Adjustment of Sentinel-3 Spectral Bands With Sentinel-2 to Enhance the Quality of Spatio-Temporally Fused Images

Meryeme Boumahdi, Angel García-Pedrero, Mario Lillo-Saavedra, and Consuelo Gonzalo-Martin

***Abstract*—Spatiotemporal fusion (STF) methods are a paramount solution for generating high spatial and temporal time series, overcoming the limitations of spatial and temporal resolution of satellite data. STF methods typically rely on band-by-band fusion, assuming spectral similarities. However, selecting the optimal band for fusion becomes challenging when multiple narrow bands overlap with the target band, often leading to the use of only one single band. Furthermore, sensor specifications and observation configurations can further compound this challenge, reducing spectral and spatial information. We introduce a new preprocessing step that maximizes the use of spectral information from narrow bands. It minimizes radiometric differences caused by sensor variations in the STF process by considering the spectral response function (SRF). Our method generates adjusted bands that closely match the target band's spectral characteristics, leveraging all available spectral information. We evaluated this strategy at two study sites employing Sentinel 2 and Sentinel 3 data by comparing fused images from adjusted bands and the original bands using three popular STF methods. The results obtained showed that the images fused with the adjusted bands were closer to the target images and achieved better performance, improving the fusion quality compared to the original bands (SAM by 37% and RMSE by 30%). The preprocessing step offers a feasible approach to generate spectral bands that would be captured by the sensors if they had the same spectral characteristics. Importantly, this preprocessing technique is applicable to any STF method.**

***Index Terms*—Band adjustment, bands overlapping, Sentinel-2, Sentinel-3 OLCI, spatiotemporal data fusion, spectral response function (SRF).**

Manuscript received 26 August 2023; revised 30 October 2023; accepted 8 November 2023. Date of publication 15 November 2023; date of current version 29 November 2023. This work was developed within the framework of the Fire CCi+ project, funded by the ESA under Grant 4000126706/19/I-NB. The work of Mario Lillo-Saavedra was supported by Chilean Science Council (ANID) through the Program Anillo [ACT210080] and CRHIAM [ANID/FONDAP/15130015]. *(Corresponding author: Consuelo Gonzalo-Martin.)*

Meryeme Boumahdi, Angel García-Pedrero, and Consuelo Gonzalo-Martin are with the Center for Biomedical Technology, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Madrid, Spain, and also with the Department of Computer Architecture and Technology, School of Computer Engineering, Universidad Politécnica de Madrid, 28660 Boadilla del Monte, Madrid, Spain (e-mail: m.boumahdi@alumnos.upm.es; angelmario.garcia@upm.es; consuelo.gonzalo@upm.es).

Mario Lillo-Saavedra is with the Facultad de Ingeniería Agrícola, Universidad de Concepción, Chillán 3812120, Chile (e-mail: malillo@udec.cl).

Digital Object Identifier 10.1109/JSTARS.2023.3333275

I. Introduction

As SENSOR technologies have advanced, there has been a noticeable improvement in the spatial, spectral, and temporal resolutions of images captured by optical remote sensing sensors used for land cover studies. However, technical limitations create a fundamental tradeoff among these resolutions [1]. For instance, the multispectral instrument (MSI) on board Sentinel-2 (S2) provides images with high spatial resolutions (10 m, 20 m, and 60 m) and a frequency of revisiting of 5 days. On the other hand, the optical instruments (OLCI and SLSTR) of the Sentinel-3 (S3) mission offer images with a higher temporal frequency (less than 2 days) but lower spatial resolution (300 m). Due to this tradeoff, the available satellite images may not always be sufficient to fulfill all the requirements for studying high-frequency changes in heterogeneous landscapes [2].

Numerous research studies have focused on merging images from different sensors to obtain time series with better spatial and temporal resolutions [3], [4], [5], [6]. In this context, spatiotemporal fusion (STF) methods have received significant attention for their ability to address the tradeoff problem between the spatial and temporal resolutions of sensors [2]. STF methods combine high spatial and low temporal resolution (HSLT) images with low spatial and high temporal resolution (LSHT) images to predict time series of images with high spatial and high temporal resolution (HSHT).

In [2], five categories of STF methods are identified: unmixing-based, weight function-based, Bayesian-based, learning-based, and hybrid methods. Recently, deep learning-based approaches have been proposed for STF [7], [8], yet some papers have identified certain limitations in their use. For instance, [9] identifies the challenge of using deep learning methods to fuse images with changes in land cover. Moreover, [10] emphasizes the limitations of convolutional neural networks in capturing global features and long-distance correlations in STF. This study looks into three of the most popular STF techniques: STARFM [11], FSDAF [12], and Fit-FC [13]. STARFM is a weight function-based approach, FSDAF is a hybrid method, and Fit-FC is one of the few methods designed to combine S2 and S3 OLCI images. Although all these methods have thoroughly considered the spatial and temporal differences of the input images, spectral discrepancies have received less attention [14]. This limitation hinders their application in studies of heterogeneous landscapes [15].

