



A new benchmark for spatiotemporal fusion of Sentinel-2 and Sentinel-3 OLCI images

Meryeme Boumahdi^{1,2} · Angel García-Pedrero^{1,2} · Mario Lillo-Saavedra³ · Consuelo Gonzalo-Martin^{1,2}

Received: 17 September 2024 / Accepted: 19 March 2025
© The Author(s) 2025

Abstract

In Earth Observation, the use of multiple sensors has gained considerable attention as a tool to overcome the limitations inherent in individual data. Despite this, the operational lifespans of sensors are finite, creating an ongoing need to explore and integrate new sensors to sustain critical Earth Observation capabilities across diverse applications. Addressing this imperative, there is a need to establish a novel benchmark dataset featuring the integration of new sensors. In response, we present a new benchmark remote sensing dataset, making a significant contribution to the existing literature. This curated dataset leverages Sentinel-2 and Sentinel-3 OLCI, comprising more than 689 image pairs. It spans a wide range of temporal and spatial variations, capturing diverse landscapes, ecosystems, and weather conditions. Importantly, this dataset is publicly accessible, facilitating researchers in the development of more robust data fusion methods. Moreover, we conduct a comprehensive evaluation of widely used spatiotemporal fusion (STF) methods, conducting a detailed quantitative and qualitative comparison as an application of this dataset. The dataset is freely available at <https://doi.org/10.5281/zenodo.14860220>

Keywords Spatiotemporal fusion assessment · Sentinel-2 · Sentinel-3 OLCI · Benchmark dataset · Optical remote sensing images

Introduction

Over the past few decades, there has been an exponential surge in the utilization of remote sensing techniques, resulting in remarkable advancements in both the quality and quantity of remote sensing images. With the availability of remote sensing images with different characteristics, and free of charge, remote sensing is now a common tool for studying the dynamics of Earth systems (Zhu et al. 2018). High temporal and spatial resolution images are highly desired to ensure better monitoring and detection of temporal changes.

For example, dense time-series data are crucial in monitoring crop yields (Bolton and Friedl 2013), natural hazards and disasters (Han et al. 2010; Van Leeuwen et al. 2010), or changes in land use (Yin et al. 2014). However, due to the trade-off between the pixel size and the scanning swath of the sensors, it is almost impossible to find a satellite product with high spatial and temporal resolutions simultaneously (Tan et al. 2018). For instance, satellite sensors such as MODIS provide daily reflectance images with a spatial resolution ranging from 250 m to 1 km, while AVHRR offers a resolution of 1100 m, and Sentinel-3A/B captures images at a resolution of 300 m. Unfortunately, these spatial resolutions are generally considered low for most applications related to spatio-temporal monitoring of land use and natural hazards. On the other hand, sensors like Landsat ETM+/OLI, SPOT-5, and Sentinel-2A/B offer finer resolutions between 10 m and 60 m. However, they have longer revisiting cycles, with Sentinel-2 revisiting every 5 days, Landsat every 16 days, and SPOT-5 every 26 days (Fensholt 2004; Wang et al. 2017). The abundance of satellite images available allows for the integration of data from different sensors, enabling researchers to address this trade-off and generate dense time series data.

Communicated by: Hassan Babaie

✉ Angel García-Pedrero
angelmario.garcia@upm.es

- ¹ Center for Biomedical Technology, Universidad Politécnica de Madrid, Madrid, Spain
- ² Department of Computer Architecture and Technology, School of Computer Engineering, Universidad Politécnica de Madrid, 28660 Madrid, Spain
- ³ Facultad de Ingeniería Agrícola, Universidad de Concepción, Chillán 3812120, Chile

In recent years, significant advances have been made in data fusion within the remote sensing community. However, a thorough examination of the available methods using benchmark datasets is currently lacking. Evaluations and comparisons are made across various data sets that employ specific sensors. One primary challenge lies in achieving a consistent evaluation due to the scarcity of a suitable dataset that encompasses a variety of locations, scenarios, landscapes, and, more importantly, incorporates data from new sensors. Few remote sensing datasets suitable for data fusion could be found in the literature. A dataset was released for public use with a study comparing the performance of the well known spatio temporal fusion (STF) methods STARFM and ESTARFM, using MODIS and Landsat images from two sites with contrasting spatial and temporal variability (Emelyanova et al. 2013) and was used to test the performance of newly developed STF methods (Liu et al. 2019; Zhao et al. 2018; Song et al. 2018; Cheng et al. 2017). The authors of Li et al. (2020) published a public dataset for spatiotemporal models consisting of MODIS and Landsat pairs.

The primary concern of these datasets is its exclusive reliance on MODIS and Landsat data, despite the availability of different satellite images with varying spatial and temporal resolutions and longer useful lifetimes, which introduces a challenge in comprehending the efficacy of fusion models when applied to integrate imagery from various sensors. Furthermore, there are other shortcomings regarding available benchmark datasets, such as the limited number of paired images and the lack of image diversity, which are insufficient to make a fair comparison.

Keeping benchmark datasets updated with new satellite missions is not only essential but a fundamental necessity for ensuring the development and assessment of fusion methods. As satellite technology continues to advance, new sensors introduce higher spectral, spatial, and radiometric capabilities, offering richer datasets that can significantly improve fusion model performance. Failure to adapt to these advancements risks rendering fusion techniques obsolete and limiting their applicability in real-world scenarios. Moreover, the meticulous preparation of benchmark datasets is a critical factor influencing the quality and reliability of fusion results. Ensuring consistency between sensors by minimizing spectral, spatial, and temporal discrepancies allows fusion models to produce more precise and meaningful outputs. Proper dataset curation enhances model generalizability, ensuring that results are not biased toward a specific sensor combination but rather applicable across different satellite missions and environmental conditions. Testing spatiotemporal fusion methods with European sensors is of paramount importance in advancing the field of remote sensing and geospatial analysis to ensure the continuity of the data and products, replacing these satellites that are close to the end of their operational lives. By systematically testing these methods with different

datasets, researchers can assess their accuracy, efficiency, and robustness in capturing temporal and spatial details in images captured by different sensors.

STF methods aim to merge two or multiple data sources with similar spectral ranges to provide time series with high spatial and temporal resolutions. In most cases, STF methods integrate data from one sensor with high spatial but low temporal resolution (HSLT) and data from another sensor with low spatial and high temporal resolution (LSHT) (Tan et al. 2018). Despite the diversity of STF methods proposed in the literature, each method has shown its advantages in terms of prediction accuracy, computing efficiency, and input data requirements. Nevertheless, there are few studies focused on comparing these methods (Han and Zhao 2019; Chen et al. 2015; Liu et al. 2019; Zhou et al. 2021). On the other hand, to our knowledge, there is no comparative study of STF methods using the new European satellites (Sentinel-2 and Sentinel-3 OLCI). In fact, most of the developed methods are designed mainly to fuse MODIS and Landsat images, even though those missions will end in a few years. Therefore, it is paramount to test fusion methods not only in different scenarios, but also with different sensors and satellites.

In this context, we are contributing to the scientific community by publishing a new benchmark remote sensing dataset based on Sentinel-2 and Sentinel-3 images. The proposed dataset has four key characteristics that differentiate it from existing datasets: First, it introduces for the first time in the literature the use of new European sensors in the STF algorithms, offering different temporal resolution as well as higher spatial and spectral resolution. Second, the dataset is large-scale, not only in terms of spatial resolution, but also in terms of the total number of paired images. Third, it provides a high spatial resolution ratio between the coarse and fine images. Fourth, our dataset includes several ecosystems, landscapes, weather conditions, and temporal and spatial variances. Additionally, we present experimental results that compare five of the most commonly used STF methods in the literature, taking into account the availability of their source code. The STF methods used in this work were evaluated both quantitatively and qualitatively using our newly benchmarked dataset.

New benchmark dataset description

Data sources

The Copernicus Program has emerged as one of the leading space data providers worldwide, offering free and unrestricted access to satellite data acquired by the Sentinel constellation. The primary objective of the Sentinel program is to guarantee seamless data continuity by replacing Earth observation missions that have been retired or are

approaching the end of their operational lifespan. This proactive approach aims to prevent disruptions in ongoing studies. Specifically, users can freely download Sentinel-2 1-C level and Sentinel-3 OLCI 1-B images from the Copernicus Open Access Hub website <https://scihub.copernicus.eu/>.

Sentinel-2 1C level

Sentinel-2 is a satellite imaging mission of the European Space Agency (ESA) designed to monitor changes in land surface conditions. The first satellite, Sentinel-2A, was launched on 23 June 2015, and was later joined by its counterpart, Sentinel-2B, in March 2017 to increase the temporal frequency (Hagolle et al. 2015; Drusch et al. 2012). These twin satellites operate simultaneously, revolving in the same orbit but phased at 180°, and now regularly release data. The Sentinel-2 mission systematically acquires data over land and coastal areas within a latitude band spanning from 56° south (Isla Hornos, Cape Horn, South America) to 82.8° north (above Greenland). The primary goal of the mission is to ensure the continuity and improvement of multispectral imagery, which has previously been provided by missions such as Landsat and SPOT. Sentinel-2 data can be used in various applications, such as climate change analysis, emergency management, security, and land monitoring, including assessments of various aspects as vegetation, water cover, soil, and coastal areas. Compared to the Landsat sensor, the Sentinel-2 sensor covers a wider field of view, spanning 290 km, as opposed to 185 km, and has been widely employed for global monitoring in recent decades (Wang et al. 2016; Drusch et al. 2012). Sentinel-2 images offer spectral information across 13 bands spanning visible, near-infrared (NIR), and shortwave infrared (SWIR) wavelengths, with a spatial resolution of 10 m for most bands, six bands at 20 m resolution, and three bands at 60 m resolution (Du et al. 2016).

Sentinel-3 OLCI 1-B level

Sentinel-3 OLCI is another mission of (ESA) within the Copernicus program, operated jointly with EUMETSAT. Its purpose is to provide ocean, land, and atmospheric services. The current mission consists of two satellites: Sentinel-3A, launched on 16 February 2016, and Sentinel-3B, launched on 25 April 2018. Two additional satellites, Sentinel-3C and Sentinel-3D, are planned for future launches (Verhoef and Bach 2012; Cazzaniga et al. 2019). These twin satellites are in near-polar, sun-synchronous orbit, following the same orbital path but with a phasing difference of 140°. Sentinel-3 carried out a dual view conical instrument called Sea and Land Surface Temperature Radiometer (SLSTR), a dual frequency SAR altimeter (SARL), a Microwave Radiometer (MWR), and the Ocean and Land Color Instrument (OLCI), which is a push-broom imaging spectrometer (Prikaziuk et al. 2021).

The main objective of the Sentinel-3 mission envisages the continuity medium-resolution sensors such as MERIS, VGT, SPOT, and ENVISAT. The Sentinel-3 OLCI mission provides sea surface temperature measurements that are of comparable quality to those obtained by the Advanced Along Track Scanning Radiometer (AATSR) on ENVISAT. It also delivers global coverage of ocean color, water quality, ocean surface, and land ice/snow surface temperature products. The Sentinel-3 OLCI sensor captures data in 21 spectral bands ranging from visible to near-infrared (400 to 1020 nm). It has a spatial resolution of 300 m and a temporal resolution of less than 1.4 days (Donlon et al. 2012). Each of the 21 bands can be programmed in terms of position and width (Donlon et al. 2012).

Sites description

We generated a benchmark dataset that includes 15 sites spanning diverse ecosystems and geographic locations, with a total of 689 Sentinel-2 Sentinel-3 pairs. These sites were carefully selected to introduce a variety of challenges for fusion algorithms, ranging from rapid environmental changes to urban expansion and coastal ecosystem dynamics. Figure 1 shows the locations of the dataset sites on the world map. Each site consists of Sentinel-2 and Sentinel-3 image pairs, with Sentinel-3 providing 16 bands at a spatial resolution of 300 m, while Sentinel-2 offers 10 bands at varying spatial resolutions of 10 m, 20 m, and 60 m. The details of the bands are listed in Table 1. By including multiple bands and diverse geographic conditions, this dataset allows for rigorous testing of STF methods. The inclusion of sites with different environmental conditions makes it easier to identify limitations and failure modes in STF techniques. Understanding these constraints is crucial for guiding the development of more adaptive and generalizable fusion methods.

The site selection was based on several key factors. First, the sites are distributed across multiple continents, covering North America, South America, Africa, Europe, Asia, and Australia, ensuring global representation. The dataset also incorporates a wide range of climate zones and ecosystems, including arid deserts, temperate forests, tropical rainforests, coastal zones, and high-altitude regions, allowing STF models to be evaluated under different environmental conditions. More details about the sites are presented in Annex A. Another important consideration was the dynamic nature of land cover changes at each site. Some locations were chosen due to their susceptibility to rapid transformations, such as wildfire-prone landscapes, expanding urban areas, and vulnerable coastal ecosystems. Additionally, cloud-free data availability was a critical factor in site selection, ensuring that Sentinel-2 and Sentinel-3 observations were consistently available for each location. Lastly, each site presents challenges for STF algorithms, such as seasonal vegetation



Fig. 1 The geographic locations of the sites included in the dataset

changes, urban development, deforestation, and atmospheric disturbances like sandstorms. The dataset is designed to serve as a comprehensive benchmark for evaluating STF methods across diverse landscapes. Table 2 summarizes the locations and key characteristics of each site.

For clarity, we provide a detailed description of three representative sites that illustrate different STF challenges: Waterbank in Australia, Neom in Saudi Arabia, and Maspalomas in Spain. Details about other sites are presented in Annex A.

