# MUMUCD: A MUlti-modal MUlti-class Change Detection dataset

F. Serva, A. Sebastianelli, *Member, IEEE*, B. Le Saux, *Member, IEEE,* and F. Ricciuti

which comprises...

***Abstract*—This work introduces the MUlti-modal MUlti-class Change Detection (MUMUCD) dataset, comprising ?? globally distributed georeferenced bitemporal pairs obtained by multiple space-borne sensors. Acquisitions over heterogenous terrains (e.g., urban, rural, forests, deserts) for the period 2019-2024 are processed to provide the first large scale curated dataset combining Synthetic Aperture Radar (Sentinel-1), multispectral (Sentinel-2) and hyperspectral (PRISMA) data with ancillary information. Provided with a common spatial resolution of 10 m and a size of 1536×1536 square pixels, MUMUCD allows to extract several thousands non-overlapping patches with size 128×128 square pixels, enabling data-intensive applications. The choice of dates and locations is based on the availability of PRISMA pairs of images, taken approximately between 1 and 4 years apart in the same time of year to minimize seasonal effects. While scenes with low cloudiness are prioritized, cloudy pixels are not excluded as different data combinations (e.g., SAR and optical) can be exploited to mitigate the atmospheric effects. Beyond coregistered data, we provide surface elevation, land cover and binary change maps obtained by processing the Dynamic World dataset, that we use as ground truth for benchmarking machine and deep learning change detection algorithms. MUMUCD gives easy access to multispectral, SAR and hyperspectral data (a unique feature of this dataset) supporting a number of tasks like semantic change detection, cloud removal, image translation, spectral compression/super-resolution, and more.**

***Index Terms*—Change detection, multimodal dataset, deep learning, hyperspectral data**

3 (sentence to add)

## I. Introduction

CHANGE detection (CD) approaches are widely used with remotely sensed images, starting more than four decades ago with seminal studies like [1] who focused on vegetation status. The availability of multispectral (MS) sensing instruments increased the range of possible applications of CD methods, as the information obtained by combining multiple bands can provide deeper insights on surface conditions. One limitation of supervised methods is the need for ground truth (GT) data, which can be strongly application-specific and generally require either ground measurements (which is sparse) or manual annotation. The latter is moreover resource-intensive and prone to error, especially when applied to large, multimodal archives. Since reliable GT data are not always available, significant attention has been given in the last years [2] to the use of machine and deep learning (ML and DL), given their generalization capabilities upon a sufficiently complete training step.

F. Serva is with the National Research Council, Rome, Italy. A. Sebastianelli, F. Serva and B. Le Saux (now at the European Commission) are with Φ-lab, European Space Agency, Frascati, Italy. F. Ricciuti is with TRE-ALTAMIRA, Milan, Italy.

F. Serva and A. Sebastianelli contributed equally to this work (e-mail: federico.serva@terrarum.eu).

Manuscript received MMM DD, YYYY; revised MMM DD, YYYY.

Initially used mostly for surface characterization, for example mineral mapping, hyperspectral (HS) data from space-borne sensors are nowadays becoming increasingly available [3]. HS measurements improve upon the spectral resolution by a factor ten or more, enabling a number of downstream tasks, including CD applications [4]. While some CD studies relying on HS data appeared in the last few years [5]–[7], they have generally used a limited number of images or resorted to unsupervised approaches [8]. With this work we aim to provide a large scale objectively annotated dataset including multimodal remote sensing data, which will help tackling the problem of sparse annotation for land cover CD applications.

4 (citations) or a sentence at the end of the paragraph, here above.

## II. Methods and data

useless to say

### A. Data collection and preprocessing

As mentioned, the MUMUCD (multimodal multiclass change detection) dataset combines data from different space-borne sensors with ancillary data to provide a global and diverse archive useful for multiple applications.