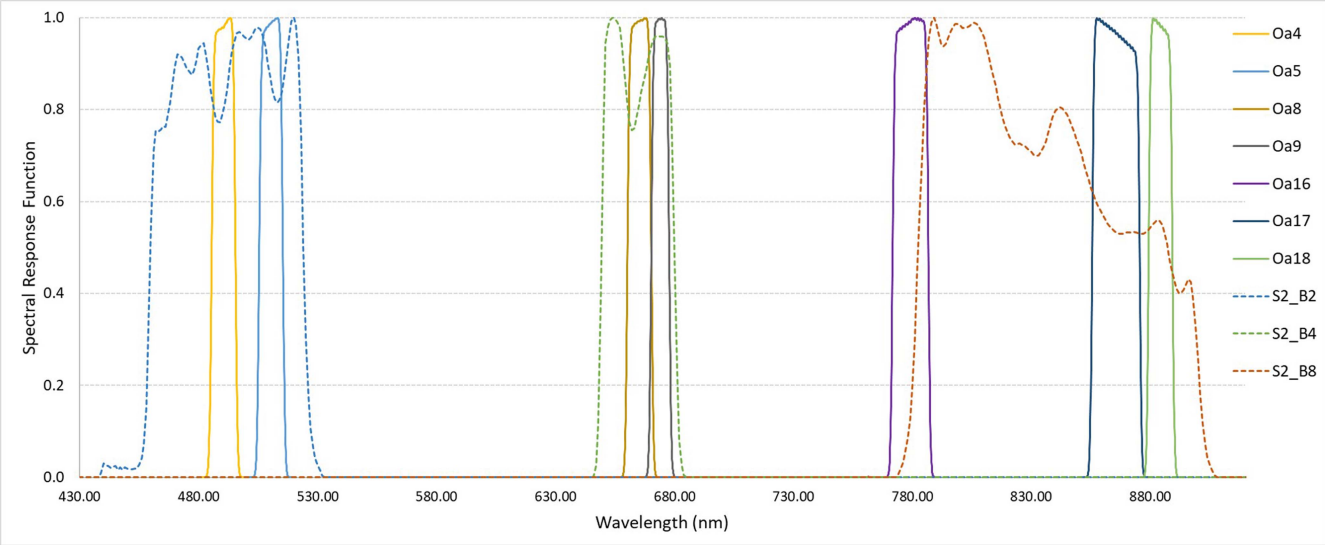


Fig. 1. SRFs corresponding to S2 (MSI) and S3 (OLCI). S2 SRF is presented by a dashed line and S3 SRF is presented by a continuous line.

Differences in spectral resolution between sensors arise from variations in the number of spectral bands and their widths, resulting in situations where the wider spectral band of one sensor might overlap with multiple narrow bands of another sensor. Fig. 1 illustrates this scenario for S2 (MSI) and S3 (OLCI) imagery, where the spectral response function (SRF), also known as relative spectral responsivity, of the S2 spectral bands overlaps with all or part of the SRF of the S3 spectral bands. Although several studies have shown that even a limited overlap between the two sensors can influence the outcomes of fusion [16], [17], the optimal selection of bands for fusion remains ambiguous. Currently, there are no established criteria or guidelines to assist in this selection process.

Most of the STF methods assume that the input images have the same spectral characteristics and ignore other bands that might overlap with the wide band considered in the fusion. This implies that important information provided by some spectral bands could be neglected. As a consequence, the spectral characteristics of fused images can be distorted with respect to the source images [18], [19], [20].

To mitigate this type of distortion, it is essential to incorporate as much spectral information as possible during the STF process. This can be achieved by taking into account all narrow bands within the same spectral range as the wide band [21]. However, accomplishing this requires ensuring spectral consistency across data collected by different sensors, which are characterized by different SRFs.

To address the difference in spectral characteristics among sensors, numerous band adjustment approaches have been reported in the literature. For instance, [22] introduced WiSpeR, a wavelet-based technique for spatio-spectral fusion, particularly in pansharpening. This method takes into account the difference in SRF to determine how much information from the panchromatic image should be injected into the multispectral image bands. In a similar vein, Aiazzi et al. [23] quantified the SRF of the sensors by calculating linear regression coefficients between panchromatic and multispectral images, enhancing the spectral quality of pansharpened images. In the case of the STF methods, Cao [24] considered differences in sensor spectral information by using linear regression models (RMs) that accounted for various land covers and spectral bands. These models were used to compute the weight function of the STARFM method, thereby enhancing the spatial consistency of the fused image. Another linear regression-based strategy was proposed in [25], in which fitting coefficients were derived between the MODIS and Landsat bands. These coefficients were then employed to generate fine-like images, which were subsequently used in the fusion process. The use of fine-like images, as opposed to the original bands, aimed to mitigate the uncertainty errors between finer and coarser sensors. Results from this study demonstrated that this preprocessing step substantially improved the fusion accuracy of typical STF methods. In general terms, statistical regression emerges as the most widely used approach in band adjustment to address such issues. Regression coefficients serve to compensate the differences in multispectral sensor spectral responses, thereby facilitating more precise cross-calibration across these sensors [26], [27]. However, it is worth noting that STF methods, in general, primarily focus on band-to-band fusion, often disregarding the remaining overlapped bands.