Table 1 Description of Sentinel-2 and Sentinel-3 bands

Sentinel-2			Sentinel-3		
Band name	Wavelength (nm)	Resolution (m)	Band name	Wavelength (nm)	Resolution (m)
Band 2	458-523	10	Oa1	392.5-407.5	300
			Oa2	407.5-417.5	300
			Oa3	437.5-447.5	300
			Oa4	485-495	300
			Oa5	505-515	300
Band 3	543-578	10	Oa6	555-565	300
Band 4	650-680	10	Oa7	615-625	300
			Oa8	660-670	300
			Oa9	670-677.5	300
Band 5	698-713	20	Oa10	677.5-685	300
Band 6	733-748	20	Oa11	703.75-713.75	300
Band 7	773-793	20	Oa12	750-757.5	300
			Oa16	771.25-786.25	300
Band 8	785-900	10	Oa17	855-875	300
Band 8A	855-875	20			
Band 11	1565-1655	20			
Band 12	2100-2280	20	Oa18	880-890	300
			Oa21	1000-1040	300

Waterbank site

The reason for selecting this site is to evaluate the performance of various STF methods in capturing rapid changes caused by natural processes such as forest fire events. Waterbank (17.706° S, 122.7259° E) is a rural location in northern western Australia. This region has a hot semi-arid climate characterized by two distinct seasons. The wet season in Waterbank typically begins around December and ends around June. Throughout this period, the Waterbank receives an average of approximately 580.2 mm of rainfall. On the contrary, during the dry season from July to November, the total rainfall is less than 29.1 mm. In February, which is the peak of the wet season, it rains an average of 11.5 days within that particular month. In fact, Waterbank has even experienced record-breaking rainfall of 476.6 mm on a single day during January. The month of January holds the record for being the wettest month, with 910.8 mm of rain. Regarding the temperatures, during the wet season, the average daytime temperature ranges between 29.2 and 34.30 °C, while the overnight temperature ranges from 15.3 to 26.40 °C. In the dry season, the average daytime temperature varies from 28.8 to 33.5 °C, while the nighttime temperature ranges from 13.6 to 25.0 °C. During this period, vegetation becomes dry and highly susceptible to ignition, resulting in a significant fuel load.

The Northern Territory in Australia records one of the highest frequencies of early-season fires. Many land managers carry out prescribed fires to remove grasses that could

act as fuel for more destructive wildfires later in the dry season. As a result, the area experiences frequent fires throughout the year, leading to rapid changes in vegetation patterns. This site serves as an ideal testing ground to evaluate the accuracy of STF algorithms in a heterogeneous landscape characterized by sudden changes. We collected a total of 38 cloud-free image pairs from Sentinel-2/Sentinel-3 OLCI, spanning from 09 January 2019, to 14 November 2020, and covering a surface of 324 km². Figure 2 displays four pairs of Sentinel-2/Sentinel-3 images, illustrating the pronounced and abrupt changes in the area resulting from fire events, accompanied by gradual changes in vegetation.

Neom city site

The objective of this site is to evaluate the precision of STF algorithms in urban changes within a desert region. Neom City, located in the northwest part of the Kingdom of Saudi Arabia on the Red Sea coast in Tabouk province (28.102° N, 35.207° E), is one of the world's newest smart cities, constructed from scratch as part of Saudi Arabia's Vision 2030 initiative. The climate in this location is relatively cooler than typical temperatures across the Gulf Cooperation Council, due to the presence of mountains and cool winds originating from the Red Sea. January holds the record for being the coldest month, with temperatures ranging from 10 to 21 °C. On the other hand, August is the hottest month, with temperatures ranging between 17 and 40 °C. The area

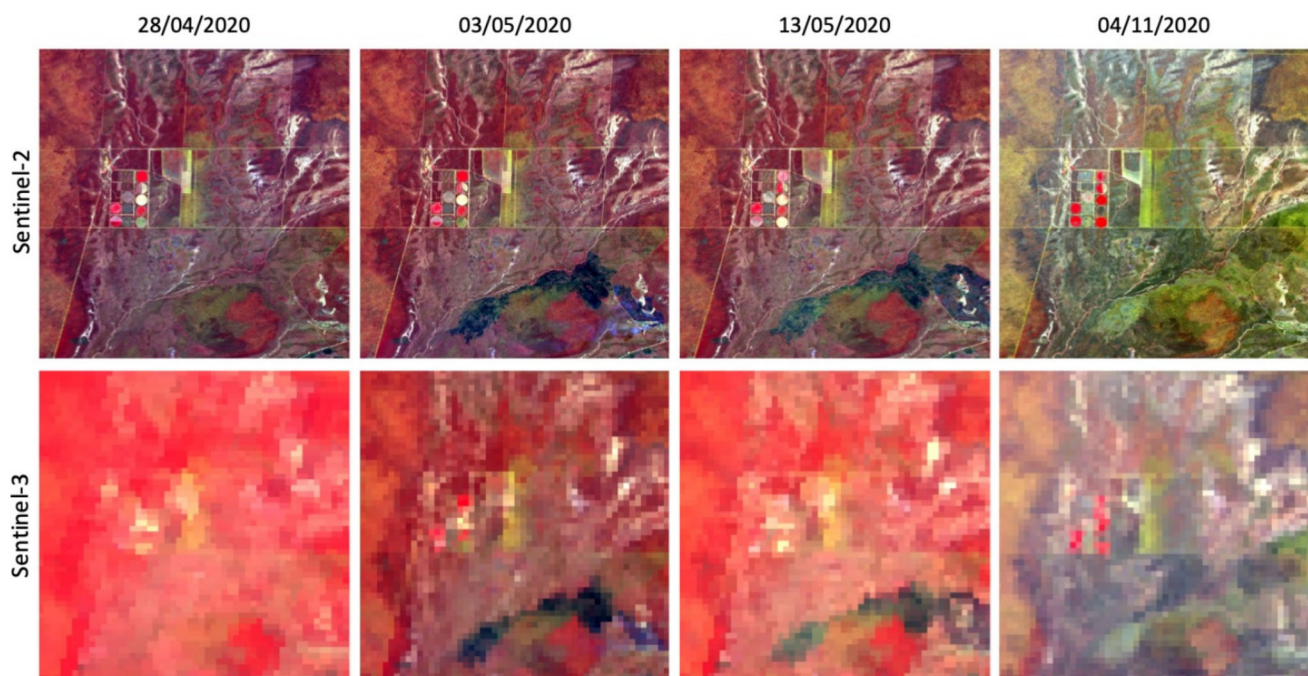


Fig. 2 Example pairs S2/S3 from Waterbank dataset (NIR-Red-Blue color composite)

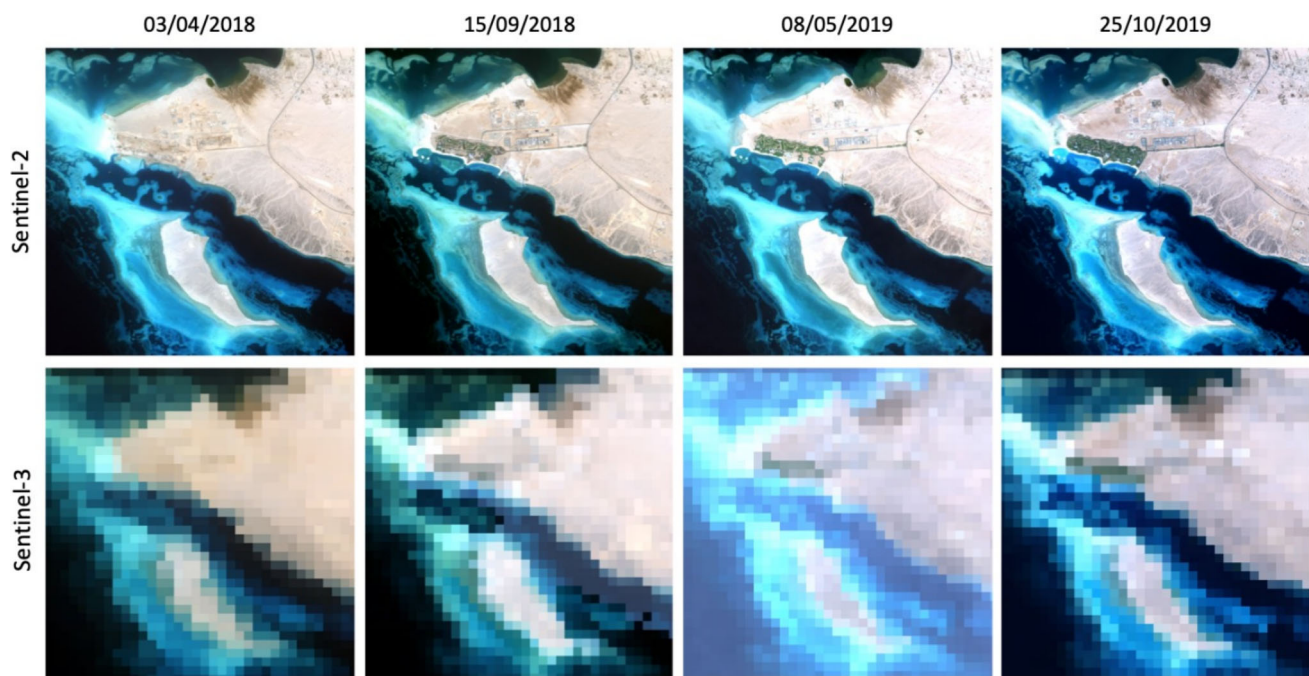


Fig. 3 Example pairs S2/S3 from the Neom City dataset (NIR-Red-Blue color composite)

experiences high sun exposure throughout the year, with an average duration of 9 to 11 hours per day. Additionally, there are only a few cloudy days, resulting in a larger number of cloud-free images available for analysis. The site exhibits a gradual transformation in land cover due to extensive urban development, including the construction of buildings and

infrastructure. In addition, there are phenological changes that occur in both the land and coastal areas. This dataset comprises 65 cloud-free image pairs from Sentinel-2/Sentinel-3 OLCI, captured between 14 March 2018, and 18 December 2020 and covers 144 km². These images document the construction progress of various projects, such as a helipad

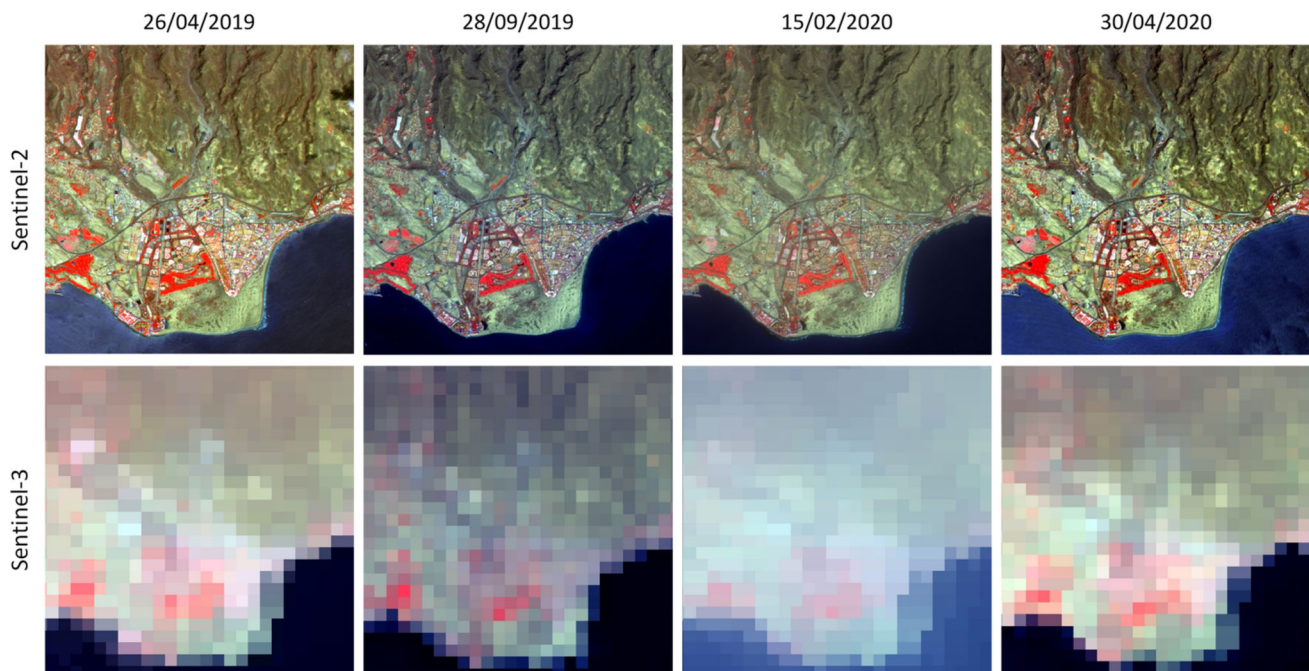


Fig. 4 Example pairs S2/S3 from the Gran Canaria dataset (NIR-Red-Blue color composite)

Table 2 Description of the sites of the dataset

Site	Location	Pairs	Images size	Timespan	Main characteristics
Waterbank	North Australia	38	1800x1800	2019-2020	Rapid change (Forest Fire)
Neom City	Saudi Arabia	65	1200x1200	2018-2021	Desert, spatial & temporal changes in urban area
Maspalomas	Gran Canarias (Spain)	36	900x900	2019-2020	Coast, Vulnerable eco-system and sand storm

and a golf course, among the initial development endeavors. Figure 3 displays four pairs of images from this site, providing visual evidence of the progress made in the construction process within the city of Neom.

Maspalomas site

The primary objective of this site is to evaluate the precision of STF methods in coastal and vulnerable ecosystems. The Canary Islands, known for their exceptional biodiversity (Ibarrola-Ulzurrun et al. 2017), include Maspalomas (27.760° N, 15.587° E), a natural reserve located in the southern part of Gran Canary Island. Maspalomas has a sub-tropical desert climate characterized by dry, hot summers and mild winters. The average annual temperature ranges from approximately 20 to 25 °C. Winter in Maspalomas sees an average rainfall of 35 mm, and the absence of rain extends for 10 months with average highs varying between 21 and 23 °C. Warm summers are marked by average high temperatures ranging from 26 to 29 °C. Maspalomas has several ecosystems, including a coastal dune with significant tourist activities, a terrestrial ecosystem, and an inner-water ecosystem known as the Maspalomas lagoon. The region is also prone to sandstorms that occur at various times throughout the year. This presents a challenging scenario for STF algorithms due to the complex interplay between different ecosystems, vulnerable areas, and human activities. The dataset for this site comprises 36 cloud-free image pairs from Sentinel-2/Sentinel-3 OLCI, covering the period from 06 January 2019, to 11 December 2020 and covers 81 km². Figure 4 presents four pairs of images, highlighting the striking heterogeneity of the landscapes surrounding this site. Table 2 presents the specifications of each site.