Synthetic aperture radar (SAR) data is obtained from the Sentinel-1 (S1) mission [9] of the European Space Agency (ESA). The mission includes multiple satellites, S1A launched in 2014 and still operational, S1B operating between 2016 and 2021, and S1C launched in late 2024. Acquisitions from the first two satellites are considered for MUMUCD. The S1 C-band radar has all-weather capabilities, but given the strong dependence on sensing geometry, in MUMUCD we select data taken along the same relative orbit for the Ground Range Detected (GRD) mode, with a spatial resolution of 5×20 $m^2$. Backscattering in two polarizations and observation geometry are provided.

The Sentinel-2 (S2) mission [10] provides MS data in the visible-near infrared (VNIR) and short-wave infrared (SWIR) ranges sensed by the Multi-Spectral Imager instrument. The first satellite, S2A, is operational since 2015, followed by S2B in 2017 and recently by S2C. After atmospheric correction (L2A collection), reflectances for 12 bands are available, with spatial resolution between 10 and 60 m, and data from S2A and S2B is considered in this work.

A key asset of MUMUCD is the inclusion of HS data from PRISMA (Precursore Iperspettrale della Missione Operativa), launched by the Italian Space Agency and operational since 2019 [11]. Data is collected by two HS sensors in the VNIR and SWIR ranges, with 230 non-overlapping bands between 400 and 2500 nm, with an approximate resolution of 10 nm and a spatial resolution of 30 m. VNIR data are prioritized over those from the SWIR sensor due to their higher quality, and atmospheric-corrected (L2D) surface reflectances are processed for MUMUCD.

words higlighted in blue could be put in bold for easier reading; HS could be written hyperspectral to ensure it.

The dataset is further complemented with data from the Dynamic World (DW) dataset [12], providing global and near-real time land cover (LC) classes based on L2A Sentinel-2 data. Available DW maps for a given location are collected around the target date, and the dominant class for each pixel is taken. Pixels not classified in the nine default DW classes (coded 0-8) are assigned to class 9 (no data). Binary change maps are then computed by differencing bitemporal LC maps (no data pixels are conservatively set to *no change*), and further filtered over a disk with three pixels radius to focus on large scale changes. Surface elevation derived from mosaicking of the digital elevation models from SRTM and ASTER is provided as a static field, available at 30 m spatial resolution.

The overall procedure followed to generate MUMUCD is illustrated in the upper portion of Fig. 1. PRISMA data are obtained from the mission data portal, while other Sentinel and ancillary datasets are available from the Google Earth Engine (GEE) platform. GEE data for Sentinel missions are gathered over a period of at least 8 weeks (longer for cloudy regions, such as the tropics) centered on the target date. Least cloudiness and closeness in time to the corresponding PRISMA acquisition are the criteria used to select S2 and S1 images, respectively.

All data are interpolated from their native spatial resolution to 10 m, the resolution of visible bands of the S2 sensor. Different datasets are harmonized by means of a coregistration step with the gefolki tool [13], by using the S2 red band of the after image as target. The different modalities are provided for the same area of interest (1536×1536 squared pixels), with HS data is interpolated to 10m resolution with a nearest neighbour approach. The coregistration procedure depends on a number of independent parameters to be determined empirically. Values used for MUMUCD, reported for reference in Table I, have been found by manual testing and visual inspection, aiming at reducing artifacts (such as noise in coastal areas for optical and spurious shifts in SAR acquisitions) across different scenes. As known from previous works [14], coregistration of SAR-optical data is particularly challenging, thus a less aggressive approach was taken.

