SARSAR

CONTRACT NR SR/00/372

WP3:

Validation and platform integration

D3.1:

Performance assessment report

Mattia Stasolla (RMA)
Sophie Petit (ISSeP-CTG)

Table of contents

List of figures

List of tables

1. Introduction

The goal of this document is to provide a performance assessment for the change detection tool that has been developed in WP2 for the automatic monitoring of the RDSs.

The documents is organized as follows:

- In Section 2, a description of the processing blocks that compose the tool is provided;
- In Section 3, the ground truth used for the validation is described;
- In Section 4, the results are reported and discussed;
- In Section 5, the conclusions are drawn.

2. Tool description

The processing blocks of the tool, whose main goal it to automatically analyze the redevelopment sites (RDSs) and estimate their probability of change, are shown in Figure 1. First, the extraction of the features based on the pre-processed images available in Terrascope is needed to create the necessary temporal profiles. Next, the methodology for the change detection is applied. This part provides information on whether or not a change is present and an estimate of the change date(s). Then, a rule-based classification is performed in order to provide additional information on the presence of a change and its type: vegetation, building or soil. Finally, a csv file is delivered to the operator. For each RDS, this report includes the information that a change has occurred (or it remained unchanged), the type of change, the estimated date of the change (if available), and the amplitude and probability of change.