Image preprocessing

To create the dataset, the first step in the preprocessing workflow is atmospheric correction, which plays a critical role in ensuring the accuracy and consistency of surface reflectance

values. Atmospheric correction is essential for removing atmospheric effects such as scattering and absorption caused by gases, aerosols, and water vapor, thereby standardizing the data across different acquisition times and sensor platforms. While there are several atmospheric correction algorithms available in the literature, only a few are designed to work with Sentinel-2 1-C and Sentinel-3 OLCI 1-B level data. For this study, atmospheric correction was performed using the Image Correction for Atmospheric Effects (iCOR) algorithm (Renosh et al. 2020). Unlike many existing datasets that either lack atmospheric correction or apply different algorithms to fine and coarse images, leading to inconsistencies, we ensured a uniform correction process across both sensors. This approach minimizes potential biases and enhances the reliability of the dataset for fusion applications.

iCOR is capable of processing satellite images collected on various types of water and land surfaces, and it involves multiple processing steps. First, it identifies water and land pixels by applying a band threshold. The aerosol optical thickness (AOT) is then estimated from the land pixels and extrapolated over the water. Next, an adjacency correction is applied using the SIMilarity Environmental Correction (SIMEC) approach (Sterckx et al. 2015). Finally, iCOR utilizes solar and view angles (Sun zenith angle (SZA), view zenith angle (VZA), and relative azimuth angle (RAA)), along with a Digital Elevation Model (DEM), to perform atmospheric correction. This correction is carried out by referencing MODTRAN5 Look-Up-Tables (LUTs) (Berk et al. 2006). The iCOR algorithm is available as a plugin in the SNAP software.¹ Subsequently, the Sentinel-3 data was reprojected to the UTM/WGS84 projection to match the projection of the Sentinel-2 data. The co-registration process involved two steps: co-registration of the Sentinel-2 and Sentinel-3 time series, and co-registration of the image pairs. While some datasets omit co-registration for simplification, we implemented a robust two-step process to ensure precise spatial alignment, a critical factor in reducing fusion

¹ https://blog.vito.be/remotesensing/icor_available Consulted on February 2025.

errors. In the first step, ENVI 5.6.1 software was used to co-register a single pair of Sentinel-2/Sentinel-3 images. Then, each image was used to co-register the entire time series using AROSICS software (Scheffler et al. 2017). AROSICS employs cross-correlation in the Fourier space for accurate coregistration. This step is particularly crucial, as even within a single sensor time series, geolocation misalignments can occur (Rufin et al. 2020). Without proper time series coregistration, these errors would significantly degrade fusion quality, especially when generating temporally consistent. Finally, the Sentinel-3 OLCI images were resampled using the Nearest Neighbor method to match the spatial resolution of the Sentinel-2 data. Additionally, we ensured that all spectral bands were preserved, unlike some datasets that reduce bands to optimize storage or computational efficiency. Retaining the full spectral information enhances the dataset's versatility for future applications, including spectral-based analyses and machine learning approaches in remote sensing. The preprocessing workflow is illustrated in Fig. 5. The generated dataset is organized by sites and date and is freely available at Boumahdi et al. (2023). Each folder contains the preprocessed coarse/fine image pair. The dataset is provided using a WGS 84 / UTM zone 48N projection, with a tag image file format (TIFF).

Comparison of spatiotemporal fusion methods using sentinel-2 and sentinel-3 OLCI

We performed a benchmark analysis on our dataset, assessing various STF methods that serve as representatives. The objective of this undertaking was to provide a thorough examination of the current state-of-the-art performance, specifically within the framework of utilizing Sentinel imagery. The insights obtained from this benchmarking exercise not only improve our current research comprehension, but also lay the foundation for future endeavors in this field.

STARFM

The Spatial and Temporal Adaptive Reflectance Fusion Method (STARFM) is one of the first STF methods in the literature (Gao et al. 2006). It is widely used and has become the basis for the development of many other methods. STARFM is the first weighted function-based method, and requires one fine resolution image and a coarse resolution image at the base date t_1 , as well as a coarse resolution image at the prediction date t_2 . The algorithm assumes that there is a uniform

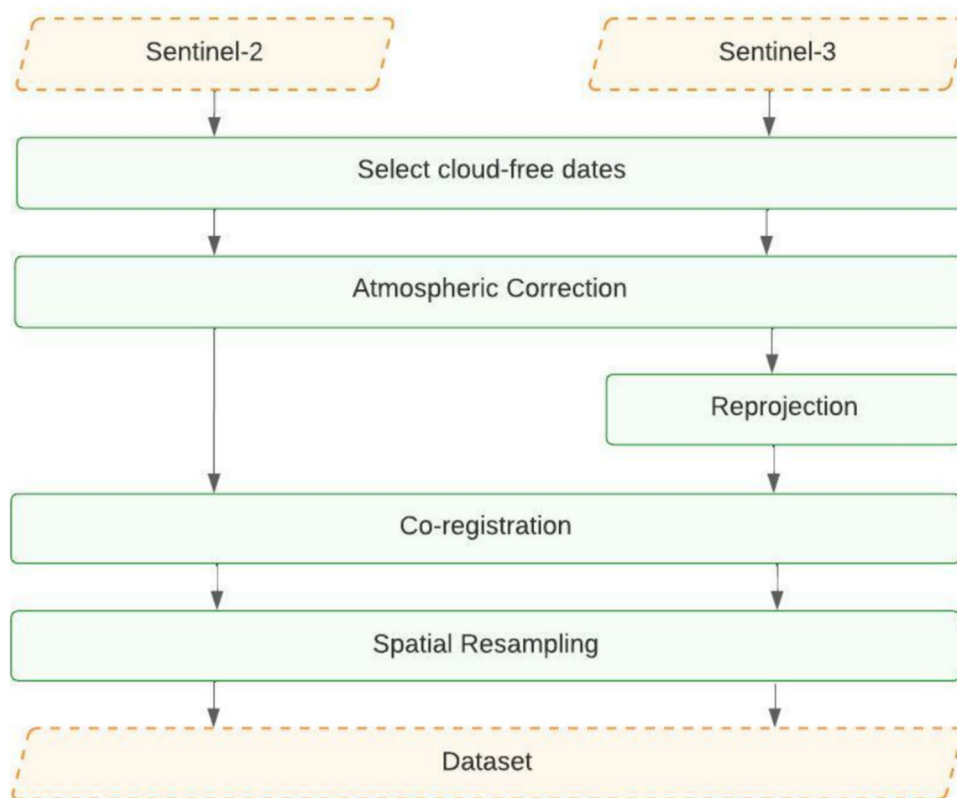


Fig. 5 Flowchart of the dataset preprocessing

temporal change between all classes in the coarse and fine images, biased by a constant error. A moving window is used to identify the spectrally similar neighbour pixels in the fine image at t_1 . When the coarse pixel is homogeneous, the spectral changes between t_1 and t_2 are consistent, so they are added directly to the fine image in t_1 . However, in the case where the coarse pixel is mixed, the weight function is calculated based on neighbouring similar pixels using a moving window. The weight function determines the contribution of each neighboring pixel to the reflectance estimation of the central pixel in the moving window. The spectral difference, the temporal difference, and the spatial Euclidean distance between the neighbor and the central pixel are combined to determine its distance. Finally, the predicted image at t_2 is generated using the previously calculated weight function.

ESTARFM

ESTARFM was proposed mainly to enhance the performance of STARFM in heterogeneous landscapes (Zhu et al. 2010). It requires two pairs of coarse-fine images from prior and posterior dates, as well as a coarse image at the prediction date. The two fine images are used to search for similar pixels within a moving window centered on the target pixel. The spectral similarity between the target pixel and similar pixels is used to determine the weights for all similar pixels. To increase accuracy in the heterogeneous area, the conversion coefficients are calculated using a linear regression analysis for each similar pixel in a search window. Subsequently, the weight function is combined with the conversion coefficients to predict the fused images at date t_2 .

FSDAF

Flexible Spatiotemporal Data Fusion (FSDAF) is a hybrid method presented in Zhu et al. (2016). It requires at least a coarse-fine pair image on the base date t_1 and a coarse image on the prediction date t_2 . FSDAF combines the idea of weighted function-based and unmixing-based methods with spatial interpolation. Firstly, the fine resolution image at the base time t_1 is used to generate a classification map. The coarse images are then used to estimate the temporal changes for each class between t_1 and t_2 . A temporal prediction of the fine resolution image at t_2 is made based on the class-level changes from the previous step, and the residual at each coarse pixel is also calculated. Another prediction, called spatial prediction, is made using a Thin Plate Spline (TPS) interpolator by downscaling the coarse resolution image at t_1 to match the size of the fine resolution image. The residuals from the temporal prediction are distributed using a weighted function based on the TPS prediction and geographical homogeneity. The final step consists of estimat-

ing the final prediction at t_2 by introducing information from the neighborhood.

Fit-FC

Fit-FC is another spatiotemporal method presented in Wang and Atkinson (2018) with the main objective of fusion of Sentinel-2 MSI and Sentinel-3 OLCI images. Fit-FC is a linear regression-based method that requires one pair of coarse-fine images at the base date t_1 and a coarse image at the prediction date t_2 . The first step involves fitting local linear regression models between the two coarse images at t_1 and t_2 within a moving window. The regression coefficients are assigned to the center pixel of the moving window and then applied to the fine image at t_1 . The second step is a spatial filtering (SF) technique that is used to address the blocky artifact problem that may arise from the previous step. Spectrally similar pixels within the moving window are considered to define a weighted function for each center pixel. The final step is Residual Compensation. It involves downscaling the residuals from the first step's coarse image to the size of the fine image using cubic interpolation to avoid the smoothing effect caused by spectrally similar pixels. To preserve spectral information, the updated residual is added back to the predicted image generated in the SF prediction step.

SFSDAF

SFSDAF is an enhanced FSDAF method that incorporates subpixel class fraction change information (Li et al. 2020). SFSDAF requires one pair of coarse-fine images at a prior date. In the first step, the ISODATA algorithm is used to generate a land cover map from the fine image at the base date. The endmembers are calculated as the average of the fine resolution image pixels for each class, and then a soft classification is applied to the fine image at the date t_1 . In the second step, an image of a coarse resolution class fraction is calculated as the difference between the coarse resolution class fractions at t_1 and t_2 . This image is then downscaled from t_1 to t_2 using bicubic and TPS interpolation. The downscaled image is added to the fine resolution class fraction at t_1 to obtain the fine resolution class fraction at t_2 . The final prediction uses a spatial interpolation, similar to that used in FSDAF, to combine the temporal and spatial predictions obtained in the previous steps.

Performance metrics

Several indices were calculated to facilitate a comprehensive comparison of the performance of the STF method, encompassing different aspects of accuracy. All predicted images were compared with the Sentinel-2 image on the prediction

date t_2 . Four indices were used for this comparison. The root mean square error (RMSE) (Zhou et al. 2016) was employed to provide an overall measure of the differences between the predicted and original reflectance. A lower RMSE indicates a higher accuracy. Cross-Correlation (CC) (Loncan et al. 2015), with an ideal value of 1, was used to evaluate the geometric distortion present in the predicted image. The third metric, Structural Similarity (SSIM) (Wang et al. 2004), quantified spatial similarities in the overall structure between the predicted and original images. A higher SSIM value indicates better results. Peak Signal to Noise Ratio (PSNR) was also employed as a quantitative assessment (Hore and Ziou 2010). It evaluated the spatial reconstruction quality and provided an overall measure of bias between the predicted and original images. A higher PSNR value indicates a higher quality spatial reconstruction.

Results and discussion

As described in the previous section, we have selected five of the most commonly used STF methods in the literature, each of which has public available codes. These methods include STARFM (Gao et al. 2006), ESTARFM (Zhu et al. 2010), FSDAF (Zhu et al. 2016), Fit-FC (Wang and Atkinson 2018), and SFSDAF (Li et al. 2020). We compared their performance when fusing Sentinel-2 and Sentinel-3 images from our dataset. The results of this comparison provide a baseline for future studies to be conducted.

For comparison, six different bands of the Sentinel-2 and Sentinel-3 OLCI sensors were utilized. These bands are the blue band (Band-2/Oa4), the green band (Band-3/Oa6), the red band (Band-4/Oa8), the red-edge 1 band (Band-5/Oa11), the red-edge 3 band (Band-7/Oa16), and the near-infrared (NIR) band (Band 8A/Oa17). Table 3 presents the details of the image pairs used in this work, including the dates and the time difference in days between the previous and prediction dates.

Visual evaluations

A coarse/fine (Sentinel-3/Sentinel-2) pair of images from the base date t_1 is used as input for the fusion methods, along

with a coarse image at the prediction date t_2 . The fine image at t_2 is used as the ground truth to compare the prediction results.

Figure 6 displays both predicted images and original Sentinel-2 images (HR0 in the figure) over the Waterbank site on 08 May 2020, the Neom city site on 12 February 2019, and the Maspalomas site on 28 September 2019. The pair of Sentinel-3 OLCI/Sentinel-2 images on the base date (Prior LR0 & HR0) were acquired on 05 May 2020, 03 January 2019 and 29 August 2019 from Waterbank, Neom city, and Maspalomas, respectively. These images were chosen for visual evaluation because of their notable spatial and temporal changes between the base and predicted dates. The Waterbank site exhibits a large and rapid change, manifested as fire and agriculture activities in the form of center-pivot irrigation circles. Neom City features significant urban construction and a change in land cover type, transitioning from desert to green areas or water bodies. The intensity of the color of the buildings in this area has changed from grey to white, indicating ongoing urban development at the location. The Maspalomas site displays changes in vegetation intensity, high heterogeneity, and meteorological conditions described as sandstorms.

For visual examination, we used different color compositions for each site to highlight the features considered in this study, such as fire, burned areas, agriculture fields, and construction. Specifically, we used the NIR-Red-Green color composite for both the Waterbank and Maspalomas sites, while the RGB color composite was used for the Neom city site.

Figure 6 shows that all methods were generally successful in capturing spatial details, despite the high-scale factor (300 to 10).