TABLE I
PARAMETERS USED FOR THE COREGISTRATION PROCEDURE WITH TARGET S2

| Modality | Radius | Level | Rank | Iteration |
|---|---|---|---|---|
| HS | [64,32] | 3 | 4 | 4 |
| MS | [36,18] | 4 | 4 | 4 |
| SAR | [64,32,16] | 1 | 2 | 2 |

A small number of pairs includes ephemeral changes (such as snow in the Bavaria, Cukotka, Eyjafjöll, Kitami and Kirtland scenes) which can be used to test the advantage of multimodal data. Files are provided with geographical coordinates and detailed metadata, such as information on physical units, licences, and information on the source files. Scenes of the MUMUCD database are named based on notable geographical features (e.g., a town or a river delta) close to the area of interest. An example of the modalities provided by MUMUCD is given in Fig. 1 (bottom) for the area of Beirut, Lebanon, showing the different sensitivities of optical and radar sensing, and how bitemporal land cover maps can be combined to provide an objective binary change map.

For each scene, before/after pairs for all modalities, DEM, land cover and change maps are provided in NetCDF format with metadata.

TABLE II
OVERVIEW OF AVAILABLE DATASETS PROVIDING HYPERSPECTRAL (HS), MULTISPECTRAL (MS), SYNTHETIC APERTURE RADAR (SAR) AND LAND COVER (LC) DATA. A † SYMBOL INDICATES MULTITEMPORAL DATASETS. THE HIGHEST SPATIAL RESOLUTION IS 10 M FOR ALL EXCEPT FOR HYSPECNET11K (30 M).

| Name | Mode | Area (km$^2$) | Span |
|---|---|---|---|
| OSCD† [18] | MS | 1728 | 4 yrs |
| OSCD+S1† [19] | MS/SAR | 1728 | 4 yrs |
| SEN12MS [20] | MS/SAR/LC | 1183986 | 1 yr |
| SEN12MS-CR-TS† [15] | MS/SAR | 2544000 | 1 yr |
| BigEarthNet [21] | MS/LC | 850069 | 1 yr |
| BigEarthNetMM [16] | MS/SAR/LC | 850069 | 1 yr |
| HySpecNet11k [22] | HS | 169323 | 1 wk |
| **MUMUCD**† | MS/SAR/HS/LC | 40108 | 6 yrs |

## B. Related works

With the advent of DL architectures, a number of datasets appeared to support model training. As shown in Table II, the construction of these archives greatly benefited from the availability of the Sentinel data, providing near-global coverage of MS and SAR acquisitions. More recently, with the ENMap mission, HS datasets begun to be produced. Due to the geometry of PRISMA acquisitions (30×30 km$^2$) and the need for nearly cloud-free bitemporal acquisitions, the spatial extent of MUMUCD is smaller than datasets based on Sentinels such as SEN12MS-CR-TS [15] or BigEarthNetMM [16]. The yet unreleased SpectralEarth dataset [17] plans to further increase the scale of HS datasets, but with limited multitemporal acquisitions and without labels. In contrast, MUMUCD represents a much larger labelled dataset compared to OSCD [18] with detailed spectral information thanks to the PRISMA data.

HS datasets specifically used for CD are generally limited in terms of spatial coverage and number of acquisitions [5], [23], and existing multimodal datasets (Table II) do not provide HS data. Compared to MS data, HS imaging provides much more detailed spectral information useful for various applications such as water and land surface characterization [3], but data are not acquired continuously, hampering their use for CD.

Similar to OSCD, the maximum time difference between acquisitions in MUMUCD is 4 years.

## C. Benchmarks

We have employed the MUMUCD dataset to verify the performances of widely used approaches for CD, namely a $k$-means clustering [24] method, a random forest (RF) network [25] and a siamese UNet (sUNet) architecture [18]. Since the clustering approach is unsupervised, we simply assign the *change* label to the minority class (based on the number of