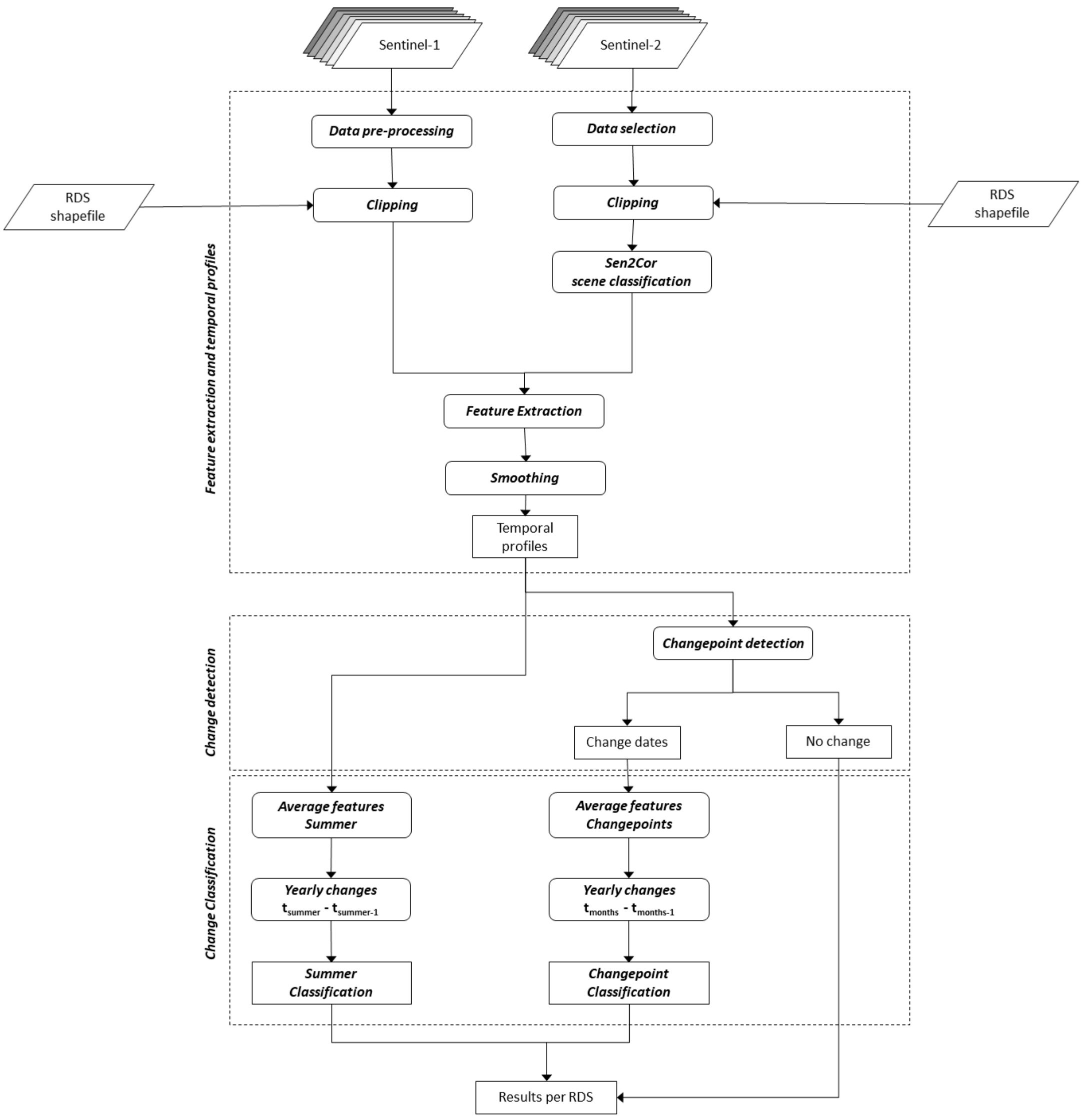

Figure 1: Workflow of the automatic change detection tool for the automatic monitoring of the RDSs.

2.1. Feature extraction and temporal profiles

For each Sentinel-1 acquisition (more specifically, the VH band) that contains the site of interest, the average sigma0 for that site is computed and used to populate the corresponding temporal profile. Since a site can be typically seen from 3 to 4 different viewing angles (considering both ascending and descending orbits), separate profiles are created for each satellite pass and then averaged to obtain a unique 'VH' feature.

Regarding Sentinel-2 data, all the L2A tiles over the area are analyzed. Only the one presenting less than 25% of cloud cover are selected to avoid as much as possible cloud pixels. Then, each image is clipped based on the RDSs vector polygons. Image co-registration is ensured during this process. Following, the pixel classified as no data, shadows, clouds and snow are removed.

For each site, the following six spectral indexes are extracted:

$$\text{Built} - \text{Up Areas Index (BAI)} = \frac{(\text{B02}-\text{B08})}{(\text{B02}+\text{B08})} \tag{1}$$

$$\text{Brightness Index (BI)} = \sqrt{\frac{(\text{B04}*\text{B04})+(\text{B03}*\text{B03})}{2}} \tag{2}$$

$$\text{Brightness Index 2 (BI2)} = \sqrt{\frac{(\text{B04}*\text{B04})+(\text{B03}*\text{B03})+(\text{B08}*\text{B08})}{3}} \tag{3}$$

$$\text{Normalized Vegetation Index (NDVI)} = \frac{(B08-B04)}{(\text{B08}+\text{B04})} \tag{4}$$

$$\text{Normalized Difference Water Index (NDWI)} = \frac{(B03-B08)}{(\text{B03}+\text{B08})} \tag{5}$$

$$\text{Soil Brightness Index (SBI)} = \sqrt{(B04 * B04) + (B08 * B08)} \tag{6}$$

where B0n corresponds to the n-th Sentinel-2 band used for the calculation.

Both Sentinel-1 and Sentinel-2 features are finally linearly interpolated to fill in the gaps in the data and smoothed using a Gaussian kernel with a standard deviation of 61.

2.2. Change detection

The second processing block is the change detection, where the features extracted from the Sentinel images are jointly analysed using the Pruned Exact Linear Time (PELT), a well-known changepoint detection method that provides an exact segmentation of the time series with a linear time complexity.

Given a time series $s = (s_1,\ldots,s_k)$, the number n and time position $t_{1:n}= (t_1,\ldots,t_n)$ of the changepoints are obtained by solving the following penalized minimization problem:

$$Q_n(s_{1:k}, p) = \min_{n,t_{1:n}} \left\{ \sum_{i=1}^{n+1} [C(s_{(t_{i-1}+1):t_i})] + p \right\}$$

where C is the segment-specific cost function

$$C(s_{a:b}) = \sum_{i=a+1}^{b} \left\| s_i - \bar{s}_{a:b} \right\|_2^2$$

and $p=log(k)$ a penalty term to control overfitting.

