Copernicus Access Platform Intermediate Layers Small Scale Demonstrator

D1.7 Use Case #1 Validation Report v2

Document Identification			
Status	Final	**Due Date**	30/09/2020
Version	1.0	**Submission Date**	21/09/2020

Related WP	WP1	**Document Reference**	D1.7
Related Deliverable(s)	D1.3; D1.5	**Dissemination Level (*)**	PU
Lead Participant	TerraNIS	**Lead Author**	Mohanad Albughdadi
Contributors	Anna PULAK-SIWIEC (SGIS)	**Reviewers**	Michelle Aubrun (TAS FR)

Keywords:
Agriculture, vineyards, Sentinel-1, Sentinel-2, change detection, data mining, data fusion, semantic search, validation

This document is issued within the frame and for the purpose of the CANDELA project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 776193. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission.

(*) Dissemination level: **PU**: Public, fully open, e.g. web; **CO:** Confidential, restricted under conditions set out in Model Grant Agreement; **CI:** Classified, **Int =** Internal Working Document, information as referred to in Commission Decision 2001/844/EC.

Document Information

List of Contributors	
Name	Partner
Mohanad Albughdadi	TNIS
Anna PULAK-SIWIEC	SGIS
Michelle Aubrun	TAS FR

Document History			
Version	Date	Change editors	Changes
0.1	01/09/2020	Mohanad ALBUGHDADI (TerraNIS)	First draft of the document
0.2	09/09/2020	Mohanad ALBUGHDADI (TerraNIS)	Second draft of the document
0.3	14/09/2020	Mohanad ALBUGHDADI (TerraNIS)	Incorporating TAS FR comments
0.4	15/09/2020	Mohanad ALBUGHDADI (TerraNIS)	Update DM and DF sections in the urban use case and incorporating DLR comments
1.0	21/09/2020	Jose Lorenzo (ATOS ES)	QA and Final revision before submission

Quality Control		
Role	Who (Partner short name)	Approval Date
Deliverable leader	Mohanad Albughdadi (TerraNIS)	18/09/2020
Quality manager	Juan Alonso (ATOS ES)	21/09/2020
Project Coordinator	Jose Lorenzo (ATOS ES)	21/09/2020

Table of Contents

List of Tables

List of Figures

List of Acronyms

Abbreviation / acronym	Description
CD	Change Detection
CESBIO	Center D'Etudes Spatiales de la Biosphere
CHIS	Color Histogram
CLC	Corine Land Cover
DF	Data Fusion
DIAS	Data and Information Access Service
DM	Data Mining
Dx.y	Deliverable number y belonging to WP 1
GLM	Gabor Linear Moments
GRD	Ground Range Detected
GT	Ground Truth
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared
OCS	Occupation du Sol (Land Cover)
RE	Red Edge
ROI	Region of Interest
RPG	Registre Parcellaire Graphique
S1	Sentinel-1
S2	Sentinel-2
SWIR	Short-Wave Infrared
VH	Vertical-Horizontal
VV	Vertical-Vertical
WLD	Weber Local Descriptors
WP	Work Package

Executive Summary

This document presents the final validation results using the CANDELA platform and its analytics tools. The conducted validation is divided into two main parts. The first one is a unitary validation where the user requirements reported in the traceability matrix [4] are tested. This unitary validation uses the Qase tool explained in D1.3 [1]. The four platform components are tested, namely, the platform itself, the optical change detection tool, the synthetic aperture radar change detection tool, the data mining and data fusion tool as well as the semantic search. For each component, a statistical representation is provided for the percentage of satisfied and unsatisfied requirements. The tested requirements are the outputs of first and second phases of requirements [2][3][5][6] and summarized in the traceability matrix [4].

The second part of the validation uses the two sub-use cases proposed for agriculture and macro-economic (see D1.1 [2] and D1.5 [3]) to validate all the tools. The first sub-use case aims at detecting the changes in vineyards after an episode of bad weather conditions. Such tools would allow reducing the analysis time and efforts. It also allows the farmers to justify the subsides they receive from insurance companies.

The second sub-use case aims at detecting the change in land cover especially those related to urban and agriculture. However, since the tools developed in the project are unsupervised, defining the transition type is impossible. Hence, a global change detection is studied instead.

The conducted validation aims at evaluating the performance of the tools using qualitative as well as quantitative results are provided when applicable where the steps and datasets required to reproduce the reported results are illustrated.

1 Introduction

CANDELA project aims at allowing the creation of value from Copernicus data through the provisioning of modeling and analytics tools given that the tasks of data collection, processing, storage and access will be provided by the Copernicus Data and Information Access Service (DIAS). These analytics tools adopt new developments in the domain of machine learning, data mining, data fusion and web semantics combined with the Copernicus data and non-Earth observation data sources to derive novel applications and services.

This document reports the results of the second and final validation phase of the CANDELA platform and its analytics tools after fine tuning them using the requirements of the use cases proposed in WP1 and more specifically in D1.1 [2] and D1.5 [3]. This validation verifies if all the user requirements (summarized in the traceability matrix [4]) are satisfied in all the platform functionalities and analytics tools.

1.1 Purpose of the document

The objective of this document is to validate the last versions of the platform and analytics tools against user requirements from the traceability matrix [4] using unitary tests and use cases.

1.2 Relation to other project work

The work in this document is related to both WP2 and WP3. The first component of this document validates the platform proposed in WP3, while the second component validates the analytic tools proposed in WP2.

1.3 Structure of the document

This document is structured in 3 major chapters

Chapter 2 presents the adopted test strategy and the used testing tools.

Chapter 3 describes the conducted unitary tests and their results.

Chapter 4 depicts the conducted validation using the two sub-use cases of agriculture and macro-economic.

Chapter 5 draws some conclusions.

2 Test Strategy

The testing plan adopted in this document validates the platform and the analytics tools against the user requirements in the Traceability matrix . Please note that all the requirements in the traceability matrix originated from D1.1[2], D1.2[5], D1.5[3] and D1.6[6]. Hence, a test case is designed for each of these requirements. Table 1 lists the versions of the components considered in this validation phase.

Table 1: Versions of the components considered for the second phase of validation

Component	Version
Platform	
Platform sub-component	V3
CreoDIAS	N/A
Analytic Tools	
Optical change detection sub-component	V3
SAR change detection sub-component	V3
Semantic IRIT	V2
Data mining and data fusion DLR	V3

Please note that this validation phase uses the same validation tool of phase 1[1]. This tool allows creating and composing test plans, cases and suites associated with each component.

[1] https://qase.io/

3 Unitary Test Results

User requirements were collected on two phases by the members of WP1. These requirements are the added-value of this work package and they aim at refining the CANDELA platform and analytics tools by adapting them to the proposed use cases.

Following the testing plan adopted in the first validation report, a unitary test is used to validate all the user requirements associated with CANDELA components. For the ease of exposure, four main test suites are designed, each one associated with a component to be tested. The four test suites are platform, optical change detection, SAR change detection, data mining and data fusion as well as semantic search. Note that the test cases in the test suites correspond to user requirements in the traceability matrix. Hence, these requirements will not be repeated but rather general statistics are provided and important points are clarified.

Please note that some user requirements in the traceability matrix are no longer considered in the framework of the project due to changes in the analytics tools or internal considerations due to the amount of efforts required to satisfy these requirements. This category of requirements will be referred to as "blocked" throughout the document.

3.1 Platform

This test suites consists of 15 test cases associated with the platform user requirements. Figure 1 shows the statistics of the test runs. This figure shows that 8 requirements are satisfied while the other 7 are no longer considered. To shed more light on these non-valid requirements, see the traceability matrix [4]. The most important of these non-considered requirements is the complexity of the jupyter environment for non-technical users. During discussions with the other partners, it was agreed that the CANDELA platform is destined with users with technical competencies that allow them to manipulate data using python and to adapt the proposed analytics tools to their needs.

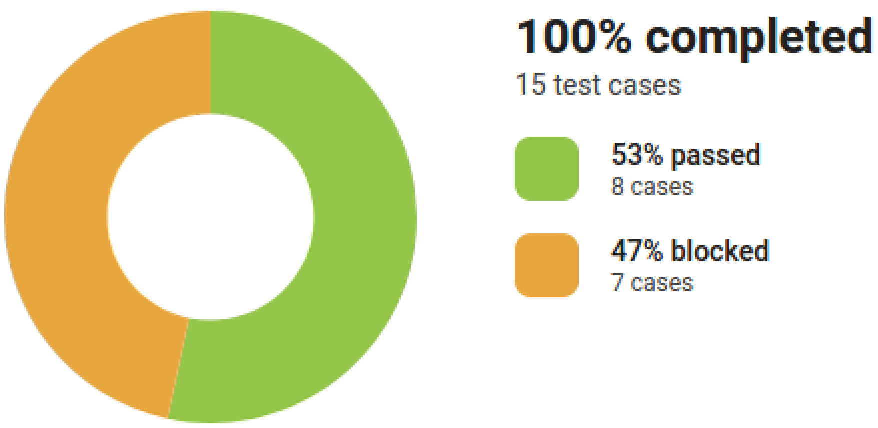


Figure 1: Statistics of the platform test runs

Another example of non-considered requirements is the functionality of searching a data catalog of the processed images using some criteria like the change-level, the Normalized Difference Vegetation Index (NDVI). This requirement was considered in the semantic search tool.

Even not considered in the traceability matrix, but it is worth noting that the availability of historical Sentinel images is a real challenge facing the development of platforms and applications based on data providers such as the DIASes. For instance, the majority of images up to 2019 are archived in the CreoDIAS, which makes it difficult for instance to detect changes that occur between historical and actual images.

3.2 Optical Change Detection

This test suite consists of 11 test cases associated with the optical change detection (CD) user requirements. Figure 2 depicts the statistics of the tests runs. This figure confirms that all the user requirements except one are satisfied. Indeed, this non-satisfied user requirement is concerned with the possibility of limiting the processing area based on land cover data provided by the user, DLR results of Corine Land Cover (CLC). However, it was discussed that this requirement will be considered if there is enough time to do it.

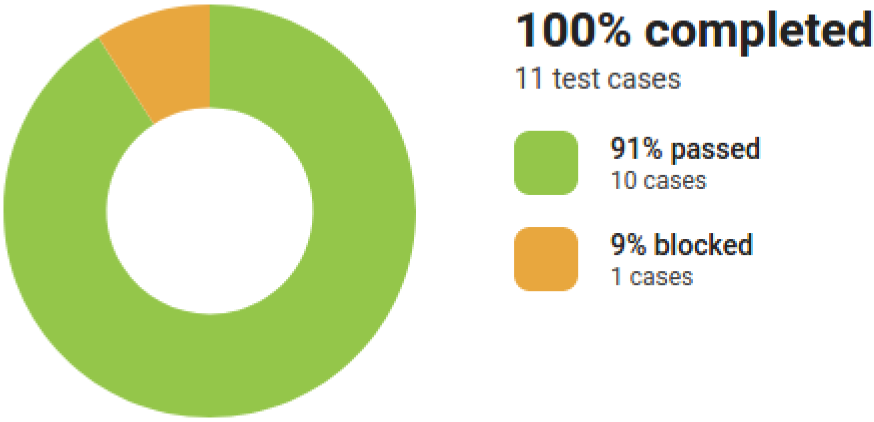


Figure 2: Statistics of the optical change detection test runs

3.3 SAR Change Detection

The SAR CD test suites consists of 13 test cases associated with the SAR CD user requirements. Figure 3 summarizes the statistics of the test runs. The figure concludes that 8 of the user requirements are satisfied and 5 are not considered. Three of the non-implemented requirements are related to the joint use of Sentinel-1 (S1) dual polarization for the analysis. Indeed, the current version of the SAR change detection tool analyses each band separately. The other non-considered requirements are associated with the ability to run the change detection tool on multiple images with different geographical extents and the assignment of a threshold for the change detection results. The later requirement is not valid any more as the current version of the change detection tool does not require a prefixed threshold as an input.

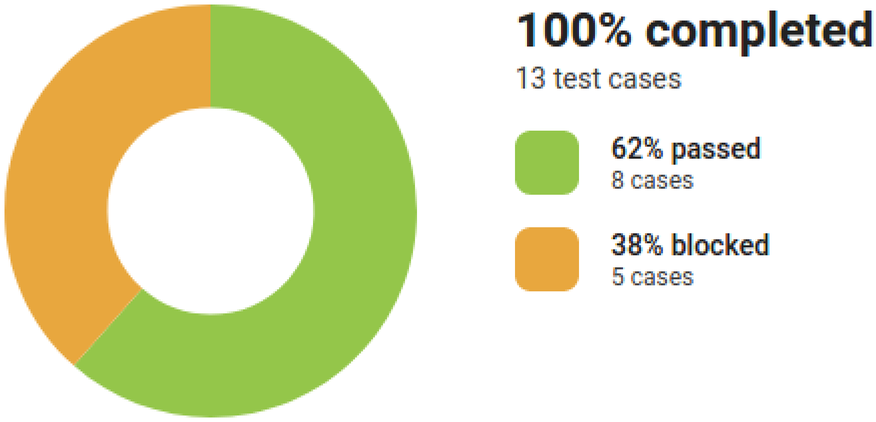


Figure 3: Statistics of the SAR change detection test runs

3.4 Data Mining and Data Fusion

Since the data mining and data fusion tools contain common components, it was decided to design one test suite for both components. The data mining (DM) and the data fusion (DF) suite consists of 12 test cases associated with the user requirements. Figure 4 shows the statistics of these tests runs. It is clear that all user requirements are met in these tools. Although one of the reviewer suggestions was to use the Big Earth Dataset for validation, this was discussed with the partners of the project and it was agreed that the data mining tool considers a different learning strategy (active learning), which only requires a very small training dataset to achieve satisfactory and comparable results with BigEarthNet [2]. The discussion on the comparison and validation criteria with BigEarthNet are presented in D2.2 data mining. Additionally, the use of the training data and labels from BigEarthNet will limit the classes to those ones in the dataset. However, the proposed data mining and data fusion tools allow custom user classes that can be of benefit for most of the use cases. The data mining tool was demonstrated on Big Data for an area larger than 1 Million km^2.

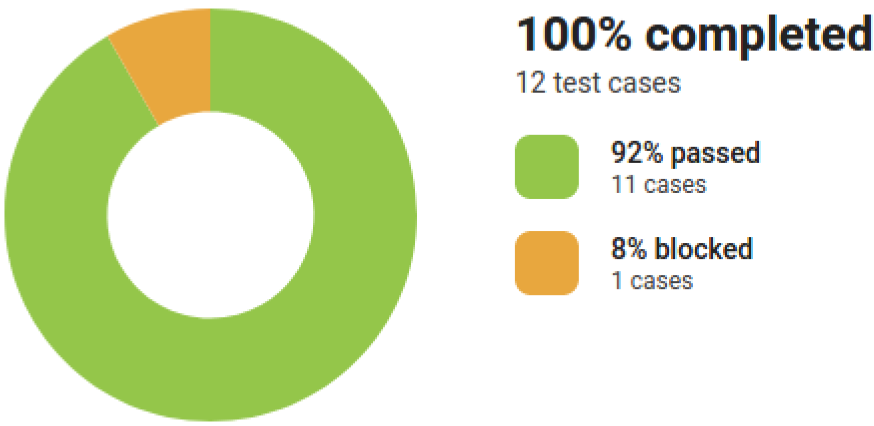


Figure 4: Statistics of the DM and DF test runs

[2] http://bigearth.net/

3.5 Semantic Search

The semantic search suite consists of 12 test cases as shown in Figure 5. This figure shows that all the user requirements are satisfied. One user requirement was not decided to be considered which is the search of the processed images indexed in the semantic catalog using the original metadata of the image.

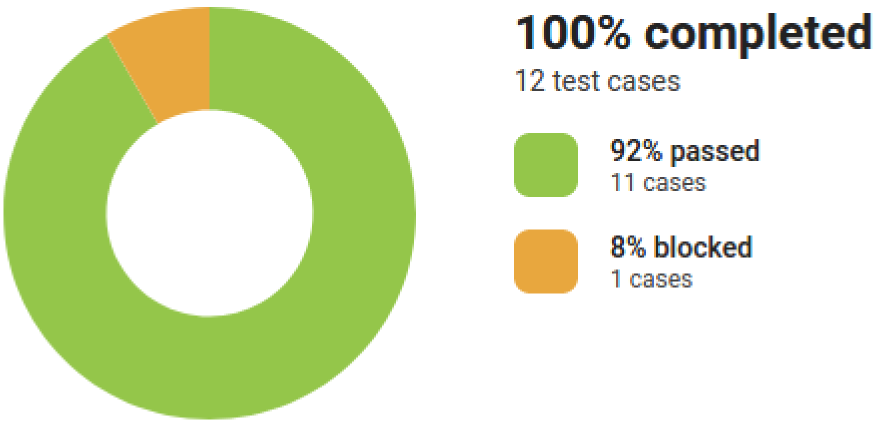


Figure 5: Statistics of the semantic search test runs

4 Use case Results

Two sub-use cases that have strong economic impacts and are aligned with Terranis' core business were proposed in this project. The first sub-use case is concerned with estimating the effect of natural hazards such as frost on vineyards. The second sub-use case is dedicated to explore the change of land cover especially urban expansion and agricultural land.

4.1 Change detection in vineyards

Severe weather conditions such as frost and hail can cause huge damage in the vineyards resulting in a significant loss in wine productions. Interprofessional associations and winemaking syndicates have shown their interest in estimating the level of damage in the affected vineyards after this kind of events, which help them to anticipate the damage and justify the subsidies they receive from the state and European insurances [1].

The following sections describe how to make use of the analytics tools provided in the CANDELA platform to achieve added value for this sub-use case. First of all, we will start by defining the region of interest (ROI), analysis dates and the required datasets. In a second step, we will show what kind of added values can be brought using each of the analytics tools. For each tool, the analysis results will be closely studied and analyzed to ground truth data and expert feedback if possible.

4.1.1 Study area and datasets

The studied area is located near Bordeaux in France. The bounding box of this area is (xMin, yMin -0.374998,44.5196: xMax, yMax 0.0950268,44.7971). This area is rich of vineyards and suffered an episode of bad weather conditions that damaged many of its vineyards between the 19th April and 29th April 2017. Figure 6 shows the extent of this ROI.

Two S1 Ground Range Detected (GRD) images were used for this sub-use case. These images were acquired on the 19th April (before) and the 29th April (after) 2017. Moreover, two Sentinel-2 (S2) images of the T30TYQ tile were considered for the analysis, even though the extent of the ROI is much smaller than the extent of the S2 tile. The first image was acquired before the episode on the 19th April 2017 while the other one was acquired after the episode on the 29th April 2017.

Figure 6: Extent of the ROI targeted for the vineyards sub-use case

For this specific sub-use case, external data sources are mandatory to define the results at the level of the vineyards. Indeed, the field boundaries coming from the French database, also known as Le Registre Parcellaire Graphique (RPG)[3], were used to quantify the results at the field-level. This field boundaries dataset contains 11355 fields. Further details on the use of these boundaries are provided in the following sections.

4.1.2 Complementarity of the analytics tools

The proposed analytics tools in the CANDELA project are based on different technologies such as deep learning, machine learning and semantic search among others. In the framework of the vineyards use case, these tools show complementarity which is shown in Figure 7.