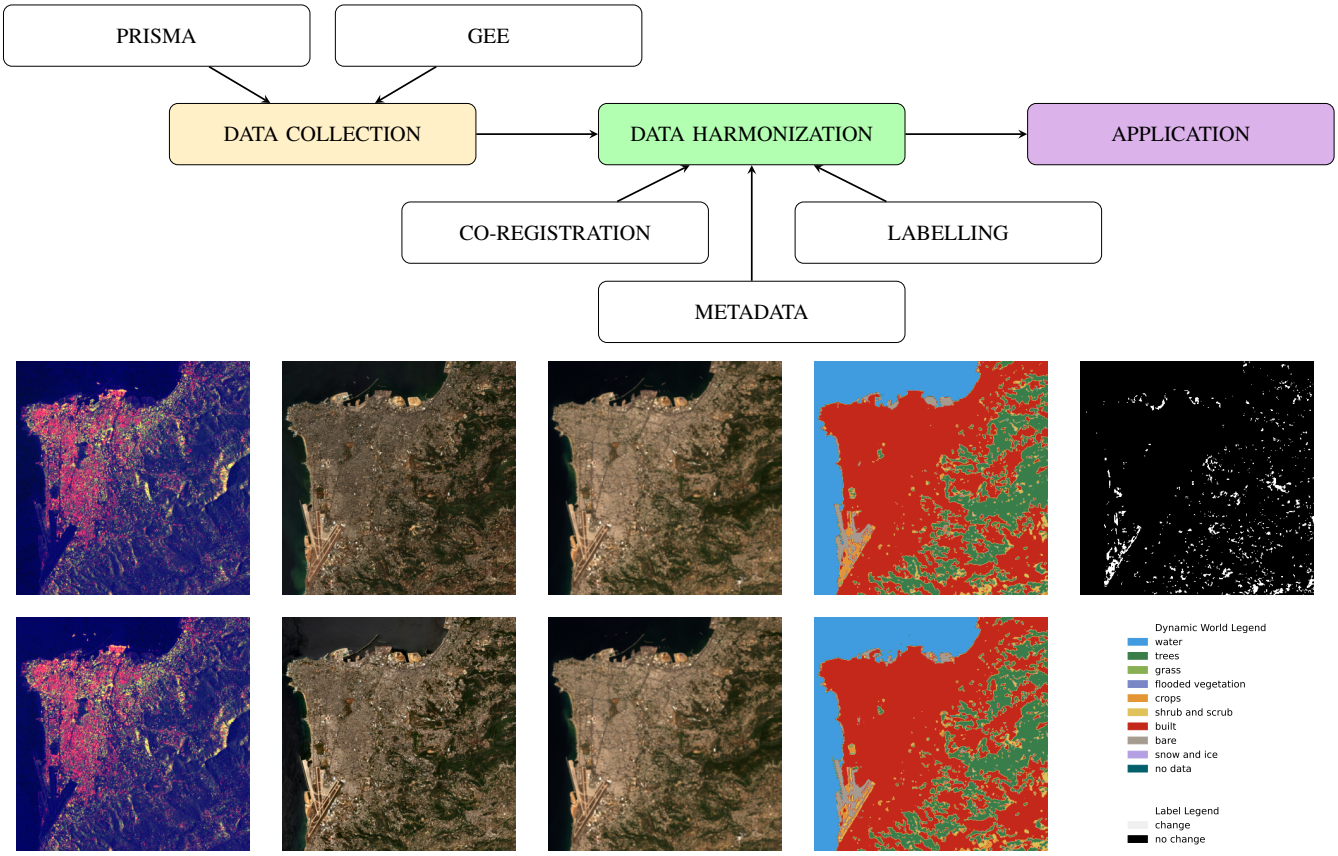


Fig. 1. Workflow of the procedure used to prepare the MUMUCD dataset (top). The bottom panel reports before (top row) and after (bottom row) images for a selected scene, with (from left to right) a Sentinel-1 VV-VH composite, Sentinel-2 and PRISMA true color composites, Dynamic World and the binary change map.

pixels), without further tuning or reliance on ground truth. While many other models could be consider for the binary CD task, with the proposed selection we aim at covering both ML and DL methods, with and without supervision.

