

Change Detection Using Sentinel-1 and Sentinel-2

Comparison of Image Processing Methods

Thibault Poux
Feddy Immoula
Simon Khan
M2 AI — Univeristé Paris-Saclay

November 2025

Abstract

In this project, we study change detection from satellite imagery by comparing the performance of two complementary sensors: Sentinel-2 (optical) and Sentinel-1 (SAR radar). Our research question is: how well can classical image processing methods detect anthropogenic changes (buildings, infrastructures, land development) from Sentinel-1 and Sentinel-2 data, and which of the two sensors is better suited for this task? Using a bi-temporal dataset with pixel-wise change / no-change annotations, we first implement non-AI methods: pixel-wise differences with Otsu thresholding (on RGB, individual bands and SAR), the urban index NDBI, multispectral Change Vector Analysis (CVA), and bi-temporal PCA (extraction of change components). On the SAR side, we exploit the VV and VH channels through weighted differences and a two-channel CVA. We also implemented an AI methods, based on Segment Any Change.

Performance is evaluated at the pixel level using the F1-score, Intersection over Union (IoU) and Cohen's kappa, in order to account for the strong class imbalance. Optical methods clearly outperform SAR-based methods: multi-channel RGB differencing reaches a mean F1-score of 0.26 ($\text{IoU} \approx 0.17$, $\kappa \approx 0.24$), while the best SAR method (VV/VH fusion) peaks at a mean F1-score of about 0.09 ($\text{IoU} \approx 0.05$). Some scenes, however, exhibit significantly better results ($F1 > 0.6$ for certain optical configurations), highlighting a strong dependence on the scene context. We conclude that, in this setting and with classical methods, Sentinel-2 is much better suited than Sentinel-1 alone for detecting anthropogenic changes, and we discuss as future work the potential of more advanced SAR processing and AI-based multi-sensor fusion.

Contents

1 Data description	4
1.1 Dataset	4
1.2 Motivation for choosing this dataset	4
1.3 If we had a 1M€ project	6
2 Description of the chosen methods	7
2.1 Non-AI algorithms	7
2.1.1 Sentinel-1 (SAR) based methods	7
2.1.2 Sentinel-2 (optical) based methods	9
2.1.3 Multiband and advanced optical method: bi-temporal PCA	11
2.1.4 Generic difference function (multi-sensor baseline)	11
2.2 AI algorithms	12
2.2.1 Constraints	12
2.2.2 Alternative aproches	12
2.2.3 Segment Anything Model	12
2.2.4 Segment Any Change	13
3 Global analysis of the results	13
3.1 Analysis of the results of non-AI methods	13
3.1.1 Summary table of results	14
3.1.2 Analysis of Sentinel-2 (optical) methods	14
3.1.3 Analysis of Sentinel-1 (SAR) methods	18
3.1.4 Comparison Sentinel-1 vs Sentinel-2	20
4 Conclusion	21

1 Data description

1.1 Dataset

In this project, we use the *multimodal change detection* dataset introduced by Ebel et al. [1] in “*Fusing Multi-Modal Data for Supervised Change Detection*”, which extends the original *Onera Satellite Change Detection (OSCD)* dataset proposed by Daudt et al. [2].

The original OSCD dataset is built from Sentinel-2 multispectral images acquired within the Copernicus programme. It consists of 24 bi-temporal image pairs located in different regions across the world (Brazil, USA, Europe, Middle East and Asia), acquired between 2015 and 2018. For each location, a pair of coregistered 13-band Sentinel-2 images is provided, with bands at spatial resolutions of 10 m, 20 m and 60 m. Pixel-wise binary change maps are available for each pair, indicating urban change (“change” vs. “no change”). For ease of use, the dataset also provides resampled versions where all bands are aligned on a 10 m grid, together with a predefined train/test split.

In our work, we focus on the bands that are most relevant for urban change detection and for comparing sensors:

- **Optical Sentinel-2 (S2):** RGB bands (B2, B3, B4), the near-infrared band (B8), and the shortwave infrared bands (B11, B12), which are sensitive to built-up areas, bare soils and vegetation.
- For each location, we consider two time points, t_1 (pre-change) and t_2 (post-change), and we compute change maps at 10 m resolution.

The multimodal extension by Ebel et al. [1] augments each OSCD pair with co-registered Sentinel-1 SAR data in VV and VH polarisations, forming a joint S1+S2 dataset. The Sentinel-1 images are temporally aligned with the Sentinel-2 acquisitions and are provided as pre- and post-change SAR intensity images, calibrated and coregistered to the Sentinel-2 grid. This makes it possible to study supervised change detection with both optical and radar data on the same scenes, using the original OSCD change maps as ground-truth.

The truth labels were originally produced in the OSCD dataset for urban change detection (e.g. new buildings, extensions of residential or industrial areas, new roads and parking lots), based primarily on the optical imagery. As a consequence, the labels emphasise anthropogenic changes rather than purely natural phenomena (vegetation phenology, seasonal water level variations, etc.). As noted by Saha et al. [3], the OSCD labels were created from optical imagery only, which can introduce inconsistencies when evaluating SAR-based change detection.