The objective of the vineyards use case is to estimate the damage at the field-level due to bad weather conditions. This use case can make use of the proposed platform and analytics tools as follows.

Since one of the main objectives is to estimate the damage in an efficient way that does not require a lot of work force and time, one can realize that satellite imagery is the best candidate for such as job. The proposed use case relies on images coming from the CreoDIAS, where additional search tools built on top of the CreoDIAS API are provided on the CANDELA platform. These tools allow users to search for S1 and S2 images using different search criteria.

In a second step, this use case requires the geolocalization of the vineyards. Although auxiliary data sources are available to define the boundaries of the fields, sometimes these boundaries are not accurate enough or do not exist at all. Finding the vineyards can be accomplished using the data mining and/or the data fusion tools. These tools use the retrieved images to provide a customizable patch-level classification of the land cover. We prove in what follows that the data mining tool allows exploring specific classes for the use case such as the vineyards. Note that at this step, the precision of classification is not crucial as the idea is to discover zones of vineyards. Once these zones are found, the user can rely on photo-interpretation to draw the field boundaries and use them for further analysis.

The change detection tools discover zones of changes in a temporal series of images. These tools are used to analyze the images in order to extract the change information. Then, the field boundaries are

[3] https://www.data.gouv.fr/fr/datasets/registre-parcellaire-graphique-rpg-contours-des-parcelles-et-ilots-culturaux-et-leur-groupe-de-cultures-majoritaire/

used to quantify the changes at the field-level and to ignore changes coming from other land cover classes.

Finally, the user can make use of the semantic search tool to look for important changes in the fields or to verify the results obtained using the change detection models. Indeed, the semantic search tool allows estimating other vegetation indicators that are correlated with the vegetation status of the crops and hence with the change level. These indicators can be used to verify the results obtained using the change detection algorithm as will be shown later in the following sections.

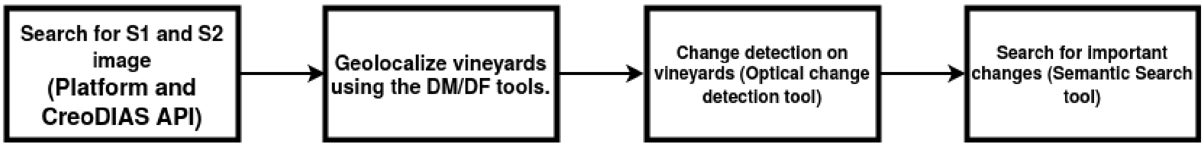


Figure 7: A diagram showing the complementarity between the analytics tools

4.1.3 Experiments using the analytics tools

4.1.3.1 Optical change detection analysis

In order to perform optical CD [7] on the CANDELA platform, multiple steps are required. These steps are details in what follows.

1. The first step consists of searching for images of interest for the ROI. This image search step is a functionality developed on the platform based on the CreoDIAS. The user is required to specify the range of dates, the level of processing, the name of the region or the bounding box directly and other metadata such as the cloud cover. The user can then filter the search results using prefixes or specific parts in the image names such as the tile.
2. It is worth noting that the change detection was performed on S2 L1C images as the L2A ones were not available on the CreoDIAS[4]. This may affect the results of the obtained results.
3. Once the user selects the images, symbolic links are created on his workspace in order to be accessed by the next processing steps.
4. In the third step, the user selects the bands to be processed. It is worth noting that bands with different resolutions can be used. The selection of these bands will be studied for the vineyards use case.
5. Next, metadata are extracted from the input images in order to be assigned to the output results.
6. Then, the user trains the change detection model using a reference image. This image is usually the first one in the multi-temporal series.
7. Finally, the changes are estimated using the trained model and the input images.

After the preparation of the input images, 3 band combinations were tested. The first combination considered the blue, green, red and near-infrared (NIR) bands. The second one considered the blue, green, red, NIR and Red Edge (RE) bands. Finally, the third combination considered the blue, green, red, NIR, RE and short-wave infrared (SWIR) bands. Indeed, the S2 bands at the NIR, RE and SWIR are known to be important for vegetation analysis and can reflect important characteristics of the vegetation phenological status.

[4] Major difference between L2A and L1C data is that the former is atmospherically corrected

After the training and prediction steps, the obtained change maps were cropped using the field boundaries in order to eliminate other land covers. Then, these clipped results are normalized using the 98th percentile in order to eliminate outlier values. Note that after the normalization, values greater than 1 are set to 1. This normalization step allows scaling the values of changes on vineyards between 0 and 1 as the change detection was run on all the land covers on the S2 image. Hence, high change probabilities are associated with stronger changes.

Figure 8, Figure 9,Figure 10 show the change detection maps obtained using the three aforementioned band combinations.

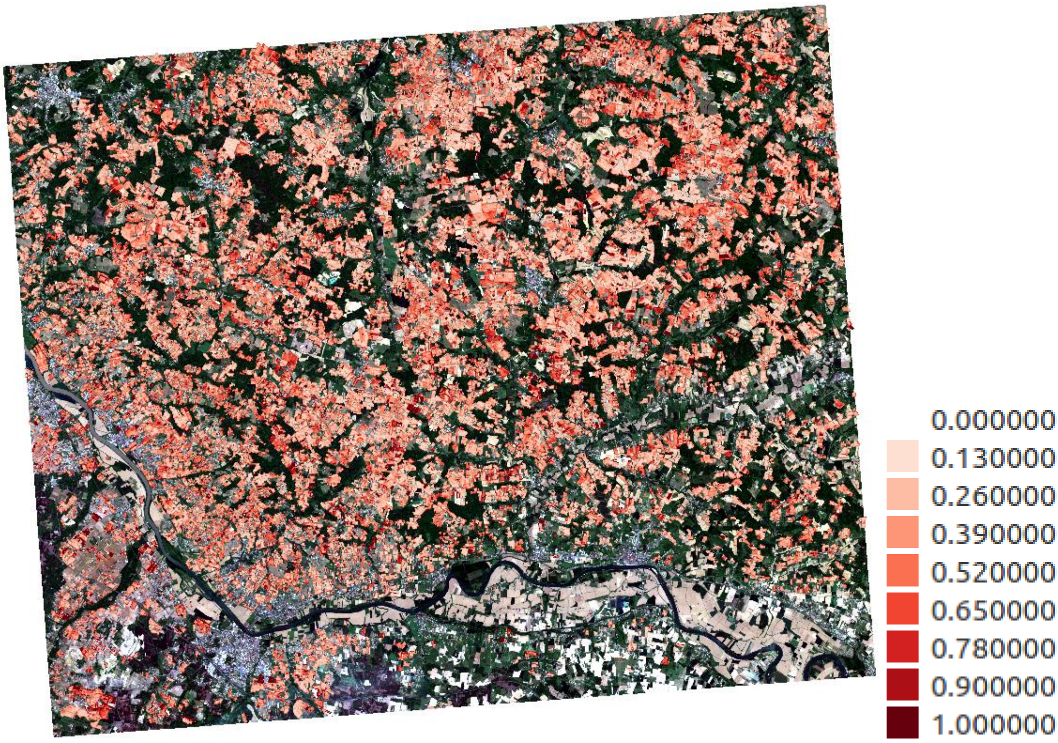


Figure 8: Change detection maps on the vineyards using the blue, red, green and NIR bands

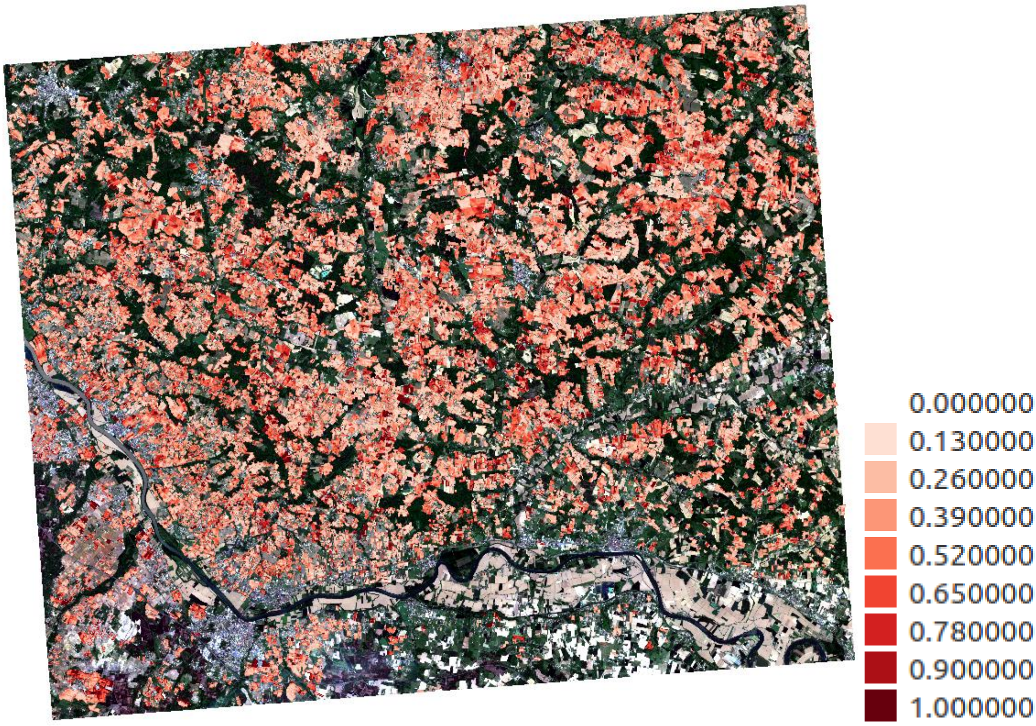


Figure 9: Change detection maps on the vineyards using the blue, red, green, NIR and RE bands

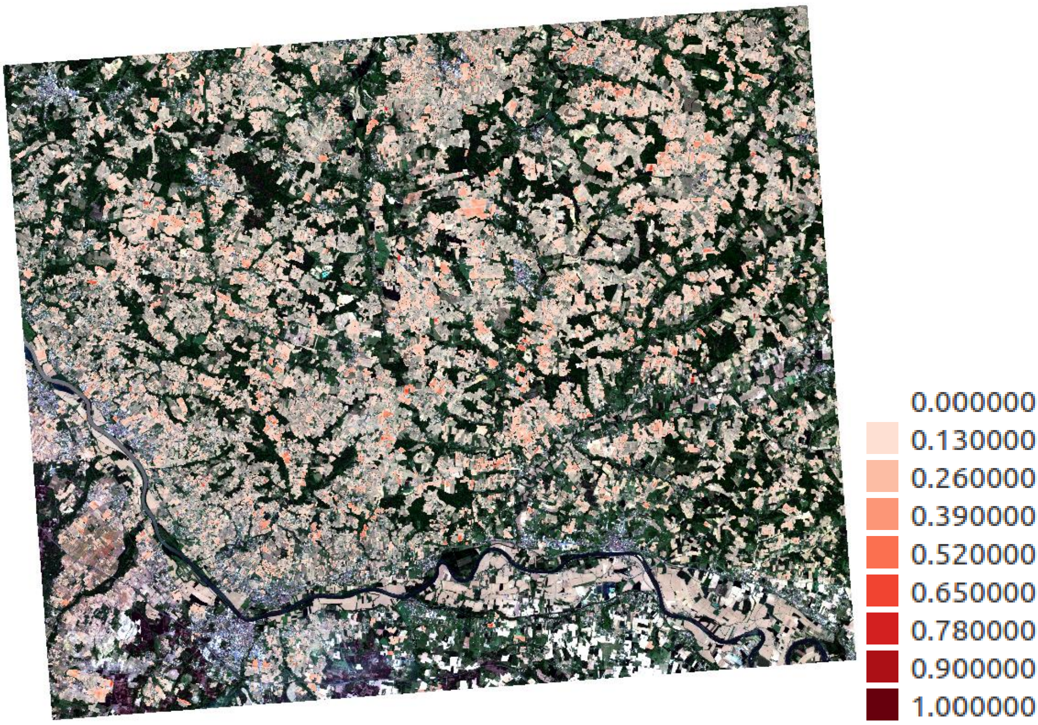


Figure 10: Change detection maps on the vineyards using the blue, red, green, NIR, RE and SWIR bands

One can notice that the results obtained 4 and 5 band combinations are similar to each other. However, using the 6 aforementioned bands reduces the absolute probability of change but keep the same trend.

To further analyze these results, two zones containing 109 and 253, respectively were annotated by an expert using photo interpretation and adequate vegetation indicators and biophysical parameters to assign a level of change for each field between 1 and 5, where 1 reflects a field with no change at all and 5 is a field with very high changes (see Figure 11).

The change detection maps obtained using the different band combinations and the optical change detection algorithm were further processed in order to quantify the change detection results at the field-level. To do so, the field boundaries were used to compute zonal statistics for each field (mean, minimum, maximum, median and standard deviation). In a second step, these zonal statistics were used as an input for the KMeans algorithm in order to cluster these results into 5 groups that correspond to the 5 levels of changes defined by the expert. The Kmeans ++ was used as an initialization for the KMeans algorithm in order to guarantee the quality of the obtained results. Finally, since the output labels of the KMeans algorithm are arbitrarily assigned (the order), these labels were adapted in order to be coherent with the change-level labels. At this point, the estimated change levels (obtained using the KMeans algorithm) and the GT labels can be compared using multiple criteria such as the accuracy, precision, recall and jaccard. Moreover, the correlation between the median values at the field-level and the expert GT was considered.

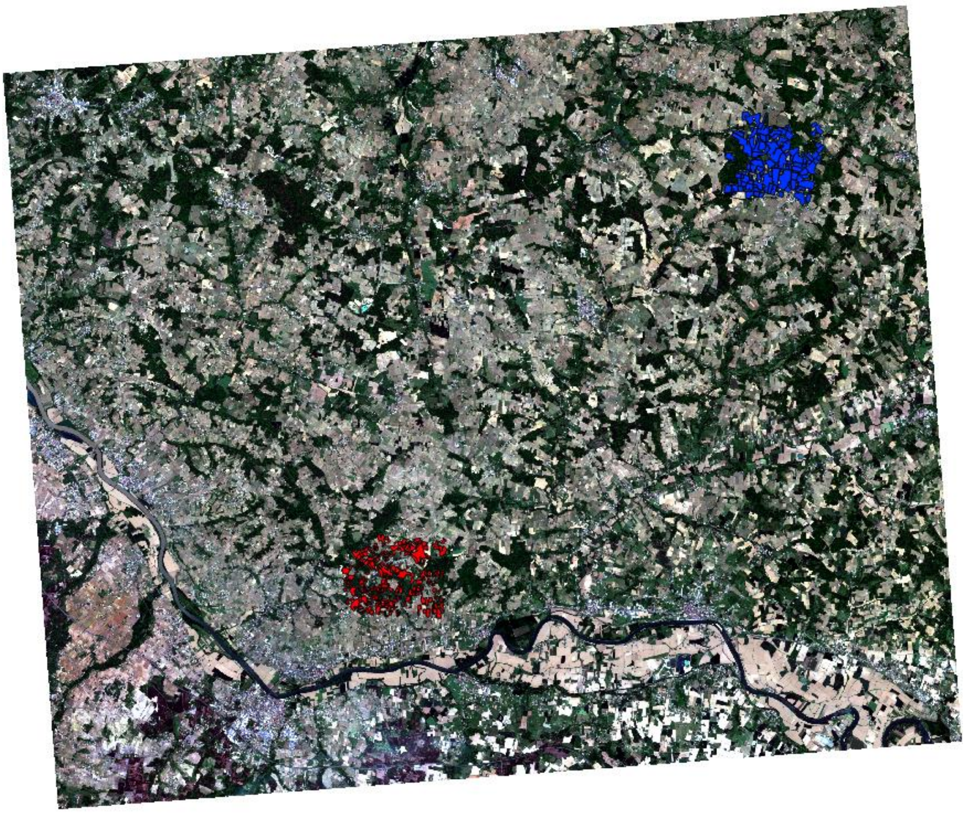


Figure 11: GT zones annotated by an expert. The blue fields are associated with zone 1, while the red ones are associated with zone 2

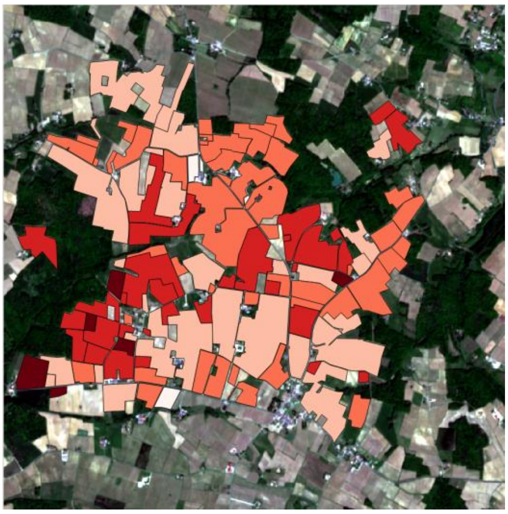
Additionally, the training and processing times were considered as part of the comparison to investigate the effect of increasing the number of bands on the computational complexity of the algorithm. Table 2 depicts the comparison between the 3 band combinations in terms of training time, processing time, accuracy, precision, recall and jaccard index when compared to the GT.

Table 2: Comparison between the three band combinations in terms of training and processing times (expressed in seconds) as well as other scoring criteria

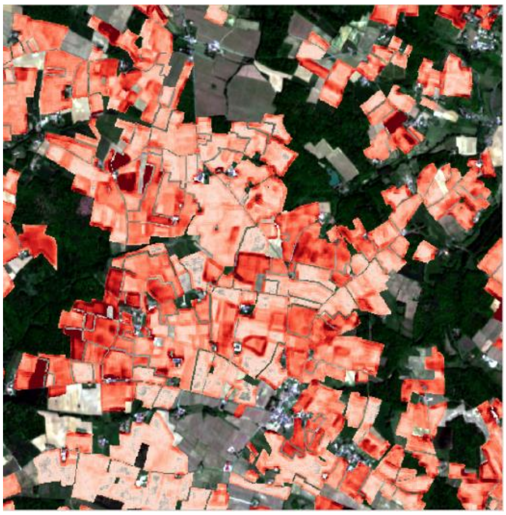
Criterion	*4 bands*	*5 bands*	*6 bands*
Training time (s)	188	187	187
Processing time (s)	371	217	248
Accuracy	0.42	**0.45**	0.40
Precision	0.31	**0.40**	0.35
Jaccard	0.28	**0.32**	0.25
Correlation	0.53	**0.55**	0.51

Regarding the training and processing times, it looks like increasing the input bands does not really affect the computational complexity of the algorithm. Regarding the performance of the algorithm compared to the GT, although the three band combinations provided similar results, the 5-band combination achieved the best performance when compared to the other two bands. Please note that the performance of the algorithm was also affected by the use of L1C images since L2A images of this zone are not available on the CreoDIAS.

Figure 12 and Figure 13 shows the GT compared to the different band combinations in zone 1 and zone 2, respectively. A visual comparison allows one to conclude the 4 and 5 band combinations are closer to the GT, which is coherent with the results reported in Table 2.

