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

Changelog

Issue	Changes	Date
1.0	First version.	29/10/2019

Detailed Change Record

Issue	RID	Description of discrepancy	Sections	Change

Contents

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 first year of activities, several comparative tests and performance analysis tasks have been carried out by the EOS team. For the current version of this document the outcome of these comparative activities is presented and indications on best performing algorithms/techniques are provided.

1.2 Purpose and scope

The Product Validation and Algorithm Selection Report (PVASR) v1.0 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. In its current status, PVASR builds upon a list of candidates as presented in the ATBD v1.1 [AD5]. Indeed, as summarized in Figure 1, a first version of ATBD was planned as preliminary source of information for algorithm choices before edition of the first PVASR version. This has been planned as such in order to be able to still produce a document related to algorithms (ATBD v1.1) before any comparative analysis took place, while having the time of gathering information from round-robin activities and internal benchmark operations to include in PVASR v1.0. In future versions (v2.0 and on), ATBD will only include details about the chosen algorithms while candidates will be moved to PVASR.

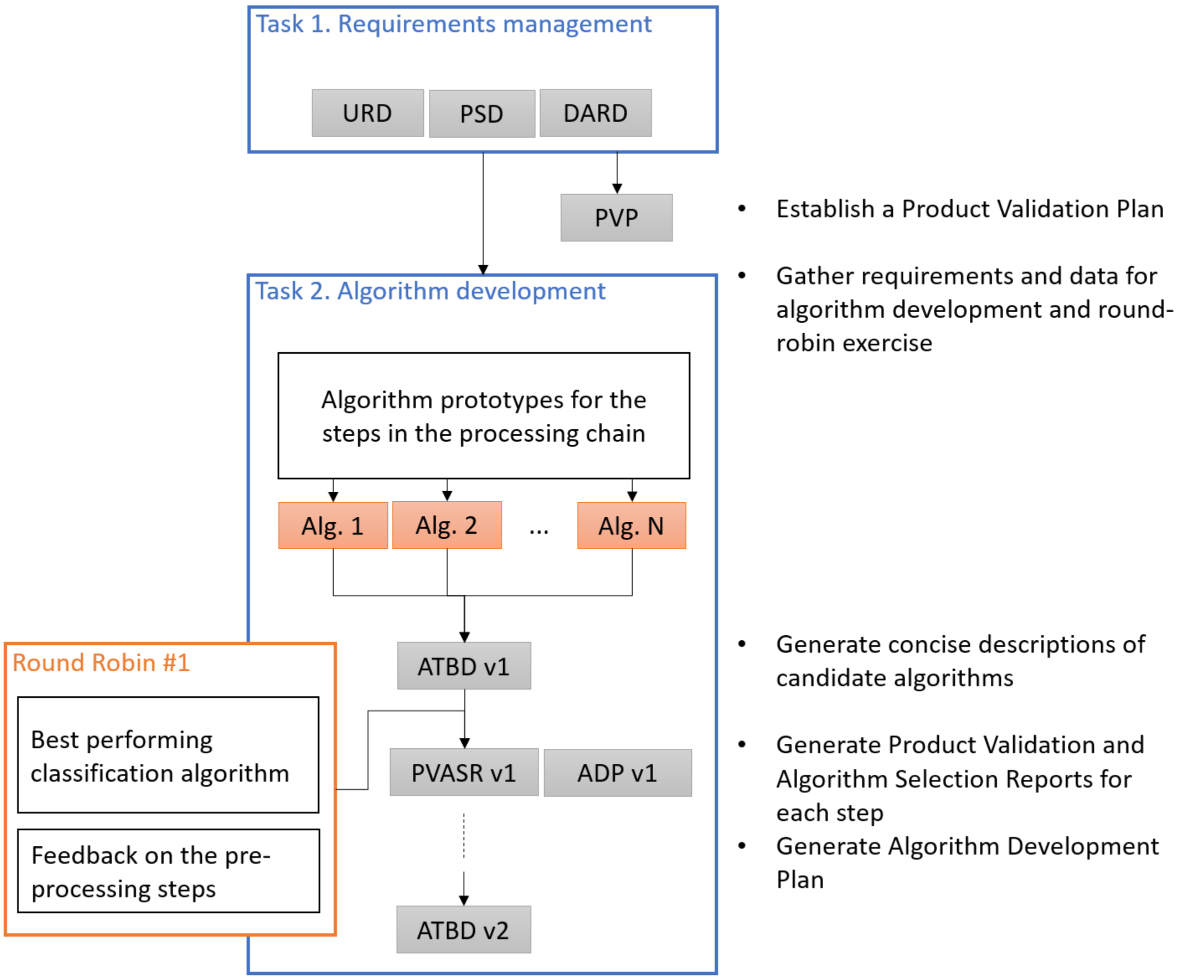

Figure 1. Concept of the PVASR v1 in the workflow of Task 2 of the CCI+ HRLC project.

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 PVASR v1.0, activities carried on in the first cycle are presented. Comparison and benchmarking activities have been devoted towards classification of optical and SAR imagery, 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.

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]	User Requirements Document, v1.1, 12/04/2019, CCI_HRLC_Ph1-D1.1_URD
[AD5]	Algorithm Theoretical Basis Document, v1.1., 25/09/2019, CCI_HRLC_Ph1-D2.2_ATBD_v1.1

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
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

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 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.);

- 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

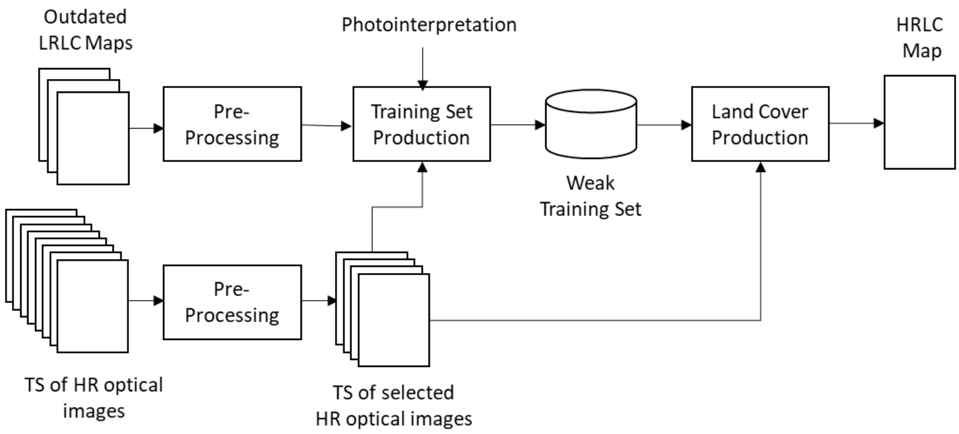

Figure 2. Optical data processing chain for the prototype production of the HR LC map obtained by classifying the time series of Sentinel 2 data.

Figure 2 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 3 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.

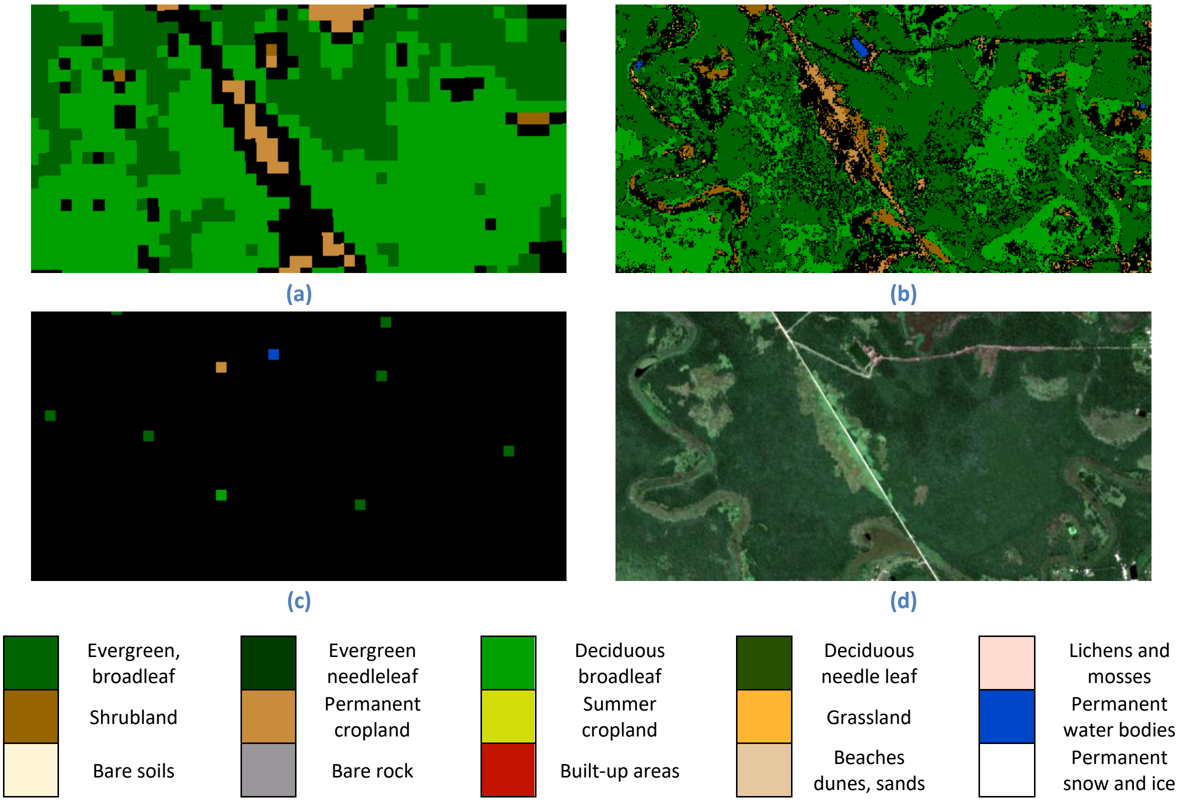

Figure 3. Visual comparison of the: (a) coarse thematic product generated by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (b) intermediate HR LC product produced by the ensemble of 5 classifiers, (c) extracted weak training samples, and (d) true color composition of the 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 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 ate 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 4 reports a comparison of classification results obtained in Amazonia (tile 21KUQ). The low geometrical resolution of the coarse thematic map (Figure 4b) 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 4a), the best LC map is obtained by using the SVM (Figure 4f). It is the only classifier that correctly extracts the build-up areas. The worst result is obtained by the ML (Figure 4c), which misclassifies evergreen broadleaf as permanent cropland.

Figure 5 presents examples of classification results obtained in Amazonia (tile 21KXT). In this case, there is small correspondence between the coarse thematic product (Figure 5b) and the LC in the scene (Figure 5a). 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.

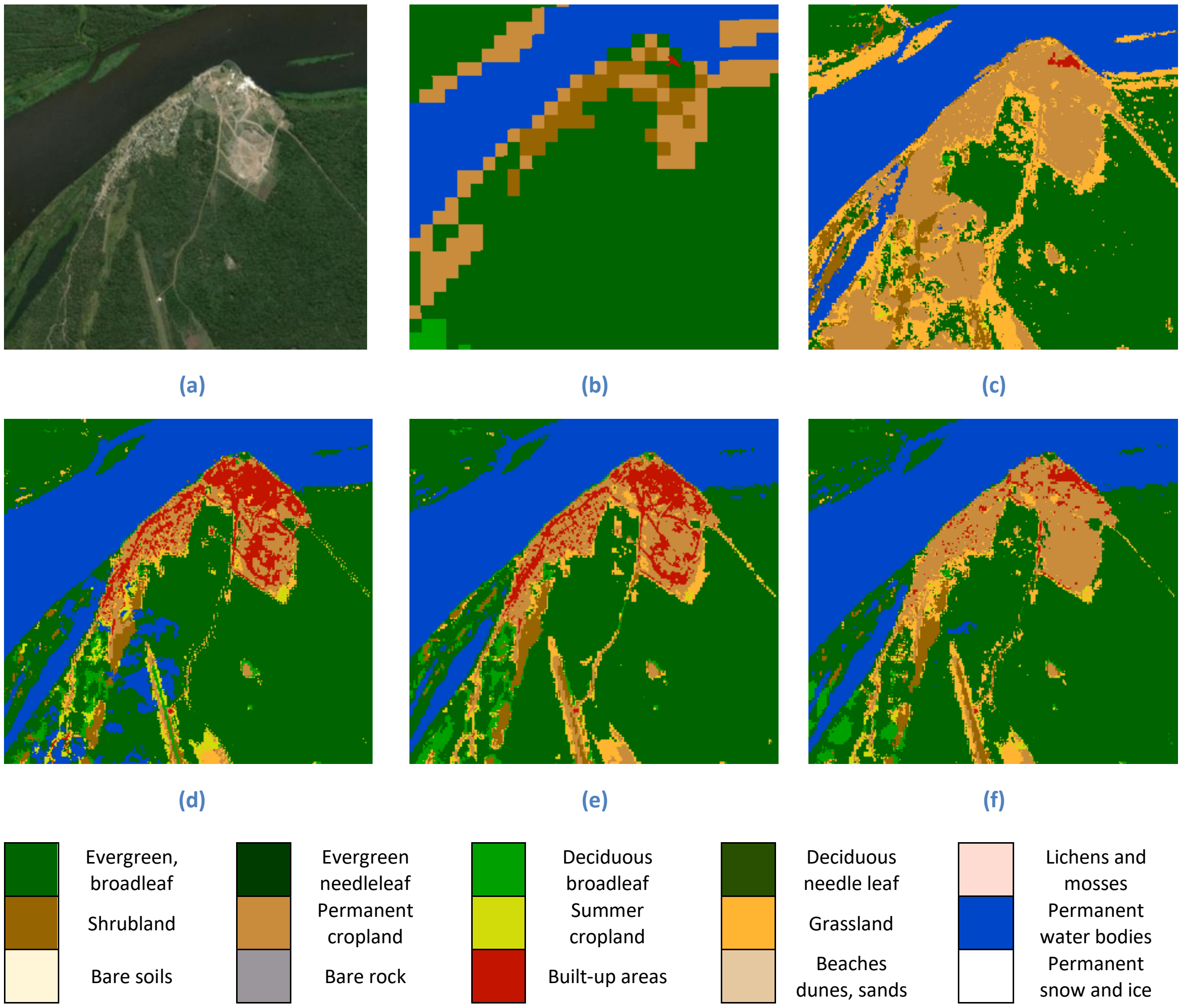

Figure 4. Visual comparison of the: (a) HR optical image used to evaluate the results obtained; (b) coarse thematic product generated by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (c) LC map obtained by using ML, (d) LC map obtained by using ANN, (e) LC map obtained by using RF, and (f) LC map obtained by using the SVM. The study area is in Amazonia (Tile 21KUQ).