1.2 Motivation for choosing this dataset

Our research question is to compare classical change detection methods applied to Sentinel-1 (SAR) and Sentinel-2 (optical) data, and to assess which sensor

is more suitable for detecting anthropogenic changes. The multimodal OSCD dataset was chosen because it matches this objective particularly well, for several reasons:

1. **Exact combination of sensors (S1 + S2).** Most public change detection datasets are mono-modal (optical only or SAR only). Here, Sentinel-1 and Sentinel-2 are both available, co-registered and temporally aligned, with a common pixel-wise change label. This enables a fair, direct comparison of SAR and optical imagery for the same scenes.
2. **Pixel-level truth for urban change.** The dataset provides high-quality binary change masks at pixel level, with a focus on urban changes such as new buildings or new roads [2]. This matches precisely the type of changes we want to detect (anthropogenic rather than natural) and allows us to evaluate methods quantitatively using metrics such as F1-score, IoU and Cohen’s kappa, despite the strong class imbalance between “change” and “no change”.
3. **Well-established benchmark in the literature.** The original OSCD dataset was introduced for CNN-based urban change detection in Daudt et al. [2], and the multimodal S1+S2 extension has been used by Ebel et al. [1] to propose a multi-modal Siamese architecture for supervised change detection. Using the same data allows us to position our work with respect to published state-of-the-art methods, even though we deliberately restrict ourselves here to non-AI, classical image processing algorithms.
4. **Multi-sensor comparison under realistic conditions.** Because Sentinel-1 and Sentinel-2 data are provided for the same locations, dates and label maps, the dataset allows us to isolate the effect of the sensor modality. We can compare what can be achieved with optical data only, with SAR only, and analyse how each behaves on the same set of real-world urban changes.
5. **Practical size and pre-processing.** The dataset is of moderate size (24 scenes split into patches) and comes with coregistration, resampling and temporal alignment already performed. This makes it feasible, within the scope of a course project, to implement and test several non-AI methods (pixel differences, CVA, PCA, indices such as NDBI) across many patches, and to run extensive metric evaluations for both S1 and S2, without spending most of the time on low-level data preparation.

In summary, the multimodal OSCD dataset provides paired Sentinel-1 and Sentinel-2 observations, pixel-level urban change labels, and a recognised benchmark framework, which together make it an excellent choice for rigorously comparing classical change detection methods on SAR versus optical imagery in a realistic supervised setting.

1.3 If we had a 1M€ project

Our research question is the following: *how well can image processing methods detect anthropogenic changes (buildings, infrastructures, land development) from Sentinel-1 and Sentinel-2 data, and which of the two sensors is better suited for this task?* In this project, we address this question using an existing bi-temporal dataset with pixel-wise change / no-change annotations and classical (non-deep) image processing methods.

If we had the resources of a 1M€ project, we would significantly extend both the *data* and the *methods* used to tackle this question.

On the data side, the budget would be devoted to designing and acquiring a dedicated, large-scale multi-modal and multi-temporal dataset specifically targeted at anthropogenic change detection:

- **Extended spatial coverage and diversity:** collect data over hundreds of cities worldwide, covering a broad range of climates, urban morphologies and socio-economic contexts (dense historical centres, suburban sprawl, industrial zones, ports, informal settlements, etc.).
- **Richer optical data:** combine free Sentinel-2 imagery (10 m) with commercial very high resolution images (e.g. 0.3–1 m) in order to capture fine-scale changes such as individual buildings, small roads, parking lots or construction sites.
- **Richer SAR data:** go beyond Sentinel-1 by including higher-resolution or multi-frequency SAR where available (e.g. X-band, L-band), in addition to multi-temporal Sentinel-1, to better characterise structural changes and reduce ambiguities inherent to low-resolution C-band data.
- **From bi-temporal to multi-temporal sequences:** move from two time points to time series (e.g. yearly or quarterly acquisitions over 5–10 years), in order to distinguish gradual urban growth from rapid events, and permanent changes from transient or seasonal phenomena.
- **Richer ground truth:** extend the binary change / no-change labels to *multi-class change maps*, distinguishing, for example, new residential buildings, industrial expansion, road construction, demolition, land artificialisation, etc. This would require combining open GIS data (building footprints, road networks, land use) with expert labelling campaigns funded by the project.

On the methodological side, instead of being limited to classical image processing, we would adopt a state-of-the-art deep learning approach for multi-modal change detection. A natural choice would be a *multi-modal Siamese encoder-decoder / Transformer* architecture that:

- jointly ingests multi-temporal Sentinel-2 and Sentinel-1 data through two branches (one for optical, one for SAR), with appropriate positional encodings to represent time;

- uses a Siamese structure to process pre-change and post-change sequences, with shared weights to learn change-invariant representations;
- employs a fusion mechanism based on cross-attention (e.g. Transformer-style decoder blocks) to:
 - fuse optical and SAR features at multiple scales,
 - exploit complementary information (texture and colour from optical, geometry and roughness from SAR),
 - and attenuate modality-specific noise (clouds and shadows in Sentinel-2, speckle and incidence-angle effects in Sentinel-1);
- outputs a pixel-wise *multi-class* change map, trained with a combination of weighted cross-entropy (to handle class imbalance) and auxiliary losses (e.g. land-cover segmentation, self-supervised pretext tasks) to stabilise learning.

The 1M€ budget would therefore support:

- large-scale data collection, curation and annotation;
- the compute infrastructure required to train deep multimodal models on hundreds of thousands of patches;
- and a systematic benchmark against existing baselines (mono-modal deep models, simple fusion strategies, and classical non-deep methods).

In summary, with a 1M€ project we would (i) build a much larger, higher-resolution and richer multi-modal dataset specifically tailored to urban and infrastructure change, and (ii) address our research question with a deep multi-modal Siamese encoder-decoder / Transformer architecture capable of jointly exploiting Sentinel-1 and Sentinel-2 time series for fine-grained, multi-class anthropogenic change detection.

