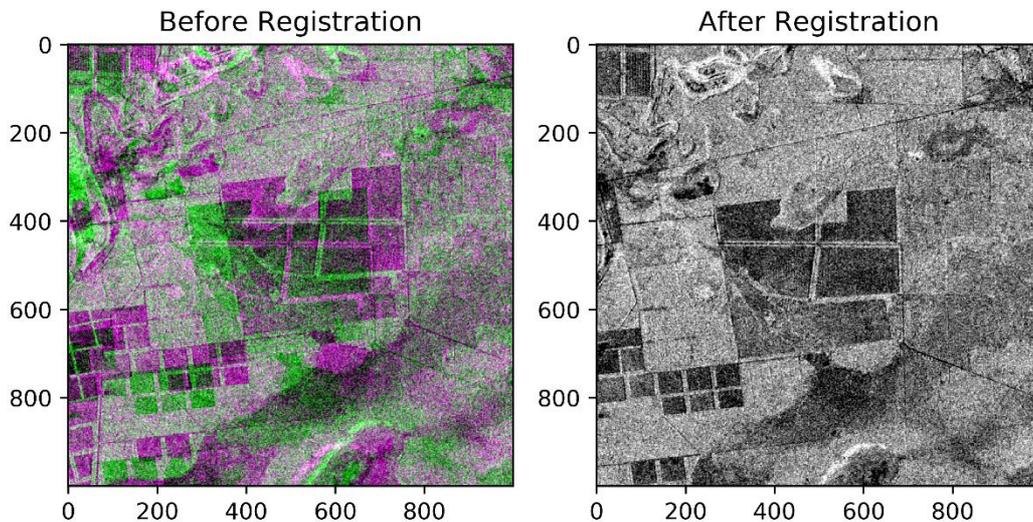


Figure 72. False color composition of the SAR images before (left) and after (right) the registration.

Mutual Information

Elapsed Time:	True Transformation:	Resulting Transformation:	Registration RMSE:
42.221 seconds	[-76, 52, 0, 1]	[-76.0, 51.9, 0.0, 1.0]	5.33 e-4 pixels

The false color composition in Figure 73 points out the accuracy of the registration result obtained. As anticipated, the accuracy when using either cross-correlation or mutual information is similar. The difference is the computation time, which is longer in the mutual information case, thus suggesting a preference for the cross-correlation metric in the single-sensor case.

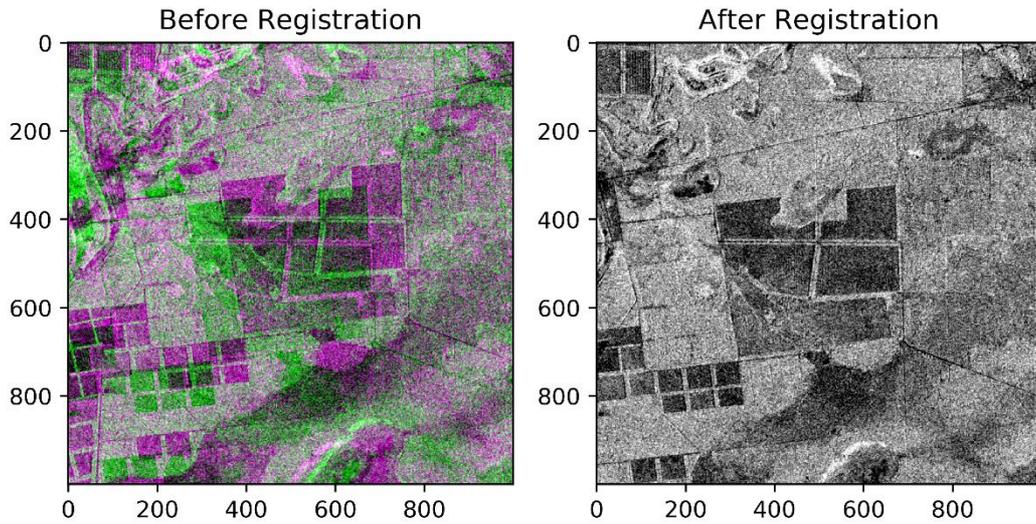


Figure 73. False color composition of the SAR images before (left) and after (right) the registration.

4.3.3.2.2 Synthetic Optical-to-SAR matching

This experiment, in combination with the one considering cross-correlation through FFT in the optical-to-SAR matching case, highlights the importance of using mutual information as a similarity measure in the multi-sensor scenario [35], [36].

Mutual Information

Elapsed Time:	True Transformation:	Resulting Transformation:	Registration RMSE:
7.59 seconds	[-76, 52, 0, 1]	[-81.5, 49.6, 0.0, 1.0]	5.99 pixels

The checkerboard representation of the images before and after registration shown in Figure 74 highlights that, even in the multi-sensor case, the mutual information is effective for determining the optimal translation.

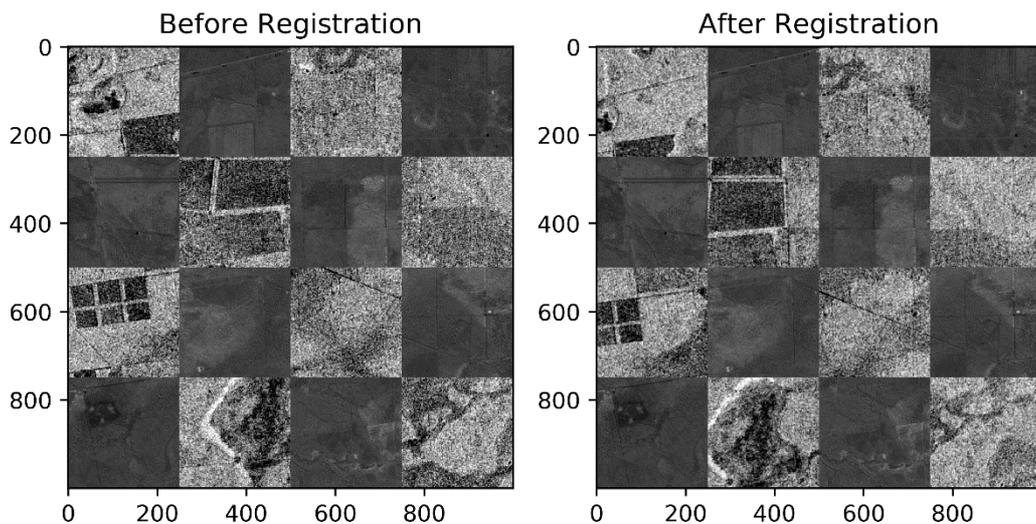


Figure 74. Checkerboard representation of the optical and SAR images before (left) and after (right) registration.

4.3.3.3 Optimal rigid transformation using cross-correlation and mutual information

As in the case above, in these experiments, the optimal transformation is found by maximizing either the cross-correlation or the mutual information using COBYLA. The difference is in the geometric transformation, since here the model is considered as a rigid transformation (i.e., translation + rotation).

For the sake of brevity, the experiments reported here consider: (i) a comparison of using cross-correlation with synthetic single-sensor data and synthetic multi-sensor data; (ii) a comparison of using mutual information with synthetic multi-sensor data and real multi-sensor data.

Experiment (i) highlights the poor performances of cross-correlation with multi-sensor data. It is worth noting that, according to these results, the use of cross-correlation will not be considered for the further multi-sensor experiments in this document. Conversely, experiment (ii) shows the satisfactory accuracy achievable in the synthetic case through the mutual information, but also confirms that the rigid transformation is often a too restrictive model for the application to real S1-S2 data.

4.3.3.3.1 Synthetic Optical-to-Optical Matching

Cross-correlation

Elapsed Time:	True Transformation:	Resulting Transformation:	Registration RMSE:
37.39 seconds	[-45, 26, 1.1, 1]	[-44.39, 25.34, 1.14, 1.0]	1.48 pixels

As expected, and as illustrated in the false color composition of Figure 75, an accurate registration was obtained in this case of rigid transformation of a single-sensor image pair.

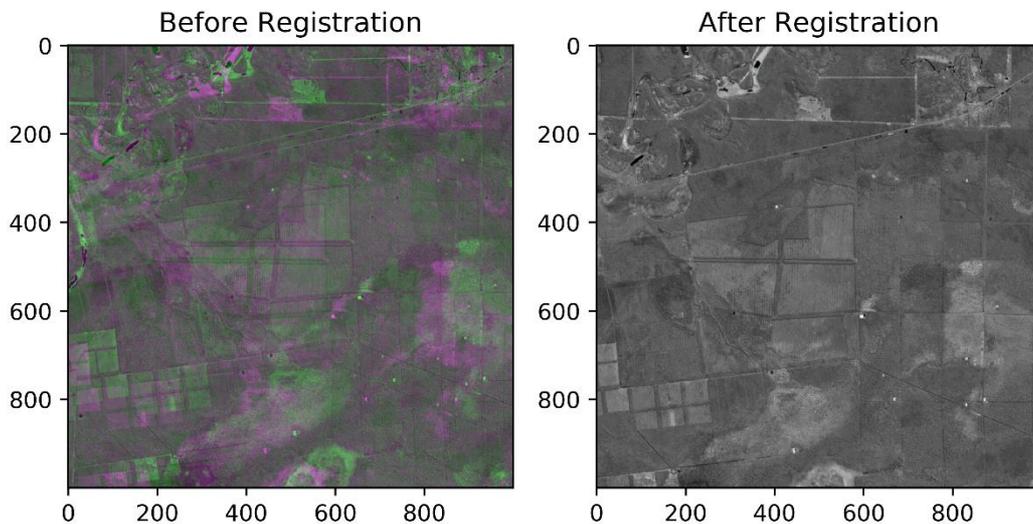


Figure 75. False color composition of the optical images before (left) and after (right) the registration.

4.3.3.3.2 Real Optical-to-SAR matching

Elapsed Time:	13.52 seconds	Resulting Transformation:	[66.44, -44.43, -0.75, 1.0]
---------------	----------------------	---------------------------	------------------------------------

Figure 76 shows a checkerboard representation of the images before and after registration. Well-registered areas are highlighted in green, while areas that still suffer from unprecise registration are highlighted in red.

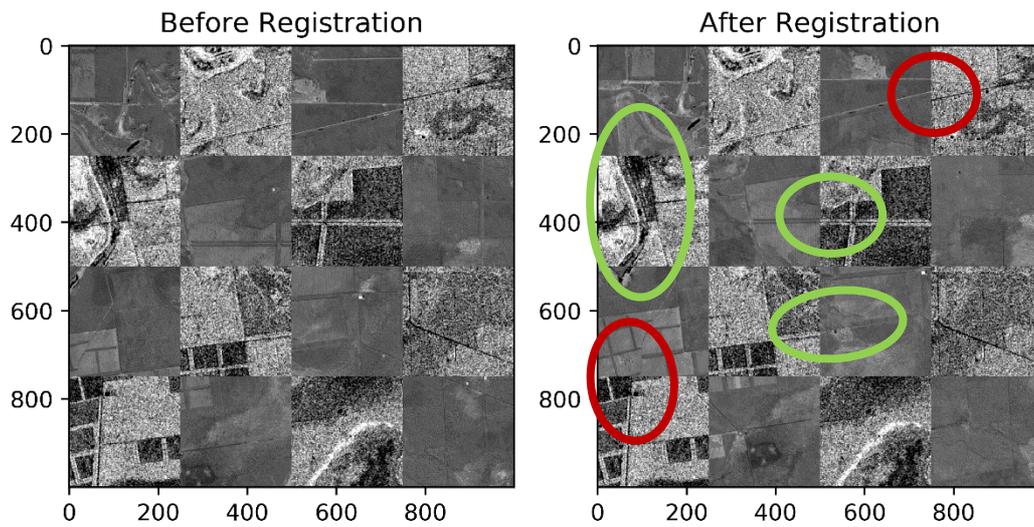


Figure 76. Checkerboard representation of the optical and SAR images before (left) and after (right) registration. Green and red circles highlight where the match is good or bad, respectively.

As anticipated above, on one hand, the registration through the cross-correlation improved as compared to the initial georeferencing. A particularly effective alignment can be noted in the green circles. Nevertheless, the resulting rigid transformation is not globally accurate, as it can be noted by inspecting the crop fields in the lower left corner of the image, together with the other areas highlighted in red.

4.3.3.4 Optimal RST transformation using mutual information

The following experiments consider the more general case of rotation-scale-translation transformations. The experiments reported in the following sections do not consider anymore the synthetic single-sensor cases to focus on the HRLC multi-sensor scenario.

Also in this case, the performances are evaluated both quantitatively, with synthetic data where the true transformation is known, and visually-qualitatively in the case of the real dataset.

The following experiments are also aimed at testing the registration method on different areas of the available Amazonian tile. Three separate areas (1000x1000 pixels each) have been registered using the RST transformation and the COBYLA and Powell's optimization strategies.

It is worth noting that the unconstrained Powell's algorithm performs well in the cases where the solution of the registration process falls in a neighborhood of the starting point. Conversely, the COBYLA method, with the initial search radius parameter, allows to better explore the search space and usually grants better convergence properties. This is shown especially in Areas from 2 to 4 in the following experiments, where the solution is significantly far away from the initialization and Powell algorithm fails to converge. Indeed, these results from the unconstrained Powell's algorithm is expected for a greedy unconstrained technique: on one hand, the lack of constraints allows exploring large portions of the parameter space; on the other hand, the greedy search approach does not guarantee that the convergence point, which is generally a local optimum of the adopted metric, is an effective solution, unless the initial configuration is already appropriate. From this perspective, a constrained reformulation of the method, which encodes prior information about the meaning of the parameters in terms of box constraints, may focus the exploration of the method to a relevant target subspace of the parameter space, thus leading to a more effective behavior.

4.3.3.4.1 Synthetic Optical-to-SAR Matching

Elapsed Time:	True Transformation:	Resulting Transformation:	Registration RMSE:
23.81 seconds	[-45, 42, 2.1, 0.98]	[-45, 42, 2.1, 0.98]	5.16 pixels

The checkerboard representation of the images before and after registration is shown in Figure 77. For all the experiments within the synthetic optical-to-SAR matching, the COBYLA optimization method has been used.

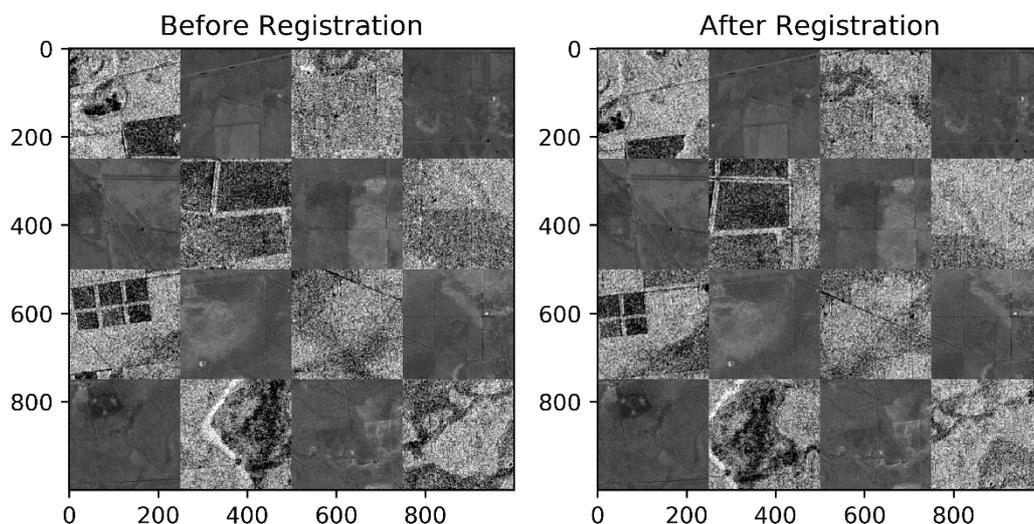


Figure 77. Checkerboard representation of the optical and SAR images before (left) and after (right) registration.

This result is comparable in accuracy with the case of rigid transformation. Nevertheless, it is worth noting that for the registration process to converge to a qualitative satisfactory result, the scaling factor should not be too large (or small). An experimental example is reported here, where a scaling factor of 0.9 prevented the registration to succeed. During the experiments, all the other parameters were kept constant and the scaling factor was reduced from 0.98 to 0.90.

With a scale factor reduced from 0.98 to 0.95, the rigid transformation still converged to a satisfactory result, as depicted in Figure 78.

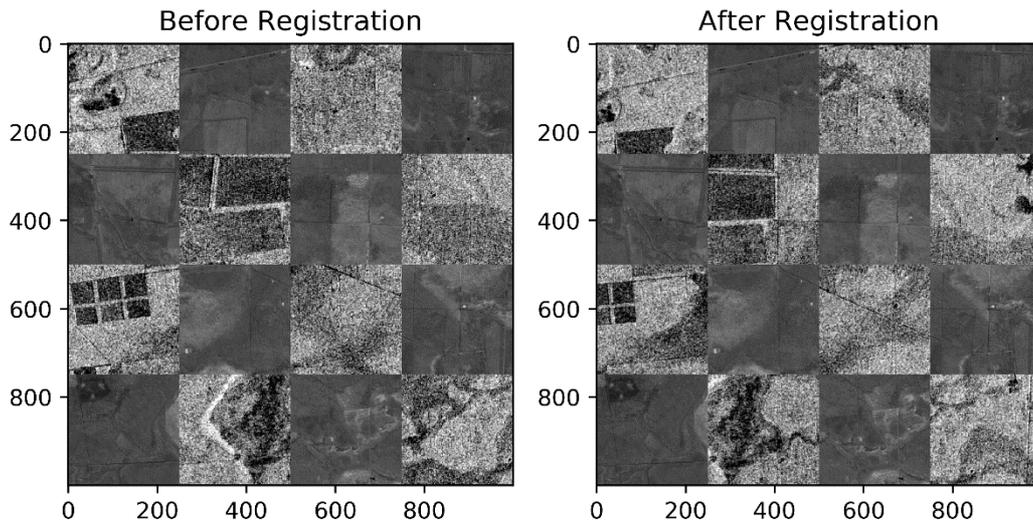


Figure 78. Checkerboard representation of the optical and SAR images before (left) and after (right) registration. Scale in range 0.98 to 0.95.

Scaling even further to 0.92 still granted acceptable results, as depicted in Figure 79.

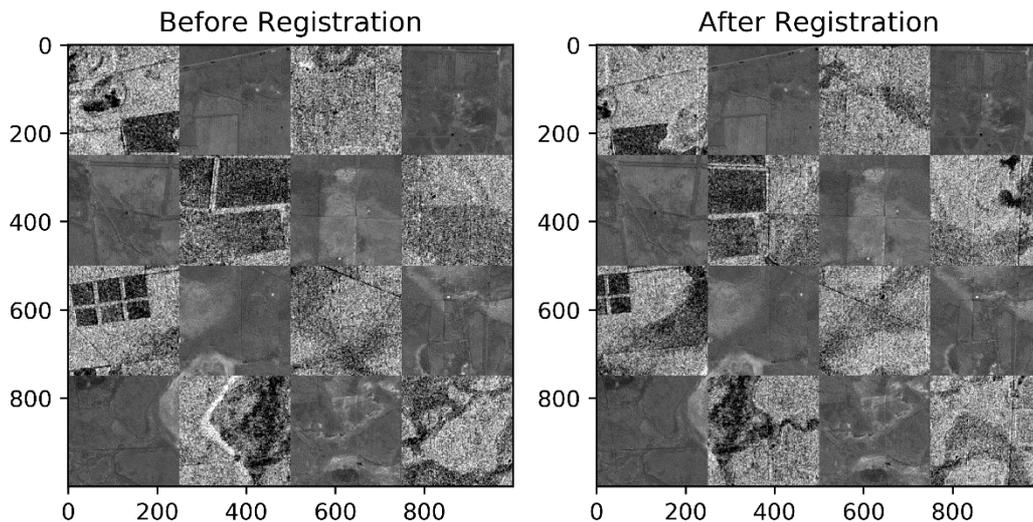


Figure 79. Checkerboard representation of the optical and SAR images before (left) and after (right) registration. Scale equal to 0.92.

However, further tests failed to converge effectively. An example is reported below in Figure 80, where the scaling factor was chosen equal to 0.9 and the other parameters were kept constant.

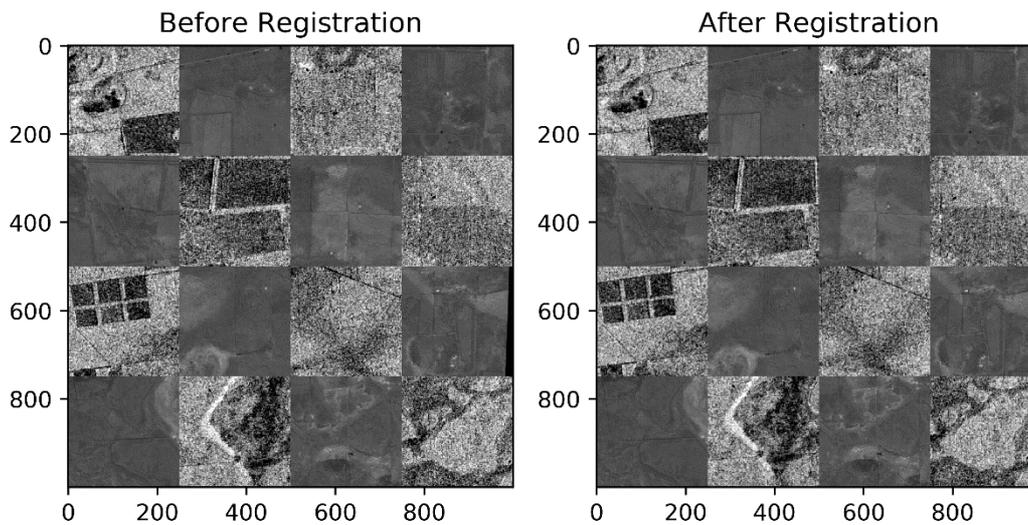


Figure 80. Checkerboard representation of the optical and SAR images before (left) and after (right) registration. With a scale factor of 0.9 the matching is not completely satisfactory.

This analysis is important especially with respect to the COBYLA minimization method. Indeed, such technique allows defining constraints in the search space, thus preventing the scale parameter to assume excessive values. This consideration is also consistent with the project, as the available images are not expected to be characterized by a large-scale difference. Indeed, as the input S1 and S2 data are first georeferenced on pixel grids associated with the same nominal pixel spacing of 10 m, the possible scaling factor is indeed expected to take values in a narrow neighborhood of unity.

4.3.3.4.2 Real Optical-to-SAR Matching: Area 1

Elapsed Time:	13.24 seconds	Resulting Transformation:	[76.40, -44.31, -0.17, 0.99]
----------------------	----------------------	----------------------------------	-------------------------------------

The checkerboard in the Figure 81 points out that this result is comparable with the one obtained in the case of rigid transformation (i.e., only translation and rotation) in the application to the real multi-sensor dataset as well. The scale factor does not have great importance in the registration process of this dataset, as it is suggested by the convergence value of 0.99 and in accordance with the aforementioned comments on the expected behavior of this parameter. COBYLA is used for minimization both for Area 1 and for all the following experiments considering other areas.

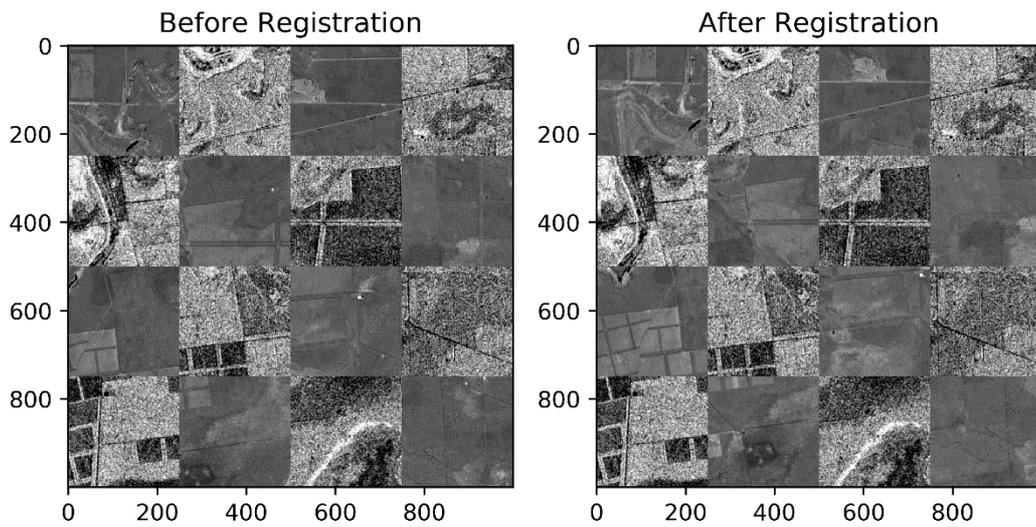


Figure 81. Checkerboard representation of the optical and SAR images before (left) and after (right) registration.

4.3.3.4.3 Real Optical-to-SAR Matching: Area 2

Elapsed Time:	13.39 seconds	Resulting Transformation:	[77.52, -62.89, 0.16, 1.03]
---------------	---------------	---------------------------	-----------------------------

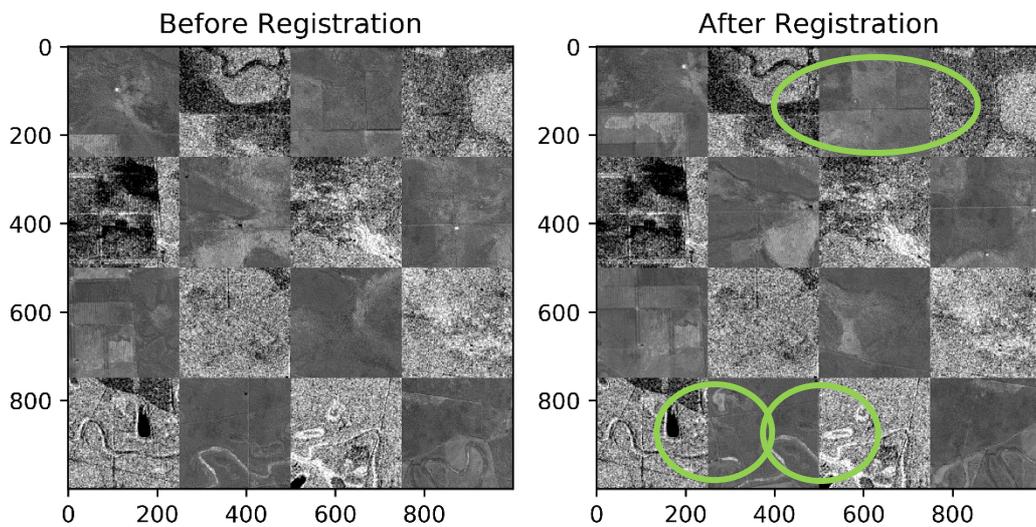


Figure 82. Checkerboard representation of the optical and SAR images before (left) and after (right) registration. The green circles highlight the good matching.

The same comments made for Area 1 hold with regard to Area 2 as well. See Figure 82 in particular, the green ellipses in the checkerboard emphasize regions where accurate registration of linear image features is especially evident.

4.3.3.4.4 Real Optical-to-SAR Matching: Area 3

In this area, the spatial difference between the reference and input images is large, especially in the translation along the x axis. COBYLA was able to reach an effective registration result, though, while the unconstrained Powell's method failed to converge due to the large distance between the solution and the initialization point. In Figure 83 and Figure 84 the green and red ellipses superimposed to the following two sets of checkerboards emphasize this significant difference in performance.

COBYLA

Elapsed Time:	15.12 seconds	Resulting Transformation:	[222.11, -3.23, 0.48, 1.00]
----------------------	----------------------	----------------------------------	------------------------------------

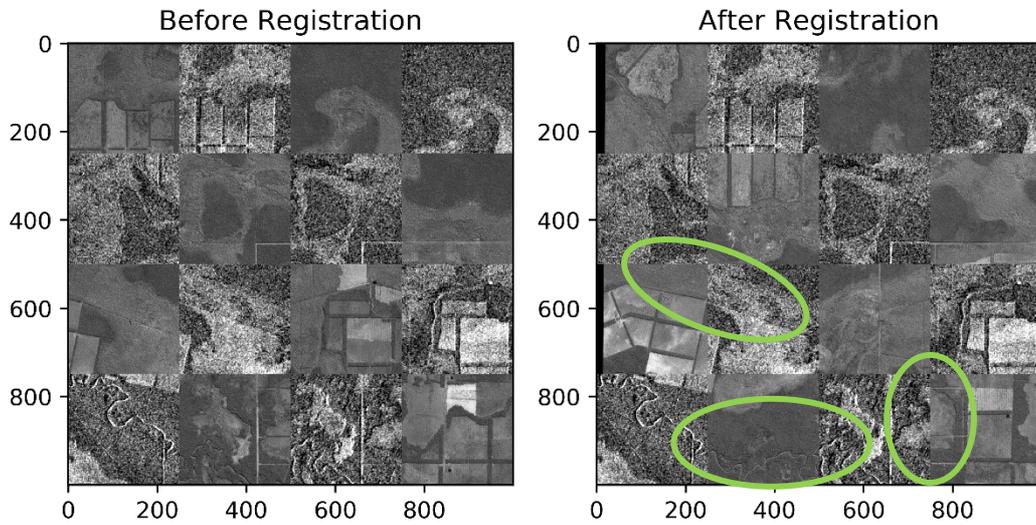


Figure 83. Checkerboard representation of the optical and SAR images before (left) and after (right) registration. The green circles highlight the good matching.

Unconstrained Powell

Elapsed Time:	46.29 seconds	Resulting Transformation:	[28.07, 16.23, 0.06, 1.00]
----------------------	----------------------	----------------------------------	-----------------------------------

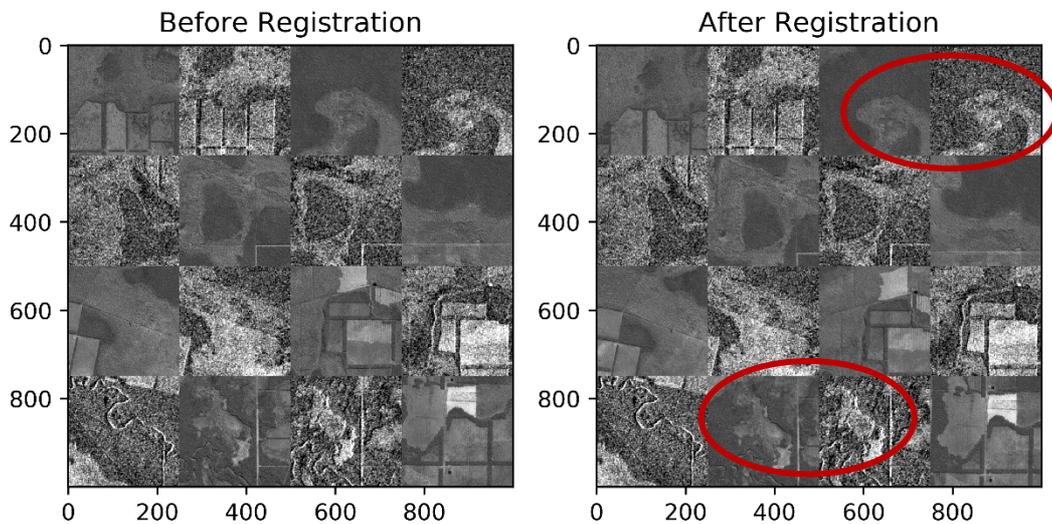


Figure 84. Checkerboard representation of the optical and SAR images before (left) and after (right) registration. The red circles highlight the major misalignments.

4.3.3.4.5 Real Optical-to-SAR Matching: Area 4

As for Area 3, also Area 4 is characterized by a large spatial difference, which is even larger than the previous case, see Figure 85 and Figure 86. Nevertheless, COBYLA succeeded to find a good transformation fitting the input and the reference images, and outperformed the unconstrained Powell's method.

Another experimental advantage of COBYLA is a lower time needed for convergence, most probably due in this case to the fact that Powell failed to find a good matching. However, the short convergence time suggests the possibility of tuning online the "initial search radius" parameter. Indeed, it should be possible to implement a

new version of such registration procedure that integrates a grid search on that parameter. The resulting similarity metrics could be compared to decide which value of such parameter is able to grant the best registration result for the dataset at hand.

COBYLA

Elapsed Time:	13.33 seconds	Resulting Transformation:	[270.68, -55.25, 0.51, 1.02]
----------------------	----------------------	----------------------------------	-------------------------------------

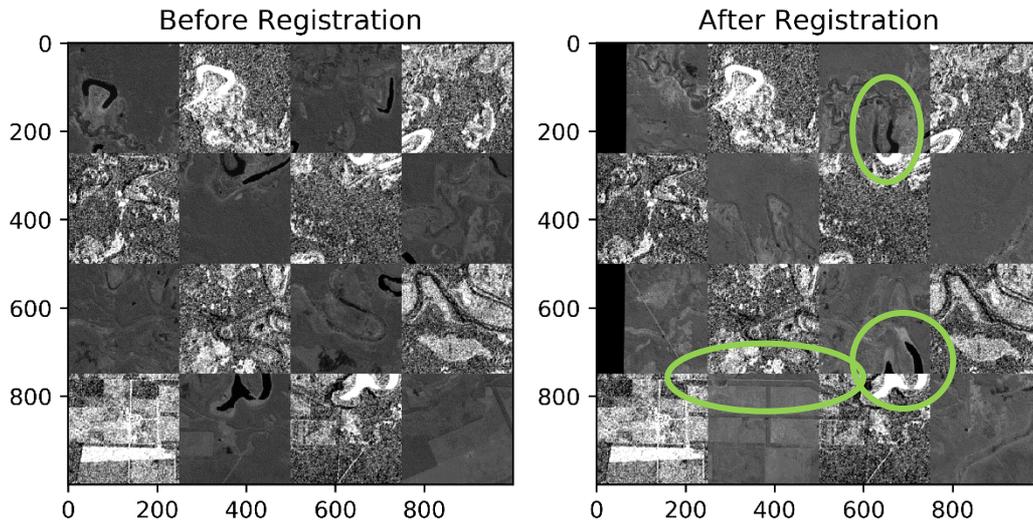


Figure 85. Checkerboard representation of the optical and SAR images before (left) and after (right) registration. The green circles highlight the good matching.

Unconstrained Powell

Elapsed Time:	38.93 seconds	Resulting Transformation:	[3.63, -1.58, 0.01, 1.00]
----------------------	----------------------	----------------------------------	----------------------------------

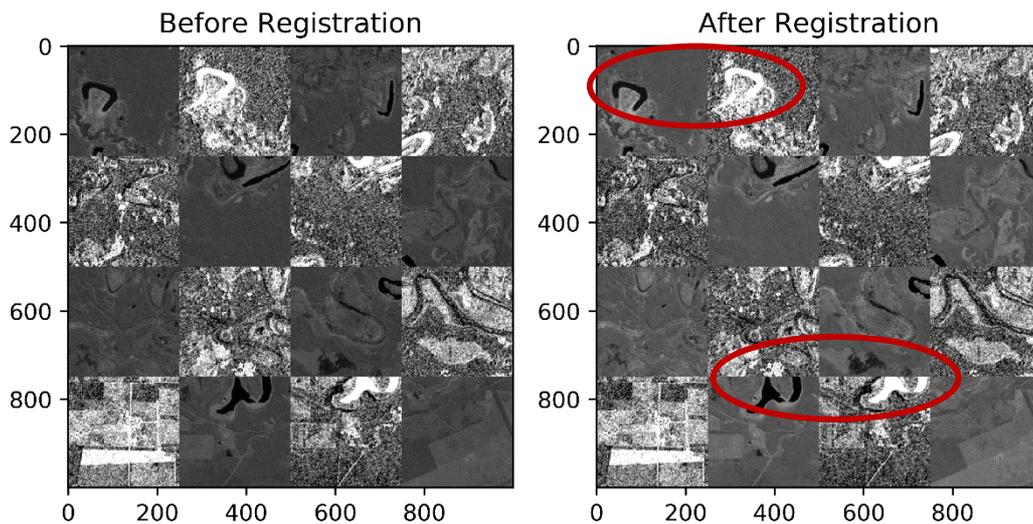


Figure 86. Checkerboard representation of the optical and SAR images before (left) and after (right) registration. The red circles highlight the major misalignments.

4.3.4 Quantitative evaluation of Powell's algorithm with barrier functions and automatic tiling

This section focus on the latest results obtained on multi-sensor data from the Round Robin areas for the modified version of Powell's method that integrates barrier functions to support box constraints on the parameters. This also allows for further comparison between COBYLA and Powell's abilities to converge to an effective solution. Moreover, integrating barrier functions into Powell's algorithm allows to avoid the need to correctly tune the search radius characteristic of COBYLA, allowing for faster convergence as compared to COBYLA itself. These experiments are also integrated with those associated to the automatic tiling algorithm that has been developed for the application to images as large as the entire S2 granule.

4.3.4.1 Quantitative Analysis on Siberia

The first analysis is performed on a Siberian area, in particular, tile 42WXS. Figure 87 represents the Sentinel-2 and Sentinel-1 seasonal composites for the summer season of 2019 in the area identified by the Sentinel-2 granule 42WXS.

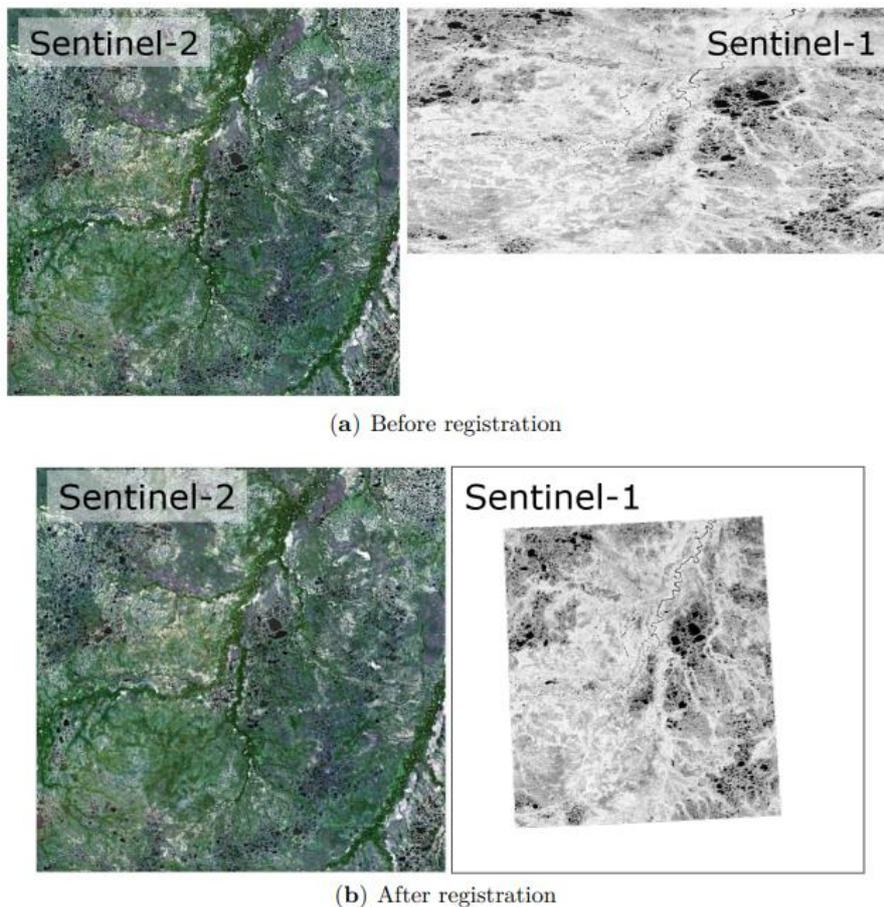


Figure 87. Sentinel-2 (left) and Sentinel-1 (right) images, as received (top) and after registration (bottom).

Panel (a) shows the images before registration, while Figure panel (b) shows the registration result. The white part of the registered Sentinel-1 image corresponds to the area where no data is present (e.g, the input S-1 image does not contain the information corresponding to the whole S-2 42WXS granule). Indeed, the output images must share the same pixel lattice for the further processing steps aiming at jointly benefiting from the input optical and SAR data. This is especially important in the context of the CCI+ HRLC processing chain (i.e., the two classification steps and the fusion step), but the very same considerations are valid for any other application aiming at jointly extracting information from optical and SAR data.

The next figures allow for a more detailed analysis of the registration results:

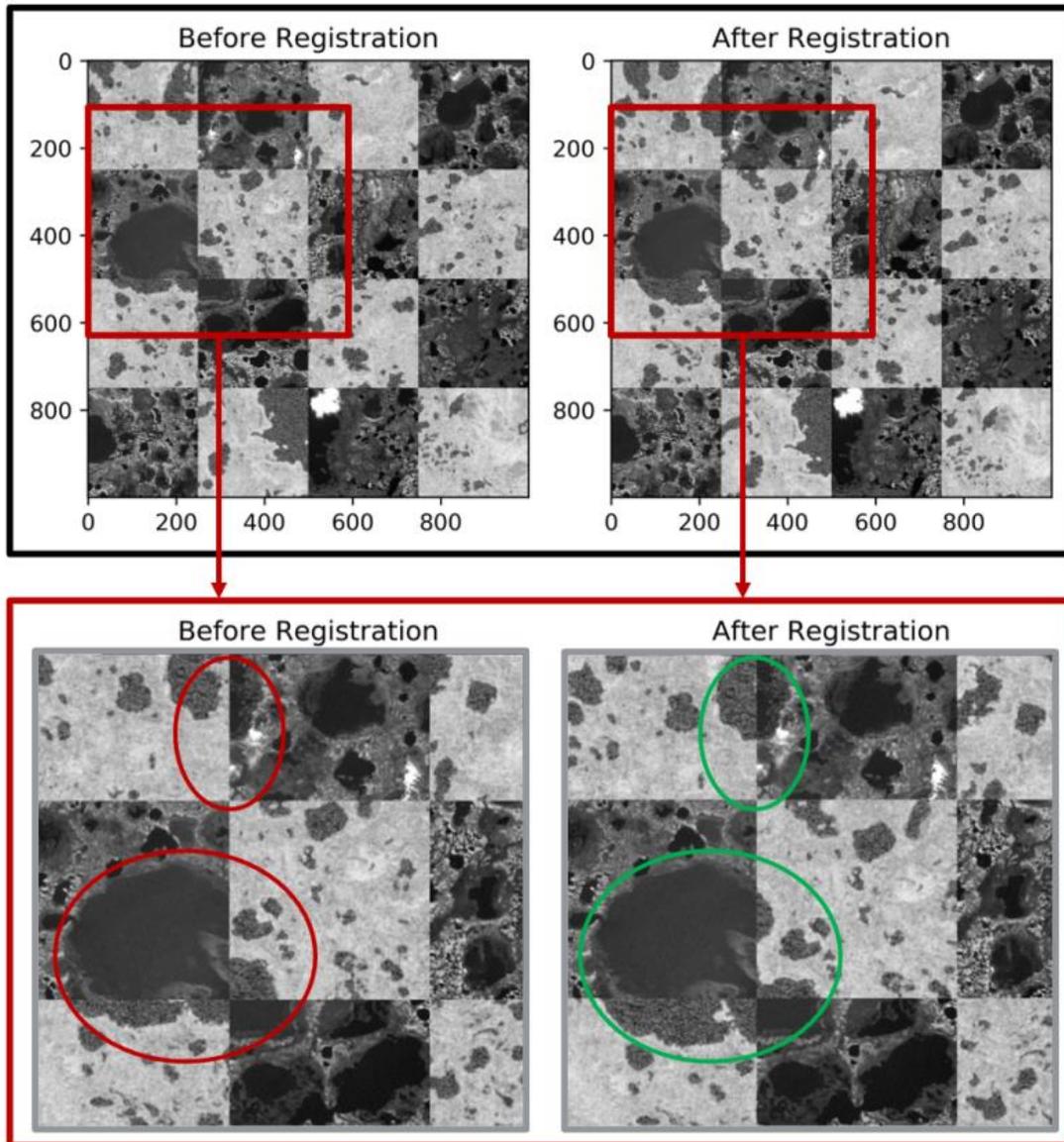


Figure 88. The top panel present a checkerboard representation of the optical and SAR image tiles before (left) and after (right) registration, for the granule 42WXS. On the bottom, a zoomed detail, with red and green circles highlighting the zones in which the effect of the registration can be better appreciated, respectively before and after.

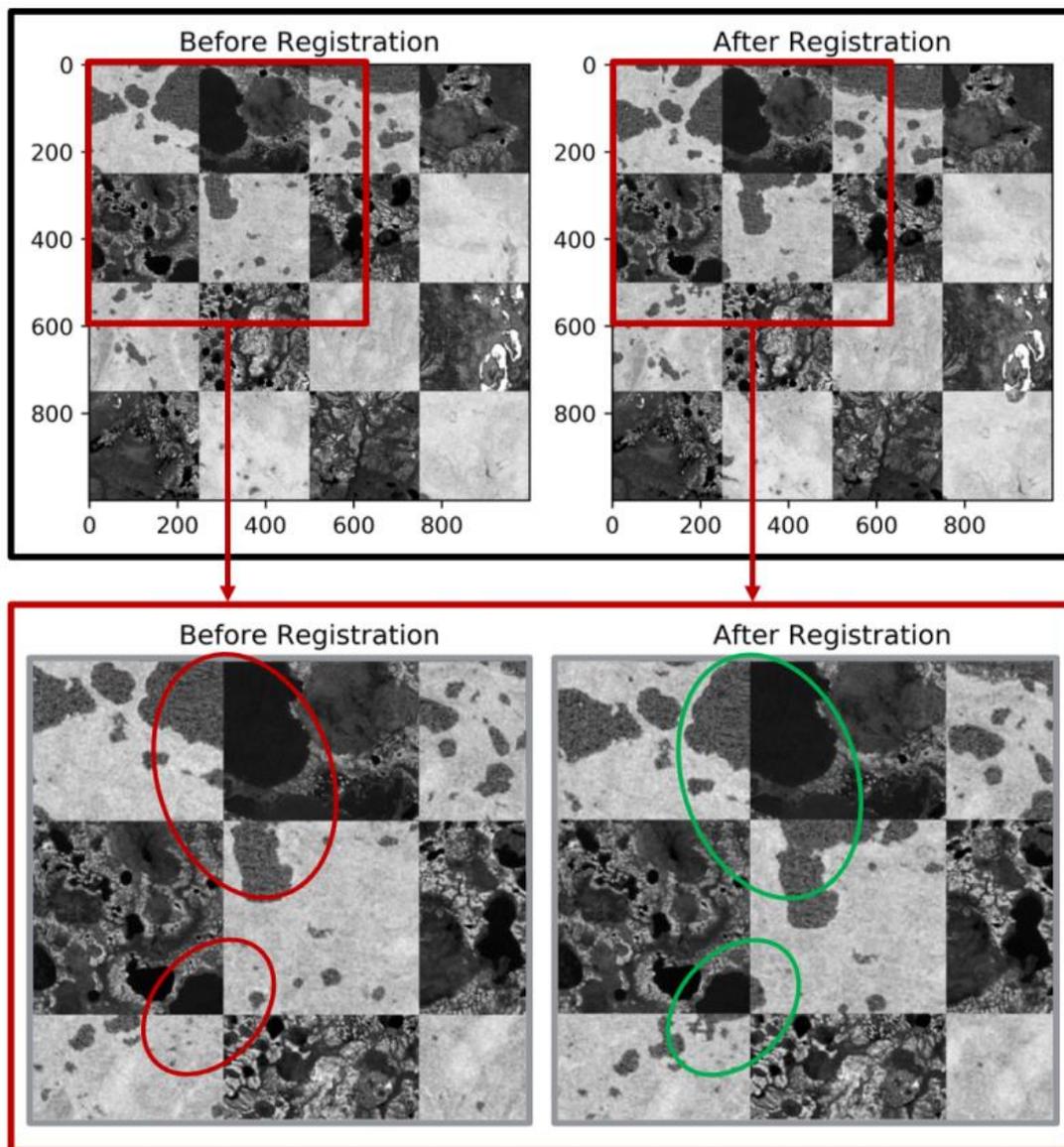


Figure 89. The top panel present a checkerboard representation of the optical and SAR image, corresponding to another tile in which granule 42WXS is divided into, before (left) and after (right) registration. On the bottom, a zoomed detail, with red and green circles highlighting the zones in which the effect of the registration can be better appreciated, respectively before and after.

Indeed, Figure 88 and Figure 89 show the registration results for two of the patches that the input images (belonging to the Siberian S-2 granule 42WXS) are divided into. In particular, in both cases, the images in the top panels show a checkerboard visualization of the situation before and after the registration is performed. Neighbouring squares show either the optical or the SAR image, allowing the reader to appreciate the matching and mismatching of the spatial features in the area. Similarly, the images in the bottom panels show the same checkerboard visualization, but considering only a zoomed detail of the patch. In this case, red and green circles have been added to the picture in order to highlight the most significant spatial features. Both figures highlight the effectiveness of the registration process, using the constrained variation of Powell's algorithm and the mutual information as metric.

Figure 90 displays the registration results for other two patches. For the sake of brevity, only the registration results are shown. Also in this case, it is possible to appreciate the matching in the rivers and in the bodies of water:

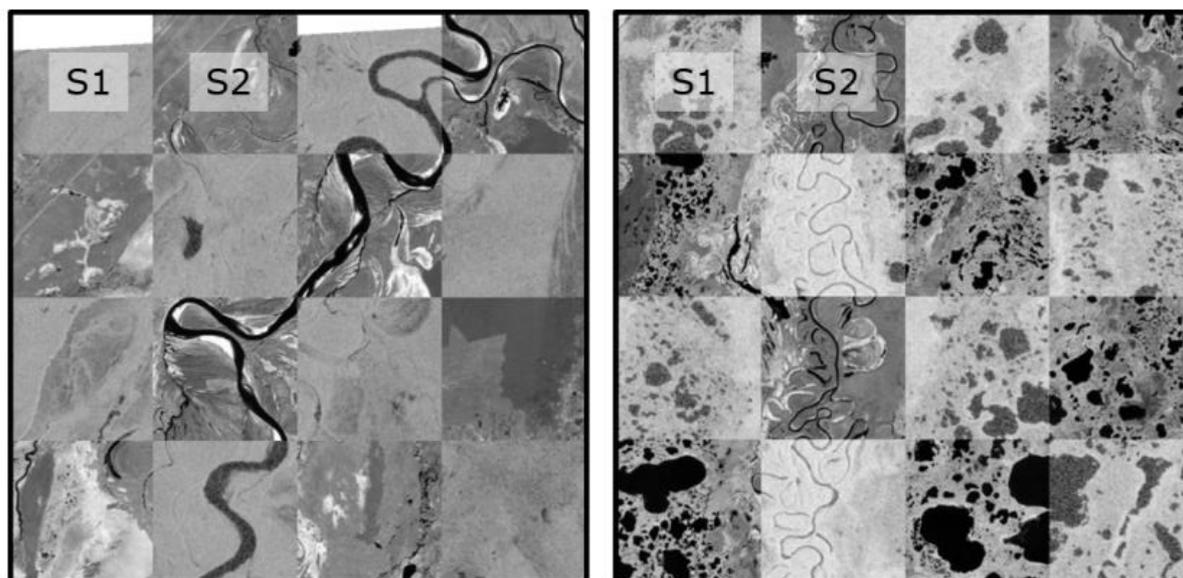


Figure 90. Checkerboard representation of the results achieved by the registration for granule 42WXS, with details of rivers.