TABLE III
DESCRIPTION OF MAIN MODELS' SETTINGS.

| Setting | k−means | RF | sUnet |
|---|---|---|---|
| Loss | - | Gini | NLLLoss |
| Estimators/weights | - | 100 | 1.1M |
| N classes | 2 | 2 | 2 |
| N init | 10 | - | - |
| Max iterations/epochs | 600 | - | 50 |

To determine the advantages of multimodality, we tested the selectd models under multiple input combinations: [S2] for only Sentinel-2, [S2+S1] for both Sentinel-2 and Sentinel-1, [S1+S2+PRS] for Sentinel-1/2 and PRISMA, [S1+(S2∩PRS)] Sentinel-1 and common bands between Sentinel-2 and PRISMA and [S1+S2+(PRS\S2)] Sentinel-1, Sentinel-2 and complementary PRISMA bands.

## III. RESULTS AND DISCUSSION

MUMUCD provides multimodal data for ?? globally distributed scenes, as illustrated in Fig. 2. The highest spatial density is found over the European region, but acquisitions in remote regions, such as the high latitudes, are included. The time between subsequent acquisitions ranges between three months and four years, with an average time span of two years.

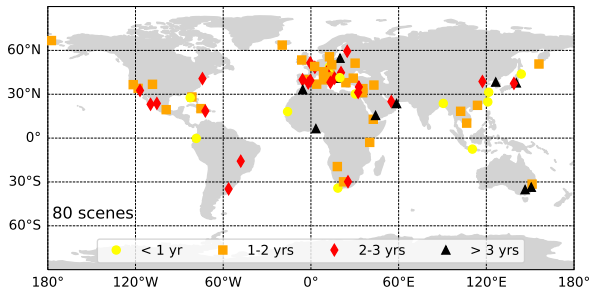


Fig. 2. Spatial distribution of the MUMUCD scenes. Markers indicate the location and the time elapsed between before/after acquisitions is color-coded.

For the results in Table IV and Fig. 3, for simplicity we split each patched image equally among training and test splits. Other possibilities could be considered by prospective users depending on the application, for example only considering urban scenes and assessing transfer learning capabilities by testing on unseen images only. The sensitivity of CD is shown in Table IV. The percentage of the *change* pixels is around XX% overall, indicating how challenging the task can be due to the strong imbalance.

TABLE IV
BENCHMARKING OF PROPOSED METHODS FOR CD. F1 SCORE FOR THE CHANGE CLASS IS GIVEN IN PERCENT, WITH MEAN AND STANDARD DEVIATION CALCULATED OVER THE WHOLE DATASET. THE PROPOSED METHODS HAVE BEEN TESTED UNDER MULTIPLE COMBINATIONS OF THE INPUT DATA: [S2] ONLY SENTINEL-2, [S1+S2] SENTINEL-2 AND -1, [S1+S2+PRS] SENTINEL-1/2 AND PRISMA, [S1+(S2∩PRS)] SENTINEL-1 AND COMMON BANDS BETWEEN SENTINEL-2 AND PRISMA AND [S1+S2+(PRS\S2)] SENTINEL-1, SENTINEL-2 AND COMPLEMENTARY PRISMA BANDS.

| Input | k−means | RF | sUNet |
|---|---|---|---|
| S2 | 0.20 (0.13) | 0.36 (0.18) | 0.45 (0.15) |
| S1+S2 | 0.19 (0.13) | 0.36 (0.18) | 0.46 (0.16) |
| S1+S2+PRS | 0.19 (0.13) | 0.35 (0.18) | 0.45 (0.16) |
| S1+(S2∩PRS) | 0.19 (0.13) | 0.35 (0.18) | 0.45 (0.16) |
| S1+S2+(PRS\S2) | 0.19 (0.13) | 0.35 (0.18) | 0.45 (0.17) |