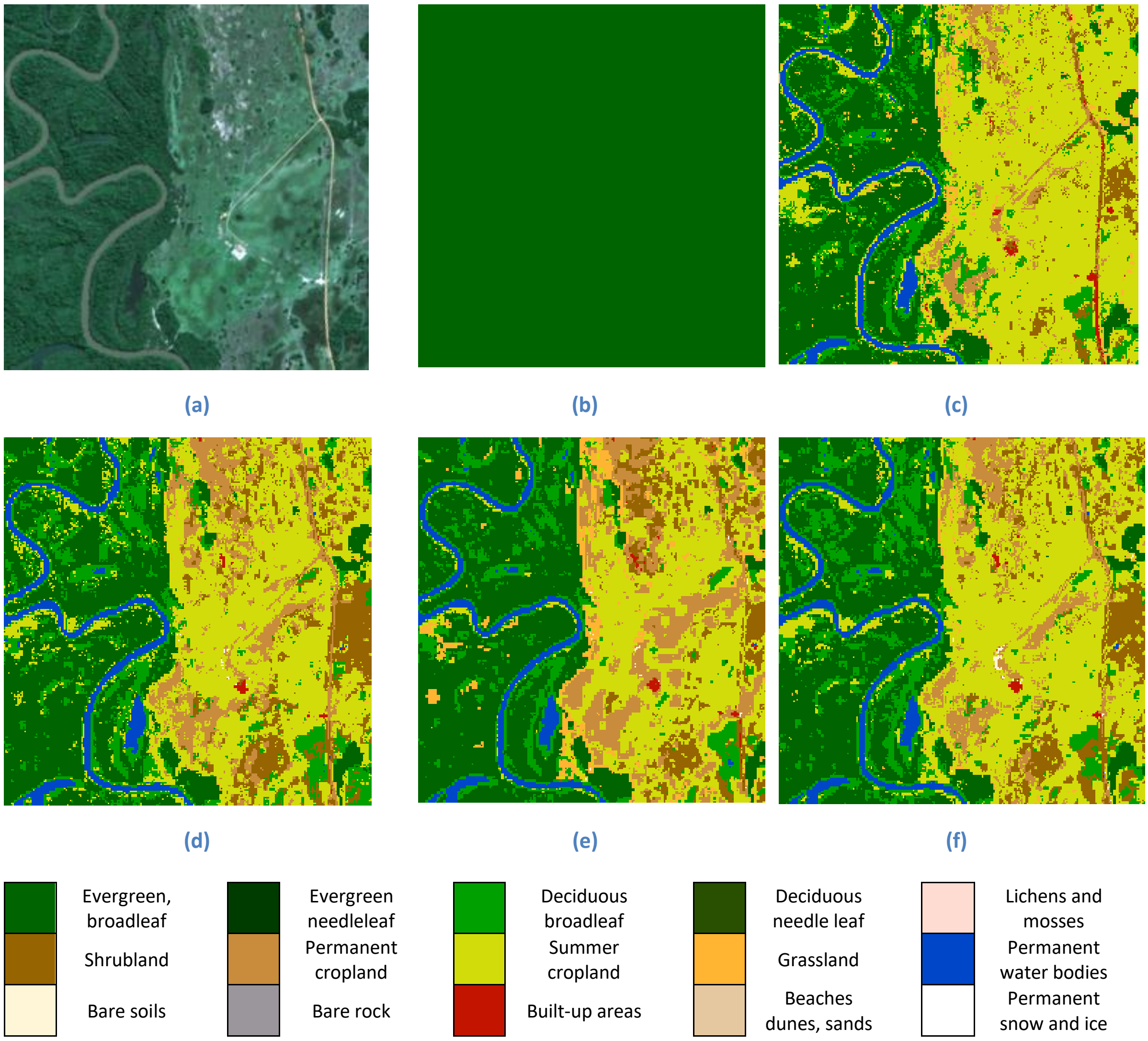

Figure 5. Visual comparison of the: (a) Sentinel 2 image acquired on the 9th May 2018; (b) coarse thematic product generated by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (c) LC map obtained by using ML, (d) LC map obtained by using ANN, (e) LC map obtained by using RF, and (f) LC map obtained by using SVM. The study area is in Amazonia (Tile T1KXT).

Figure 6 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 6c). One can notice that the build-up areas are accurately classified in all the HR maps.

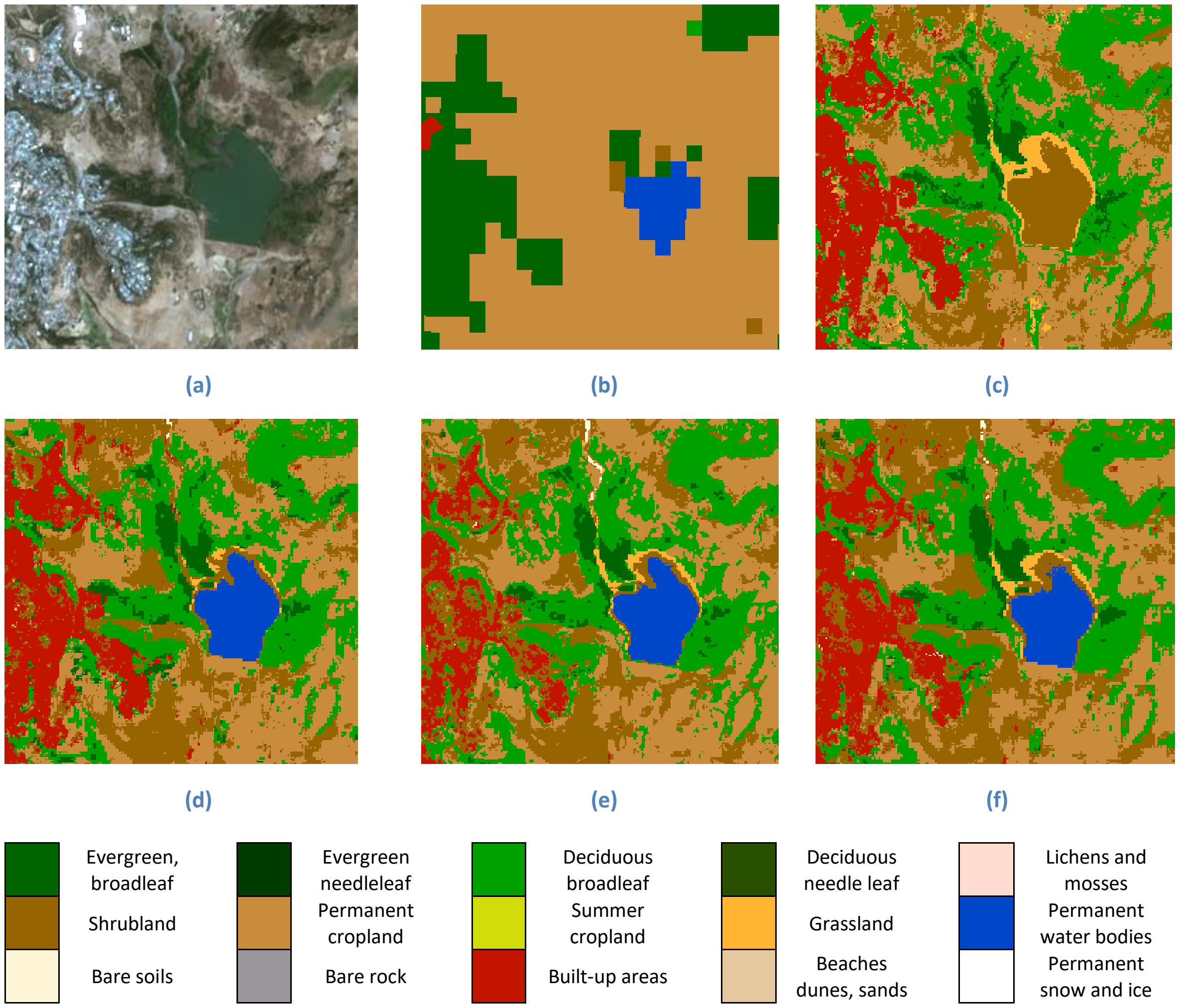

Figure 6. Visual comparison of the: (a) Sentinel 2 optical image acquired on the 3rd March 2018; (b) coarse thematic product generated by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (c) LC map obtained by using ML, (d) LC map obtained by using ANN, (e) LC map obtained by using RF, and (f) LC map obtained by using SVM. The study area is located in Africa (Tile 37PCP).

Figure 7 reports a comparison between the SVM classifier (Figure 7b and Figure 7e) and the African Prototype produced by ESA CCI (Figure 7c and Figure 7f) 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.

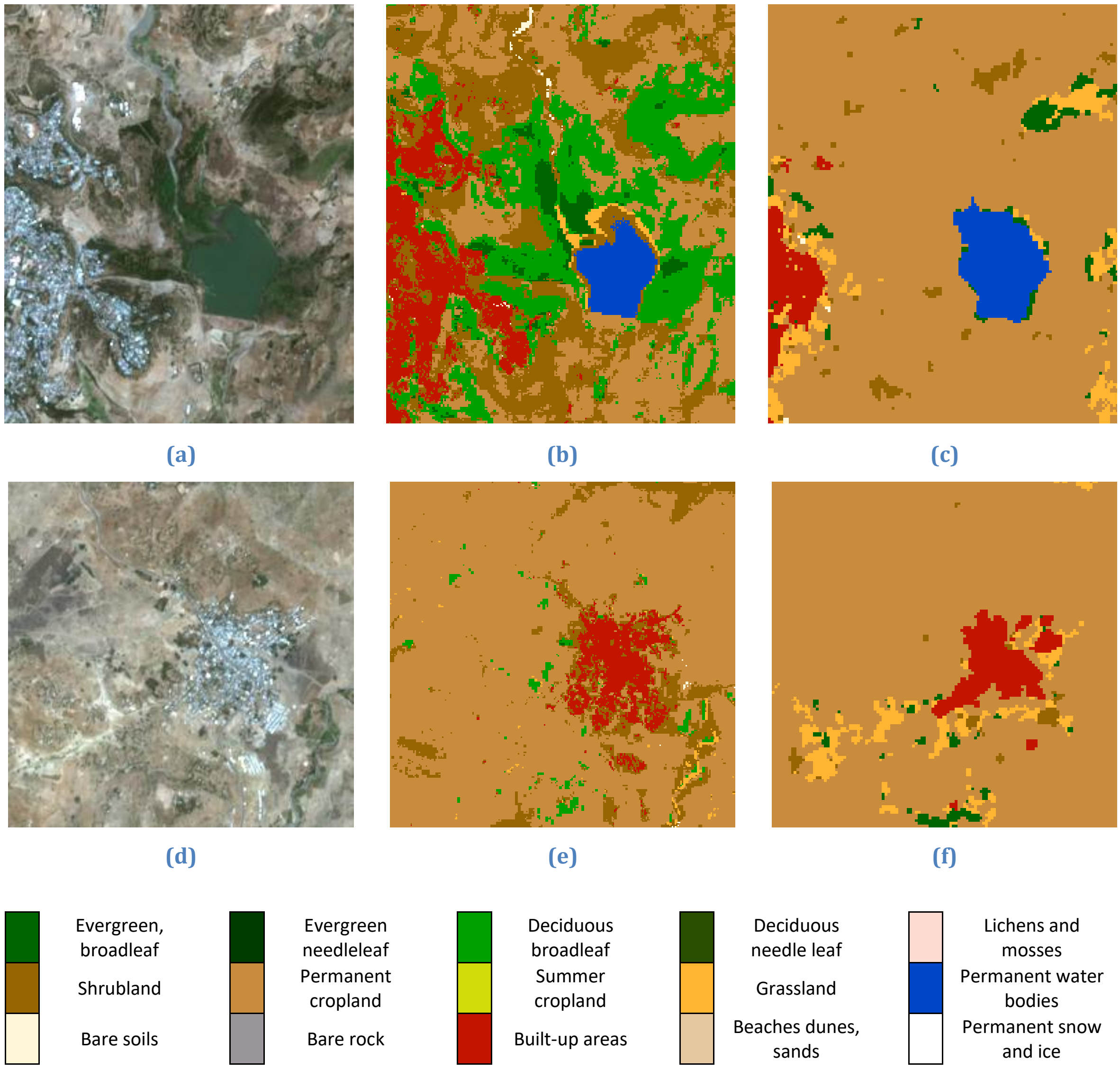

Figure 7. Visual comparison of the: (a)(d) true color composition of Sentinel 2 image acquired on the 3 rd March 2018; (b)(e) LC maps obtained by using SVM, (c)(f) African prototype LC maps produced by ESA CCI converted to the HRLC map legend. The study area is located in Africa (Tile 37PCP).