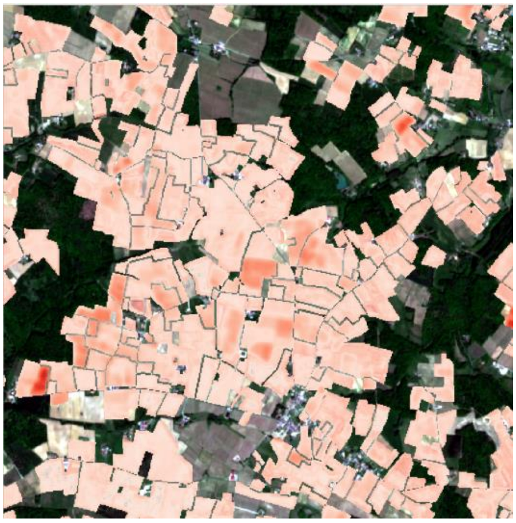


(a) Zone1 GT



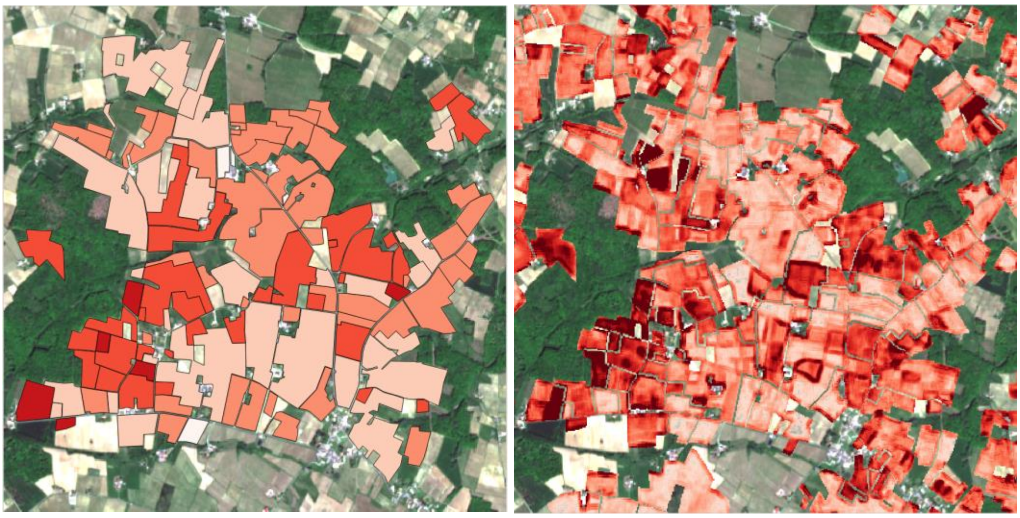
(b) Zone1 4 bands

(c) Zone1 5 bands

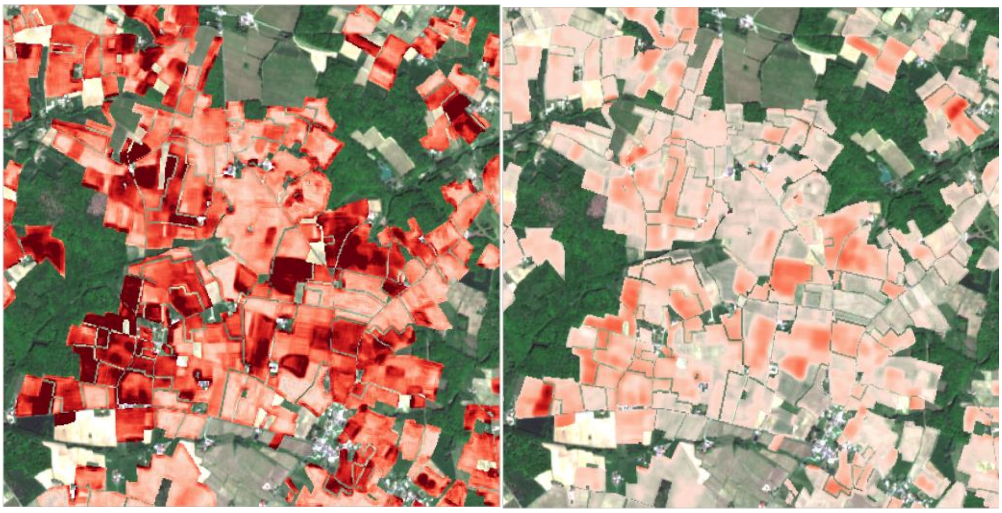


(d) Zone1 6 bands

Figure 12: Zone1 GT (a), 4-band (b), 5-band (c) and 6-band (d) combinations



(a) Zone2 GT (b) Zone2 4 bands



(c) Zone2 5 bands (d) Zone2 6 bands

Figure 13: Zone2 GT (a), 4-band (b), 5-band (c) and 6-band (d) combinations

4.1.3.2 SAR change detection analysis

Change detection using S1 images [7] was carried out for the vineyards sub-use case. Although the exploitation of SAR images for agriculture is challenging, the experiment aims at understanding if the proposed algorithm and S1 images can bring an added value for such use cases.

Multiple steps are required to run the SAR change detection on the CANDELA platform as explained below.

1. In a first step, the user search for images of interest using the search tool developed on top of the CreoDIAS API using the bounding box, the product type and the acquisition dates.
2. In a second step, the retrieved products are filtered in order to choose the ones of interest.
3. After, the preprocessing step is performed to prepare the input dataset. This step requires the north, south, east and west coordinates of the zone to be processed. Please note that it is

important to verify the results of the preprocessing step before proceeding to the next step as the temporal series of images must be covering the same geographical extent.

4. The processing step is then performed to estimate the change detection maps.

The two S1 images were chosen to cover zone 1 of the GT data generated by the expert. Although the covered zone is much larger than the GT zone, we only studied the zone covered in the GT dataset.

In terms of computational time, the preprocessing step took around 5.2 minutes while the processing step lasted for 3.1 minutes.

Figure 14 and Figure 15 show the change detection results using the VV and VH overlayed over the VV and VH polarizations, respectively, of the S1 image acquired on the 19th April 2017. The field boundaries were used to crop the obtained results in order to only highlight the changes at the field-level. Please note that in this analysis, no normalization was performed as the changes were mostly very strong on the vegetation areas, contrary to the case of the optical change detection.

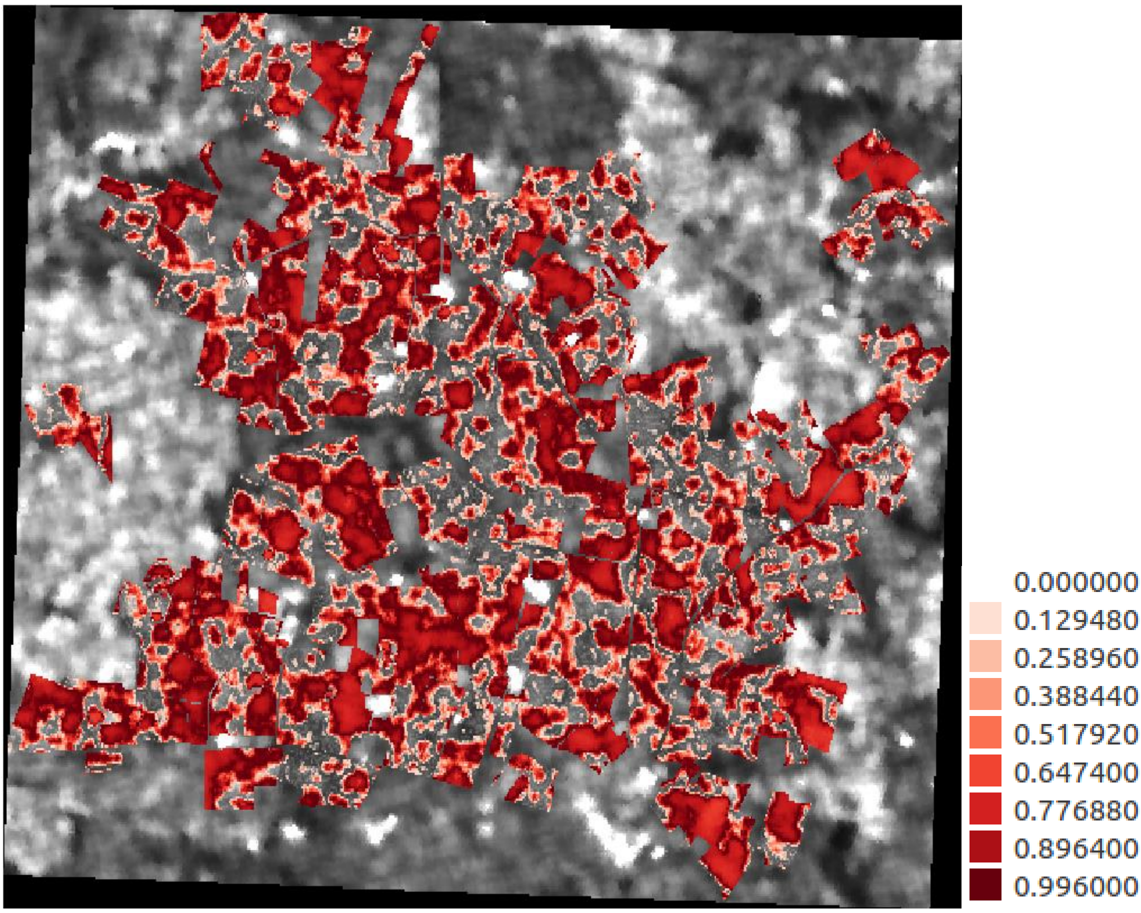


Figure 14: Change detection map obtained using the VV polarization cropped using the field boundaries

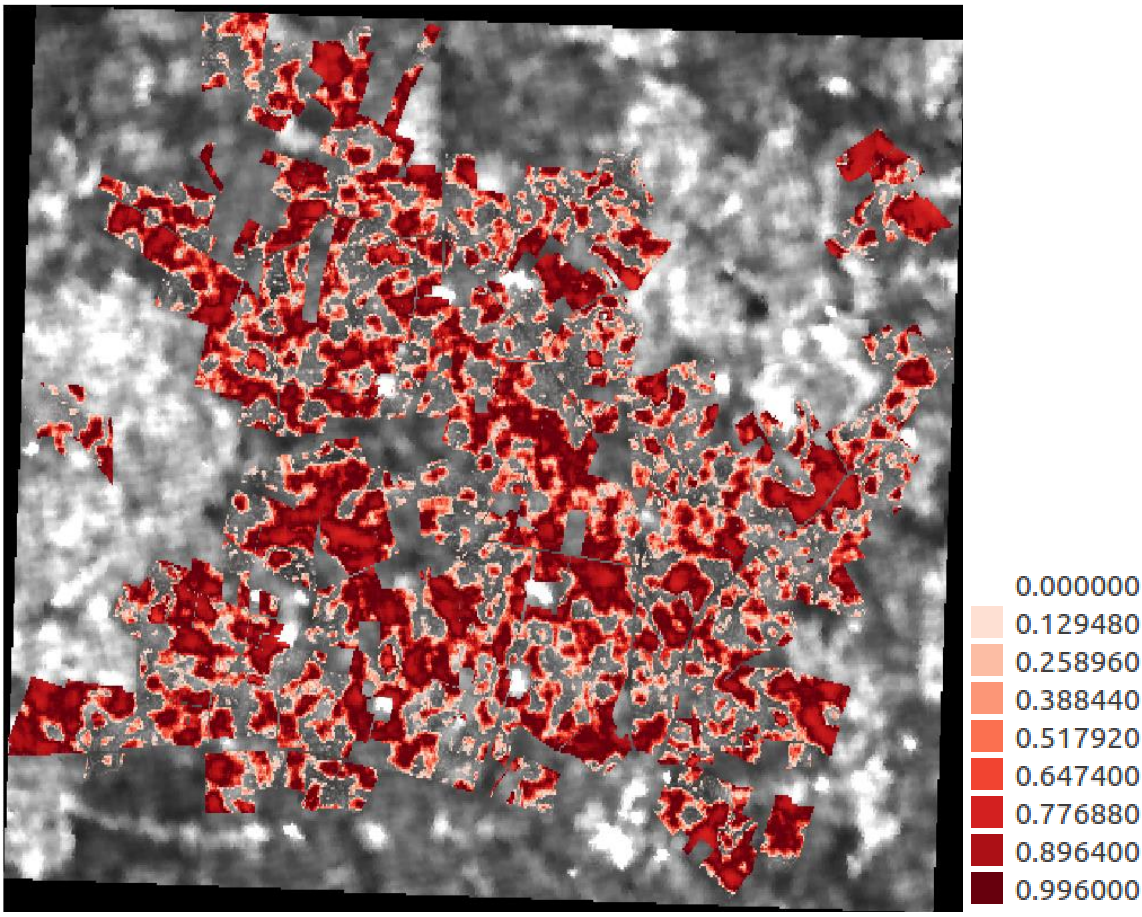


Figure 15: Change detection map obtained using the VH polarization cropped using the field boundaries

To further investigate these results, zonal statistics were computed for each field in order to be compared to the expert GT. Then, the correlation coefficient was estimated between the change detection at the field-level using the VV and VH polarizations and the GT. The reported correlation coefficient of –0.09 and –0.01 for VV and VH, respectively, did not indicate any relation between the changes estimated using S1 images and the real changes on vegetation. This can be explained by multiple factors. First of all, it is known that the SAR images are sensitive to water conditions. This period of time was characterized by bad weather conditions and hence the field can be stagnated with water. Moreover, after an episode of frost or hail where the crops are damaged, the soil becomes clearer and SAR images are also sensitive to soil conditions. These factors can cause very high changes between two images even in a short period of time.

4.1.3.3 Data mining analysis

The data mining [8] was conducted using a S2 image acquired on 19th April just before the episode of frost. The objective of this analysis is to verify if such a tool is able to correctly classify image patches and identify the location of vineyards. This analysis requires multiple steps to be run as follows.

1. The user starts by searching images for the ROI and filter the search results to select the required image.

2. The data mining module is then run on the CANDELA platform. This module requires defining some parameters such as the patch size, the bands to be processed and the feature types (see [9] for more details).
3. Once the processing is finished on the platform, the user switches to the KDD module installed on a local machine.
4. In the KDD module, the user selects the image processed on the platform, the feature type that will be used for the active learning.
5. Finally, the user starts to select positive and negative examples for one class of land cover at a time. This step ends when the user is satisfied with the results.

To analyze this image, the patch size was set to 120, the blue, red and green bands were selected for the analysis. Furthermore, four feature types were selected for the analysis, namely, Gabor Linear Moments (GLM), Weber Local Descriptors (WLD), Color Histogram (CHIS) and Gabor Local Cumulants (GLC). However, during the analysis, it was noted that training the active learning algorithm with the WLD feature is much easier than training it using any of the other features. Indeed, the training with the other features require many steps of user interaction, which might be difficult and time consuming. Hence, the results in what follows are only displayed for the WLD features.

Five classes were considered for this classification, namely, urban, vineyards, cropland, heathland forest and mixed forest. The learning of these classes adopted the strategy of one vs the rest. It means that each class was learned separately. Please note that this data mining module does not guarantee classifying all image patches as in classical multi-class classification methods. Figure 16 shows the classification results obtained at the level of the S2 tile. For this analysis, no ground truth was available for quantitative evaluation of the results. Hence, it was decided to use the pixel-wise land cover classification provided by the Centre D'Etudes Spatiales de la Biosphere (CESBIO) [5] for visual comparison. This classification is an annual classification that covers all French metropolitan territories. Figure 17 shows the pixel-wise classification for the S2 tile covered in this analysis obtained using the land cover classification of the CESBIO. The classes in this classification were adapted to match those used in the vineyards sub-use case. From Figures 15 and 16 one can notice a general agreement in the results even though no direct comparison could be performed. For instance, we can notice that urban areas are well detected in the patch classification results (see Figure 18). Moreover, the extent of the vineyards is also detected (see Figure 19). Regarding the forest areas and especially the heathland forests as well as cropland patches, it is clear that the data mining algorithm was able to detect these patches (see Figure 20).

[5] https://www.theia-land.fr/ceslist/ces-occupation-des-sols/

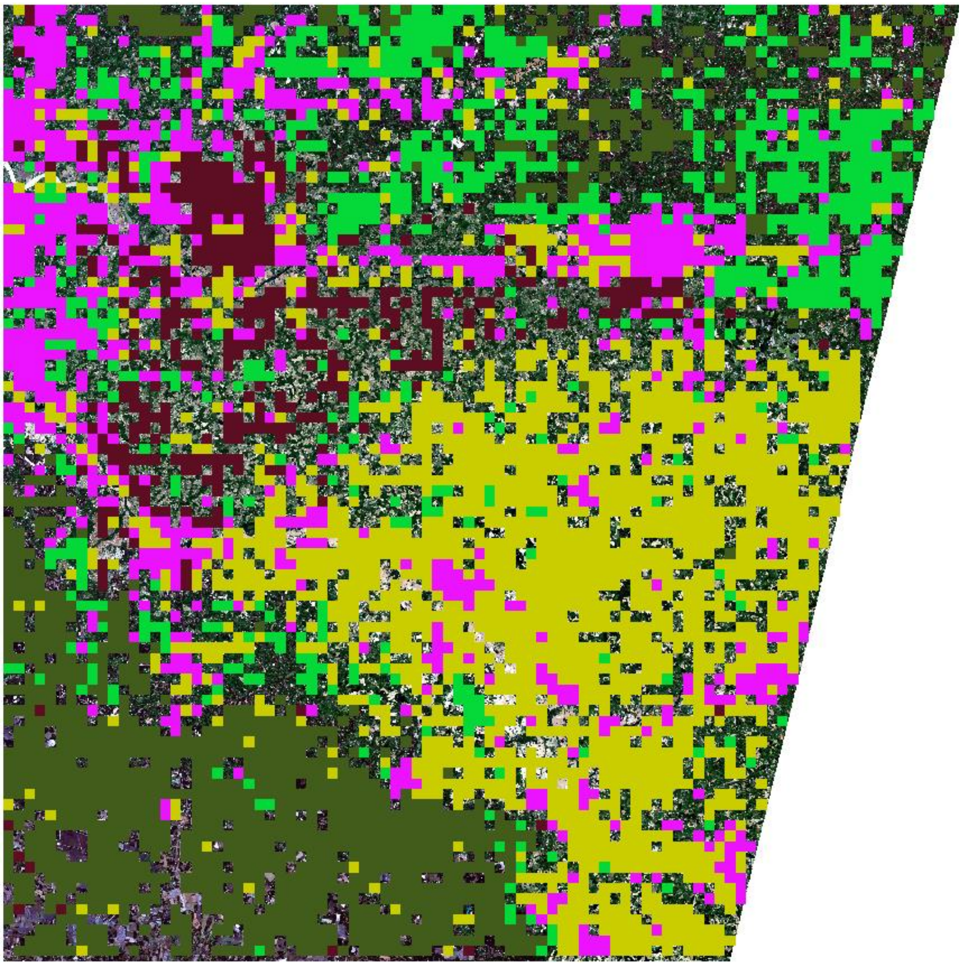


Figure 16: : Patch classification obtained using the data mining tool of the image acquired on the 19th April 2017. Vineyards, urban, cropland, heathland forest and mixed forest are highlighted in Bordeaux, pink, yellow, dark green and light green, respectively