When one or multiple changepoints are detected, these become the input of the next block, the change classification. When no changepoints are detected, this information is sent to the final block and merged into the final report to the users.

2.3. Change Classification

For each site, two separate processes are implemented that allow us, on the one hand, to provide information on the type of change for progressive changes and, on the other hand, to classify the changes associated to the detected changepoints.

The first one is solely based on Sentinel-2 data. It focuses on the summer months (t_{summer}), from May to August. This process, hereafter referred to as "summer classification", enables the detection of changes that occur gradually over one year period.

The second process, the "changepoint classification", is based on both Sentinel-1 and Sentinel-2 features, and it is performed when one or multiple change dates are available from the previous block. It takes into account the average of the data available after the change date (t_{months}), namely 2 months for Sentinel-2 data and 1 month for Sentinel-1 data).

For both processes, the difference between the average values at the time of change (t_{summer} or t_{months}) of the selected features and the average values that same features had the year before ($t_{summer-1}$ and $t_{months-1}$) is calculated and compared to specific thresholds that were found to be suitable for discriminating between changes of different land cover types (vegetation, building or soil). Additionally, the NDVI and VH features are then used to determine also the direction (increase or decrease) of the change

The rules used for the change classification are listed in Table 1.

Change classification	t_{summer} - $t_{summer-1}$	t_{months} – $t_{months-1}$
Vegetation increase	NDVI > 0.1	NDVI > 0.1
Vegetation decrease	NDVI <- 0.1	NDVI < -0.1
Building change	BI ≠ 150 *or* BI2 ≠ 150 *or* SBI ≠ 250	-
Building increase	-	VH > 0.135
Building decrease	-	VH < -0.135
Soil change	BAI ≠ 0.05	BAI ≠ 0.05

Table 1 – List of the rules used for the classification of the types of changes.

3. Ground truth

For this performance assessment, two ground truths have been created by visual analysis.

The first ground truth is based on the orthophotos (25 cm resolution) taken in summer 2016 and 2018, and focuses on the RDSs for which there are changes that can be observed from Sentinel data. This dataset was developed to account for major changes for which we do not have information about the exact date of change.

The second ground truth is based on Pleiades images (50 cm resolution) acquired monthly between January 2019 and December 2020 on two specific areas with a high concentration of RDSs. This provides complementary information compared to the orthophotos ground truth. In fact, while the orthophotos ground truth focuses on RDSs with significant changes, this dataset was created to take into account in a more balanced way the different types of change. Moreover, thanks to the more frequent coverage guaranteed by the Pleiades images it was also possible to report with a certain accuracy the change dates.

In total, 141 and 161 sites are respectively present in each ground truth. For each of the 302 RDSs, changes have been recorded for vegetation, buildings and soil (see Figure 2). Overall, 152 of the sites present at least one change and 150 no changes at all. The breakdown of the changes into the three possible types is available in Table 2 - Ground truth: number of changes and breakdown into change types.Table 2.