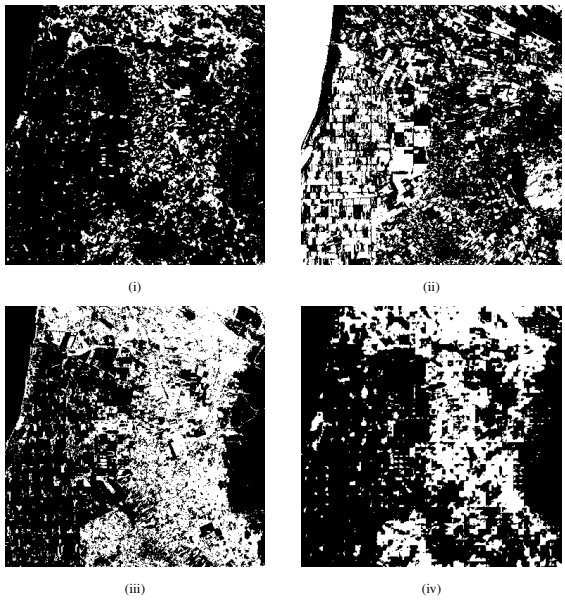


Fig. 3. An example image for the best performing input combination including i) ground truth CD, ii) k-means, iii) RF, iv) sUNet

As shown in Table IV, the unsupervised clustering approach gives the lower score (generally lower than 0.2), while the supervised ML method gives a notable improvement (up to 0.3). The performances of the sUNet model matches those reported by [18], exceeding 0.4 for the OSCD-like case. This is an interesting result since, unlike the OSCD dataset, MUMUCD provides heterogeneous conditions and therefore the model needs more generalization capabilities. The fact that the highest score is obtained with only S2 as input can be explained by the fact that DW is only based on DW, and highlights how tailored labels should be prepared to get out the most of the multimodal collection. Visual inspection of the results (Fig. 3) the $k$-means method has a relatively poor match with the (unseen) GT data, independently of the provided input data. In particular, water bodies are often mistakenly assigned to the *change* class. The RF network, while trained at the pixel level, provides qualitatively better results but still tends to overestimate the *change* class fraction. Interestingly, we found that RF performances are strongly sensitive to the model hyperparameter, with a tendency of producing spurious features for lighter configurations. As anticipated from Table IV, results for the sUNet are spatially smoother and tend to match better the GT information. The first property can be expected with convolutional networks, while the second result indicates that MUMUCD can be successfully applied to other domains beyond urban CD, which was the main purpose of the sUNet.

## IV. CONCLUSIONS

This work outlines the motivation and potential uses of the satellite-based MUMUCD dataset, designed to give easy access to multimodal and multitemporal images for multiple downstream tasks. We demonstrated that despite the varied nature of MUMUCD scenes, ranging from natural to urban environments, the performances of DL models is at par with established benchmarks when using automatically defined CD ground truth data. While ideally annotation should be done manually to support specific applications, the inclusion of multiclass land cover information would enable development of multi-class CD methods [26]. Rich metadata information and modern formats allows users to focus on the design of new architectures requiring multimodal datasets, as extensive preprocessing is performed in advance. The dataset is freely accessible online (https://zenodo.com; *editorial note: an illustrative subset is available at https:owncloud.world*).

The variety of MUMUCD makes it suitable for very different applications beyond land cover CD, as prospective users may complement it with ancillary information, such as meteorological, surface geology, infrastructural or crop type data. Owing to its large size and multiple modalities provided, this dataset can be suitable for training data-intensive foundation models [27] and serve a number of applications for societal benefit.