Figure 8 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 lynches 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.

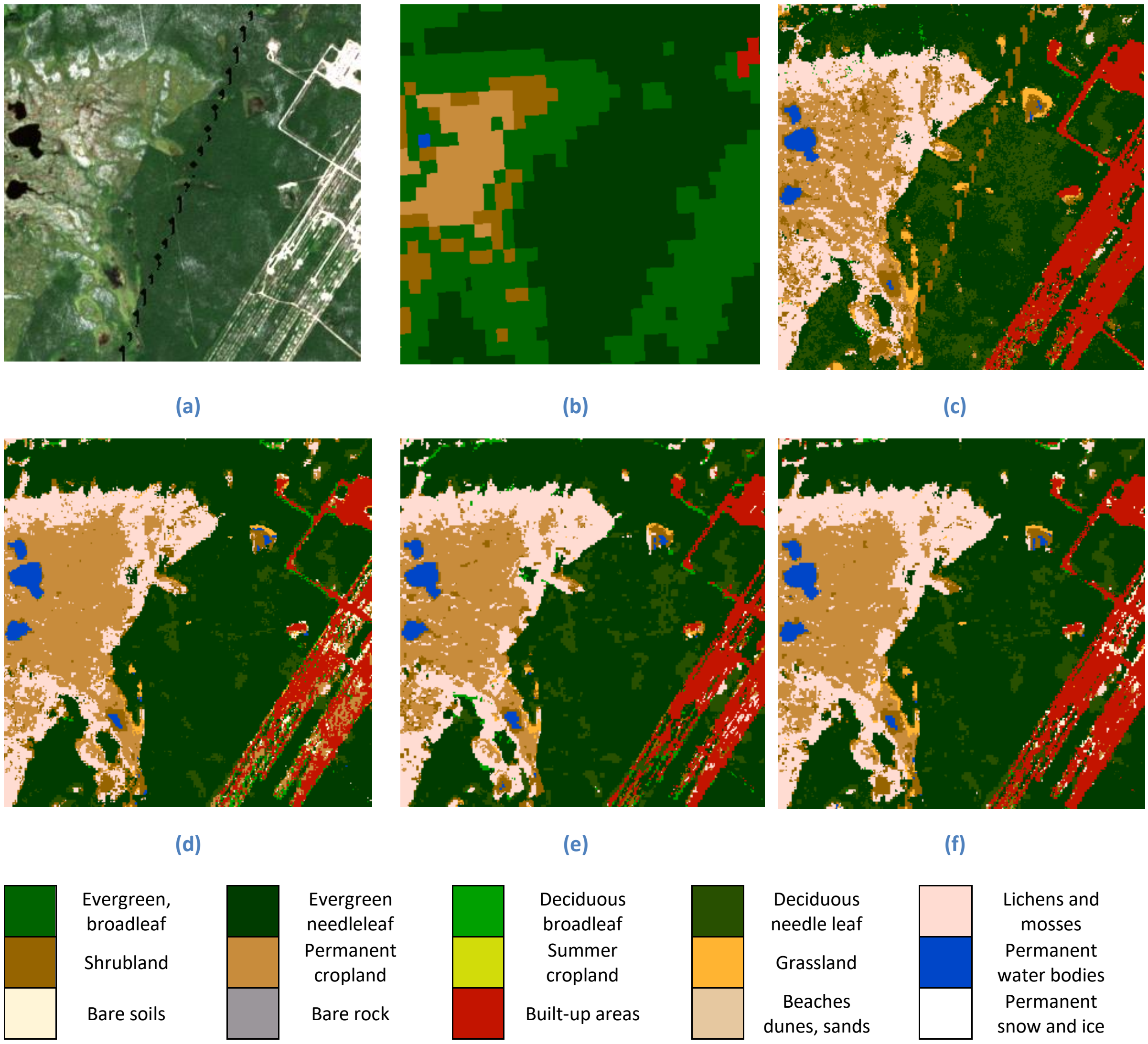

Figure 8. Visual comparison of the: (a) Sentinel 2 optical image acquired on the 21st July 2018; (b) coarse thematic product generated by merging the CGLC, ESA CCI 2015 and the GLCNMO after the legend conversion, (c) LC map obtained by using ML, (d) LC map obtained by using ANN, (e) the LC maps obtained 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

According to what is reported in the ATBD document [AD5], Figure 9 reports the general work flow for the kernel of the processing chain.

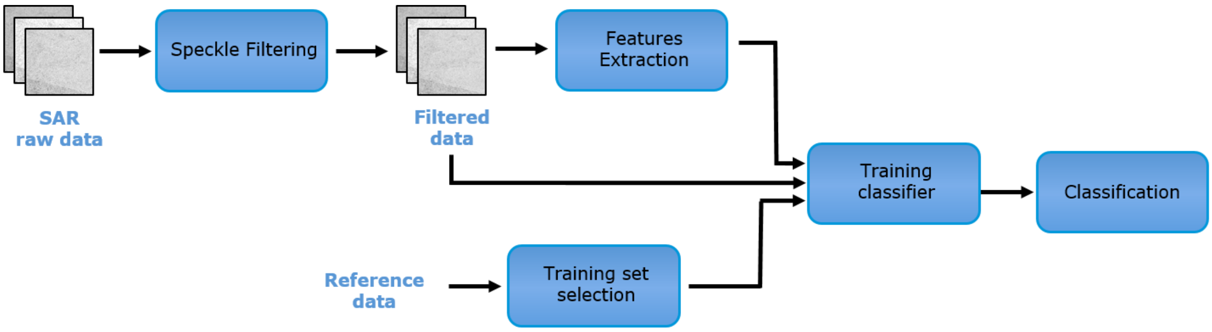

Figure 9. Block diagram for SAR data of the land cover map production procedure.

The reference data is split into a training subset. The image data is proper 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:

- Amazonia area for 21KUQ S-2 tile;
- Amazonia 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 develop and implemented in the SAR image processing chain in Figure 9.

3.2.2.1 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 10).

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 06-1406-26 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	
1	Evergreen broadleaf tree	
2	Evergreen needleleaf tree	
3	Deciduous broadleaf tree	
4	Deciduous needleleaf tree	
5	Shrubland	
6	Permanent cropland	
7	Annual summer cropland	
8	Grassland	
9	Lichens and mosses	
10	Permanent water bodies	
11	Permanent snow and ice	
12	Beaches dunes and sands	
13	Bare soils	
14	Bare rock	
15	Built-up areas	

Figure 10. Land cover classes for Sentinel-1 data classification

3.2.2.2 Speckle filtering

The Level-1 products of Sentinel-1 GRD data are calibrated and terrain corrected before any other processing. The preprocessing is accomplished by Sentinel's Application Platform (SNAP) software provided by ESA. 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.

3.2.2.3 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.

3.2.2.4 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 3, which amounts to about 20.000 Km^2.

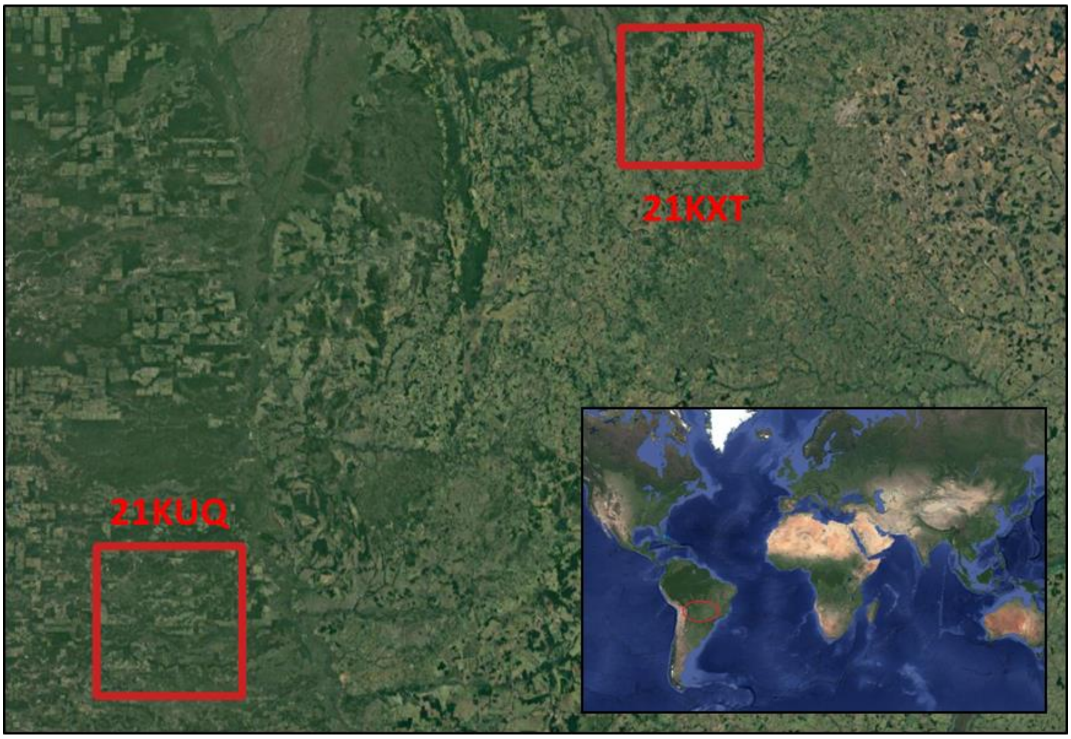

Figure 11. Study areas (highlighted in red) referring to 21KUQ and 21KXT Sentinel-2 granules.

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 3.2.2.3). 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.

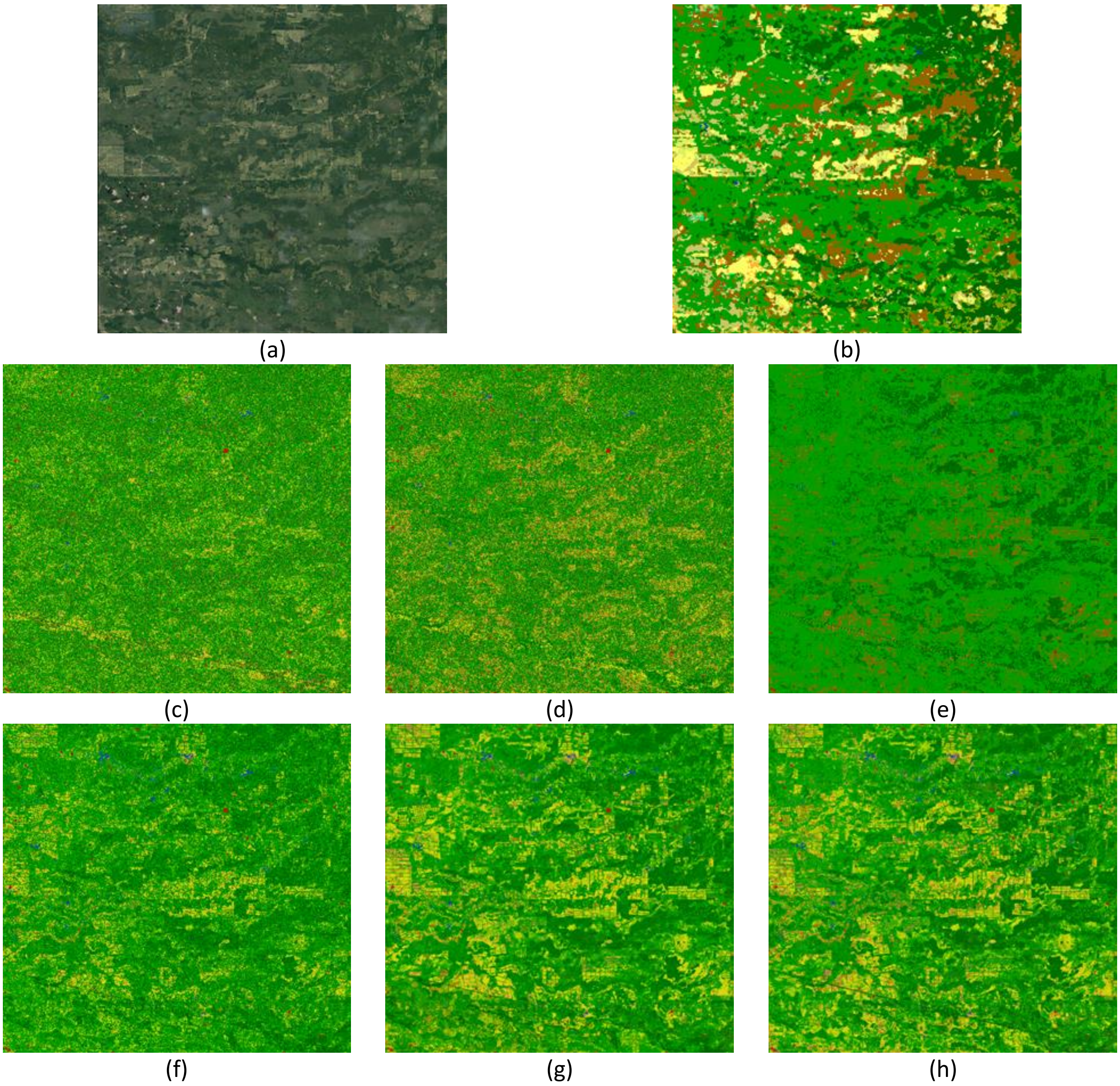

Figure 12. 21KUQ Amazonian tile: the figure shows a comparison among the satellite image (a); ESA CCI LC 2015 reference data (b), and classification maps using single image-single band (c), single image-double band (d), multitemporal sequence (e), multitemporal sequence with majority voting (f), seasonal 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 12 and Figure 13, 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 14 and Figure 15.