The Waterbank site presents the greatest challenge for all methods. First, there is high heterogeneity due to the center pivot irrigation. In the Sentinel-3 OLCI images, each of these circles occupies only one or two pixels, while in the Sentinel-2 images, more than 30 pixels occupy the same area. Second, the methods retain the smoke caused by the fire from the prior HR0, even if it disappears on the predicted date. Fit-FC fails to detect the edge of the burned area, and its blurring effect is greater compared to FSDAF. When it comes to the city of Neom, most of the methods are unable to detect spatial

Table 3 Pairs of images from the dataset used in the experiments and the time lapse between the prior and predicted date

Waterbank		Neom City		Maspalomas	
Date	Time lapse	Date	Time lapse	Date	Time lapse
14/05/2019	10	16/08/2018	35	24/08/2019	10
23/07/2019	20	12/07/2018	20	29/08/2019	5
28/04/2020	35	12/02/2019	40	28/09/2019	30
03/05/2020	5	27/07/2019	20	02/11/2019	35
08/05/2020	5	10/10/2019	30	27/11/2019	25

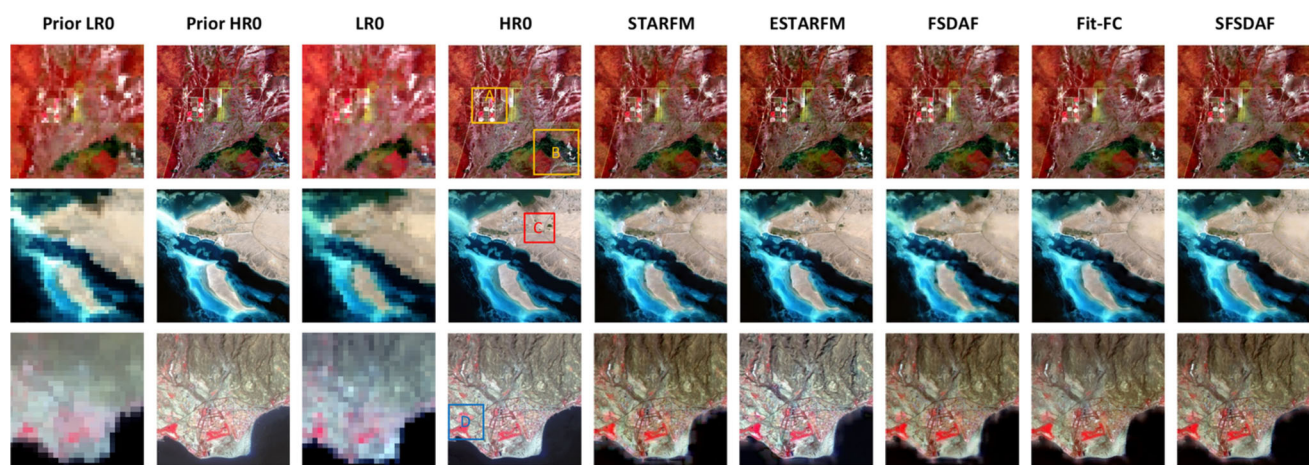


Fig. 6 Sentinel-3 images on base date (Prior LR0) and prediction date (LR0), Sentinel-2 images observed on base date (Prior HR0) and prediction date (HR0), and the Sentinel-2-like images. First row Waterbank site with (NIR, Red, Green) composition, prior HR0 was acquired on 05 May 2020, and HR0 08 May 2020. Second row the Neom city site with

(RGB) composition, prior HR0 was acquired on 03 January 2019 and HR0 was acquired on the 12 February 2019. Third row on Maspalomas site with (NIR, Red, Green) composition, prior HR0 was acquired on 29 August 2019 and HR0 was acquired on 28 September 2019

changes such as green areas and constructions. At the Maspalomas site, the methods face a greater challenge in defining the coast, especially near the Maspalomas Dunes, which are located at the bottom of the images. As a result, the methods fail to detect spatial changes within a single coarse pixel, and the predicted images have spatial details that are more similar to the base image at t_1 than the actual Sentinel-2 image at t_2 .

The general visual evaluation of the five methods in Fig. 6 did not show a significant difference in their performance. Figure 7 presents a zoomed-in area within the boxes shown in Fig. 6. At the Waterbank site, there have been changes in the details and color of the center pivot irrigations from the base to the prediction date. In the false color composite images, the color has changed from green to white, while the size and boundaries of the irrigations remained the same. This site exhibits high heterogeneity, with each circle occupying one to two pixels in the coarse image and 30 to 60 pixels in the fine image. None of the methods was able to detect the temporal changes accurately; instead, they preserved the details and intensity from the base date and some methods maintained the spatial boundaries. However, certain methods exhibited worse results, characterized by a blurring effect around the edges. For example, FSDAF and Fit-FC showed this blurring effect and FSDAF failed to define all the boundaries of the centre pivot irrigations. Regarding the burned area at the same location, the fire expanded from the base to the prediction date, increasing the surface area of the burned region, visible as a green color in the false color composite. The smoke from the fire on the base date disappeared in the prediction date. Generally, the methods generated tones similar to those of the actual image on the prediction dates, but they all retained the

smoke in the predicted image. STARFM and SFSDAF methods were able to predict the new burned areas with imprecise edges, while the ESTARFM performed better in this regard. Fit-FC and FSDAF exhibited lower performance due to their blurring effects, making them less suitable for capturing sudden changes.

Zoomed-in view of the Neom city site revealed a change in land cover between the base date (prior HR0) and the prediction date (HR0), with the emergence of green areas and water basins. Among the methods, ESTARFM demonstrated better performance by detecting most of these changes, particularly the presence of water basins, which no other method predicted. On the contrary, Fit-FC and FSDAF exhibited the worst performance. Neither of these methods was able to detect spatial details or temporal changes, and the blurring effect affected all heterogeneous pixels. However, Fit-FC generated an image that closely resembled the base date image rather than the actual date image.

The zoomed-in view of the Maspalomas site reveals a change in image color due to a sandstorm that occurred in the base date. Additionally, many areas transitioned from light to intense red in the false color composition. STARFM was able to detect some of these changes, although the spectral characteristics of the generated image differed from the actual image. ESTARFM also captured some of the changes and produced images with spectral characteristics closer to the actual image. However, similar to the previous site, FSDAF generated blurred images and lacked clear boundaries between the land and water areas. Many urban areas near the coast were not noticeable due to the blurring effect. Fit-FC, like FSDAF, struggled to capture small details in the images. It is important to note that none of these methods

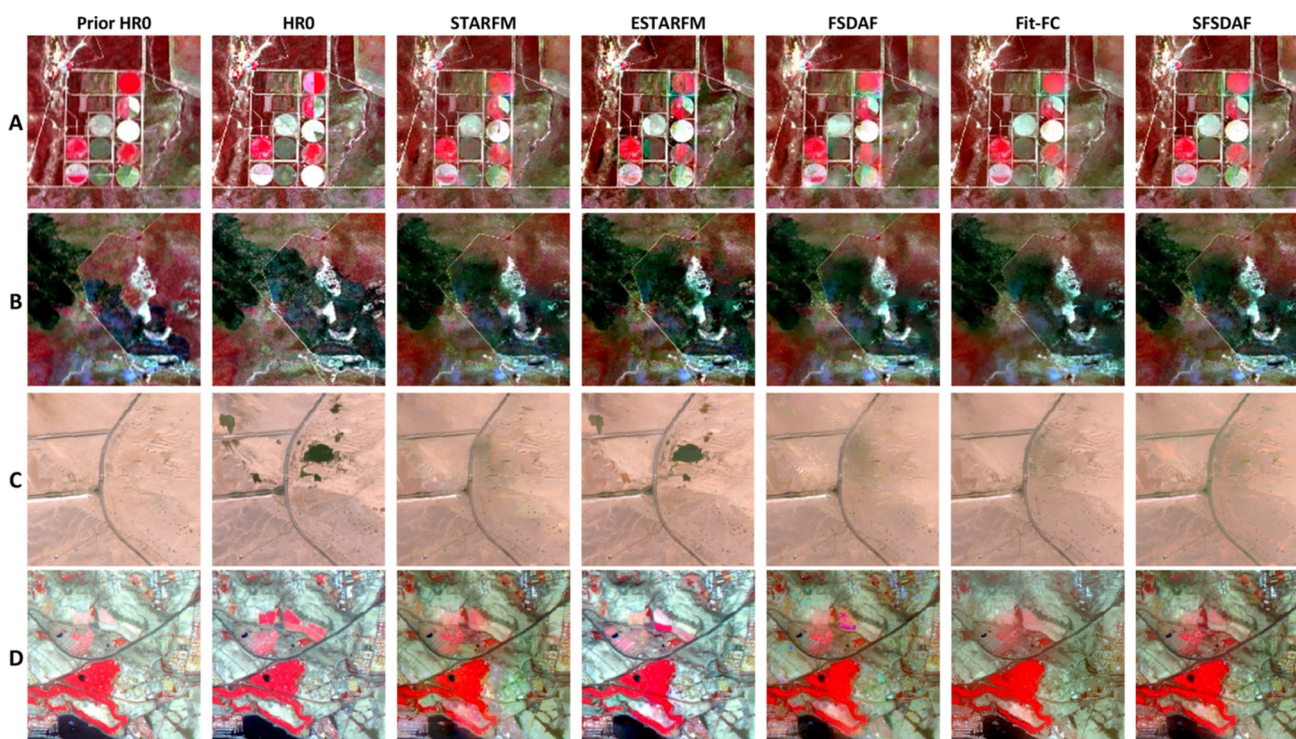


Fig. 7 Zoomed-in area (Rectangles area in Prior HR0 in Fig. 6) of Sentinel-2 images at prior and prediction date and the images generated by the methods used in the comparison

successfully predicted the small spatial details or temporal changes seen in the actual image. This highlights the challenges faced by these STF methods in predicting changes in small objects within highly heterogeneous sites, as well as sudden and continuous temporal changes in other cases.

Quantitative evaluations

To better assess the accuracy of the predictions, the metrics CC, RMSE, SSIM, and PSNR were calculated for the three sites, and the results are included in (Tables 4, 5 and 6). The best results are shown in bold. Each table presents in detail the quantitative assessments for every band used in the comparison of the aforementioned methods. All the methods were able to fuse different bands within the bandwidth of both Sentinel-2 and Sentinel-3 OLCI sensors, with a higher spatial resolution ratio between coarse and fine images. ESTARFM obtained competitive results and the metrics were closer to the optimal values than the other methods in all six bands. For example, it significantly improved the SSIM for the Neom City site from 0.4 to 0.88 for band 6. Additionally, ESTARFM consistently achieved higher PSNR values compared to other methods at all sites and bands. SFSDAF also showed good results at all three sites, with metrics that were comparable to those of ESTARFM. The qualitative metrics presented in Tables 4, 5 and 6 align with the visual evaluation discussed

Table 4 Quantitative assessment of experimental results on Waterbank site on 08 June 2019

	Band	STARFM	ESTARFM	FSDAF	Fit-FC	SFSDAF
CC	Band 1	0.9166	0.9517	0.8963	0.9363	0.9282
	Band 2	0.9231	0.9583	0.9090	0.9422	0.9393
	Band 3	0.9112	0.9540	0.8982	0.9229	0.9358
	Band 4	0.9091	0.9544	0.9127	0.9021	0.9355
	Band 5	0.9223	0.9590	0.9213	0.9053	0.9420
	Band 6	0.9228	0.9593	0.9221	0.9055	0.9434
RMSE	Band 1	0.0063	0.0047	0.0080	0.0074	0.0072
	Band 2	0.0095	0.0061	0.0115	0.0103	0.0104
	Band 3	0.0142	0.0096	0.0169	0.0152	0.0148
	Band 4	0.0185	0.0095	0.0192	0.0193	0.0175
	Band 5	0.0236	0.0124	0.0247	0.0251	0.0230
	Band 6	0.0268	0.0144	0.0277	0.0283	0.0260
SSIM	Band 1	0.9928	0.9958	0.9885	0.9906	0.9909
	Band 2	0.9908	0.9953	0.9861	0.9903	0.9899
	Band 3	0.9843	0.9909	0.9767	0.9831	0.9856
	Band 4	0.9842	0.9936	0.9819	0.9772	0.9866
	Band 5	0.9850	0.9933	0.9819	0.9749	0.9867
	Band 6	0.9843	0.9928	0.9811	0.9733	0.9863
PSNR	Band 1	44.0234	46.5277	41.9740	42.5805	42.8116
	Band 2	40.4547	44.3251	38.7530	39.7263	39.6621
	Band 3	36.9341	40.3169	35.4204	36.3649	36.5769
	Band 4	34.6772	40.4041	34.3225	34.2804	35.1436
	Band 5	32.5419	38.1264	32.1480	31.9969	32.7829
	Band 6	31.4491	36.8079	31.1442	30.9528	31.7143

Table 5 Quantative assessment of experimental results on Neom City site on 12 July 2018

	Band	STARFM	ESTARFM	FSDAF	Fit-FC	SFSDAF
CC	Band 1	0.9553	0.9727	0.9291	0.9418	0.9371
	Band 2	0.9662	0.9795	0.9471	0.9581	0.9599
	Band 3	0.9777	0.9864	0.9645	0.9741	0.9774
	Band 4	0.9743	0.9836	0.9627	0.9693	0.9745
	Band 5	0.9738	0.9807	0.9637	0.9668	0.9742
	Band 6	0.9745	0.9806	0.9646	0.9669	0.9753
RMSE	Band 1	0.0513	0.0245	0.0458	0.1010	0.0383
	Band 2	0.0563	0.0343	0.0530	0.1117	0.0464
	Band 3	0.0677	0.0447	0.0644	0.1248	0.0557
	Band 4	0.0708	0.0482	0.0673	0.1252	0.0590
	Band 5	0.0774	0.0550	0.0739	0.1295	0.0677
	Band 6	0.0777	0.0570	0.0759	0.1279	0.0684
SSIM	Band 1	0.5949	0.8705	0.6520	0.4053	0.7738
	Band 2	0.6497	0.9253	0.7283	0.4654	0.8322
	Band 3	0.5675	0.9194	0.6657	0.4326	0.8632
	Band 4	0.5598	0.9104	0.6716	0.4415	0.8660
	Band 5	0.5011	0.8913	0.6248	0.4086	0.8653
	Band 6	0.4818	0.8896	0.5568	0.4106	0.8748
PSNR	Band 1	31.8266	38.2468	32.7992	25.9366	34.3595
	Band 2	31.0133	35.3163	31.5344	25.0634	32.6932
	Band 3	29.4063	33.0109	29.8369	24.0957	31.1055
	Band 4	29.0222	32.3669	29.4567	24.0688	30.6051
	Band 5	28.2414	31.2142	28.6506	23.7737	29.4026
	Band 6	28.2069	30.9088	28.4174	23.8862	29.3235

Table 6 Quantative assessment of experimental results on Maspalomas site on 28 September 2019

	Band	STARFM	ESTARFM	FSDAF	Fit-FC	SFSDAF
CC	Band 1	0.8349	0.9466	0.8454	0.8214	0.8441
	Band 2	0.8797	0.9630	0.8946	0.8982	0.8963
	Band 3	0.9026	0.9654	0.9116	0.9175	0.9212
	Band 4	0.9197	0.9682	0.9302	0.9205	0.9385
	Band 5	0.9318	0.9691	0.9344	0.9245	0.9437
	Band 6	0.9331	0.9680	0.9331	0.9236	0.9435
RMSE	Band 1	0.0470	0.0121	0.0496	0.0531	0.0480
	Band 2	0.0459	0.0150	0.0492	0.0483	0.0487
	Band 3	0.0429	0.0197	0.0473	0.0462	0.0471
	Band 4	0.0415	0.0195	0.0454	0.0457	0.0455
	Band 5	0.0437	0.0232	0.0484	0.0486	0.0489
	Band 6	0.0427	0.0247	0.0475	0.0478	0.0478
SSIM	Band 1	0.5671	0.9716	0.5492	0.4944	0.6153
	Band 2	0.6853	0.9536	0.6876	0.6518	0.7365
	Band 3	0.7489	0.9262	0.7540	0.7088	0.7888
	Band 4	0.7812	0.9202	0.7894	0.7336	0.8123
	Band 5	0.8157	0.9086	0.8041	0.7708	0.8115
	Band 6	0.8377	0.9024	0.8265	0.7987	0.8313
PSNR	Band 1	26.5560	38.3781	26.0864	25.4952	26.3735
	Band 2	26.7549	36.4828	26.1626	26.3181	26.2419
	Band 3	27.3449	34.1307	26.5107	26.7103	26.5411
	Band 4	27.6321	34.2133	26.8500	26.8052	26.8394
	Band 5	27.1844	32.6897	26.2987	26.2699	26.2134
	Band 6	27.3976	32.1321	26.4708	26.4081	26.4121

in the previous section. On the contrary, FSDAF generally exhibited the highest RMSE among the other methods and the lowest CC.