Concerning the quantitative analysis, the proposed tiling-based strategy has also been evaluated in terms of root mean squared error by generating a semi-synthetic dataset from pairs of manually registered images. Two registered images (corresponding to the Siberian S-2 granule 42WXS) have been warped with a set of known RST transformations. Then, the proposed method has been applied to the resulting unregistered pairs.

Due to the tiling-based processing, for each one of the pairs, the root mean square errors obtained in each patch are averaged to get the final score. In order to also quantitatively assess the gain that the proposed solution allows in terms of computation time, the same semi-synthetic datasets have been used for experiments with the straightforward application of the area-based registration method, without the tiling strategy. As a consequence, the time needed for convergence, together with the obtained root mean squared error, has been compared to the one achieved by the proposed automatic tiling solution.

It is worth recalling that the registration problem faced by the area-based method (without tiling) is intrinsically less complex than the one faced by the proposed method. Indeed, the synthetic transformations applied to construct the synthetic dataset belong to the family of global RST transformation models. On the one hand, there exist a feasible solution that the area-based method can converge to, thus allowing a fair comparison. On the other hand, the dimensionality of the problem faced by the proposed solution is much higher, due to the larger search space. Nevertheless, comparing the results allow us to appreciate the level of accuracy that the proposed method is able to reach. The results of the comparison are reported in the following table:

	Entire granule	Proposed automatic tiling
Time [min]	68	31
RMSE [pixels]	2.1	2.6

The computation times are measured in minutes, while the root mean squared errors are measured in terms of pixel units. The numbers have to be considered as averaged across a synthetic dataset generated from a set of 5 transformations applied to the Siberian and the Amazonian images. The table allows us to appreciate a significant

reduction in the computation time, thus assessing the capability of the proposed method to achieve a speedup of around 2.2, while also achieving accuracy performances that are comparable to those obtained by the area-based method (which operates in the aforementioned favourable setup). We also note that, as mentioned above, the traditional approach used here for comparison could obtain low RMSE values also because a unique well-defined parametric transformation, valid across the whole scene, existed by construction. Indeed, in a registration problem on a large-scale data set (as the in CCI+ HRLC), this is normally not the case, because the geometric relation between the two input images may generally be space variant and a unique parametric transformation usually does not exist.

4.3.4.2 Quantitative analysis on Amazonia

The same analysis has been conducted on the two tiles of the Amazonian RR, in particular tile 21KUQ and 21KXT. Figure 91 allow the reader to appreciate the effectiveness of the registration method when run on different geographical areas:

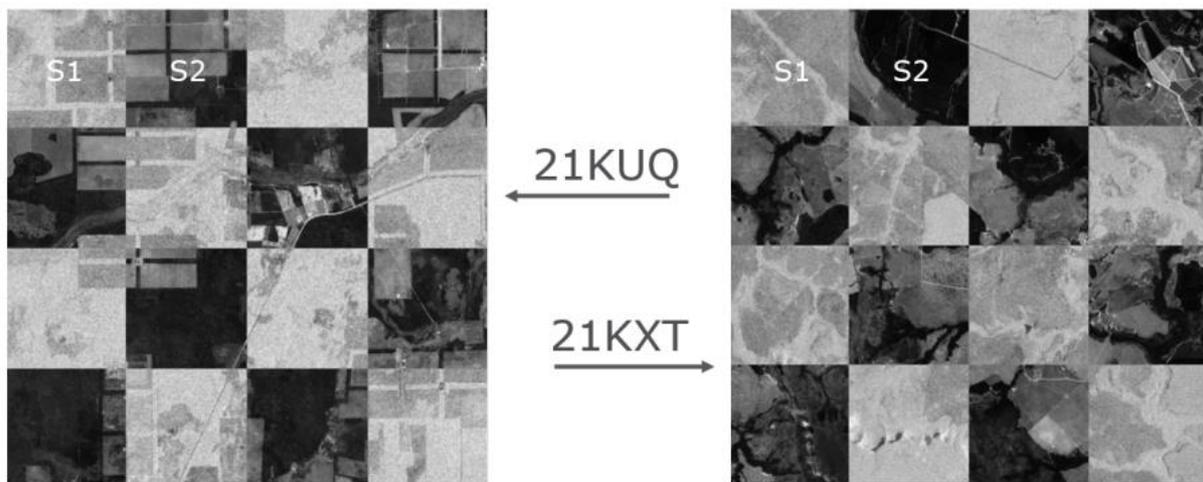


Figure 91. Checkerboard representation of the results achieved by the registration, with details for the granules 21KUQ (left) and KXT (right).

The spatial features present in the scenes are different with respect to the ones in the Siberia area, and the registration method has been proven to be robust enough to adapt accordingly. Also in this case, the proposed tiling-based strategy has been evaluated in terms of root mean squared error by generating a semi-synthetic dataset from pairs of manually registered images. Again the computational time is almost halved by the proposed method while the RMSEs are similar. The same comments on the existence of a well-defined unique global transformation in this semi-synthetic case – and generally not in the real case – hold for this image pair as well.

	Entire granule	Proposed automatic tiling
Time [min]	51	27
RMSE [pixels]	2.0	2.2

4.3.4.3 Qualitative analysis on Africa

The study on Africa was meant to investigate the robustness of the proposed method also in those cases where the spatial features are very few. This is clearly noticeable in Figure 92, in which the lake cover almost the entire patch used for the registration and consists of a nearly flat surface that is almost uninformative for registration purposes:



Figure 92. Details of the spatial feature in one of the tiles in which the granule 37PCP is divided into. On the left, Sar image; in the centre, optical image; on the right, a checkerboard representation of the results achieved by the registration, even with very few features available.

Also in this particular case the method is able to achieve accurate results. This suggests the robustness of the algorithm to a lack of spatial features.

4.3.4.4 Results on historical images

With this analysis, the goal is to investigate the effectiveness of the method also in the application to images from the sensors involved in the historical HRLC product. Indeed, Figure 93 and Figure 94 below show the results of: (i) an experiment conducted with 2005 data acquired by Landsat-7 (optical image) and ENVISAT ASAR (radar image); and (ii) an experiment considering 2015 data acquired by Landsat-8 (optical image) and Sentinel-1 (radar image). Again, the geographical area that is taken into consideration is the same used for the experiments illustrated before (i.e., the S-2 granule 42WXS in Siberia).

2005 Experiment in Amazonia

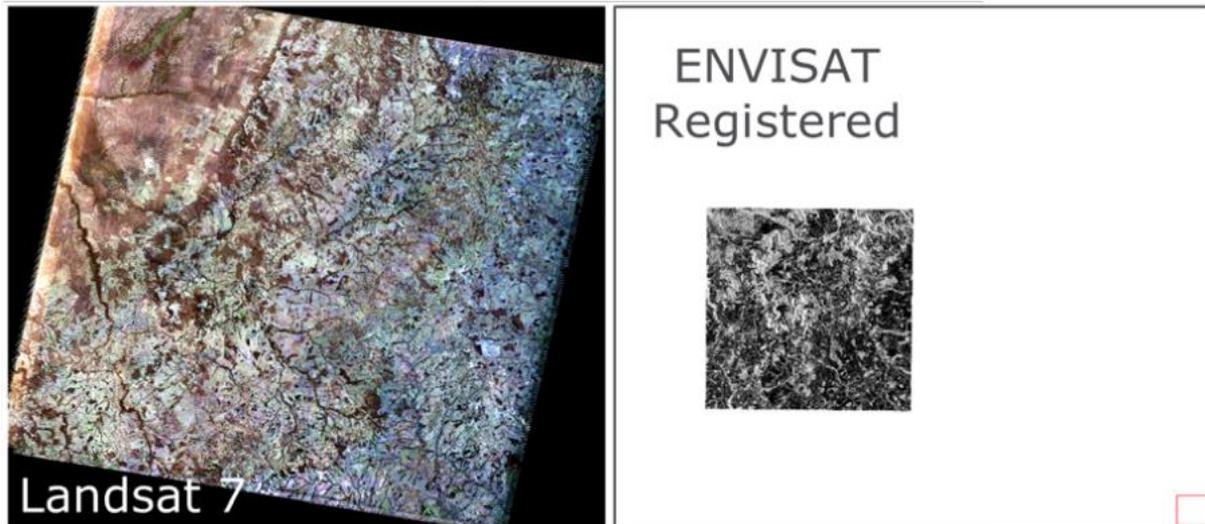


Figure 93. Registered Landsat 7 (left) and ENVISAT (right) images acquired in 2005.

2015 Experiment in Siberia

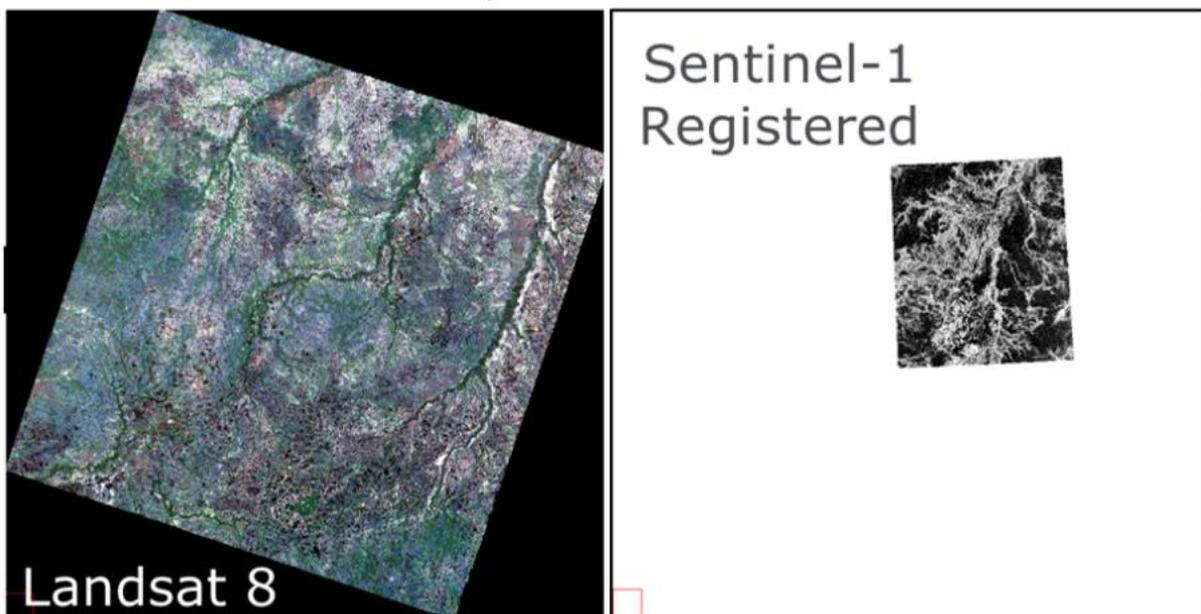


Figure 94. Registered Landsat 8 (left) and Sentinel-1 (right) images acquired in 2015.

We recall that the inputs to the multi-sensor geolocation module are the composites generated by the optical and SAR pre-processing chains. This explains the lack of artifacts in the aforementioned Landsat-7 image (post scan line corrector failure).

To investigate the registration accuracy, Figure 95 shows the result on one of the patches the images are divided into. Green circles have been added to the images to highlight the good matching obtained between the spatial features in the multisensor dataset coming from historical images as well:

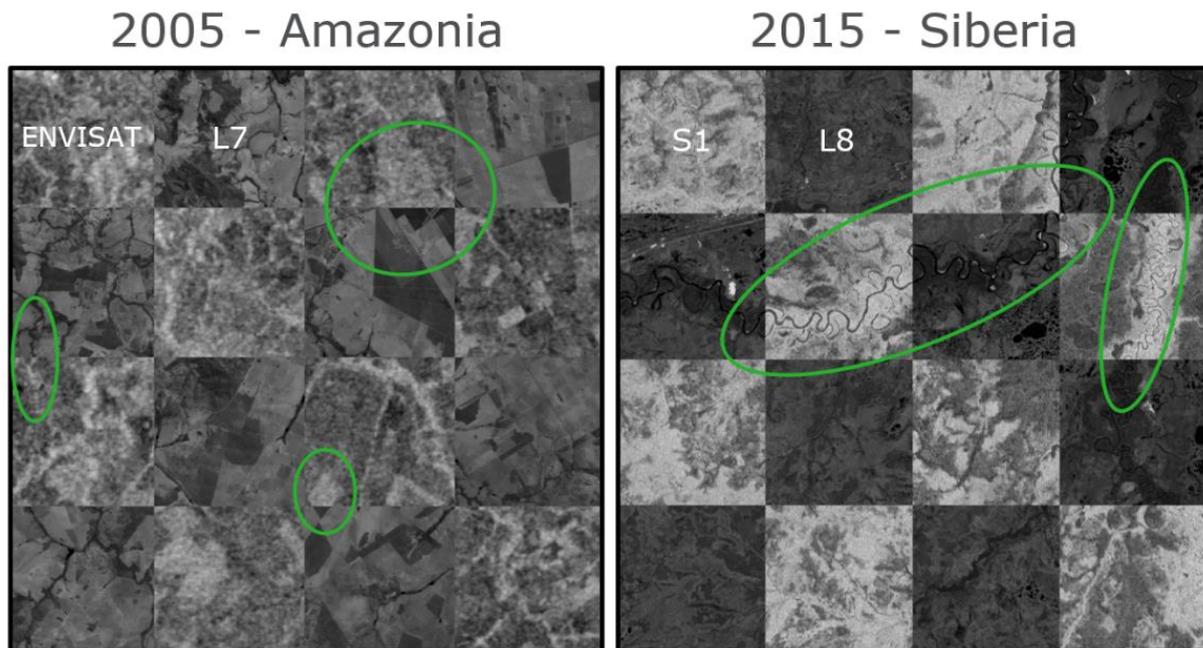


Figure 95. Checkerboard representation of the results achieved by the registration, for ENVISAT and Landsat 7 in 2005 (left) and for Sentinel-1 and Landsat 8 in 2015 (right). The green circles highlight appreciable matchings.

4.3.5 Final decision

The experimental analysis conducted with respect to multi-sensor geolocation pointed out a list of guidelines that indicated which registration strategy was pursued in the pipeline for the first production and suggested possible directions where to focus the future work on the topic. The following list enumerates such considerations and gives some insight in each of the points:

- The choice of the similarity metric.
Both cross-correlation and mutual information have been experimented with. The results pointed out that cross-correlation is valuable in the single-sensor case, while mutual information should be preferred in the multi-sensor scenarios. As a result, for the CCI+ HRLC product, mutual information have been prioritized. As discussed in ADP-v2, different implementations and approximations of such metric will be taken into account in the future (e.g., Parzen window estimation, Parzen window applied to a random selection of sample to reduce the computation time, etc.). Future work will consider the possible effectiveness of other metrics.
- The choice of the geometric transformation.
The experiments pointed out that, at least with respect to the available dataset, a simple translation transformation is not sufficient to find a good match between the input and reference images. However, rigid and RST transformations performed equally, as the available dataset was not affected by a consistent scale factor between the input couple. The last two transformations have therefore been considered and the HRLC pipeline makes use of the more general RST. The exploration of additional higher-order transformations (e.g., affine transformation) seems not to be necessary but can be considered in the future. Indeed, the proposed automatic tiling algorithm is also integrated into the HRLC pipeline to support an efficient application to large scale data. From the viewpoint of image transformation, the use of this algorithm is equivalent to the application of non-global transformation that is patchwise-RST.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	97	

- The choice of the optimization strategy.
Both the unconstrained Powell's method and COBYLA performed effectively when the spatial difference between the reference and input images was not large. On one hand, COBYLA outperformed the unconstrained Powell's method in the case of very large transformations. On the other hand, the constrained extension on the Powell's algorithm, which is based on the use of barrier functions, did not share this limitation of the original unconstrained Powell's formulation. The combination of this case-specific optimization algorithm and of the aforementioned automatic tiling method overall performed best, obtaining low errors with short computational times. Jointly, the tiling solution allows to compute non-global transformation, addressing for space variant distortion (common in large scale datasets). Future work will explore the possibility of integrating more sophisticated minimization methods (e.g., global search methods like genetic algorithms and simulated annealing). In this case, the focus will be in analyzing the time versus accuracy tradeoff.

4.3.6 Decision Fusion Methods

Decision fusion methodologies are used in the HRLC project in order to combine the posterior probabilities coming from the disjoint classification of SAR and optical images. Different families of decision fusion methods have been implemented and experimentally compared, including consensus-theoretic methods and fusion strategies based on Markovian modelling (i.e., Markov random fields), integrating a class-specific matching rule, and leading to the solutions adopted for the first production. Experimental comparisons have been focused on the round robin areas.

Specifically, the posterior probabilities coming from the optical and SAR processing chains are quantized and coded into unsigned integers using 8 bits per class and per pixel. This choice is aimed at minimizing memory requirements without any expected loss in appreciable precision. As discussed in the deliverables of the previous milestones, the sets of classes that can be accurately discriminated by using optical and SAR data exclusively generally do not coincide. In particular, SAR data are expected to well discriminate especially built-up classes and water bodies. Accordingly, SAR and optical classification algorithms may generally work on different sets of classes. Decision fusion methodologies are aimed at fusing posterior probabilities related to the common classes. Hence, a class-specific combination rule has been devised to take this into account.

4.3.6.1 Consensus theory and class-specific combination

Consensus theory [37],[38] involves general procedures with the goal of combining multiple probability distributions to summarize their estimates in a non-contextual manner. Under the assumption that both the SAR and the optical 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 most common consensus theory methods are linear opinion pool (LOP) and logarithmic opinion pool (LOGP) [37],[38]. Both these methods were implemented and tested in combination with the aforementioned class-specific rule.

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 j th 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

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	98	

associated with the corresponding sets of classes. Deriving $P(\omega_j|\underline{Q}, \Omega_C)$ and $P(\omega_j|\underline{S}, \Omega_C)$ (as well as $P(\omega_j|\underline{Q}, \Omega_O)$ and $P(\omega_j|\underline{S}, \Omega_S)$) is straightforward.

According to the same definitions, the logarithmic opinion pool (LOGP) functional can be defined as:

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

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. $P(\omega_j|\underline{Q}, \Omega_C) = 0$ 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:

$$\begin{aligned} P_{\mathcal{F}}(\omega_j|\underline{x}) &= P(\omega_j|\underline{x}, \Omega_C)P(\Omega_C|\underline{x}) + P(\omega_j|\underline{x}, \Omega_O)P(\Omega_O|\underline{x}) + P(\omega_j|\underline{x}, \Omega_S)P(\Omega_S|\underline{x}) = \\ &= P_{\mathcal{F}}(\omega_j|\underline{x}, \Omega_C)P(\Omega_C|\underline{x}) + P(\omega_j|\underline{Q}, \Omega_O)P(\Omega_O|\underline{x}) + P(\omega_j|\underline{S}, \Omega_S)P(\Omega_S|\underline{x}), \end{aligned}$$

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 $P(\omega_j|\underline{Q}, \Omega_O)$ and $P(\omega_j|\underline{S}, \Omega_S)$ are used for the two sets of exclusive classes.

4.3.6.2 Markov Random Fields

Markov random fields (MRFs) can include contextual information in the form of class interactions between neighbouring pixels. An MRF is determined by an energy function, composed of two main terms: one characterizing class likelihood at the pixel level (depending on per-class scores obtained by the optical and SAR classification methods), and another promoting label smoothness in a local neighbourhood. This means that the model encourages two neighbouring pixels to be labelled with the same class and that the fusion processing stage allows incorporating spatial-contextual information as well.

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 consider y_i as a 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 [39]. In this case, ∂i was chosen to be made of the four pixels adjacent to the i -th pixel (four-connected).

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

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

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	99	

where $X = \{x_i\}_{i \in I}$ is the random field of all optical and SAR observations ($x_i = [O_i, S_i]$ on each pixel $i \in I$), $P_{\mathcal{F}}(y_i|x_i)$ is the pixelwise fusion output described in the previous section, and the spatial energy contribution has been modelled using the Potts model [39].

The local contextual pixelwise probability $P(y_i|x_i, \{y_j\}_{j \in \partial i})$, i.e., the distribution of the class label of each pixel, conditioned to its observations from all sources and to the labels of the neighbouring pixels, can be derived from the energy and provides a spatial-contextual measure of uncertainty of the predicted land cover. While LOGP directly conveys uncertainty in the form of $P_{\mathcal{F}}(\omega_j|x)$, the MRF formulation is based on an energy function. Therefore, the uncertainty in MRF should be estimated according to its energy. In order to do that, a softmax is applied:

$$\tilde{P}_{MRF}(y_i|x_i, \{y_j\}_{j \in \partial i}) = \frac{\exp[-U(y_i|x_i, \{y_j\}_{j \in \partial i})]}{\sum_{\omega_k \in \Omega} \exp[-U(\omega_k|x_i, \{y_j\}_{j \in \partial i})]}$$

However, the MRF energy depends on the weight γ that tune the tradeoff among the various contributions to the energy function. These parameters also have an impact on the resulting uncertainty – an undesired behaviour given the different meaning of these parameters. In order to output an uncertainty distribution that both complies with the posteriors coming from LOGP and reflects the spatial structure of the MRF output, the uncertainty estimation is also parametrized with a further parameter μ so that the overall distribution remains similar, although with enhanced spatial regularity:

$$\tilde{P}_{MRF}(y_i|x_i, \{y_j\}_{j \in \partial i}) = \frac{\exp[-\mu U(y_i|x_i, \{y_j\}_{j \in \partial i})]}{\sum_{\omega_k \in \Omega} \exp[-\mu U(\omega_k|x_i, \{y_j\}_{j \in \partial i})]}$$

The value for μ is determined by comparing the distributions (histograms) of the LOGP and MRF uncertainty.

In the application of MRF-based methods to decision fusion, special focus has been devoted to the minimization of the energy function U with respect to the random field Y of the class labels. For that reason, the iterated conditional mode (ICM) algorithm has been applied, since it represents an efficient tradeoff between accuracy and computational burden [40]. Moreover, a lot of effort was spent to make the implementation of ICM as fast as possible. This was achieved by making use of convolution-like operations only, which have a very fast implementation in the Python environment. That way it was possible to speedup the computations, at a cost of a slight increase in memory occupation, and make MRF feasible to be applied to large scale dataset as the one of HRLC.

4.3.7 First experimental analysis on the Amazon round robin area

The first phase of the experimental analysis of the developed decision fusion methods was initially pursued by using the posterior probabilities estimated by the optical and SAR chains applied to the preliminary draft class legend including 8 classes, corresponding to numbers 1, 3, 5, 6, 7, 8, 10, 15 of the table illustrated in Figure 96.

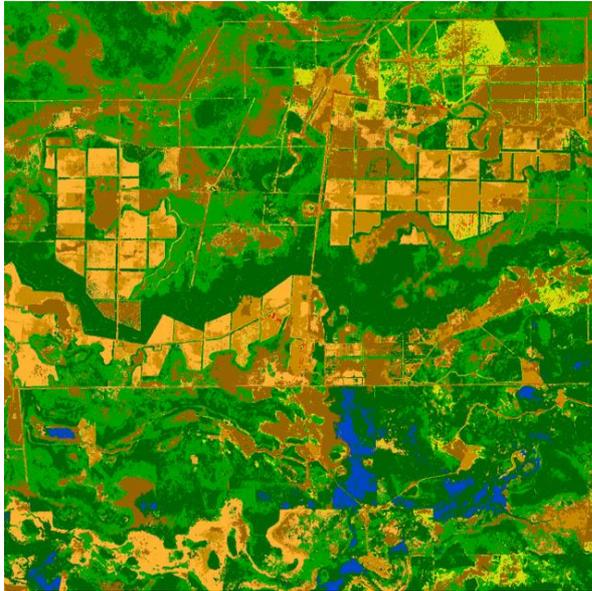
Value	Label	Color
0	No data	Black
1	Evergreen broadleaf tree	Dark Green
2	Evergreen needleleaf tree	Dark Green
3	Deciduous broadleaf tree	Bright Green
4	Deciduous needleleaf tree	Dark Green
5	Shrubland	Brown
6	Permanent cropland	Light Brown
7	Annual summer cropland	Yellow-Green
8	Grassland	Orange
9	Lichens and mosses	Pink
10	Permanent water bodies	Blue
11	Permanent snow and ice	White
12	Beaches dunes and sands	Tan
13	Bare soils	Yellow
14	Bare rock	Grey
15	Built-up areas	Red

Figure 96. Draft of the class legend used during first year. Note that this class legends has been updated and does not correspond with the final one.

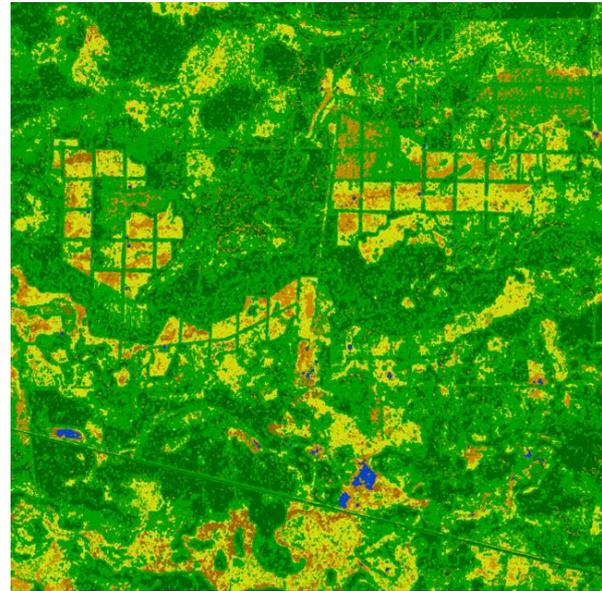
In that preliminary phase, the focus was initially concentrated on the Amazon round robin area, i.e., the S2 tile 21kuq and the S1 images that spatially overlapped with it. Then, similar results, omitted for brevity, were obtained in this preliminary experimental analysis on the African and Siberian round-round areas as well.

The set of classes in common to the optical and SAR chains is $\Omega_C = \{1, 3, 5, 6, 7, 8, 10\}$. There were one optical-exclusive class, i.e., no. 15 ($\Omega_O = \{15\}$) and no SAR-exclusive classes ($\Omega_S = \emptyset$). The next sections focus on the consensus (non-contextual) and Markovian (contextual) approaches. The class-specific combination is explained together with the former.

Figure 97 collects a detail of the classification maps obtained by the optical and SAR processing chains (the color legend is reported at the beginning of the current section):



Optical classification map (detail)



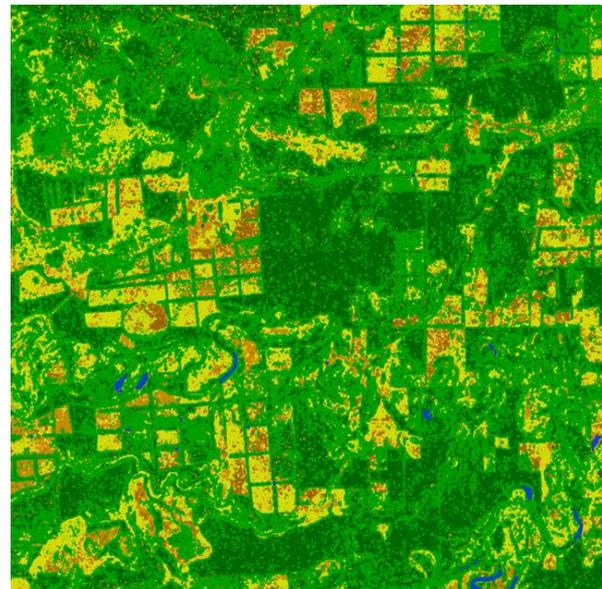
SAR classification map (detail)

Figure 97. Detail of the optical (left) and SAR (right) classification maps.

As expected, the two results differed in spatial regularity (smoother in the optical-based case than in the SAR-based case, because of the influence of residual speckle), class legend (see above) and in the labels assigned by the classification of the two sources to several areas. While optical data were expected to be fundamental in discriminating the considered land cover classes, Figure 98 and Figure 99 provide an example of a case in which residual impact of cloud cover may have affected the optical-based map but obviously not the SAR-based map.



Optical classification map (detail of cloudy area)



SAR classification map (detail of cloudy area)

Figure 98. Detail of the optical (left) and SAR (right) classification maps, with a residual cloudy area.

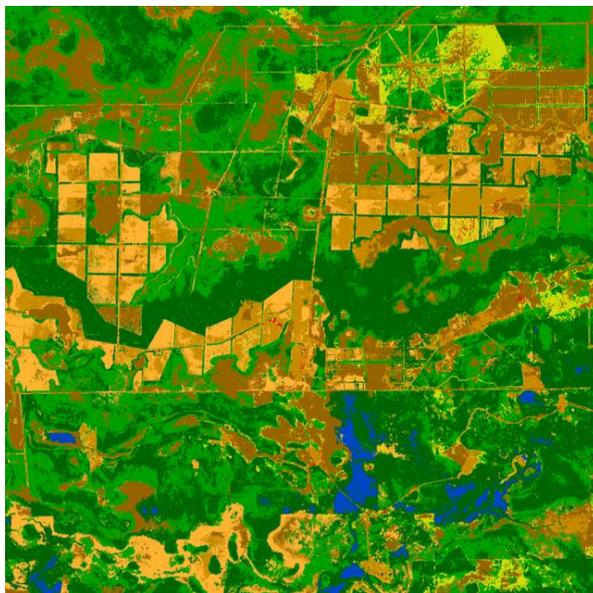


Optical data (detail of cloudy area)

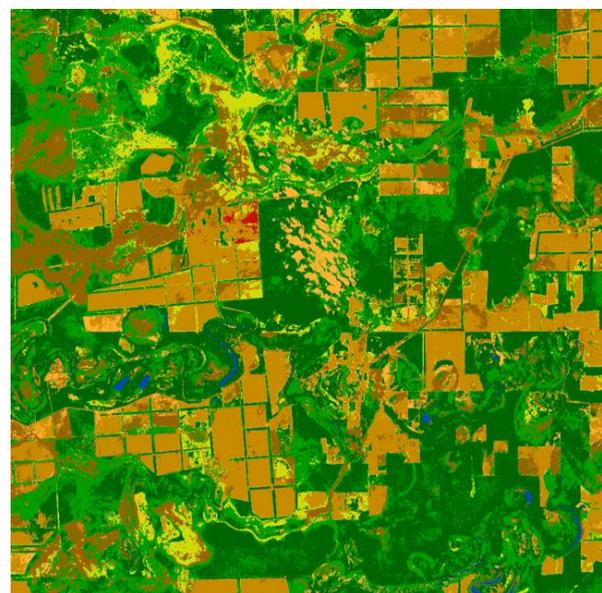
Figure 99. Detail of the optical image, with the residual cloudy area.

4.3.7.1 Results - Consensus Theory

The results of applying linear opinion pool to the posterior probabilities associated with the aforementioned maps is shown in Figure 100.



LOP classification map (detail)



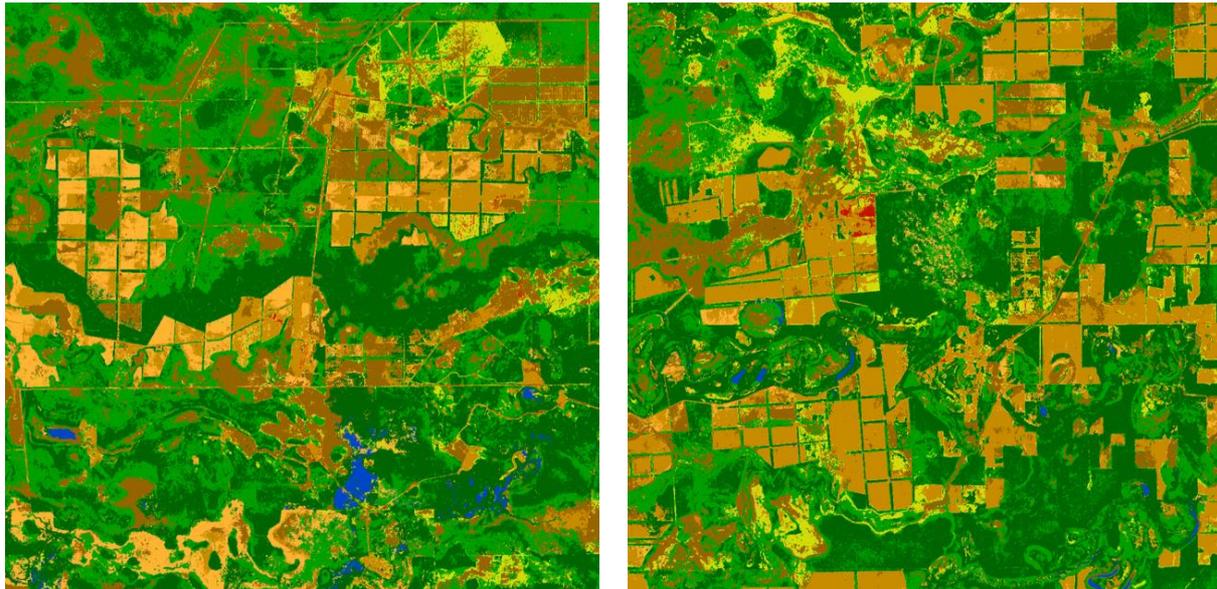
LOP classification map (detail of cloudy area)

Figure 100. Details of the LOP classification maps, with the residual cloudy area on the right.

In particular, LOP was applied by giving slightly larger weight to the optical source than to the SAR source ($\alpha = 0.6$ and $\beta = 0.4$, uniformly on all classes in Ω_C), consistently with the expected reliability of the corresponding land cover classification output. Indeed, linear opinion pool made minor changes as compared to the optical-based classification map. It was however the fastest method among those described above, requiring around 30 seconds on a 8200x8200 tile on a standard desktop machine (no GPU). In particular, it did not compensate for

the impact of the aforementioned residual cloud on the classification map. Indeed, the stronger influence of the optical-based result than of the SAR-based result on the fusion map was consistent with both the expected contribution of each type of sensor to class discrimination and with the different behaviours of the corresponding posteriors. The posterior probabilities obtained from the optical source on a given pixel most often indicated the most probable class membership with higher confidence as compared to the posteriors obtained from the SAR source. Accordingly, they more strongly affected the pixelwise decision fusion outcome.

The results of logarithmic opinion pool are shown in Figure 101.



LOGP classification map (detail)

LOGP classification map (detail of cloudy area)

Figure 101. Details of the LOGP classification maps, with the residual cloudy area on the right.

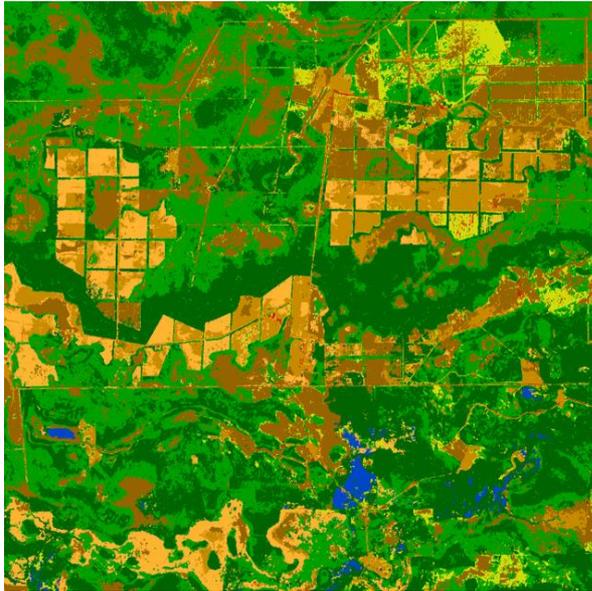
In the case of the logarithmic opinion pool, while the overall classification output was still significantly influenced by the optical source, it is worth noting that the erroneously labelled residual cloud was mostly corrected thanks to the SAR source, thus confirming the relevance of fusing the two separate classification results. LOGP required around 50 seconds to run on the 8200x8200 tile. Hyperparameter setting was addressed for LOGP analogously to the case of LOP.

In both cases of LOP and LOGP, the spatial regularity of the output classification map was similar to that of the optical-based result. On one hand, this suggests that the two fusion strategies are not negatively affected by the impact of residual speckle on the SAR-based result. On the other hand, further smoothness is obviously not achieved by LOP or LOGP, because they are fully non-contextual approaches.

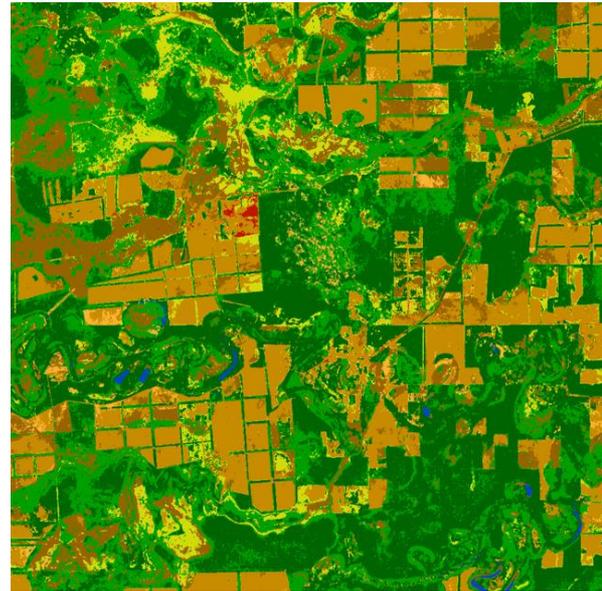
4.3.7.2 Results – Markov Random Fields

While the previous methods make no use of contextual information, Markov random fields model this type of information explicitly. Here, consistently with the results discussed in the previous subsection, the adopted MRF used a unary energy term derived according to LOGP, while the contextual term depending on a 4-connected neighbourhood was given by the Potts model. A unary based on LOP is not presented for brevity as it could benefit less than LOGP from the fusion of the two sources, as argued in the previous subsection.

The classification map obtained by applying ICM to this Markovian energy is as in Figure 102.



MRF classification map (detail)



MRF classification map (detail of cloudy area)

Figure 102. Details of the MRF classification maps, with the residual cloudy area on the right.

MRF required around 7 minutes to run on the 8200x8200 tile and applying fusion on 7 classes. It is possible to appreciate the spatial regularization achieved with MRF as compared to the previous classification map. On one hand, the comments made in the previous section with regard to LOGP and its relation to the optical-based and SAR-based individual outputs hold in this case as well. In particular, the impact of the residual cloud is significantly mitigated in the MRF result, too. On the other hand, the Markovian output is overall smoother than the aforementioned pixelwise results. In particular, these maps were obtained by setting the spatial parameter γ to 0.8. Stronger or less spatial regularization can as well be achieved increasing or decreasing the value of γ . That way, the tradeoff between spatial smoothness and small-scale spatial details can be controlled explicitly. On one hand, this tradeoff can be tuned by applying automatic hyperparameter optimization methods that determine an optimal value for γ through mean-square-error or Bayesian approaches [5], [6]. Moreover, alternate spatial models, including contrast sensitive conditional random fields, mitigate the possible degradation in small-scale details. On the other hand, the opportunity to explicitly tune the tradeoff between spatial smoothness and detail provides further flexibility in the generation of an output HRLC product that meets the requirements of the climate community.

4.3.8 Experimental analysis on all round robin areas

The experimental phase leading to the first production was focused on all round robin areas with regard to the final formulation of the class legend defined within the project in accordance with the Climate Group and the Validation Team. Input posteriors associated with the first level of this final class legend and coming from the optical and SAR classifiers were used.

The final class legend is reported in Figure 103.

HRLC CLASSES											
CODE	DN	1 st LEVEL	CODE	DN	2 nd LEVEL	CODE	DN	3 rd LEVEL	CODE	DN	4 th LEVEL
--	0	No data									
10	1	Tree cover evergreen broadleaf									
20	2	Tree cover evergreen needleleaf									
30	3	Tree cover deciduous broadleaf									
40	4	Tree cover deciduous needleleaf									
50	5	Shrub cover evergreen	51	17	Broadleaf						
			52	18	Needleleaf						
60	6	Shrub cover deciduous	61	19	Broadleaf						
			62	20	Needleleaf						
70	7	Grasslands	71	21	Natural						
			72	22	Managed						
80	8	Croplands	81	23	Winter	811	24	Rainfed	8121	26	Sparkling
						812	25	Irrigated	8122	27	Flooding
			82	28	Summer	821	29	Rainfed	8221	31	Sparkling
						822	30	Irrigated	8222	32	Flooding
			83	33	Multicropping	831	34	Rainfed	8321	36	Sparkling
						832	35	Irrigated	8322	37	Flooding
90	9	Woody vegetation aquatic or regularly flooded									
100	10	Grassland vegetation aquatic or regularly flooded									
110	11	Lichens and Mosses									
120	12	Bare areas	121	38	Unconsolidated	1211	39	Sands			
			122	41	Consolidated	1212	40	Bare soils			
130	13	Built-up	131	42	Buildings						
			132	43	Artificial Roads						
140	14	Open Water seasonal									
150	15	Open Water permanent									
160	16	Permanent snow and/or ice	161	44	Snow						
			162	45	Ice						

Figure 103. Final class legend

The set of classes with associated available posteriors was different from tile to tile, in accordance to the actual covers that could be present within each region.

For what regards the posterior probability map, the values are represented as grey scale with black corresponding to 0% probability and white to 100%, according to the pictured colorbar:



In the following, we shall focus first on the African round robin tile 37PCP, and then, we will move to the other round robin areas. For the sake of brevity, the results will be discussed for LOGP only, comparing them with optical and SAR classification maps. In fact, similar comments hold also in the case of MRF about the overall structure of the map and the meaning of the pixelwise posteriors. Then, a comparison between these results and those of the MRF will be shown on a smaller crop of tile 37PCP to focus on the impact of MRFs on spatial properties. MRF required around 20 to 30 minutes to run on the 10980x10980 tiles and applying fusion on a varying number of classes, ranging from 10 to 14.

4.3.8.1 African tile T37PCP

4.3.8.1.1 LOGP

The next figures show the results obtained by LOGP on the African tile 37PCP, in particular the fused classification map and the corresponding uncertainty products.

The top panel consist in optical classification map (left) and Sar classification map (right). The middle panel consists in the final fused classification map (left) and the posterior probability corresponding to the most probable class (right). On the bottom, the classification map and the posterior probability corresponding to the second best class are shown in Figure 104.

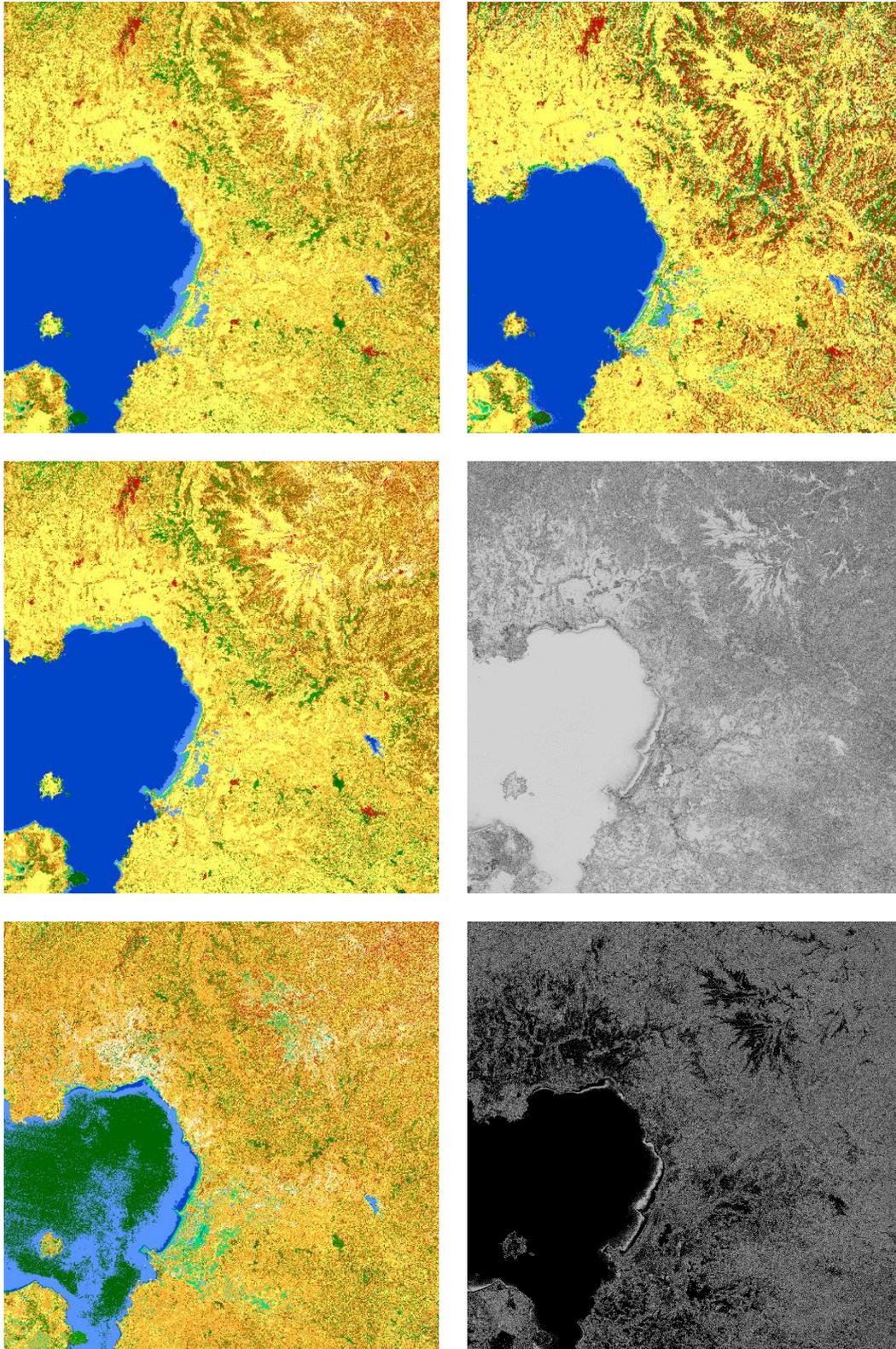


Figure 104. Results of LOGP on RR granule T37PCP, together with optical and SAR classification maps.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	107	

Looking at the top row of the figure, it is clear that the SAR map (top right panel) over estimates the presence of urban class. However, the fusion map (middle left panel) correctly enforces the correct classification obtained by optical map (top left panel).

Then, for example, the area of the lake allows for a clear interpretation of the results. The first class, in this case permanent water, is assigned with high confidence, for that reason the corresponding area in the top right panel is very bright. On the contrary, the area corresponding to the lake on the bottom right panel is very dark since the second most probable class (in this case seasonal water or trees) is unlikely. This is an example of how the posterior probability maps that come together with the classification maps emphasize the confidence / uncertainty associated with the labelling of each pixel. In turn, this uncertainty measures comes from the fusion of the uncertainty encoded in probabilistic terms in the outputs of the optical and SAR classification chains.

It is also interesting to note how the map of the second-best labels exhibits a significantly more noisy spatial behaviour than the map of the best labels and more poorly captures the spatial features in the scene. This is consistent with the meaning of these two labels in relation to the target land cover.

4.3.8.1.2 MRF

Figure 105 shows the results obtained using MRF on the same tile, Africa 37PCP. On top, the final fused classification map (left) and the posterior probability corresponding to the most probable class (right). On the bottom, the classification map and the posterior probability corresponding to the second best class are shown:

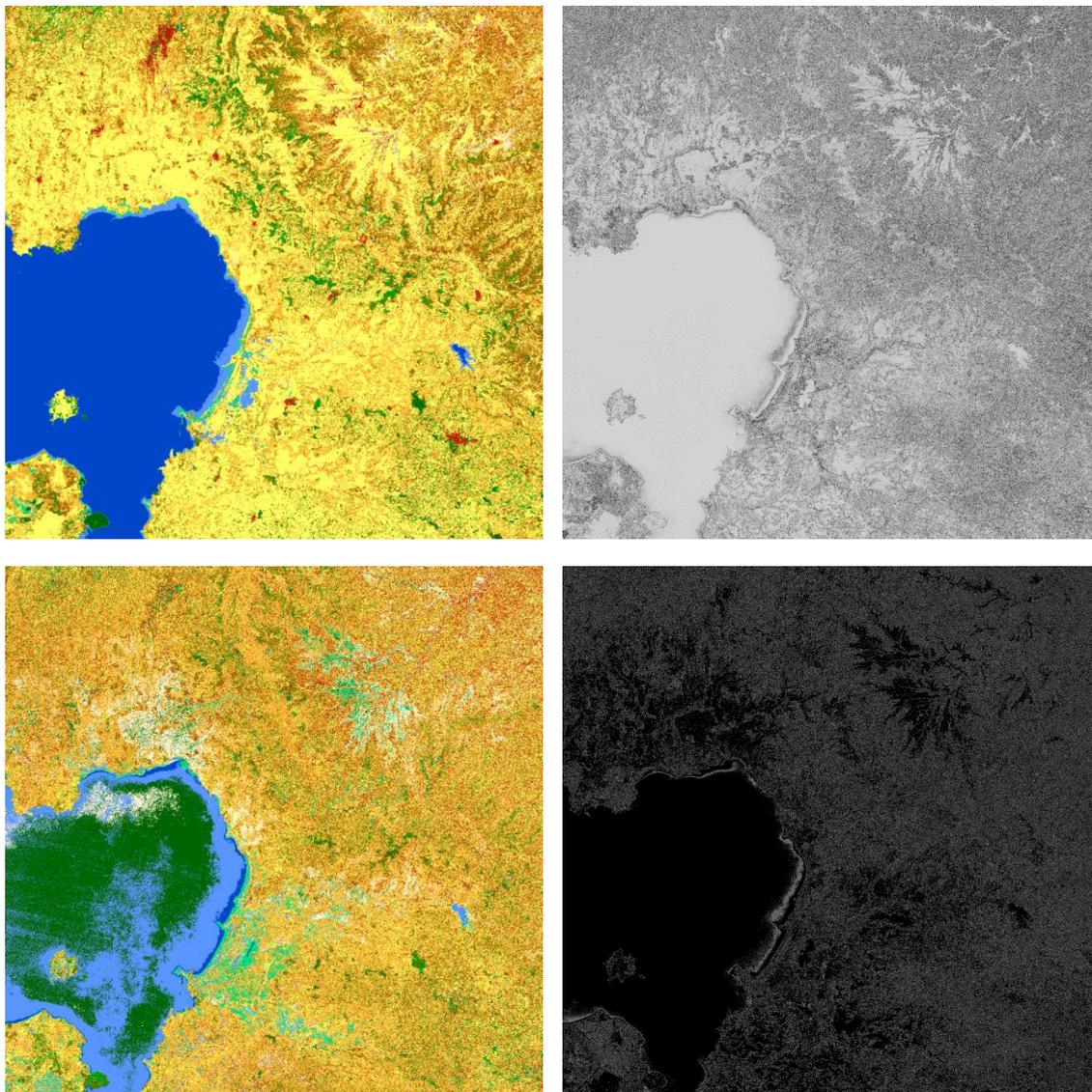


Figure 105. Results of MRF on RR granule T37PCP.