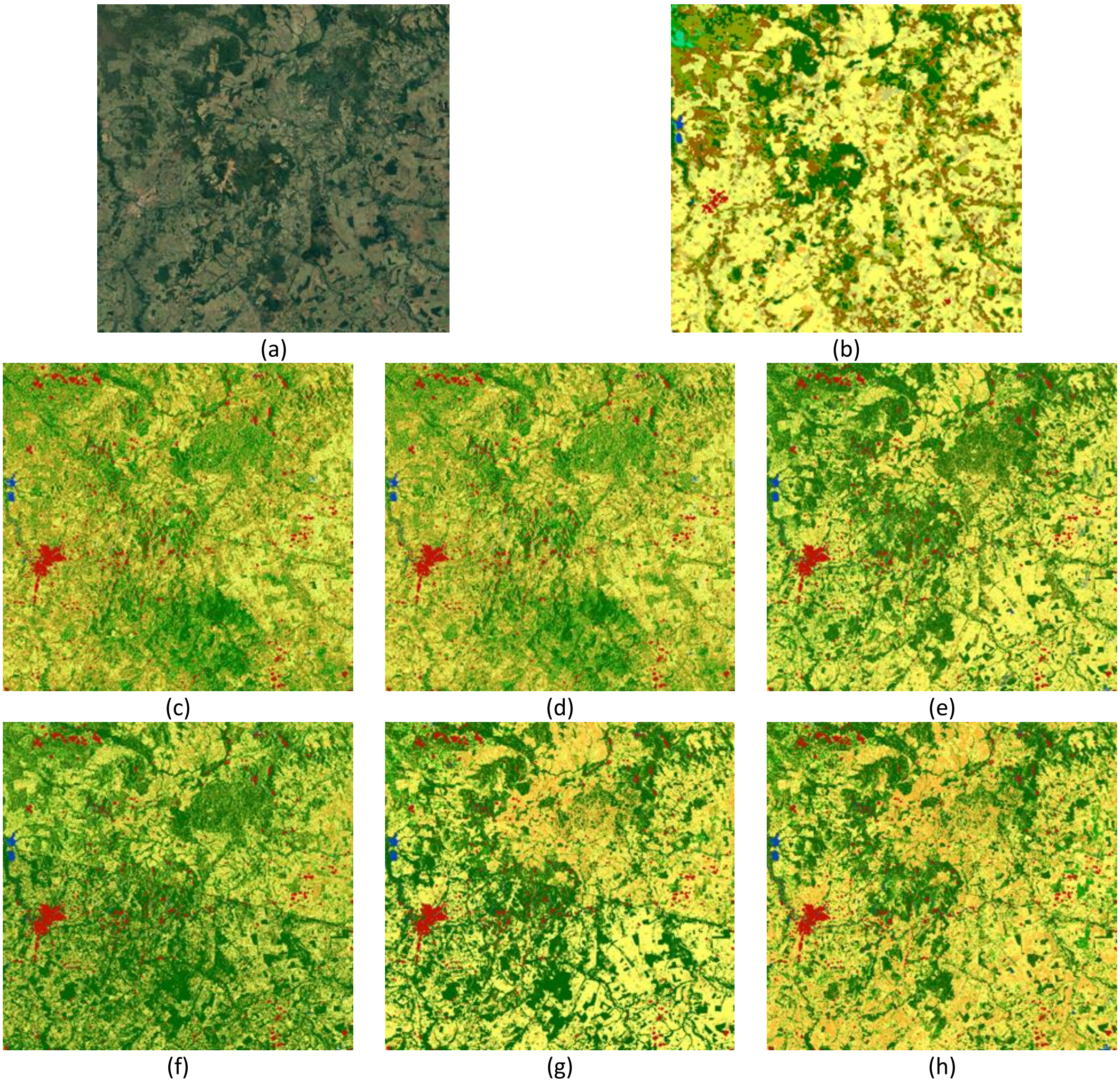

Figure 13. 21KXT Amazonian tile: the figure shows a comparison among the satellite image (a); ESA CCI LC 2015 reference data (b), and classification maps using single image-single band (c), single image-double band (d), multitemporal sequence (e), multitemporal sequence with majority voting (f), seasonal multitemporal sequence with three images per season (g) and seasonal multitemporal sequence with five images per season (h).

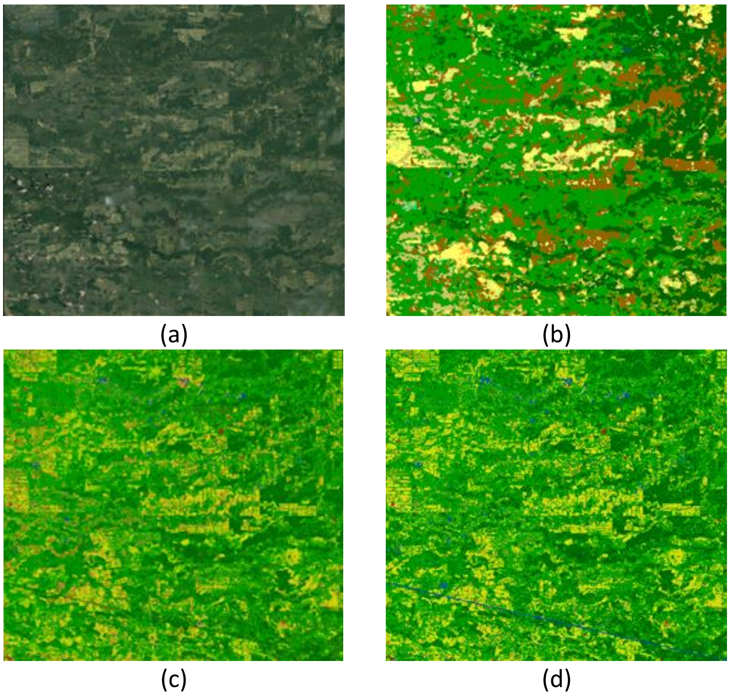

Figure 14. Comparison of RF and SVM classification for the 21KUQ Amazonian tile: the satellite image (a), ESA CCI LC 2015 reference data (b), RF map using a seasonal multitemporal sequence (five image per season) (c) and SVM map with the same input (d).

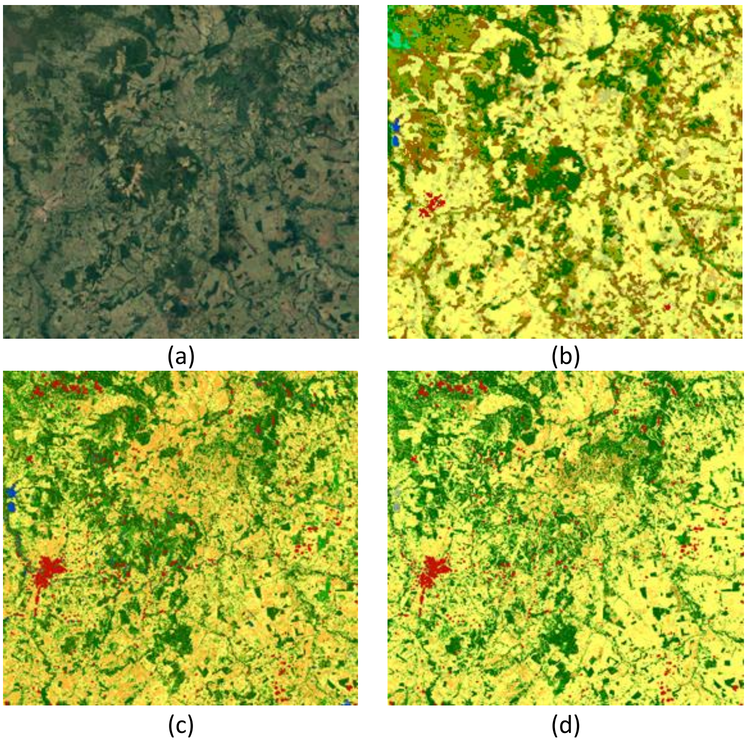

Figure 15. Comparison of RF and SVM classification in 21KXT Amazonian tile: the satellite image (a), ESA CCI LC 2015 reference data (b), RF map using a seasonal multitemporal sequence (five image per season) (c) and SVM map with the same input (d).

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 16 and Figure 17.

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].

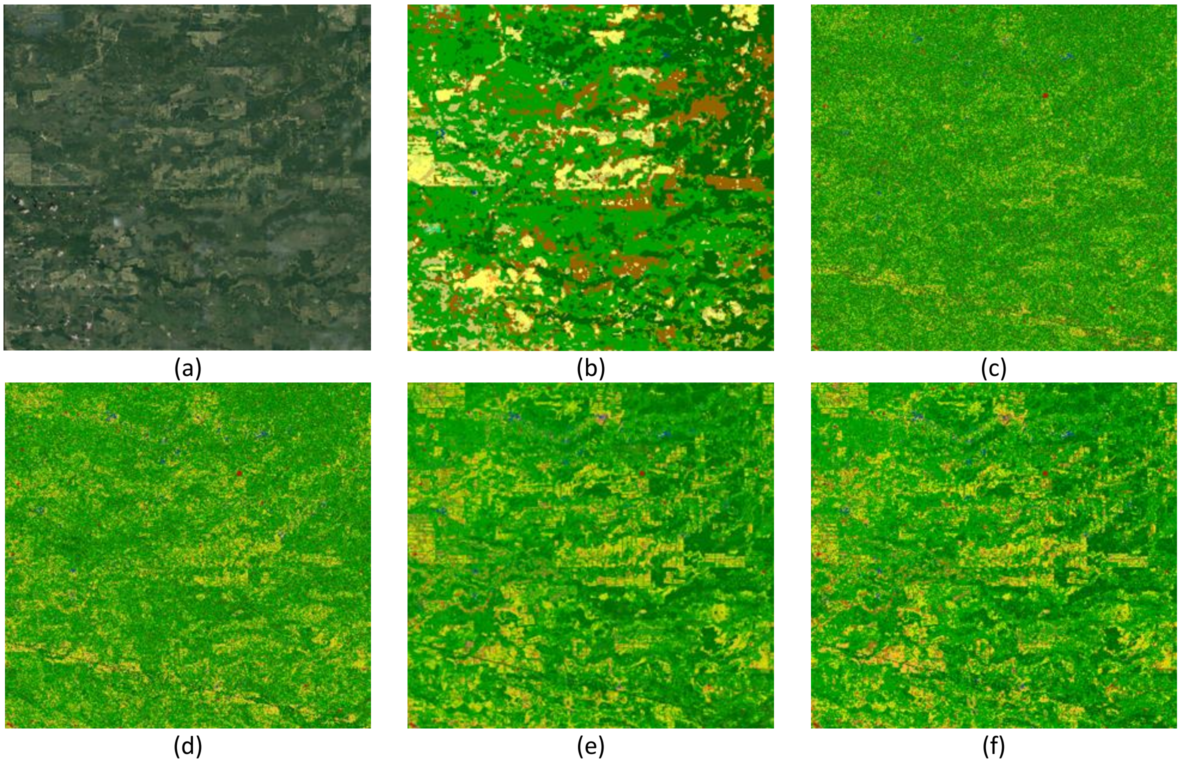

Figure 16. Lee filter applied in RF classification for the 21KUQ Amazonian tile: the satellite image (a), ESA CCI LC 2015 reference data (b), single image classification with multitemporal despeckling (c), single image classification with Lee despeckling (d), multitemporal classification with multitemporal filtering (e) and multitemporal classification with Lee filtering (d).