The variation in performance, both in visual and quantitative assessments, can be attributed to the number of pairs used in the fusion process. ESTARFM, for instance, requires two pairs of data, both from the prior and posterior dates. This additional information, particularly regarding temporal changes, enables for more accurate predictions. However, it is important to note that even with the use of multiple pairs, all methods experienced challenges in certain pixels, resulting in inaccurate predictions in some cases.

Different pairs were utilized in this comparison, with five dates tested for each site. Table 3 provides the specific dates for each pair. The selection of these dates aimed to ensure a diverse range of temporal variations in the input data, with the time difference between the base and predicted dates ranging from 5 to 40 days. Figure 8 shows the quantitative evaluations of the results of the experiments mentioned above. We should note that CC, RMSE, SSIM and PNSR in the Fig. 8 corre-

spond to the average of the 6 bands mentioned previously in Table 1.

The results indicate that the time gap between the base and prediction dates does not have as significant an impact as the image heterogeneity and temporal changes between the two images. It is notable that the results for the Neom City site are consistently similar, which can be attributed to the continuous changes occurring in the site over time. ESTARFM demonstrated promising results compared to other algorithms. For example, Fit-FC exhibited a high RMSE for the generated image, indicating lower consistency between the prior and predicted images. In all experiments, ESTARFM outperformed the other methods in terms of PSNR, indicating an overall bias between the actual and predicted images. STARFM showcased a stronger correlation between the base date image and the predicted images in most of the experiments, with correlations exceeding 0.9. On the other hand, Fit-FC exhibited lower correlations, such as 0.65 in the first pair of the Maspalomas site. RMSE values were generally low in the majority cases, with some outliers observed, par-

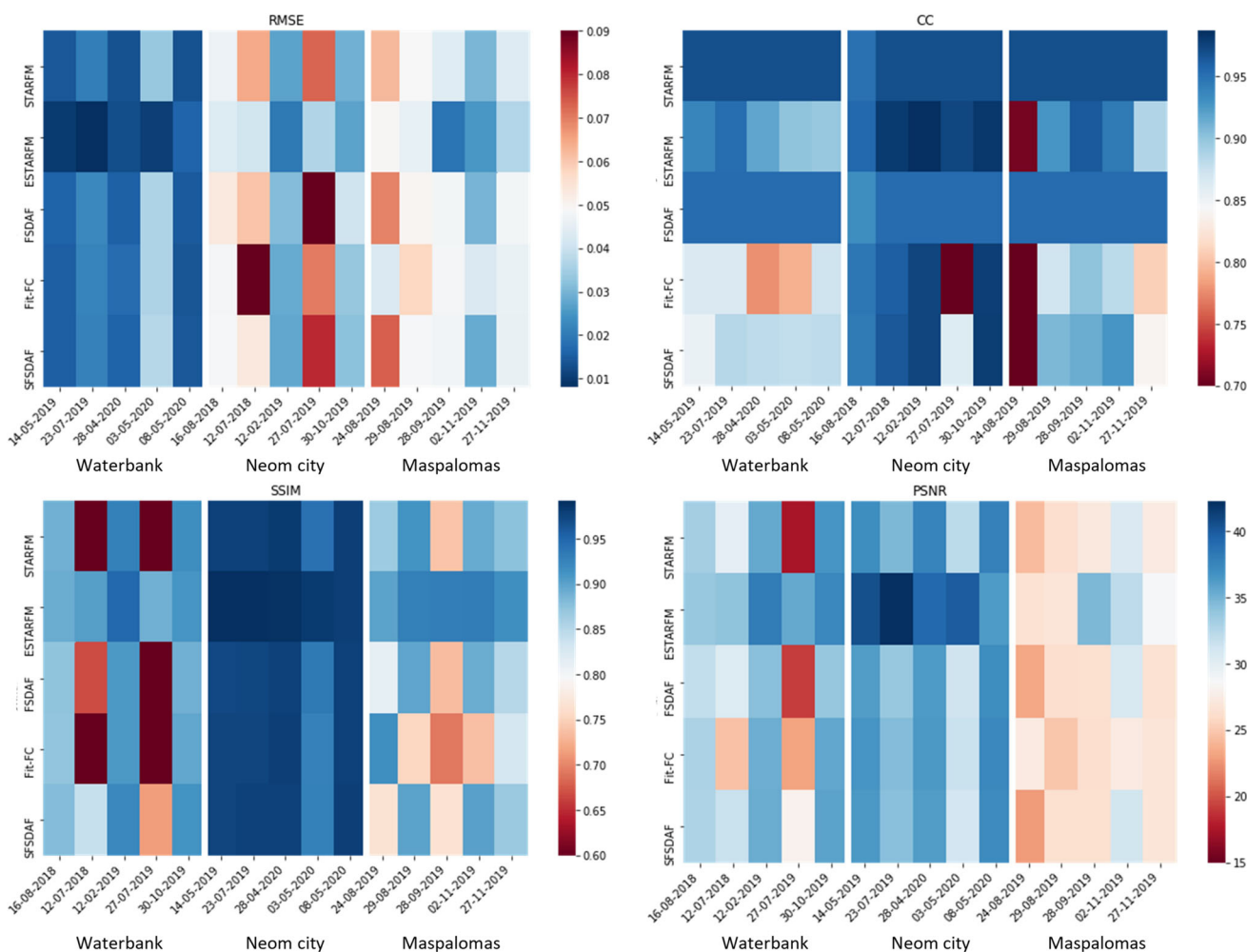


Fig. 8 Average RMSE, CC, SSIM and PSNR for the six bands of each method for the three sites of the dataset between the actual and the predicted image

ticularly in the second and fourth pair of the Neom City site, which experienced greater temporal changes during that period (0.11 for Fit-FC and 0.1 for FSDAF).

Moreover, the results show that there were no significant differences between the five STF methods in landscapes with low heterogeneity and low temporal variance. However, their performance varied in landscapes with higher heterogeneity and greater temporal variance, highlighting their sensitivity to such conditions. Through visual assessment and quantitative metrics, ESTARFM was found to be more suitable for heterogeneous landscapes and sudden events in terms of accurately capturing the structural details between the generated image and the ground truth image, even though it may not always have the lowest errors. While the availability of coarse/fine image pairs for fusion is not always guaranteed, SFSDAF or STARFM demonstrated competitive results and were able to work with a minimum number of image pairs.

Discussion

STF techniques have been widely explored in the literature, with most benchmark datasets focusing on the fusion of MODIS and Landsat due to their complementary temporal and spatial resolutions. Studies such as Zhu et al. (2010) and subsequent research have extensively evaluated the performance of STF methods using these datasets, where the typical spatial resolution ratio between the coarse and fine resolution images is 8:1. However, this study introduces a benchmark dataset with a significantly different configuration, where the Sentinel-3 to Sentinel-2 resolution ratio is approximately 30:1. This larger ratio presents both a challenge and an opportunity for the development of more advanced STF methods. Previous STF evaluations have primarily relied on MODIS-Landsat datasets, where MODIS provides high temporal frequency but at a coarser 250 m

to 1 km resolution, and Landsat offers finer 30 m resolution but with longer revisit intervals of 16 days. While these datasets have been instrumental in advancing STF methods, the Sentinel-2/Sentinel-3 fusion scenario presents a fundamentally different problem. The greater disparity in spatial resolution between Sentinel-3 (300 m) and Sentinel-2 (10–60 m) introduces new interpolation challenges, requiring more sophisticated spatial downscaling strategies to preserve spectral and structural integrity.

The 30:1 resolution disparity between Sentinel-3 and Sentinel-2 presents a greater spatial extrapolation gap than MODIS-Landsat fusion. This larger scale difference increases uncertainty in reconstructed images, particularly in areas with high spatial heterogeneity, such as urban environments, agricultural fields, and regions undergoing rapid land cover changes. While ESTARFM and FSDAF demonstrated robust performance in our dataset, the results indicate that current STF algorithms may not be fully optimized for handling such extreme resolution differences. The larger Sentinel-2/Sentinel-3 gap necessitates the development of enhanced downscaling strategies, possibly incorporating machine learning or physics-informed models to improve spatial coherence. Conversely, this challenge also represents an opportunity to advance STF research by pushing the limits of existing methods. Given the need to reconstruct high-resolution observations from extremely coarse inputs, the Sentinel-2/Sentinel-3 dataset could drive new methodological developments.

Beyond technical challenges, there is a growing need for new benchmark datasets tailored to modern Earth observation missions. MODIS and Landsat, the most widely used sensors in STF research, are expected to be retired in the coming years, necessitating the transition to alternative sensor constellations. The Sentinel series, with its open-access policy and systematic acquisitions, provides an ideal successor for developing next-generation STF methods. However, Sentinel-2 and Sentinel-3 were not originally designed with STF applications in mind, making it essential to establish a robust benchmark dataset to support research in this area. This dataset provides a new standard for evaluating STF methods by incorporating a diverse range of geographic locations, including snow, forests, agricultural lands, deserts, mountainous regions, and coastal environments. The heterogeneity of the selected sites ensures that STF models can be tested under varied environmental conditions, facilitating a broader assessment of their performance.

Previous studies have highlighted that the performance of STF methods is highly dependent on the spatial and temporal variances of the study area. Li et al. (2020); Emelyanova et al. (2013) with their dataset and comparison demonstrated that ESTARFM generally outperformed STARFM in heterogeneous landscapes but was not always the best performer. In

regions where spatial variance was dominant over temporal variance, STARFM exhibited lower errors. Similarly, studies have shown that the choice of fusion algorithm should consider the underlying spatial and temporal characteristics of the dataset rather than assuming a universally optimal method (Zhou et al. 2021). Moving forward, STF research should focus on developing adaptive models capable of handling extreme resolution disparities, integrating multi-source data such as SAR and topographic information, and improving computational efficiency. By establishing a new benchmark dataset, this work paves the way for next-generation fusion techniques that will support high-resolution, temporally continuous Earth observation applications.

Conclusions

In this paper, we have generated a new benchmark dataset, Sentinel-2/Sentinel-3, with the aim of contributing to the scientific community. For the first time in the literature, we present a dataset consisting of pairs of new European satellites, Sentinel-2/Sentinel-3 OLCI, exhibiting various temporal, spatial, and spectral characteristics. Furthermore, this dataset encompasses distinct sites from different regions of the world with diverse spatial features. Moreover, the dataset offers a wide range of spectral bands, thereby enriching the available data options for various applications. One of the applications of this dataset is to evaluate spatiotemporal fusion methods. The dataset's diverse sites offer a good challenge for testing the STF techniques across various landcover types and assess their ability to handle spatial and temporal changes. In addition, we conducted a comparative analysis of the STF methods using European sensors. Specifically, we demonstrated the performance of five of the most widely used STF methods in the literature. The experiments conducted in this study reveal that all STF methods suffer from adequately capturing spatial details in highly heterogeneous landscapes, as well as capturing sudden changes like forest fires. The results obtained from these experiments serve as valuable baselines for future studies and applications involving this dataset.

Appendix A: Supplement description

A.1 Relevant location

This annex provides additional details about specific locations included in the dataset, highlighting their geographical characteristics, environmental conditions, and relevance for STF.

A.1.1 Andorra

Andorra is a small, mountainous country located in the eastern Pyrenees, between France and Spain. The region is characterized by rugged terrain, high-altitude valleys, and a temperate mountain climate. It features dense forests, alpine meadows, and glacial landforms, making it a highly diverse landscape. The area experiences significant seasonal changes, with cold, snowy winters and mild summers, making it an ideal site for evaluating STF methods in mountainous environments. Snow cover dynamics, vegetation changes, and variations in cloud cover present challenges for spatiotemporal fusion techniques. The dataset captures seasonal transitions, enabling the assessment of STF performance in areas with strong topographic influence, where shadow effects, abrupt elevation changes, and weather variability must be accounted for. Figure 9 presents examples of the images in this location.

A.1.2 Fùcino Plateau, Italy

The Fùcino Plateau is a large agricultural region in central Italy, located in the Abruzzo region. Historically, it was a lake that was drained in the late 19th century to create fertile farmland. Today, the plateau is an important agricultural hub, with vast fields dedicated to the cultivation of potatoes, carrots, and other crops. The area experiences different seasonal vegetation cycles, with planting, growth, and harvest periods creating dynamic changes in land cover. This makes it an ideal site for assessing the performance of STF in agricultural landscapes, where accurate time-series reconstruction is critical to monitoring crop development, detecting anomalies,

and improving land management practices. Figure 10 presents examples of the images in this location.

A.1.3 Beni Mellal, Morocco

Beni Mellal is a city located in central Morocco in the foothills of the Atlas Mountains. The region has a Mediterranean climate with semi-arid characteristics, featuring hot, dry summers and mild, wetter winters. The area is known for its agriculture, particularly the cultivation of olive trees, citrus fruits and cereals, supported by irrigation from nearby reservoirs. Seasonal variations in vegetation cover and water availability make Beni Mellal a valuable site for testing STF methods in semi-arid agricultural regions. The dataset captures vegetation dynamics, urban expansion, and land use changes, providing insights into STF performance in areas with complex interactions between natural and human-modified landscapes. Figure 11 presents examples of the images in this location.