Similar comments hold also in this case about the overall structure of the map and the meaning of the pixelwise posteriors. However, a more effective spatial regularization can be noted in the MRF result rather than in the LOGP output: more details on this aspect can be found later in Section 0. We also note how the uncertainty measures represented by the pixelwise posteriors are overall consistent with those of the LOGP result, yet with the spatial structure determined by the MRF model. This confirms the effectiveness of the proposed dedicated algorithm that has been developed for quantifying the pixelwise uncertainty associated with the MRF-based classification product.

4.3.8.2 Amazon tile T21KUQ

4.3.8.2.1 LOGP

Figure 106 shows the results obtained by MRF for the Amazonian tile T21KUQ. The structure of the panels is the same of the previous LOGP section:

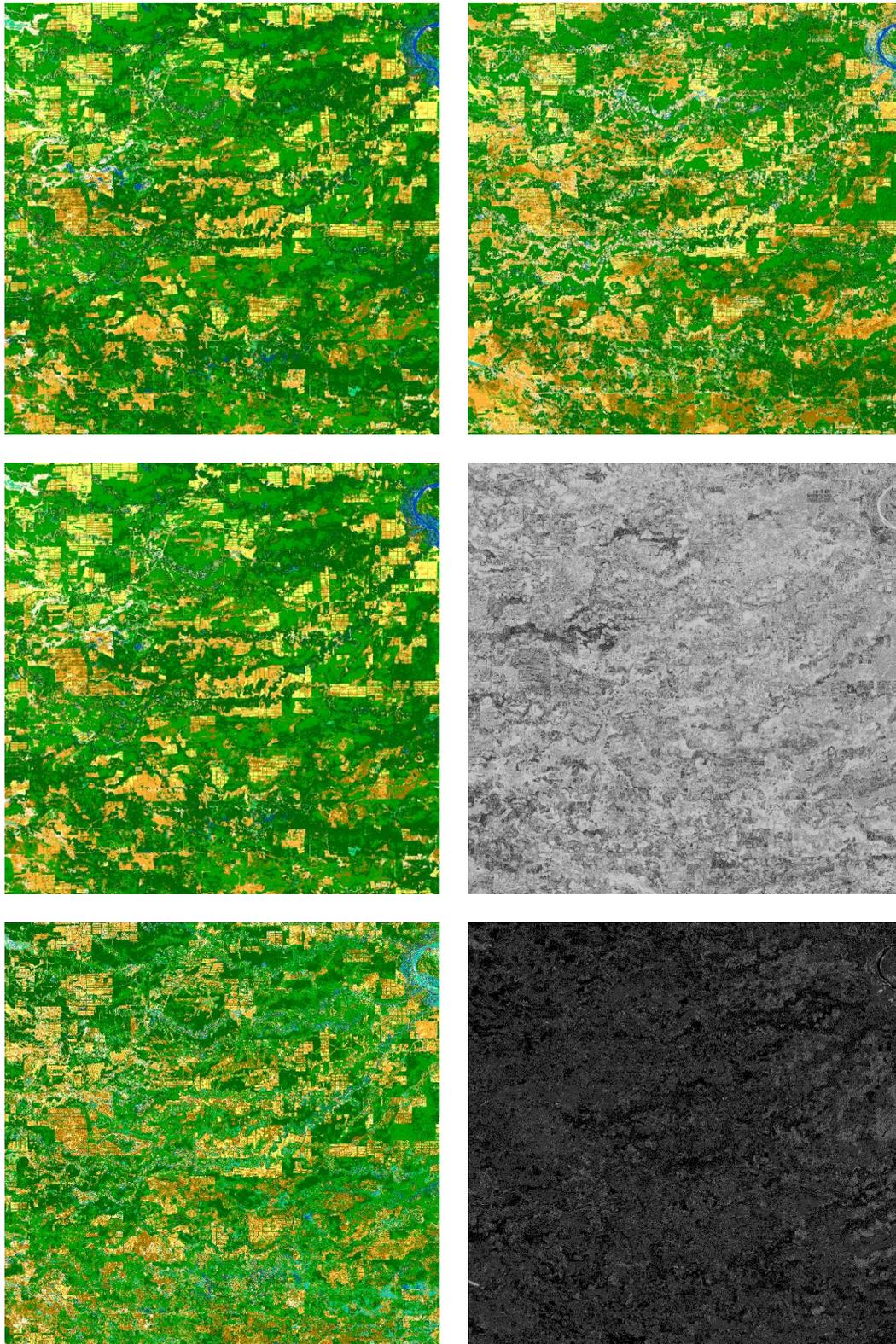


Figure 106. Results of LOGP on RR granule T21KUQ, together with optical and SAR classification maps.

The top panel consist in optical classification map (left) and Sar classification map (right). The middle panel consists in the final fused classification map (left) and the posterior probability corresponding to the most probable class (right). On the bottom, the classification map and the posterior probability corresponding to the second best class are shown.

SAR classifier over estimates the presence of bare areas but the fusion step is able to correctly address for this. On the other hand, some missing watery areas in the SAR map had an impact on the final fused map.

Looking at the top right panel it is possible to see some areas which are classified with lower confidence. In fact, on the left part of the posterior probability map some darker areas are present, some of which are assigned to bare areas. Looking then at the posterior probabilities for the second best class, it can be noticed that the same points corresponds to brighter areas, meaning that the posterior probabilities for the first and second best class are similar – an intrinsic indication of uncertainty in the labeling of those pixels.

4.3.8.2.2 MRF

Figure 107 shows the results obtained by MRF for the Amazonian tile T21KUQ (same structure of the panels of MRF as before):

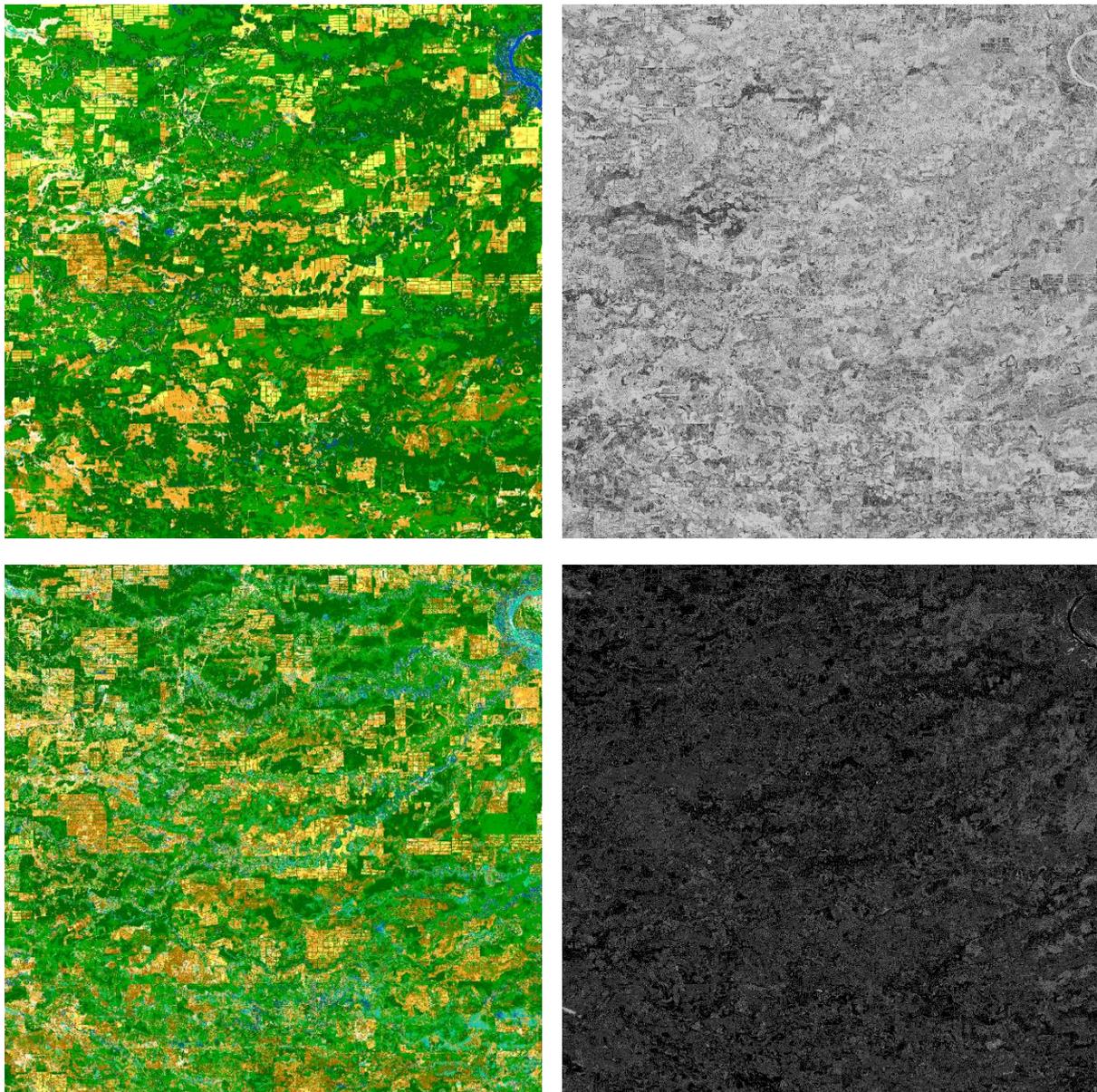


Figure 107. Results of MRF on RR granule T21KUQ.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	111	

The discussion of the results obtained by the MRF-based approach shown above are omitted for brevity, since similar comments as in the previous LOGP section hold as well.

4.3.8.3 Amazonian Tile T21KXT

4.3.8.3.1 LOGP

Figure 108 shows the results obtained by LOGP for the Amazonian tile T21KUQ. The structure of the panels is the same of the previous LOGP section:

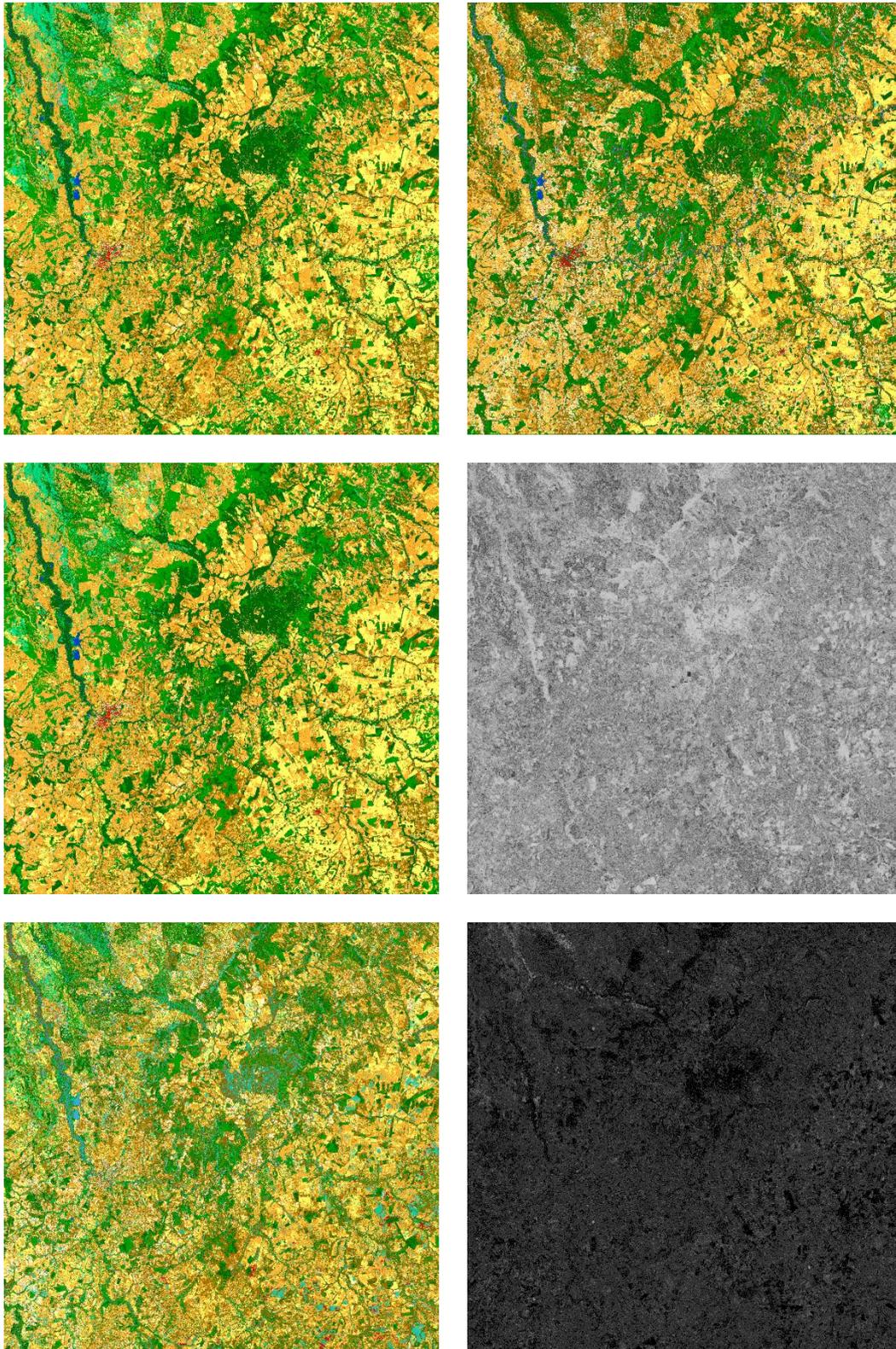


Figure 108. Results of LOGP on RR granule T21KXT, together with optical and SAR classification maps.

In this case, the confidence of the classification, indicated by the pixelwise posterior measures, is quite uniform all over the tile, without major uncertain areas.

4.3.8.3.2 MRF

Figure 109 shows the results obtained by MRF for the Amazonian tile T21KXT:

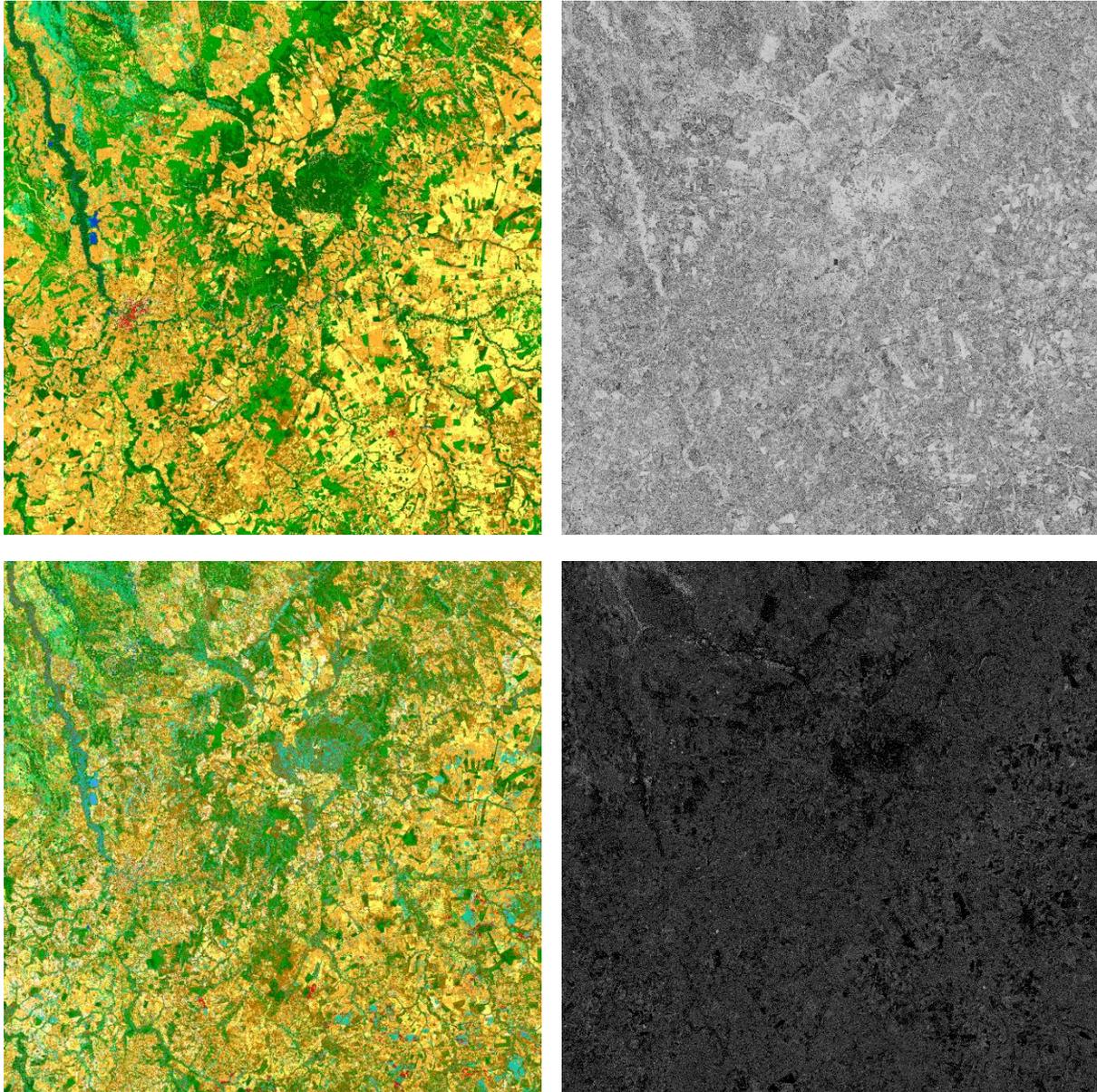


Figure 109. Results of MRF on RR granule T21KXT.

In this case as well, the map obtained by the MRF approach are shown but the discussion is omitted for brevity.

4.3.8.4 *Siberian Tile T42WXS*

4.3.8.4.1 LOGP

The last round robin tile is the Siberian T42WXS, which results obtained from LOGP are shown in Figure 110.

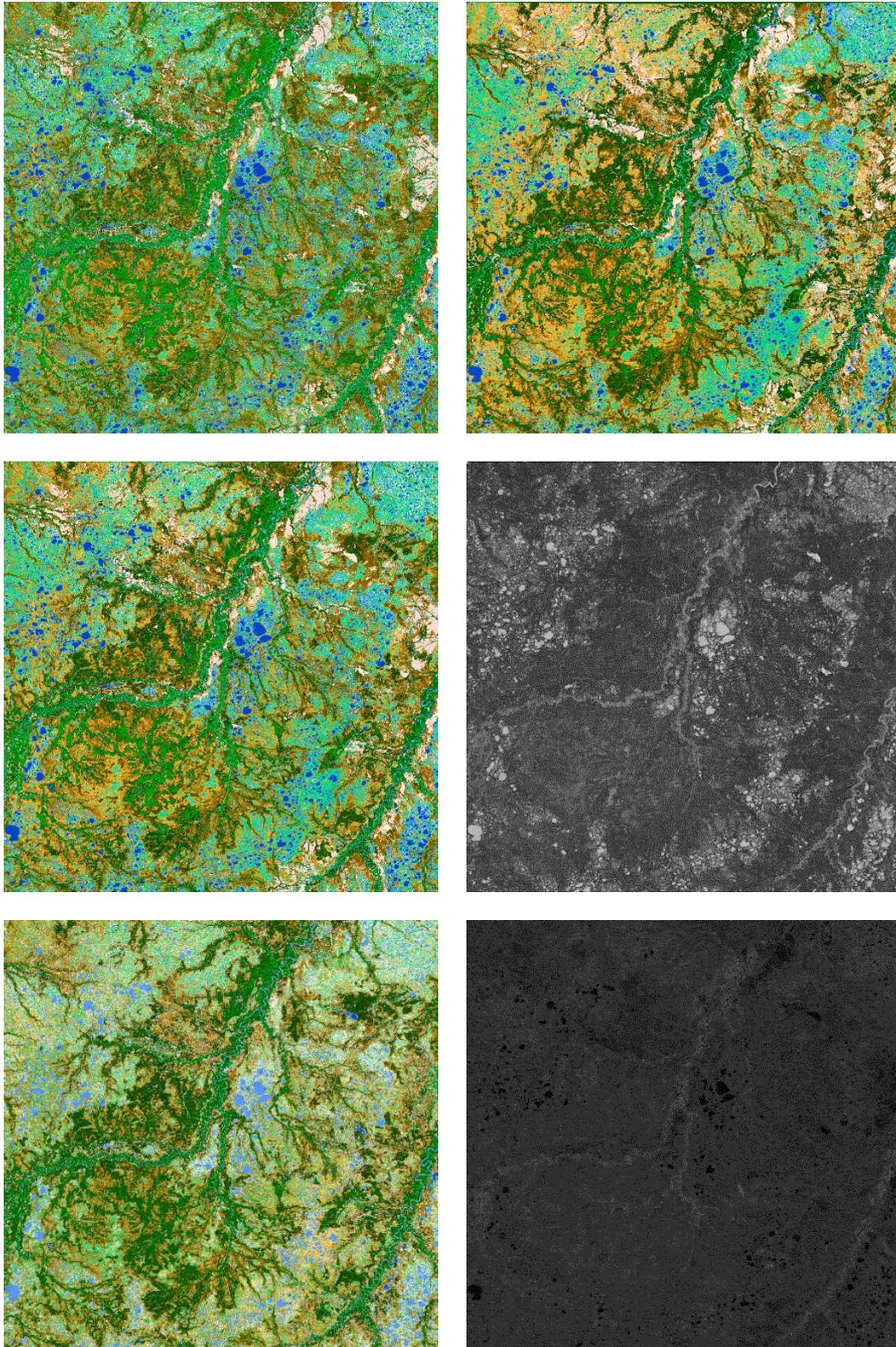


Figure 110. Results of LOGP on RR granule T42WXS, together with optical and SAR classification maps.

The overall confidence of the final classification is lower as compared to the previous tile, with peaks of “brightness” (i.e., higher confidence) in the areas corresponding to small lakes, which are correctly assigned to the permanent water class. This indicates an overall higher uncertainty in the labeling associated with this round robin tile, as compared to those of the previous tiles.

4.3.8.4.2 MRF

Concluding the analysis of the tile, the Figure 111 shows the results for MRF on the Siberian tile T42WXS:

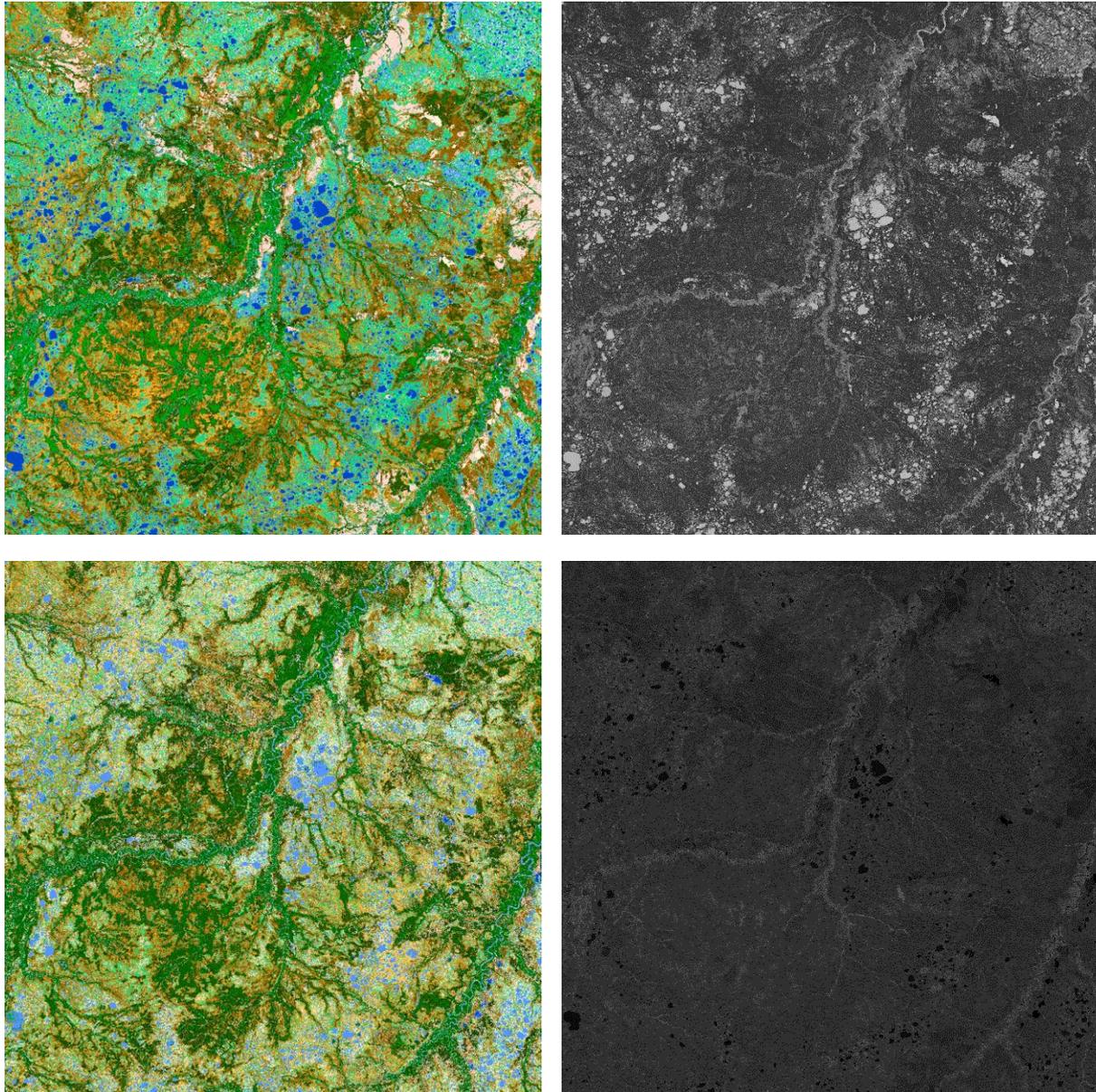


Figure 111. Results of MRF on RR granule T42WXS.

In this case as well, we show the map associated with the MRF output, but omit the discussion for brevity. As anticipated at the beginning of Section 4.3.8, we shall focus in the next subsection precisely on the effect of spatial regularization that can be obtained using the MRF approach and on the capability to tune it to meet the user requirements.

4.3.8.5 Parameterizable spatial regularization through the MRF approach

As mentioned above, this section presents a brief comparison of the effect of MRF as compared to LOGP in terms of spatial structure of the output map. Among the advantages of MRF modelling, we recall the possibility to tune the spatial regularization by mean of appropriate parameters (γ in the proposed formulation). Figure 112 shows a crop taken from the African tile T37PCP – on top, the results of LOGP; on the bottom, the results of applying the MRF-based method with two different regularization levels, heavier on the left and softer on the right:

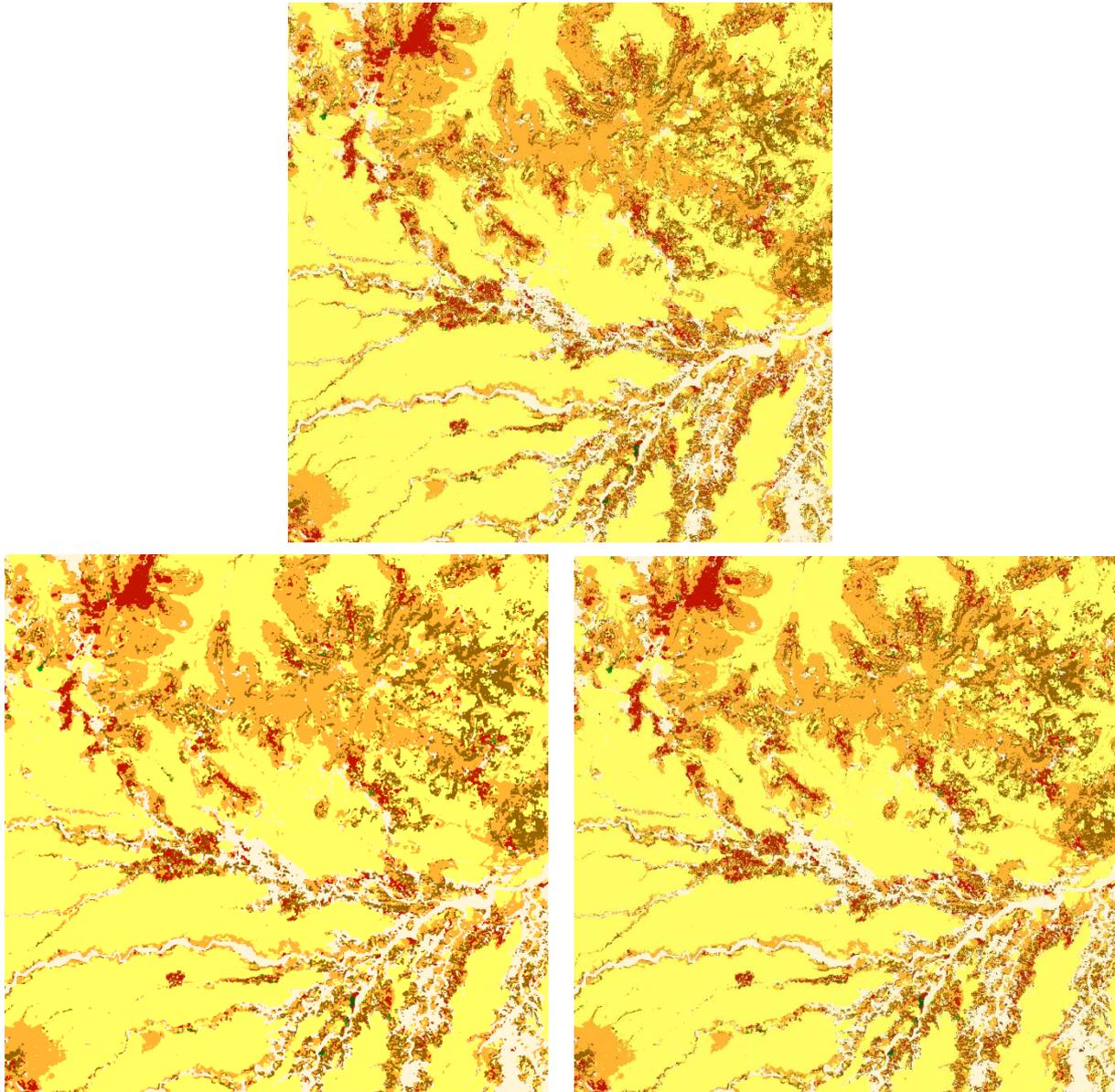


Figure 112. LOGP map (top) and different spatial regularizations achieved using MRF (bottom). On bottom left, the old heavier spatial regularization. On the bottom right, the new parametrization with very light regularization.

The use of the MRF removes errors in the form of salt and pepper classification noise, thus favouring a higher classification accuracy. Simultaneously, MRF-based processing may generally tend to remove small spatial structures. Especially when these structures have an important significance in relation to the HRLC goals, like the dry rivers structures on the bottom left parts of the crops, their preservation can be recommended. For that reason, also taking into account intermediate feedback from the Climate Group and the Validation Team, a low regularization level has been chosen in the first production.

4.3.9 Final decision

Both LOP and LOGP are fast since they are one-shot non-iterative pixelwise methods. However, data fusion performed that way does not benefit from contextual information, which can be important in order to remove salt-and-pepper classification noise and favour the spatial regularity of the output map. Between LOP and LOGP, the experiments also suggested a stronger potential of the latter in taking benefit from both input sources. Markov Random fields are able to also capture local spatial information, thus providing more – and tuneable – spatial regularization at a cost of a comparatively higher, yet still quite short, computational time.

Overall, LOGP and MRF-ICM have been identified as appropriate solutions for HRLC, in combination with the aforementioned class-specific combination strategy. In particular, the default choice in the first production is the MRF approach with low regularization, thanks to its ability to promote spatial regularity up to the desired level.

In this framework, further development may regard the automatic optimization of the hyperparameters of these methods (weights of the various information sources), the integration of edge-preserving / contrast-sensitive spatial terms in the Markovian approach, and the evaluation of the potential of alternate energy minimization algorithms based on graph theory. The last item is possibly promising from the viewpoint of classification accuracy, although at the cost of significantly longer computation time. This time-vs-accuracy tradeoff will be addressed carefully.

5 Algorithms and procedures (year 3)

5.1 Optical data processing

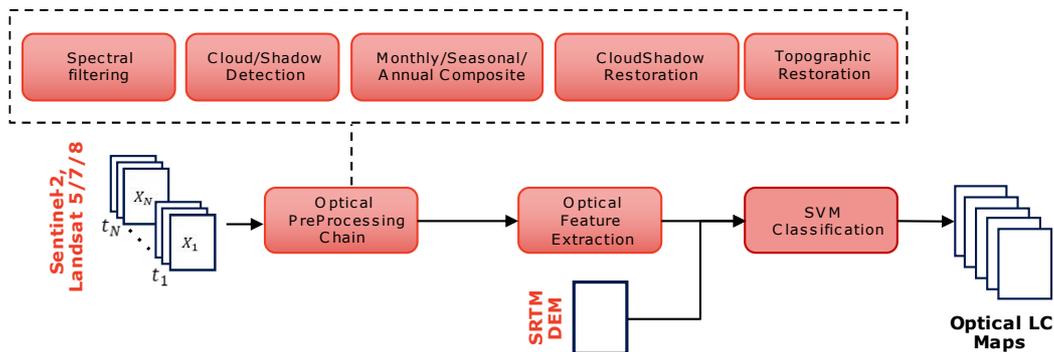


Figure 113. Optical data processing chain for the production of the HR LC map obtained by classifying the time series of Sentinel 2 and Landsat data.

Figure 113 depicts the optical data processing chain for the production of the HR LC maps obtained by classifying the time series of Sentinel 2 L2A data for the static map of 2019 and the time series of Landsat L2 data for the historical maps of 1990, 1995, 2000, 2005, 2010 and 2015. The images are first pre-processed in order to remove spectral outliers and to detect clouds and shadows. Then, monthly, seasonal and annual composites are generated according to the availability of cloud free images. The missing data (caused by cloud and shadow coverage) are restored. Finally, topographic shadows are detected and corrected. Due to the missed availability of training data, the team has integrated the pixel-based photo-interpretation activity with the inclusion of samples from the agreement of available land cover products (static map production only). To provide spatial information to the classifier the altitude and textural features (GLCM-based) have been computed. For the production of the static maps, the ecological region of the pixel is taken into account by splitting the areas into ecoregions. Finally, we performed the ecoregion-based classification by using SVM on the times series of the considered composites.

5.1.1 Satellite images

The considered study area for the static production analysis production is Amazon, specifically the Sentinel-2 tile 21KXT. Figure 114 shows the selected patches used for the analyses.

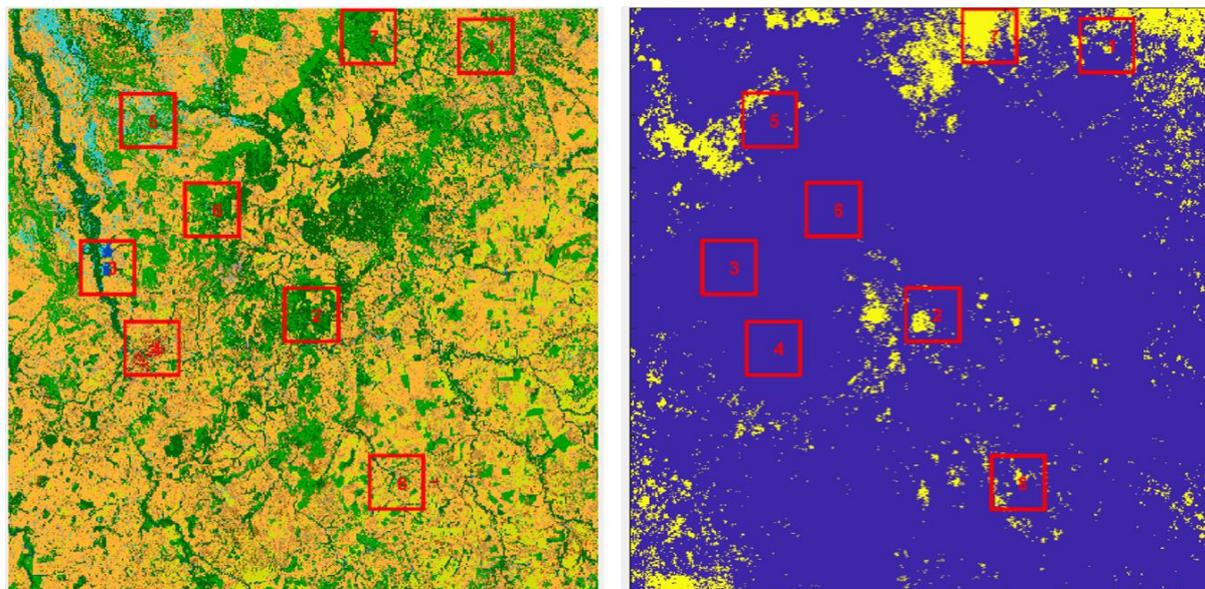


Figure 114. Considered patches for the analysis on the cloud masks improvement and topographic shadow reconstruction. (left) prototype HR LC map for tile 21KXT, (right) cloud mask of an acquisition on this area.

The historical analysis related to the presence of Landsat 7 data started from an analysis of the Landsat data available in the historic regions in the considered areas. Figure 115 shows the number of available images for each area for each Landsat sensor. Africa is the most critical region, as the number of images is low, and Landsat 7 is the only source of images in 2005 and the main source in 2010. Note that in Siberia an annual composite is produced instead of four seasonal composites, thus Siberia has on average a larger number of images (without considering their quality) used in the composite generation, thus the use of Landsat 7 data in Siberia (and Amazonia) has been avoided.

Therefore, the considered study area for the historical production analysis is located in Africa, specifically in the Sentinel-2 tile 37PCP. For the tests on single acquisitions on this area, we considered the following Landsat-7 products cropped to the 37PCP Sentinel-2 tile:

- LE07_L2SP_169051_20050225_20200914_02_T1
- LE07_L2SP_169052_20050225_20200914_02_T1
- LE07_L2SP_170051_20050131_20200914_02_T1
- LE07_L2SP_170052_20050216_20200914_02_T1

Africa historic				Siberia historic				Amazon historic			
Year	#L5	#L7	#L8	Year	#L5	#L7	#L8	Year	#L5	#L7	#L8
1990	1274			1990	851			1990	1183		
1991	1224			1991	628			1991	1191		
1992	1162			1992	414			1992	1086		
1993	1007			1993	496			1993	1464		
1994	1615			1994	496			1994	960		
1995	1642			1995	781			1995	1015		
1996	1540			1996	542			1996	1369		
1997	1733			1997	706			1997	674		
1998	1698			1998	720			1998	933		
1999	1728	240		1999	23	528		1999	1223	892	
2000	1636	490		2000		1346		2000	1325	1648	
2001	1691	699		2001		644		2001	1390	1888	
2002	807	837		2002		492		2002	459	1992	
2003		863		2003		364		2003	858	1376	
2004	3	1248		2004		443		2004	1836	1839	
2005		1342		2005		599		2005	1666	1871	
2006		1158		2006	592	497		2006	1649	1826	
2007		1462		2007	610	543		2007	1421	1772	
2008	236	1322		2008	108	525		2008	1538	1878	
2009	342	1303		2009	812	538		2009	1649	1757	
2010	444	891		2010	340	459		2010	1345	1714	
2011	466	983		2011	508	673		2011	1134	1725	
2012		1432		2012		1117		2012		1910	
2013		1503	1534	2013		964	1471	2013		1628	1418
2014		1837	2188	2014		1116	1530	2014		1801	1929
2015		2036	2304	2015		1177	1764	2015		1929	2062
2016		2054	2257	2016		1251	1997	2016		1871	2012
2017		2078	2259	2017		1114	1568	2017		1856	2040
2018		2066	2220	2018		1238	1664	2018		1846	1978

Corrupted Landsat 7 data
 Corrupted Landsat 7 data combined with reduced Landsat 5 data or non-problematic Landsat 7
 Sufficient Landsat 5, 7 and 8 data.

Figure 115. Number of Landsat images available for each historic region for each Landsat sensor.

5.1.2 Method/algorithm/technique

Within this Section, the final developed methods during the third year in the optical image processing chain are presented and compared.

5.1.2.1 Composite input quality improvement

The goal of this year was to improve the composite quality. In the case of Sentinel-2, the sen2cor cloud and cloud shadow masks have been refined to better filter out noise in the composite generation step (see ATBD). In the case of Landsat data, we defined a strategy to handle the presence of Landsat-7 data with SLC-off. Indeed, the considered composite strategy showed to be affected by it when Landsat-7 is the main or only source of optical data used in the composite. This is particularly relevant for Africa, where Landsat-7 is scarce and is the only source in 2005 and a main source in 2010. For this reason, we adopted a gap-filling strategy (see ATBD) for retrieving the missing information, which is then used in the composite for generating spectrally consistent composites.

5.1.2.2 Topographic shadows correction

A residual problem in the composites (for both Sentinel and Landsat) is related to topographic shadows, which is associated to low reflectance values. The proposed approach is able to identify the topographic shadows and then reconstruct the information contained in them (see ATBD).

5.1.3 Qualitative evaluation

5.1.3.1 Results – Composite input quality improvement in static map production

Figure 116 shows examples of the residual cloud cover in the composites due to missed detection in the sen2cor masks. Figure 117 and Figure 118 show the improvement with the proposed strategy.

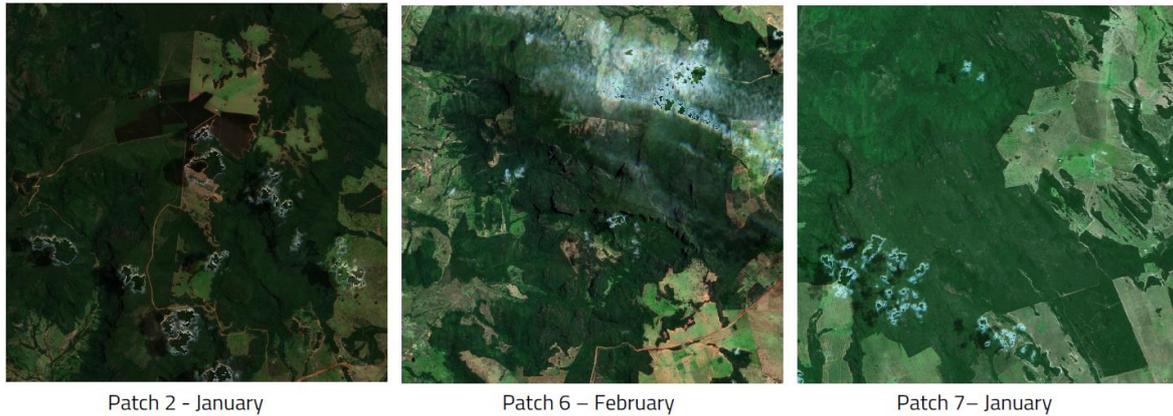


Figure 116. Three composite examples on the considered patches showing the residual cloud cover in the composites.

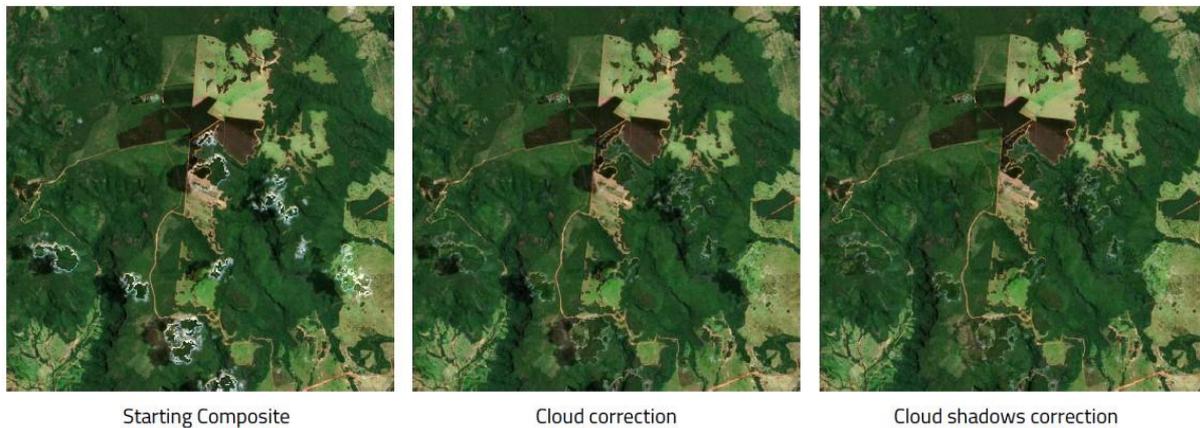


Figure 117. Improvement in the January composite of patch 2 after using the proposed cloud and cloud shadow detection strategy.

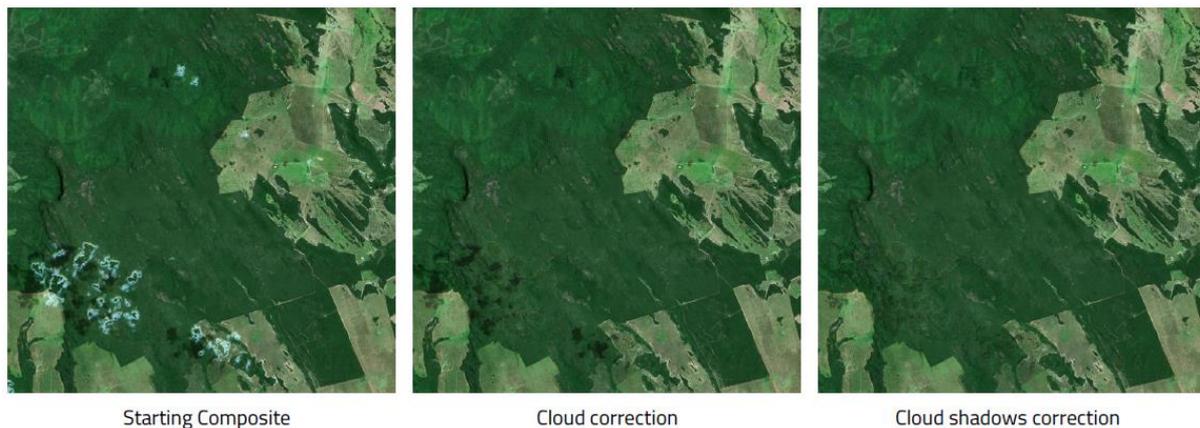


Figure 118. Improvement in the January composite of patch 7 after using the proposed cloud and cloud shadow detection strategy.

5.1.3.2 Results – Composite input quality improvement in historic map production

The experiments were performed on the RR tile 37PCP for year 2005. Figure 119 shows the effect of the gap filling operation on the single acquisition image. We can see in Figure 120 the corresponding effect on the

seasonal composite. Note that the gap filling strategy is not able to reconstruct accurately the spatial information, hence this can result in a loss of detail in the composite. However, we can see in Figure 121 that the classification map is significantly improved. This can be understood in better spectral uniformity with gap filling, as the composite generation with few images can result in discontinuities due to the systematic oscillation of the number of valid acquisitions used for the composite. Indeed, the composite generation strategy works better when enough images are used.

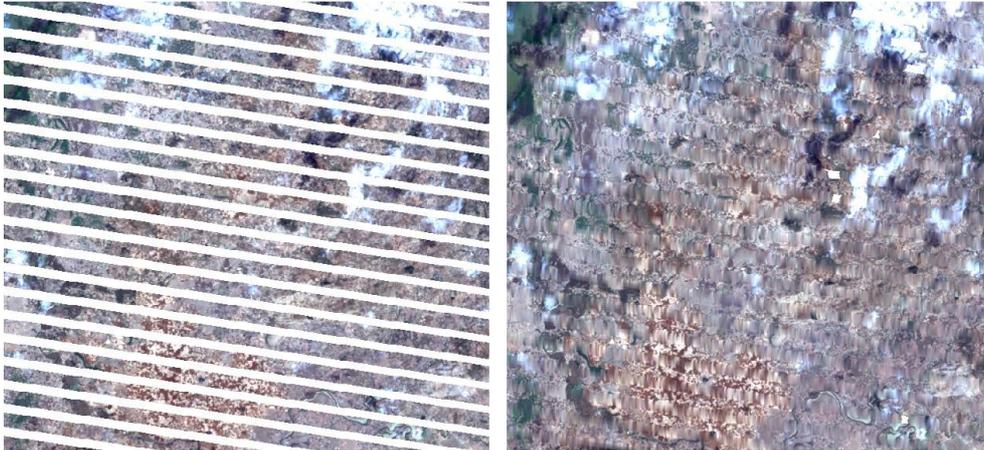


Figure 119. Example of the effect of the gap filling strategy on the single acquisitions: (left) before and (right) after.

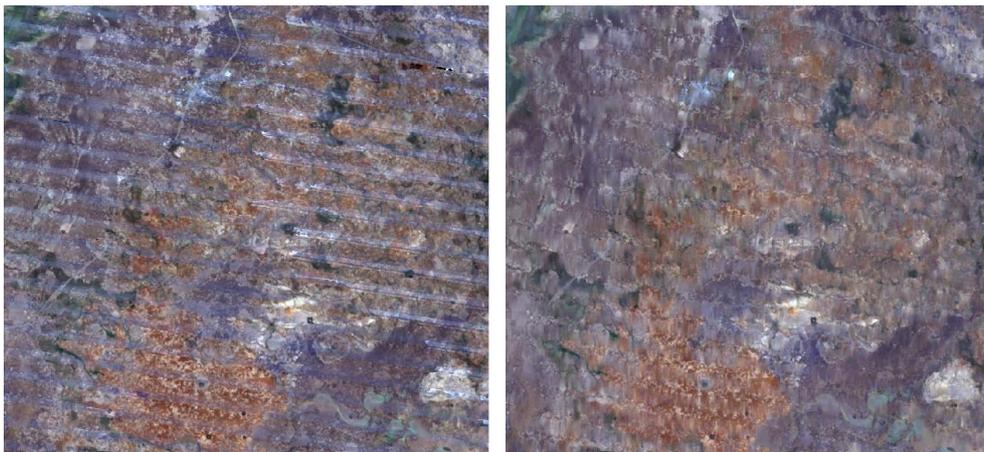


Figure 120. Example of the effect of the gap filling strategy on the composite: (left) before and (right) after.