2 Description of the chosen methods

2.1 Non-AI algorithms

In this section, we describe the classical (non-AI) change detection algorithms applied to the Sentinel-1 (SAR) and Sentinel-2 (optical) data. All methods produce a *change map*, i.e. a 2D image where each pixel encodes the likelihood of change between a pre-change time t_1 and a post-change time t_2 . Unless otherwise stated, the final binary mask is obtained by applying a Gaussian smoothing followed by Otsu's automatic thresholding [4] on the change magnitude.

2.1.1 Sentinel-1 (SAR) based methods

The Sentinel-1 data are provided as two co-registered polarimetric channels, VV and VH, at times t_1 and t_2 . We exploit these channels both jointly and individually.

Weighted VV/VH difference with Otsu (compute_change_map_VVVH).

We first separate the two SAR channels (VV and VH) and compute their temporal differences:

$$d_{\text{VV}} = |\text{VV}_{t_2} - \text{VV}_{t_1}|, \quad d_{\text{VH}} = |\text{VH}_{t_2} - \text{VH}_{t_1}|. \quad (1)$$

These two difference maps are then combined linearly with weights α and β :

$$M_{\text{SAR}} = \alpha d_{\text{VV}} + \beta d_{\text{VH}}. \quad (2)$$

The resulting change magnitude M_{SAR} is normalised to $[0, 255]$, smoothed with a Gaussian filter, and binarised using Otsu's method.

In order to choose the weights α and β in a principled way, we performed a simple grid search:

- for each image pair, we evaluated all combinations (α, β) with

$$\alpha \in \{0.0, 0.1, \dots, 0.9\}, \quad \beta \in \{0.0, 0.1, \dots, 0.9\},$$

- for each pair (α, β) we computed the corresponding change map and F1-score with respect to the -truth,
- we then averaged the F1-scores over all images for each (α, β) and visualised the result as a heatmap.

Figure 1 shows the mean F1-score as a function of (α, β) . The heatmap exhibits a broad plateau of similar performance, with the brightest (best) region located approximately around

$$\alpha \in [0.6, 0.9], \quad \beta \in [0.1, 0.4].$$

This confirms the intuition that the VV channel should be given more weight than VH for urban change detection, while still benefiting from a smaller contribution of VH (texture and vegetation information). At the same time, the colour variations are very smooth: many neighbouring combinations of (α, β) yield almost identical F1-scores. This limited sensitivity can be explained by the fact that

1. the VV channel carries most of the useful signal for anthropogenic changes, so moderate reweightings of VH act mainly as a small correction term;
2. after linear fusion, the change map is normalised and thresholded by Otsu's method, which makes the result largely invariant to global rescaling of the change magnitude.

In practice, we therefore select a pair (α, β) inside this plateau region (we use $\alpha = 0.8$, $\beta = 0.2$ in our experiments) and fix it for all subsequent Sentinel-1 runs.

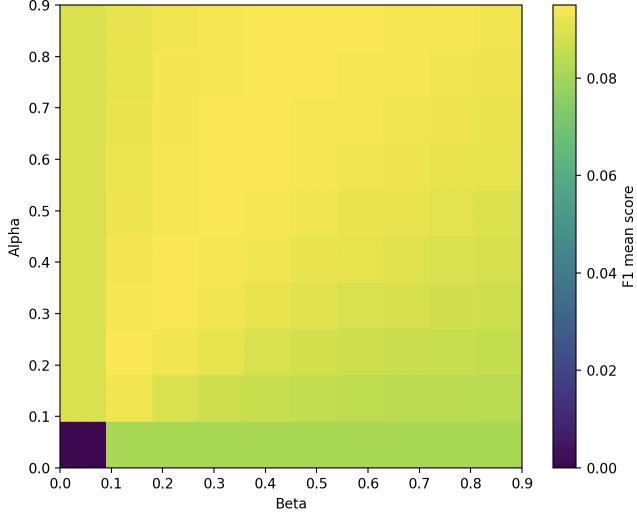


Figure 1: Mean F1-score as a function of the weighting parameters α (VV) and β (VH) for the weighted VV/VH difference method. The best-performing region lies approximately in $\alpha \in [0.6, 0.9]$ and $\beta \in [0.1, 0.4]$, and the broad plateau indicates that the method is relatively insensitive to the precise choice of (α, β) .

Single-polarisation SAR baseline (compute_change_map). As a baseline for Sentinel-1, we also apply a generic change computation function independently to each polarisation (VV or VH). Given two single-channel images I_{t_1} and I_{t_2} , we consider two variants:

- **L1 difference (int_distance = 1):**

$$d(x, y) = |I_{t_2}(x, y) - I_{t_1}(x, y)|, \quad (3)$$

- **L2 norm for multi-channel inputs (int_distance = 2):**

$$d(x, y) = \|\mathbf{I}_{t_2}(x, y) - \mathbf{I}_{t_1}(x, y)\|_2, \quad (4)$$

where $\mathbf{I}_{t_1}, \mathbf{I}_{t_2}$ are vectors when multiple channels are present.

The difference image d is then normalised, optionally smoothed by a Gaussian filter, and thresholded via Otsu’s method. This constitutes our *baseline* SAR method for VV-only and VH-only change detection.

2.1.2 Sentinel-2 (optical) based methods

For Sentinel-2, we use several bands and indices that are known to be informative for urban and land-cover changes: RGB (B2, B3, B4), NIR (B8) and SWIR (B11, B12).

Pixel-wise RGB difference (baseline for S2). The simplest optical method is to compute the difference between pre- and post-change RGB images. Let $\mathbf{I}_{t_1}(x, y)$ and $\mathbf{I}_{t_2}(x, y)$ be the RGB vectors at pixel (x, y) . We define the change magnitude as the L2 norm of the difference:

$$M_{\text{RGB}}(x, y) = \|\mathbf{I}_{t_2}(x, y) - \mathbf{I}_{t_1}(x, y)\|_2. \quad (5)$$

The map M_{RGB} is normalised, smoothed and thresholded by Otsu's method to produce a binary change mask. This method is very simple and serves as a reference, but it is sensitive to vegetation dynamics and illumination changes, which can lead to many false positives.