A.1.4 Akhaltsikhe Region, Georgia

Akhaltsikhe is a town in southern Georgia, located near the border with Turkey. The region is characterized by a continental climate, with cold winters, warm summers, and moderate precipitation throughout the year. The landscape includes mountainous terrain, river valleys, and patches of forested areas, creating a complex environment for remote sensing applications. The area is known for its agricultural activities, historical sites, and strategic location within the Lesser Caucasus range. The dataset captures seasonal changes in land cover, agricultural cycles, and the impact of climate variability, making it an important site to evaluate

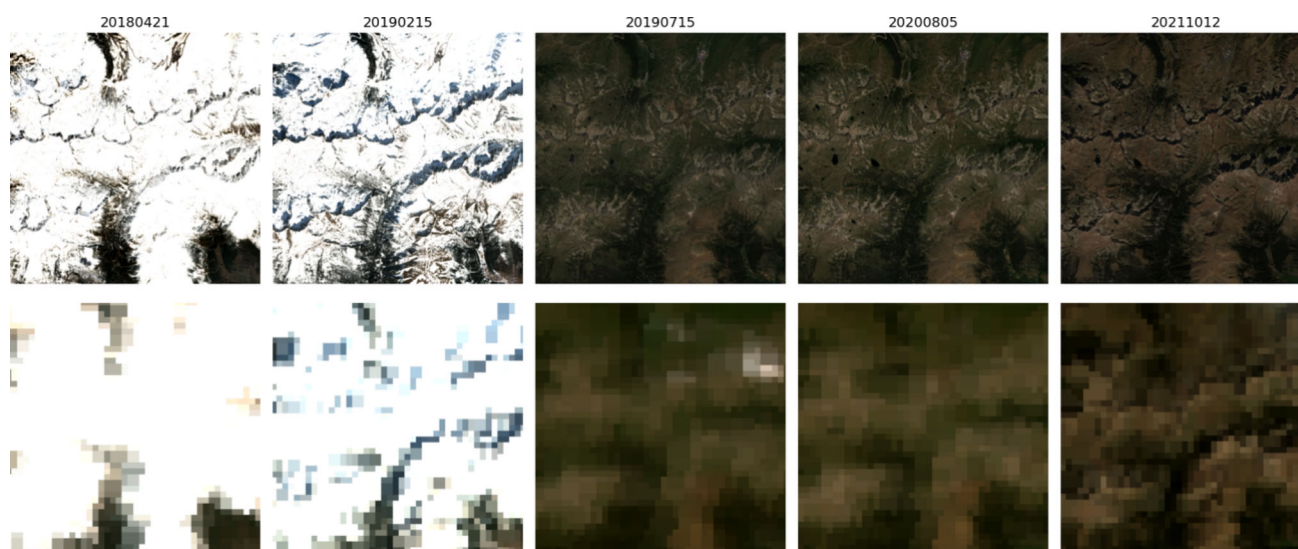


Fig. 9 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Andorra location

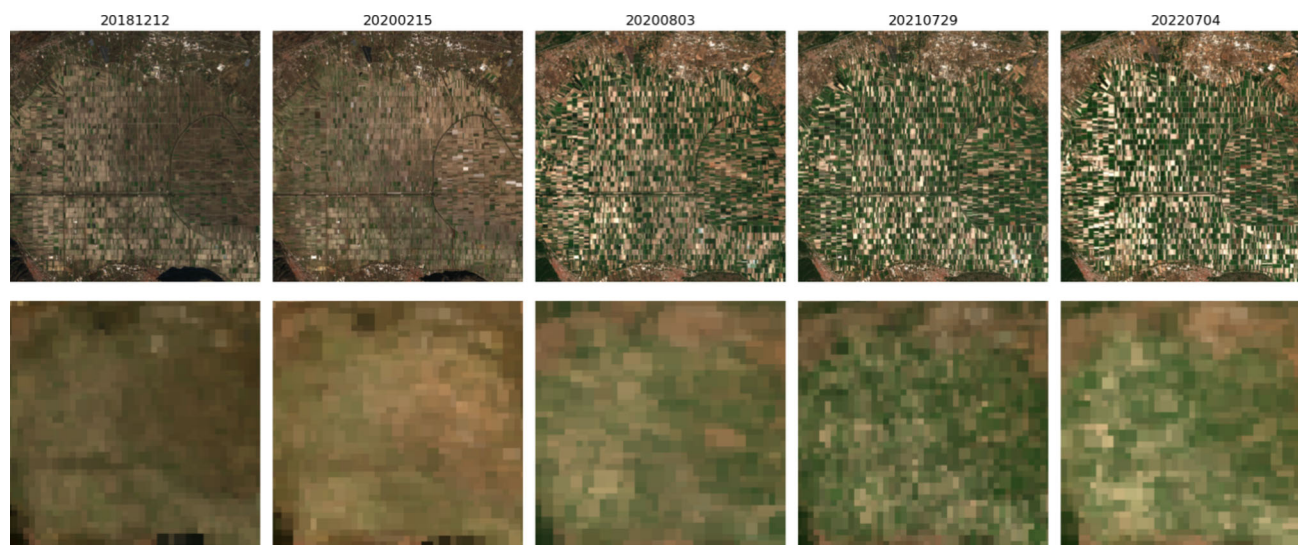


Fig. 10 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Fucino Plateau location

STF methods in a mixed mountain and agricultural landscape. Figure 12 presents examples of images in this location.

A.1.5 Range national monument, Nevada

Basin and Range National Monument is a vast protected area in southeastern Nevada, characterized by a rugged desert landscape, isolated mountain ranges, and arid valleys. The region exemplifies the Basin and Range physiographic province, a geological formation known for its alternating high mountain ridges and low desert basins. The climate is arid to semi-arid, with hot summers, cold winters, and low annual precipitation. The vegetation in the monument consists mainly of sagebrush, desert shrubs, and scattered

juniper woodlands, with minimal water availability shaping the ecological dynamics. This site is particularly relevant for evaluating STF methods in arid environments with sparse vegetation, extreme seasonal variability, and complex topography. The dataset captures long-term land cover stability and subtle seasonal variations, providing valuable insights for STF applications in desert and semi-arid ecosystems. Figure 13 presents examples of the images in this location.

A.1.6 Parc National de la Lopé, Gabon

Parc National de la Lopé, located in central Gabon, is a UNESCO World Heritage site known for its unique blend of dense tropical rainforest and savannah ecosystems. The

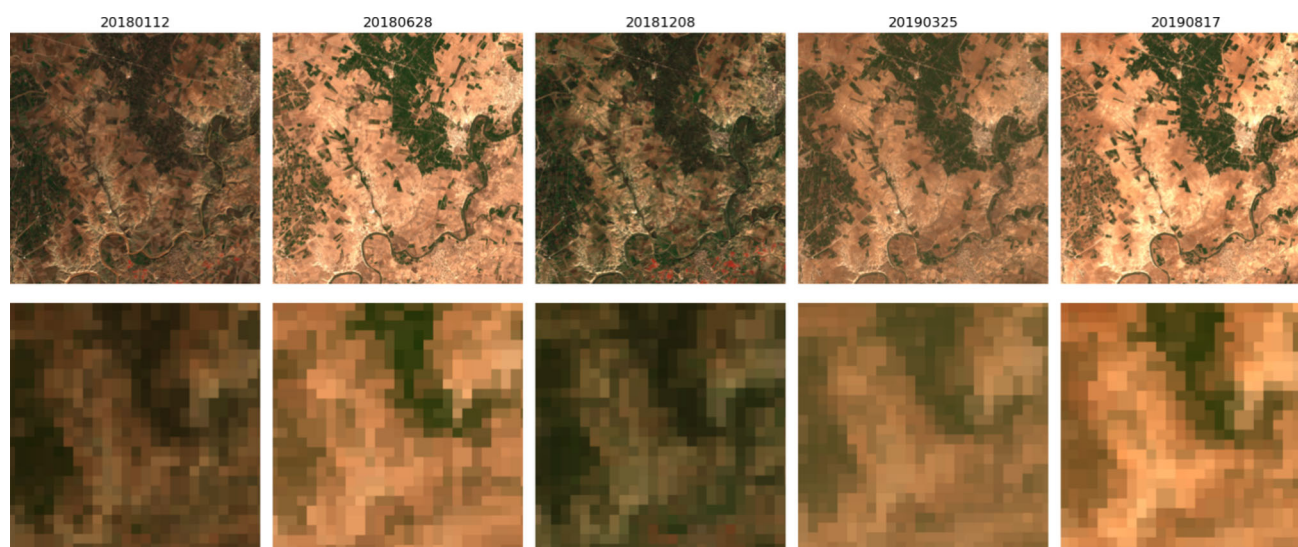


Fig. 11 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Beni Mellal location

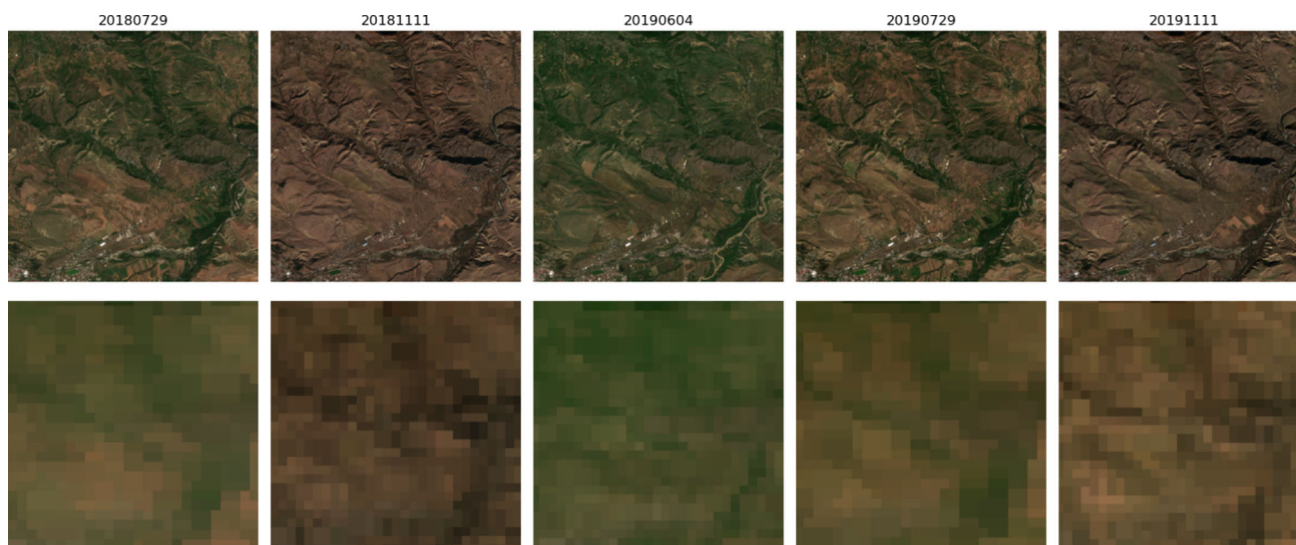


Fig. 12 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Akhaltsikhe location

park is home to a rich biodiversity, including forest elephants, gorillas, and numerous bird species. Its landscape is characterized by rolling hills, thick forests, and open grasslands, creating a diverse and challenging environment for STF methods. However, due to the cloud coverage of the tropical area, we have few images of this site.

Figure 14 presents examples of the images in this location.

A.1.7 Burggen, Germany

Burggen is a small town located in the Bavaria region of southern Germany, with a combination of rural, natural

and semi-urban landscapes. The area includes agricultural fields, dense forests, and a river that runs through the town, providing a mixture of types of land cover. Burggen experiences snowfall during winter, leading to significant seasonal changes in land cover, with snow-covered fields and forests in winter transitioning to lush greenery in spring and summer. This seasonal variability presents a challenge for STF methods, as the models must accurately capture temporal transitions between snow-covered and snow-free periods, making Burggen an excellent site for evaluating STF performance in dynamic seasonal conditions.

Figure 15 presents examples of images in this location.