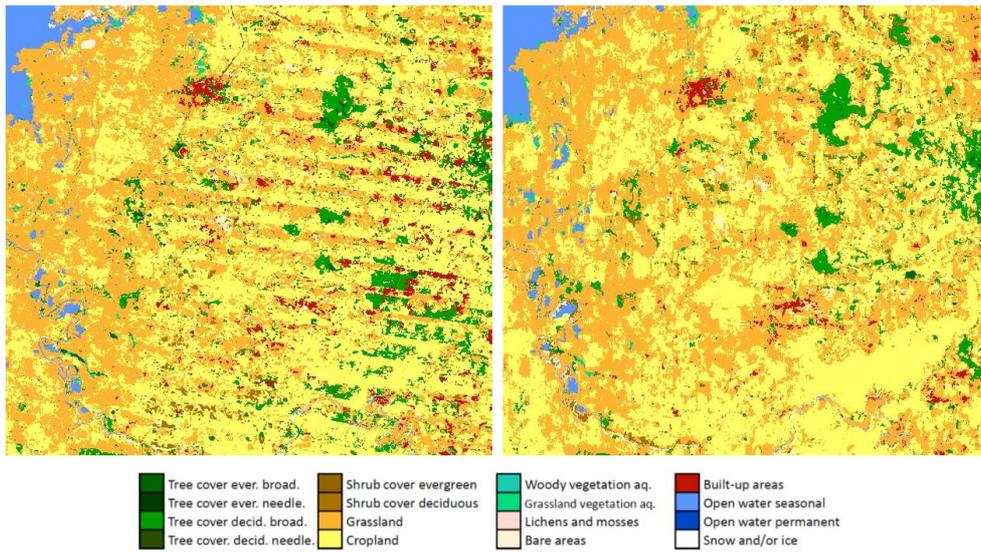


Figure 121. Example of the effect of the gap filling strategy on the classification map: (left) before and (right) after.

5.1.3.3 Results – Topographic shadow reconstruction

Figure 122 shows examples of topographic shadows in the composites. They affect the classification as they recall classes with low reflectance values as seen in Figure 123. Figure 124 shows the result after reconstructing the reflectance values in the shadows. The improvements in the classification maps can be clearly seen in Figure 125.

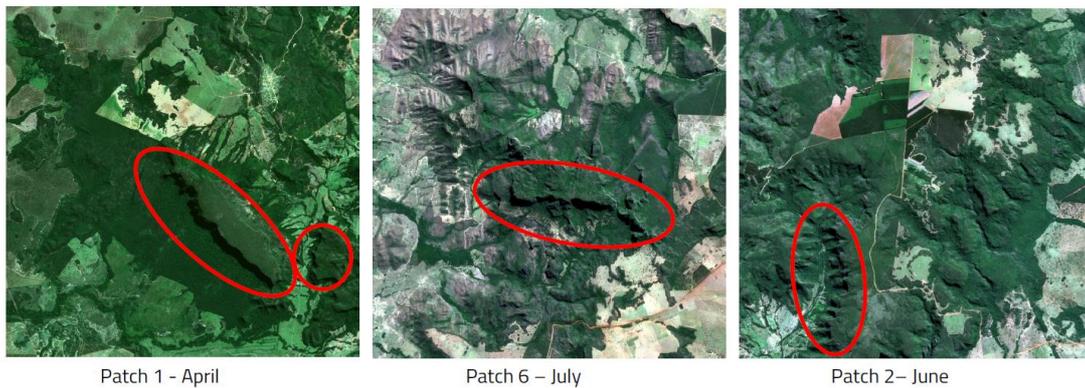


Figure 122. Examples of topographic shadows in the composites.

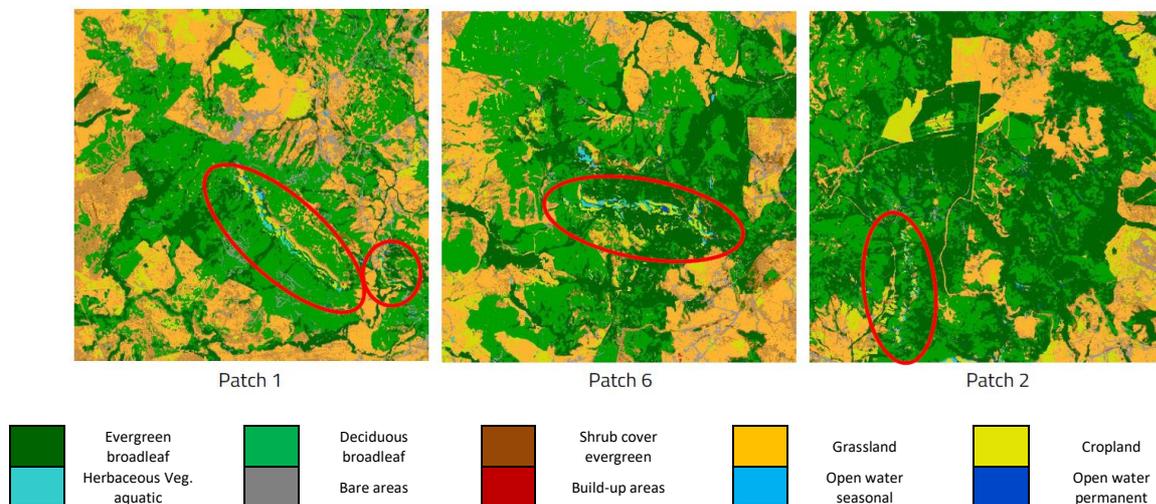


Figure 123. Examples of the effect of the topographic shadows on the prototype classification



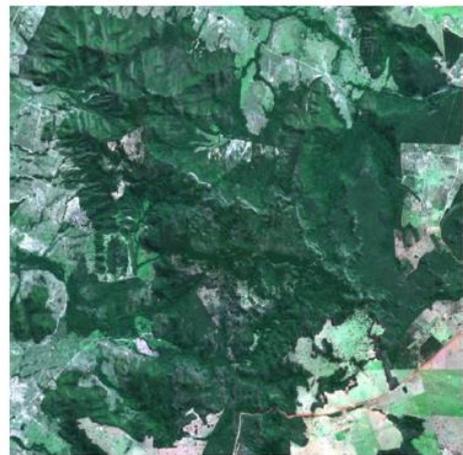
Cloud shadows correction



Topographic shadows correction



Cloud shadows correction



Topographic shadows correction

Figure 124. Examples of the effect on the composite of the topographic shadow reconstruction

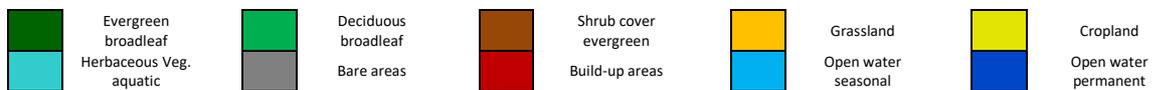
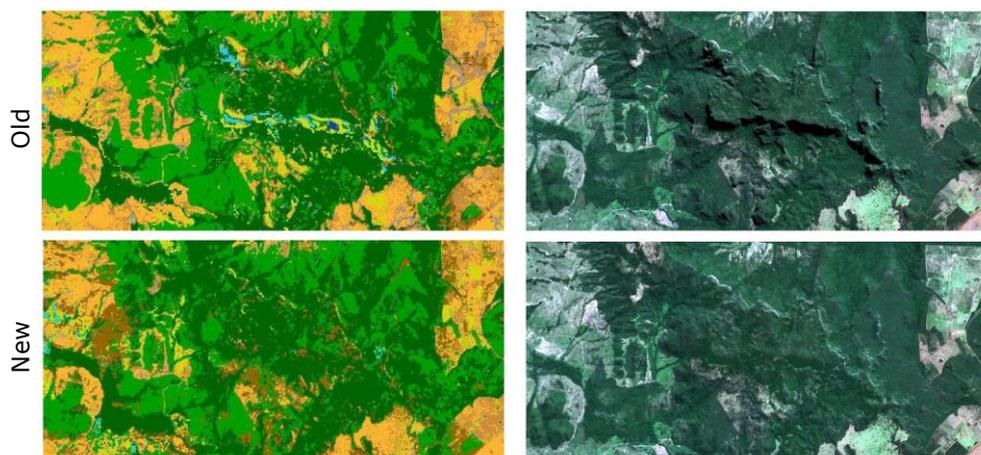


Figure 125. Overall improvement on the prototype classification when using the topographic shadow reconstruction (below) vs. not using it (above).

5.1.4 Final decision

According to the qualitative analysis, the proposed approaches are successful in improving the composite quality and the performance of the classifier: the sen2cor cloud and cloud shadow mask refinements successfully remove residual noise in the composites; the gap filling strategy for Landsat 7 allows the Africa historic composites to be spectrally consistent, resulting in improved classification maps; the topographic shadow reconstruction approach is able to retrieve the information in the shadows, reducing the misclassification of the classifier.

In the case of Landsat 7, residual striping in the maps may be present in the classification maps. However, they are minimized and the multitemporal cascade model showed to be effective in removing the residual striping.

5.2 SAR data processing

The main goal of this year was to improve and release the final version of the processing chain designed for the high resolution and global land cover mapping. The processing chain was tested and used for two specific tasks:

- the extraction of the static map for the year 2019;
- the extraction, whenever possible because of the availability of SAR data in the past, of additional historical land cover maps every five years from 2015 backwards.

For the production of the static map, a time series of Sentinel-1 images was considered as input, while for the second task, data for the ASAR sensors on board of the ENVISAT satellite or by the ERS-1 and 2 satellites were considered, with a much more limited time series and geographical coverage.

The processing chain for the static and historical maps exploiting the SAR data is realized according to the structure highlighted in Figure 126.

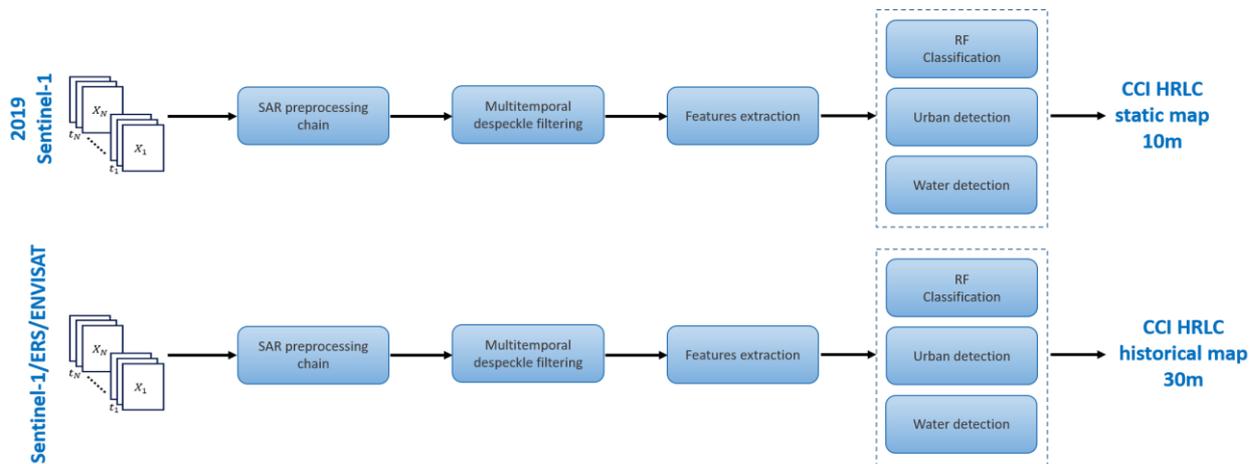


Figure 126. The final version of the SAR processing chain for the production of static and dynamic HRLC products

5.2.1 Satellite images

5.2.2 Method/algorithm/technique

The chain in Figure 126 is composed by the following tasks, which are described in detail in the ATDB document:

- SAR preprocessing, obtained by the standard SNAP chain, to correct radiometrically and geometrically the SAR sequence, and to coregister the data sets not perfectly aligned.
- Multitemporal despeckling, according to the approach described in [22] and applied separately to four temporal segments of the yearly sequence used as input to the chain. The rationale for this choice is to reduce the overall set of data and make the procedure less computationally complex, retaining at the same time the possibility to exploit the temporal trajectory of land cover samples, which is particularly important for vegetation -related classes. As the output of this chain, the original sequence was reduced to four seasonal super-images, extracted as intermediate product of the despeckling method.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	125	

3. SAR feature extraction, which aims at adding spatial features to the already extracted temporal features. In this case, as mentioned in [41], simple statistical features corresponding to the neighborhood of each pixel have been considered.
4. Classification, which is in turn subdivided into three parts:
 - an unsupervised water extraction routine as presented in [42], applying a K-means procedure to discriminate areas with low backscatter and low variance along the year from other potential areas of interest for the water class.
 - An unsupervised urban extent extraction approach based on the extraction of a single super-image for the whole year to which the algorithm described in [26] is applied;
 - A supervised classification implemented by means of a RF classifier, trained by samples manually extracted by the team on the basis of the hierarchical approach (see ATDB).
5. A reconciliation step, aimed at composing the final land cover map by spatially combining the three maps extracted in the previous step.

The novelty introduced this year concerns in the improvement of the classification task, to make the best use of the SAR signal backscattered in discriminating the urban settlements and the water bodies. Two dedicated routines in Python language have been properly developed and tested.

The random forest classifier was trained with the ground truths collected by the team via the hierarchical approach (see ATDB) according to the high resolution land cover classes in Figure 53, excluding training points of the urban and the water classes.

5.2.3 Qualitative evaluation

The experiments were performed on three regions of interest, both for static and historical production, in different portions of the areas in order to check the performance of the approach with respect to very different land cover environments.

The standard implementation of UEXT has been adapted to the two chains (static and historical) as follow:

- The parameters for each of the three static/historic regions (Siberia, Amazonia and Africa) are different, to take into consideration the different environments and the effect of vegetations, level of water and humidity and so on.
- Additionally, the algorithm is applied with two different sets of parameters depending on the 3D structure of the tile, specifically whether or not the tile includes mountainous areas, defined by means of a threshold on the standard deviation of the DEM. The latter method has been implemented for Amazonia and Africa but has not been applied to Siberia, since SAR is not used in Siberia.
- The basic implementation of the water extraction procedure has been improved to deal with desert areas, or areas which have a strong similarity with water because of their dominant backscattering mechanism. Specifically, water areas that are extracted but belong to classes that are desert or bare soil in coarser resolution existing land cover maps (e.g., the Copernicus 100 m spatial resolution Land Cover Map) are masked out. This approach may cause a “staircase effect” on the border of water surfaces that are adjacent to areas of this type (e.g., a lake bordering a desert). To avoid this issue, the masking layers are spatially filtered to reduce this effect (see examples in Figure 127 and Figure 128).

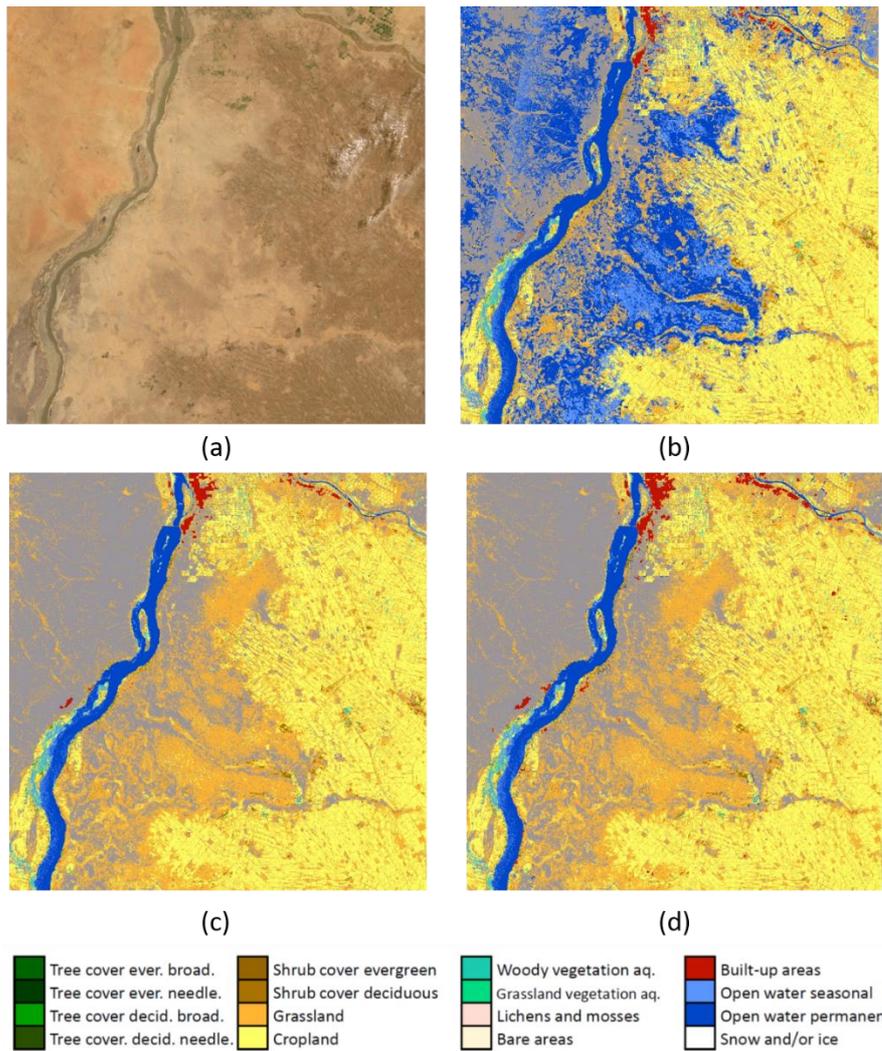


Figure 127. Tile 36PVB: ESRI satellite image (a), water layer without masking out desert/bare soil areas (b), water layer with masking using coarser maps (c) and the final result including all steps described in the text (d).

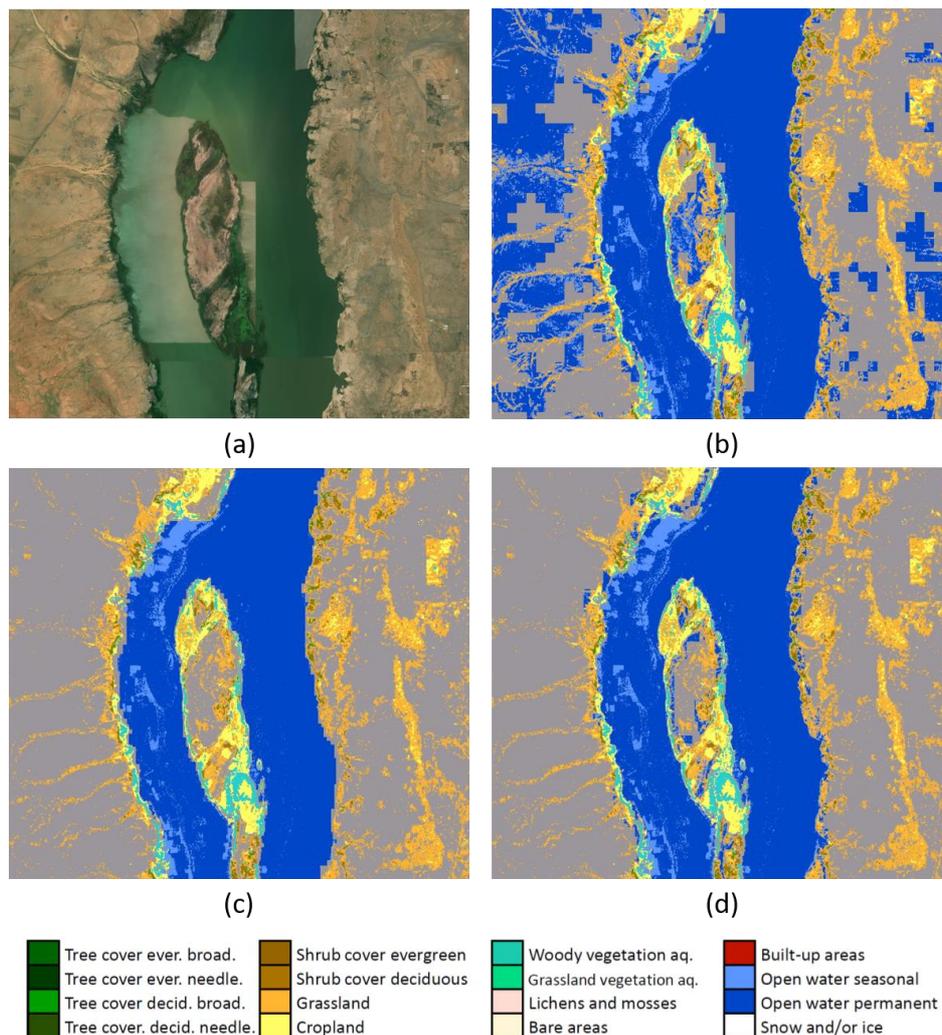
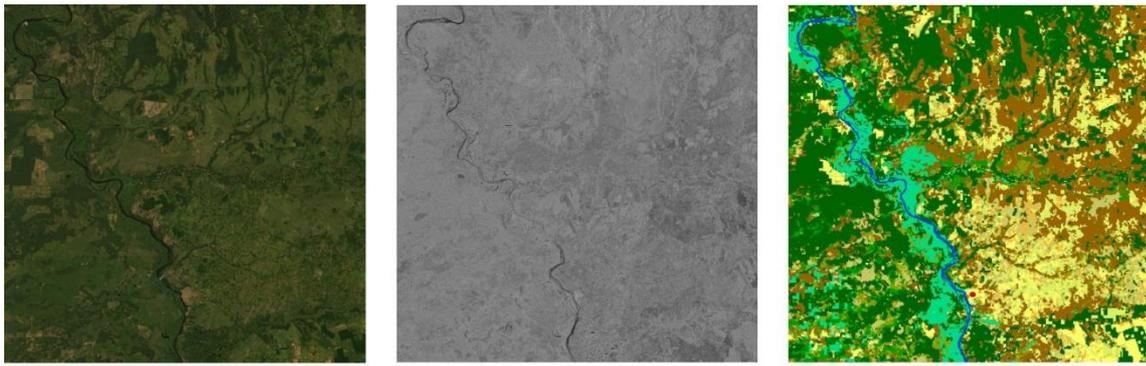


Figure 128. Detail of tile 36PVB: ESRI satellite image (a), water layer without masking out desert/bare soil areas (b), water layer with masking using coarser maps (c) and the final result including all steps described in the text (d).

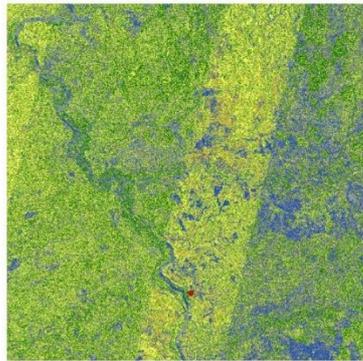
Additionally, this implementation of the water extraction procedure has been further changed in the processing chain for the historical maps, because of the lower quality of the SAR data sets, due both to the different resolution and signal to noise ratio of the past sensors and, more importantly, to the small number of images that have been collected on each tile. Accordingly, for historical mapping the cluster which is assigned to the “water” class is not the one which has more points superimposed with the water class in coarser resolution existing water maps (e.g., the water class in the above-mentioned Copernicus map). Instead, the cluster with the lower minimum average value is considered as water. This ensures that the cluster which is most likely the water one is extracted even when the superposition between it and the coarser map is not ensured, see examples in Figure 129 and Figure 130.



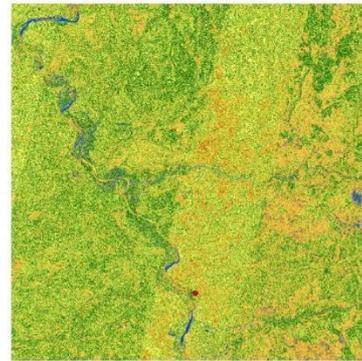
(a)

(b)

(c)



(d)



(e)

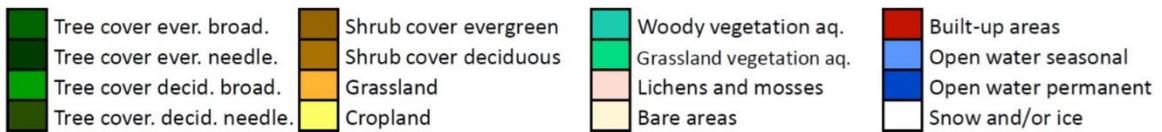


Figure 129. 1995 historical land cover map, tile 22KVQ: ESRI satellite image (a), super image (b), blue: coarse water layer (c) blue: water layer extracted from SAR historical data using the standard approach (d) and the final result exploiting the alternative, in blue (e).

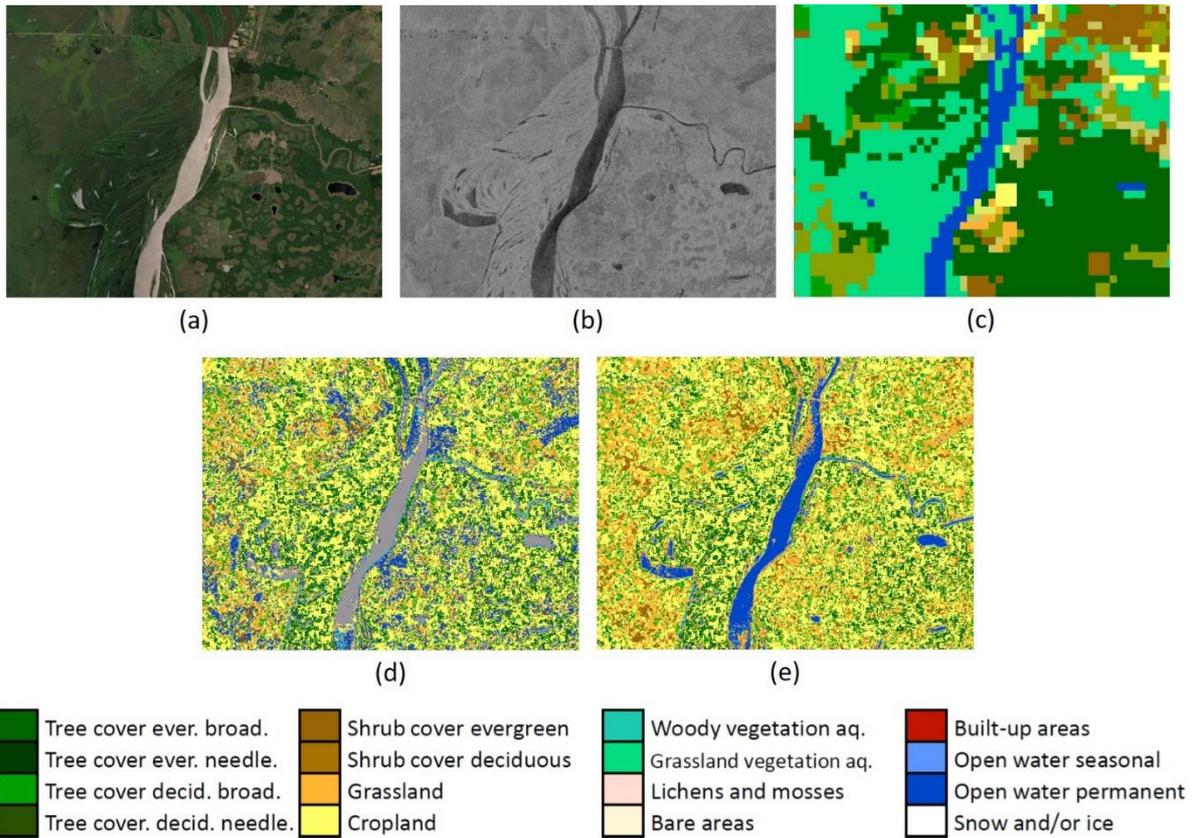


Figure 130. An enlargement of the 1995 historical land cover map, tile 22KVQ: ESRI satellite image (a), super image (b), blue: coarse water layer (c) blue: water layer extracted from SAR historical data using the standard approach (d) and the final result exploiting the alternative, in blue (e).

5.2.3.1 Amazonia: static and historical classification

As mentioned in the previous section, the whole classification task is given by the merging of three algorithms: the Random Forest classifier, the urban extent and the water detectors. The RF classifier is trained with training samples collected by the team deprived of urban and water ground truths. Then, the trained model is applied to SAR features in a dedicated task. Concerning the two detectors, they accept in input a series of parameter to classify the study area, according to suitable instructions to be typed in the Python command prompt. About urban detector, the parameters to set are listed above:

- *m*, the threshold value applied to the image given by the average of the SAR time sequence (the default value is 3);
- *dg*, value in degree used to compute the mask from the DEM image (the default value is 2);
- *t*, threshold value applied to the cost function (the default value is 40);
- *w*, size of the filter kernel (the default value is 3);
- *b*, number of iterations for the closing and opening morphological operations (the default value is 2);
- *oo*, flag to perform the opening operation to the output image (post-processing);
- *oi*, number of closing and opening iterations to be applied in the post-processing phase (the default value is 1);
- *ow*, dimension of the opening kernel (the default value is 3).

The last three parameters are used for the post-processing of the resultant urban map. To compute the Amazon static land cover map, the urban parameters “-dg 0 -b 10 -t 45 -m 3 -oo” are used, while “-dg 20 -b 10 -t 15 -m 3 -oo” sequence is applied for the historical production. If the parameters are not directly specified, default values are used.

For the water extractor, Instead, the following parameters are used:

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	130	

- a , value used to compute the altitude mask (the default value is 400);
- cp , geographical coverage (in percentage) that the SAR data must have to be processed (the default value is 70%);
- dg , value in degree used to compute the mask from the DEM image (the default value is 8);
- w , size of the erosion filter kernel (the default value is 3);
- e , number of erosion iterations (the default value is 1);
- m , max number of water points used to train the K-means classifier (these points are extracted by the coarser land cover maps, ESA CCI LC map 300m and CGLC at 100m) (the default value is 1000);
- kt , parameter used to specify the type of K-means algorithm to use: “auto”, “full” or “elkan”, according to the routine available in the Scikit learn tool [43];
- kn , value used to set the number of cluster for the K-means (the default value is 4);
- ki , parameter used to initialize the K-means (the default value is “k-means++”);
- sn , minimum water pixel number extracted by the coarser maps: this value helps to understand if water bodies are present or not in the area under investigation (the default value is 200).

Taking into the account the list of parameters mentioned above, to carry out the water detection in Amazonia both in the static and historical areas, the following values are used: “-ki random -kt full -sn 50 -cp 10”.

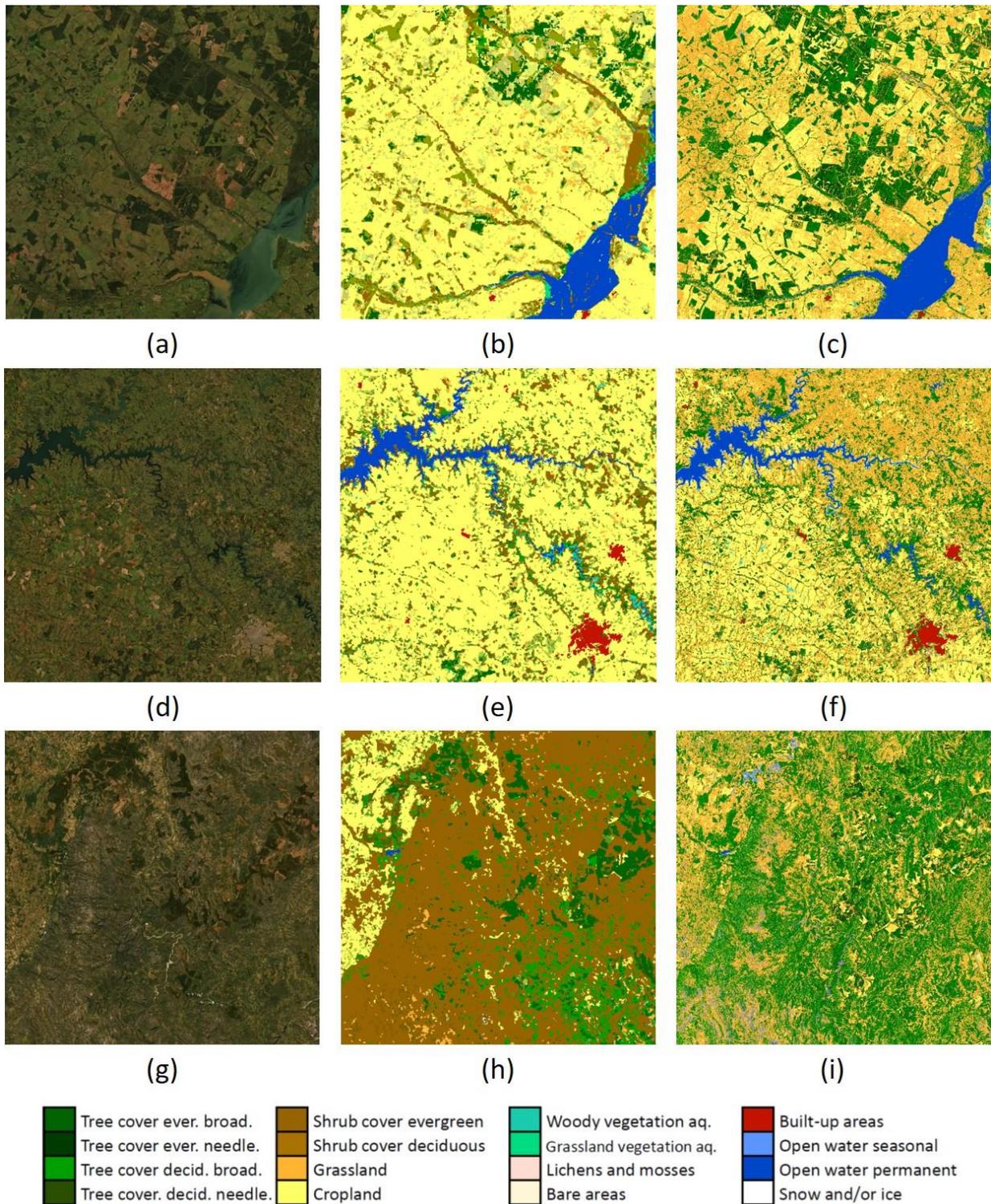


Figure 131. Static land cover maps in Amazonia: 22KCB tile ESRI satellite image (a), 22KCB tile coarser map (b), 22KCB tile S1 2019 land cover map (c), 22KGE tile ESRI satellite image (d), 22KGE tile coarser map (e), 22KGE tile S1 2019 land cover map (f), 23KPA tile ESRI satellite image (g), 23KPA tile coarser map (h) and 23KPA tile S1 2019 land cover map (i).

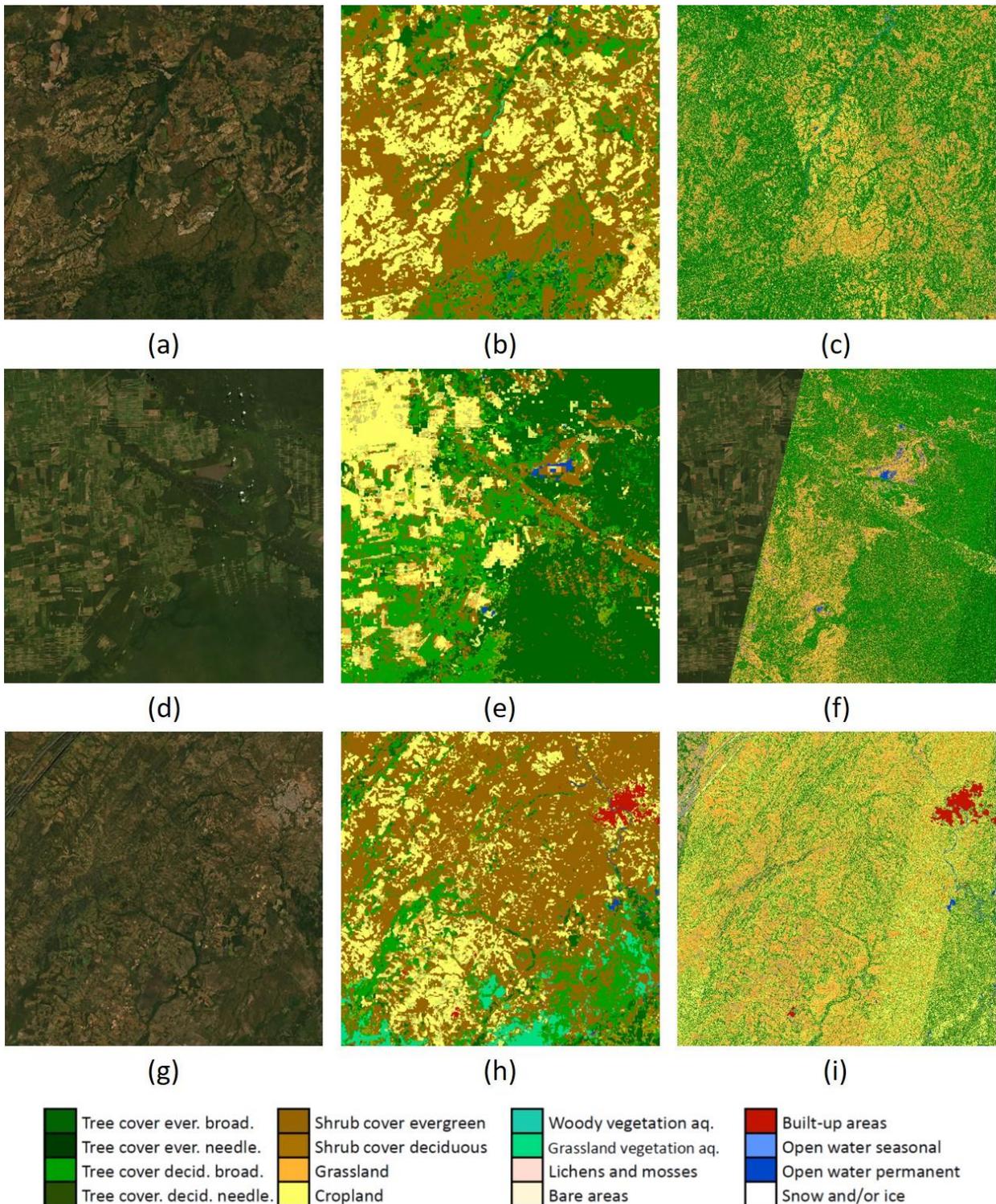


Figure 132. Historical land cover maps in Amazonia: 22LBK tile ESRI satellite image (a), 22LBK tile coarser map (b), 22LBK tile ENVISAT_ASA.IMP.1P 2005 land cover map (c), 22KGE tile ESRI satellite image (d), 22KGE tile coarser map (e), 22KGE tile ERS_SAR.IMP.1P 2000 land cover map (f), 21LWC tile ESRI satellite image (g), 21LWC tile coarser map (h) and 21LWC tile ERS_SAR.IMP.1P 1995 land cover map (i).

By looking at Figure 131, the static maps show a better classification performance than the coarser ones. The classes in the high-resolution maps well represent the geographical scenario observed in the reference image, and more details (e.g., crops, narrow roads, water streams) are depicted. This is very easy to see by comparing Figure 131 (h) and v (i), for instance. About historical classification, Figure 132 shows that urban and water

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	133	

detectors work well also with the historical SAR data, while the classification of the other land covers is not as accurate as in the static maps, because of the much smaller number of SAR images in the sequence.

Moreover, within the maps, several “orbit” artefacts are present. They arise because of the low and not homogeneous geographical coverage of the historical SAR datasets. In fact, some portions of the tile are covered by more images than other ones, and this is emphasized with the average operation applied to the time series, which generates different SNR values.

5.2.3.2 Africa: static and historical classification

To compute the static maps in Africa, the parameter set “dg 20 -b 10 -t 15 -m 3 -oo” is used for the urban recognition, and “-ki random -kt full -sn 50 -cp 10” is applied for the water extraction.

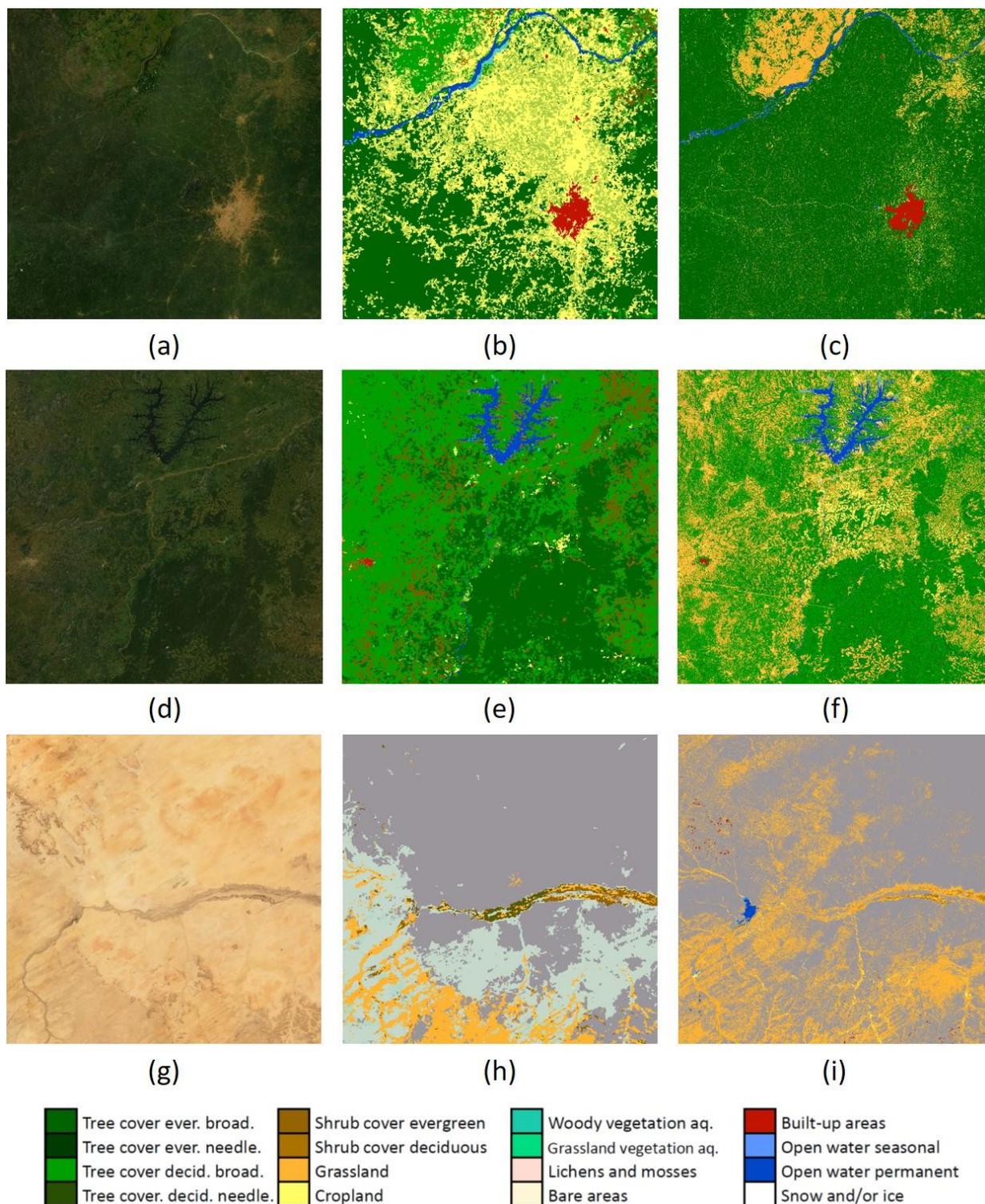


Figure 133. Static land cover maps in Africa: 32NQG tile ESRI satellite image (a), 32NQG tile coarser map (b), 32NQG tile S1 2019 land cover map (c), 32NQM tile ESRI satellite image (d), 32NQM tile coarser map (e), 32NQM tile S1 2019 land cover map (f), 34PGC tile ESRI satellite image (g), 34PGC tile coarser map (h) and 34PGC tile S1 2019 land cover map (i).

Same classification performances already encountered in Amazonia can be appreciated also in Africa. Once again, urban and water classifiers can detect details that are not present in the coarser reference map. This is easily seen in Figure 133 (i), for instance, where a lake, not present in the medium resolution map (see Figure 133 (h)), is depicted. However, several urban outliers are also evident in Figure 133 (i), which correspond to the mountainous reliefs. These are removed easily by a proper thresholding during the urban detector application:

the slopes image is computed by the DEM, and its mean value is analysed. If it is more than 7, the following parameters are applied “-dg 0 -b 10 -t 45 -m 3 - oo”. For the historical production, “-dg 20 -b 10 -t 15 -m 3 -oo” is used for the urban, and ‘-ki random -kt full -sn 50 -cp 10’ for the water algorithm, respectively.

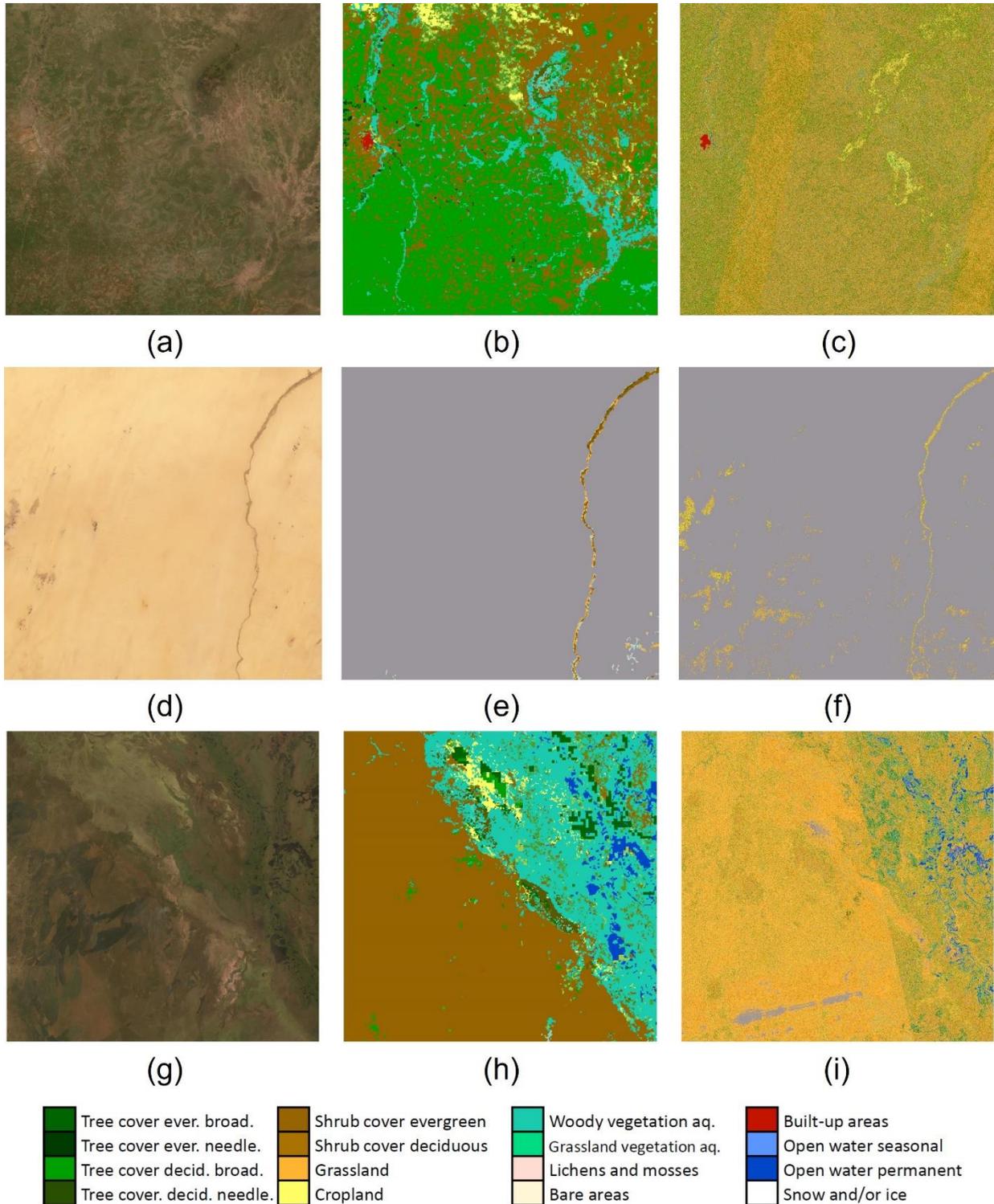


Figure 134. Historical land cover maps in Africa: 35NPJ tile ESRI satellite image (a), 35NPJ tile coarser map (b), 35NPJ tile ERS_SAR.IMM.1P 2005 land cover map (c), 35PPT tile ESRI satellite image (d), 35PPT tile coarser map (e), 35PPT tile ERS_SAR.IMM.1P 2000 land cover map (f), 35NRJ tile ESRI satellite image (g), 35NRJ tile coarser map (h) and 35NRJ tile ERS_SAR.IMM.1P 1995 land cover map (i).

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	136	

5.2.3.3 *Siberia: static and historical classification*

To produce the 2019 land cover map in Siberia, the combination of the urban and water classifier parameters are properly adjusted. Concerning the urban detector, the series “-b 10 - oo” is used, while “-ki random -kt full - sn 50 -cp 10” is used for water detection. The following figure reports several maps for different regions of Siberia static area, according to the S2 tiling grid.