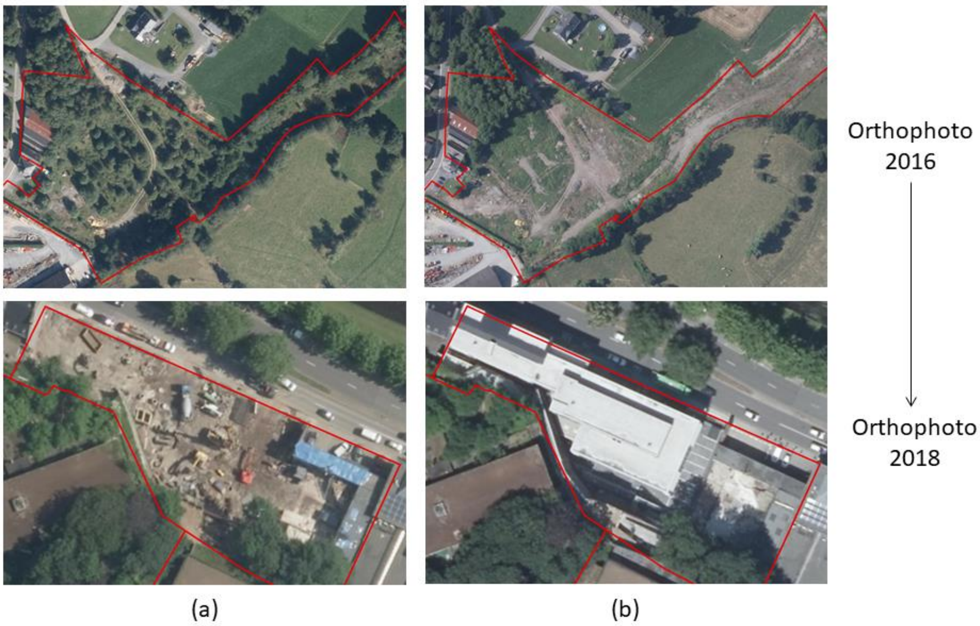

Figure 2 - Close-ups of two RDSs showing (a) vegetation decrease and soil change ("Scierie Renard"); (b) building increase and soil change ("Immobilière Bouchoms").

Ground truth	Building	Vegetation	Soil	Total changes	Total RDSs
Orthophotos	60	97	125	**282**	**141**
Pleiades	8	13	15	**36**	**161**
Total	***68***	***110***	***140***	***318***	***302***

Table 2 - Ground truth: number of changes and breakdown into change types.

4. Results

4.1. Change detection

The performance is assessed in terms of true positive rate (TPR) and false positive rate (FPR). The overall problem can be in fact seen as a binary classification where either a 'change' (1) or a 'no change' (0) has to be detected. In order to compare the results with the ground truth, the latter has been coded so that any change in any of the 3 classes (building, vegetation, soil) is assigned the value 1; in the case of no change for all the 3 classes, the ground truth is given the value 0. A confusion matrix is then generated so that the number of true positives (TPs), true negatives (TNs), false positives (FPs) and false negatives (FNs) could be used to compute the TPR and FPR. To provide a unique measure that takes into account both detection and miss rates, the F_1-score has also been calculated. For the sake of completeness, the overall accuracy (OA) is also reported.

It is worth mentioning that, due to the specific way in which the ground truth is constructed, in order to generate the confusion matrix we have made the arbitrary assumption that only one change per site has occurred in the considered period of time. This is a simplification that helped us to compare in a more straightforward way the results, but might not fully reflect the real situation, especially for the sites belonging to the Orthophotos ground truth, as for a certain number of them it is more likely that multiple changes have occurred at different times.

The change detection has been performed using the VH and NDWI features, which amongst the other features ultimately provided the highest accuracy. The use of both Sentinel-1 and Sentinel-2 data, which provide complementary information (VH mostly about buildings and NDWI mostly about vegetation and soil), allows a more effective identification and classification of changes. The results for the entire dataset are shown in the first row of Table 3. The number of sites for which we had an estimated change is 108, 91 of which correctly classified. Among the unchanged sites, we missed 46 of them, resulting in an OA of 79%. In terms of correct and miss detection rates, we therefore obtained a TPR of 66% and an FPR of 10%, with an F_1-score of 0.74.

In order to better understand the results of the following block, the change classification, it is helpful to separate the Pleiades detections from the full dataset. The results are provided in the second row of Table 3. For this dataset, the number of sites that have been flagged as changed is 26, with 9 FPs, whereas the correct detections of the unchanged sites are 125. As a result, the TPR and FPR decrease to 55% and 7%, respectively, and consequently the F_1-score drops to 0.59. The OA, instead, increases to 87%, mainly due to the fact that the dataset is rather unbalanced.

	TP	FP	FN	TN	TPR	FPR	F_1-score	OA
Full dataset	91	17	46	148	66%	10%	0.74	79%
Pleiades	15	9	12	125	55%	7%	0.59	87%

Table 3 - Changepoint analysis: confusion matrix and performance metrics.