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 9 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 12 (c)) shows poor performance in terms of class recognition, as well as attained in double band analysis (Figure 12(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 12 (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 13, a similar behaviour can be appreciated, up to the map in Figure 13 (h).

In Figure 14 and Figure 15, 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 14 (c) and Figure 15 (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.

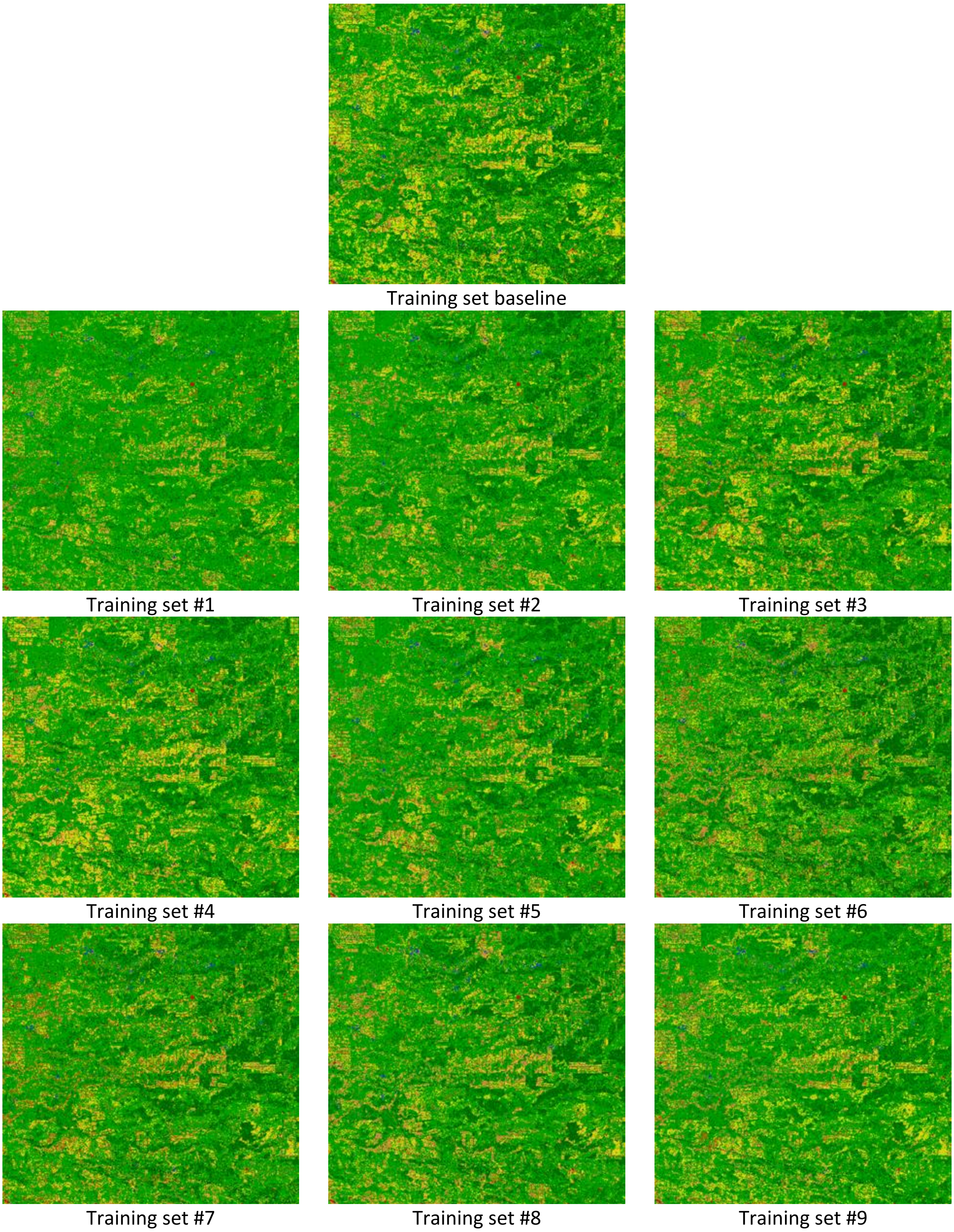

Figure 17. Random Forest maps for to the 21KUQ Amazonian tile using randomly selected different training sets.

Figure 16 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 16 (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 strong. In comparison, the multitemporal filter, applied to a temporal series of Sentinel-1 images (Figure 16(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 17), 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 18), 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 18. 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

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

- Mutual Information

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

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, $\mathrm{CC}(\cdot)$ is their cross-correlation evaluated on a given pixel location, $\mathrm{MI}(\cdot)$ 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:

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

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].

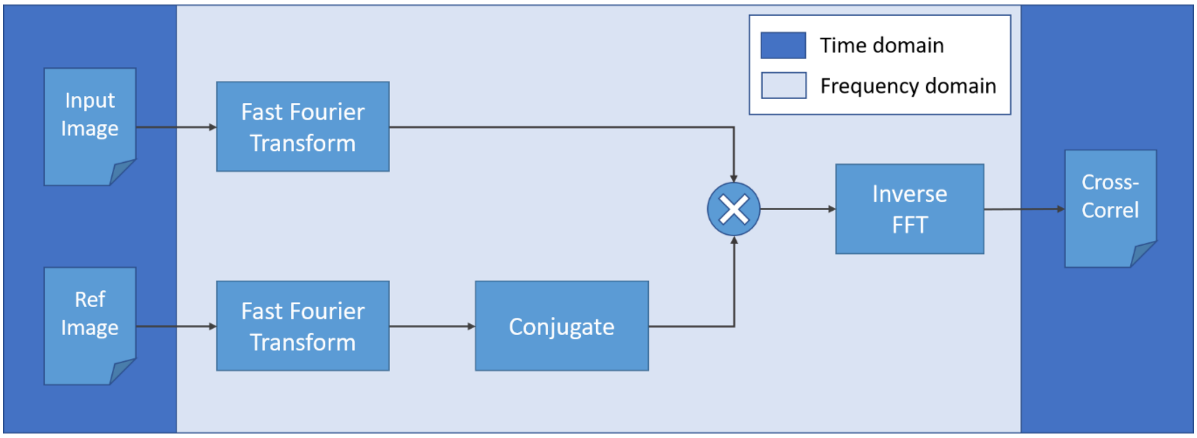

Figure 19. Flowchart of the cross-correlation computation between two images

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 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 19.

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 geo-referencing 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 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 20 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).

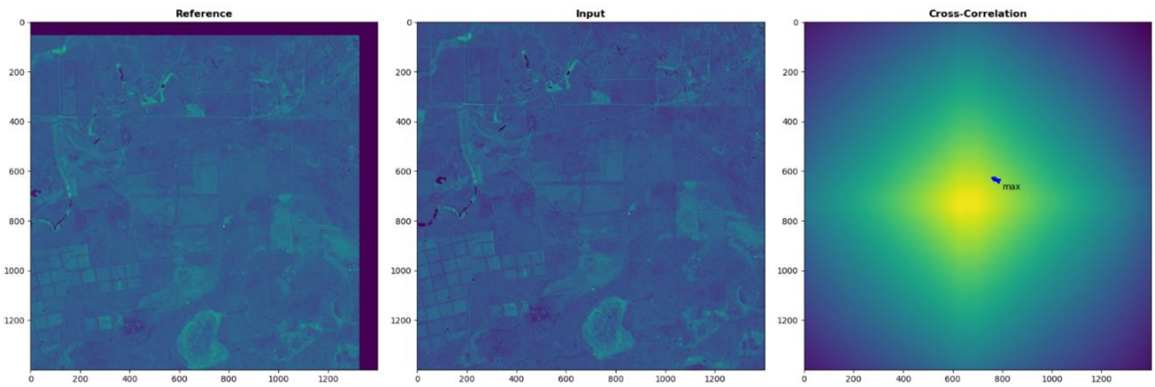

Figure 20. Synthetic optical-to-optical matching: reference optical image (left), input optical image to be registered (center), and cross-correlation (right).

Figure 21 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.

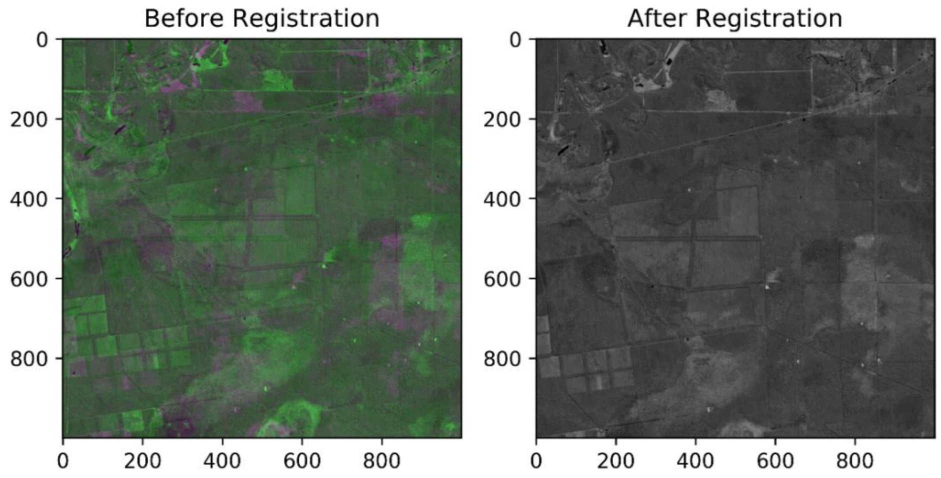

Figure 21. Synthetic optical-to-optical matching: false color composition of the images before (left) and after registration (right) based on translation and cross-correlation metric.

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

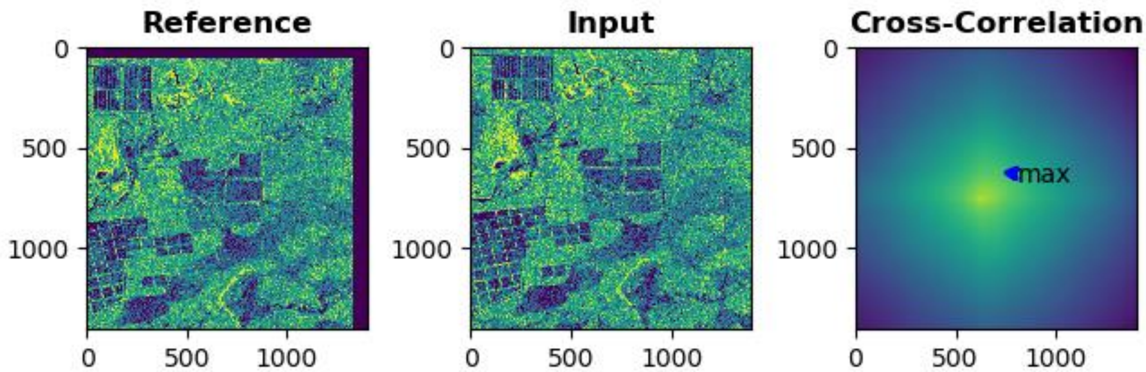

Figure 22. Synthetic SAR-to-SAR matching: reference SAR image (left), input SAR image to be registered (center) and cross-correlation (right).

Figure 23 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.

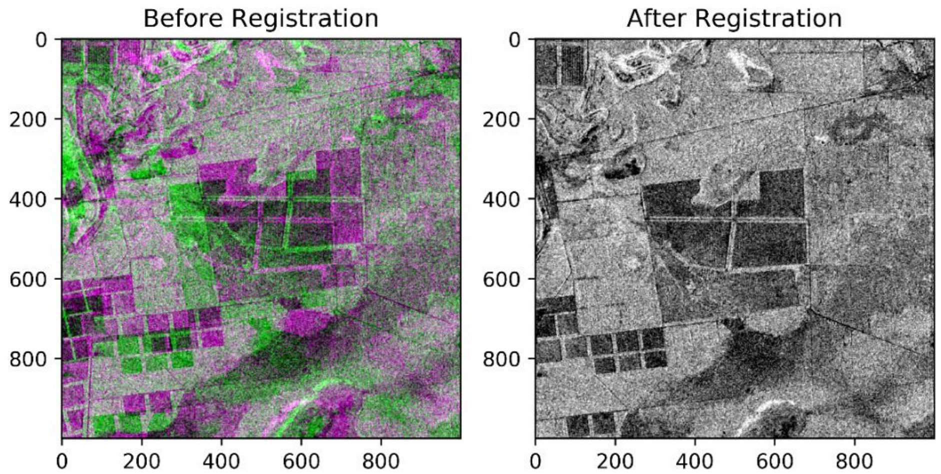

Figure 23. Synthetic SAR-to-SAR matching: false color composition of the images before (left) and after registration (right) based on translation and cross-correlation metric.