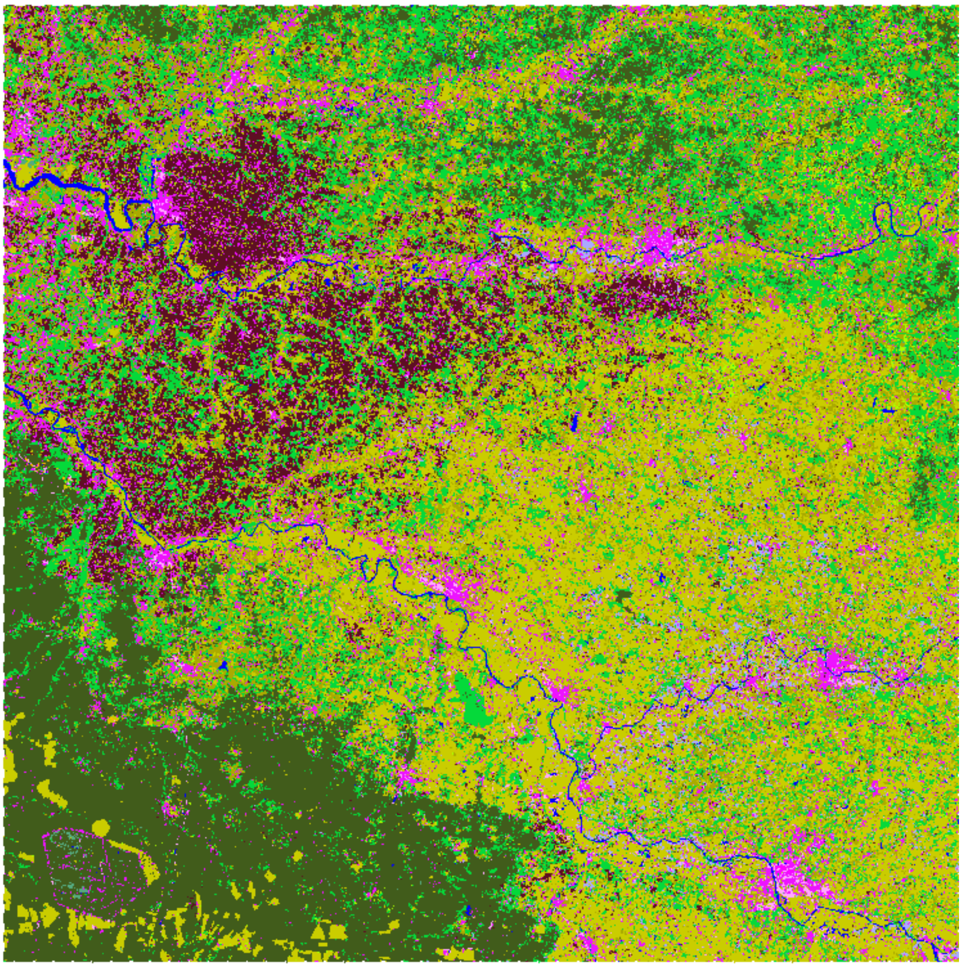
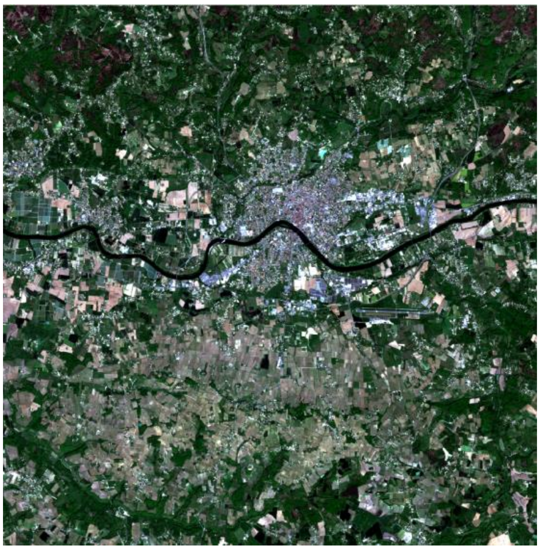
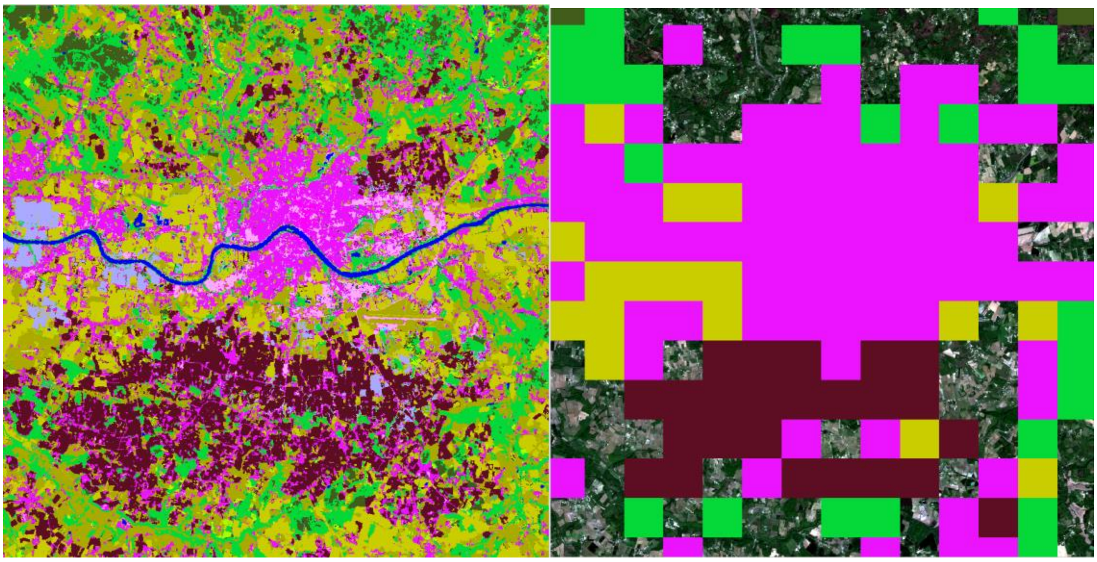


Figure 17: Pixel-wise classification obtained using the classification of land cover of 2017 provided by CESBIO. Vineyards, urban, cropland, heathland forest and mixed forest are highlighted in Bordeaux, pink, yellow, dark green and light green, respectively



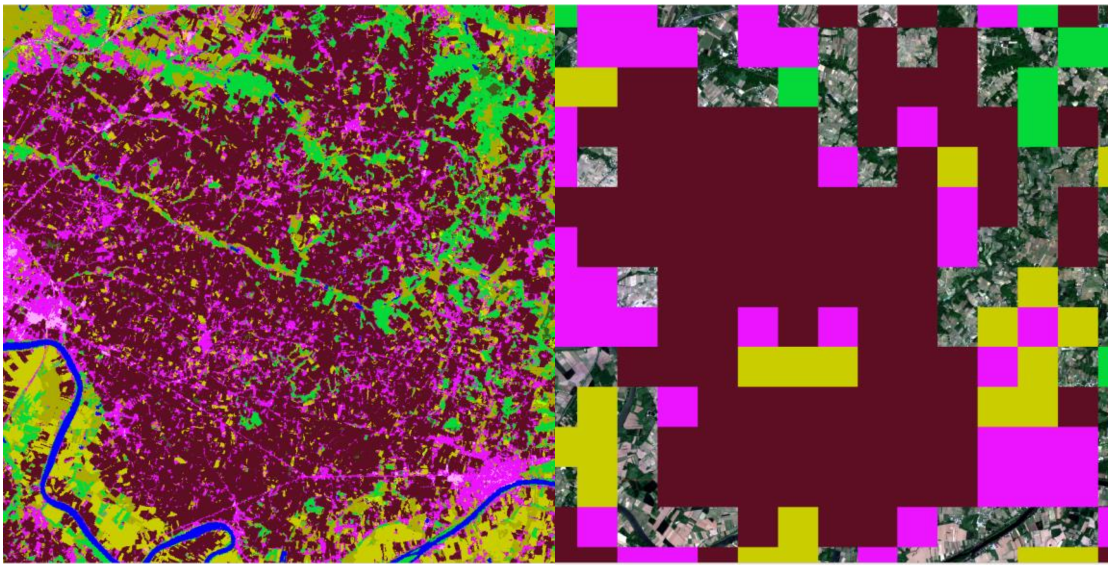
(a) S2 image



(b) Pixel-wise classification (c) Patch-level classification

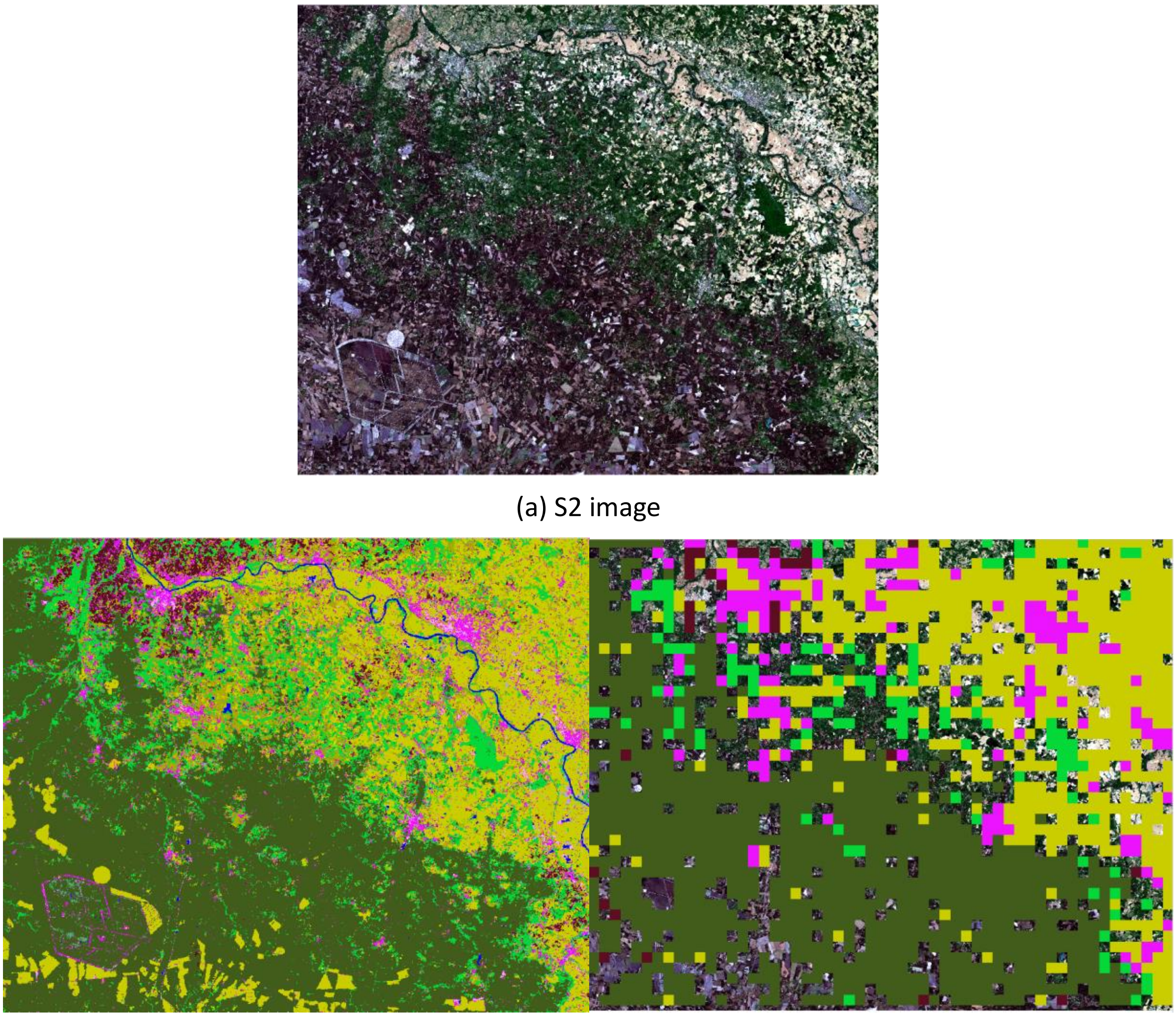
Figure 18: S2 image (a), pixel-wise classification (b) and patch-level classification (c) of an urban area. Vineyards, urban, cropland, heathland forest and mixed forest are highlighted in Bordeaux, pink, yellow, dark green and light green, respectively

(a) S2 image



(b) Pixel-wise classification (c) Patch-level classification

Figure 19: S2 image (a), pixel-wise classification (b) and patch-level classification (c) of a vineyards area. Vineyards, urban, cropland, heathland forest and mixed forest are highlighted in Bordeaux, pink, yellow, dark green and light green, respectively



(a) S2 image

(b) Pixel-wise classification (c) Patch-level classification

Figure 20: S2 image (a), pixel-wise classification (b) and patch-level classification (c) of a heathland forest and cropland area. Vineyards, urban, cropland, heathland forest and mixed forest are highlighted in Bordeaux, pink, yellow, dark green and light green, respectively

It is worth noting that the purpose of the data mining tool is to explore the image datasets rather than providing a full classified scene of patches. In the context of the vineyards use case, the data mining tool can help to easily locate vineyards areas where an intervention is needed in order to perform further analysis. This can be of crucial importance when no auxiliary datasets are available such as the field boundaries.

4.1.3.4 Data fusion analysis

Data fusion analysis was not conducted for this sub-use case. Indeed, and as shown in the next sub-use case, data fusion can be difficult to train with the fixed feature types and especially for a the vineyards. Please note that the idea is not to apply all tools for every use case but to use those tools that are adapted to the use case.

4.1.3.5 Semantic search

The semantic search tool [10] was used to analyze the fields of interest using two S2 images acquired on the 19th April and 29th April 2017. To run the tool, the following steps are required.

1. The user search for S2 images using the ROI. Then, these results are filtered to find the images before and after the episode of bad weather conditions.
2. The user can estimate the NDVI or directly use the change detection maps generated using one of the change detection algorithms for the next step.
3. The user runs the triplification step using the raster file generated using the previous step and another auxiliary file (shapefile) that defines the boundaries of objects. For the vineyards use case, the shapefile comes from the RPG to define the field boundaries.
4. The user can then search the results using change levels, NDVI levels, search weather data from stations nearby the fields of interest, search satellite images associated with the processed fields, etc.

The field boundaries of the vineyards extracted from the RPG were used along with the NDVI calculated from the two S2 images. For this use case, the semantic search was used to verify the results obtained using the change detection tools since the data and results are already known for us as a user. However, we think that the semantic search tool can be of great interests for users that do not perform the analysis themselves but those who are looking for specific fields with some characteristics such as field with a lot of changes, the satellite images that cover these fields, etc.

Two main experiments were conducted with the results obtained using the semantic search tool. The first experiment aimed at finding all the vineyards that had high NDVI values before the episode of bad weather conditions (using the S2 image acquired on the 19th April). In a second step, we retrieved the NDVI levels for the same fields after the episode (using the S2 image acquired on the 29th April). We compared the retrieved NDVI levels for each field and kept those that showed decreased NDVI levels. It is worth noting that the NDVI is correlated with the vegetation status in the fields and hence must be correlated with the change detection level. The semantic search using the images of the 19th April resulted in 37 fields with high NDVI values. 29 fields of them showed lower NDVI levels after the bad weather conditions. Figure 21 shows the 29 fields showing changes after the episode. A zoom on some fields is shown in Figure 22 as well as their change detection results obtained using the optical change detection tool.

Another experiment was conducted by looking for fields with medium NDVI levels before the episode and compared to their NDVI levels after the episode. Since the number of fields is very large for this use case, we only studies 100 fields. 46 of these fields showed NDVI levels lower than their levels before the episode (see Figure 23). Figure 24 shows a sample of these fields and their optical change detection results. These figures show that there were actual changes detected by the optical change detection tool in the fields retrieved using the semantic search tool.

These results show that the semantic search can be complementary to the optical change detection tool in the sense that it can be used to search for fields suffering from a lot of damages. But we also prove that using some field experience, other indicators of the vegetation vigor can be used and compared to the change detection results.

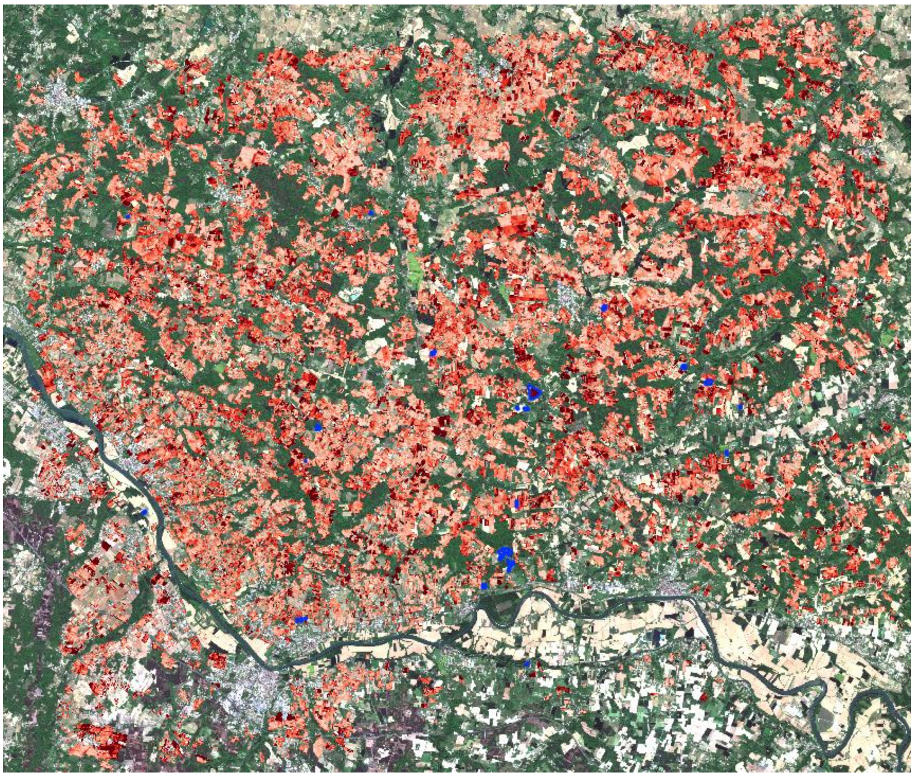


Figure 21: Vineyards with high NDVI before the bad weather conditions that showed damage detected by the semantic search tool

Figure 22: Fields with high NDVI levels before the episode and lower NDVI levels after the episode overlayed on the optical change detection results

Figure 23: Vineyards with medium NDVI before the bad weather conditions that showed damage detected by the semantic search tool

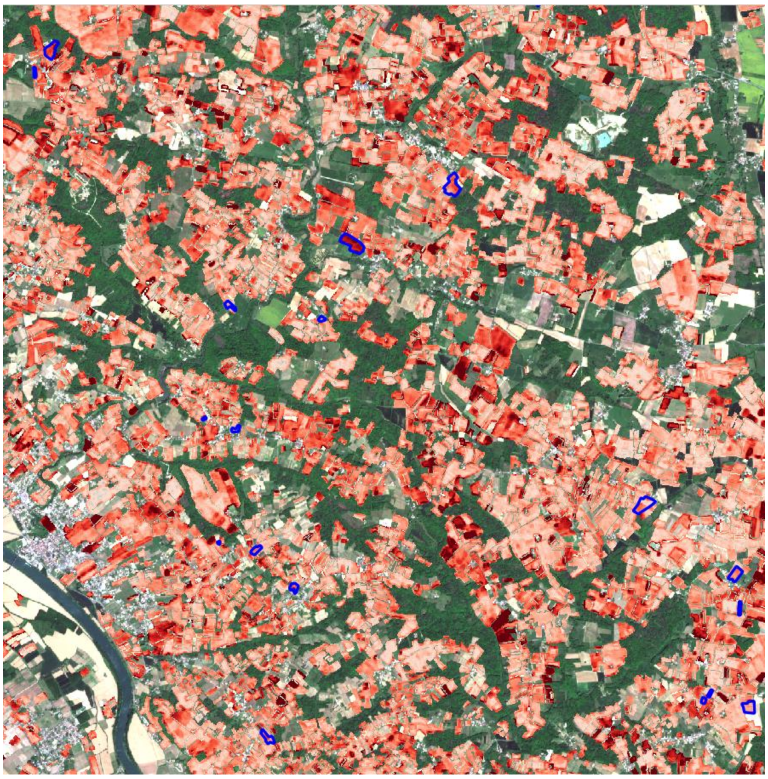


Figure 24: Fields with medium NDVI levels before the episode and lower NDVI levels after the episode overlayed on the optical change detection results

4.2 Urban expansion and agriculture

This sub-use case aims at investigating the changes related to land cover such as urban and agriculture land expansion and shrinking. This kind of changes can be of interest for decision makers in order to make informed decisions related to land management.