4.2. Change classification

Summer classification

The “summer classification” takes into account, for each of the 302 sites, the summer comparison between 2016 and 2018 for the orthophotos dataset and between 2019 and 2020 for the Pleiades dataset. The performance has been assessed combining the two datasets and computing the TPR, FPR and the F_1-score for each class, along with the overall accuracy (see Table 4).

The overall performance of the yearly classification based on summer values is satisfactory. The best results are obtained for the ‘vegetation’ class, for which the OA is 90% and the TPR and FPR are 87% and 9%, respectively. The resulting F_1-score is 0.80. The performance for the ‘building’ and ‘soil’ classes are slightly lower, with an OA of 76% and 79%, respectively, yet still good, with an F_1-score that is above 0.7.

Type of change	TP	FP	FN	TN	TPR	FPR	F_1-score	OA
Vegetation	59	21	9	213	87%	9%	0.80	90%
Building	87	49	23	143	79%	26%	0.71	76%
Soil	103	26	37	136	74%	16%	0.77	79%

Table 4 - "Summer classification" (full dataset): confusion matrix and performance metrics.

To look deeper into the ‘vegetation’ class, Table 5 also shows the results disaggregated by ‘increase’, ‘decrease’ and ‘no change’ types, with the corresponding overall accuracy and omission/commission errors. As can be seen, for both the increase and decrease of vegetation, around 1 in 4 detections is a false alarm, whereas the percentage of missed changes are 20% and 12%, respectively. It is worth noting that there is no confusion between the two classes, as all the errors fall in the ‘no change’ class. For this class, instead, the commission and omission errors are much lower, namely 4% and 9%.

	Vegetation				
	Increase	**Decrease**	**no change**	**Total**	**Commission errors**
Increase	8	0	3	11	27%
Decrease	0	51	18	69	26%
no change	2	7	213	222	4%
Total	10	58	234	302	
Omission Errors	20%	12%	9%		**OA = 90%**

Table 5 - “Summer classification” (full dataset): detailed confusion matrix for the ‘Vegetation’ class.

Changepoint classification

The “changepoint classification” takes into consideration only the RDSs for which at least one changepoint date has been estimated within the change detection process. As multiple changes can occur in the same site during the considered time period, the yearly comparison has to be done for each estimated change date. This was only possible using the Pleiades dataset, as only for this ground truth the exact change dates are available. The results are provided in Table 6.

Even though some changes happened during winter months, the results for the vegetation changes remain good, with an OA for of 85% and an F_1-score of 0.75. With respect to the full dataset, the main difference here is in the TPR, which is lower by 20 percentage points (67%). As regards the 'building' class, we have the opposite trend, with both OA and F_1-score higher than those obtained for the full dataset. Although the TPR is slightly lower, the significant drop off in the FPR brings an improvement in the performance. Finally, for the 'soil' class, all the metrics show a drop in the performance, especially as far as the FPR is concerned.

Type of change	TP	FP	FN	TN	TPR	FPR	F_1-score	OA
Vegetation	6	1	3	16	67%	6%	0.75	85%
Building	7	1	3	15	70%	6%	0.78	85%
Soil	11	4	4	7	73%	36%	0.73	69%

Table 6 - "Changepoint classification" (Pleiades dataset): confusion matrix and performance metrics.

To complete the analysis, the detailed confusion matrices for the classes 'vegetation' and 'building' are provided in Table 7 and Table 8. Once again, the results are disaggregated by 'increase', 'decrease' and 'no change' types.

For the 'vegetation' class no increase was reported within any site of the ground truth, therefore no metric was calculated. Instead, out of 9 'decrease' changes, 6 were correctly identified, resulting in a commission error of 14% and an omission error of 33%. If we look at the 'no change' class, we have a similar false alarm rate, but a much lower miss rate.