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

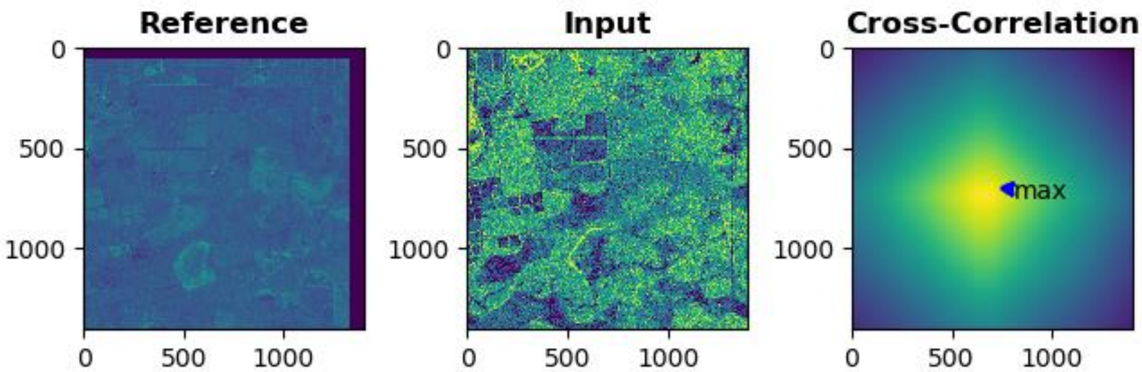

Figure 24. Synthetic Optical-to-SAR matching: reference optical image (left), input SAR image to be registered (center) and cross-correlation (right).

In this case, Figure 25 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.

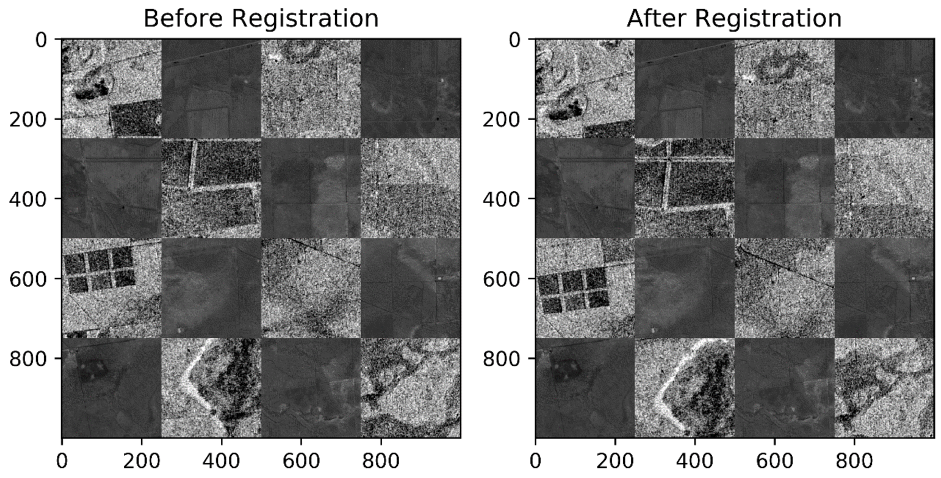

Figure 25. Synthetic Optical-to-SAR matching: checkerboard of the optical and SAR images before (left) and after registration (right) based on translation and cross-correlation.

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 26 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 27 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.

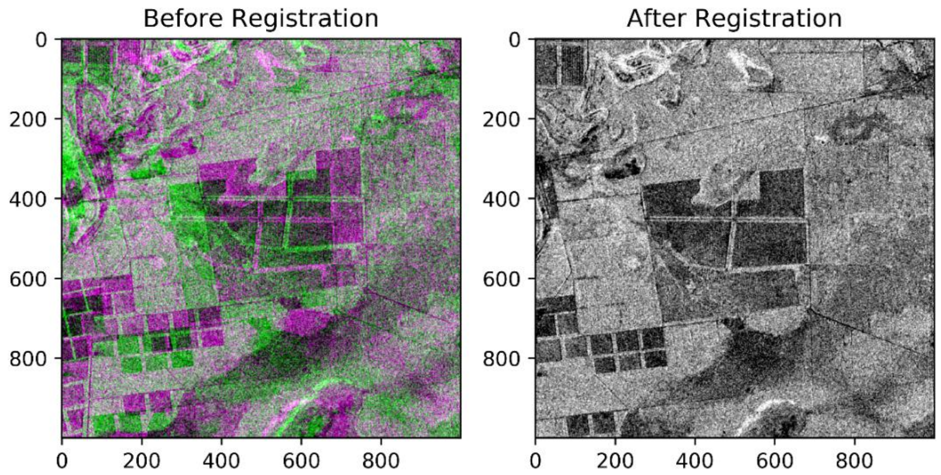

Figure 26. Synthetic SAR-to-SAR matching: false color composition of the images before (left) and after registration (right) based on translation, COBYLA, and cross-correlation metric.

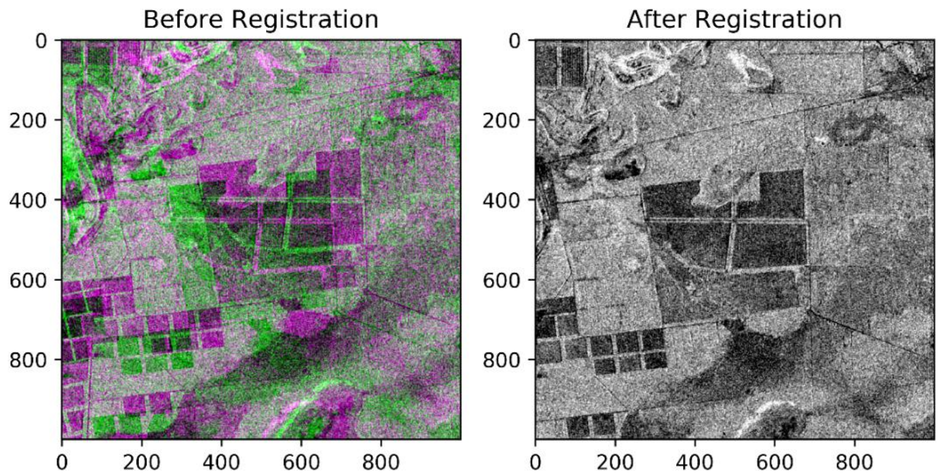

Figure 27. Synthetic SAR-to-SAR matching: false color composition of the images before (left) and after registration (right) based on translation, COBYLA, and mutual information metric.

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 28 highlights that, even in the multi-sensor case, the mutual information is effective for determining the optimal translation.

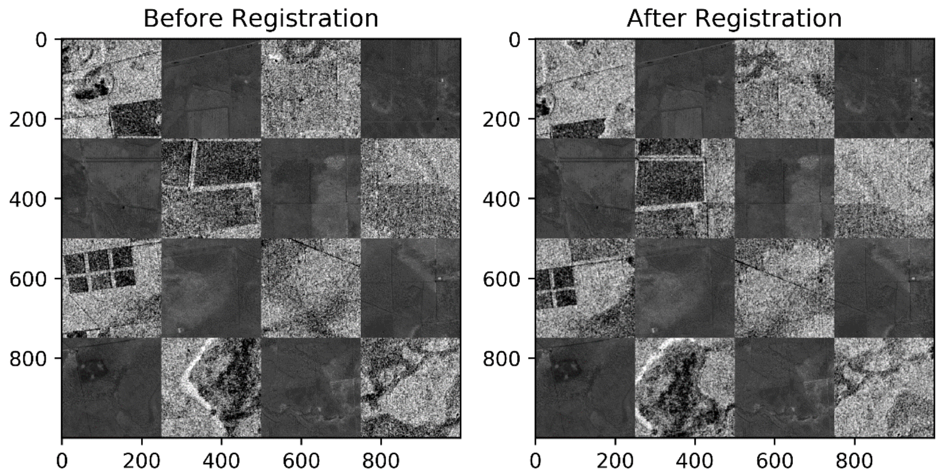

Figure 28. Synthetic Optical-to-SAR matching: checkerboard of the images before (left) and after registration (right) based on translation, COBYLA, and mutual information metric.

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 29), an accurate registration was obtained in this case of rigid transformation of a single-sensor image pair.

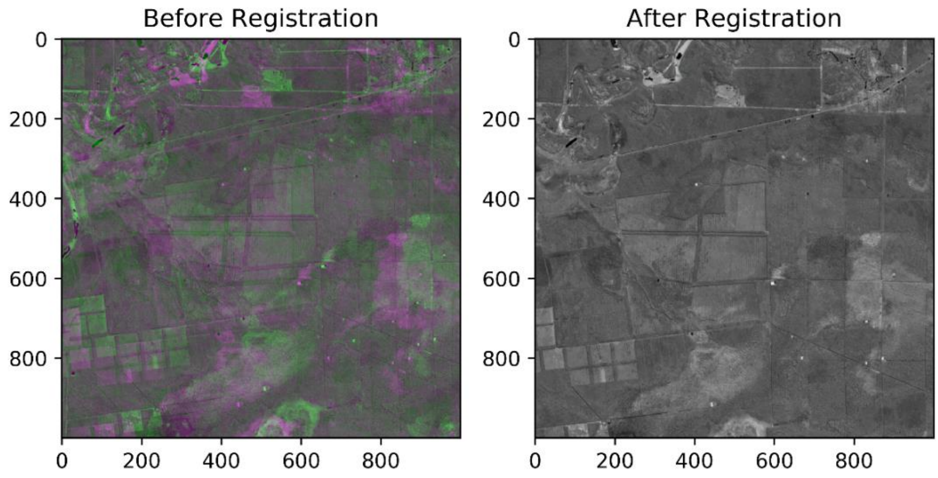

Figure 29. Synthetic Optical-to-Optical matching: false color composition of the images before (left) and after registration (right) based on rigid transformation, COBYLA, and cross-correlation metric.

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 30 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.

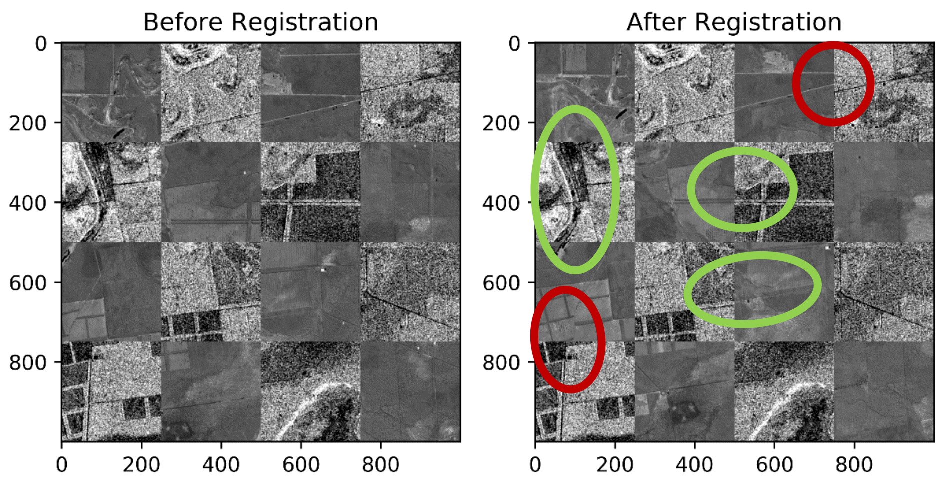

Figure 30. Real Optical-to-SAR matching: checkerboard of the images before (left) and after registration (right) based on rigid transformation, COBYLA, and mutual information metric.

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 31. For all the experiments within the synthetic optical-to-SAR matching, the COBYLA optimization method has been used.

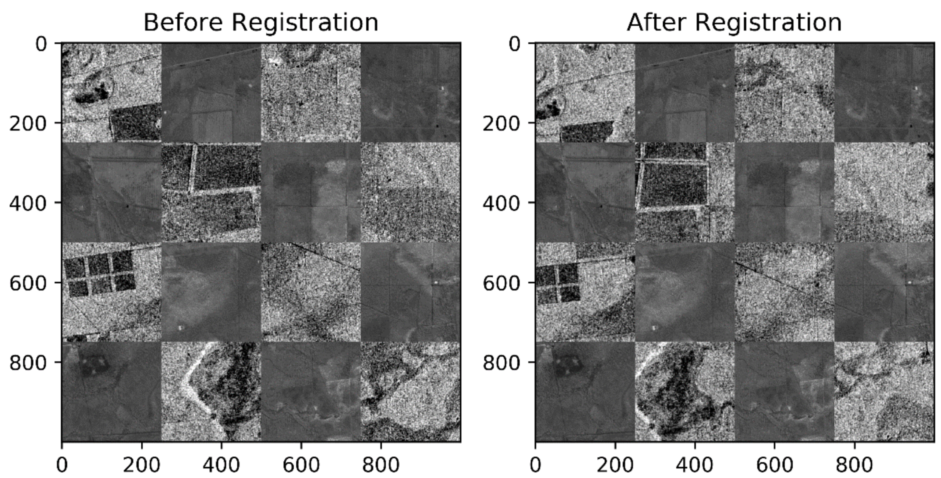

Figure 31. Synthetic Optical-to-SAR matching: checkerboard of the images before (left) and after registration (right) based on RST transformation, COBYLA, and mutual information metric.