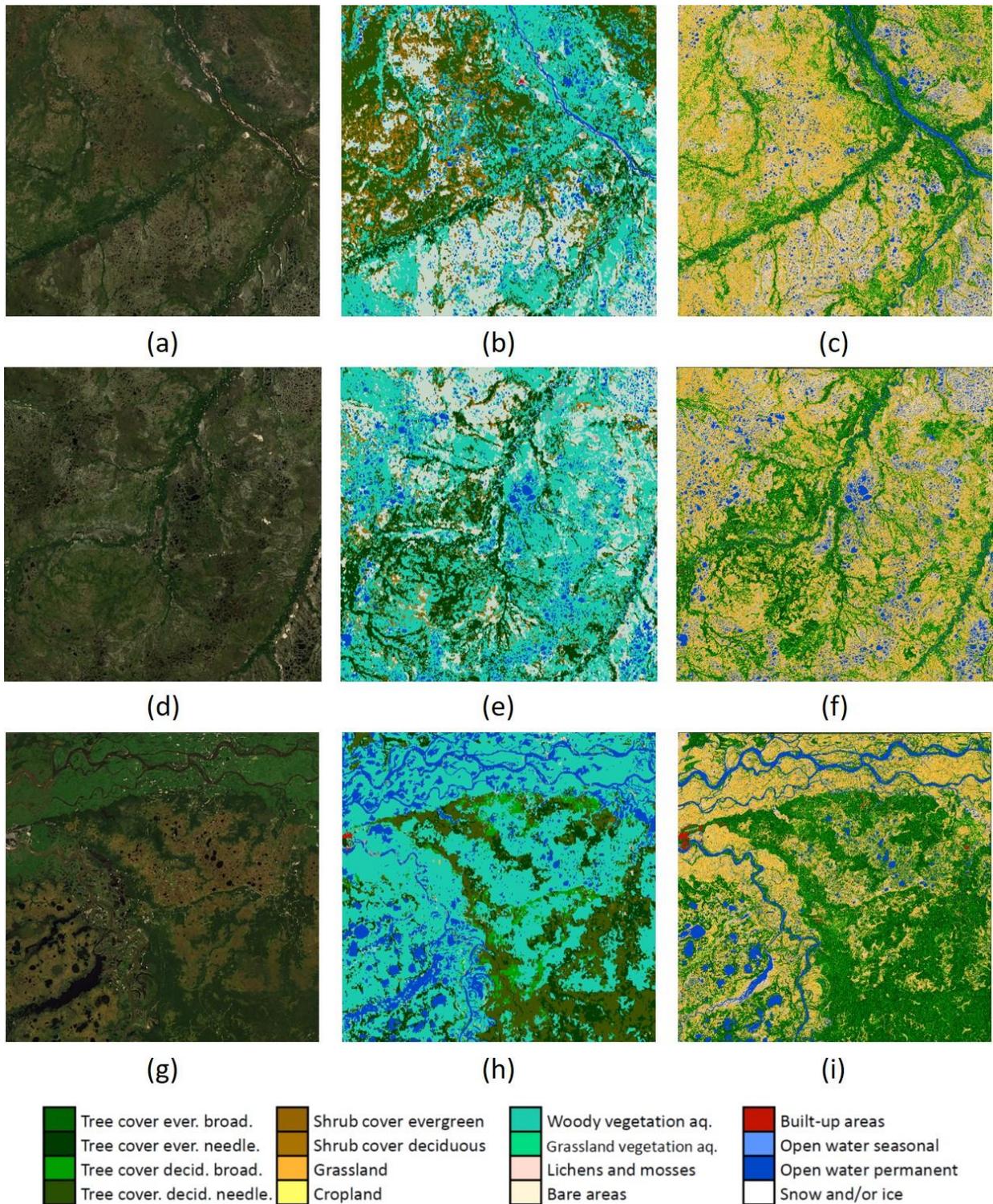


Figure 135. Static land cover maps in Siberia: 42WXT tile ESRI satellite image (a), 42WXT tile coarser map (b), 42WXT tile S1 2019 land cover map (c), 42WXS tile ESRI satellite image (d), 42WXS tile coarser map (e), 42WXS tile S1 2019 land cover map (f), 42VWN tile ESRI satellite image (g), 42VWN tile coarser map (h) and 42VWN tile S1 2019 land cover map (i).

The classification task does not perform well where the number of images is low, which is increasingly the case moving backwards from 2015. This is what happens in this test site, where the SAR historical coverage of the tile is incomplete.

Figure 136 shows the historical maps obtained using SAR only in 2010, 2005 and 1995 years, for the following test tiles: 42WXS, 43WET and 43WFV.

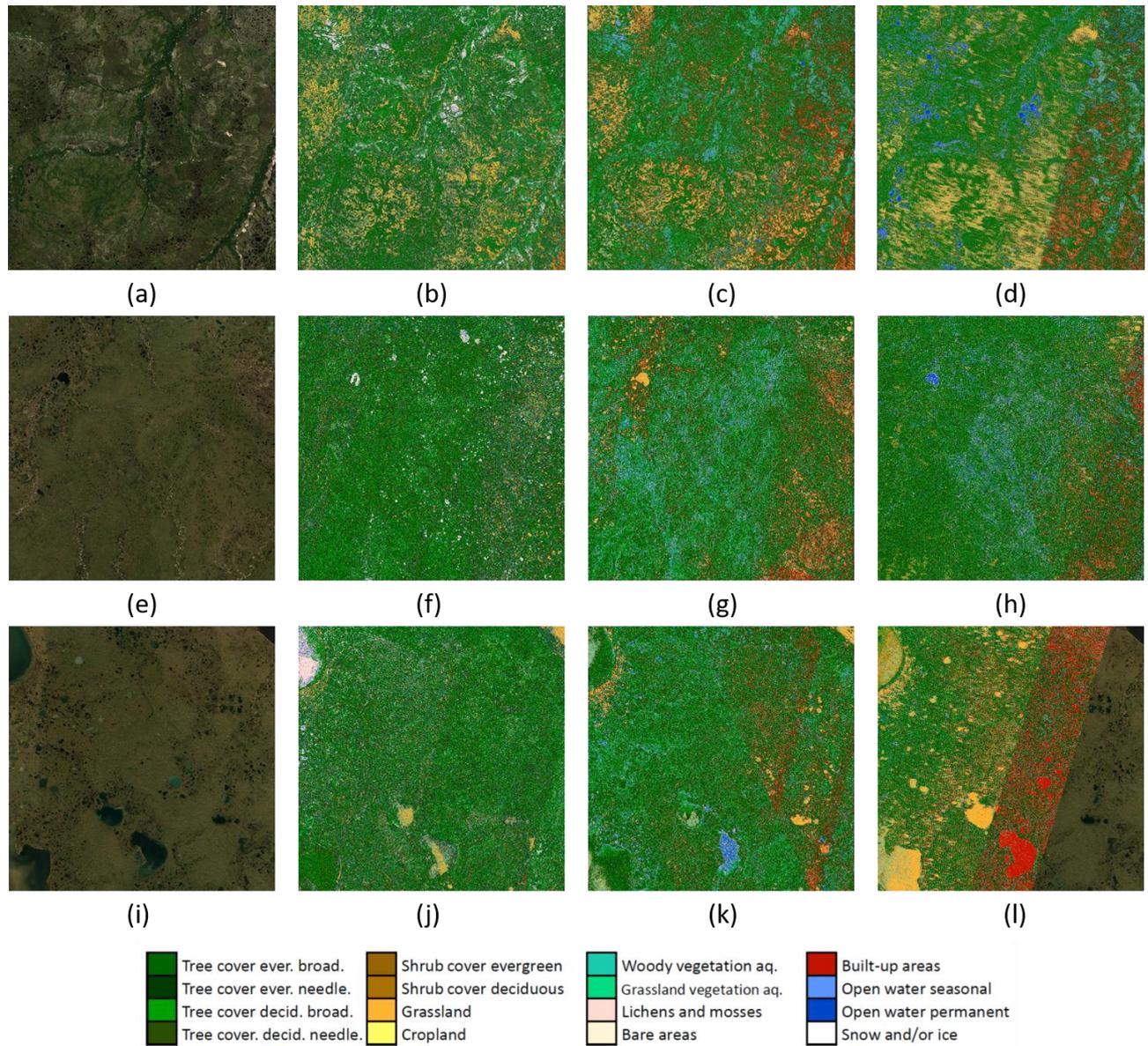


Figure 136. Historical land cover maps in Siberia: 42WXS tile ESRI satellite image (a), 42WXS tile ERS_SAR.IMP.1P 2010 land cover map (b), 42WXS tile ENVISAT_ASA.IMP.1P 2005 land cover map (c), 42WXS tile ERS_SAR.IMP.1P 1995 land cover map (d), 43WET tile ESRI satellite image (e), 43WET tile ERS_SAR.IMP.1P 2010 land cover map (f), 43WET tile ENVISAT_ASA.IMP.1P 2005 land cover map (g), 43WET tile ERS_SAR.IMP.1P 1995 land cover map (h), 43WFV tile ESRI satellite image (i), 43WFV tile ERS_SAR.IMP.1P 2010 land cover map (j), 43WFV tile ENVISAT_ASA.IMP.1P 2005 land cover map (k) and 43WFV tile ERS_SAR.IMP.1P 1995 land cover map (l).

In this test site the classification was performed with the RF classifier only, given the scarcity of SAR images. Indeed, the water and urban extent extraction unsupervised techniques do generate accurate result. However, due the severe data lack, the RF classification performances are not so appealing, too, and a lot of classification errors are evident in the final products. The maps in the figure above present serious urban outliers and the water bodies are not correctly recognized. In addition, we should note in Figure 136 that the historical sequence is not consistent since the classes reported into the maps are very different. In fact, by observing the optical image reference in Figure 136 (i) there is a water target (a lake) in the bottom part of the picture. Considering the 2010 and 2005 land cover maps in Figure 136 (j) and Figure 136 (k) respectively, the lake is not correctly classified in 2010, where it is represented as a mixture of the grassland and cropland classes. In the 2005 map,

the lake is approximately recognized, but results much more performing cannot be achieved because of the scarcity of the SAR geographical coverage in the area.

5.2.3.4 UEXT case study: slums detection in the SAR map

Upon request by ESA, the urban extraction algorithm was tested to verify its capability in detecting slums, very poor and crowded areas. Three Sentinel-2 test sites has been identified, in Africa and Amazonia, within three important cities:

- Nairobi, Kenya, Africa (37MBU S2 tile);
- Rio de Janeiro, Brazil, Amazonia (23KPQ S2 tile);
- Buenos Aires, Argentina, Amazonia (21HUB S2 tile);

Where it is well known that social and economic gaps led to the creation of important districts of barracks and dilapidated buildings. As ground truths, a database with the most important slums in India, Africa and South America was considered, freely available at the following web link: <https://github.com/jkkersey/Mapping-Urban-Informality>**Error! Hyperlink reference not valid..**

As mentioned earlier, the urban detector exploits the mean backscatter coefficient of the 2019 S1 time series, and a thresholding operation is applied. Apparently, the module seems to work well, but not all slums sites are properly delineated.

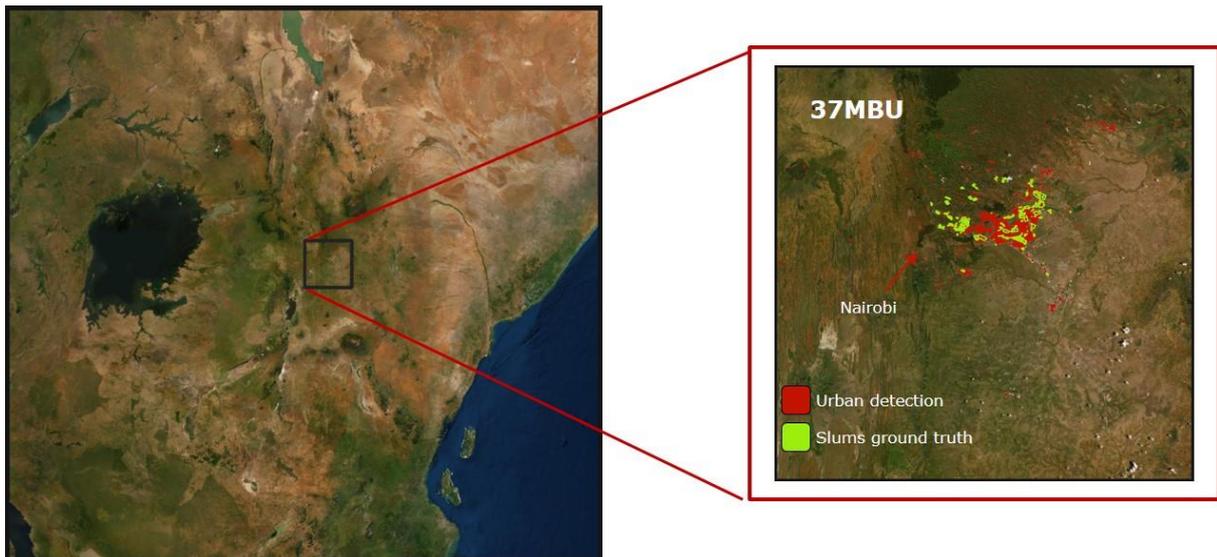


Figure 137. Slums detected in Nairobi city, tile 37MBU.

Error! Reference source not found. reports the results for the Nairobi test site, where the slums ground truth are in green while in red is the extracted urban map.

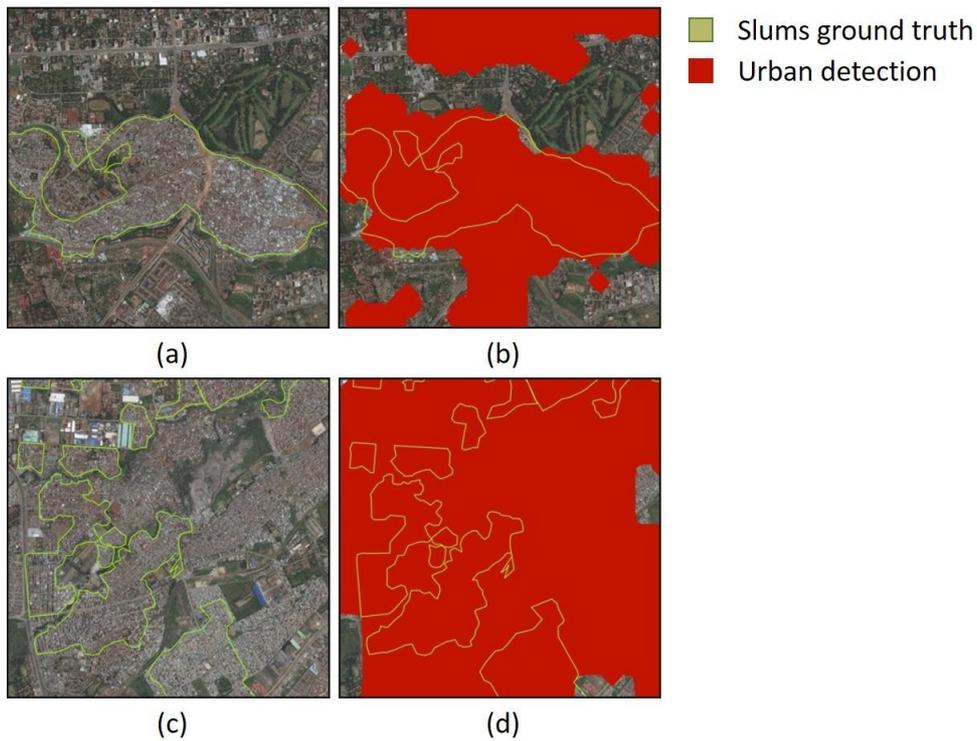


Figure 138. Slums in Nairobi city, tile 37MBU: ground truth of the Kibera district (a), urban detection of the Kibera district (b), ground truth of the Dandora/Korogocho district (c) and urban detection of the Dandora/Korogocho district (d).

In **Error! Reference source not found.** two of the most important and well known slums in Nairobi, Kibera and D andora/Gorogocho districts, are reported. Comparing the ground truth and urban map. **Error! Reference source not found.** (a) and **Error! Reference source not found.** (b), for instance, the slums are detected and reported in the urban map in both cases, although small areas are not well covered due the spatial limitation of the S1 sensor. However, not all settlements are correctly depicted in the urban map, as reported in **Error! Reference source not found.**. This can be easily explained since the urban algorithm works based on a strong thresholding applied to mitigate those large-scale outliers arising from SAR distortions. Anyway, the application of a DEM can solves this issue.

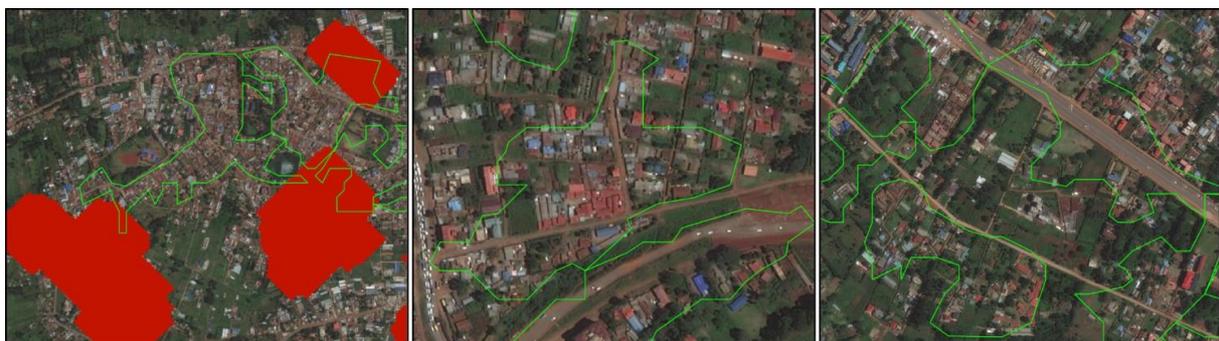


Figure 139. Examples of slums in Nairobi not detected by the UEXT module.

To strengthen this result, the detection has been validate on two additional cities, Rio de Janeiro, for example, and good capability of urban detector to recognize the slums was find. In Rio, for instance, the “complexo do Alemão” is completely covered, and, in the same way, this is true also for the other barrios/districts.

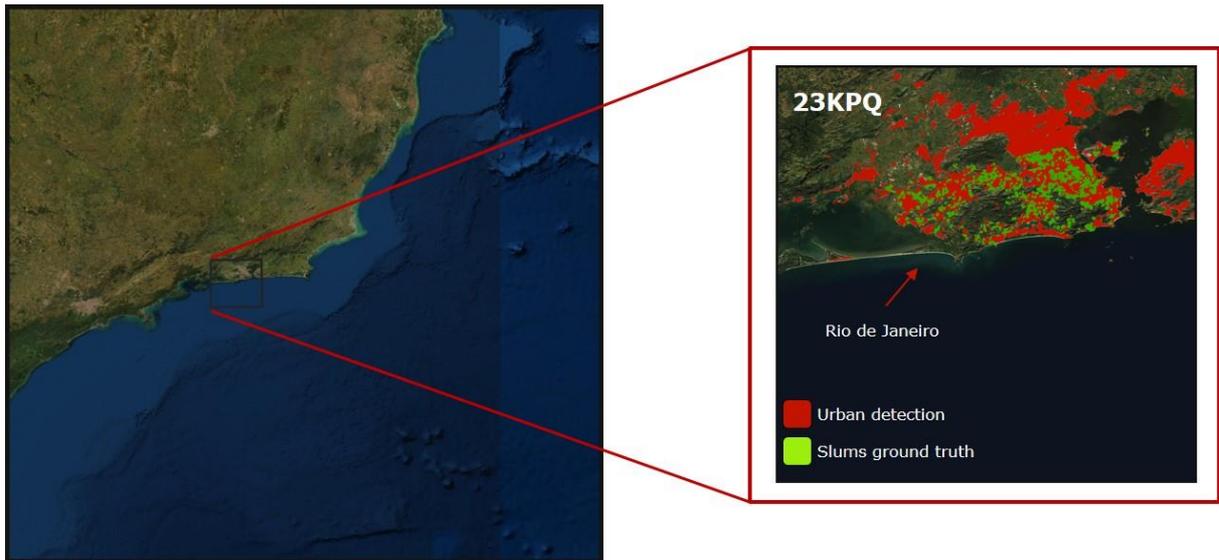


Figure 140. Slums detected in Rio de Janeiro city, tile 23KPQ.

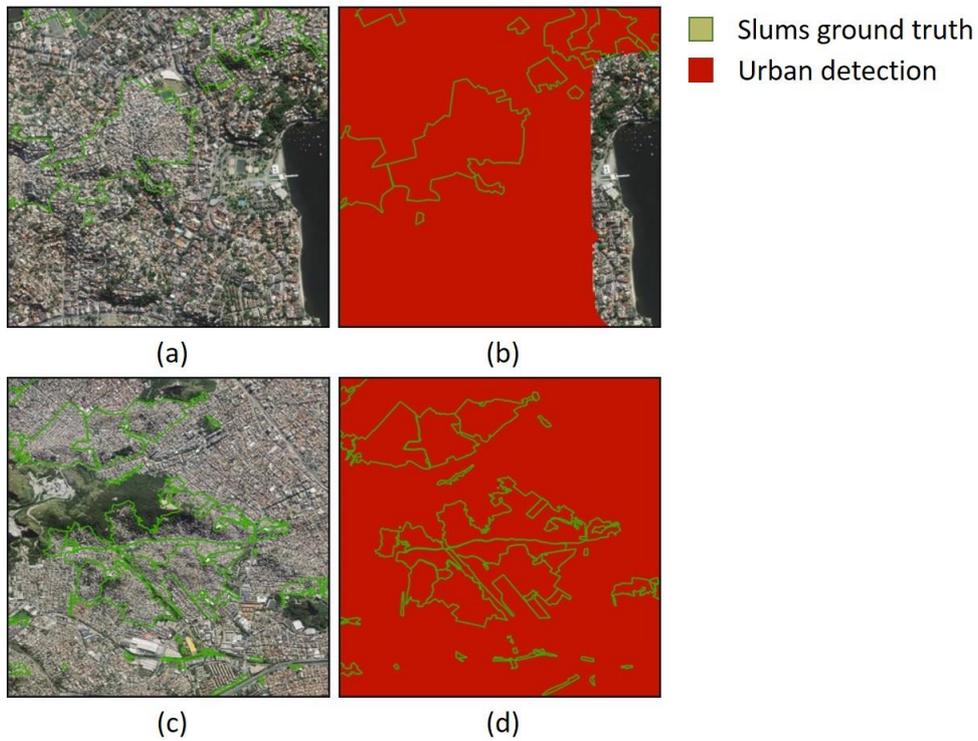


Figure 141 Slums in Rio de Janeiro city, tile 23KPQ: ground truth of the Jardim Carioca barrio (a), urban detection of the Jardim Carioca barrio (b), ground truth of the Complexo do Alemão (c) and urban detection of the Complexo do Alemão (d).

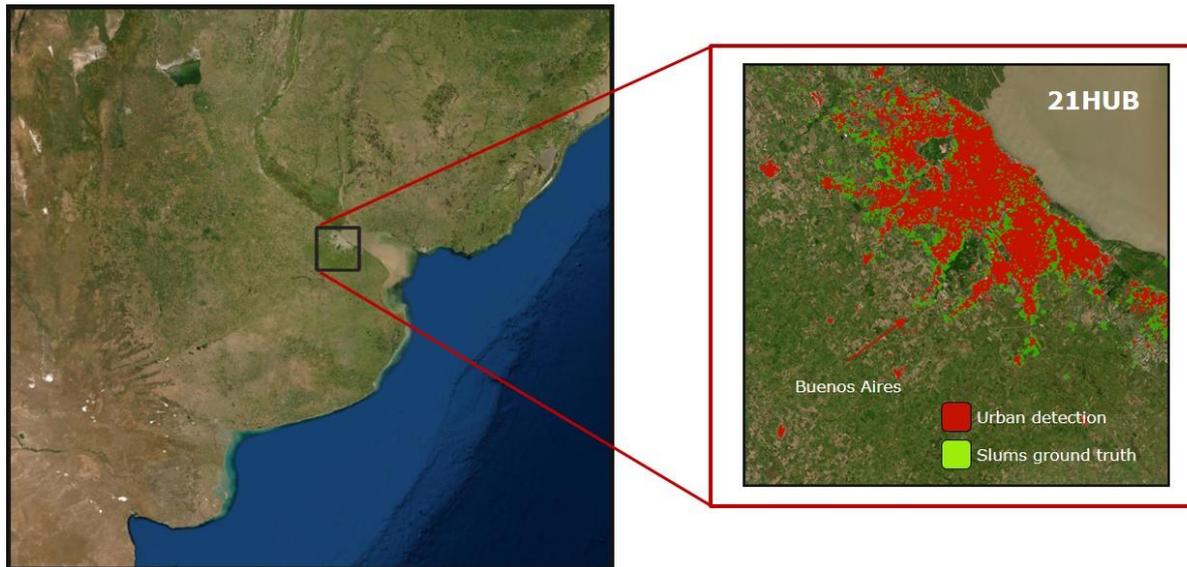


Figure 142. Slums detected in Buenos Aires city, tile 21HUB

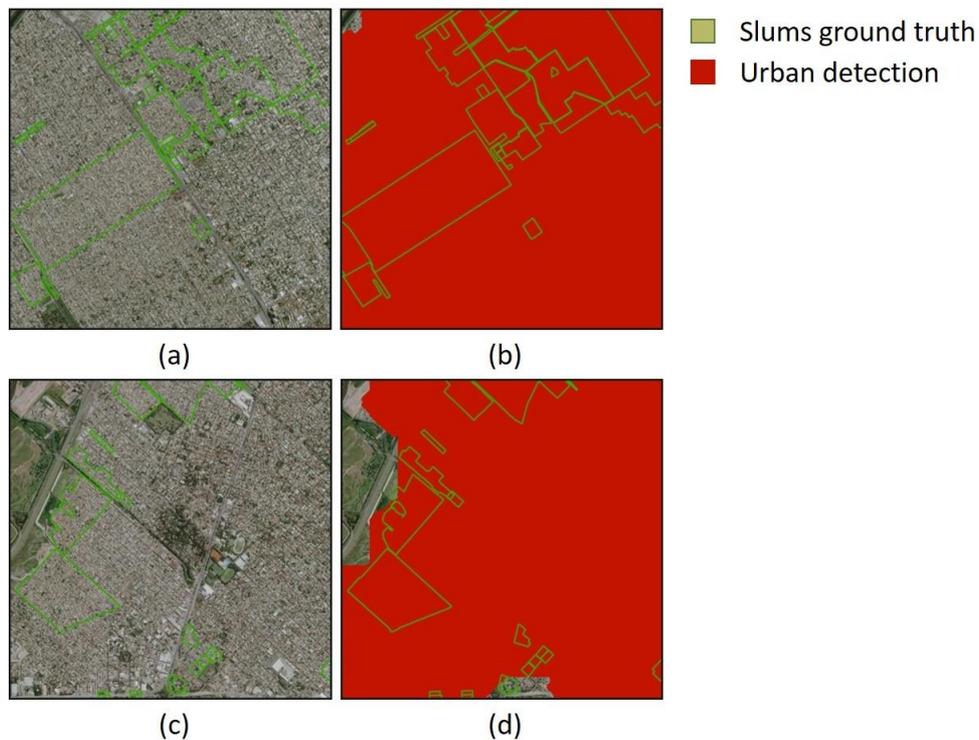


Figure 143. Slums in Buenos Aires city, tile 21HUB: ground truth of the 17 de Noviembre barrio/ Villa Fiorito (a), urban detection of the 17 de Noviembre barrio/ Villa Fiorito (b), ground truth of the Loma Hermosa district (c) and urban detection of the Loma Hermosa district (d).

5.2.3.5 UEXT case study: slums detection in the HRLC10 map

The performance of the UEXT algorithm about the slums detection have been evaluated also for the HRLC10 maps, given by the fusion of the optical S2 and the SAR S1 classification maps. The following test sites have been considered for the evaluation:

- **Kampala city**, Uganda, Africa (36NVF S2 tile).
- **Youndè city**, Cameroon, Africa (32NQG S2 tile).

- **Rio de Janeiro city, Brazil, Amazonia (23KPQ S2 tile).**

As concerns the urban sites in Uganda and Cameroon a qualitative assessment only is provided since that areas are not covered by the ground truth collected in the considered database, which is free accessible at the following link: <https://github.com/jkkersey/Mapping-Urban-Informality>**Error! Hyperlink reference not valid..** For the Rio de Janeiro city both the qualitative and the quantitative assessments have been evaluated.

The figures below show the slum districts identified via photointerpretation of the Google images in the cities of Kampala and Youndè, which were also recognized by the urban detection algorithm and depicted into the HRLC10 maps. In Kampala (Uganda, Africa), four slums have been identified: Bwaise, Katanga, Kasanvu, and Kasokoso (Figure 144).

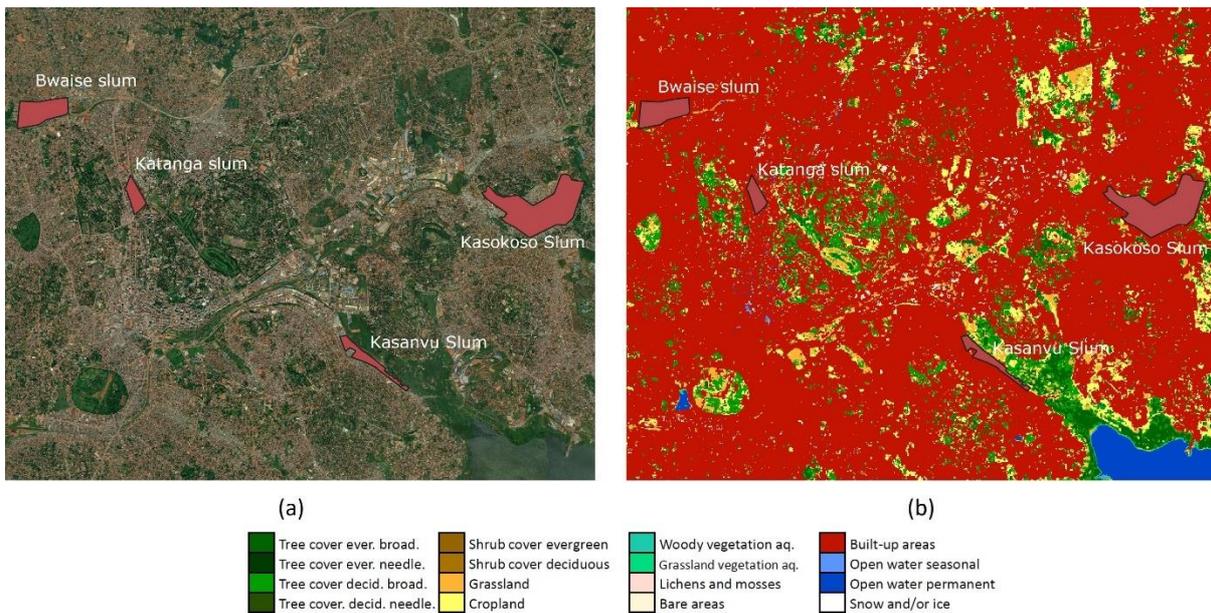


Figure 144. An example of slums distribution in Kampala (Uganda, Africa): slum settlements boundaries are reported on the Google optical reference image (a) and in the HRLC10 map (b).

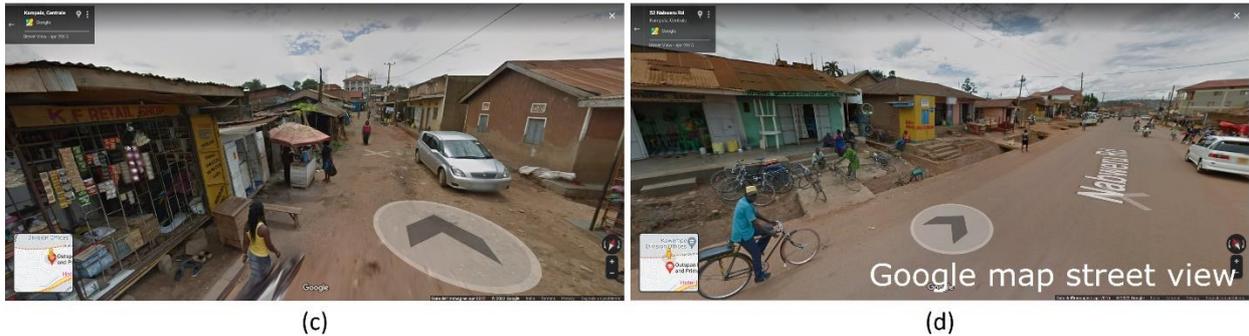


Figure 145. Enlargement of the Bwaise slum in Kampala (Uganda, Africa): Google optical reference image (a), HRLC10 map (b) and photointerpreted ground truths acquired with the Google map street view tool (c), (d).

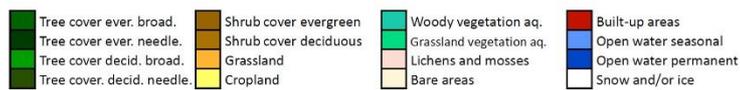
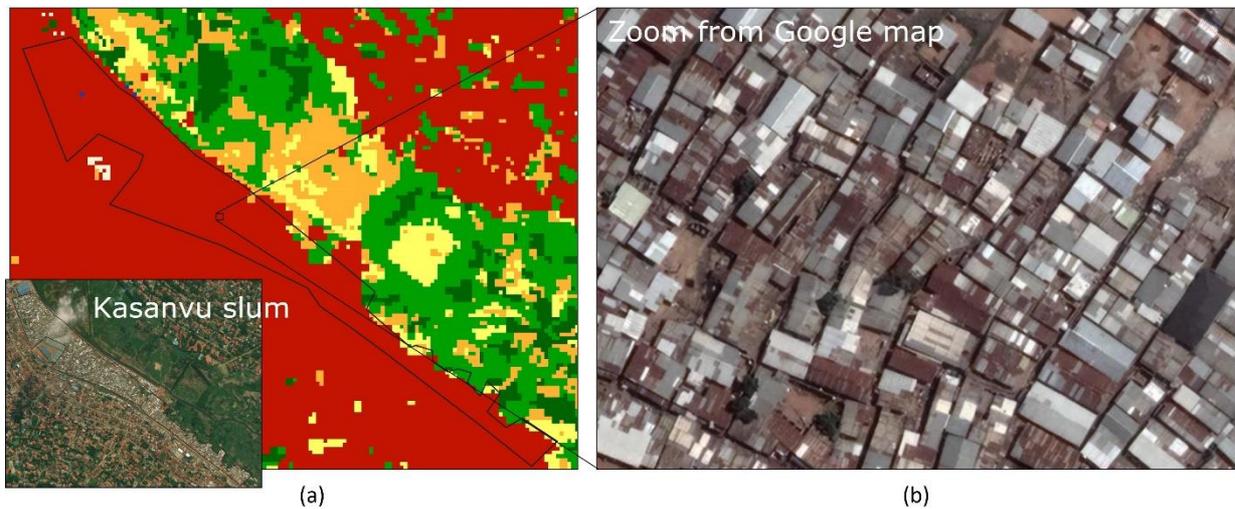


Figure 146. Enlargement of the Kasanvu slum in Kampala (Uganda, Africa): Google optical reference image and HRLC10 map (a) and a zoom from Google map of a little area in the Kasanvu slum (b).

In the following, the slum settlements of the Youndé city, in Cameroon (Africa) are shown (Figure 147 and Figure 148).

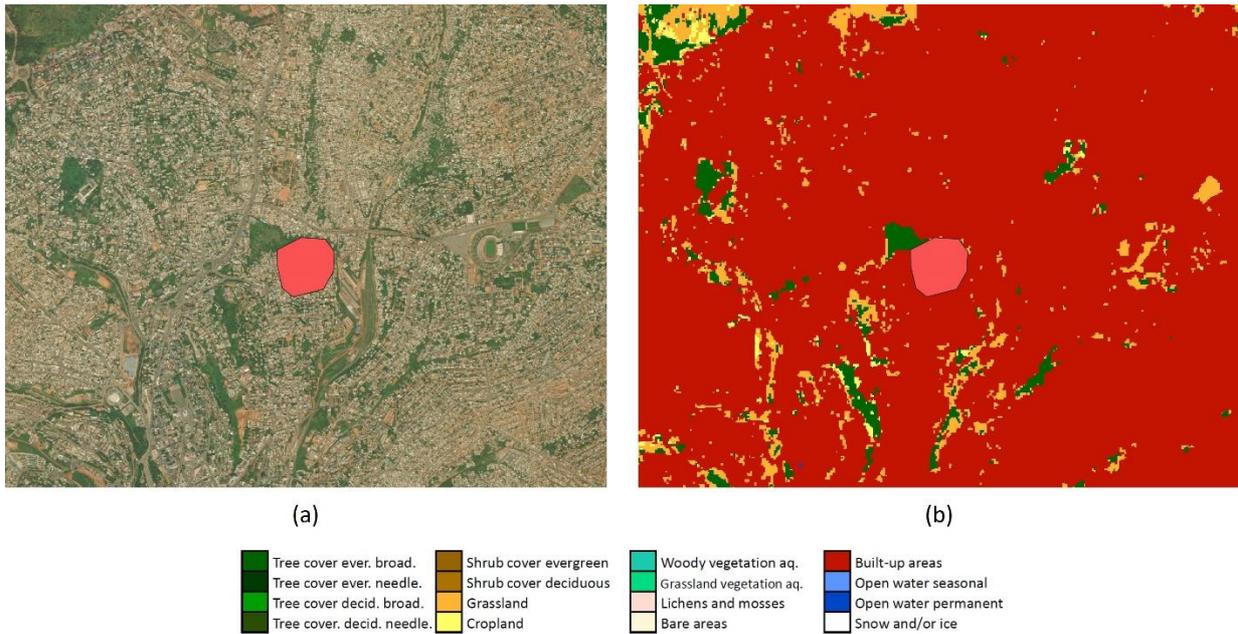


Figure 147. An example of a slum in Youndè (Cameroon, Africa): slum settlement boundary is reported on the Google optical reference image (a) and in the HRLC10 map (b).

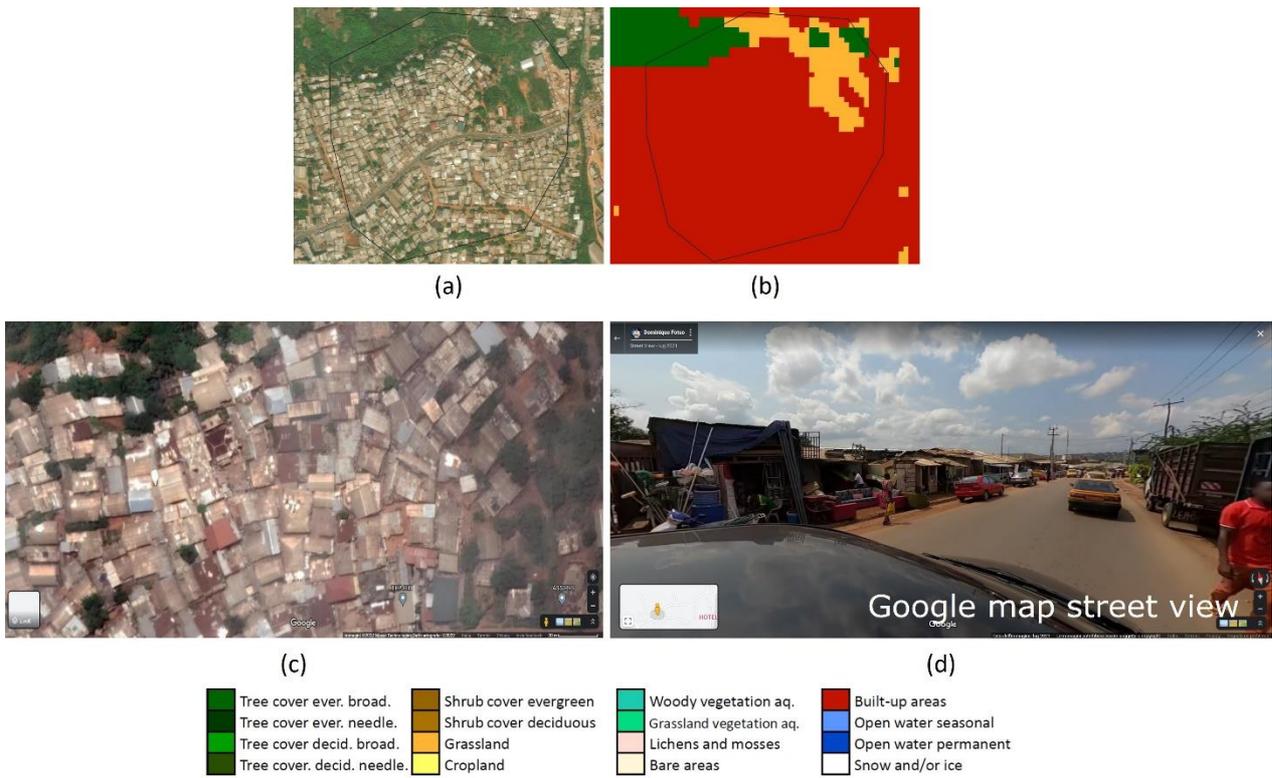


Figure 148. Enlargement of a slum in Youndè (Cameroon, Africa): Google optical reference image (a), HRLC10 map (b) and photointerpreted ground truths acquired with the Google map street view tool (c), (d).

Concerning the city of Rio de Janeiro in Brazil (Amazon area) the slums extraction from HRLC10 map have been visually compared according to the geospatial data for urban informality downloaded from the open web repository.

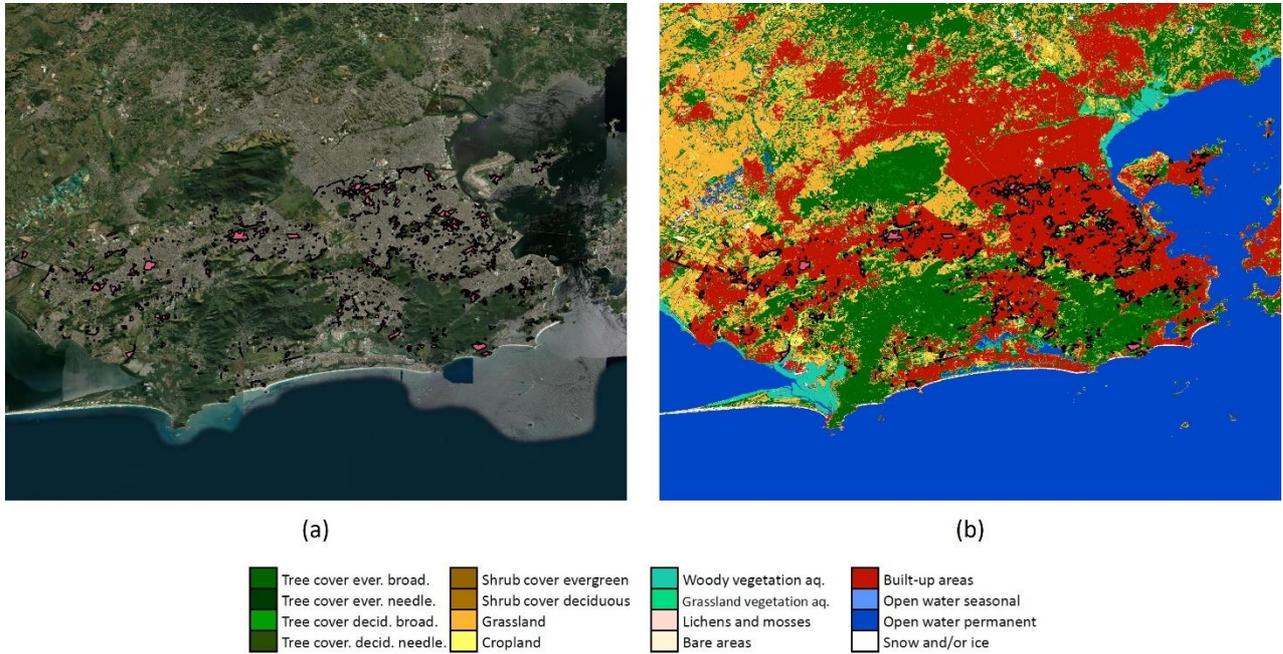


Figure 149. Slums distribution in Rio de Janeiro (Brazil, Amazon): slum settlements boundaries are reported in pink on the Google optical reference image (a) and in the HRLC10 map (b).

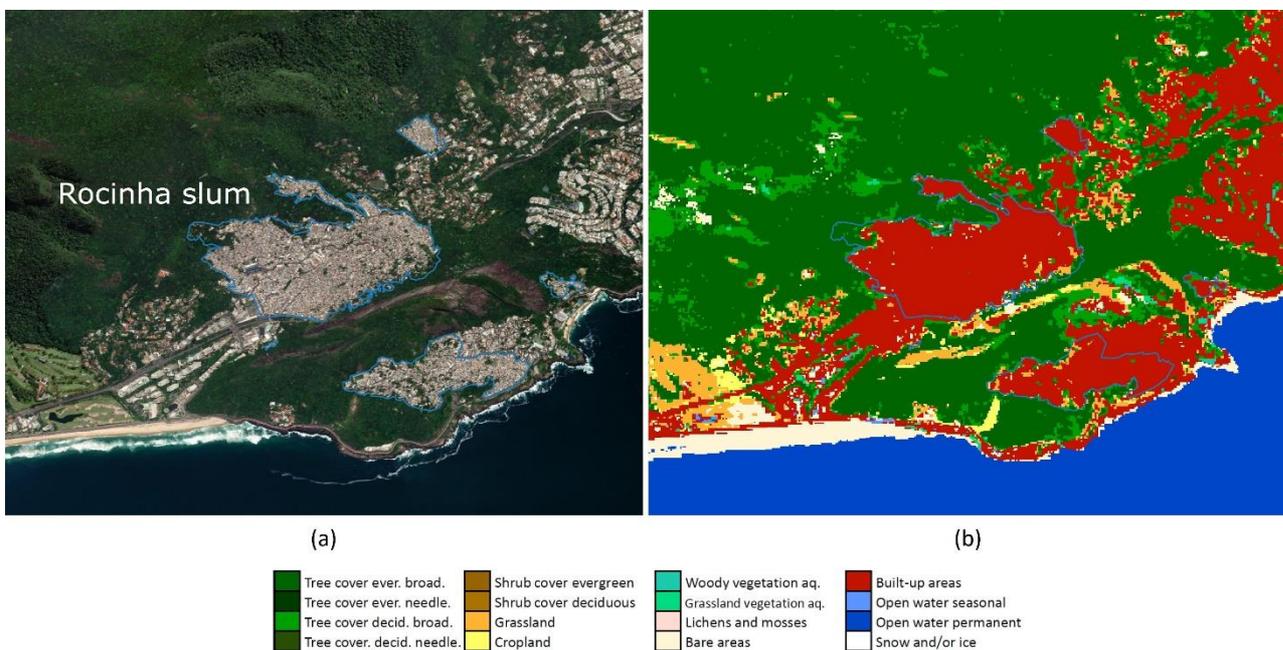


Figure 150. Enlargement of the Rocinha slum in Rio de Janeiro (Brazil, Amazon): the Google optical reference image (a) and the HRLC10 map (b) are compared.

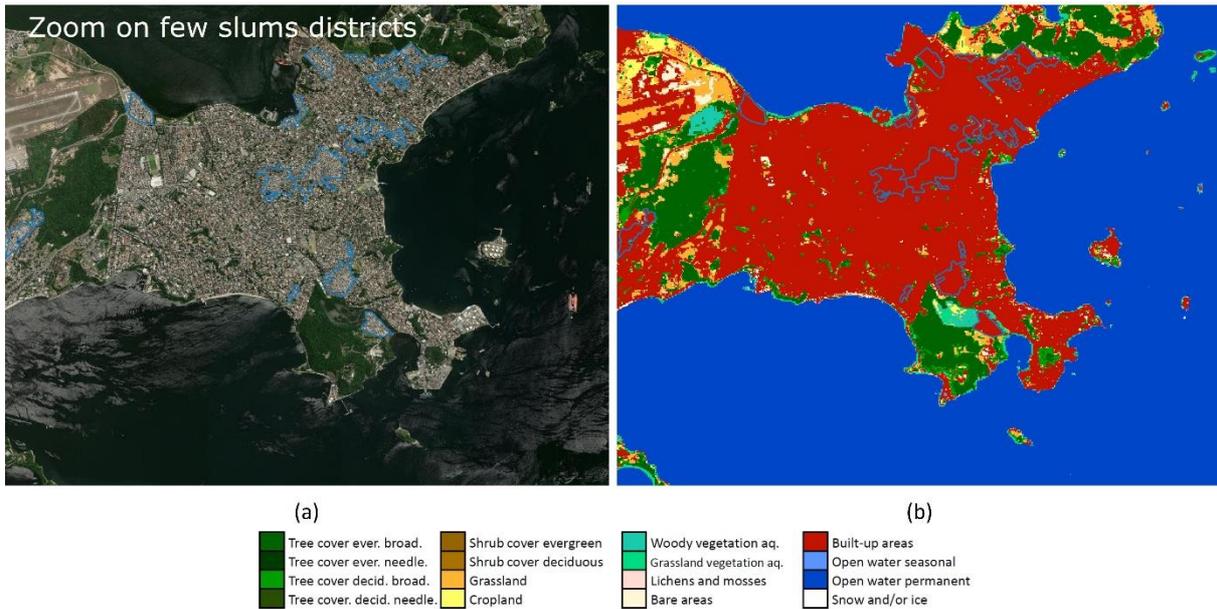


Figure 151. Enlargement of a slums distribution (in blue) in Rio de Janeiro (Brazil, Amazon): the Google optical reference image (a) and the HRLC10 map (b) are compared.

Given the availability of the ground truth, also a quantitative analysis was provided. A validation dataset with 2000 points have been randomly extracted from the slums ground truth to achieve the following balanced distribution: 1000 slum points and 1000 no-slum points (see). Then, the confusion matrix and the overall accuracy (OA) have been computed and reported in the Table 7.

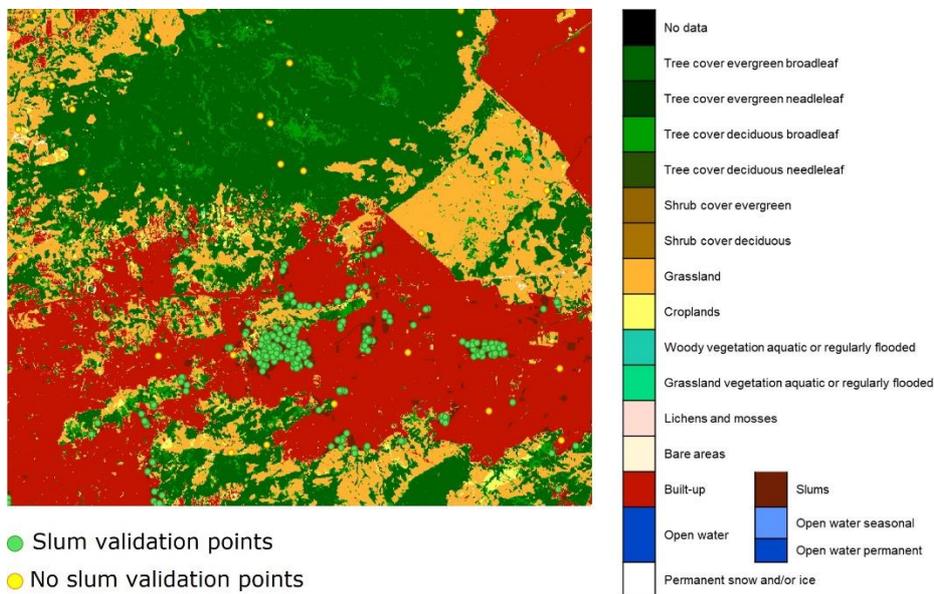


Figure 152. Distribution of a balanced validation dataset extracted for the Rio de Janeiro city (Brazil, Amazon). Slum validation points are in green, while the no slum ones are in yellow.