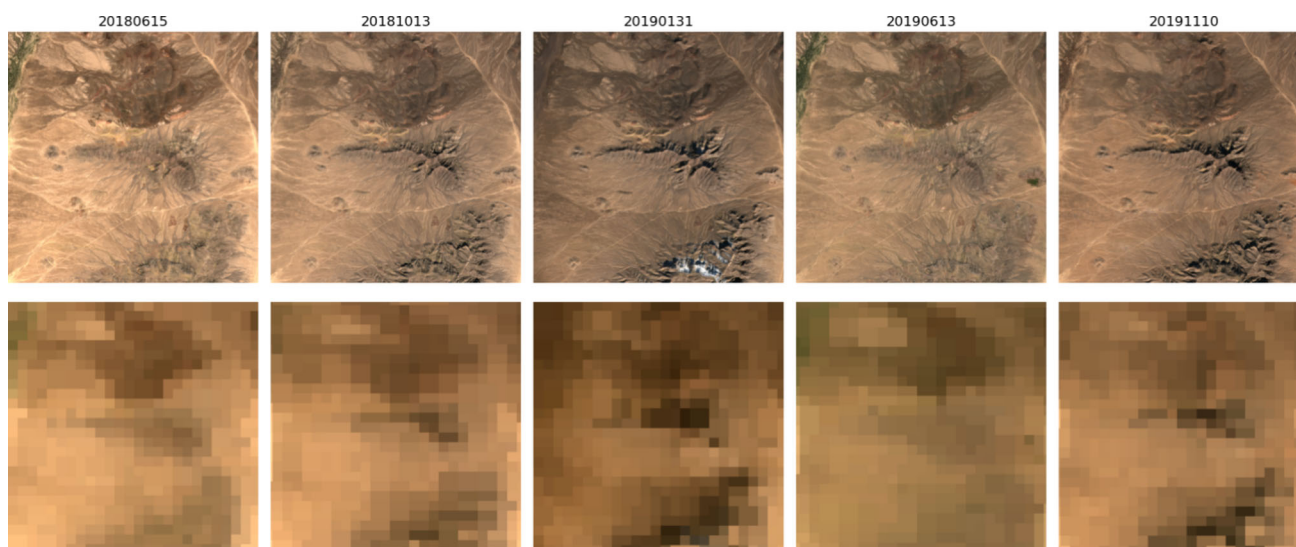


Fig. 13 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Nevada location

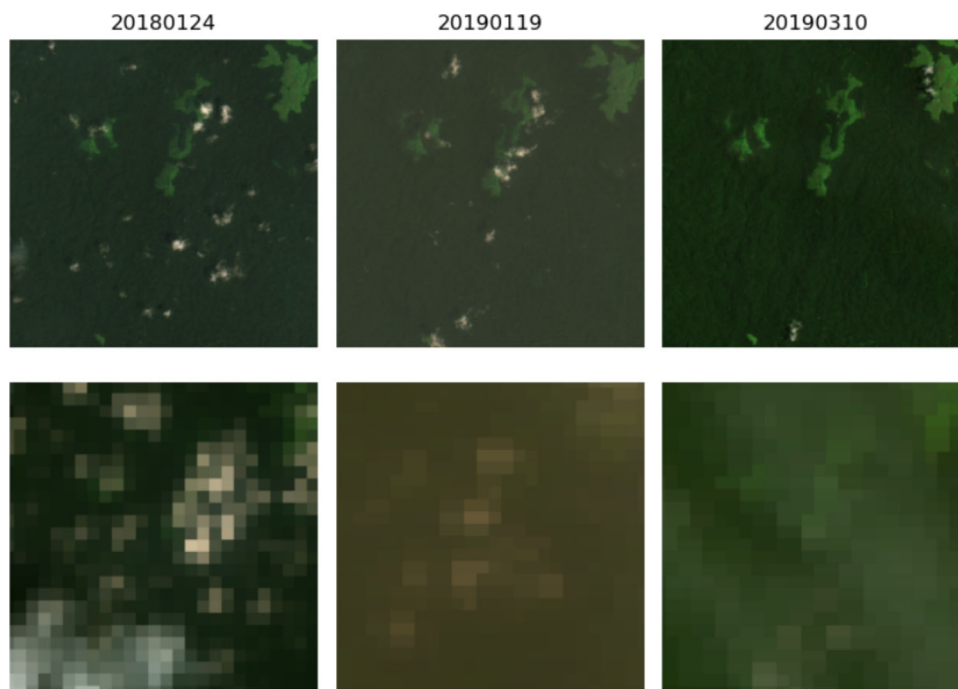


Fig. 14 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for De la Lopé Park

A.1.8 Amazon Region, Brazil

The selected site in the Amazon region of Brazil is part of the world's largest tropical rainforest, renowned for its biodiversity and complex ecological processes. The Amazon rainforest experiences high rainfall, dense vegetation, and frequent cloud cover. This region is characterized by vast expanses of tropical forests, winding rivers, and wetlands, with rapid changes in vegetation growth, and seasonal flooding. Persistent cloud coverage in the Amazon often limits the

availability of cloud-free optical images, thereby reducing the number of usable image pairs. This makes the Amazon region a critical site for evaluating the robustness of STF methods, particularly in their ability to reconstruct missing data and maintain temporal consistency in regions with limited clear-sky observations. Dynamic changes in land cover, combined with cloud-induced data gaps, create a challenging but essential test environment for advanced STF algorithms.

Figure 16 presents examples of images in this location.

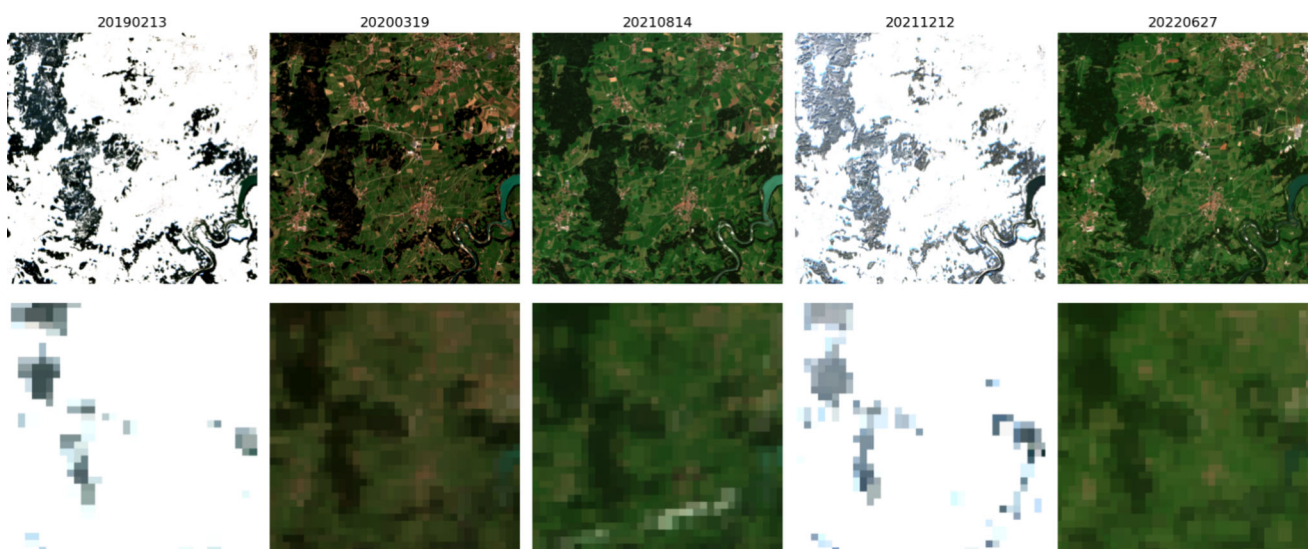
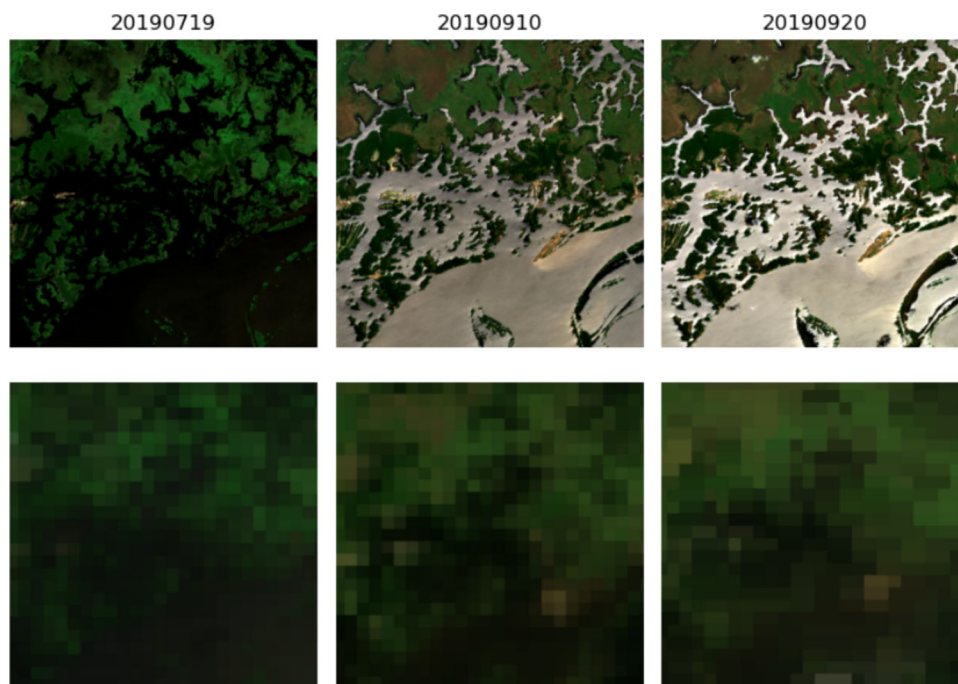


Fig. 15 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Burggen site

Fig. 16 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Amazon location



A.1.9 Tagleft, Morocco

Tagleft, situated in the Moyen Atlas region of Morocco, is known for its mountainous terrain, dense cedar forests, and agricultural terraces. The area experiences a Mediterranean climate with wet winters and dry summers, resulting in distinct seasonal variations in vegetation and water availability. The mountainous landscape of Tagleft presents challenges for spatiotemporal fusion methods, particularly in capturing seasonal changes across varying elevations and complex topography. The combination of steep slopes, forested areas,

and cultivated lands in the region makes it an ideal site to test the adaptability of STF models to mountain environments where terrain-induced variations in reflectance and illumination can affect the accuracy of image fusion processes.

Figure 17 presents examples of images in this location.

A.1.10 Asturias, Spain

Asturias, located in northern Spain, is characterized by its lush green landscapes, rugged coastline, and mountainous terrain. The region has an oceanic climate, with high rainfall

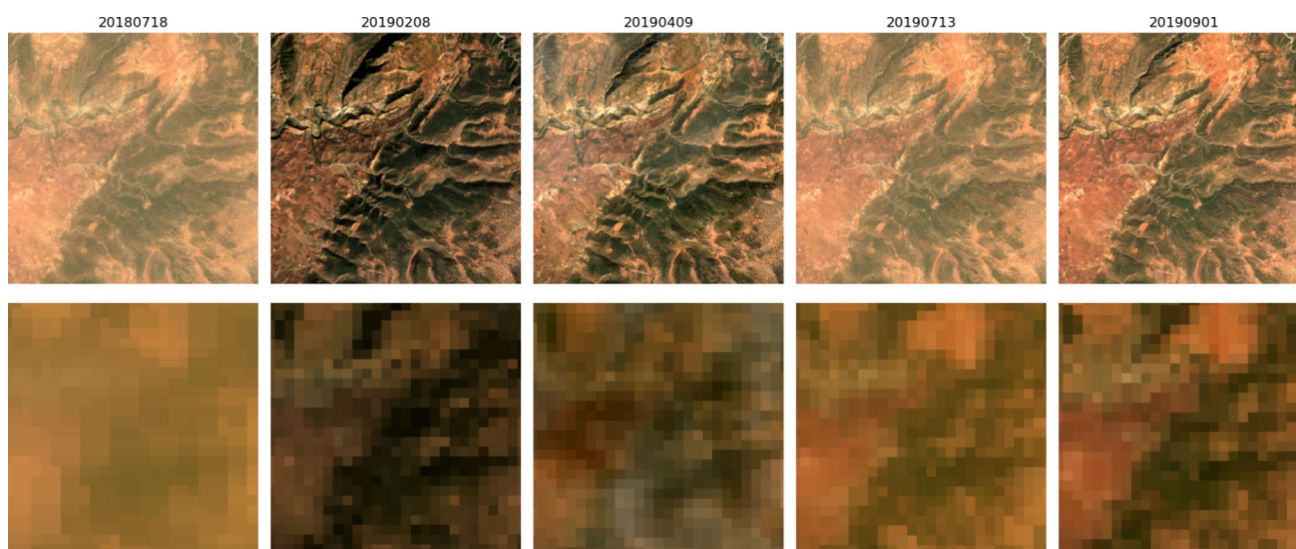


Fig. 17 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Tagleft location

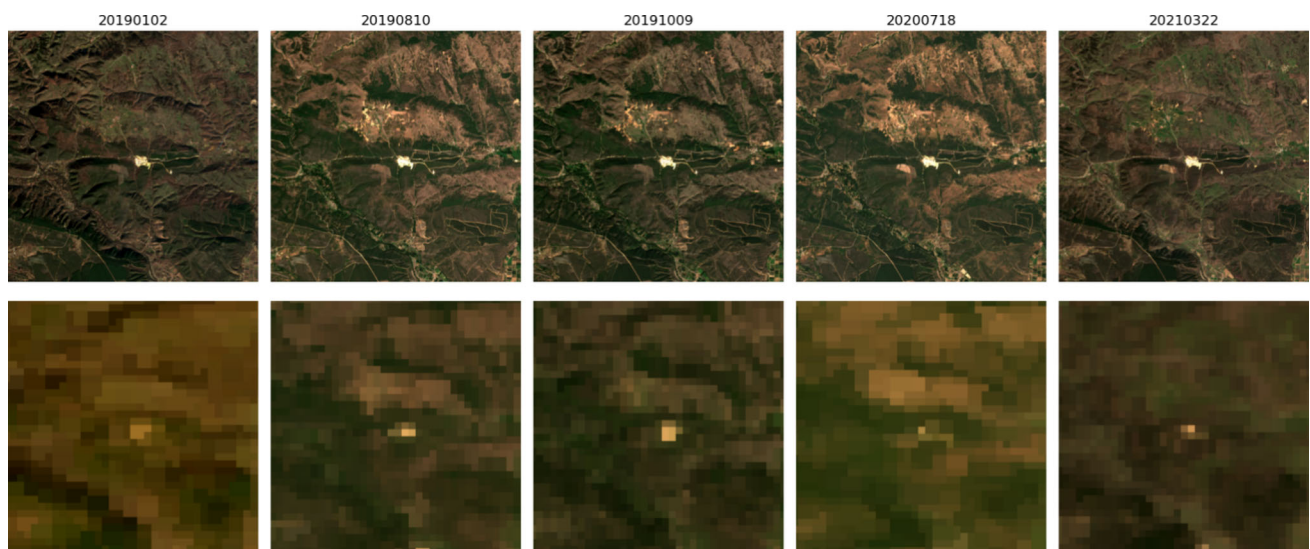


Fig. 18 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Asturias location

and mild temperatures throughout the year. The varied topography of Asturias, including coastal cliffs, dense forests, and agricultural valleys, provides a diverse environment for the STF testing. The frequent cloud cover and rapid changes in vegetation make it an ideal site for evaluating the temporal and spectral accuracy of fusion methods.

Figure 18 presents examples of images in this location.

A.1.11 Fiano Romano, Italy

Fiano Romano, a town near Rome, is a highly heterogeneous site that includes urban areas, agricultural fields, an airport, and a river. This combination of land cover types poses challenges for STF methods, as models must accurately capture

both built-up areas and natural landscapes. The site experiences seasonal agricultural changes and continuous urban development, making it an excellent testbed for assessing the adaptability of STF algorithms to complex and dynamic environments.

Figure 19 presents examples of images in this location.

A.1.12 Pontecorvo, Italy

Pontecorvo, located in the Lazio region of central Italy, is a historic town surrounded by agricultural land, rolling hills, and river valleys. The Mediterranean climate and agricultural activity of the area result in seasonal changes in land cover, including crop growth cycles and irrigation patterns.