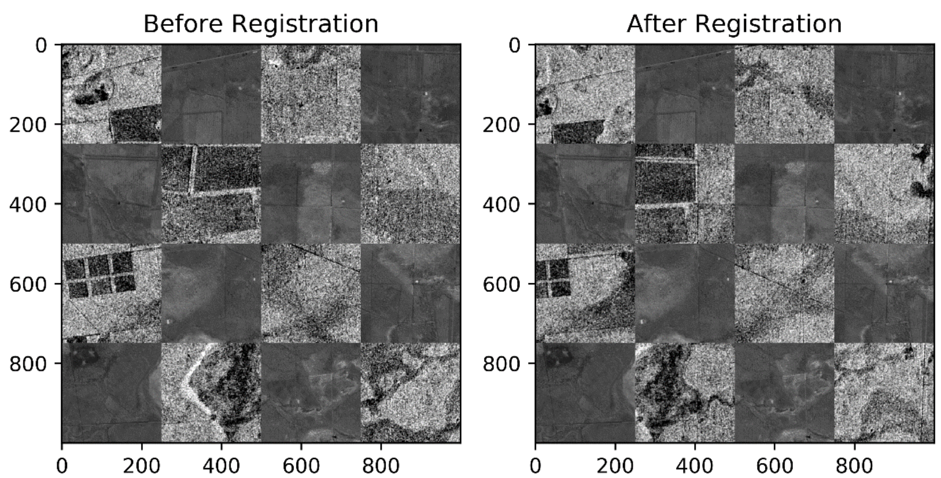

Figure 32. Synthetic Optical-to-SAR matching (scaling factor of 0.98): checkerboard of the images before (left) and after registration (right) based on RST transformation, COBYLA, and mutual information metric.

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 32), the rigid transformation still converged to a satisfactory result.

Scaling even further to 0.92 still granted acceptable results (Figure 33).

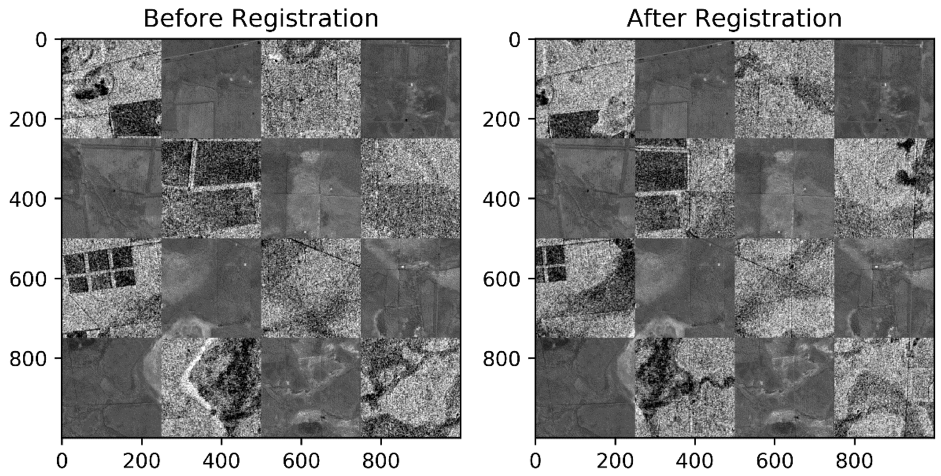

Figure 33. Synthetic Optical-to-SAR matching (scaling factor of 0.92): checkerboard of the images before (left) and after registration (right) based on RST transformation, COBYLA, and mutual information metric.

However, further tests failed to converge effectively. An example is reported in Figure 34, where the scaling factor was chosen equal to 0.9 and the other parameters were kept constant.

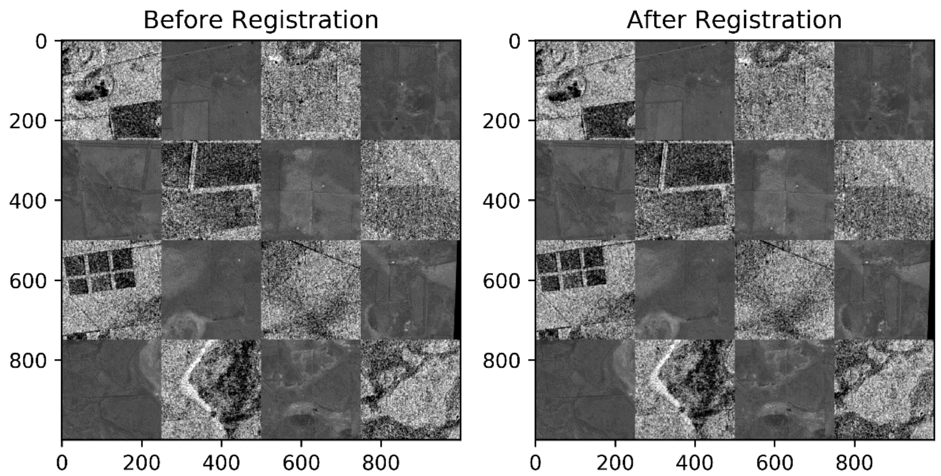

Figure 34. Synthetic Optical-to-SAR matching (scaling factor of 0.90): checkerboard of the images before (left) and after registration (right) based on RST transformation, COBYLA, and mutual information metric.

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 35 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.

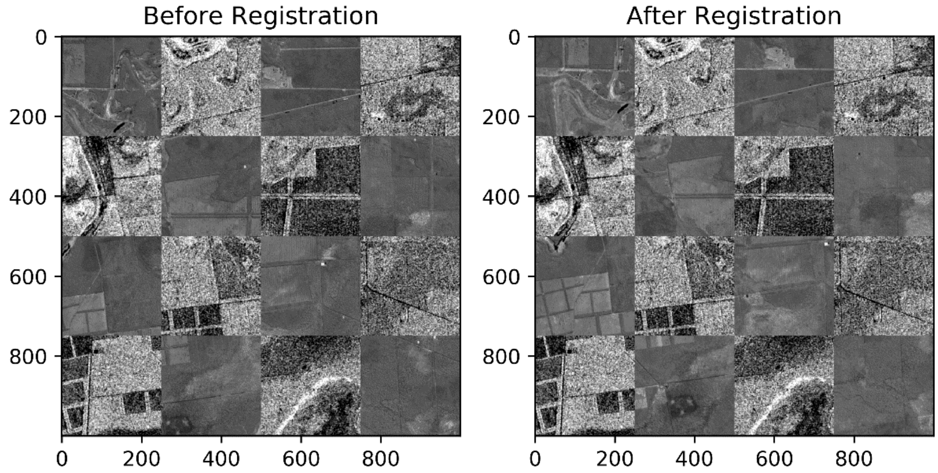

Figure 35. Real Optical-to-SAR matching – Area 1: checkerboard of the images before (left) and after registration (right).

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]

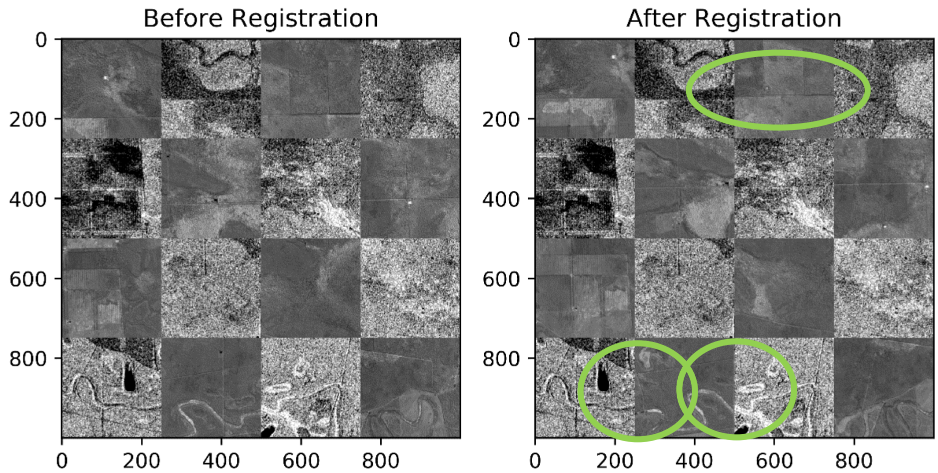

Figure 36. Real Optical-to-SAR matching – Area 2: checkerboard of the images before (left) and after registration (right).

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 36) 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]

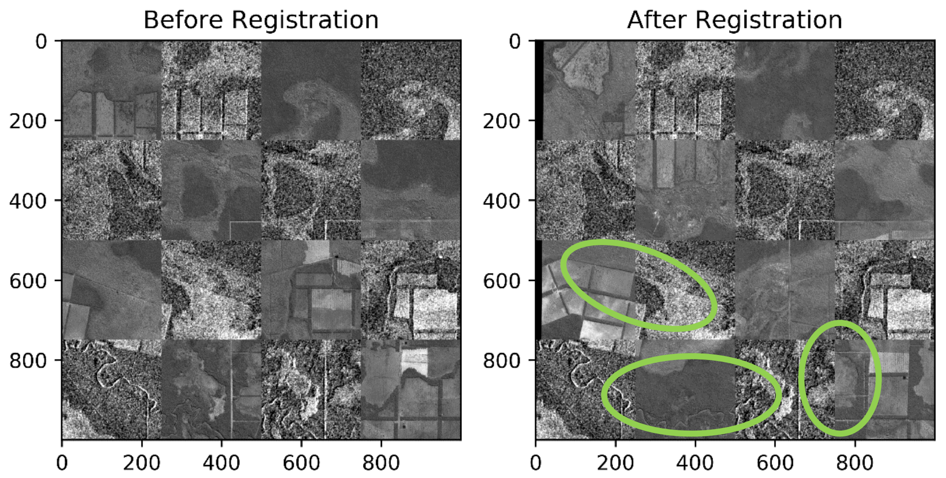

Figure 37 Real Optical-to-SAR matching – Area 3: checkerboard of the images before (left) and after registration (right) using COBYLA.

POWELL

Elapsed Time: 46.29 seconds

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

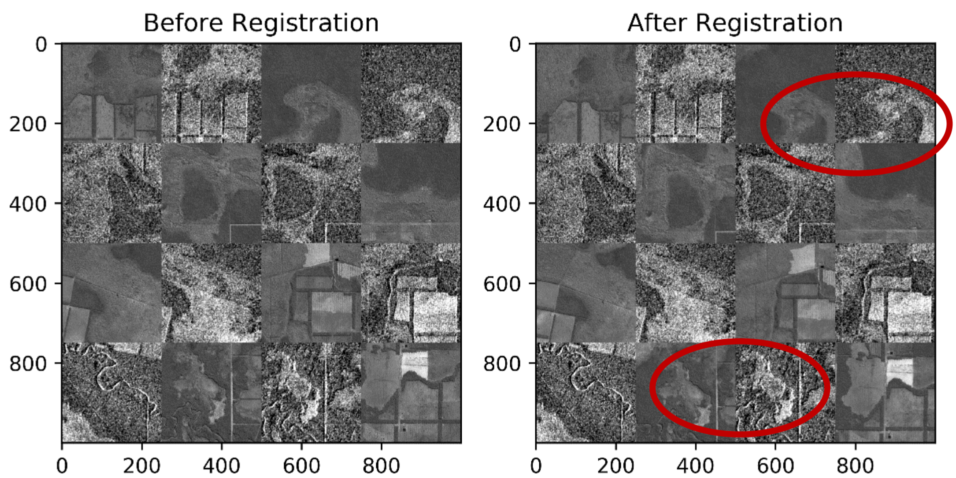

Figure 38 Real Optical-to-SAR matching – Area 3: checkerboard of the images before (left) and after registration (right) using Powell's algorithm.

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]

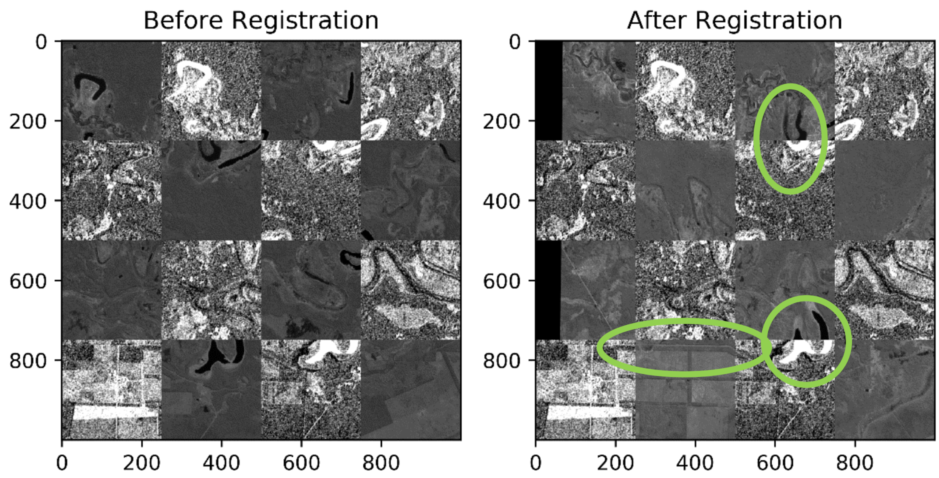

Figure 39 Real Optical-to-SAR matching – Area 4: checkerboard of the images before (left) and after registration (right) using COBYLA.