Table 7. Quantitative assessment for the slum detection in Rio de Janeiro (Brazil, Amazon).

	Slum	No slum	PA (%)	UA (%)	OA (%)
Slum	1903	97	99,95	95,15	97,55
No slum	1	1999	95,37	99,95	

The obtained values confirm what via visual inspection has been seen, with an OA value of about 97% that asserts the good capability of the UEXT approach in identifying the dilapidated urban districts as well.

5.2.4 Final decision

SAR images for years prior to 2013 (and the launch of Sentinel-1) have a very different quality in terms of signal to noise ratio and there are very different amounts of acquisitions per year in different geographic locations. The preliminary analysis performed by UNIPV shows that there are tiles in the historic areas that are not covered by any acquisitions in one or more of the years where the mapping should be performed. The remaining tiles have from 1 to more than 20 acquisitions per year, making their quality very different. Additionally, in some cases the images are obtained from ascending and descending orbits, or from different orbits. As a result, and to reduce the differences, SAR has been used in different ways depending on the geographic location:

- In Siberia, no SAR data sets has been considered for the historic map, because of the low quality.
- In Amazonia, only tiles that are covered by at least six images form the same orbit are considered, although it has been found that 8 to 10 images are necessary to ensure a consistent quality and a contribution by SAR-based maps to the fusion that can be reasonably considered as always constructive.
- In Africa, tiles that are covered by at least 5 images form the same orbit are considered. The number of images is different in this case because the overall number of images per year form the same orbit is in most of the cases smaller than 6 (the threshold used in Amazonia).

5.3 Multi-Sensor Optical-SAR Data Fusion and Decision Fusion

5.3.1 Experimental analysis on the static and historical area

This section presents a combined analysis of the static map of 2019 and of the historical maps, for the years 2015, 2010, 2005, 2000, 1995 and 1990, as generated by the pipeline of the second production. In particular, the effect of the multitemporal cascade model is investigated.

A fundamental aspect to keep in mind when analysing the historical maps is the focus of the study on climate changes. For that reason, while the advantages given by the increase in the spatial resolution of Sentinel-1/2 acquires a central role, it is necessary to keep coherence with the historical data acquired by Landsat 7, Landsat 5, etc. Therefore, for each tile, the static map of 2019 is analysed and then used as a comparison.

A main issue when going back in time is the scarcity in the quantity and quality of the input data available for the classification at that time. This is especially true for Sentinel-1 data, as compared to previous missions with SAR payloads. This can make traditional fusion approaches less effective in the historical case than in the static case. The proposed cascade approach is aimed at propagating the information through time (as MRF does in space) in order to improve the classification accuracy and favour consistency along the time axis.

5.3.1.1 Africa Tile T37PCP

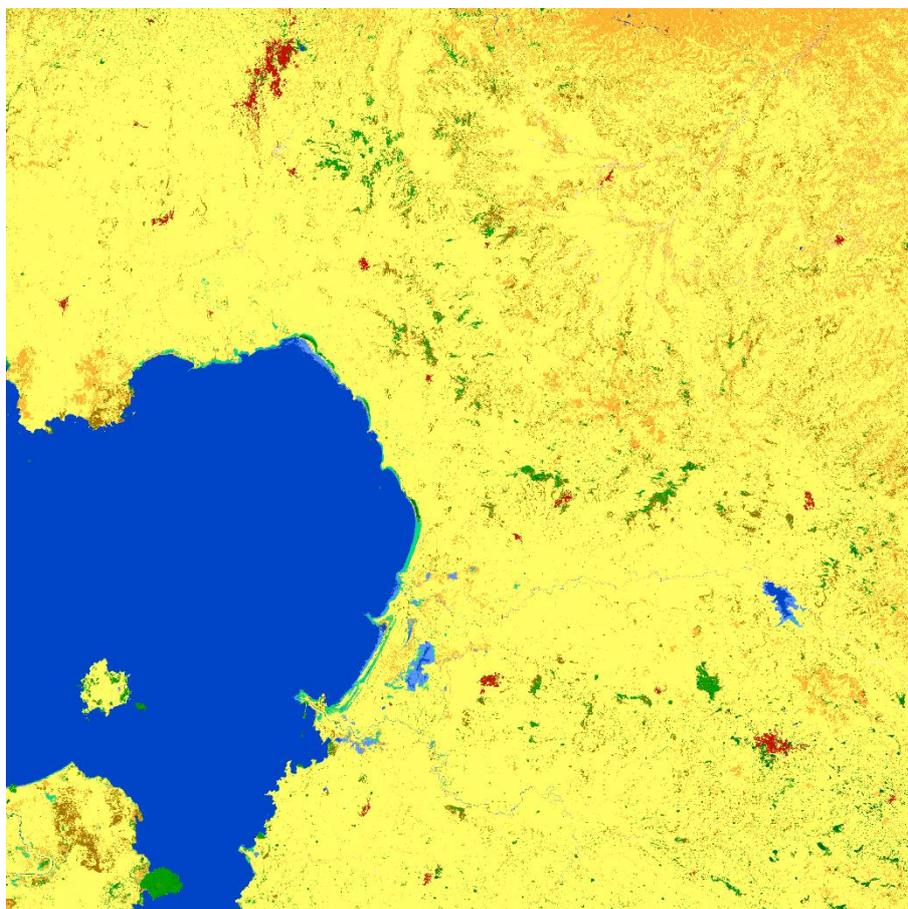


Figure 153. Static map for tile 37PCP

The 2019 classification map for African tile 37PCP is shown in Figure 153. With respect to the previous production, the overestimation of urban and shrub classes is greatly reduced.

The comparisons of optical and cascade maps for years 2015, 2010 and 2005 are shown in Figure 154, while Figure 155 shows the comparison for 2000, 1995 and 1990. Comparing the maps, the effectiveness of cascade in the reduction of false changes across the considered years is evident. In particular, the propagation of multitemporal information using cascade can improve the consistency of the maps in time. In particular, different typologies of errors, which exclusively impact some years or areas, are fixed thanks to the cascade contribution.

As an example, the optical map for 2010 (central row, left panel of Figure 154) is affected by a striping effect. This is caused by the well-known striping issue in the acquisitions coming from Landsat-7. Thanks to the cascade module, this issue is almost completely removed. The maps for 2000, 1995 and 1990, shown in Figure 155, all show a certain overestimation of the urban class in the upper right part of the tile. In all the cases, with the multitemporal cascade model it is possible to strongly mitigate or even removing the issue entirely.

On one hand, this confirms the ability of the cascade approach to correctly propagate the information backward in time and to correctly fix changes in the classification maps which are not reflected by actual changes in the landscape. On the other hand, it is also important to assure that the multitemporal cascade does not influence the cases in which a change actually happened. Figure 156 shows a detail of tile 37PCP and illustrates a well-known case of change, in order to verify the aforementioned case. The figure shows a detail of the artificial lake which can be seen in the middle-low area in the right of Figure 153. This lake was originated by the building of a

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	150	

dam over the Rib river in Ethiopia, which was completed in 2017, providing 234 million cubic meters in water conservation ability. As can be noticed in , in this case, the information propagated by cascade is enough to partially recover the parts of the river which are missed in the optical map while no part of the artificial lake is present in 2015. This also confirms that the ability of cascade to work at a Bayesian level allows to avoid strict constraints which may end up censoring changes.

In other words, the conducted experimentation indicates that the multitemporal cascade approach is advantageous in the reduction of false changes across consecutive LC maps without leading to missed changes. We also recall that, among the hyperparameters of the cascade module, the JPM (joint probability matrix) is included and that the JPM is defined according to prior knowledge on the expected temporal transitions in each area (Siberia, Amazon, Sahel), as formulated by the Climate Group (v. ATBD-v3). In this respect, the tuning of the JPM on each area has been performed precisely in such a way to minimize risks of censoring changes or propagating artifacts from 2019 to the past. The aforementioned experimental results did confirm this rationale and the opportunity of taking benefit from prior knowledge on each target area within the multitemporal modelling. Similar comments holds with regard to the Amazon and Siberian areas as well.

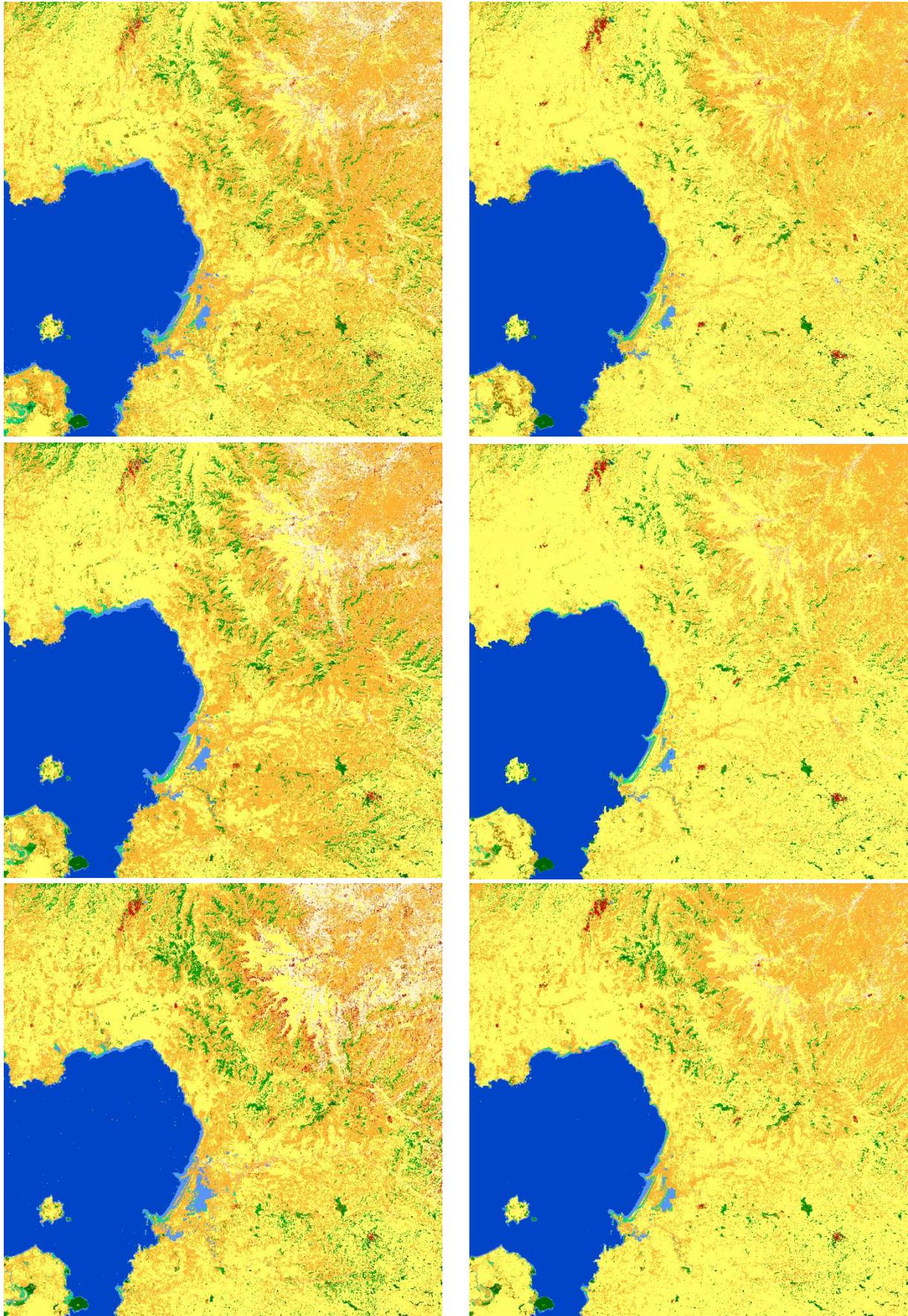


Figure 154. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 37PCP. Year 2015 on the top row, 2010 in the centre and 2005 in the bottom.

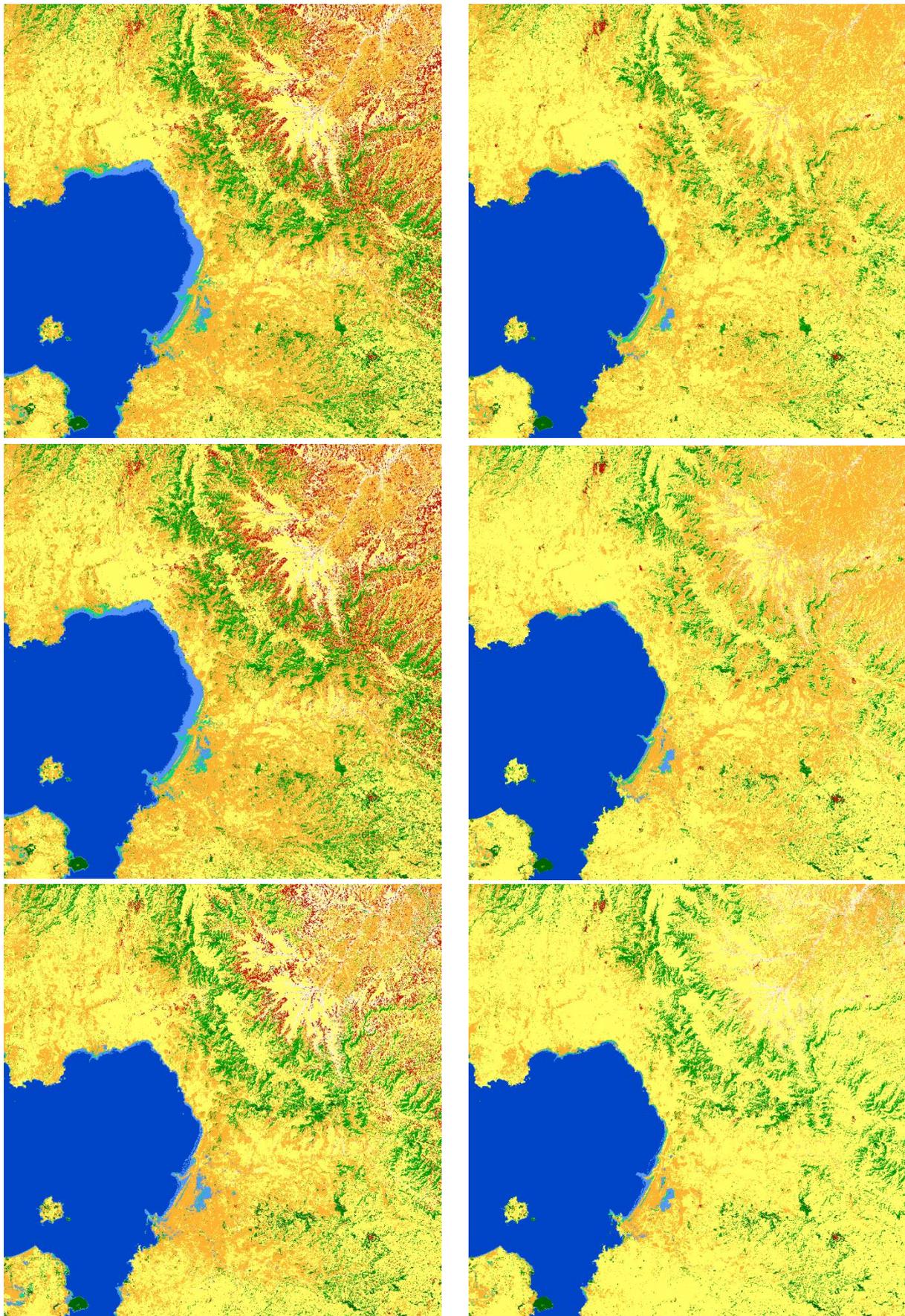


Figure 155. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 37PCP. Year 2000 on the top row, 1995 in the centre and 1990 in the bottom.

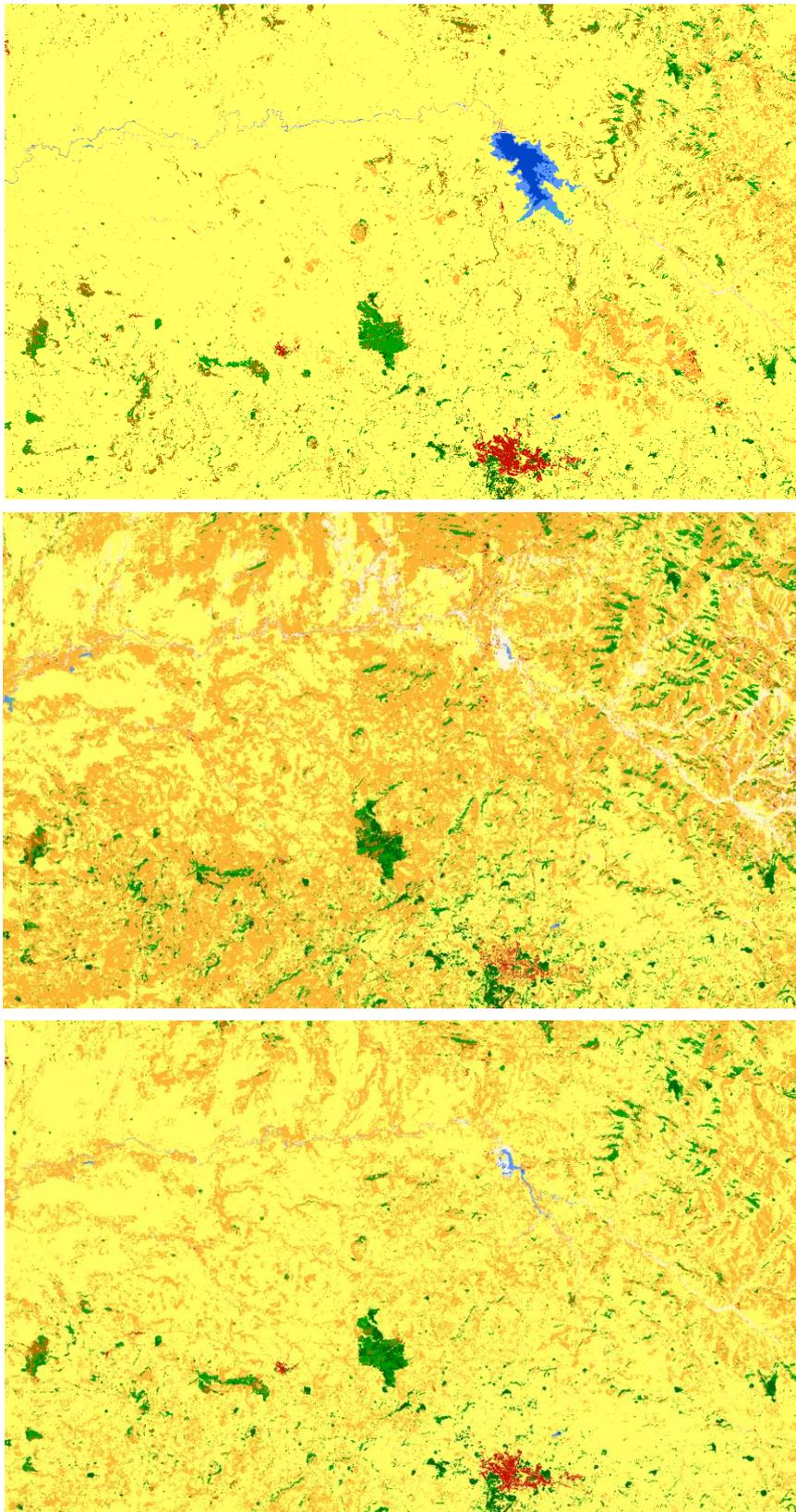


Figure 156. Comparison of a detail of an artificial lake over the Rib river in Ethiopia. On top, the static map for 2019. The optical map of 2015 is in the middle while the final cascade map is on the bottom.

5.3.1.2 Africa Tile T37PES

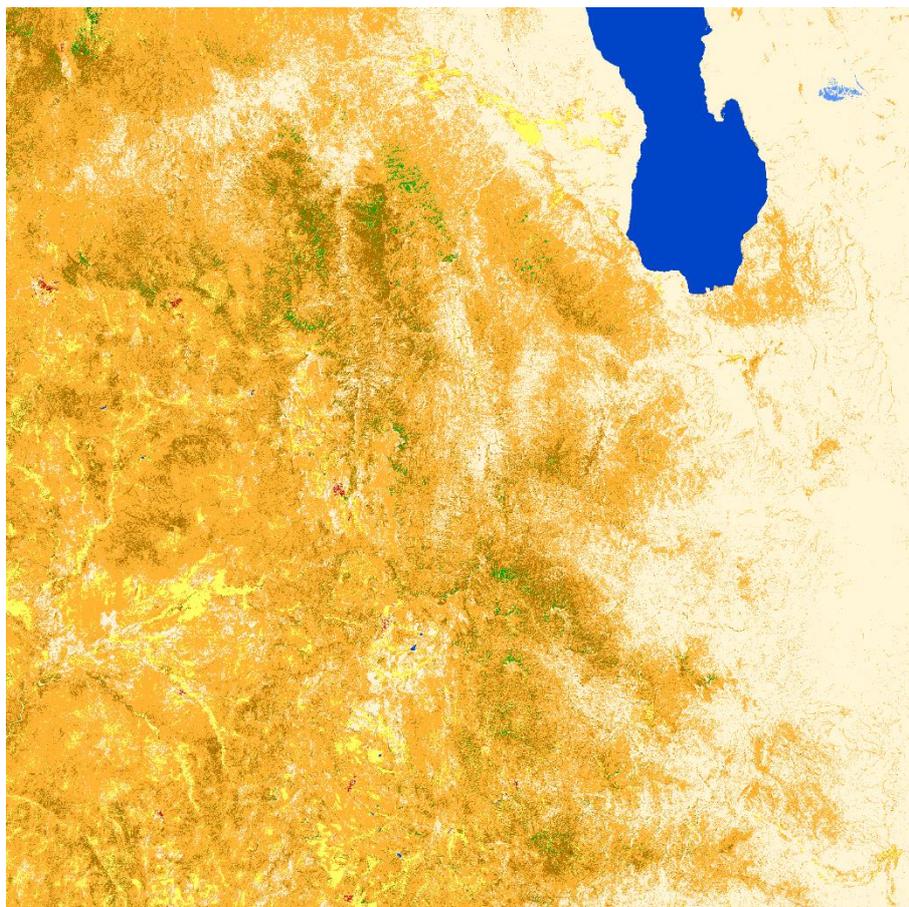


Figure 157. Static for tile 37PES.

Figure 157 shows the 2019 classification maps for African tile 37PES. The comparisons of optical and cascade maps for years 2015, 2010 and 2005 are shown in Figure 158, while Figure 159 shows the comparison for 2000, 1995 and 1990.

Also in this case, the effect of cascade is evident and causes a great increase in the consistency of the maps across the years. In particular, the overestimation of urban class in the mountainous area in the central-lower part of the tile which affected 2015, 2005 is greatly reduced. The same issue also affected the top-left part of the tile in 2000, 1995 and 1990 and, also in this case, the cascade module can mitigate the issue. The same is true for an overestimation of the shrubs and tree classes which partially affected the previous versions of all the historical maps.

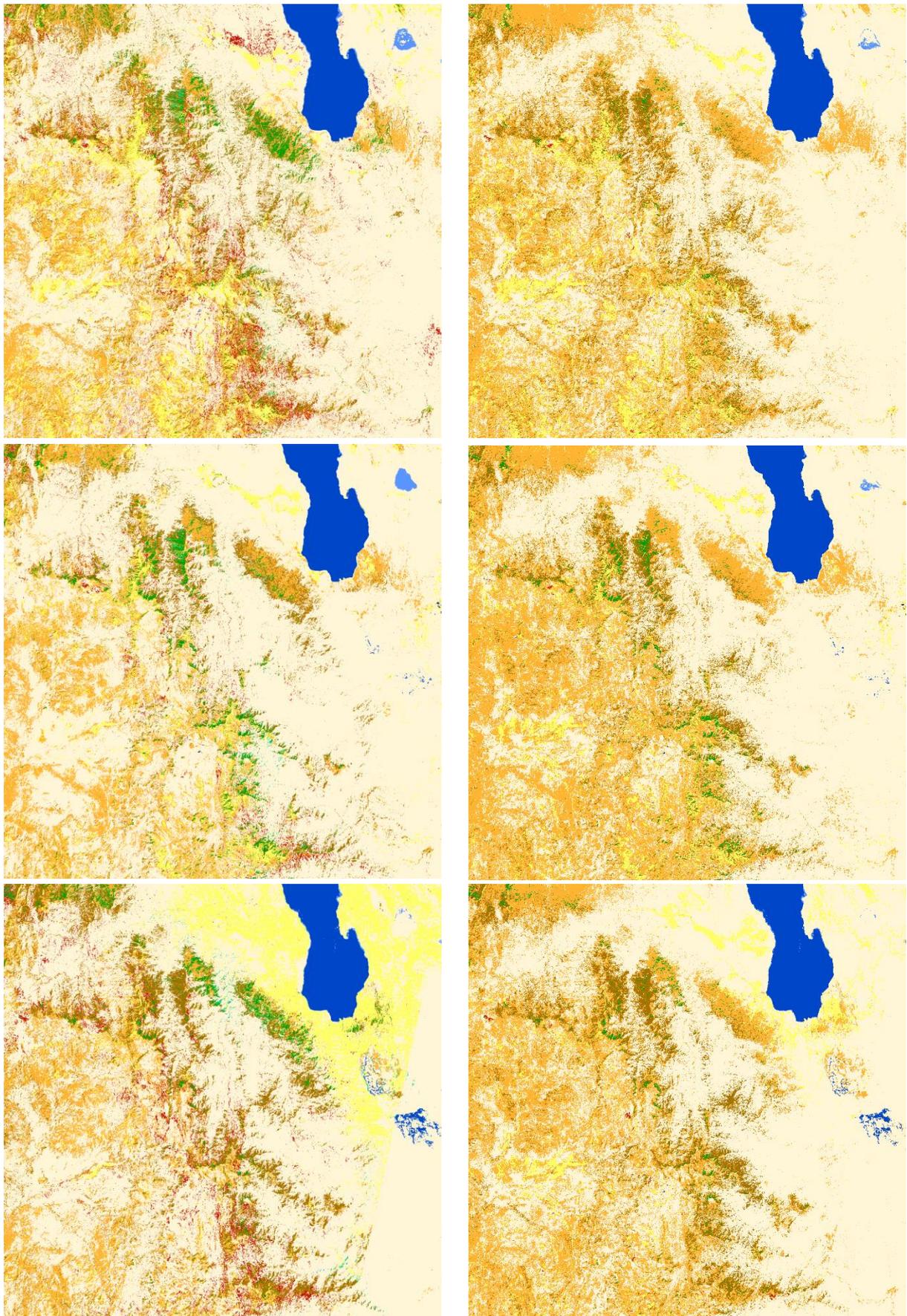


Figure 158. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 37PES. Year 2015 on the top row, 2010 in the centre and 2005 in the bottom.

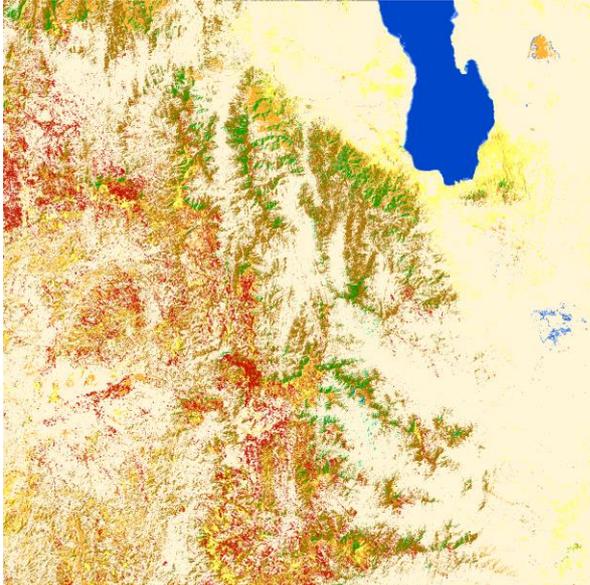
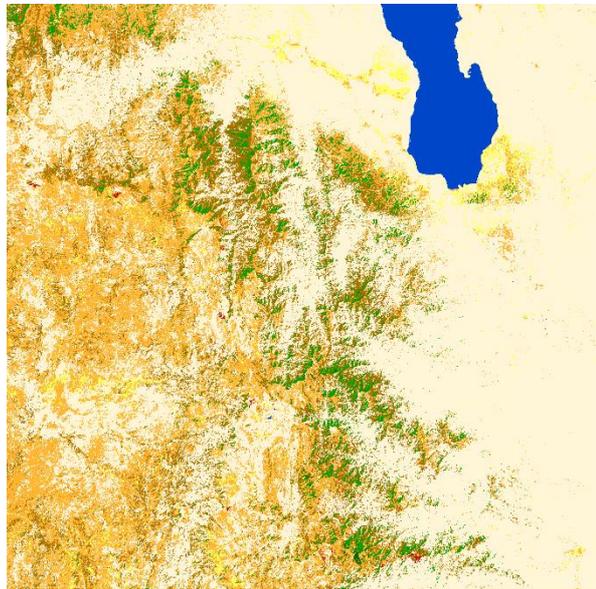
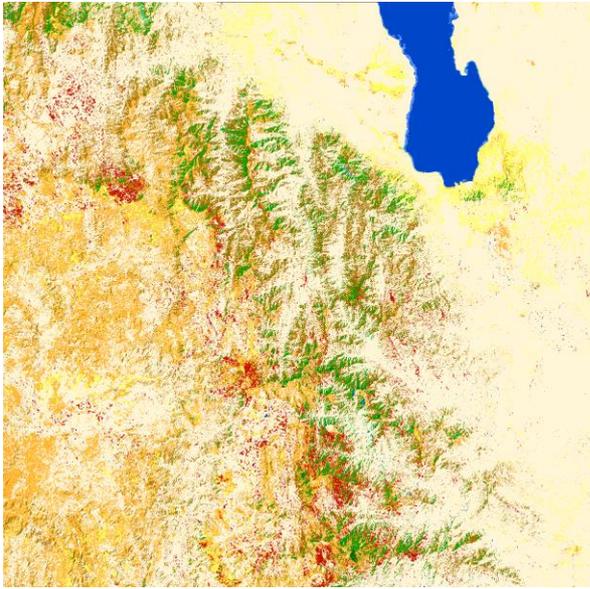
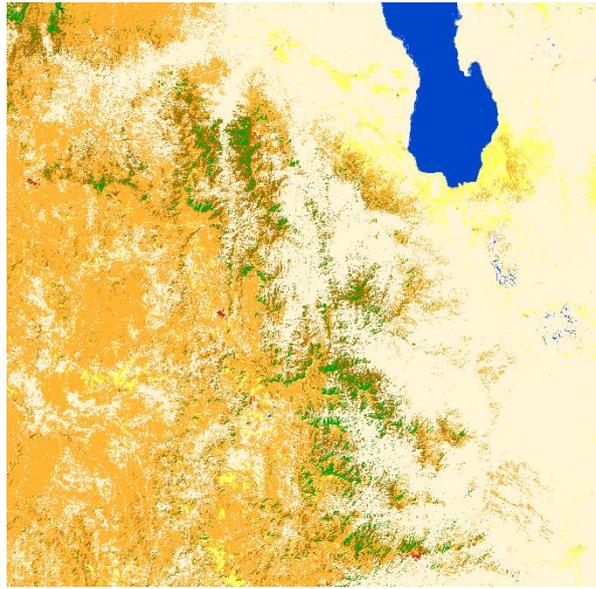
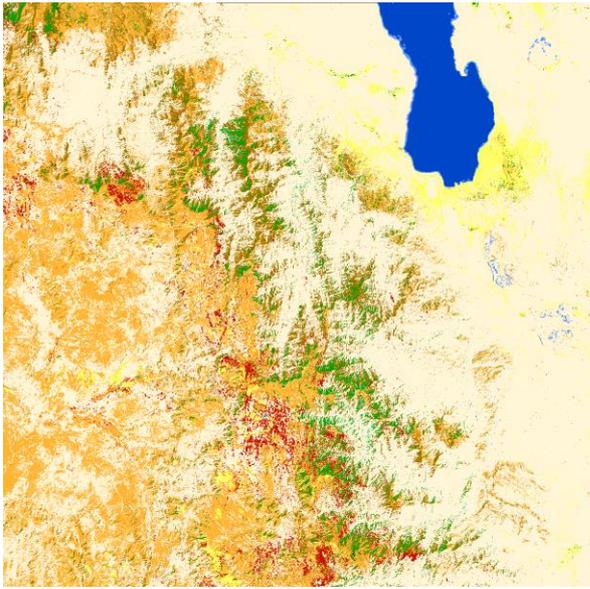


Figure 159. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 37PES. Year 2000 on the top row, 1995 in the centre and 1990 in the bottom.

5.3.1.3 Amazon Tile T21KUQ

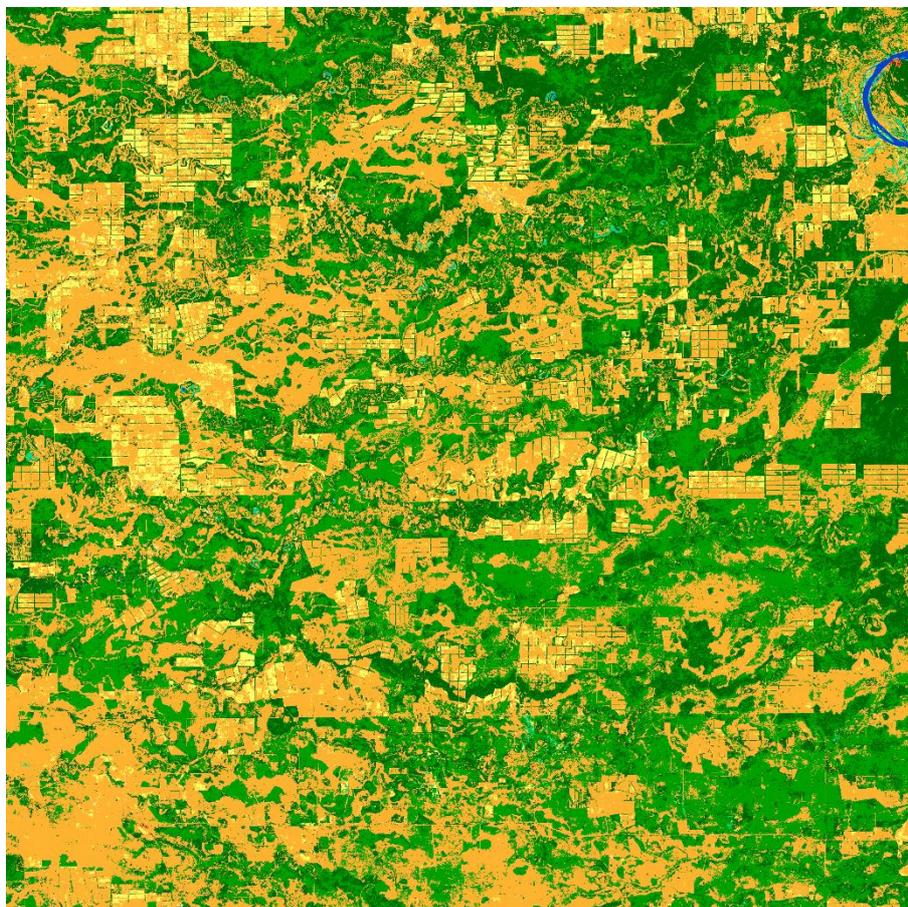
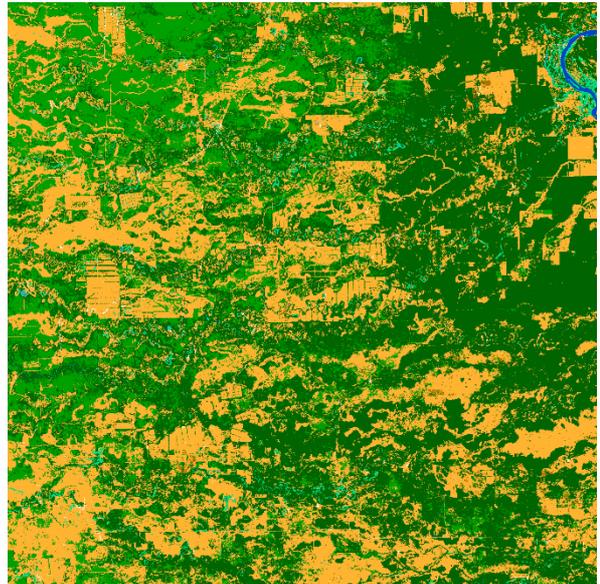
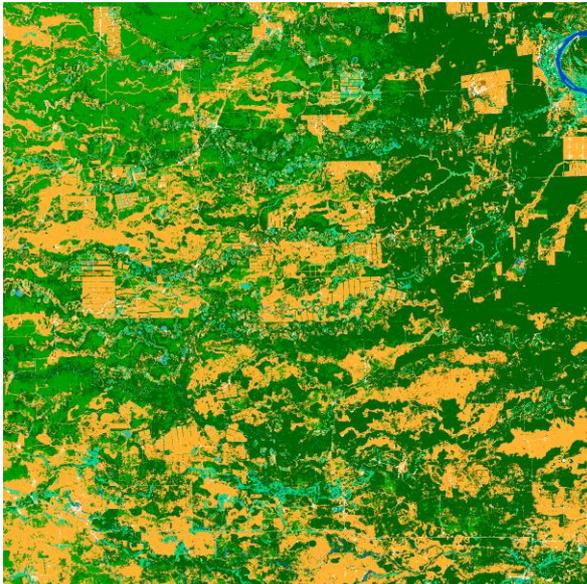
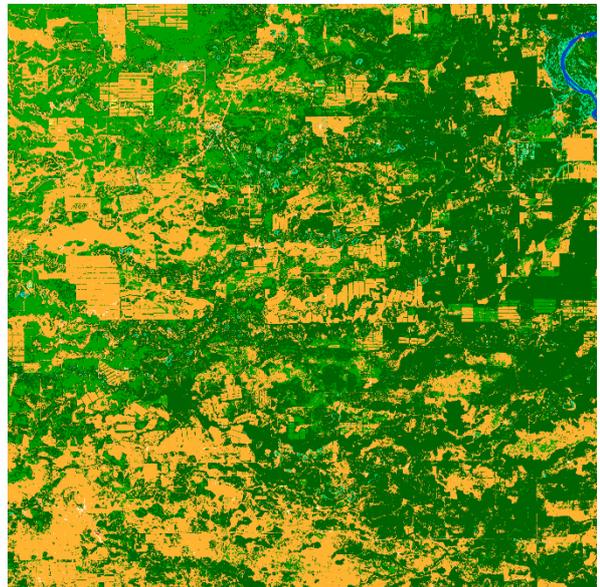
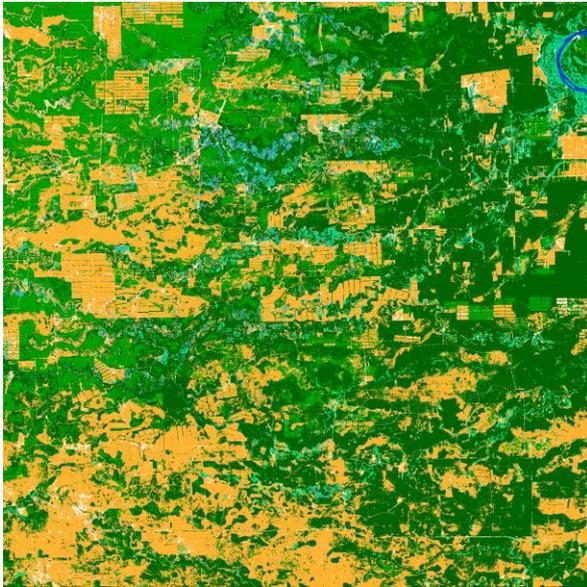
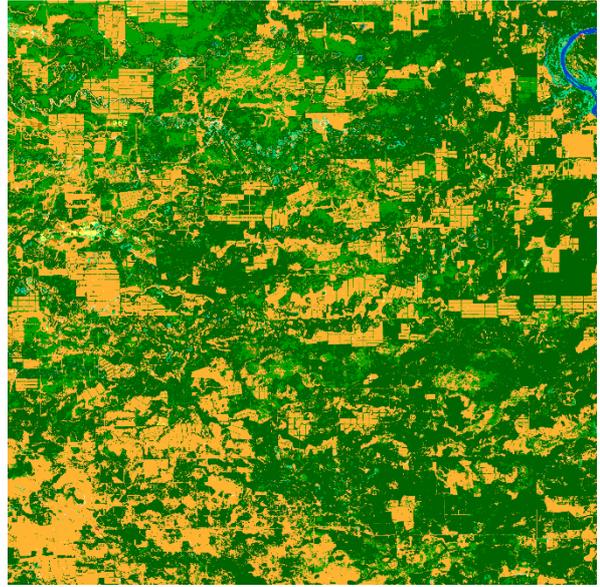
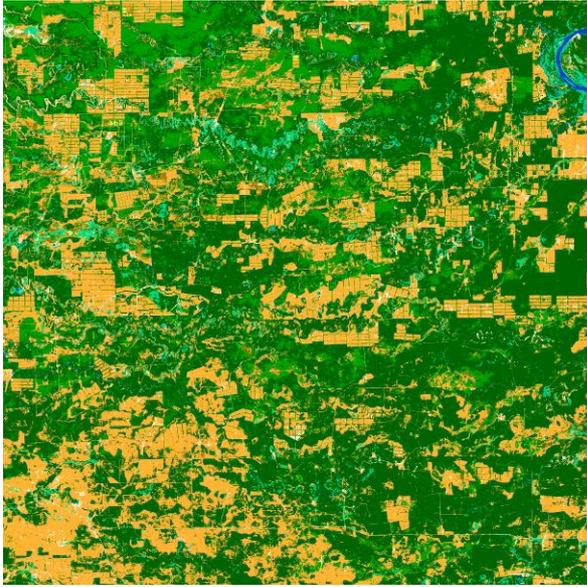


Figure 160. Static for tile 21KUQ.

The 2019 classification maps for Amazon tile 21KUQ is shown in Figure 160. The comparisons of optical and cascade maps for years 2015, 2010 and 2005 are shown in Figure 161, while Figure 162 shows the comparison for 2000, 1995 and 1990.

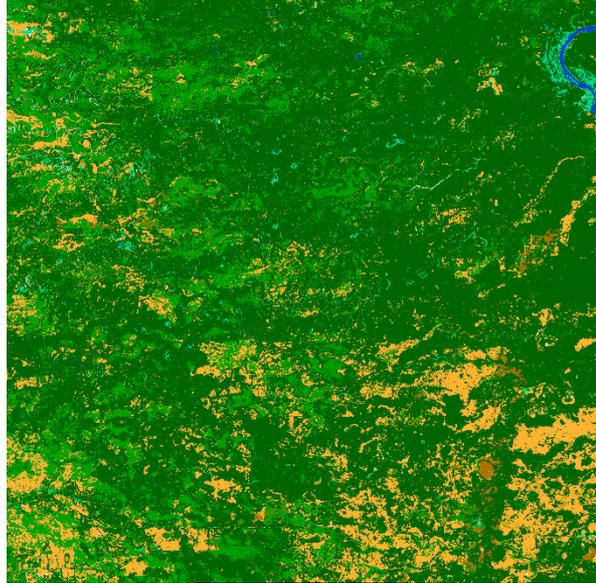
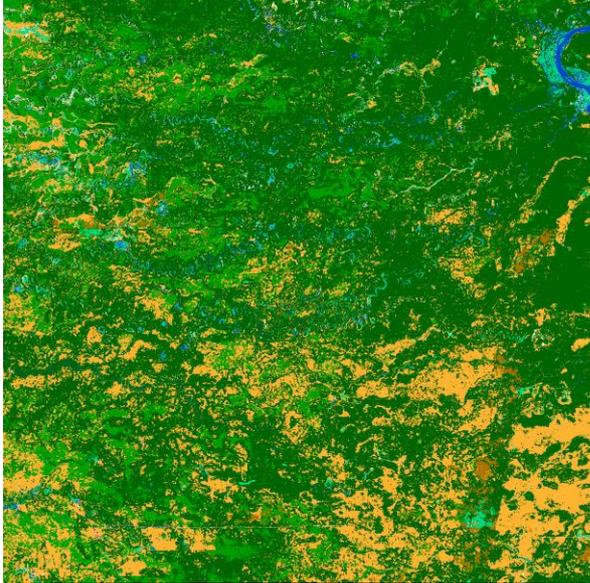
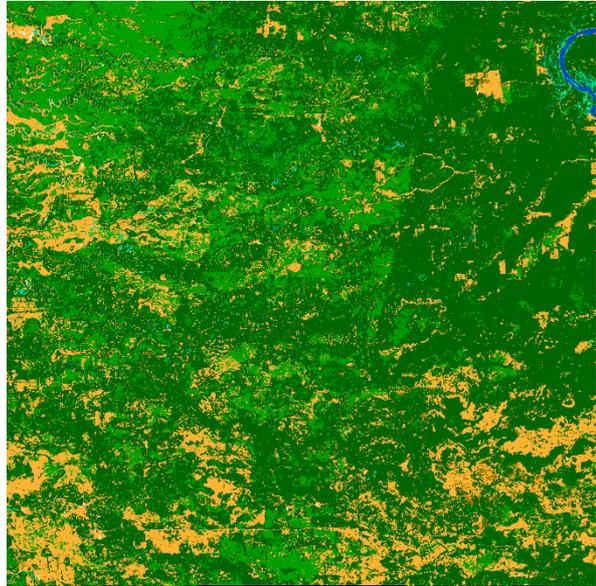
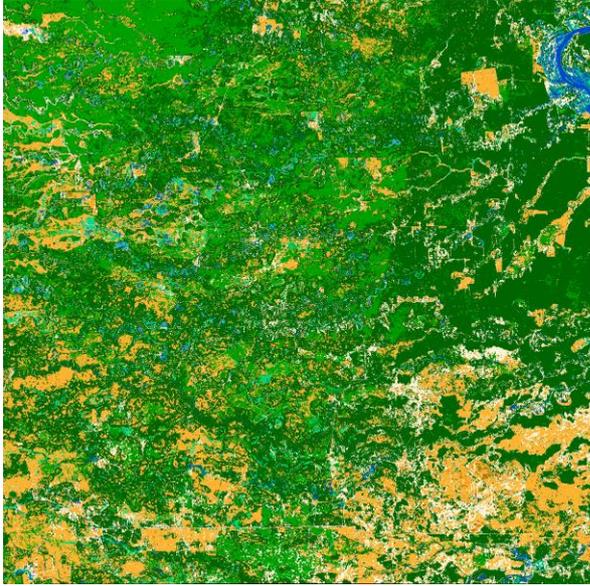
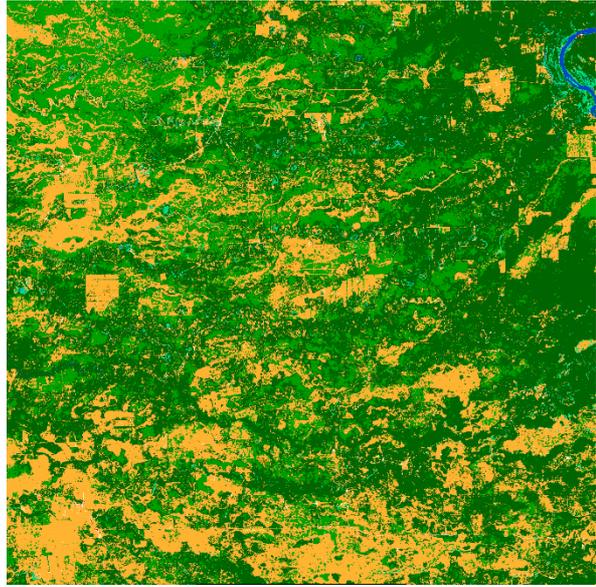
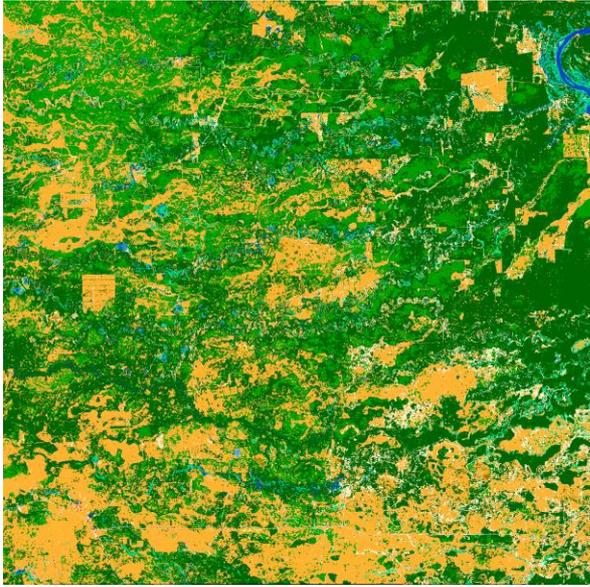
Also in this case, it is possible to confirm the ability of the cascade approach to correctly propagate the information and to correctly remove changes in the classification maps which are not reflected by actual changes in the landscape while preserving the actual ones. This can be noticed by looking at the evolution of the crops, which are increasingly present in the tile, moving from 1990 to 2019.

The use of cascade allows to solve some overestimation of watery and aquatic vegetation classes in 2015, 2010 and 2005 and also to reduce the overestimation of bare in 2000 and 1995 without impacting the correct evolution of the crops over time.



	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	159	

Figure 161. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 21KUQ. Year 2015 on the top row, 2010 in the centre and 2005 in the bottom.



	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	161	

Figure 162. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 21KUQ. Year 2000 on the top row, 1995 in the centre and 1990 in the bottom.