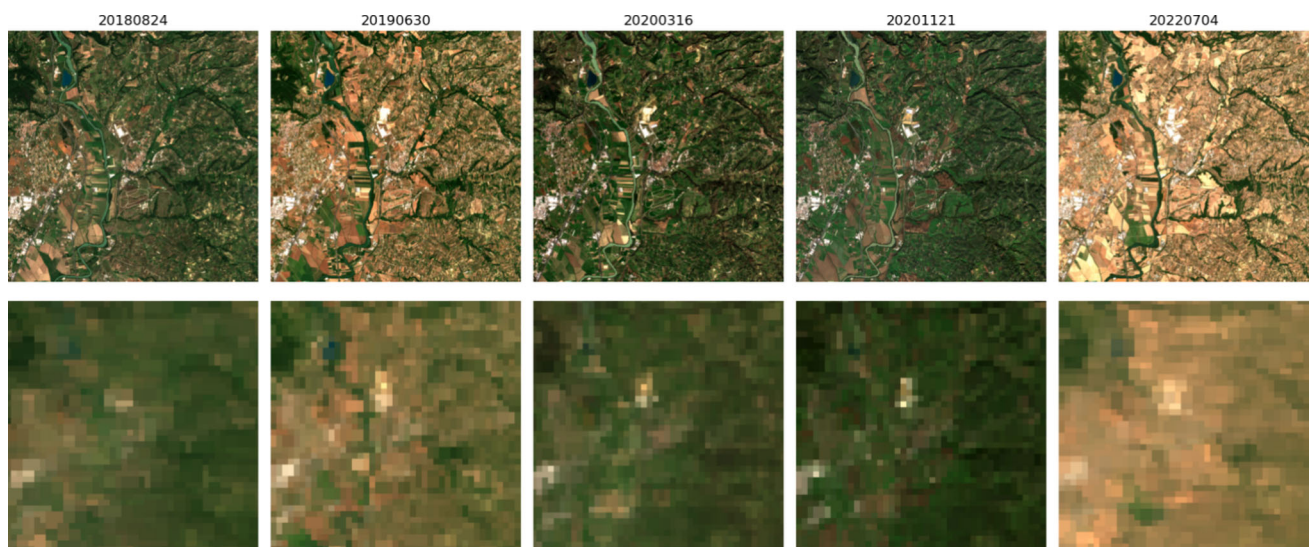


Fig. 19 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Fiano Romano location

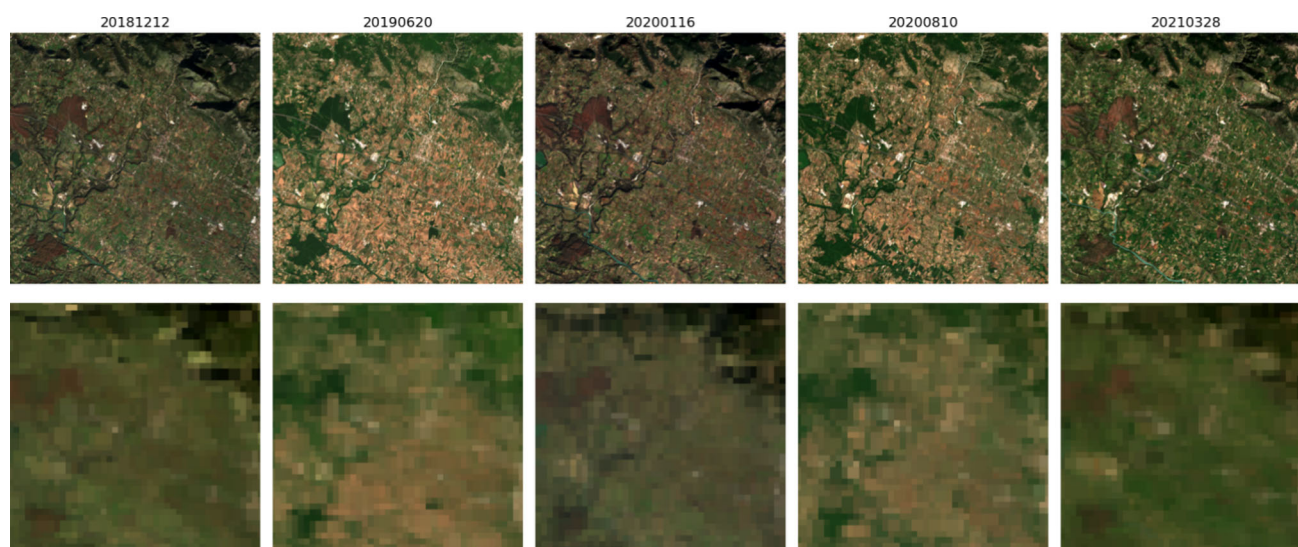


Fig. 20 Example of Sentinel-2 (top row) and Sentinel-3 (bottom row) pairs for Pontecorvo location

Pontecorvo's mix of historical architecture, agricultural landscapes and natural features makes it a valuable site for testing STF methods in regions with moderate spatial heterogeneity and seasonal variability.

Figure 20 presents examples of images in this location.

Acknowledgements M.B. thanks GMV and the Fundación Mujeres por Africa for their support for her doctoral thesis under the Learn Africa program. M.L.-S. acknowledges the support of the Chilean Science Council (ANID) through the Program Anillo (ACT210080) and CRHIAM (ANID/FONDAP/1523A0001). We thank the European Space Agency for providing us with the pre-processed Sentinel-2 and Sentinel-3 data and the respective software tools. The authors would like to thank the developers of STARFM, ESTARFM, FSDAF, Fit-FC, and SFSDAF algorithms for sharing their codes.

Author Contributions M.B: Planning, methodology, analysis, experiments, initial draft writing. A.G-P: Methodology, conception, supervision, original draft writing and revision. M.L-S: Validation, review and editing. C.G-M: Funding procurement, supervision, review and editing. All authors have read and agreed to the published version of the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work was developed within the framework of the Fire CCI+ project, funded by the ESA (no. 4000126706/19/I-NB).

Data Availability The dataset is freely available at <https://zenodo.org/records/14860220>

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the

source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Berk A, Anderson GP, Acharya PK, Bernstein LS, Muratov L, Lee J, Fox M, Adler-Golden SM, Chetwynd Jr, JH, Hoke ML et al (2006) Modtran5: 2006 update. In: Algorithms and technologies for multispectral, hyperspectral, and ultraspectral Imagery XII, vol 6233. International Society for Optics and Photonics, p 62331
- Bolton DK, Friedl MA (2013) Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric For Meteorol* 173:74–84
- Boumahdi M, Garcia-Pedrero A, Lillo-Saavedra M, Gonzalo-Martin C (2023) Evaluation of spatiotemporal fusion methods using sentinel-2 and sentinel-3: a new benchmark dataset and comparison. Zenodo. <https://doi.org/10.5281/zenodo.7970846>
- Cazzaniga I, Bresciani M, Colombo R, Della Bella V, Padula R, Giardino C (2019) A comparison of sentinel-3-olci and sentinel-2-msi-derived chlorophyll-a maps for two large Italian lakes. *Remote Sens Lett* 10(10):978–987
- Chen B, Huang B, Xu B (2015) Comparison of spatiotemporal fusion models: a review. *Remote Sens* 7(2):1798–1835
- Cheng Q, Liu H, Shen H, Wu P, Zhang L (2017) A spatial and temporal nonlocal filter-based data fusion method. *IEEE Trans Geosci Remote Sens* 55(8):4476–4488
- Donlon C, Berruti B, Buongiorno A, Ferreira M-H, Féménias P, Frerick J, Goryl P, Klein U, Laur H, Mavrocordatos C et al (2012) The global monitoring for environment and security (gmes) sentinel-3 mission. *Remote Sens Environ* 120:37–57
- Donlon C, Berruti B, Mecklenberg S, Nieke J, Rebhan H, Klein U, Buongiorno A, Mavrocordatos C, Frerick J, Seitz B et al (2012)

- The sentinel-3 mission: Overview and status. In: 2012 IEEE International geoscience and remote sensing symposium. IEEE, pp 1711–1714
- Drusch M, Del Bello U, Carlier S, Colin O, Fernandez V, Gascon F, Hoersch B, Isola C, Laberinti P, Martimort P et al (2012) Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote Sens Environ* 120:25–36
- Du Y, Zhang Y, Ling F, Wang Q, Li W, Li X (2016) Water bodies' mapping from sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the swir band. *Remote Sens* 8(4):354
- Emelyanova IV, McVicar TR, Van Niel TG, Li LT, Van Dijk AI (2013) Assessing the accuracy of blending landsat-modis surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. *Remote Sens Environ* 133:193–209
- Fensholt R (2004) Earth observation of vegetation status in the sahelian and sudanian west africa: comparison of terra modis and noaa avhrr satellite data. *Int J Remote Sens* 25(9):1641–1659
- Gao F, Masek J, Schwaller M, Hall F (2006) On the blending of the landsat and modis surface reflectance: predicting daily landsat surface reflectance. *IEEE Trans Geosci Remote Sens* 44(8):2207–2218
- Hagolle O, Sylvander S, Huc M, Claverie M, Clesse D, Dechoz C, Lonjou V, Poulain V (2015) Spot-4 (take 5): simulation of sentinel-2 time series on 45 large sites. *Remote Sens* 7(9):12242–12264
- Han Z, Zhao W (2019) Comparison of spatiotemporal fusion models for producing high spatiotemporal resolution normalized difference vegetation index time series data sets. *J Comput Commun* 7(7):65–71
- Han P, Wang PX, Zhang SY et al (2010) Drought forecasting based on the remote sensing data using arima models. *Math Comput Model* 51(11–12):1398–1403
- Hore A, Ziou D (2010) Image quality metrics: Psnr vs. ssim. In: 2010 20th International conference on pattern recognition. IEEE, pp 2366–2369
- Ibarrola-Ulzurrun E, Gonzalo-Martin C, Marcello-Ruiz J, Garcia-Pedrero A, Rodriguez-Esparragon D (2017) Fusion of high resolution multispectral imagery in vulnerable coastal and land ecosystems. *Sensors* 17(2):228
- Li X, Foody GM, Boyd DS, Ge Y, Zhang Y, Du Y, Ling F (2020) Sfsdaf: An enhanced fsdaf that incorporates sub-pixel class fraction change information for spatio-temporal image fusion. *Remote Sens Environ* 237:111537
- Li J, Li Y, He L, Chen J, Plaza A (2020) Spatio-temporal fusion for remote sensing data: an overview and new benchmark. *Sci China Inf Sci* 63(4):1–17
- Liu M, Ke Y, Yin Q, Chen X, Im J (2019) Comparison of five spatio-temporal satellite image fusion models over landscapes with various spatial heterogeneity and temporal variation. *Remote Sens* 11(22):2612
- Loncan L, De Almeida LB, Bioucas-Dias JM, Briottet X, Chanussot J, Dobigeon N, Fabre S, Liao W, Licciardi GA, Simoes M et al (2015) Hyperspectral pansharpening: a review. *IEEE Geosci Remote Sens Mag* 3(3):27–46
- Prikaziuk E, Yang P, Tol C (2021) Google earth engine sentinel-3 olci level-1 dataset deviates from the original data: Causes and consequences. *Remote Sens* 13(6):1098
- Renosh PR, Doxaran D, Keukelaere LD, Gossn JI (2020) Evaluation of atmospheric correction algorithms for sentinel-2-msi and sentinel-3-olci in highly turbid estuarine waters. *Remote Sens* 12(8):1285
- Rufin P, Frantz D, Yan L, Hostert P (2020) Operational coregistration of the sentinel-2a/b image archive using multitemporal landsat spectral averages. *IEEE Geosci Remote Sens Lett* 18(4):712–716
- Scheffler D, Hollstein A, Diedrich H, Segl K, Hostert P (2017) Arosics: An automated and robust open-source image co-registration software for multi-sensor satellite data. *Remote Sens* 9(7):676
- Song H, Liu Q, Wang G, Hang R, Huang B (2018) Spatiotemporal satellite image fusion using deep convolutional neural networks. *IEEE J Sel Top Appl Earth Obs Remote Sens* 11(3):821–829
- Sterckx S, Knaeps S, Kratzer S, Ruddick K (2015) Similarity environment correction (simec) applied to meris data over inland and coastal waters. *Remote Sens Environ* 157:96–110
- Tan Z, Yue P, Di L, Tang J (2018) Deriving high spatiotemporal remote sensing images using deep convolutional network. *Remote Sens* 10(7):1066
- Van Leeuwen WJ, Casady GM, Neary DG, Bautista S, Alloza JA, Carmel Y, Wittenberg L, Malkinson D, Orr BJ (2010) Monitoring post-wildfire vegetation response with remotely sensed time-series data in Spain, USA and Israel. *Int J Wildland Fire* 19(1):75–93
- Verhoef W, Bach H (2012) Simulation of sentinel-3 images by four-stream surface-atmosphere radiative transfer modeling in the optical and thermal domains. *Remote Sens Environ* 120:197–207
- Wang Q, Atkinson PM (2018) Spatio-temporal fusion for daily sentinel-2 images. *Remote Sens Environ* 204:31–42
- Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612
- Wang Q, Shi W, Li Z, Atkinson PM (2016) Fusion of sentinel-2 images. *Remote Sens Environ* 187:241–252
- Wang Q, Blackburn GA, Onojeghuo AO, Dash J, Zhou L, Zhang Y, Atkinson PM (2017) Fusion of landsat 8 oli and sentinel-2 msi data. *IEEE Trans Geosci Remote Sens* 55(7):3885–3899
- Yin H, Pflugmacher D, Kennedy RE, Sulla-Menashe D, Hostert P (2014) Mapping annual land use and land cover changes using modis time series. *IEEE J Sel Top Appl Earth Obs Remote Sens* 7(8):3421–3427
- Zhao Y, Huang B, Song H (2018) A robust adaptive spatial and temporal image fusion model for complex land surface changes. *Remote Sens Environ* 208:42–62
- Zhou J, Kwan C, Budavari B (2016) Hyperspectral image super-resolution: a hybrid color mapping approach. *J Appl Remote Sens* 10(3):035024
- Zhou J, Chen J, Chen X, Zhu X, Qiu Y, Song H, Rao Y, Zhang C, Cao X, Cui X (2021) Sensitivity of six typical spatiotemporal fusion methods to different influential factors: a comparative study for a normalized difference vegetation index time series reconstruction. *Remote Sens Environ* 252:112130
- Zhu X, Chen J, Gao F, Chen X, Masek JG (2010) An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens Environ* 114(11):2610–2623
- Zhu X, Helmer EH, Gao F, Liu D, Chen J, Lefsky MA (2016) A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sens Environ* 172:165–177
- Zhu X, Cai F, Tian J, Williams TK-A (2018) Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. *Remote Sens* 10(4):527