Single-band SWIR difference (B11). To reduce sensitivity to vegetation, we also apply the generic difference method to the SWIR1 band (B11) alone. Let $B11_{t_1}$ and $B11_{t_2}$ denote the pre- and post-change B11 images. We compute:

$$M_{B11} = |B11_{t_2} - B11_{t_1}|, \quad (6)$$

followed by normalisation, Gaussian smoothing and Otsu thresholding. The SWIR band is less affected by vegetation phenology and responds better to mineral and built-up surfaces, which helps reduce false positives due to seasonal changes in vegetation.

Urban index (NDBI) difference and masking. We also exploit the classical Normalised Difference Built-up Index (NDBI), defined from B11 (SWIR1) and B8 (NIR) as:

$$\text{NDBI} = \frac{B11 - B8}{B11 + B8 + \varepsilon}, \quad (7)$$

which was originally proposed for automated mapping of urban built-up areas from multispectral imagery [5], where ε is a small constant added for numerical stability. We compute NDBI at t_1 and t_2 , then derive a mask of built-up areas from the post-change image by selecting pixels with

$$\text{NDBI}_{t_2} > 0. \quad (8)$$

This mask is used to restrict change detection to urban or mineralised zones, effectively filtering out purely vegetated regions.

This approach separates built-up from vegetation and is particularly useful for focusing on anthropogenic changes (new buildings, extensions of residential or industrial areas, new roads or parking lots).

Change Vector Analysis (CVA) on B4, B11, B12. Change Vector Analysis is a multispectral method that measures the spectral distance between pre- and post-change feature vectors and was originally introduced for Landsat data by Malila [6]. We construct three-band vectors using the red band B4, SWIR1 (B11) and SWIR2 (B12):

$$\mathbf{V}_1 = [B4_{t_1}, B11_{t_1}, B12_{t_1}], \quad \mathbf{V}_2 = [B4_{t_2}, B11_{t_2}, B12_{t_2}]. \quad (9)$$

The change magnitude is then defined as the Euclidean norm of the difference:

$$M_{\text{CVA}} = \|\mathbf{V}_2 - \mathbf{V}_1\|_2. \quad (10)$$

After normalisation and Otsu thresholding, we obtain a binary change map. CVA is inherently multispectral and less sensitive to global radiometric shifts; it captures the spectral structure of changes more effectively than single-band differences.

2.1.3 Multiband and advanced optical method: bi-temporal PCA

Bi-temporal PCA (make_PCA_features). Principal Component Analysis (PCA) has long been used for multitemporal change detection by enhancing the variance related to change in stacked images [7]. In our case, for each pixel we stack several bands from t_1 and t_2 into a 6-dimensional vector:

$$\mathbf{x}(x, y) = [B4_{t_1}, B11_{t_1}, B12_{t_1}, B4_{t_2}, B11_{t_2}, B12_{t_2}]. \quad (11)$$

PCA is then applied over all pixels of a patch, yielding principal components PC_k . The first component typically captures the common information between t_1 and t_2 , while subsequent components encode differences. We use the second and third components as *change components* and define the change magnitude as:

$$M_{\text{PCA}} = \sqrt{\text{PC}_2^2 + \text{PC}_3^2}. \quad (12)$$

The map M_{PCA} is normalised and thresholded using Otsu's method. This approach is a robust spectral change detector: PC1 mainly represents invariant content, whereas PC2 and PC3 highlight variations, making the method relatively insensitive to illumination differences and global radiometric biases.

2.1.4 Generic difference function (multi-sensor baseline)

Finally, the generic `compute_change_map` function described above is applied as a common baseline to different inputs:

- RGB (Sentinel-2),
- individual spectral bands (B8, B11, etc.),
- SAR channels (VV, VH),
- any stacked multi-spectral or multi-sensor vector.

With its two variants (L1 difference and L2 norm), this function provides a unified baseline for comparing the *intrinsic* ability of each sensor or band combination to support classical change detection.

2.2 AI algorithms

In this section, we focus on describing *AI-based* methods for change detection algorithms applied to Sentinel-2 optical data. These methods share a common goal: to produce a *change map*, i.e., a 2D image where each pixel represents the likelihood of change between a pre-change time t_1 and a post-change time t_2 . The objective of this analysis is to evaluate the extent to which AI-based approaches (specifically deep learning) can enhance change detection performance in this remote sensing task compared to traditional, non-AI methods. To this end, we will review state-of-the-art algorithms in the field, carefully selecting models based on our specific requirements and constraints.

2.2.1 Constraints

To select the most suitable model for our study case, we must first define the constraints that limit our field of action. Given our very small dataset (24 bitemporal acquisitions), our options for deep learning are significantly restricted. Indeed small dataset severely limits—or even makes impossible—the task of training a model, as it would face a high risk of overfitting due to the limited availability of data. Additionally, the data is highly heterogeneous, consisting of satellite images from cities around the world, including deserts, highly urbanized areas, and vegetated regions. This heterogeneity further complicates the task. Even if the model could learn from such a limited amount of data, evaluating its performance in a robust way would remain a challenge. The scarcity of data would make it difficult to obtain reliable evaluation metrics. We considered using another dataset with more image pairs to enable training, but unfortunately, we were unable to find a larger dataset that includes both Sentinel-1 and Sentinel-2 bitemporal acquisitions. Using a dataset with only Sentinel-2 images would have allowed us to explore AI and *non-AI* techniques, but this would have required abandoning the comparison with *non-AI* techniques for Sentinel-1 acquisitions—an approach we deemed unsuitable, given our original focus on analyzing and comparing SAR and Optical images. Furthermore, our limited computational resources (GPU capacity) make the training task unfeasible.