5.3.1.4 Amazon Tile T21KXT

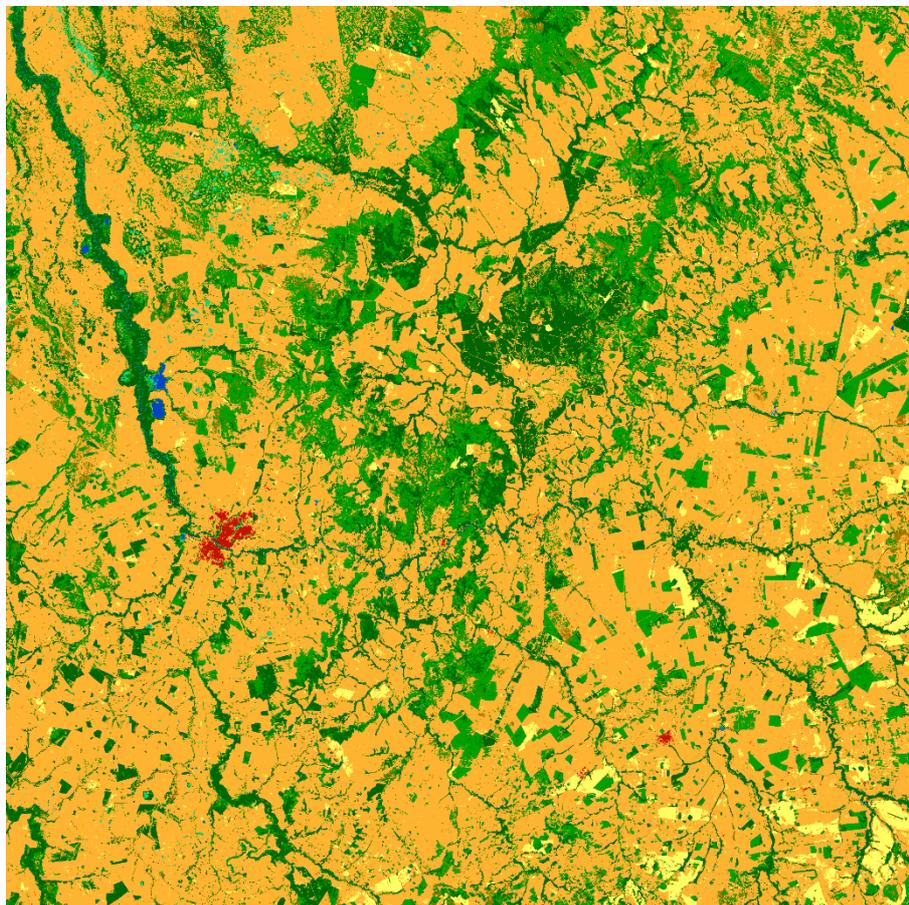
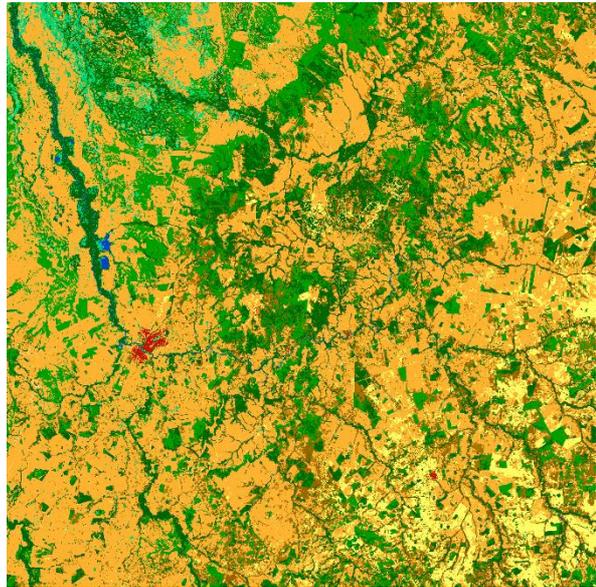
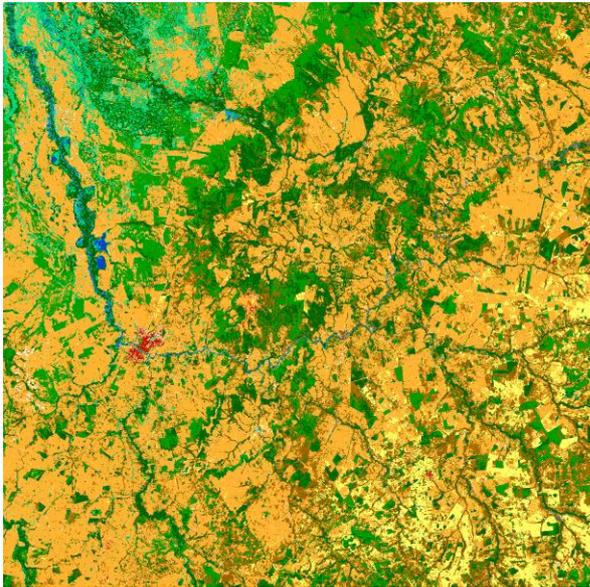
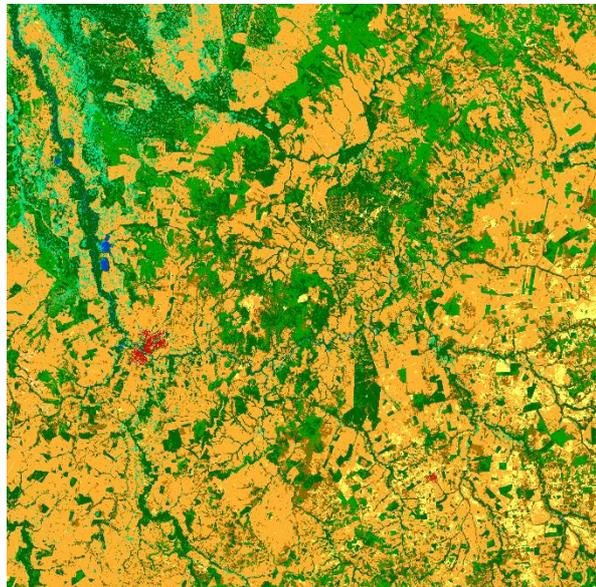
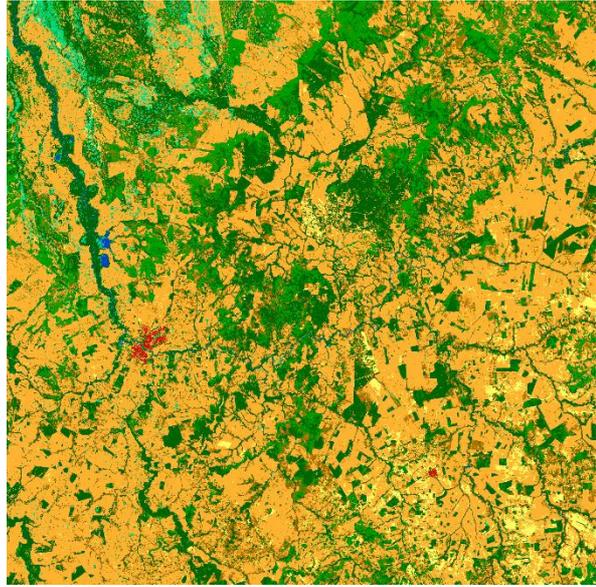
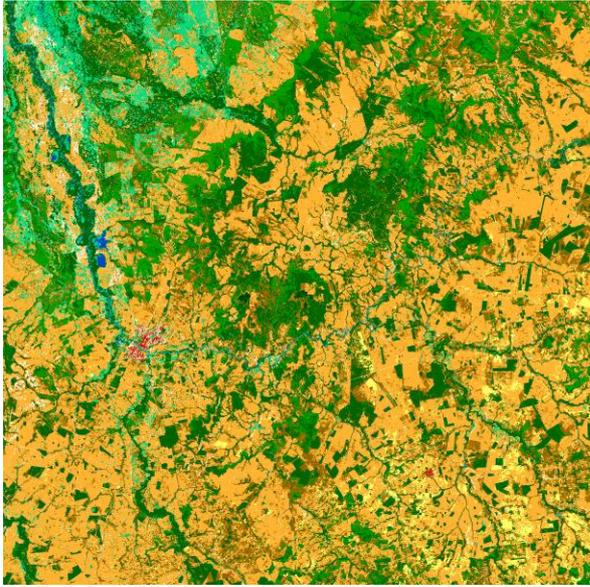


Figure 163. Static for tile 21KXT.

Figure 163 shows the 2019 classification maps for Amazon tile 21KXT. The comparisons of optical and cascade maps for years 2015, 2010 and 2005 are shown in Figure 164, while Figure 165 shows the comparison for 2000, 1995 and 1990.

With the use of the cascade approach the overestimation of aquatic vegetation class, especially relevant in 2005 and 1995, is mitigated. Moreover, the consistency of the discrimination of the city, present in the central-left area of the time, across the years is greatly improved.



	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	163	

Figure 164. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 21KXT. Year 2015 on the top row, 2010 in the centre and 2005 in the bottom.

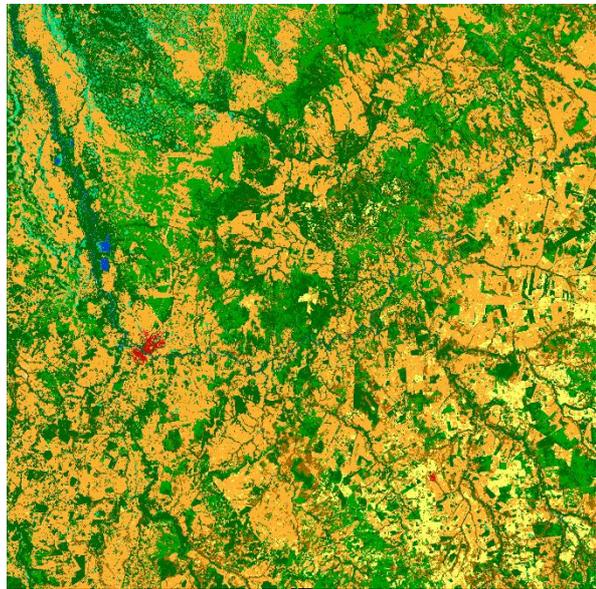
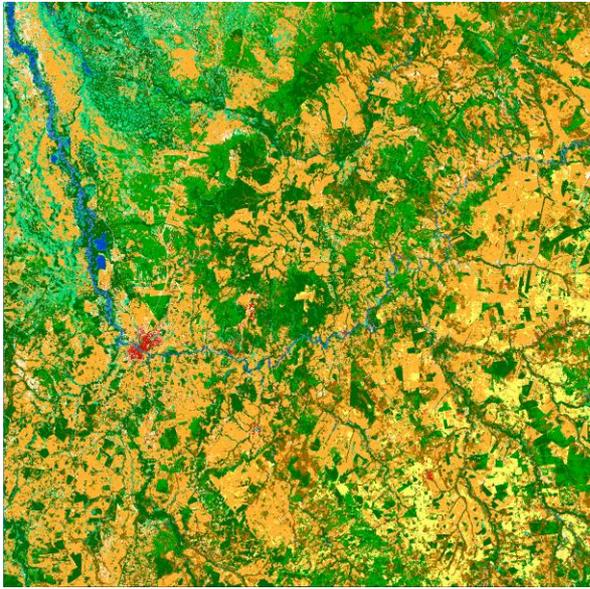
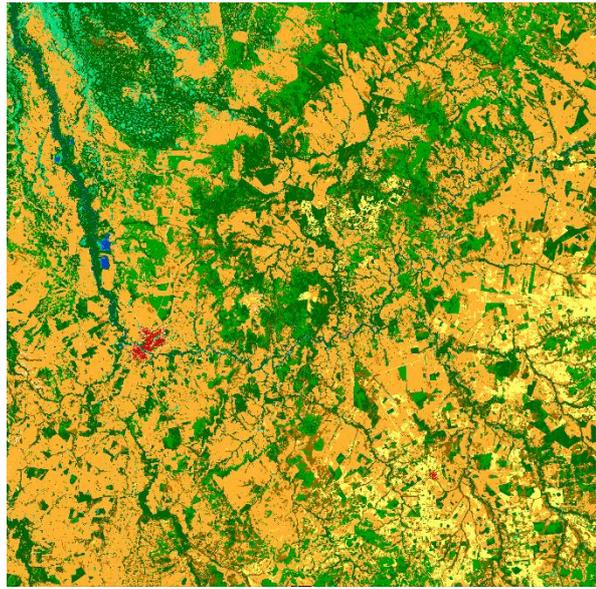
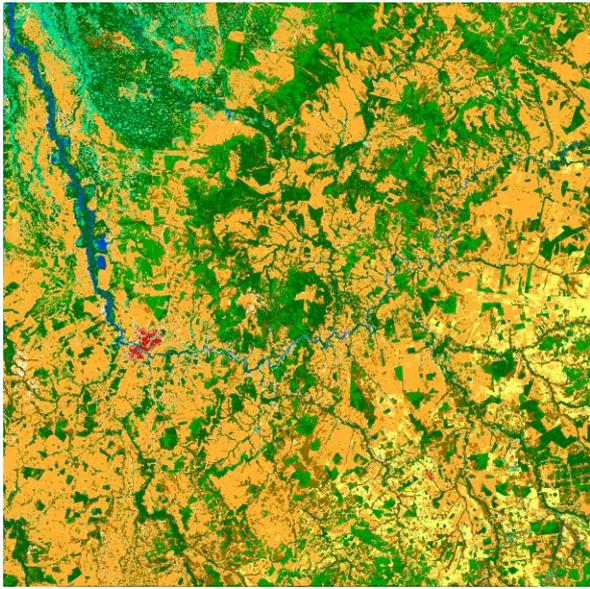


Figure 165. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 21KXT. Year 2000 on the top row, 1995 in the centre and 1990 in the bottom.

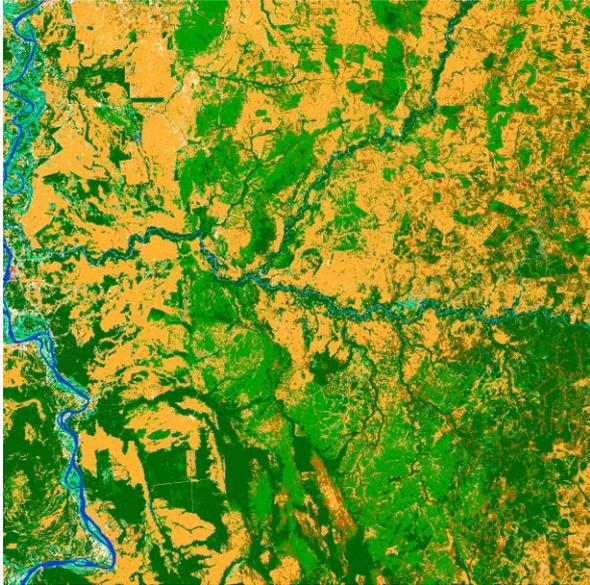
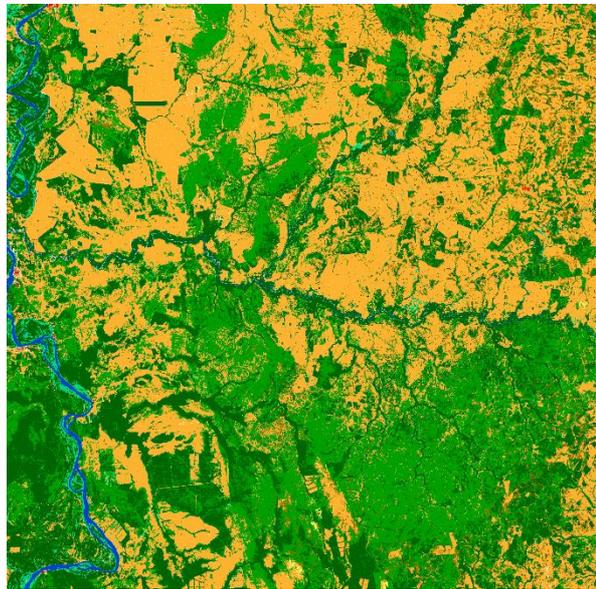
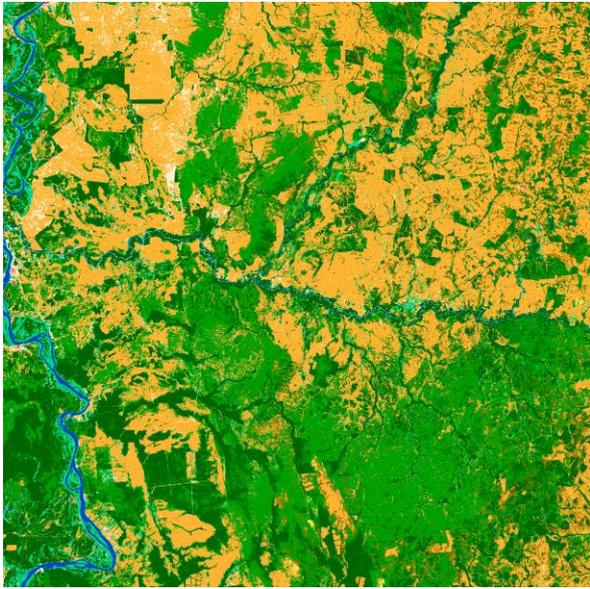
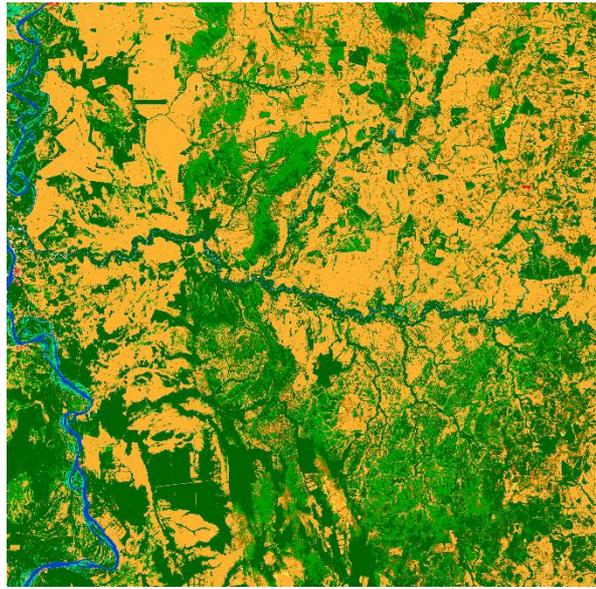
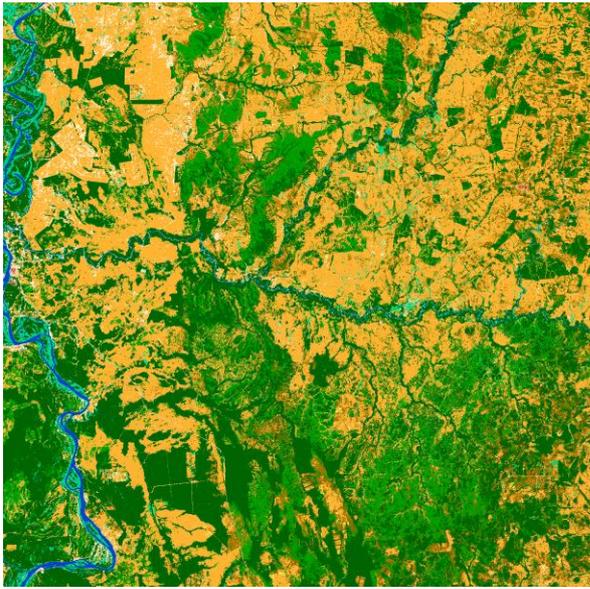
5.3.1.5 Amazon Tile T21KVR



Figure 166. Static for tile 21KVR.

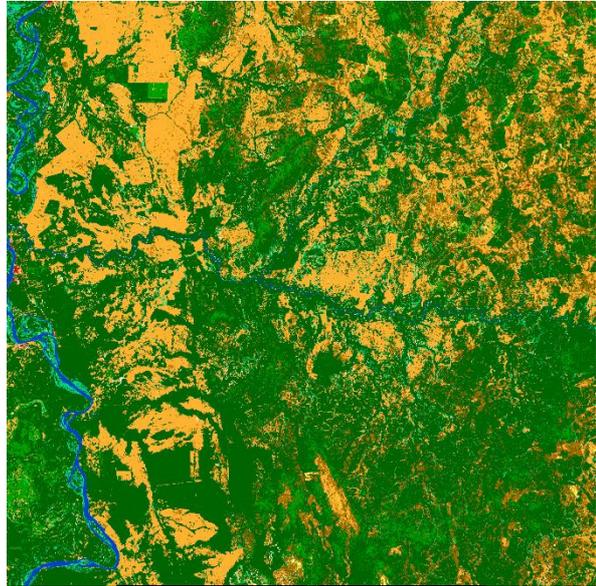
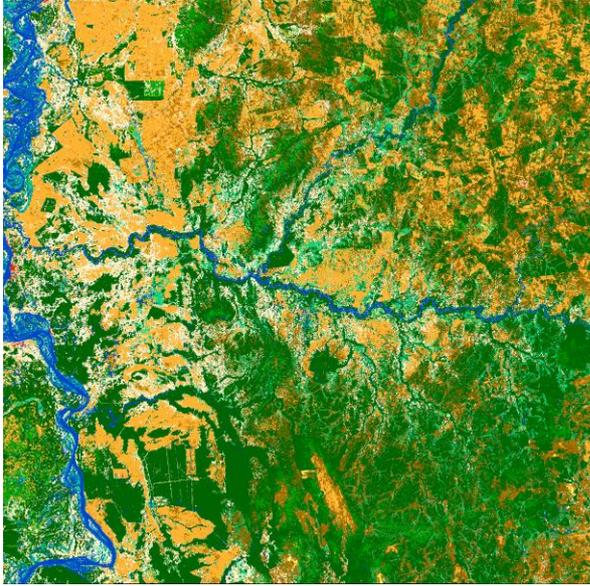
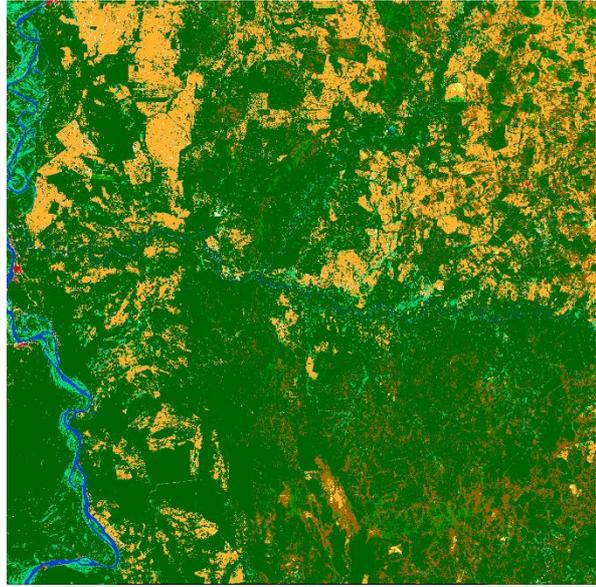
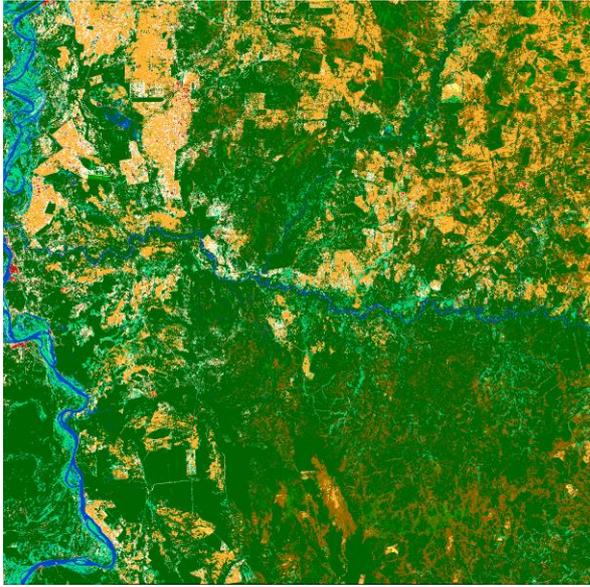
The 2019 classification maps for Amazon tile 21KVR is shown in Figure 166. The comparisons of optical and cascade maps for years 2015, 2010 and 2005 are shown in Figure 167, while Figure 168 shows the comparison for 2000, 1995 and 1990.

In this case, the impact of cascade is mainly visible in 2000 and 1995. In those years, the bare class was overestimated and, thanks to cascade, the final multitemporal classification is improved. The same holds for the overestimation of the vegetation aquatic class in 1990.



	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	167	

Figure 167. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 21KVR. Year 2015 on the top row, 2010 in the centre and 2005 in the bottom.



	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	169	

Figure 168. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 21KVR. Year 2000 on the top row, 1995 in the centre and 1990 in the bottom.

5.3.1.6 Siberia Tile T43WET

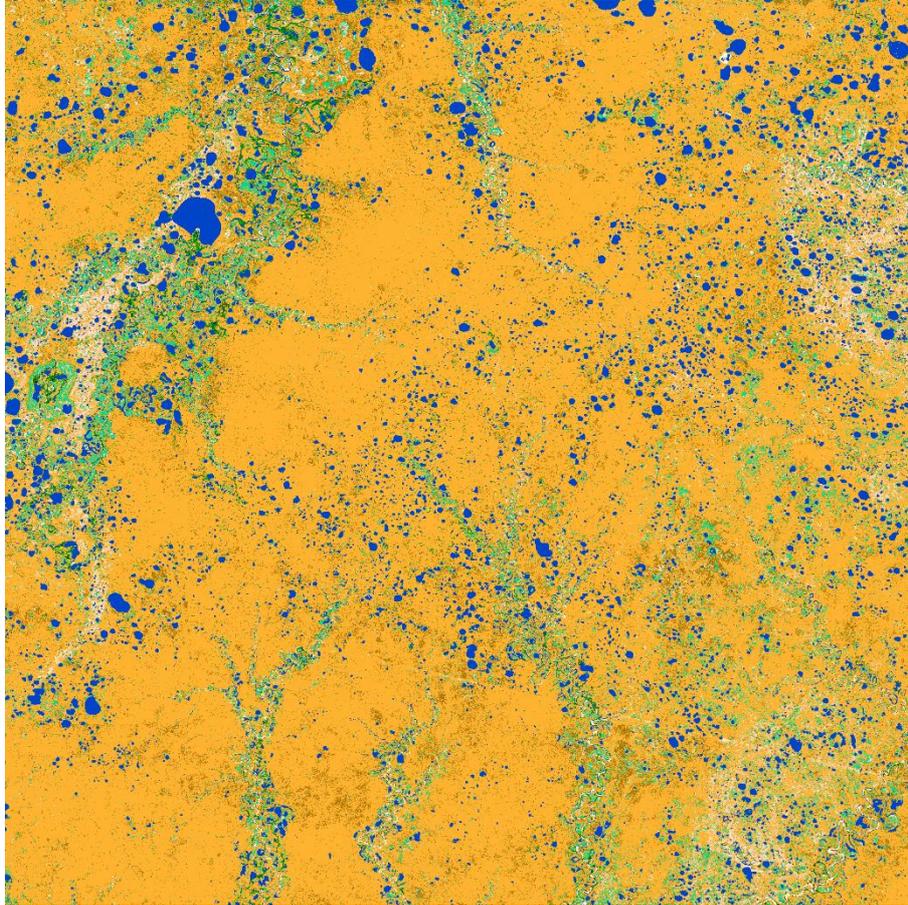
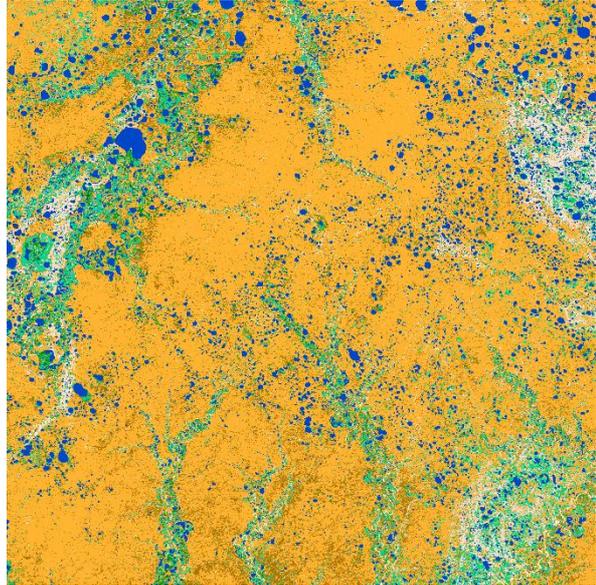
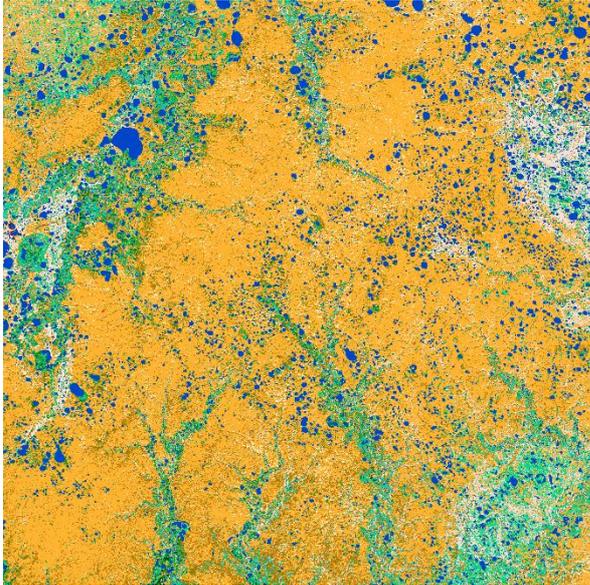
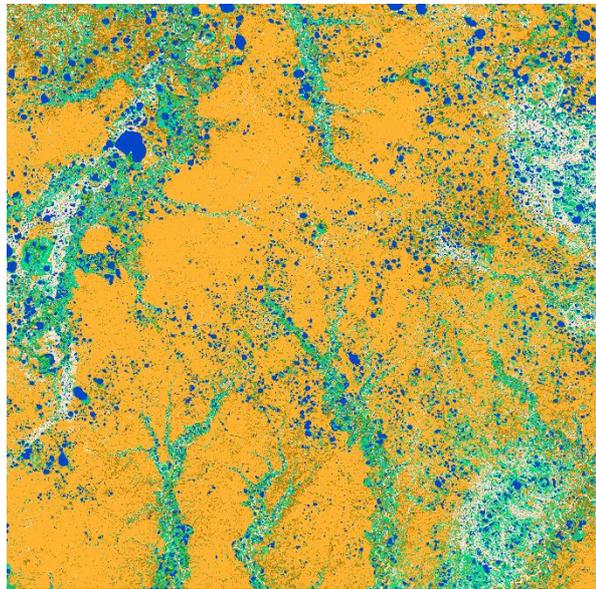
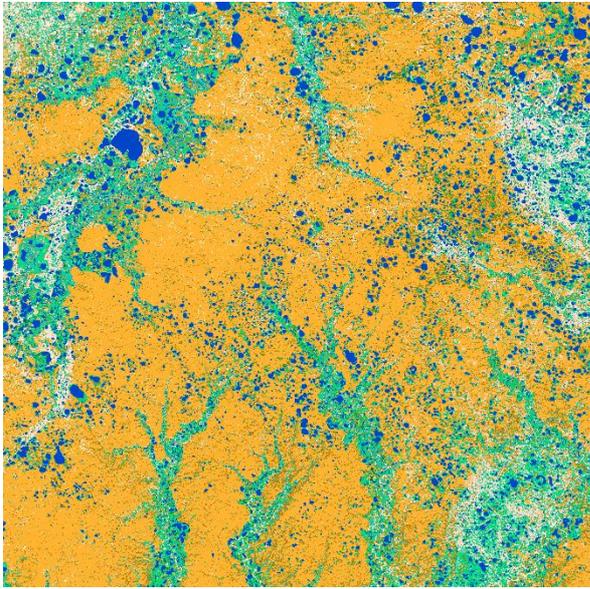
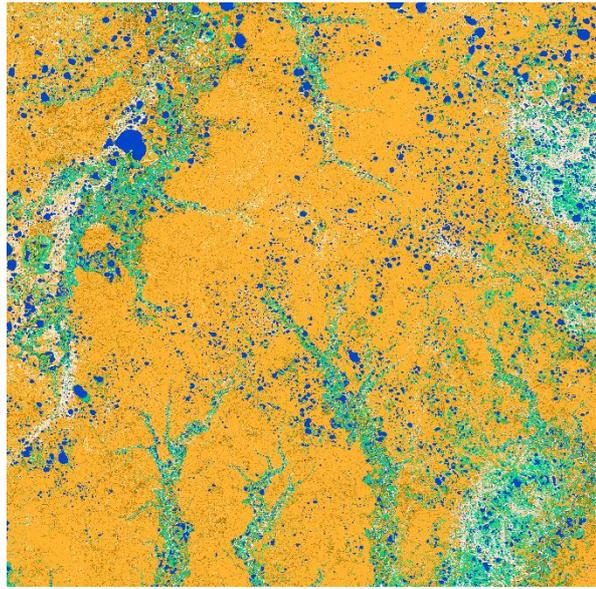
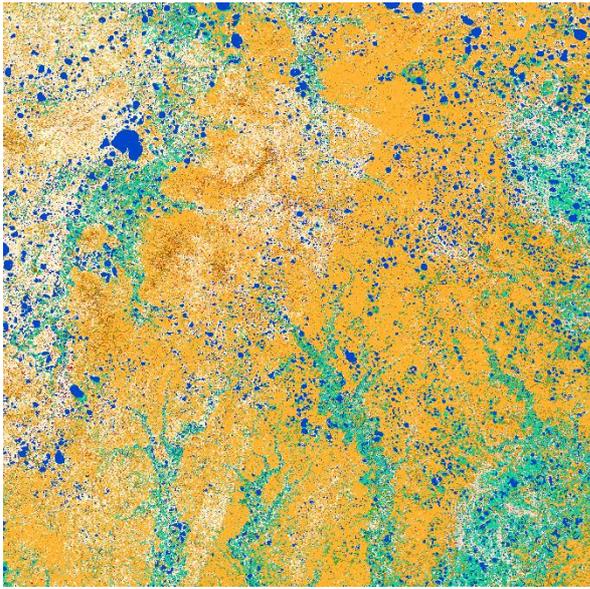


Figure 169. Static for tile 43WET.

Figure 169 shows the 2019 classification maps for Siberia tile 43WET. The comparisons of optical and cascade maps for years 2015, 2010 and 2005 are shown in Figure 170, while Figure 171 shows the comparison for 2000, 1995 and 1990.

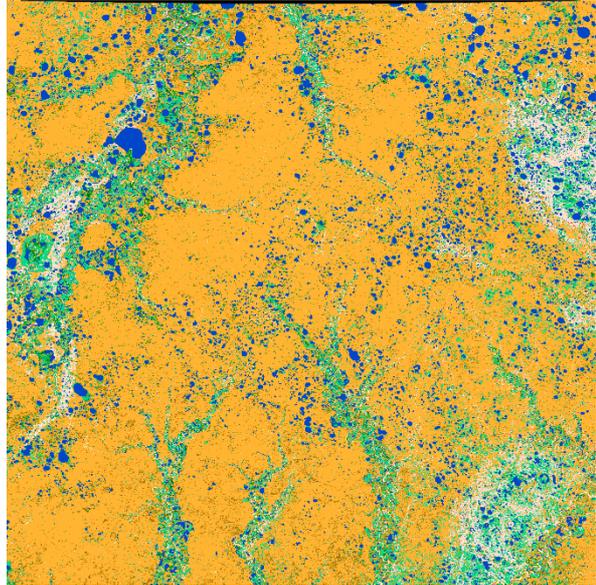
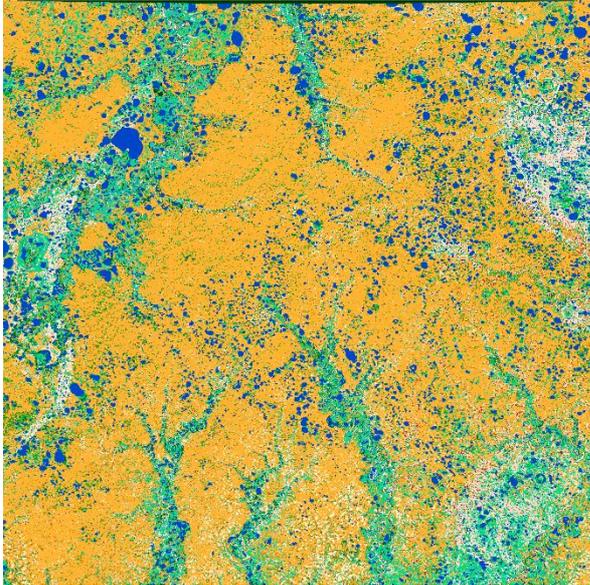
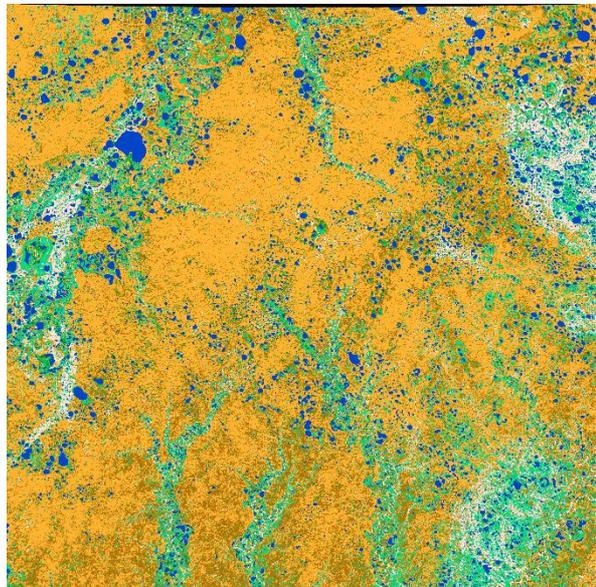
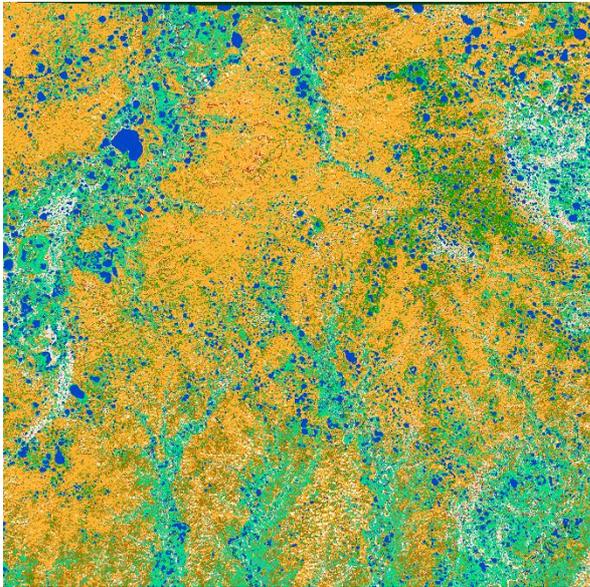
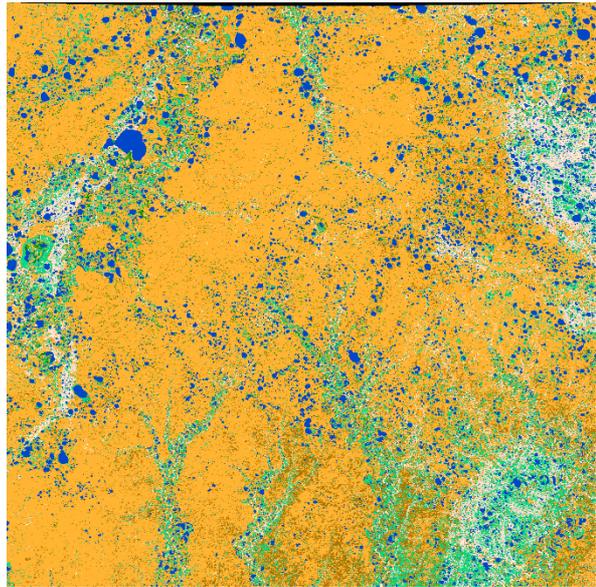
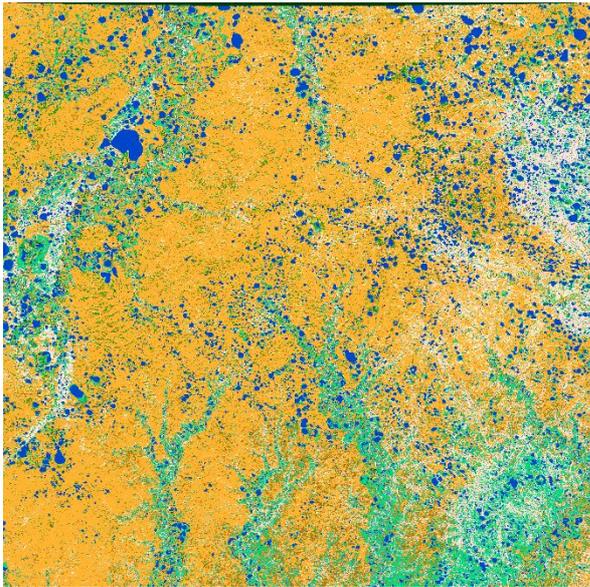
In this tile, the advantage of using the cascade approach can be noticed especially in 2015. In this case, in fact, a polygonal-shaped artifact, which is due to the input images, affects the left side of the tile. With cascade, the artifact is completely removed.

Moreover, the consistency of the maps for the other years is generally improved.



	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	171	

Figure 170. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 43WET. Year 2015 on the top row, 2010 in the centre and 2005 in the bottom.



	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	173	

Figure 171. Comparison of the optical map (left panels) with the final cascade map (right panels) for tile 43WET. Year 2000 on the top row, 1995 in the centre and 1990 in the bottom.

5.3.1.7 Conclusion

Summing up, the proposed cascade approach, integrated with the specific MRF formulation, is capable of recovering the vast majority of the errors and inconsistencies which are present in the historical maps. In fact, while 4-connected MRF mainly deals with isolated points or smaller areas – since it acts as neighbourhood spatial context –, the cascade model can also recover huge areas that were densely classified in the wrong way by propagating the information through the multitemporal series.

Moreover, it is important to consider one of the main aspects of the proposed implementation, which is the required computational time. In fact, each historical image at 30 meters resolution, corresponding to a Sentinel-2 tile, has an extension of 3660×3660 pixels. The whole data fusion process, integrated with the cascade method, which both integrates the information coming from the already computed posterior probabilities of year 2019 and applies the MRF model, requires an average of 30 to 40 s/granule to be completed.

5.3.2 Results of the spatial harmonization

The spatial harmonization module is needed in order to ensure consistency across the granule borders in the final mosaic. 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. Therefore, the final mosaic may present edge artifacts at the interface between the two granules.

In order to prevent such possible artifacts and obtain a spatially smooth mosaic, a spatial harmonization module is run on the pairs of neighbouring granules.

Figure 172 shows the area across two adjacent tiles before (top) and after (bottom) the spatial harmonization module. Looking at the areas highlighted by the ellipses it can be noticed that the edge among the two granules is completely removed by the spatial harmonization module.

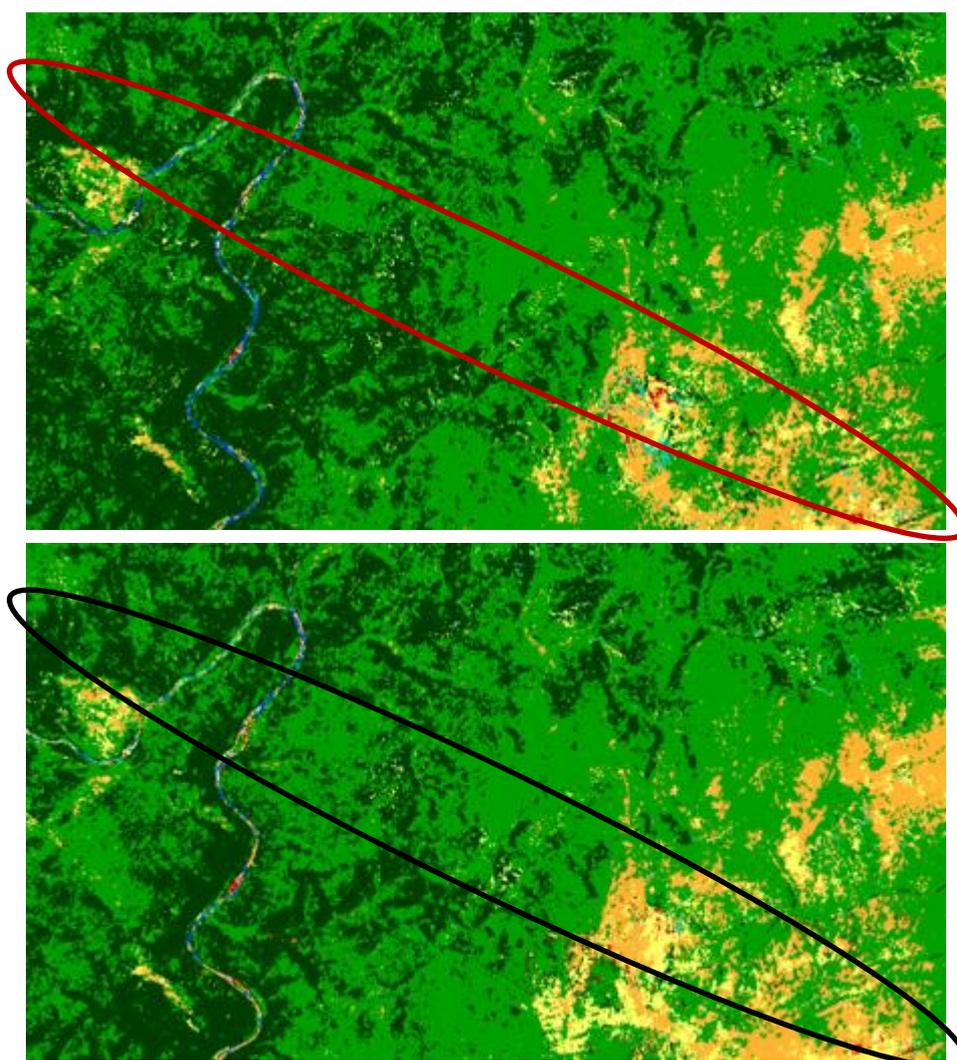


Figure 172 . Detail of the result obtained before (top) and after (bottom) the spatial harmonization module. The ellipses highlight the edges of the two tiles.

5.3.3 Results of the temporal harmonization

The temporal harmonization module is needed to check the temporal consistency of the land cover products and to update the maps based on the results of the change detection module. It first analyses the land-cover time series to check for inconsistencies and then applies the proper corrections. Then, the module works iteratively with the change detection module to remove the high-priority changes that are identified as wrong. The output of the temporal harmonization is the whole land-cover time series from 1990 to 2019 (every 5 years) at 30m resolution together with the change detection products. Figure 173 shows an example where the output of the multitemporal cascade model (left column) is compared with the output of the temporal harmonization (right column). We can see that in 2005 the output of the multitemporal cascade model is underestimating the urban area. In many cases, the cascade model is able to solve these problems. However, it is unable to characterize temporal trends that involve more than two years. Thus, it is possible that some inconsistency might remain. In the temporal harmonization, a backward moving window of three or more consecutive years is considered, thus allowing to analyse longer temporal trends than in the cascade model. Only clearly wrong patterns are corrected. Examples are the tree cover seasonality, which showed to be difficult to detect for historic classifiers; built-up, whose presence is not allowed to oscillate in time; and generally other highly oscillating behaviours that are not

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	175	

likely to happen. In the case of end-point years, i.e., 1990 and 2019, additional regularization was necessary to make them consistent with the whole time series (as the moving window can never be centred on these years). Moreover, 1990 represents the most difficult year to map with respect to data availability and quality, whereas 2019 is produced using data from a different sensor with a different resolution. In the example below, Figure 173 (h) shows the correct reconstruction of the built-up area, while Figure 173 (d) shows the removal of shrubs, which was present only in 1995 in the original cascade output (due to overestimation in the 1995 classification). Note that this is in addition to the use of the change detection results for the removal of wrong detected changes.

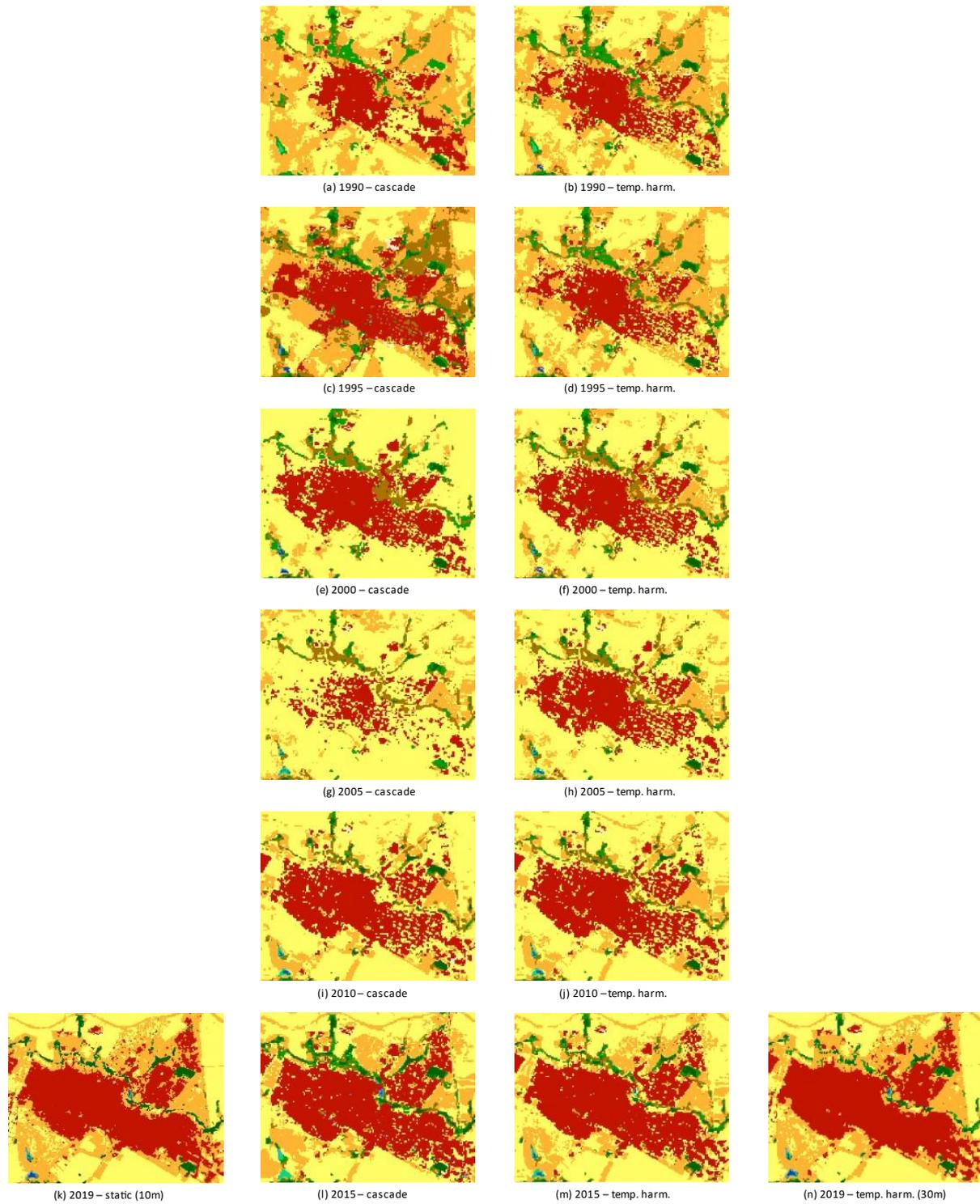


Figure 173. Example of result from the temporal harmonization module: (left) result from the multitemporal cascade model; (right) result from the temporal harmonization module.