The following sections describe how to apply the analytics tools proposed in the CANDELA project on S1 and S2 data in order to extract useful indicators. First of all, we will start by defining the ROI, the analysis dates and the required datasets. In a second step, we will demonstrate the added value brought by using each of the analytics tools. For each tool, the results will be studied and compared to ground truth data if available.

4.2.1 Study area and datasets

The ROI is centred around Bordeaux in France. The bounding box of this area is (xMin,yMin -0.703984,44.7483: xMax,yMax -0.46762,44.9188). Bordeaux is one of the largest cities in France and it is known to be surrounded with agricultural areas.

Two S1 images were used for the SAR change detection. The first image was acquired on the 4th August 2017 while the second one was acquired on the 23rd of August 2018.

For S2 images, the dates to be analyzed are the 2nd of August 2017, the 23th of February 2019 and the 22nd of August 2019. Please note that the idea was to analyze a temporal series of images > 2. Three S2 images of the T30TXQ tile were considered for the analysis. Although the extent of the ROI is much smaller than the tile extent, the analysis was performed on the whole S2 tile. Please note that no S2 images were available for this tile in 2018 on the CreoDIAS. For the semantic search analysis, the boundaries of the French communes are used [6].

4.2.2 Complementarity of the analytics tools

In what follows, we will show how the platform and the analytics tools can be used for such as use case. Please note that this use case has different particularities when compared to the one of the vineyards. For instance, the time period between images is longer since urban changes or agriculture changes are considered as significant ones and do not happen in a short period of time. Hence, it is more reasonable to use a temporal series of images for this use case, which is possible due to the image search tool implemented on the CANDELA platform. Moreover, defining the transition type is much more complex in this use case. Indeed, for the vineyards we knew that changes happen only in vegetation, which is not the case for this use case. However, the change detection tools are able to capture generic changes. Defining the change type can then be done using photo-interpretation or the data mining and/or data fusion tools in a condition of having very large areas of changes. These tools work at the patch-level and are not adapted to pixel-level analysis. Finally, the semantic search tool is used to analyze the vegetation in the communes covering the ROI, where a statistical analysis is conducted between images of different years to show complementary information at a larger granularity than the changes at the pixel-level.

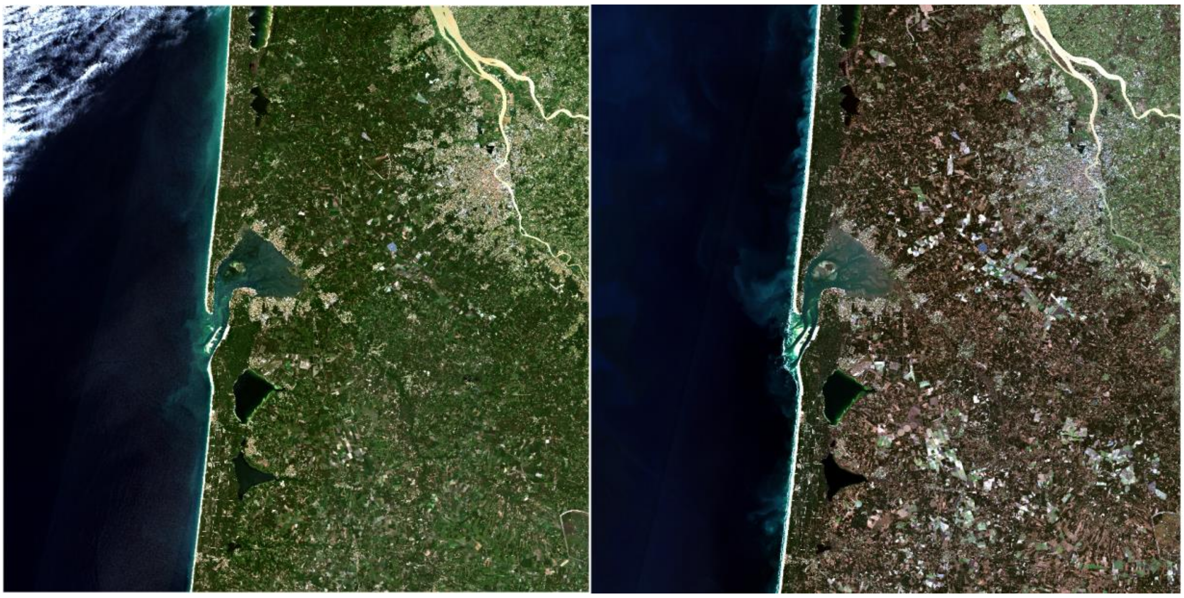
[6] https://www.data.gouv.fr/fr/datasets/decoupage-administratif-communal-francais-issu-d-openstreetmap/

4.2.3 Experiments using the analytics tools

4.2.3.1 Optical change detection analysis

The analysis using the optical change detection algorithm was conducted using the same steps in Section 4.1.2.1. For this analysis, the blue, red, green and NIR bands of the L2A products were used. The change detection algorithm was trained on the image of the 2nd of August 2017, where it took around 187 seconds for the training step. Then, it was applied to detect the changes resulting in a change detection map for all the S2 tile. The processing time for 3 S2 images took around 400 seconds.

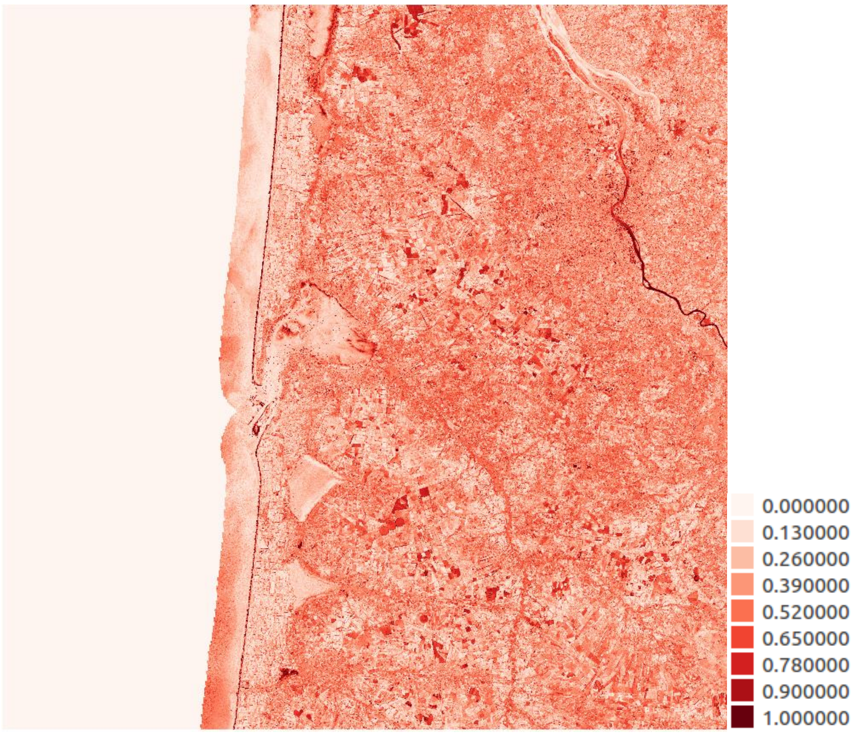
Figure 25 and Figure 26 show the 3 S2 images and the two change detection maps between the pairs of images, respectively. The change detection maps show the importance of selecting the proper dates for changes detection. For instance, and when looking at the change detection results between the 2nd August 2017 and 23rd February 2019, we can notice that more changes were detected compared to the change detection map between the 2nd August 2017 and the 22nd August 2019. Indeed, selecting the same period of time during the year can guarantee that more similarities exist for the same land cover.



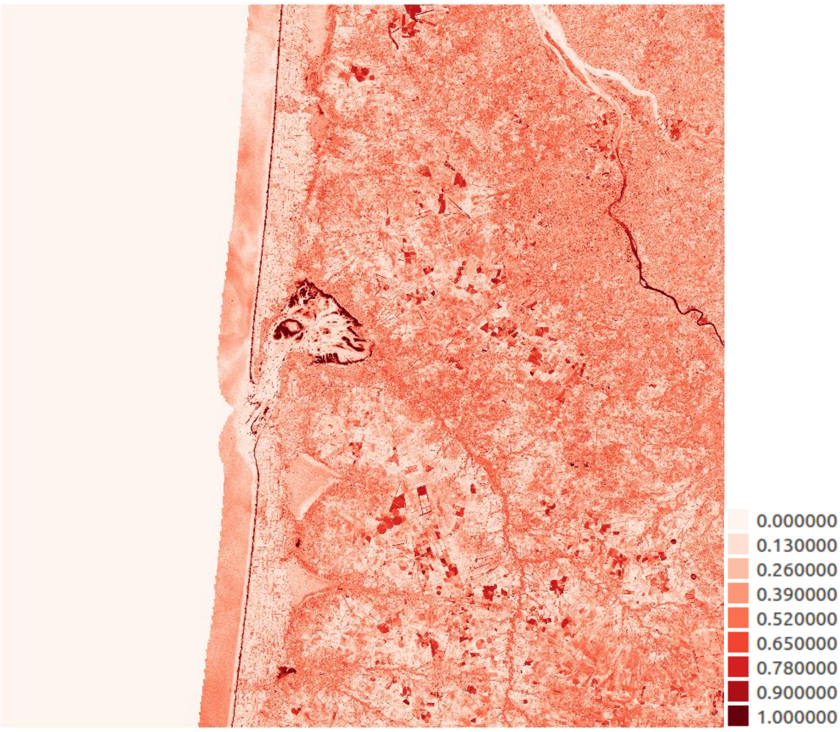
(a) 2nd August 2017 (b) 23rd February 2019

(c) 22nd August 2019

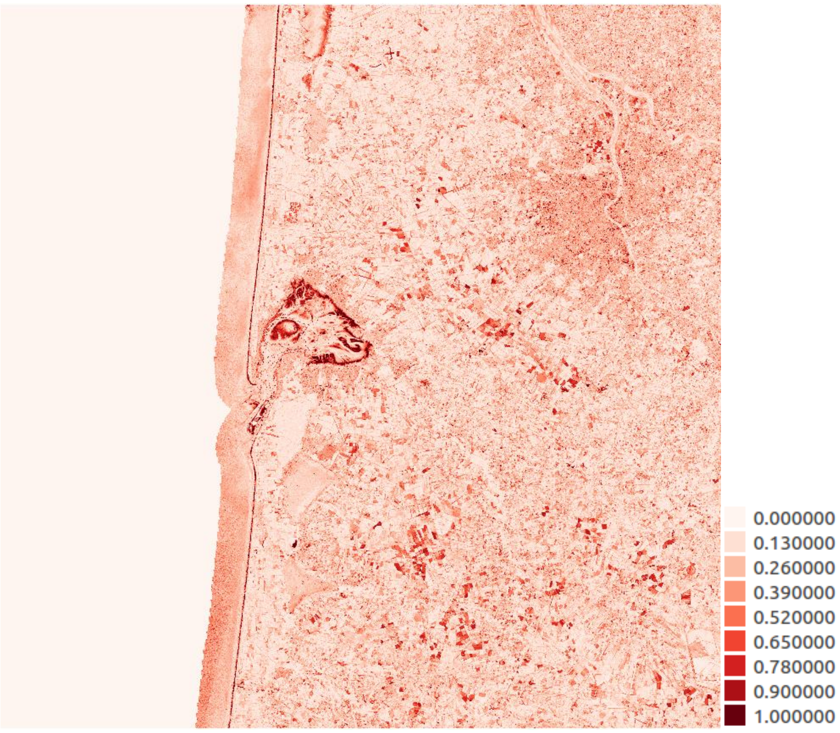
Figure 25: S2 images used for the change detection analysis in the urban sub-use case



(a) Change detection between the 2nd August 2017 and 23rd February 2019



(b) Change detection between the 23rd February 2019 and 22nd August 2019

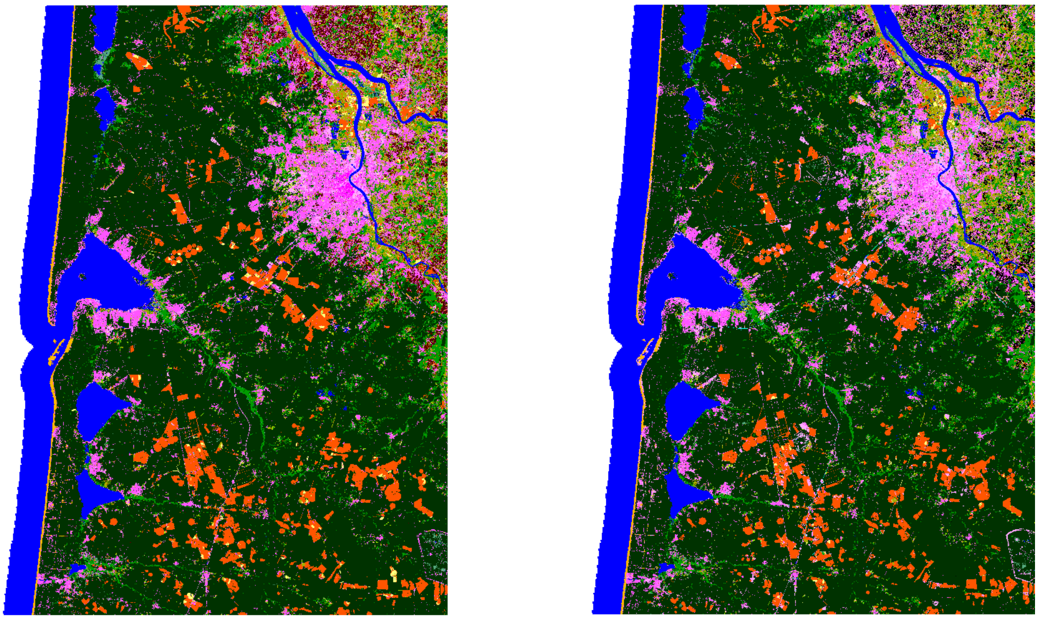


(c) Change detection between the 2nd August 2017 and 22nd August 2019

Figure 26: Change detection maps for the urban sub-use case

To evaluate the quality of the results, the land cover classification maps of 2017 [7] and 2018 [8] produced by the CESBIO were used. These maps were clipped using the extent of the T30TXQ tile (see Figure 27). Please note that these two land cover classification maps contain some different classes. So, we had to pass through a step of class matching between the two classifications. This land cover classification contains multiple classes. In this S2 tile, the classes that can be found are summer crops, winter crops, deciduous forests, coniferous forests, lawns, woody lands, dense urban areas, sparse urban areas, industrial and commercial zones, routes, mineral surfaces, beaches, water, grassland, orchards and vineyards.

To quantitatively compare these results, the classification maps between the two years were used to estimate a binary change detection map, where a change is indicated when a pixel has different classes in the two years. In a second step, thresholds were used to detect changed and unchanged pixels. These thresholds range between [0.01,096] with a step size of 0.05. For each threshold, all pixels that have a change probability value larger than this threshold are considered as changed ones. Additionally, the F1-score was estimated for each threshold value resulting in curves similar to those of precision recall curves. For this analysis, 4 different strategies were considered to estimate the F1-score, namely, the binary, micro, macro and weighted methods. Figure 28 (a) shows the F1-score for all the considered thresholds for the change detection results between the 2nd of August 2017 and 23rd of February 2019, while Figure 28 (b) shows the F1-score for all the considered thresholds for the change detection results between the 2nd of August 2017 and 22nd of August 2019.



(a) Land cover classification 2017 (b) Land cover classification 2018

Figure 27: Land cover classification of the CESBIO of the T30TXQ used for evaluating the change detection algorithm on the urban use case

[7] http://osr-cesbio.ups-tlse.fr/~oso/posts/2018-04-09-carte-s2-2017/
[8] http://osr-cesbio.ups-tlse.fr/~oso/posts/2019-03-25-carte-s2-2018%20/

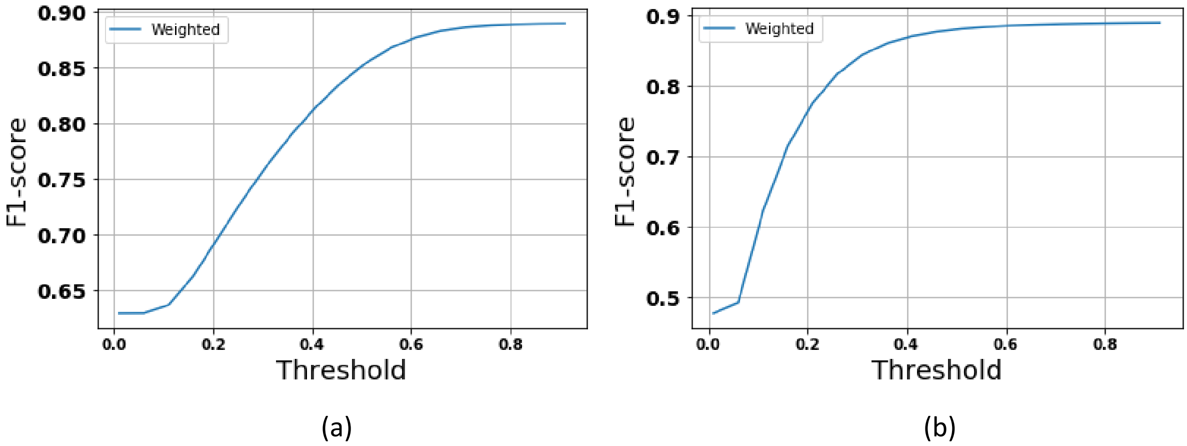


Figure 28: F1 score using 4 different strategies for the considered thresholds between 0.01 and 0.96 for the change detection results between the 2nd of August 2017 and 23rd of February 2019 (a) as well as between the 2nd of August 2017 and 22nd of August 2019 (b)

The previous F1 score results show the importance of choosing the appropriate dates to conduct the change detection analysis. We can notice that the F1-score in Figure 28 (b) stabilizes starting from 0.4 as in practical binary classification problems, the threshold is usually selected between [0.4, 0.5]. Please note that the previous evaluation is based on a land cover classification, which itself contains some errors.