## REFERENCES

[1] W. A. Malila, "Change Vector Analysis: An Approach for Detecting Forest Changes with Landsat," in *LARS Symposia*, 1980. [Online]. Available: http://docs.lib.purdue.edu/lars_symp/385
[2] D. Peng and coauthors, "Deep learning change detection techniques for optical remote sensing imagery: Status, perspectives and challenges," *International Journal of Applied Earth Observation and Geoinformation*, vol. 136, p. 104282, 2025.
[3] S.-E. Qian, "Hyperspectral Satellites, Evolution, and Development History," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7032–7056, 2021.
[4] S. Liu and coauthors, "A Review of Change Detection in Multitemporal Hyperspectral Images: Current Techniques, Applications, and Challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 140–158, 2019.
[5] S. Arjasakusuma and coauthors, "Change Detection Analysis using Bitemporal PRISMA Hyperspectral Data: Case Study of Magelang and Boyolali Districts, Central Java Province, Indonesia," *Journal of the Indian Society of Remote Sensing*, 2022.
[6] J. F. Amieva and coauthors, "Deep-Learning-based Change Detection with Spaceborne Hyperspectral PRISMA data," in *Proceedings of the 2023 conference on Big Data from Space*. Publications Office of the EU, 2023.
[7] G. Settembre and coauthors, "A land cover change framework analyzing wildfire-affected areas in bitemporal PRISMA hyperspectral images," *Mathematics and Computers in Simulation*, vol. 229, pp. 855–866, 2025.
[8] S. Saha and coauthors, "Change Detection in Hyperdimensional Images Using Untrained Models," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 11 029–11 041, 2021.
[9] R. Torres and coauthors, "GMES Sentinel-1 mission," *Remote Sensing of Environment*, vol. 120, pp. 9–24, 2012.

[10] M. Drusch and coauthors, "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, 2012.

[11] S. Cogliati and coauthors, "The PRISMA imaging spectroscopy mission: overview and first performance analysis," *Remote Sensing of Environment*, vol. 262, p. 112499, 2021.

[12] C. F. Brown and coauthors, "Dynamic World, Near real-time global 10 m land use land cover mapping," *Scientific Data*, vol. 9, no. 1, p. 251, 2022.

[13] G. Brigot and coauthors, "Adaptation and Evaluation of an Optical Flow Method Applied to Coregistration of Forest Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 7, pp. 2923–2939, 2016.

[14] L. Charrier and coauthors, "Analysis of dense coregistration methods applied to optical and SAR time-series for ice flow estimations," in *2020 IEEE Radar Conference (RadarConf20)*, 2020, pp. 1–6.

[15] P. Ebel and coauthors, "SEN12MS-CR-TS: A Remote-Sensing Data Set for Multimodal Multitemporal Cloud Removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[16] G. Sumbul and coauthors, "BigEarthNet-MM: A Large-Scale, Multimodal, Multilabel Benchmark Archive for Remote Sensing Image Classification and Retrieval [Software and Data Sets]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 3, pp. 174–180, 2021.

[17] N. A. A. Braham and coauthors, "SpectralEarth: Training Hyperspectral Foundation Models at Scale," 2024, arXiv:2408.08447.

[18] R. Caye Daudt, B. Le Saux, and A. Boulch, "Fully Convolutional Siamese Networks for Change Detection," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 4063–4067.

[19] S. Hafner and coauthors, "Sentinel-1 and Sentinel-2 Data Fusion for Urban Change Detection Using a Dual Stream U-Net," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[20] M. Schmitt and coauthors, "SEN12MS – A curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2/W7, pp. 153–160, 2019.

[21] G. Sumbul and coauthors, "Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding," in *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 2019, pp. 5901–5904.

[22] M. H. P. Fuchs and B. Demir, "HySpecNet-11k: a Large-Scale Hyperspectral Dataset for Benchmarking Learning-Based Hyperspectral Image Compression Methods," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023, pp. 1779–1782.

[23] J. López-Fandiño and coauthors, "Stacked Autoencoders for Multiclass Change Detection in Hyperspectral Images," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 1906–1909.

[24] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[25] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[26] Q. Zhu and coauthors, "A review of multi-class change detection for satellite remote sensing imagery," *Geo-spatial Information Science*, vol. 27, no. 1, pp. 1–15, 2024.

[27] C. Fibaek and coauthors, "PhilEO Bench: Evaluating Geo-Spatial Foundation Models," in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2024, pp. 2739–2744.