POWELL

Elapsed Time: 38.93 seconds

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

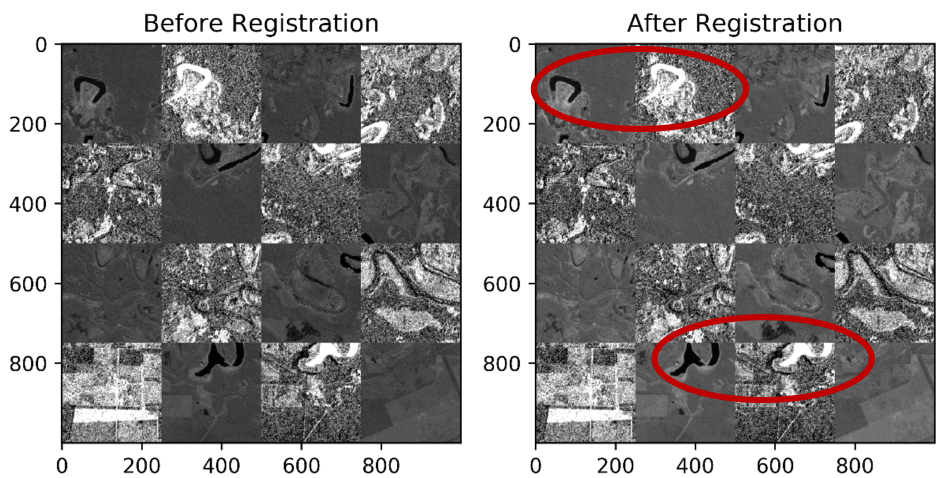

Figure 40 Real Optical-to-SAR matching – Area 4: checkerboard of the images before (left) and after registration (right) using Powell's algorithm.

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 8 classes, corresponding to numbers 1, 3, 5, 6, 7, 8, 10, 15 in the following table.

Value	Label	Color
0	No data	
1	Evergreen broadleaf tree	
2	Evergreen needleleaf tree	
3	Deciduous broadleaf tree	
4	Deciduous needleleaf tree	
5	Shrubland	
6	Permanent cropland	
7	Annual summer cropland	
8	Grassland	
9	Lichens and mosses	
10	Permanent water bodies	
11	Permanent snow and ice	
12	Beaches dunes and sands	
13	Bare soils	
14	Bare rock	
15	Built-up areas	

Figure 41 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).
- $\overline{\Omega}_s = \Omega_o - \Omega_s$, 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 jth information class ($j = 1,2,\dots,M$). The LOP functional can be expressed as:

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

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:

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

where the quantities associated with $\overline{\Omega}_s$ 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 42 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).

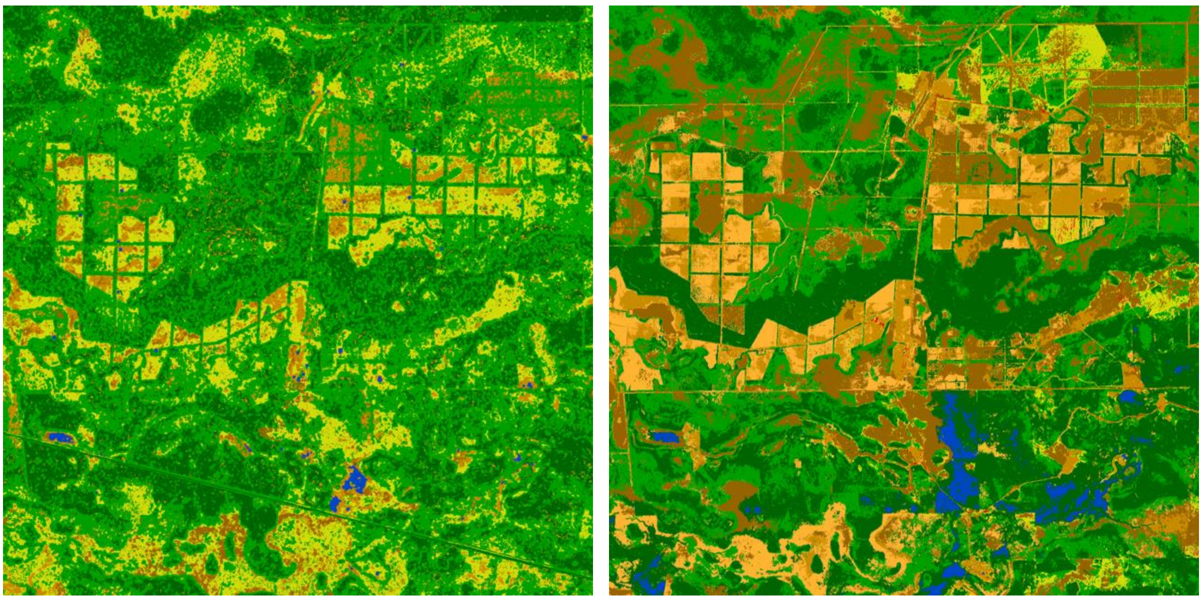

Figure 42 Details of the classification maps obtained from the classification of optical (left) and SAR data (right).

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 43 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 44.

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 45.

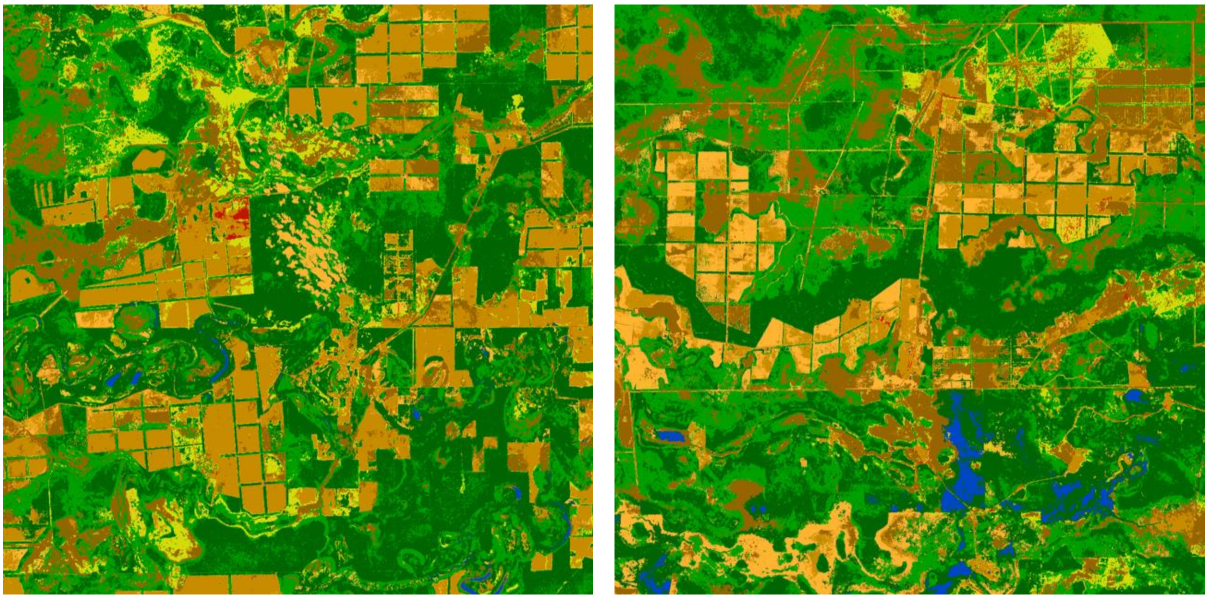

Figure 44 Details of the classification map obtained by LOP in the same areas as in Figure 42 (left) and Figure 43 (right; cloudy area on the right)

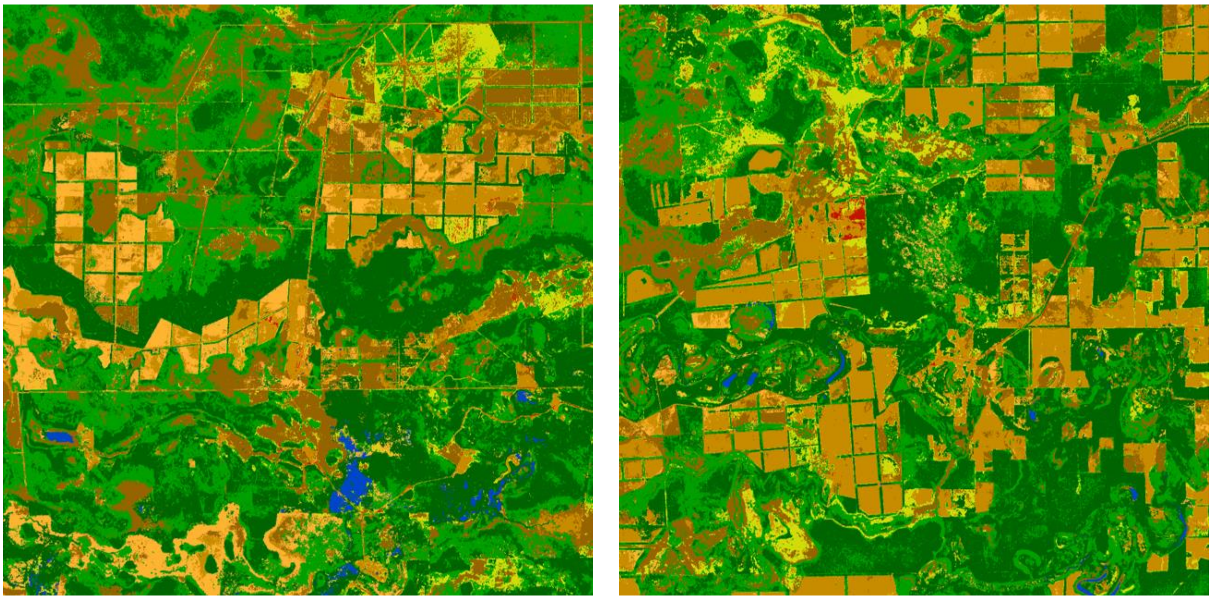

Figure 45 Details of the classification map obtained by LOGP in the same areas as in Figure 42 (left) and Figure 43 (right; cloudy area on the right)

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.

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 46.

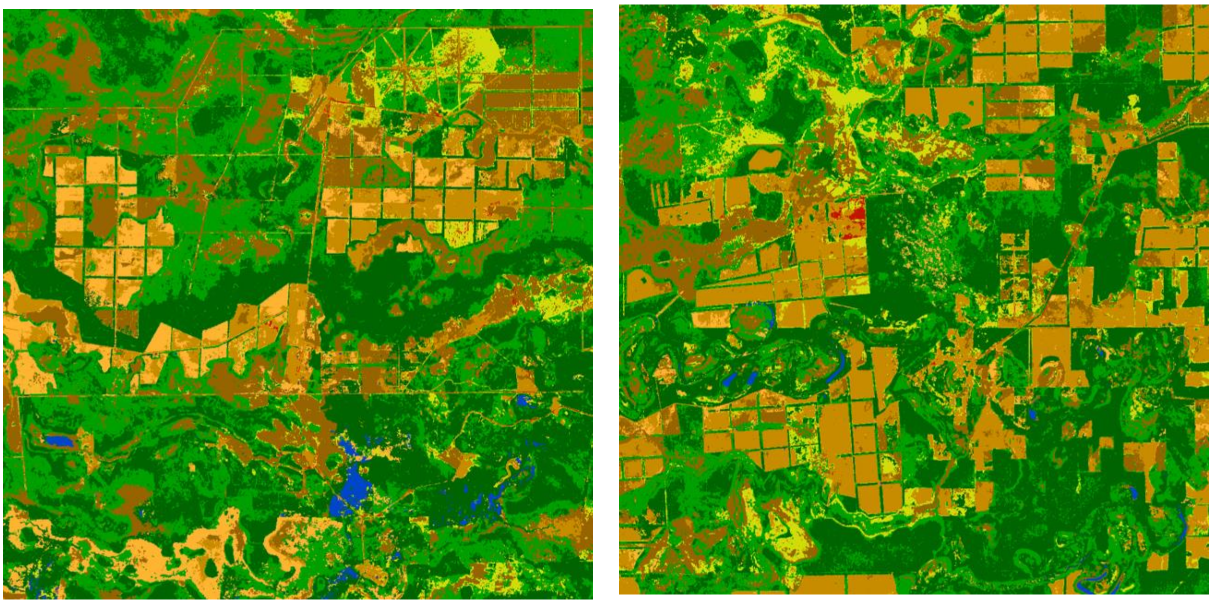

Figure 46. Details of the classification map obtained by MRF-ICM in the same areas as in Figure 42 (left) and Figure 43 (right; cloudy area on the right).

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.

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 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.

[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.

[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.

[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.

[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.

[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.

[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)*, 2013, pp. 1175–1177.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm,” 1976.

[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.

[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.” [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/0143116031000150077. [Accessed: 29-Oct-2019].

[20] ESA, “Land Cover CCI Product User Guide Version 2.0.” .

[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.

[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.

[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, 2018, p. 6.

[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)*, 2008, pp. 3853–3859.

[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.

[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.

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

[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.

[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.

[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.

[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.

[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.

[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.

[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.

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

[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.

[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.

[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.