For such a use case, the use of unsupervised change detection tool is not enough to achieve the required results as these tools estimate changes on all land cover classes. However, with the help of the tools, the manual operator work is significantly reduced in order to focus on specific areas of interest. Considering as an example the change detection results between the images acquired on the 2nd August 2017 and 2nd August 2019, we show the evolution of the percentage of changed pixels in the S2 tile by changing the change probability threshold. Figure 29 and Figure 30 show respectively the binary mask and the percentage of changed pixels obtained when changing the threshold.

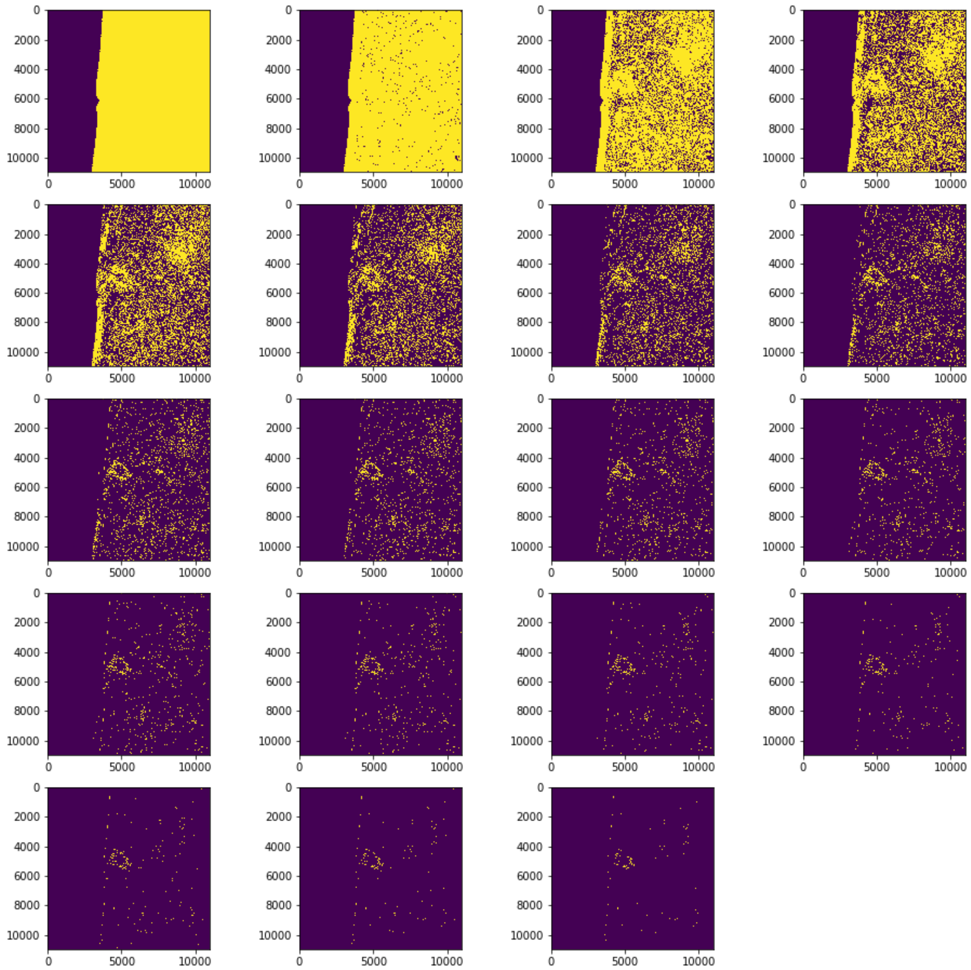


Figure 29: Binary change detection map obtained using the images acquired on the 2nd August 2017 and 2nd August 2019 when changing the threshold of change probability between [0.01,0.96] with a step size of 0.05

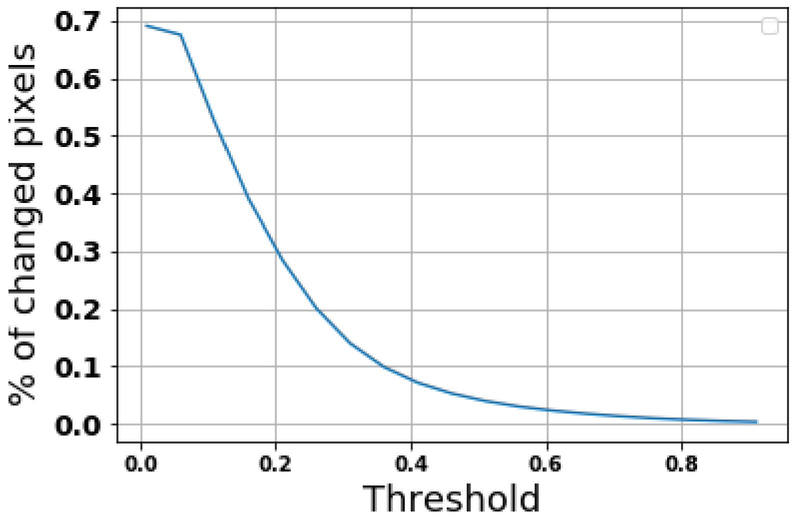


Figure 30: Percentage of changed pixels between the images acquired on the 2nd August 2017 and 2nd August 2019 when changing the threshold of change probability between [0.01,0.96] with a step size of 0.05

Figure 31 shows the changes mostly related to areas of vegetation that were replaced by urban construction. These results were obtained by concentrating on change with high probability (>= 0.75), which shows how the optical change detection tool can be used to reduce the amount of effort required to scan a whole S2 tile.

(a) S2 image acquired on the 2nd August 2017

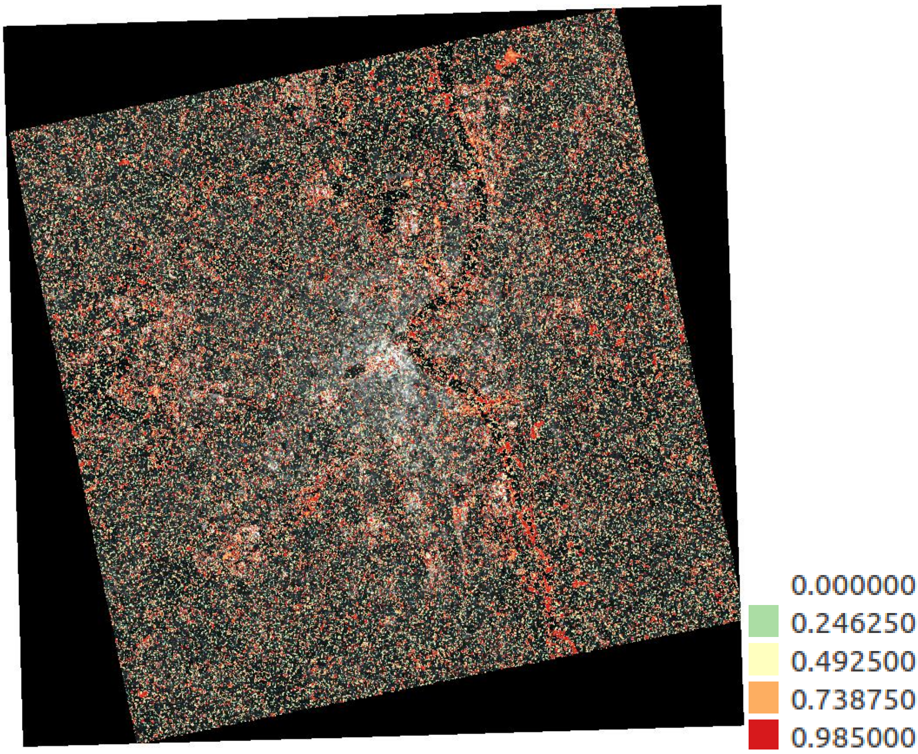
(b) S2 image acquired on the 22nd August 2019

(b) Change between the two images

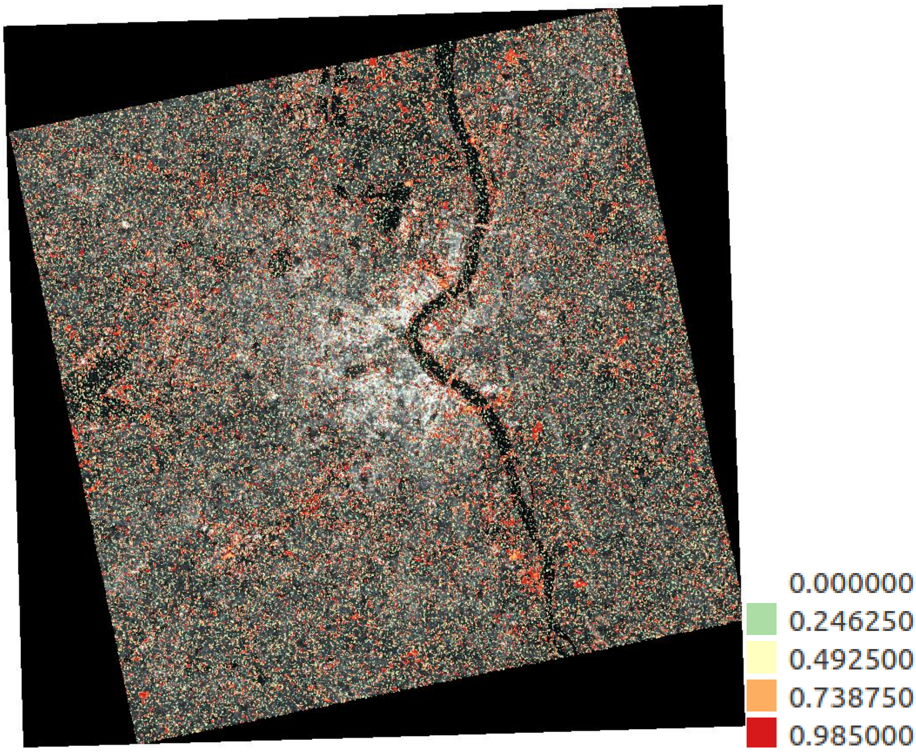
Figure 31: S2 images acquired on the 2nd August 2017 (a) and 22nd August 2019 (b) and their corresponding change detection with a threshold >= 0.75

4.2.3.2 SAR change detection analysis

This analysis was run using the same steps in Section 4.1.2.2. Two S1 images were analyzed acquired on 2017 and 2019, respectively (see Section 4.2.1). The preprocessing of these two images took around 310 seconds while the processing step took around 125 seconds. Figure 32 shows the change detection maps obtained using the VV and VH overlayed over the VV and VH polarizations, respectively.



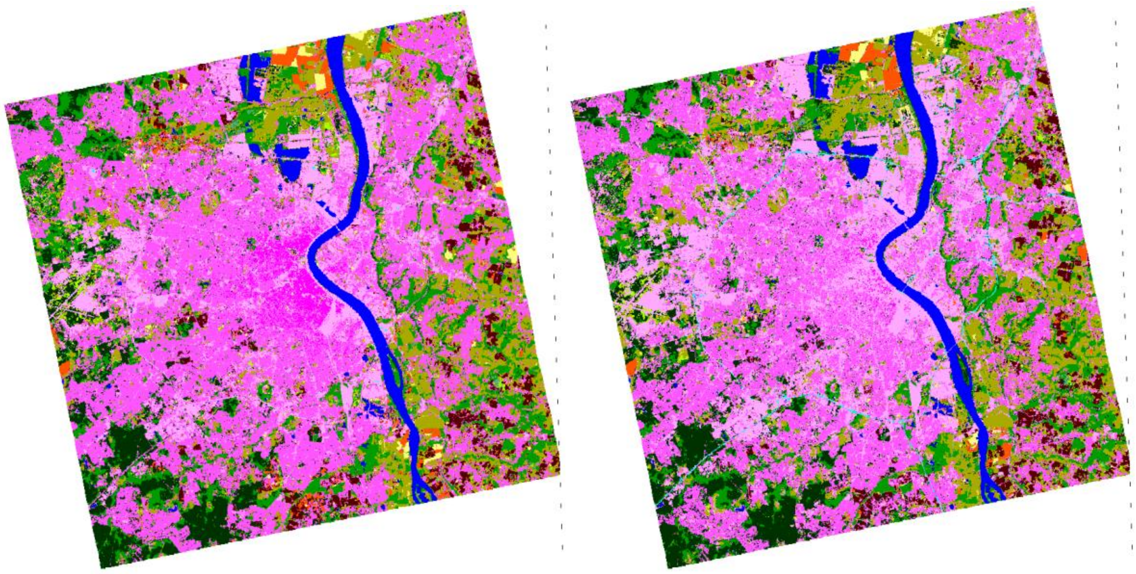
(a) Change detection map using the VV polarization



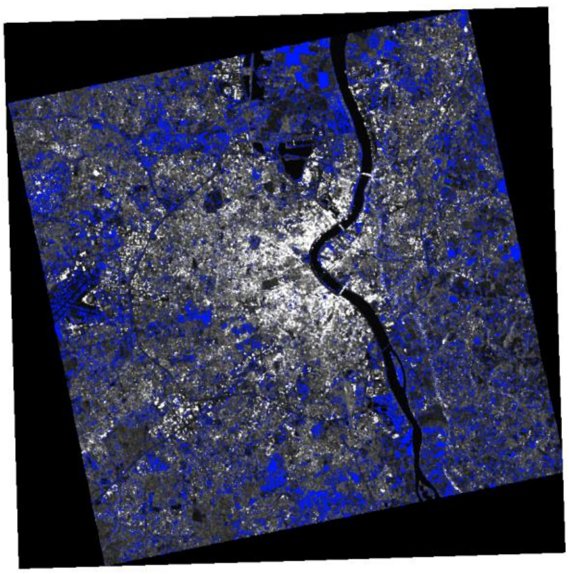
(b) Change detection map using the VH polarization

Figure 32: Change detection using the VV (a) and VH (b) polarizations of two S1 images acquired during 2017 and 2018, respectively

To further investigate these results, the land cover classification of this ROI described in the previous section was used (see Figure 33 (a) and (b)). A binary change detection map (see Figure 33 (c)) was generated using the classification of 2017 and 2018.



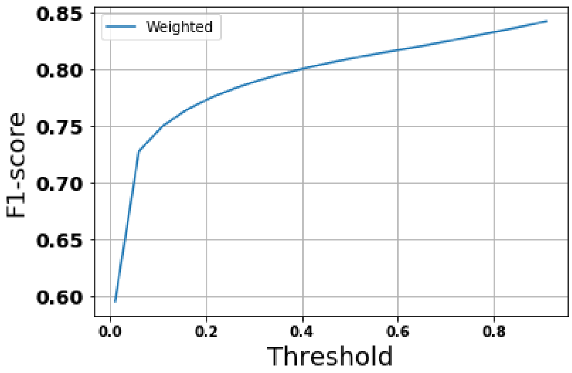
(a) Land cover classification 2017 (b) Land cover classification 2018



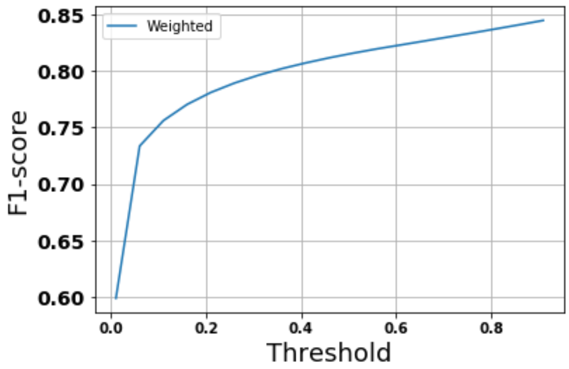
(c) Binary change detection map

Figure 33: Land cover classification of the ROI in 2017 and 2018 as well as the GT binary change detection map generated using the two classifications

Thresholds were used to detect changed and unchanged pixels. These thresholds range between [0.01,096] with a step size of 0.05. For each threshold, all pixels that have a change probability value larger than this threshold are considered as changed ones. Additionally, the F1-score was estimated for each threshold value resulting in curves similar to those of precision recall curves. For this analysis, 4 different strategies were considered to estimate the F1-score, namely, the binary, micro, macro and weighted methods. Figure 34 shows the F1-score for all the considered thresholds for the change detection results using the VV and VH.



(a) VV polarization



(b) VH Polarization

Figure 34: F1 score using 4 different strategies for the considered thresholds between 0.01 and 0.96 for the change detection results using the VV (a) and VH (b) polarizations between two images acquired during 2017 and 2018, respectively

Similar to the analysis conducted for optical change detection, the previous results were obtained by comparing change detection results to a binary mask obtained using land cover classification of the compared two years, which may propagate the classification errors. A further analysis was conducted by varying the change probability and calculating the number of changed pixels. Figure 35, Figure 36 and Figure 37 show the binary change mask and the evolution of changed pixels when increasing the change probability. Please note that the change detection results obtained using the SAR images were challenging to interpret as other factors beside the changes in land cover can affect the acquired signal in SAR images.