2.2.2 Alternative aproches

In our aim to incorporate *ai* techniques, we chose a different approach from training or fine-tuning a model. Instead, we focused on models capable of zero-shot inference on our images. We relied on various research papers that propose pre-trained, open-source models.

2.2.3 Segment Anything Model

This model, developed by Meta AI Research, is presented in the publication [8]. The paper demonstrates the model’s ability to perform remote sensing tasks as a generalist model. It enables us to approach image segmentation through zero-shot inference on our data. The model can segment images either by specifying

regions (via pixel selection) or by automatically generating segmentation masks for the entire image. Our initial goal was to segment both images separately (at times t_1 and t_2) and then analyze the differences between the two resulting segmentations. We first explored the automatic segmentation method. However, this approach posed a significant challenge: differences in illumination between images can cause the model to segment identical regions inconsistently, making it unsuitable for our change detection task. An alternative approach with this model involves providing point query in the input context —i.e., manually selecting points in regions we want to segment. This yields more useful segmentation masks for comparing changes between t_1 and t_2 . While this method only requires light re-annotation (placing a few points on the images), it shifts the task from fully automatic change detection to a semi-supervised process.

2.2.4 Segment Any Change

The model introduced in [9] is particularly well-suited to the goals of our study. Building on the foundational architecture of the Segment Anything Model (SAM) [8], this approach harnesses SAM’s powerful zero-shot segmentation capabilities to perform automated change detection between two images of the same location acquired at different times, t_1 and t_2 . SAM’s ability to generate high-quality segmentation masks for arbitrary images—without requiring task-specific fine-tuning—makes it an ideal tool for our application. The model in [9] extends SAM’s functionality by specifically adapting it to identify and isolate regions of change between bitemporal images.

By leveraging SAM’s robust segmentation framework, this model produces detailed segmentation masks for both t_1 and t_2 . These masks highlight the areas where changes have occurred, such as urban expansion, vegetation loss, or other land-use transformations. The next critical step involves combining or comparing these masks to generate a comprehensive 2D change map. This map serves as a visual and quantitative representation of the differences between the two time points, effectively pinpointing regions of interest for further analysis.

The resulting 2D change map can then be systematically evaluated against our ground truth data, enabling us to assess the model’s accuracy, precision, and reliability. This evaluation is essential for benchmarking the performance of this *ai*-driven technique against traditional, *non-ai* change detection methods.

3 Global analysis of the results

3.1 Analysis of the results of non-AI methods

In this section, we analyse the quantitative performance of all classical (non-AI) change detection methods implemented in this work. The methods are based on standard image processing techniques such as pixel-wise differences, Change Vector Analysis (CVA), bi-temporal PCA and spectral indices. Overall, the average performance remains modest: F1-scores range between approximately 0.17 and 0.26 for Sentinel-2, and around 0.09 for Sentinel-1. However, the

relatively large standard deviations and the high maximum values show that, for some patches, certain methods can reach quite convincing performances ($F1 \approx 0.7$ and $\text{IoU} \approx 0.5$ for the best Sentinel-2 configurations). This suggests that these approaches are highly dependent on the scene context (type of change, cloud cover, geometry, class imbalance, etc.).

3.1.1 Summary table of results

Table 1: Summary of quantitative results ($F1$ -score, IoU and Cohen’s κ) for all change detection methods on Sentinel-2 (optical) and Sentinel-1 (SAR). Values are reported as mean \pm standard deviation and [min, max] over all patches.

Sensor	Method	$F1$ mean \pm std	$F1$ [min, max]	IoU mean \pm std	IoU [min, max]	κ mean \pm std	κ [min, max]
Sentinel-2 (optical) methods							
S2	RGB diff + Otsu	0.2627 \pm 0.2007	[0.0125, 0.6802]	0.1680 \pm 0.1463	[0.0063, 0.5154]	0.2352 \pm 0.1972	[0.0037, 0.6560]
S2	B11 diff + Otsu	0.1770 \pm 0.2080	[0.0000, 0.7183]	0.1141 \pm 0.1501	[0.0000, 0.5605]	0.1485 \pm 0.2025	[-0.0065, 0.6958]
S2	B11 + NDBI mask	0.0741 \pm 0.0552	[0.0043, 0.1866]	0.0394 \pm 0.0305	[0.0022, 0.1029]	0.0217 \pm 0.0291	[-0.0014, 0.1257]
S2	B8 diff + Otsu	0.1584 \pm 0.1718	[0.0111, 0.6891]	0.0974 \pm 0.1229	[0.0056, 0.5257]	0.1231 \pm 0.1655	[-0.0113, 0.6660]
S2	CVA (B4, B11, B12)	0.2024 \pm 0.2215	[0.0000, 0.7322]	0.1325 \pm 0.1627	[0.0000, 0.5776]	0.1755 \pm 0.2170	[-0.0011, 0.7105]
S2	PCA (B4, B11, B12)	0.1674 \pm 0.1420	[0.0218, 0.5966]	0.0988 \pm 0.0967	[0.0110, 0.4251]	0.1320 \pm 0.1300	[0.0132, 0.5632]
S2	AI RGB	0.235 \pm 0.1638	[0.0232, 0.5739]	0.1438 \pm 0.1158	[0.0117, 0.4024]	0 \pm 0	[0, 0]
Sentinel-1 (SAR) methods							
S1	VV+VH weighted	0.0947 \pm 0.0729	[0.0057, 0.2951]	0.0513 \pm 0.0420	[0.0028, 0.1731]	0.0498 \pm 0.0562	[-0.0031, 0.2376]
S1	SAR CVA (VV, VH)	0.0905 \pm 0.0689	[0.0065, 0.2720]	0.0488 \pm 0.0393	[0.0033, 0.1574]	0.0441 \pm 0.0500	[-0.0023, 0.2062]
S1	VV diff + Otsu	0.0695 \pm 0.0511	[0.0015, 0.2157]	0.0367 \pm 0.0283	[0.0008, 0.1209]	0.0180 \pm 0.0359	[-0.0639, 0.1411]
S1	VH diff + Otsu	0.0662 \pm 0.0466	[0.0052, 0.1849]	0.0348 \pm 0.0255	[0.0026, 0.1019]	0.0162 \pm 0.0306	[-0.0804, 0.0982]