5.4 Change Detection

5.4.1 Change Detection data processing

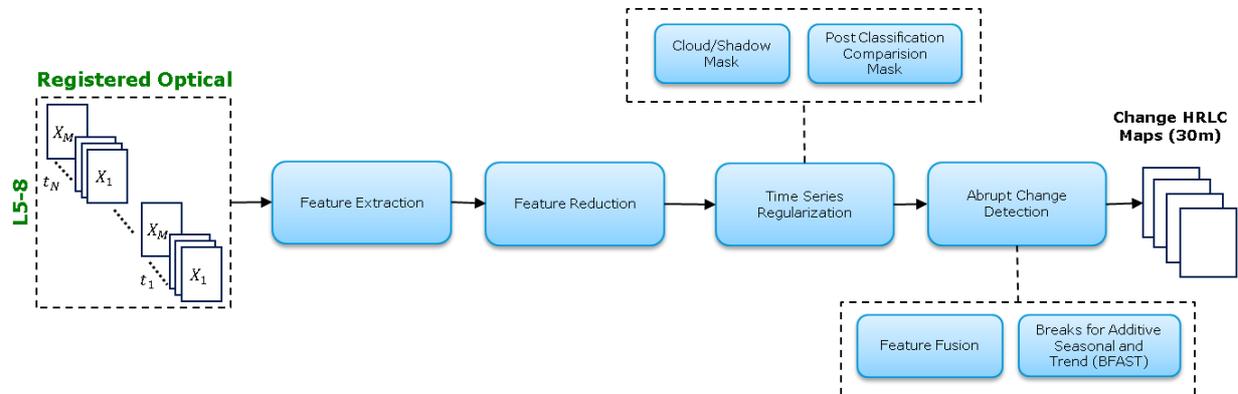


Figure 174. Change Detection processing chain.

Figure 174 depicts the change detection processing chain for producing the CD maps. First the features are extracted considering all possible couples of the available original sensor bands to compute a set of Normalized Difference Indices NDI_f . Then, a feature reduction strategy is implemented to preserve the most informative features [44]. A time series regularization is considered to generate a sequence of values, denser than the source signal that is uniformly sampled [45]. In this stage the cloud/shadow and Post Classification Comparison (PCC) masks are imposed to filter out cloudy pixels and the changed pixels, respectively. Finally, a binary CD method based on Breaks For Seasonal and Trend (BFAST) [46] is implemented to effectively discriminate changed and unchanged pixels. In order to reduce the computational burden, a feature fusion strategy is considered to fuse informative features. The final product is a four-channel image, one is related to the years 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 distance among the years in which the feature magnitude is calculated (some years does not have enough data for yearly change information thus more than a couple of years is considered in the calculation), and the last channel is the PCC map giving the information of the high and low priority pixels. The processing chain was applied consecutively to a five years Landsat 5, 7 and 8 SITS from 1990 to 2019 over three different areas being Amazonia, Africa and Siberia.

5.4.1.1 Time series reconstruction

Clouds, cloud shadows, and radiometric effects generate missing data and reduce images regularity in the SITS. The gaps in the SITS decrease the data quality and thus the capability of performing LCC detection. To have continuous and regular SITS in time (thus denser sequences than the source signal), a time series reconstruction technique is applied over extracted features from SITS. The choice of the appropriate interpolation approach is critical in this case since it has a direct impact on the CD accuracy. As a result, for the non-vegetation profiles the strategy is based on an upper-envelope piecewise cubic interpolation (PCHIP interpolation) and for the vegetation classes an adaptive non-parametric regression is performed that considers a General Regression Neural Network (GRNN). A non-parametric regression captures the temporal signatures trend reducing the profile complexity and arithmetic dependency. An analysis is performed using Upper-envelope PCHIP with General Regression Neural Network and another popular time series reconstruction methods (Univariate Spline) using a test area in Amazonia between 2013 and 2018 (see Figure 175) to evaluate the performance of each method. Looking at Figure 175 of GRNN and upper-envelope PCHIP interpolation have nearly same performance. However GRNN is more time consuming than PCHIP interpolation (these aspects have been analysed in details in [AD7]). By considering the performance of two reconstruction methods that and the computational burden of the two methods, the upper envelope PCHIP methodology was selected for all the classes.

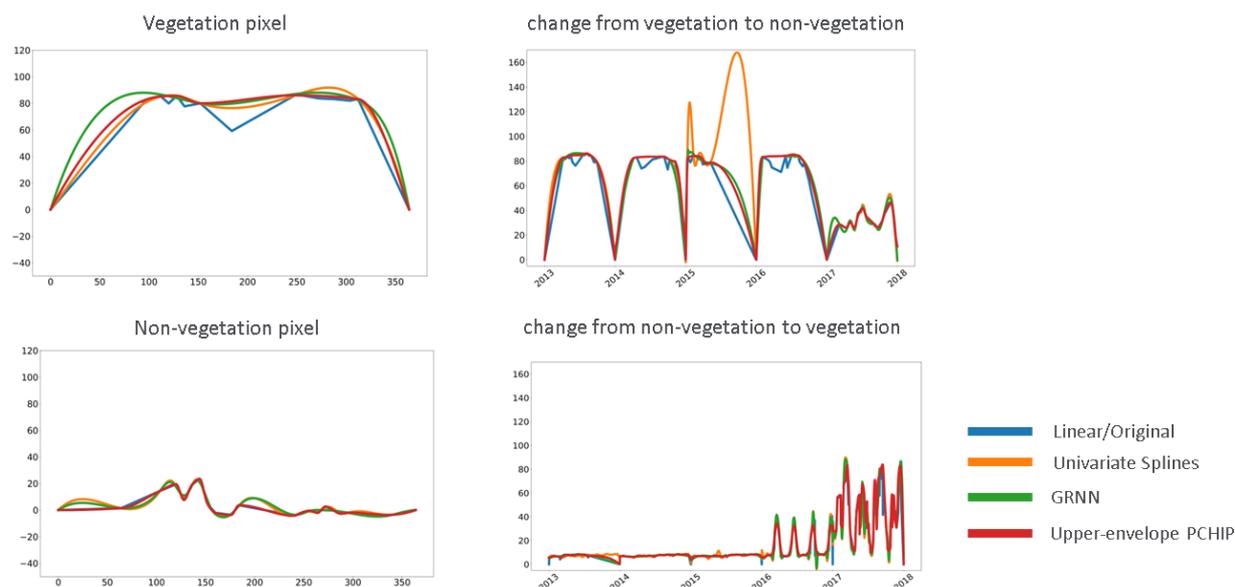


Figure 175. Comparison of different time series reconstruction methods.

5.4.1.2 Abrupt change detection

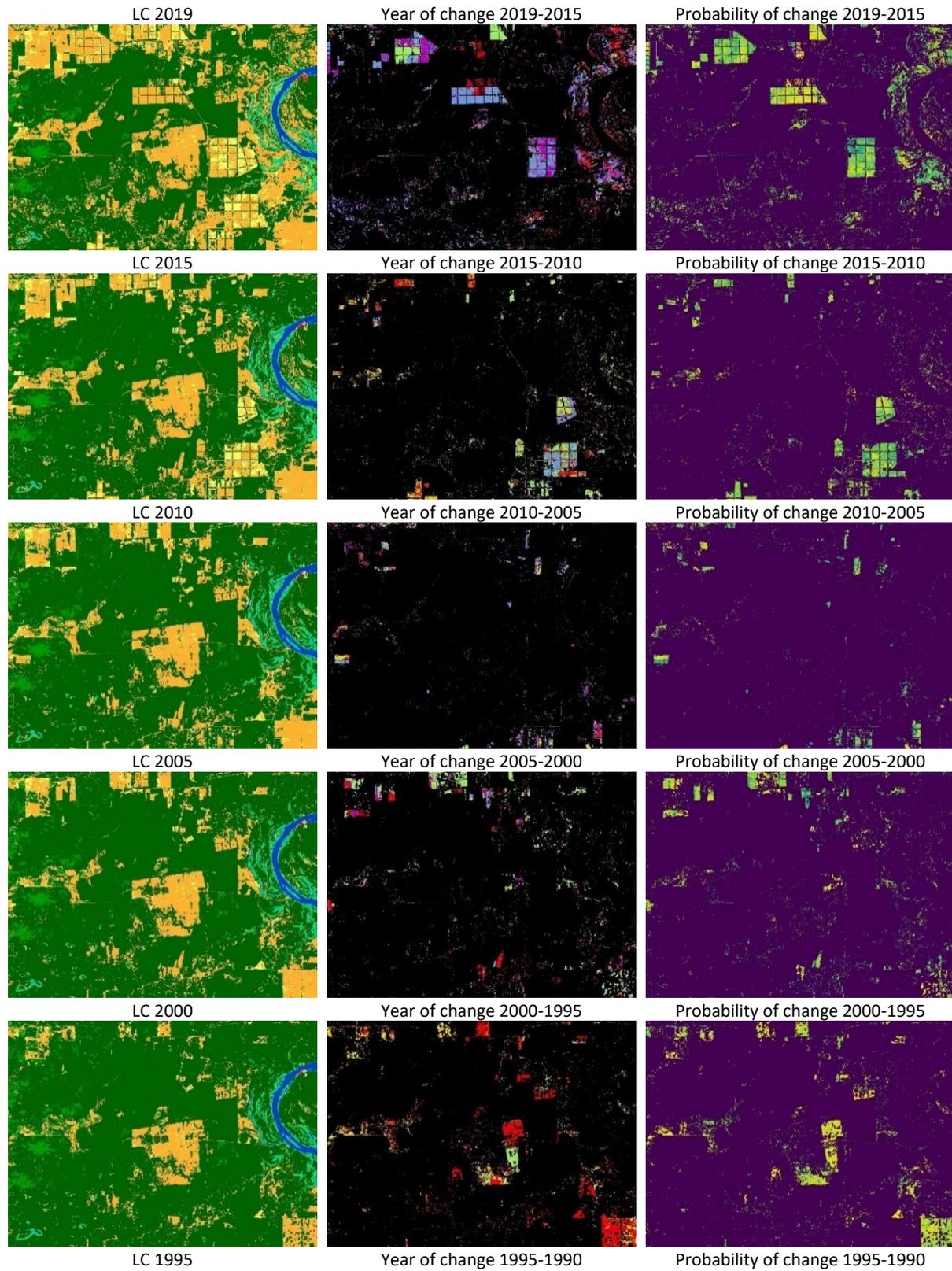
A limited number of methods have been developed in the literature that allow the analysis of long time series (with 16 days acquisitions) and can be considered as scalable to the spatial resolutions of the available sensors in this project. To detect the abrupt changes BFAST is selected that models both linear trends and seasonal variations without the need to select a reference period, set a threshold, or define a change trajectory that makes it suitable for detecting changes in different areas with different type of change. BFAST in an unsupervised manner recursively evaluates the possibility of every single time point being a breakpoint, and then chooses the most optimal set of breakpoints. BFAST is mainly exploited on medium spatial resolution SITS and in the most cases it is analysed for the Normalized Difference Vegetation Index (NDVI) or a few vegetation indices. But it is modified to deal with multi-feature data using HR SITS. As an output, BFAST provides the information regarding the time of the occurrence of the of the abrupt changes and the probability of change. Considering the huge amount of data to be processed by BFAST methodology for different areas (Amazonia, Africa and Siberia [AD4]), a feature fusion step based on the calculation of the feature magnitude is implemented to meliorate the computation burden of BFAST for different features extracted in feature extraction step.

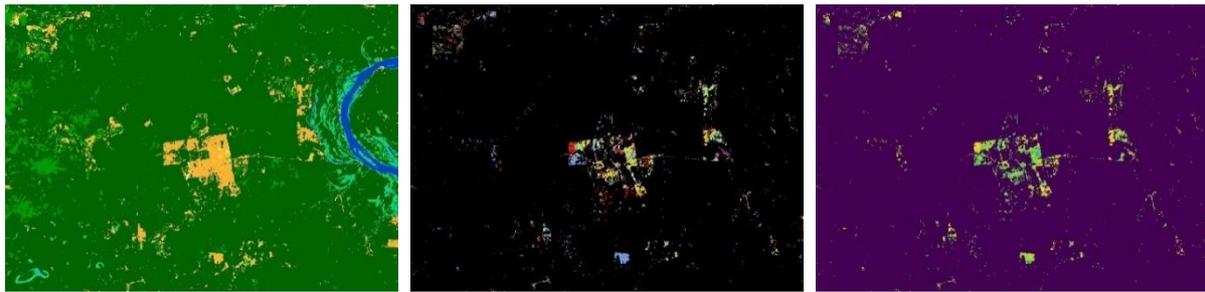
5.4.1.2.1 Qualitative and Quantitative evaluation

For a zoomed area in the tile defined below the LC maps for the years 1995 to 2019, the CD maps of the years 1990 to 2019 have visualised.

- Amazonia area for 21KUQ Sentinel-2 tile.

Moreover, a detailed quantitative analysis is provided in this section considering three processing areas (Amazonia, Africa, and Siberia) over the whole processing period (1990 to 2019). The first analysis has been carried out to determine the percentage of the pixels that have changed one to six times during 30 years of analysis in three areas. It is worth mentioning that considering the available LC maps every five years, it is assumed that each pixel can change once every five years. As a result, the maximum number of changes for each pixel from 1990 to 2019 will be six times. Another analysis provides the percentage of high priority changes for each year, the percentage of low priority changes every five years, the percentage of the high priority changed pixels without enough data to report the year of change (NoData) and the total amount of changes every five years. Please refer to Table 8 for Amazon, Table 9 for Africa and Table 10 for Siberia.





CCI HRLC Legend	
	Tree Cover Evergreen Broadleaf
	Tree Cover Evergreen Needleleaf
	Tree Cover Deciduous Broadleaf
	Tree Cover Deciduous Needleleaf
	Shrub cover Evergreen
	Shrub cover Deciduous
	Grassland
	Cropland
	Woody Vegetation Aquatic
	Grassland Vegetation Aquatic
	Lichens and Mosses
	Bare areas
	Built-up
	Open Water Seasonal
	Open Water Permanent

Year of Change	
	1996-2001-2006-2011-2016
	1997-2002-2007-2017
	1998-2003-2008-2018
	1999-2004-2009-2019
	2000-2005-2010-2015

Probability of Change	
	0
	0.1
	0.2
	0.3
	0.4
	0.5
	0.6
	0.7
	0.8
	0.9
	1

Figure 176. Examples of historical LC maps (1st column) and year of change maps (2nd column) change probability (3rd column) of the period from 1990 to 2019 for an area in tile 21KUQ in Amazon.

Looking at Figure 177 the percentage of the pixels that have changed one (21.97%) and two times (4.89%) in Amazon are significantly higher than the other percentages that are 0.99%, 0.11%, 0.0085% and 0.004%, respectively for three, four, five and six times. The same happens for the other areas being Africa and Siberia.

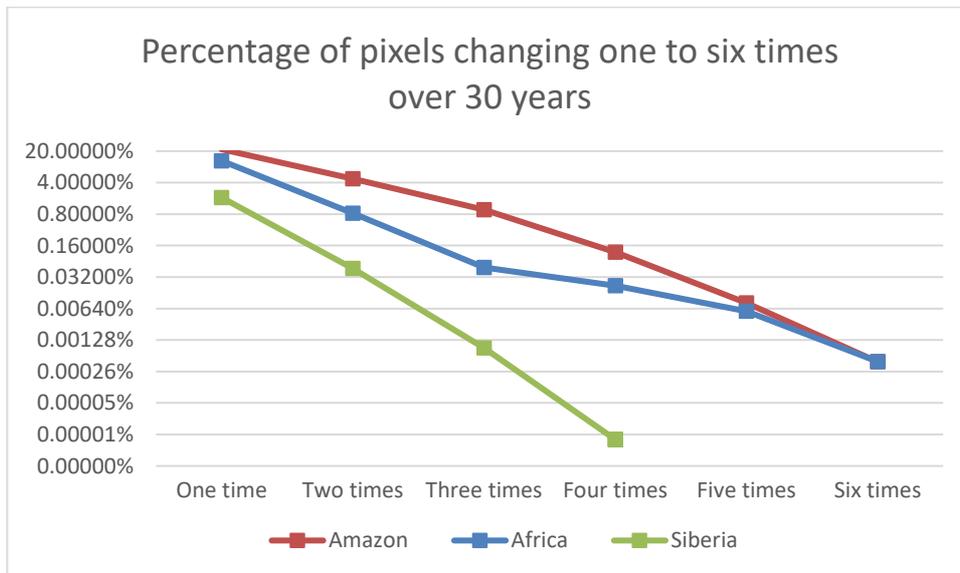


Figure 177. Percentage of the pixels changing one-six times in Amazon, Africa and Siberia.

Table 8. The percentage of high/low priority and overall changes in Amazon from 1990 to 2019

Year of Change	High Priority Changes	Low priority Changes	NoData	Overall Changes
1991	0.40%	1.71%	0.47%	4.62%
1992	0.32%			
1993	1.04%			
1994	0.36%			
1995	0.32%			
1996	0.29%	3.60%	0.71%	6.58%
1997	0.24%			
1998	0.20%			
1999	0.86%			
2000	0.69%			
2001	0.47%	4.01%	0.59%	7.40%
2002	0.63%			
2003	0.64%			
2004	0.68%			
2005	0.39%			
2006	0.29%	3.48%	0.61%	5.77%
2007	0.43%			
2008	0.45%			
2009	0.30%			
2010	0.21%			
2011	0.36%	3.08%	0.54%	5.57%
2012	0.06%			
2013	0.57%			
2014	0.53%			
2015	0.43%			
2016	0.60%	2.17%	0.57%	5.31%
2017	0.58%			
2018	0.75%			
2019	0.65%			

Table 9. The percentage of high/low priority and overall changes in Africa from 1990 to 2019

Year of Change	High Priority Changes	Low priority Changes	NoData	Overall Changes
1991	0.01%	0.18%	0.17%	0.50%
1992	0.01%			
1993	0.02%			
1994	0.06%			
1995	0.04%			
1996	0.07%	0.20%	0.97%	1.44%
1997	0.05%			
1998	0.04%			
1999	0.06%			
2000	0.05%			
2001	0.19%	0.81%	1.60%	2.87%
2002	0.22%			
2003	0.04%			
2004	0.01%			
2005	0.01%			
2006	0.01%	0.69%	3.32%	4.12%
2007	0.01%			
2008	0.01%			
2009	0.02%			
2010	0.05%			
2011	0.03%	0.71%	2.78%	4.14%
2012	0.02%			
2013	0.10%			
2014	0.24%			
2015	0.27%			
2016	0.02%	0.94%	0.03%	1.07%
2017	0.02%			
2018	0.02%			
2019	0.03%			

Table 10. The percentage of high/low priority and overall changes in Siberia from 1990 to 2019

Year of Change	High Priority Changes	Low priority Changes	NoData	Overall Changes
1991	0.007%	0.31%	0.11%	0.46%
1992	0.008%			
1993	0.002%			
1994	0.003%			
1995	0.019%			
1996	0.001%	0.22%	0.04%	0.27%
1997	0.001%			
1998	0.003%			
1999	0.000%			
2000	0.005%			
2001	0.001%	0.40%	0.11%	0.51%
2002	0.002%			
2003	0.000%			
2004	0.000%			
2005	0.000%			
2006	0.000%	0.23%	0.04%	0.28%
2007	0.001%			
2008	0.000%			
2009	0.007%			
2010	0.001%			
2011	0.000%	0.30%	0.03%	0.35%
2012	0.000%			
2013	0.006%			
2014	0.009%			
2015	0.006%			
2016	0.002%	0.07%	0.00%	0.08%
2017	0.001%			
2018	0.001%			
2019	0.001%			

To have a better overview of the quantity and temporal behaviour of the changes in different areas (Amazon, Africa and Siberia) the additional figures for the high priority yearly changes and 5-year changes have been provided. Figure 178 and Figure 179 visualise the temporal behaviour of the percentage of high priority changes and 5-year changes for the Amazon area. As it is clear, the amount of annual high priority change has a significant fluctuation during 1990 to 2019 and has a peak of 1.04% in 1992, with an increasing trend in the last 10 years. Considering Figure 179, the overall 5-year changes has a rising trend from 1991-1995 to 2001-2005 but we record a decreasing trend till 2016-2019. This is mainly because of the decrease of the amount of low priority changes, that compensate for the increasing trend of the high priority ones.

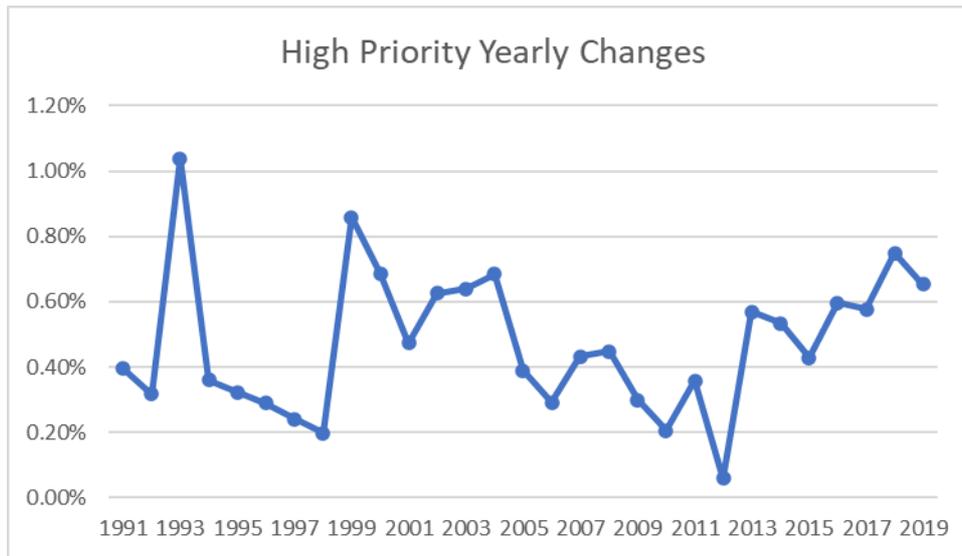


Figure 178. Percentage of high priority change of Amazon.

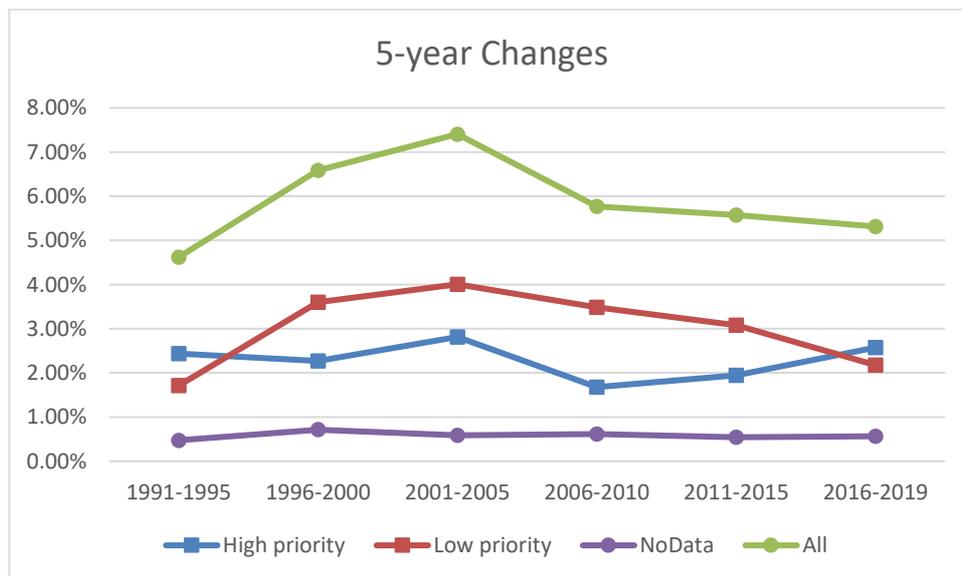


Figure 179. Percentage of 5 year changes in Amazon.

Considering Figure 180, the percentage of yearly high priority changes of Africa has two peaks of about 0.22% and 0.27% in the years 2002 and 2015, respectively. Figure 181 shows that the 5-year low priority changes increase dramatically from 1991-1995 to 2006-2010 while the high priority changes without reported the year of change (NoData) has a peak in 2006-2010 and decreases in 2016-2019. As it is shown, because of the lack of

data to reveal the year of change, a majority of the high priority changes detected by post classification comparison mask (pcc) has been represented as NoData.

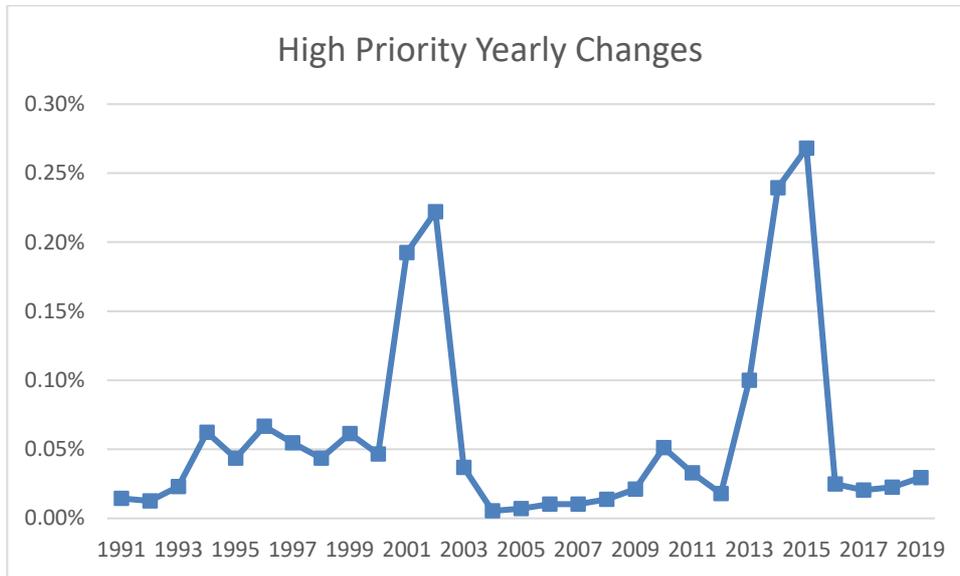


Figure 180. Percentage of High priority yearly changes of Africa.

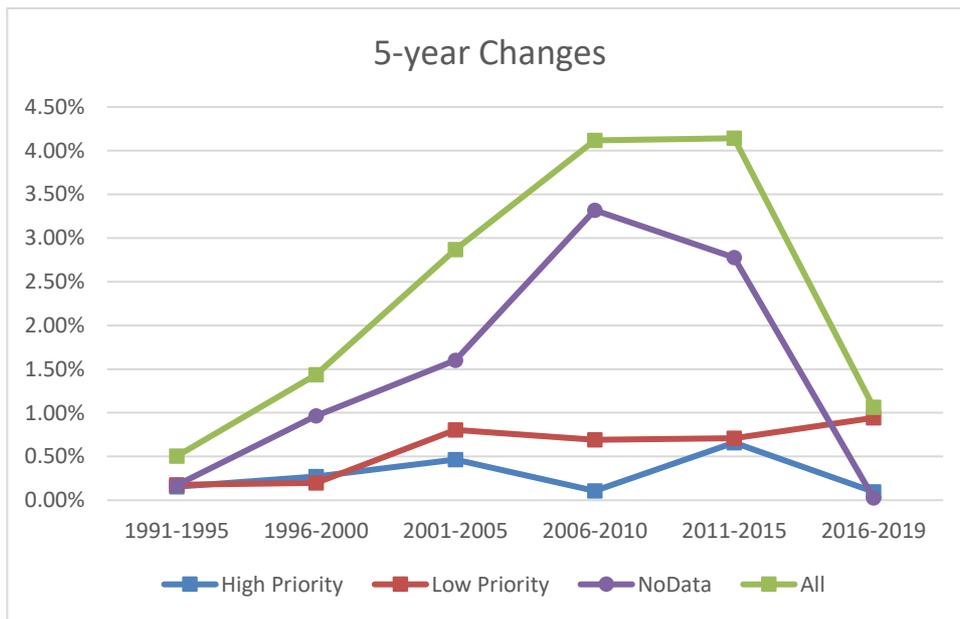


Figure 181. Percentage of 5-year changes of Africa.

The number of changes in Siberia is mainly affected by data availability especially for earlier years and it is considerably low in comparison with the other areas (see Figure 182 and Figure 183). Looking at Figure 183, the overall 5-year changes is aligned to the 5-year low priority changes since they represent the highest percentage of the overall changes. The percentage of low priority changes has the highest value of 0.4% for the years 2001-2005 and the lowest value for the years 2016-2019 of 0.08%.

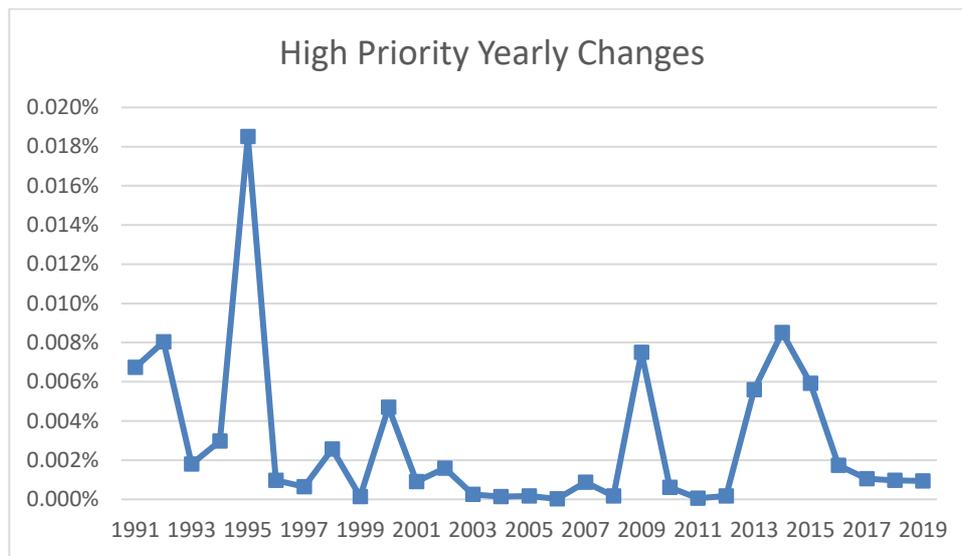


Figure 182. Percentage of High priority yearly changes of Siberia.

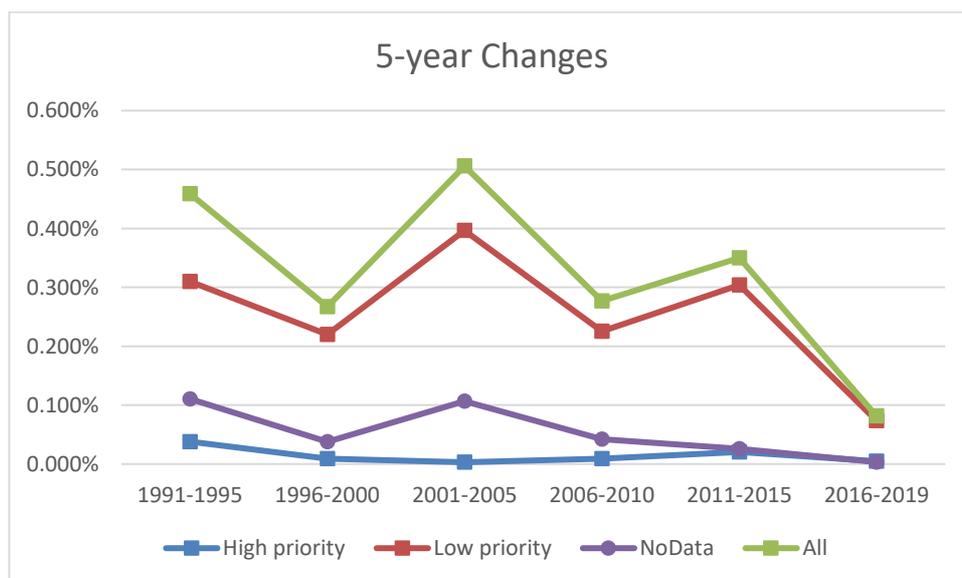


Figure 183. Percentage of 5-year changes of Siberia.

In order to have a better understanding of the number of LC transitions per year, another analysis has been done considering the yearly high priority changes for the years 2010 to 2019. Figure 184, Figure 185 and Figure 186, shows the amount of LC transitions on a yearly basis for the Amazonia, Africa and Siberia, respectively. From each plot it is possible to understand which kind of transition (among the high priority ones) contributes more to the LCC. For Amazonia (Figure 184) we can observe that there is an increasing trend for almost all kinds of high priority transitions except “Tree cover deciduous broadleaf to Crop” that shows a peak in 2015 and after that tends to decrease. “Grass to Shrub cover deciduous” is the main LC transition in Africa (Figure 185). It has increased dramatically in 2014 and then started a decreasing trend yet dominating the others. In general, we can observe that most of the changes show an increasing trend over time. Siberia (Figure 186) has the lowest number of high priority changes. Different kinds of tree cover to grass and shrubs to grass are the main changes that have happened in this area. Some of the changes show a peak between 2013 and 2015 and then their incidence reduces, others show a more regular increasing trend over time.

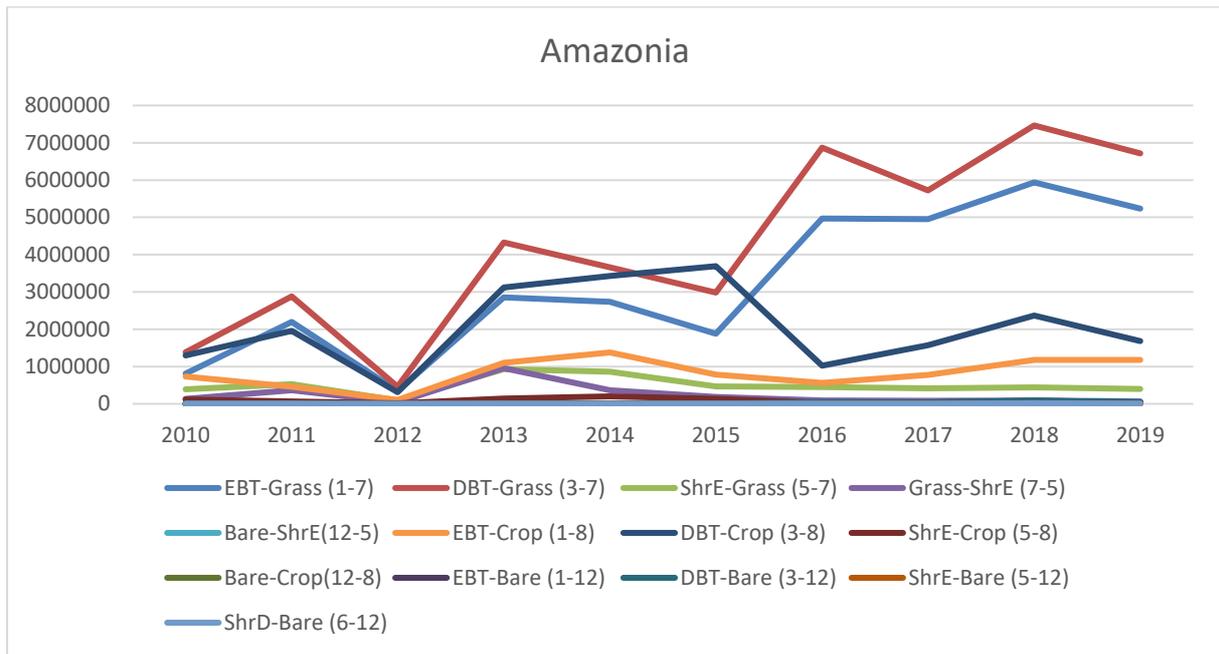


Figure 184. The number of LC transitions for Amazonia from 2010 to 2019

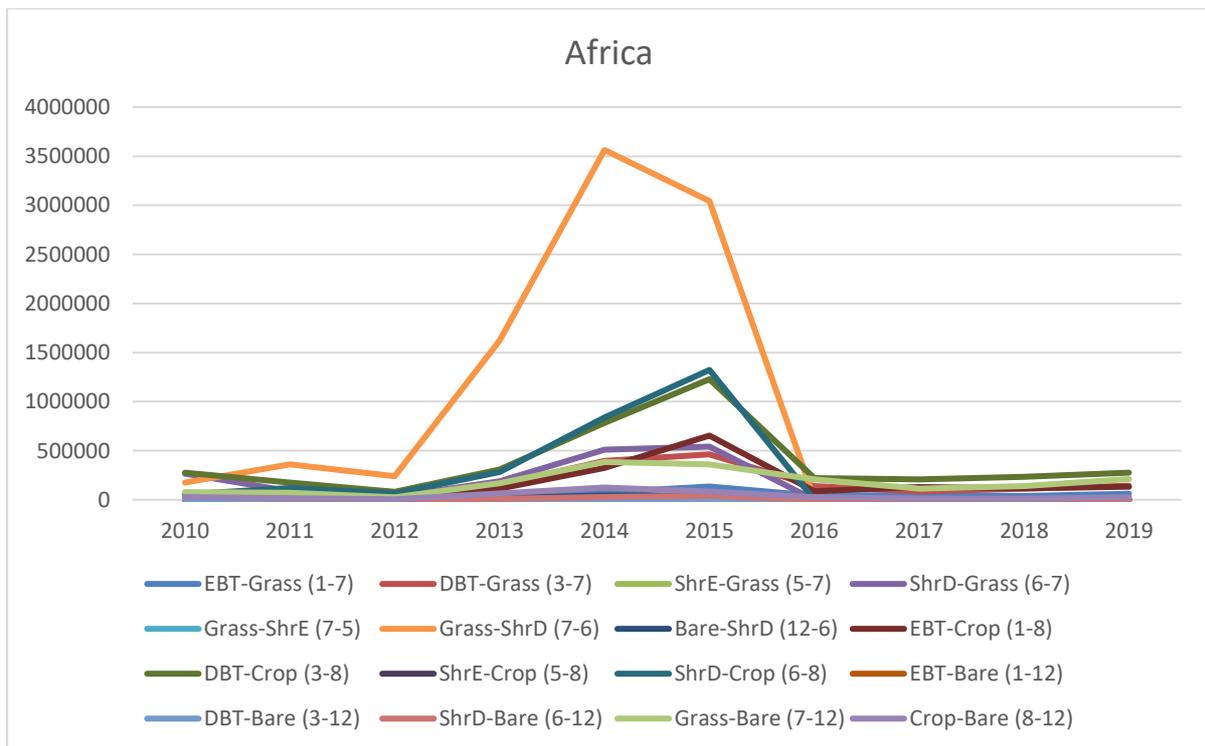


Figure 185. The number of LC transitions for Africa from 2010 to 2019

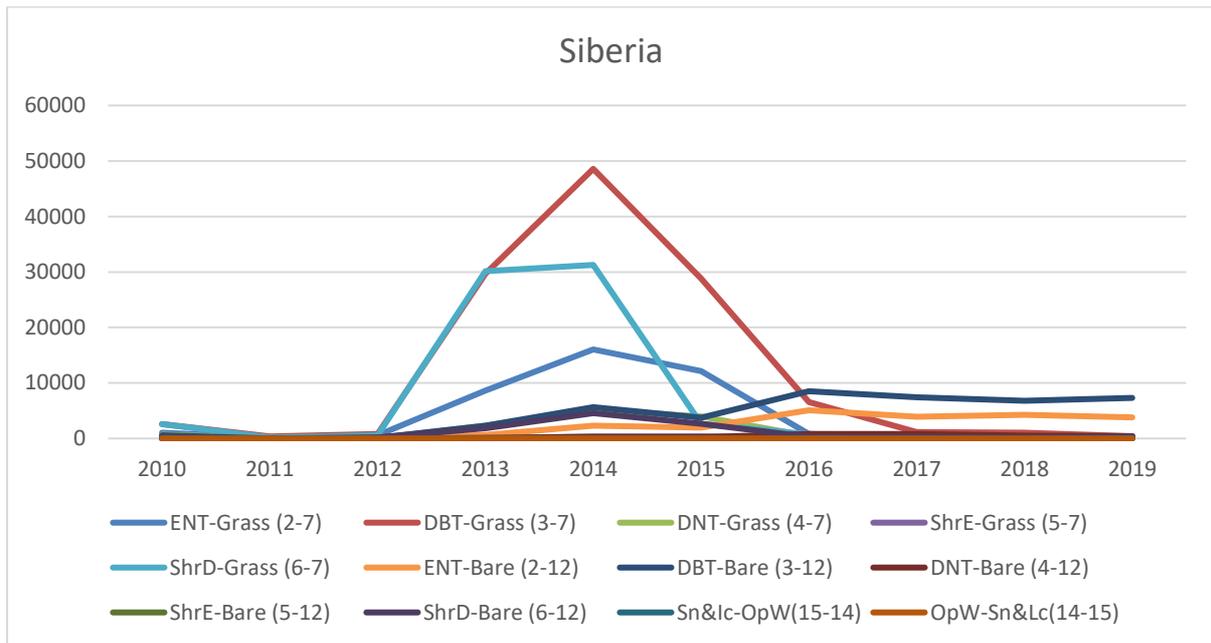


Figure 186. The number of LC transitions for Siberia from 2010 to 2019

5.5 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. For example, changes related to road construction are always kept, even if they are associated to small changes.

5.5.1 Results of post processing

Figure 187 shows an example of residual misregistration between the 2015 LC map and the 2019 map, which is mostly related to the down sampling of the 2019 map to the resolution of 30m. The post processing is able to remove the misregistration noise as well as other noisy small changes without affecting the real line-like changes associated to roads.

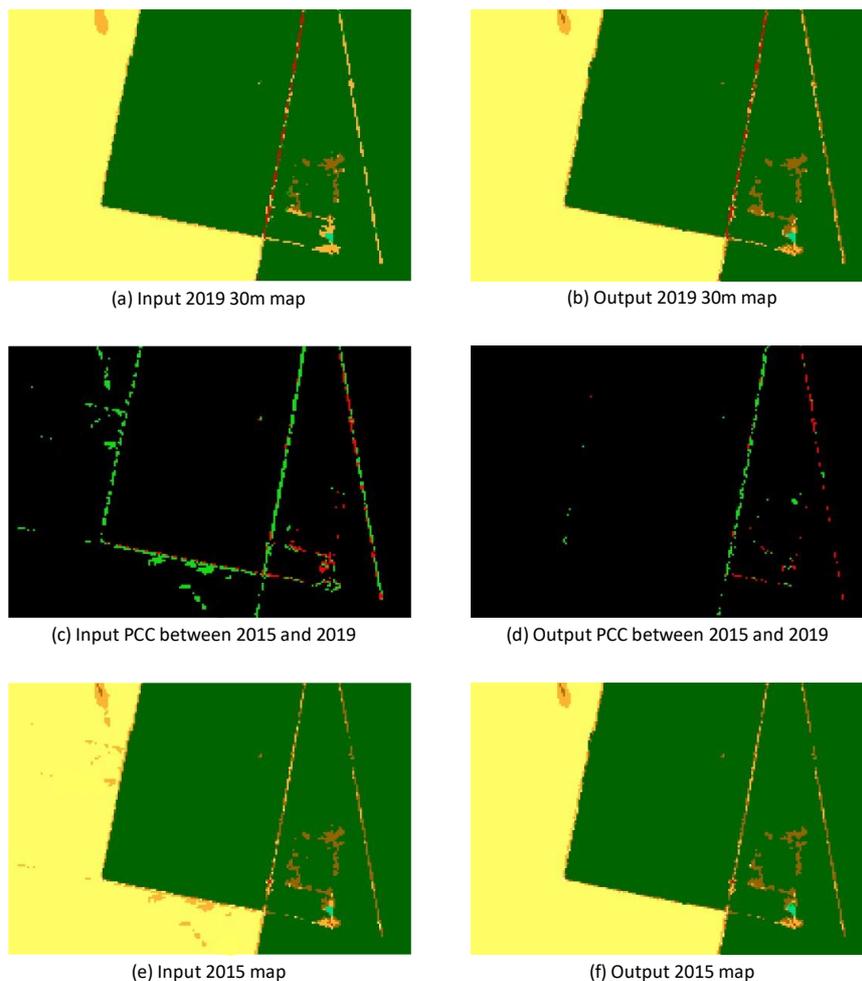


Figure 187. Example of results of post processing. The left column shows the input data for 2015 and 2019, while the right column shows the corresponding output of post processing. In the middle, the PCC is shown, where the green color is used for low-priority changes and the red color is used for high-priority changes.

6 References

- [1] M. A. Friedl *et al.*, "MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets," *Remote Sensing of Environment*, vol. 114, no. 1, pp. 168–182, Jan. 2010, doi: 10.1016/j.rse.2009.08.016.
- [2] 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.
- [3] X. Niu and Y. Ban, "Multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using an object-based support vector machine and a rule-based approach," *International Journal of Remote Sensing*, vol. 34, no. 1, pp. 1–26, Jan. 2013, doi: 10.1080/01431161.2012.700133.
- [4] N. Longepe, P. Rakwatin, O. Isoguchi, M. Shimada, Y. Uryu, and K. Yulianto, "Assessment of ALOS PALSAR 50 m Orthorectified FBD Data for Regional Land Cover Classification by Support Vector Machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2135–2150, Jun. 2011, doi: 10.1109/TGRS.2010.2102041.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	190	

- [5] H. Balzter, B. Cole, C. Thiel, and C. Schmullius, "Mapping CORINE Land Cover from Sentinel-1A SAR and SRTM Digital Elevation Model Data using Random Forests," *Remote Sensing*, vol. 7, no. 11, pp. 14876–14898, Nov. 2015, doi: 10.3390/rs71114876.
- [6] J. Soria-Ruiz, Y. Fernandez-Ordoñez, and I. H. Woodhouse, "Land-cover classification using radar and optical images: a case study in Central Mexico," *International Journal of Remote Sensing*, vol. 31, no. 12, pp. 3291–3305, Jun. 2010, doi: 10.1080/01431160903160777.
- [7] J. Liu and J. Li, "Land-use and land-cover analysis with remote sensing images," in *2013 IEEE Third International Conference on Information Science and Technology (ICIST)*, Mar. 2013, pp. 1175–1177. doi: 10.1109/ICIST.2013.6747746.
- [8] Y. Li, J. Chen, Q. Ma, H. K. Zhang, and J. Liu, "Evaluation of Sentinel-2A Surface Reflectance Derived Using Sen2Cor in North America," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 6, pp. 1997–2021, Jun. 2018, doi: 10.1109/JSTARS.2018.2835823.
- [9] J. Chen, X. Zhu, J. E. Vogelmann, F. Gao, and S. Jin, "A simple and effective method for filling gaps in Landsat ETM+ SLC-off images," *Remote Sensing of Environment*, vol. 115, no. 4, pp. 1053–1064, Apr. 2011, doi: 10.1016/j.rse.2010.12.010.
- [10] M. Drusch *et al.*, "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, May 2012, doi: 10.1016/j.rse.2011.11.026.
- [11] G. Georgievski and S. Hagemann, "Characterizing uncertainties in the ESA-CCI land cover map of the epoch 2010 and their impacts on MPI-ESM climate simulations," *Theor Appl Climatol*, vol. 137, no. 1, pp. 1587–1603, Jul. 2019, doi: 10.1007/s00704-018-2675-2.
- [12] M. Buchhorn *et al.*, "Copernicus Global Land Service: Land Cover 100m: epoch 2015: Globe. Dataset of the global component of the Copernicus Land Monitoring Service.," 2019.
- [13] 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.
- [14] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004, doi: 10.1109/TGRS.2004.831865.
- [15] M. Rußwurm and M. Körner, "Multi-Temporal Land Cover Classification with Long Short-Term Memory Neural Networks," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42W1, p. 551, May 2017, doi: 10.5194/isprs-archives-XLII-1-W1-551-2017.
- [16] J. Inglada, A. Vincent, M. Arias, B. Tardy, D. Morin, and I. Rodes, "Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series," *Remote Sensing*, vol. 9, no. 1, p. 95, Jan. 2017, doi: 10.3390/rs9010095.
- [17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," 1976, Accessed: Feb. 27, 2017. [Online]. Available: <https://dash.harvard.edu/handle/1/3426318>
- [18] C. Pelletier, S. Valero, J. Inglada, N. Champion, C. Marais Sicre, and G. Dedieu, "Effect of Training Class Label Noise on Classification Performances for Land Cover Mapping with Satellite Image Time Series," *Remote Sensing*, vol. 9, no. 2, p. 173, Feb. 2017, doi: 10.3390/rs9020173.
- [19] "Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities: International Journal of Remote Sensing: Vol 25, No 9." <https://www.tandfonline.com/doi/abs/10.1080/0143116031000150077> (accessed Oct. 29, 2019).
- [20] ESA, "Land Cover CCI Product User Guide Version 2.0." [Online]. Available: http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf
- [21] 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.
- [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] 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.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	191	

- [24] A. Folleco, T. M. Khoshgoftaar, J. Van Hulse, and L. Bullard, "Software quality modeling: The impact of class noise on the random forest classifier," in *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, Jun. 2008, pp. 3853–3859. doi: 10.1109/CEC.2008.4631321.
- [25] A. Mellor, S. Boukir, A. Haywood, and S. Jones, "Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 155–168, Jul. 2015, doi: 10.1016/j.isprsjprs.2015.03.014.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] A. B. Carlson and P. Crilly, *Communication Systems*, 5 edizione. Boston: McGraw-Hill Education, 2009.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] I. Zavorin and J. Le Moigne, "Use of multiresolution wavelet feature pyramids for automatic registration of multisensor imagery," *IEEE Transactions on Image Processing*, vol. 14, no. 6, pp. 770–782, Jun. 2005, doi: 10.1109/TIP.2005.847287.
- [35] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2941–2958, 2017, doi: 10.1109/TGRS.2017.2656380.
- [36] M. Costantini *et al.*, "Automatic coregistration of SAR and optical images exploiting complementary geometry and mutual information," presented at the International Geoscience and Remote Sensing Symposium (IGARSS), 2018, vol. 2018-July, pp. 8877–8880. doi: 10.1109/IGARSS.2018.8519242.
- [37] 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.
- [38] 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.
- [39] 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: 10.1561/20000000035.
- [40] R. Szeliski *et al.*, "A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008, doi: 10.1109/TPAMI.2007.70844.
- [41] A. Sorriso, D. Marzi, and P. Gamba, "A General Land Cover Classification Framework for Sentinel-1 SAR Data," in *2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI)*, Sep. 2021, pp. 211–216. doi: 10.1109/RTSI50628.2021.9597319.
- [42] 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.
- [43] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *MACHINE LEARNING IN PYTHON*, p. 6.

	Ref	CCI_HRLC_Ph1-PVASR		
	Issue	Date	Page	
	3.rev.1	13/01/2023	192	

- [44] 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.
- [45] 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,” *Image and Signal Processing for Remote Sensing XXVI*, vol. 11533, 2020.
- [46] 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.