			Vegetation		
	Increase	**Decrease**	**no change**	**Total**	**Commission errors**
Increase	0	0	0	0	-
Decrease	0	6	1	7	14%
no change	0	3	16	19	16%
Total	0	9	17	26	
Omission Errors	-	33%	6%		**OA = 85%**

Table 7 - "Changepoint classification" (Pleiades dataset): detailed confusion matrix for the 'Vegetation' class.

For the 'building' class, half of the 'increase' changes in buildings were missed (50% omission error). However, all the changes that were flagged as an increase were correct (0% commission error). More accurate was instead the classification of a decrease, with only 1 false alarm and 1 missed detection. Finally, the 'no change' classification was the one providing the best performance, with a commission error of 17% and an omission error of 6%.

			Building		
	Increase	**Decrease**	**no change**	**Total**	**Commission errors**
Increase	2	0	0	2	0%
Decrease	0	5	1	6	17%
no change	2	1	15	18	17%
Total	4	6	16	26	
Omission Errors	50%	17%	6%		**OA = 85%**

Table 8 - "Changepoint classification" (Pleiades dataset): detailed confusion matrix for the 'Building' class.

5. Conclusions

Four main conclusions can be drawn:

- The proposed method has provided satisfactory results for the change detection and the change classification for both ground truth datasets. As far as the change detection is concerned, thanks to the complementary information provided by the VH and NDWI features (the former mainly for buildings, the latter mainly for vegetation/soil), we were able to achieve an overall accuracy for the full dataset of 79%. As far as the change classification is concerned, the OA ranges from 79% to 90%, depending on the type of change that is considered (vegetation, building, soil). The very high OA and F_1-score obtained for the vegetation "summer classification" illustrate the robustness of the selection of the NDVI as a vegetation indicator, especially in summer conditions. Regarding the classification of buildings, the results revealed the suitability of combining the BI, BI2 and SBI indices. Finally, the BAI has proven to be useful for the detection of soil changes.
  However, there are some caveats. First, the results here shown only focus on those sites that are bigger than 500 m², as the spatial resolution of the Sentinel data poses some problems with smaller sites. Second, when comparing the "summer classification" and the "changepoint classification" results it should be taken into account that the size of the two datasets is very different (302 vs 26), and this has an impact on the results both in terms of representativeness and numerical accuracy.

- The "summer classification" is better suited for the detection of gradual changes. Figure 3 illustrates an ongoing vegetation growth leading to a soil decrease. This was not captured by the changepoint detection method but was classified as vegetation increase and soil change thanks to the summer 2016-2018 comparison. The "summer classification" also provides better vegetation classification when change dates occur during winter, as seasonality strongly impacts the performance.

Figure 3 - Close-ups of an RDS showing a gradual vegetation increase ("Ets Biernaux").

- The use of the RDSs vector polygons to group image pixels in the change analysis constitutes a limitation. In fact, it averages the information over the whole sites. This might lead to the non-detection and/or non-classification of either small changes or bigger changes occurring on large sites. Nevertheless, reducing the polygon size by applying segmentation based either on a fixed grid or on external sources can lead to several other problems. First, some polygons might be smaller than 500 m², which is the minimum size per polygon used to cope with the spatial resolution of the Sentinel data. Second, even though external sources could provide prior information on the type of change, they need to be up-to-date in order to avoid the propagation of errors.
- The use of Sentinel data has also its limitations. First, as already mentioned, the spatial resolution reduces the number of RDSs that can be analyzed. In total 83% of the RDSs are larger than 500 m². Second, even though most of the sites are former large industrial areas, changes may occur only in minor parts of the site, as illustrated in Figure 4.

  However, Sentinel images offer major advantages compared to orthophotos, which are taken once a year, or Pleiades images, which can be obtained on demand. In fact, not only can they guarantee a much higher temporal coverage (especially if we consider the Sentinel-1 all-weather capabilities), but also they are completely free, which means that the operational costs of the tool are significantly reduced.

Orthophoto 2019 Orthophoto 2020

Figure 4 - Close-ups of an RDS showing a building increase that is too small to be detected with Sentinel data ("S.A.N.I. Carrelages").