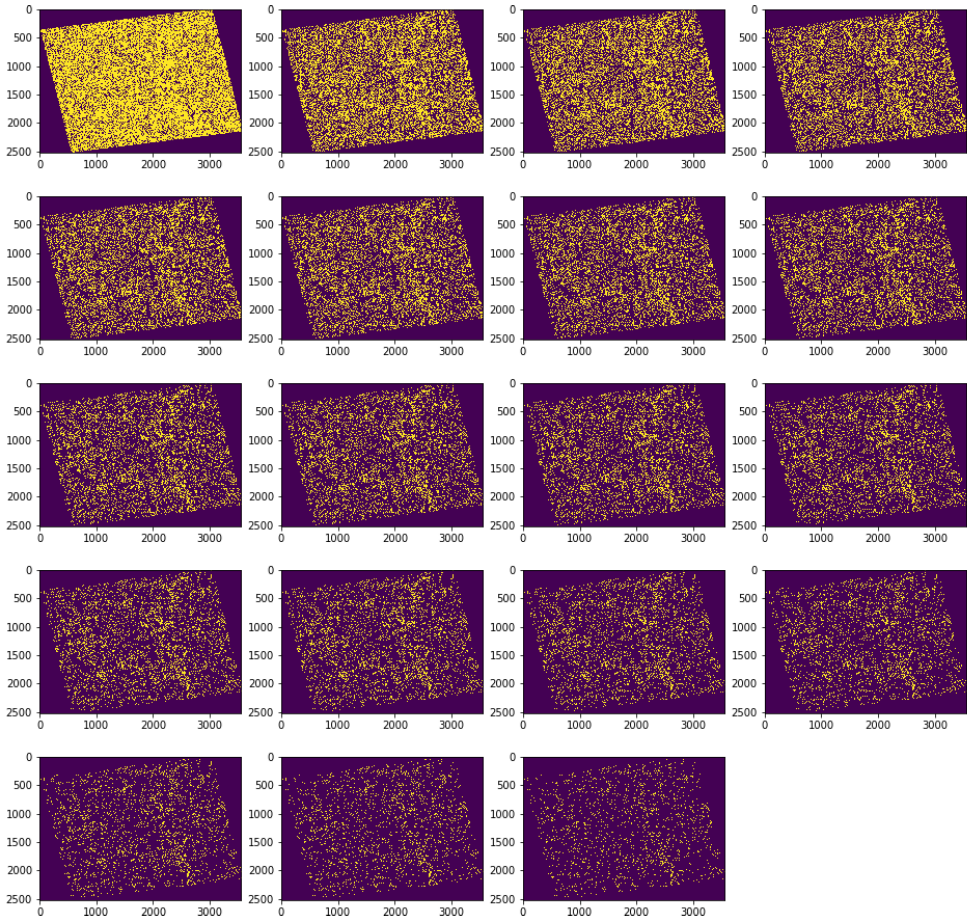


Figure 35: Binary change detection map obtained using the vv polarization of the images acquired during 2017 and 2019 when changing the threshold of change probability between [0.01,0.96] with a step size of 0.05

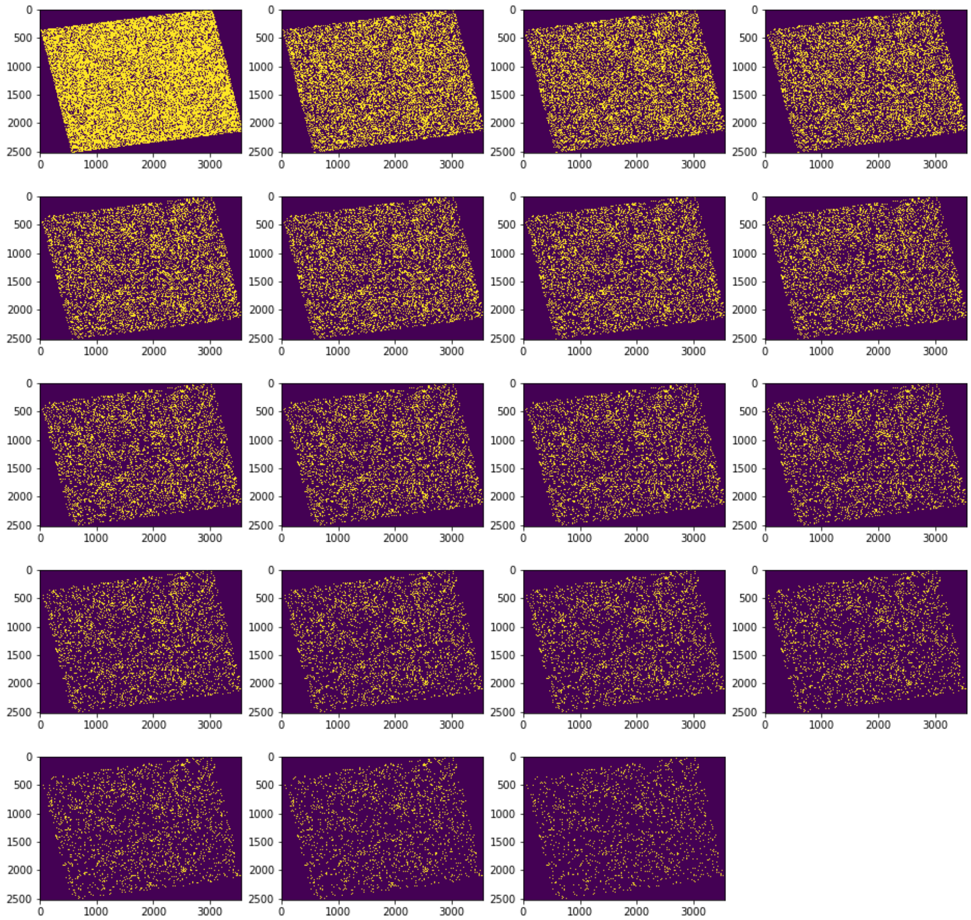
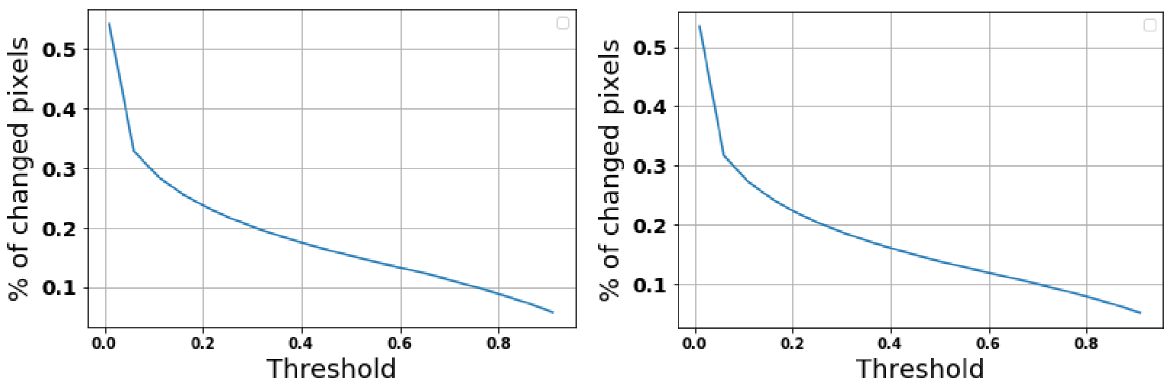


Figure 36: Binary change detection map obtained using the vh polarization of the images acquired during 2017 and 2019 when changing the threshold of change probability between [0.01,0.96] with a step size of 0.05



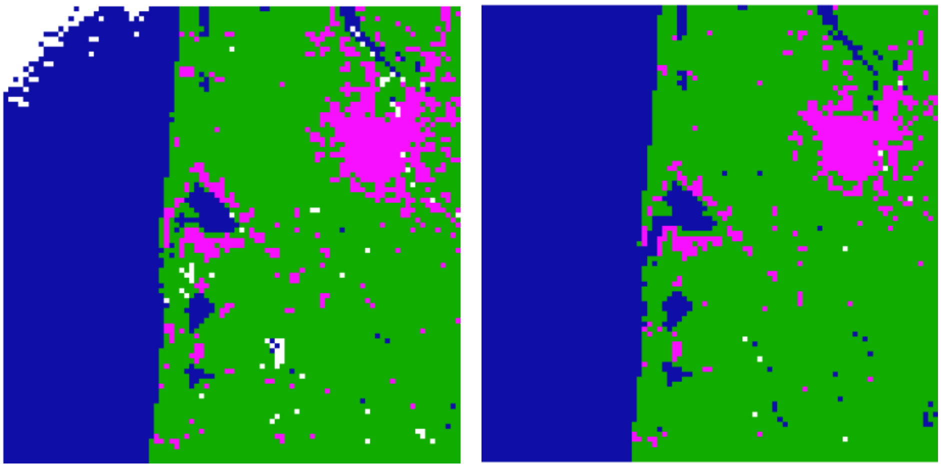
(a) VV (b) VH

Figure 37: Percentage of changed pixels between the images acquired during 2017 and 2019 when changing the threshold of change probability between [0.01,0.96] with a step size of 0.05

4.2.3.3 Data mining analysis

The data mining analysis was conducted using the same steps in Section 4.1.2.3. Two S2 L1C images were used. These images were acquired on the 2nd August 2017 and the 22nd August 2019. The data mining module was run for each image by setting the tile size to 120 and extracting the GLM, WLD, CHIS and GLC features. However, the same observation as before was noticed, the WLD features were the easiest to use when training the active learning algorithm.

Three classes were considered when training the active learning algorithm (mixed urban, mixed vegetation and water). Figure 38 shows the patch classification results on the whole tile for the two images.



(a) 2nd August 2017 (b) 22nd August 2019

Figure 38: : Patch-level classification of the 2nd August 2017 (a) and the 22nd August 2019 obtained using the data mining tool. Pink, green and blue represent the mixed urban, mixed vegetation and water classes, respectively

From this figure, we can notice that there are no significant changes between the two years, which is logical since the time difference between the two images is only two years. Please note that the data mining tool can be used to explore the land cover classes of an image but it can also give an idea of significant changes in terms of change size between two images. However, pixel-level changes are very hard to be detected as the tool works at the patch-level.

To elaborate on the previous visual comparison, the bar plot in Figure 39: Bar plot showing the number of patches for each class obtained using the data mining tool for two images acquired in 2017 and 2019shows the number of parches for each class. This graph concludes that the changes between the two years are very minor, where the differences in the number of patches between the corresponding classes in the two years are negligible.

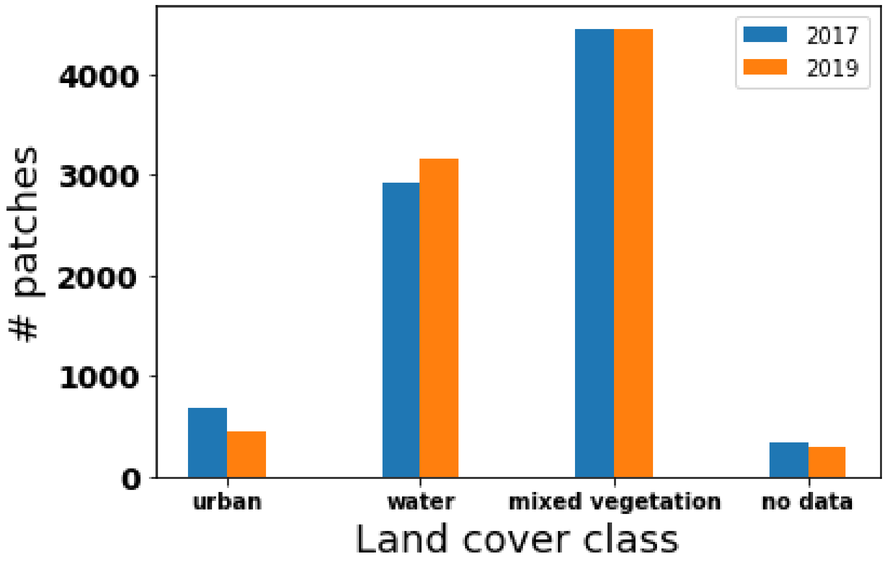


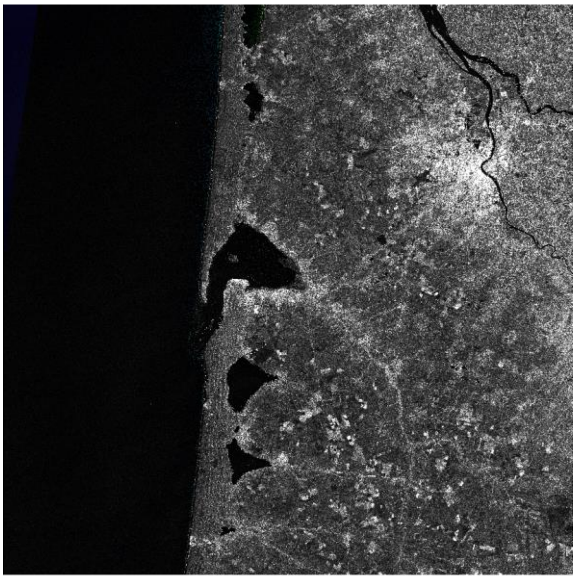
Figure 39: Bar plot showing the number of patches for each class obtained using the data mining tool for two images acquired in 2017 and 2019

4.2.3.4 Data Fusion analysis

The data fusion [9] analysis requires multiple steps to be conducted. These results are explained in what follows.

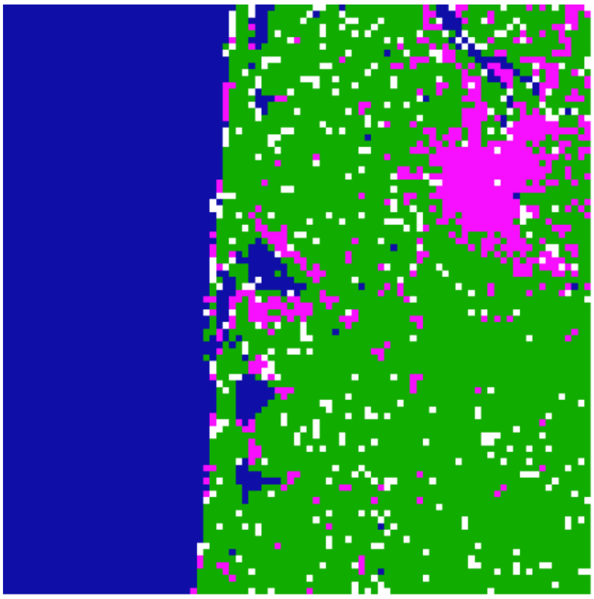
1. The user searches for S1 and S2 images of the ROI then filters these images based on some criteria with a preference to have close dates for both S1 and S2 images. At the end of this step, one S2 image is selected for the tile and multiple S1 images that cover the geographical extent of the ROI.
2. In a second step, the S1 images are tiled to the grid of S2 images in order to co-register both images to the same geographical extent. Please note that this step results in multiple tiled S1 images, where each polarization is separated in a Geotiff file.
3. In this step, the user visualizes all the tiled polarizations of S1 images resulting from the previous step in order to select one of them for the analysis.
4. Once the user selects an image, the identifier of this image is used to create an S1-like image with selected polarization of a specific date all the metadata and auxiliary files.
5. This is the last step on the platform, which consists of jointly analyzing the S1 and S2 images. The user defines multiple parameters such as the tiles size, the bands of interest for S2 analysis and the feature types.
6. The user switches to his local machine in order to perform the active learning using the KDD interface.

The selected S1 (Figure 40 (a)) and S2 (Figure 40 (b)) images were acquired on the 29th August 2019 and 22nd August 2019, respectively. The tile size was set to 120 and the WLD, CHIS features were selected to be extracted from the VH polarization of the S1 image and the blue, green and red bands of the S2 image. Please note that for the data fusion, the feature types are fixed to WLD for S1 images and CHIS for S2 images at the algorithm implementation. Three classes were considered for this analysis, namely, mixed urban, mixed vegetation and water. Figure 40 (c) shows the results obtained when analyzing the S1 and S2 patches jointly. For these classes, the addition of S1 tiles did not change the results drastically and we noticed that the learning process was longer when compared to the data mining using S2 alone.



(a) VH polarization of S1 image

(b) RGB composition of S2 image



(c) Patch-level classification using the data fusion tool

Figure 40: S1 image acquired on the 29th August 2019 (a) S2 images acquired on the 22nd August 2019 (b) and Patch-level classification of the obtained using the data fusion tool (c). Pink, green and blue represent the mixed urban, mixed vegetation, respectively

To further compare the results obtained the data mining and the data fusion tools, the patch classification obtained using the data mining tool on the image acquired during 2019 was compared to the results obtained using the data fusion in Figure 41. This figure shows that both tools lead to similar results for this use case.

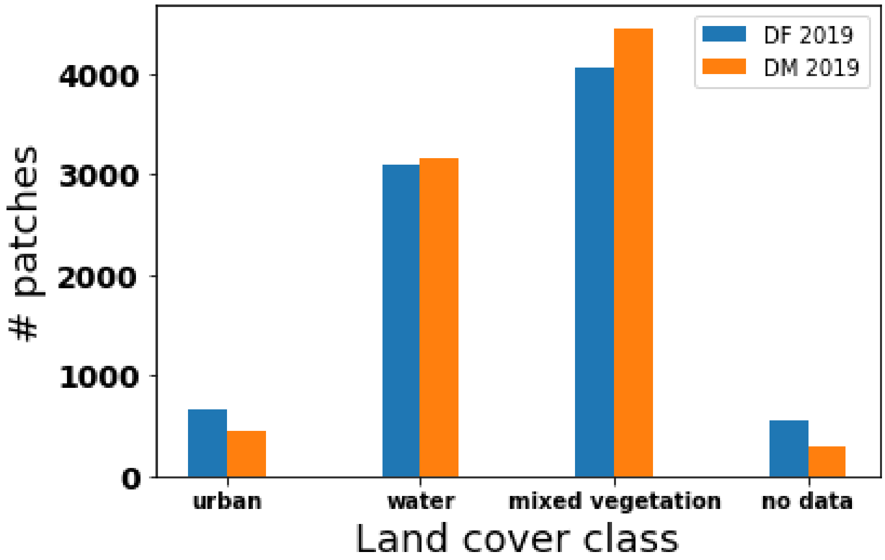
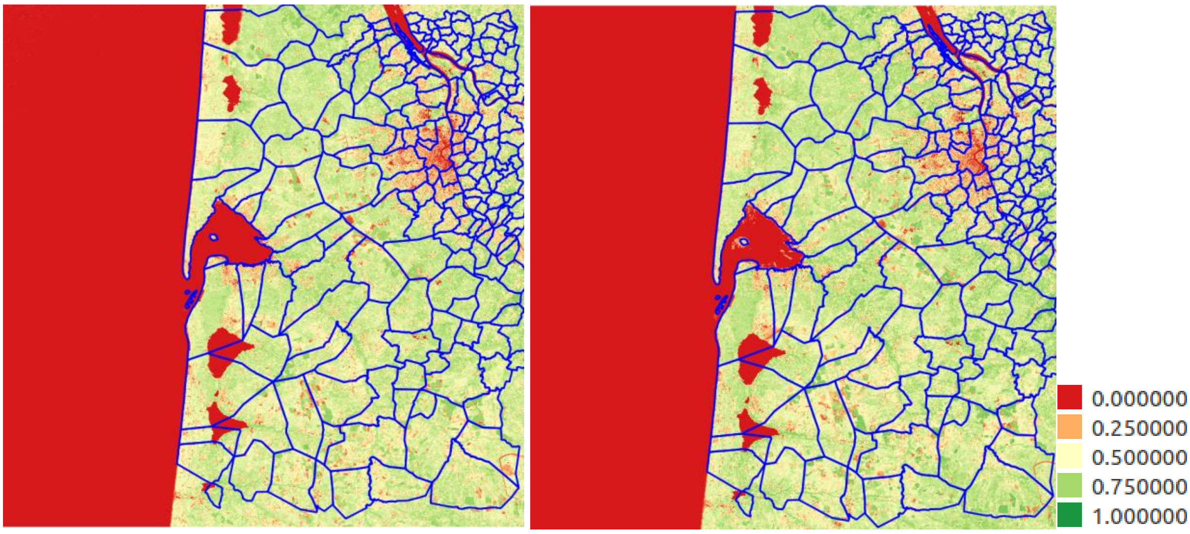


Figure 41: Bar plot showing the number of patches for each class obtained using the data mining and the data fusion tools on the image acquired during 2019

4.2.3.5 Semantic search

The semantic search for this sub-use case uses the same steps in Section 4.1.3.5. The NDVI of two S2 images acquired on the 2nd August 2017 and 22nd August 2019 are used along with the administrative boundaries of French communes that are covered using the T30TXQ S2 tile. The objective of the conducted semantic search is to study the changes in vegetation at the commune-level using the NDVI level as a search criterion. Figure 42 shows the two NDVI rasters estimated using the two S2 images as well as the commune boundaries overlayed on both rasters.