3.1.2 Analysis of Sentinel-2 (optical) methods

RGB method (pixel-wise difference + Otsu). The RGB method is the best-performing approach on average among all Sentinel-2 methods:

- mean $F1$ -score: ≈ 0.26 (max ≈ 0.68),
- mean IoU : ≈ 0.17 (max ≈ 0.52),
- mean Kappa: ≈ 0.24 .

The method itself is very simple (multi-channel difference on RGB followed by Otsu thresholding), but it benefits from two important factors:

- the change maps were annotated based on optical data,
- anthropogenic changes (buildings, roads, urban areas) have a fairly visible signature in RGB space.

The high standard deviation ($F1$ std ≈ 0.20) and the very low minimum $F1$ (≈ 0.01) nevertheless indicate that the method is unstable: it performs very well on some scenes (clear changes, little vegetation), and much worse on others (seasonal variations, shadows, illumination differences).

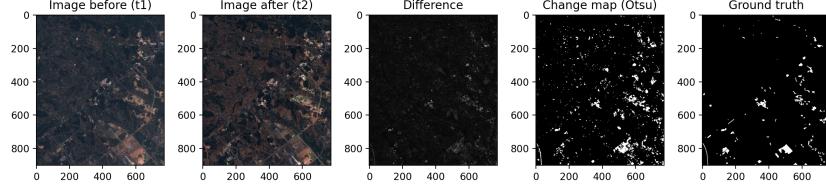


Figure 2: Example of change detection with the RGB difference + Otsu method: pre-change image (t_1), post-change image (t_2), difference map, predicted change map, and ground-truth mask.

B11 method (single SWIR band). For the B11 (SWIR1) method, we obtain:

- mean F1-score: ≈ 0.18 ,
- mean IoU: ≈ 0.11 ,
- mean Kappa: ≈ 0.15 .

The average performance is lower than RGB, but the maximum values (F1 max ≈ 0.72 , IoU max ≈ 0.56) show that B11 can be highly discriminative for certain types of scenes (strongly mineralised areas, strong contrast between built-up and non-built-up surfaces). Since SWIR is less sensitive to vegetation variations than NIR, this method appears better suited to cases where urbanisation dominates and vegetation remains relatively stable.

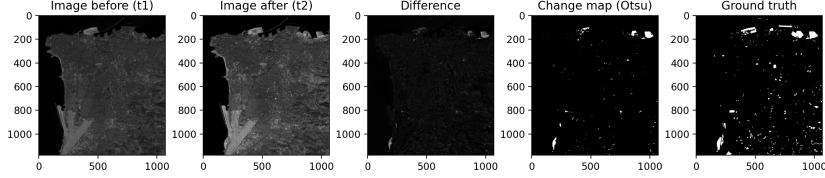


Figure 3: Example of change detection with the B11 (SWIR1) difference method.

B11 masked by NDBI (B11_mask_NBDI). For the B11_mask_NBDI method, we obtain:

- mean F1-score: ≈ 0.07 (very low),
- mean IoU: ≈ 0.04 ,
- mean Kappa: ≈ 0.02 .

This method applies a mask based on the built-up index (NDBI) to focus on urban areas. The global performance is clearly degraded, which indicates that:

- the mask is probably too restrictive,
- a large fraction of truly changed pixels is excluded, leading to a very low recall and thus a collapse of F1 and IoU.

This method illustrates well the precision vs. recall trade-off: by focusing too much on strongly urban areas ($\text{NDBI} > 0$), many real changes are missed (extensions of semi-urban areas, bare soils, construction sites, etc.). Conceptually, the idea is sound, but in its current implementation it discards too much information.

B8 method (NIR). For the B8 (NIR) method, we obtain:

- mean F1-score: ≈ 0.16 ,
- mean IoU: ≈ 0.10 ,
- mean Kappa: ≈ 0.12 .

NIR is highly sensitive to vegetation. In practice, B8 introduces:

- false positives due to seasonal variations and vegetation phenology (leaf-on / leaf-off),
- effects related to soil moisture.

The results are therefore logically worse than RGB on average, and fairly close to those of B11 / PCA, with a large variability from one scene to another.

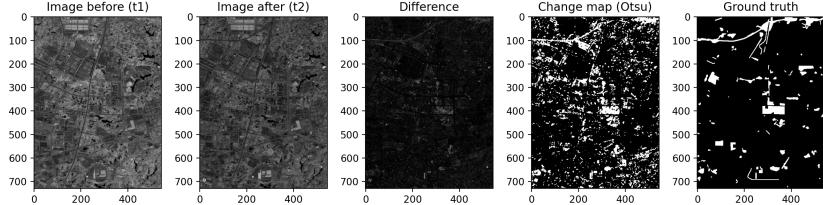


Figure 4: Example of change detection with the B8 (NIR) difference method.

CVA on B4, B11, B12. For the multispectral CVA method on B4, B11 and B12, we obtain:

- mean F1-score: ≈ 0.20 ,
- mean IoU: ≈ 0.13 ,
- mean Kappa: ≈ 0.18 ,
- maximum F1-score: ≈ 0.73 , maximum IoU: ≈ 0.58 .

The multispectral CVA yields intermediate results:

- better than B8 alone or B11_mask_NBDI,
- slightly worse than RGB on average,
- but with very high maxima, comparable to or even higher than RGB.

This shows that CVA, by exploiting several bands (red + SWIR), can perform very well when the changes translate into a clear spectral variation. However, it remains sensitive to:

- variations in bare soils,
- global radiometric differences,
- and the absence of any explicit modelling of spatial context.

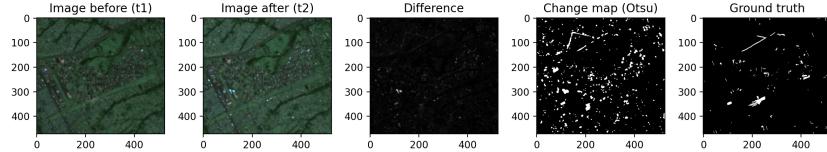


Figure 5: Example of change detection with the multispectral CVA method on bands B4, B11 and B12.

Bi-temporal PCA. For the bi-temporal PCA method, we obtain:

- mean F1-score: ≈ 0.17 ,
- mean IoU: ≈ 0.10 ,
- mean Kappa: ≈ 0.13 .

Bi-temporal PCA, which was intended to isolate change components (PC2, PC3), produces results comparable to B11 or B8, but inferior to RGB and CVA on average. Several factors may explain this:

- the use of a global Otsu threshold on the magnitude of the components may be too coarse,
- there is no spatial regularisation (no morphological post-processing or CRF).

Nevertheless, on some patches ($F1 \max \approx 0.6$), the method demonstrates that PCA can effectively isolate changes when they are sufficiently strong.

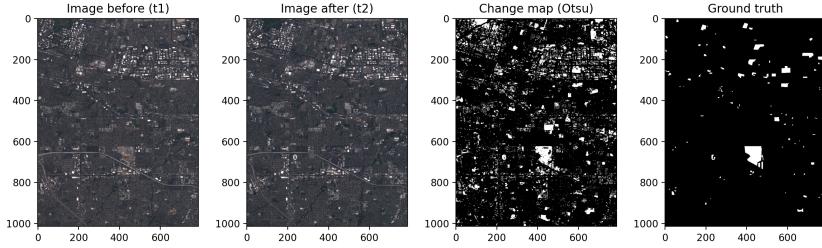


Figure 6: Example of change detection with the bi-temporal PCA method.

Segment Any Change, AI based method For the segment any change model, we obtain:

- mean F1-score: ≈ 0.24 ,
- mean IoU: ≈ 0.14 ,
- mean Kappa: ≈ 0 .

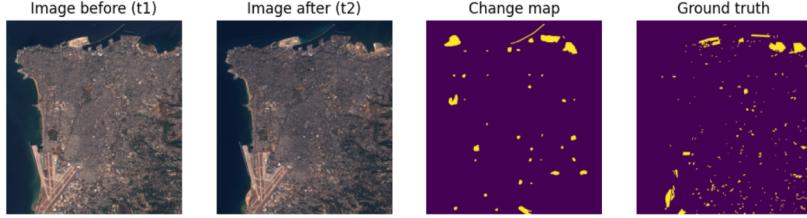


Figure 7: Example of change detection with the Segment Any Change model.

The technique, which aims to use a zero-shot deep learning model to outperform standard algorithms, fails to do so. Upon analyzing the metrics, it is clear that the model struggles to detect changes accurately. And it does not perform better than $RGB\ diff + Otsu$. The F1-score and IoU are low, and Kappa is zero, indicating no meaningful agreement with the ground truth. Additionally, the wide ranges in F1 and IoU scores suggest that the model’s performance is highly inconsistent across different areas. The primary issues likely stem from inherent limitations of SAM, particularly its difficulty in controlling segmentation granularity, which causes it to miss fine details. Compounding this problem is the limited resolution of the images, which further reduces the model’s precision and overall accuracy in change detection.

3.1.3 Analysis of Sentinel-1 (SAR) methods

Overall, all SAR-based methods are significantly less effective than their optical counterparts, with mean F1-scores between approximately 0.06 and 0.09 and IoU values around 0.04–0.05.

Weighted fusion VV+VH (S1_VVVH). For the VV+VH fusion (weighted difference + Otsu), we obtain:

- mean F1-score: ≈ 0.095 (best SAR method),
- mean IoU: ≈ 0.051 ,
- mean Kappa: ≈ 0.05 .

The combination of the two polarisations slightly improves performance compared to using each channel independently. This confirms that:

- VV is sensitive to built-up surfaces,
- VH brings complementary information on texture and vegetation,
- the fusion better captures the global signature of urban changes.

However, even this method remains far below the best optical methods.

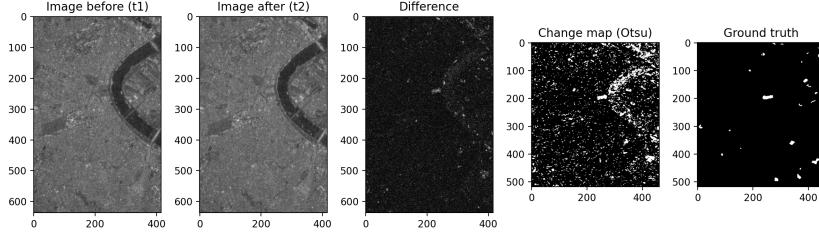


Figure 8: Example of change detection with the weighted VV+VH fusion method (S1_VVVH).