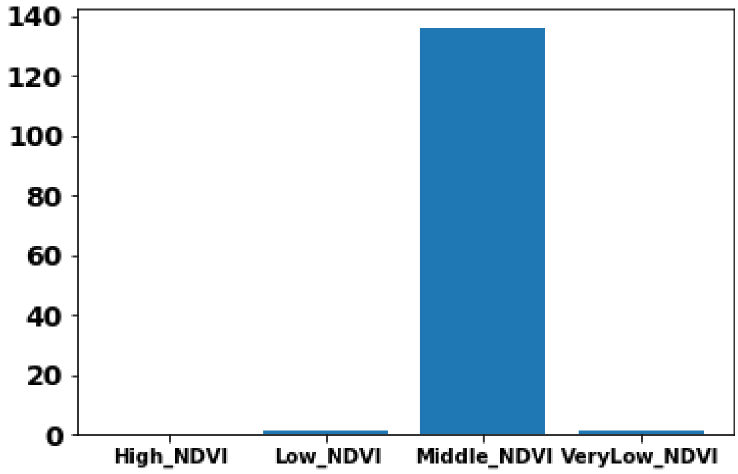


(a) 2nd August 2017 (b) 22nd August 2019

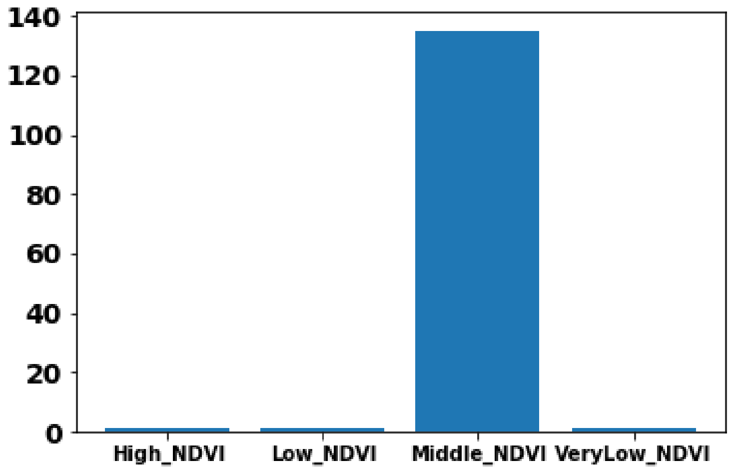
Figure 42: NDVI of two S2 images used for the semantic search analysis as well as the commune boundaries

When using the semantic search, all NDVI levels were considered where 50% of the pixels inside the polygon of a commune hold the same NDVI level in order to extract the dominant NDVI level only. This search was conducted on the two images and only communes that appeared in the two search results were considered.

Figure 43 shows the statistical analysis of 138 communes resulting from the semantic search analysis using the two images. From this figure, it is clear that the area covered by this S2 tile is characterized by a rich vegetation cover, where the majority of communes are characterized by a middle level of NDVI. Additionally, these results showed that the communes of Bordeaux and Carbon-Blanc were always characterized with low vegetation as there is a high concentration of urban structures.



(a) Histogram of the NDVI levels using the semantic search and the S2 image acquired on the 2nd August 2017



Histogram of the NDVI levels using the semantic search and the S2 image acquired on the 22nd August 2019

Figure 43: Histogram analysis for the NDVI levels using the semantic search tool on two S2 images and the boundaries of French communes in the area

5 Conclusion

This report is dedicated to the final validation phase of the CANDELA platform and its analytics tools. The validation process was conducted using two mains steps. The first step consisted of unitary tests where the tools were validated as separated components against the use requirements gathered in the traceability matrix [4], whose objective was to refine the developed tools in the projects for specific use cases. Test cases were created using these user requirements and a statistical analysis on satisfied and unsatisfied requirements was shown, where important points regarding the unsatisfied requirements were discussed. The second validation step tested the proposed use cases in D1.1 [2] and D1.5 [3] using the platform and the used analytics tools, where we showed how these tools can complement each other to provide information of value for the users. The objective of the first sub-use case is to detect changes in vineyards caused by bad weather conditions. The other sub-use case detects changes in urban and agricultural areas. For each sub-use case and each tool, the used datasets and the steps to reproduce the results were illustrated. Additionally, comparison with other data sources was performed.

There are multiple points that can summarize our experience with such a platform. As a user for this platform, one of the main advantages is being connected to a computational and analytical platform where the data sources (Sentinel-1 and Sentinel-2 images) are directly available. To this extent, and during our work on the platform, we found that historical images and especially Sentinel-2 Level-2A images are not available most of the time. Indeed, this is a known challenge for DIAS services and it is worth being taken into consideration when building tools on top of a DIAS. The non-availability of data makes choosing a regular cloud environment a more economical choice, where images can be searched from other sources such as Amazon or Google. Due to the aforementioned challenge of data unavailability, the experiments conducted in this report uses Sentinel-2 L1C images, which affects the performance of the processing algorithms.

Another point that worth stressing out is the audience type of the CANDELA platform. As it has been clear in other reports [1], jupyter lab is the main environment of the platform. This means that there are no graphical interfaces. For a user to interact with the tools, he/she needs to interact with the code. Although the amount of code shown for the user is minimized, having a technical background is mandatory to run the tools and also to conduct some post-processing analysis. Additionally, a remote sensing expertise is needed to handle input/output data (GeoTiffs and shapefiles), their geographical system projection, etc.

Finally, we insist on the importance of technical support for the platform in case it will be commercialized, especially at the first period (early adopters). Some technical issues may occur and there is a need to find fast solutions for the users in order to improve the provided services and adapt to user needs if needed. We also think that making data available for more geographical zones and the tool more generic especially for the semantic search tool is very important for the commercializing step.

6 References

[1] M. Albughdadi (2019), D1.3-4 Use Case Validation Report v1, Deliverable of the CANDELA project

[2] M. Albughdadi (2018), D1.1 Use Case#1 Requirements v1, Deliverable of the CANDELA project

[3] M. Albughdadi (2019), D1.5 Use Case#1 Requirements v2, Deliverable of the CANDELA project

[4] M. Albughdadi, A. Pulak-Siwiec (2020), Traceability matrix v2.8, Internal deliverable of the CANDELA project. Provided as Annex to this Deliverable

[5] A. Pulak-Siwiec (2018), D1.2 Use Case#2 Requirements v1, Deliverable of the CANDELA project

[6] A. Pulak-Siwiec (2019), D1.6 Use Case#2 Requirements v2, Deliverable of the CANDELA project

[7] M. Aubrun (2020), D2.4 Deep learning v2, Deliverable of the CANDELA project

[8] M. Dactu (2020), D2.2 Data mining v2, Deliverable of the CANDELA project

[9] W. Yao (2020), D2.8 Data fusion v2, Deliverable of the CANDELA project

[10] C. Trojhan (2020), D2.6 Semantic search v2, Deliverable of the CANDELA project

Annex I. Traceability Matrix

The Traceability Matrix has been the tool shared among partners (between WP2-3 and WP1) to keep track of the fulfillment of the user requirements. It was introduced in the User Requirements documents [3] [6] and served to collect the possibility of implementing a requirement on the platform and if the requirement is already implemented.

Last version of the document is provided as a common Annex to D1.7 and D1.8, the final Validation Reports.

	ID REQ	FEEDBACK	TNIS	SGIS	V2	SGIS commnent	Status
CD TAS FR	**ANALY-OPTIC-PRE-1**	Remove the limitation of preprocessing only Red, Green, Blue and NIR bands and allow flexible selection of the input bands beside the default ones.	X	X	Done	Done, works well.	
	ANALY-OPTIC-PRE-2	Allowing the integration of bands at 20 m and 60 m which requires and a band resampling module.	X	X	Done	Done, works well.	
	ANALY-OPTIC-PRE-3	Following a naming convention of the outputs identical to the S2 products on CreoDIAS to make retrieving the preprocessed images easier.	X	X	Done	Done, works well.	
	ANALY-OPTIC-PRE-4	Integrating a module that allows using a cloud mask in order to exclude clouds from processing.	X	X	Done	Done, works well.	
	ANALY-OPTIC-CD-1	Adding information in the metadata of the change detection results the S2 images used to produce each change map and the image considered as t0.	X	X	Done	Done, information is included in the file name.	
	ANALY-OPTIC-CD-2	Adding a naming convention to the outputs using names of the input images, which allows running the same module with different input images.	X	X	Done	Done	
	ANALY-OPTIC-CD-3	The ability to run the change detection on multiple S2 tiles to process large geographical zones.	X	X	Done	Done	
	ANALY-OPTIC-CD-4	Normalizing the change probability between 0 and 1 in order to be able to choose a threshold. The maximum of the change probability is currently unknown.	X	X	Done	Done	
	ANALY-OPTIC-CD-5	Allow the user to retrieve the change detection as a full raster (stripe results are only available now that need to be concatenated).	X	X	Done	Done	
	ANALY-OPTIC-CD-6	Possibility of limiting the processing area based on user data, land use layer delivered by DLR tool or CLC		X	If we have time to do it	Not implemented	
	ANALY-OPTIC-CD-7	Removing the stripe effect and/or dark lines.	X	X	Done	Done	

	ID REQ	FEEDBACK	TNIS	SGIS	FEASIBLE	V3	TNIS comment	Status
CD TAS IT	**ANALY-SAR-PRE-1**	Implementing a preprocessing component for SAR images.	X	X		Done	Done, works well.	
	ANALY-SAR-PRE-2	Allow speckle noise filtering for SAR images.	X	X	GRD products are already multi-looked for despeckle. An addional filter is feasible, but will lower spatial resolution even more	Done	Ok	
	ANALY-SAR-PRE-3	Merge the dual polarization of SAR images to be used for further processing.	X	X	Currently not planned. Each polarization is treated as a single product		Not yet, but it is very important. Change detection results significantly differ using each band separately	
	ANALY-SAR-PRE-4	Following the convention of S1 products in naming the preprocessed images.	X	X		Done	Done	
	ANALY-SAR-PRE-5	Adding tools that allow cropping retrieved S1 images to the same geographical extent in order to be processed with the change detection module.	X	X		Done	Done, works well.	
	ANALY-SAR-CD-1	Adapting the SAR change detection module to S1 products available on CreoDIAS (GRD-IW).	X	X		Done	Done	
	ANALY-SAR-CD-2	The ability to use the dual polarization of S1 products in order to exploit all the available information.	X	X	Currently not planned. Each polarization is treated as a single product		Done but separately (see comment on ANALY-SAR-PRE-3)	
	ANALY-SAR-CD-3	The ability to run the change detection on multiple S1 images acquired in different dates (temporal aspect).	X	X		Done	Tested and works well	
	ANALY-SAR-CD-4	The ability to run the change detection on multiple S1 images (geographical aspect).	X	X	Not planned. Time-series acquired over a distinct area require indivudal runs (PRE+CD)		Will not be considered	
	ANALY-SAR-CD-5	Allowing change detection results at the pixel-level (the current version works at the patch-level).	X	X		Done	Done	
	ANALY-SAR-CD-6	Generating only the probability of change in output.	X	X		Done	Change index between 0 and 1	
	ANALY-SAR-CD-7	Making the change detection threshold a parameter to be fixed a posteriori (the results are generated with the prefixed threshold).	X	X				No longer a topic
	ANALY-SAR-CD-8	Highlight only relevant changes	X	X				

	ID REQ	FEEDBACK	TNIS	SGIS	FEASIBLE	V2	TNIS comment	Status
DLR	**ANALY-DM-1**	Make use of the large amount of data available on [http://bigearth.net/]	X		Not planned for the moment		Review comment	No longer a topic
	ANALY-DM-2	The existence of technical support for the installation and user manual that explains how to perform queries, semantic annotations, selection of training patches, etc.	X	X	V2 user guide and videos are on Owncloud Readme files will be updated	Done	Done	
	ANALY-DM-3	Receiving accuracy and precision report that allows them to evaluate the classification model they create using the selected patches and hence enhancing the model if needed (on the KDD side)	X	X		Done	Done	
	ANALY-DM-4	Tool that allows using data fusion results on the CANDELA platform	X	X			Scripts of data fusion will be prepared soon and then tested	
	ANALY-DM-5	Resolve MonetDB bug	X	X	This is not a monetdb bug as it is working locally but the problem was on the platform.		Done	
	ANALY-DM-6	Create the required log and output folders automatically on the platform. This will allow avoiding errors associated with the errornous	X				Done for data fusion and will be added to data mining	
	ANALY-DM-7	Run the data mining module on multiple images	X				Will be done on the DMG	
	ANALY-DM-8	Explain the following issue. On the user side, annotating a given class takes a lot of time up to 10 mins. This causes the GUI to freeze before it starts working again	X				On local docker is much faster. Can be due to connection problems	
	ANALY-DM-9	Defining the way how patches with multiple labels are assigned a specific class.	X				The last annotated label is used to export to tiff. From the database, all the labels are available	
	ANALY-DM-10	Show the unlabled patches in the classification report	X		To be verified once having access to the scripts exporting geotiff		This is done at the export of the geotiff file. Unlabeled patches are assigned a label of -1	
	ANALY-DM-11	Is there away to export the model created on one image to be used on another?	X				due to the variability of the datasets and the scarcity of training data, doing this is not a good idea	
	ANALY-DM-12	Add the scripts that allow exporting a geotiff from the database	X		Also realted to ANALY-DM-10. There was a problem on the platform when trying to duplicate the notebook and the associated python files			

	ID REQ	FEEDBACK	TNIS	SGIS	FEASIBLE	V1	V2	SGIS comment	Status
IRIT	**ANALY-SEM-1**	Allowing users to search images using metadata of the images such as the S2 tile.	X	X	Already done by the platform				No longer a topic
	ANALY-SEM-2	Allowing users to search images using a defined period (start and end dates).	X	X		Done		Done	
	ANALY-SEM-3	Put a jupyter-notebook example of radiometric indice (e.g. NDVI)	X	X			Done	Done	
	ANALY-SEM-4	Allowing users to search images using a criteria (e.g. change, NDVI) level.	X	X		Done		Done	
	ANALY-SEM-5	Allowing users to search images using a specified land cover (LC) coming from the DLR tool or from the CLC.	X	X	Will be developed and implemented in V2 The CLC2018 will be used.			Done	
	ANALY-SEM-6	Allowing users to search images using administrative units.	X	X		Done		Done, it is possibility to search images based on villlage names	
	ANALY-SEM-7	Allowing users to search images using weather data	X		Not planned for the moment			Done	
	ANALY-SEM-8	Put a jupyter-notebook example to check the different sources of information (e.g. CD fromm TAS F and TAS I)		X	Already feasible with the web interface. Will be developed in V2			Done	
	ANALY-SEM-9	Put a jupyter-notebook example to combine the different sources of information (e.g. CD and radiometric indice)		X	Already feasible with the web interface. Will be developed in V2			Done	
	ANALY-SEM-10	Quantify positive and negative change percentage of a criteria according to a use case.		X	Will be developed and implemented in V2			Done	
	ANALY-SEM-11	Ability to preview raster (e.g. S1, S2, CD results) and vector in one window.		X	Already done for S1 & S2 images and vector data. Need some help from the platform to do it for others raster data			Done	
	ANALY-SEM-12	Allowing user to search using some user defined units.		X	Will be developed and implemented in V2			Done	

	ID REQ	FEEDBACK	TNIS	SGIS	FEASIBLE	Version	TNIS commnent	SGIS commnent	Status
PLATFORM	**PLAT-ACC-1**	Assigning a URL to access the platform instead of using its IP address.	X	X		Done		Done, works well.	
	PLAT-ACC-2	Implement a functionality that allows the user to retrieve his/her password if lost.	X	X	Feasible Not planned yet			Not yet	
	PLAT-ACC-3	Implement a functionality that allows the user to sign out from the platform.	X	X		Done		Done, works well.	
	PLAT-JUP-1	Integrating new functionalities to the platform that allows easy deleting, moving and manipulating data stored on the user workspace as the used jupyter environment makes it difficult to achieve that.	X	X	The juptyer lab interface allow to create, delete, rename, cut and paste files and folder in the file browser. The delete non empty folder function has been added to the testing Jupyter lab environment		Done	Copying entire folders and their content from public is not possible and it only works by copying each file separately	
	PLAT-DAM-1	Including new modules that allows easy selection from images retrieved from CreoDIAS without coding.	X	X	Easier functions for searching products have been implemented in the creodias_lib Developement of a GUI is ongoing		Added code to do this task (Done)	Done	
	PLAT-DAM-2	The ability to retrieve cloud masks for S2 images.	X	X	A new function in the creodias_lib that would return the path to the cloud mask for a product would be ok?		Implemented in TAS FR CD	Available in TAS FR scripts	No longer a topic
	PLAT-DAM-3	Add functionalities that allow sharing data between different users.	X	X	Users can share data in the public/SHARED folder	Done	Done, use the public folder	Done via public folder	
	PLAT-DAM-4	Create a data catalog in order to index processed images and search them using the semantic search tool.	X	X	New module will be available soon to convert change detection result to geojson. The geojson will be ingested to the PostGIS database		The IRIT tool allows searching processed images	IRIT tool	No longer a topic
	PLAT-DAM-5	Allow users to search the catalog with different criteria such as start and end dates, weather, changes, NDVI and administrative units.	X	X	I think this is what semsearch does		IRIT tool	Developed in IRIT tools	No longer a topic
	PLAT-1	The jupyter lab environment is complicated for non-technical users and the need to explore other alternatives that are more user-friendly.	X	X	We propose to develop GUI inside jupyter lab with ipywidget	X	As agreed, the platform is destined to users with minimum technical knowledge	Not yet	
	PLAT-2	Build the different components of the analytic tools are independent ones and allow the user to build an analysis pipeline by defining a workflow through an XML file for example without the need to interact with the code directly. If the latter is complicated, it might be simpler for the users to call the python module and passing inputs as arguments.	X	X	Can you provide an example of an XML file describing such pipeline? And how the XML should be executed? From a Jupyter notebook?		As agreed, the platform is destined to users with minimum technical knowledge	Not yet	No longer a topic
	PLAT-3	Permanent installation of python packages that allow manipulating auxiliary datasets such as vector data, weather data, etc.	X	X	Requested python packages have been installed	Done	Done, packages added such as geopandas, tiffffile, fiona, shapely	Done, works well.	
	PLAT-4	Adding visualization functionalities to the platform that allow visualizing Copernicus products (at least the outlook) and the obtained outputs.	X	X	In developement with ipywigets and ipyleaflet. To be tested on images coming from CreoDIAS (does it work to visualize Sentinel products before preprocessing or we need to preprocess the products first (JP2toGeoTiff))	X	Implemented using ipyleaflet for the results	It is possible to visualize results	
	PLAT-5	Creating an example notebook that shows how the proposed analytic tools are complementary to each other. For the moment, each analytic tool is a separate component. This can add more value to the proposed analytic tools.	X	X	I can create a notebook that shows how to run the different algorithms on the same areas at the same dates. Is it what you expect?		Same comment as SGIS. This is the task of tools developers (WP2)	A script combining TAS FR and IRIT tools was created during the heckathon.It would be helpful if there were scrips and connections in TAS IT and DLR tools with IRIT semantic search.	No longer a topic

	PLAT-6	Adding a script to search for images before and after a given date. This given date is associated with a specific event such as a date of a natural disaster in abrupt natural disasters or change detection in vineyards.		X	The creodias_lib allow to search for S1 and S2 products between two dates with any search criteria available on creodias (cloudcover, type of acquisition...). The result of the search is a llist of product that satisfy all the constraints. If the date of the event is between those two dates you can easily spot the two closest images before and after the event. If needed I can also code a function takes in parameter only one date and return only the two closest products before and after the date.		Done	Tested. The around date tool allows searching for images before and after a given date (around date)	