CVA on VV and VH. For the CVA applied to the two SAR channels (VV and VH), we obtain:

- mean F1-score: ≈ 0.09 ,
- mean IoU: ≈ 0.049 ,
- mean Kappa: ≈ 0.044 .

The SAR CVA yields results very close to the simple VV+VH fusion. It confirms that the (here: polarimetric) spectral information is useful, but:

- speckle noise,
- differences in incidence angle,
- and radiometric variations between acquisitions

introduce a substantial amount of noise in the change map.

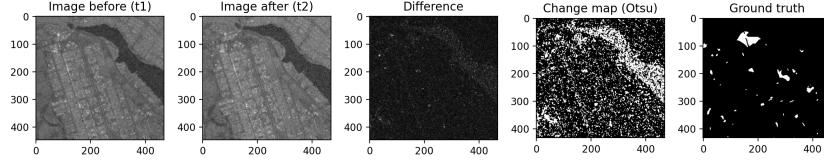


Figure 9: Example of change detection with the CVA method on the VV and VH SAR channels.

VV only and VH only. For VV-only and VH-only methods, we obtain:

- mean F1-score VV: ≈ 0.07 ; VH: ≈ 0.066 ,
- mean IoU VV/VH: ≈ 0.036 – 0.035 ,
- Kappa values are very low, sometimes even slightly negative for some scenes.

The channels taken separately are clearly insufficient for reliable change detection. We observe:

- many false positives related to speckle,
- strong sensitivity to variations in soil moisture, building orientation, etc.

These results confirm that raw SAR, exploited only via intensity differences and CVA, is difficult to use for fine-grained change detection without:

- advanced speckle filtering,
- rigorous radiometric calibration,
- or more sophisticated methods (Bayesian models, Markov Random Fields, deep learning).

3.1.4 Comparison Sentinel-1 vs Sentinel-2

The comparison of the metrics clearly shows that:

- Sentinel-2 (optical) methods are substantially superior to Sentinel-1 (SAR) methods in this experimental setting.
- The best S2 method (RGB difference + Otsu) reaches a mean F1-score of ≈ 0.26 (max ≈ 0.68) and a mean IoU of ≈ 0.17 ,
- whereas the best S1 method (VV+VH fusion) reaches only a mean F1-score of ≈ 0.095 and a mean IoU of ≈ 0.05 .

Kappa values confirm this gap: $\kappa \approx 0.23$ for RGB vs. $\kappa \approx 0.05$ for VV+VH, which means that SAR-based models behave almost like near-random classifiers on the change class, while optical models capture a real structure in the changes.

Several factors can explain these results:

Nature of the annotations. The change maps were initially derived from optical data; consequently, the visible changes are aligned with the Sentinel-2 signature and less with SAR-specific artefacts.

Readability of urban changes. Anthropogenic changes (constructions, roads, mineralised areas) produce a very clear signature in optical imagery (colour, texture), whereas in SAR they are mixed with other effects (surface roughness, object orientation, moisture).

Processing of SAR data. In this project, the SAR methods tested remain relatively simple (difference, CVA, linear fusion) and do not fully correct for:

- speckle,
- variations in incidence angle,
- fine radiometric calibration issues.

Consequently, limited performance is expected without more advanced SAR processing.

4 Conclusion

Overall, the results obtained with classical image processing methods remain relatively low, highlighting the limitations of approaches that rely solely on simple radiometric operations, or zero-shoot AI technique, especially when compared to deep learning or advanced modeling techniques trained extensively on specific data. Although some scenes occasionally show good performance, the very strong variability across contexts, the sensitivity to acquisition conditions and the lack of explicit modelling of spatial or complex spectral context lead to modest global scores. These methods therefore provide an interesting first insight, but are insufficient for robust large-scale change detection. This conclusion naturally motivates the next section, where AI- and deep-learning-based approaches can be explored to assess to what extent trained models can overcome these structural limitations.

References

- [1] P. Ebel, S. Saha, and X. X. Zhu, “Fusing multi-modal data for supervised change detection,” in *ISPRS Congress*, (Nice, France / Virtual), July 2021.
- [2] R. Caye Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, “Urban change detection for multispectral earth observation using convolutional neural networks,” in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, (Valencia, Spain), July 2018.
- [3] S. Saha, P. Ebel, and X. X. Zhu, “Self-supervised multisensor change detection,” *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

- [4] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [5] Y. Zha, J. Gao, and S. Ni, “Use of normalized difference built-up index in automatically mapping urban areas from tm imagery,” *International Journal of Remote Sensing*, vol. 24, no. 3, pp. 583–594, 2003.
- [6] W. A. Malila, “Change vector analysis: An approach for detecting forest changes with landsat,” in *Proceedings of the 6th Annual Symposium on Machine Processing of Remotely Sensed Data*, (West Lafayette, IN, USA), pp. 326–335, 1980.
- [7] T. Fung and E. LeDrew, “Application of principal components analysis to change detection,” *Photogrammetric Engineering and Remote Sensing*, vol. 53, no. 12, pp. 1649–1658, 1987.
- [8] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, “Segment anything,” *arXiv:2304.02643*, 2023.
- [9] Z. Zheng, Y. Zhong, L. Zhang, and S. Ermon, “Segment any change,” in *Advances in Neural Information Processing Systems*, 2024.