Disregarding such data can lead to the loss of valuable spectral information and a significant increase in spectral differences between the sensors. Numerous studies on spatio-spectral fusion have consistently demonstrated the effectiveness of incorporating multiple bands into the fusion process [16], [28]. In a related approach, Wang et al. [29] introduced Area-To-Point Regression Kriging (ATPRK) to enhance the quality of spatio-spectral fusion by using multispectral images as input to predict a single band.

However, to the best of our knowledge, no studies have considered the inclusion of more than one narrow spectral band that overlaps with a wide band in the context of STF. This constitutes the main contribution of our work.

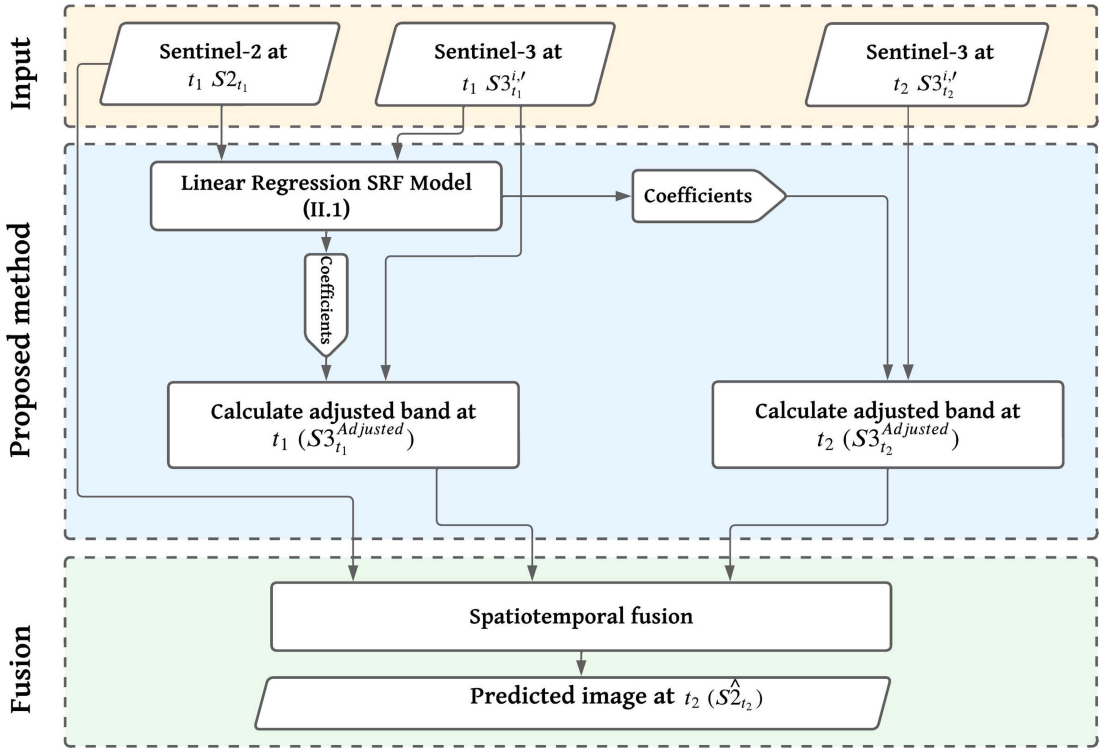


Fig. 2. Flowchart of the proposed band adjustment method for one single S2 band with multiple S3 bands.

In this work, we present a new preprocessing step named SRF-adjustment, designed to enhance the spectral quality of predicted images produced by STF methods. Since this step occurs prior to the STF process, the proposed strategy can be seamlessly integrated into any STF approach. Our method revolves around generating adjusted bands, which involves incorporating all available spectral information from bands within the same spectral range as the target band (high spatial resolution image). This effectively merges information from multiple narrow bands (low spatial resolution image) with the broader bands. The adjusted bands exhibit spectral similarity to what would have been recorded if the source sensors had similar spectral characteristics. This approach helps mitigate the spectral disparities caused by sensor specifications and retains a substantial portion of the spectral details. Specifically, our research addresses the inconsistencies in spectral data arising from differences in the SRFs of the European sensors S2 and S3.

SRF-adjustment method has been evaluated at two study sites using real data, comparing the fused images obtained by the adjusted bands and the original bands using three different STF methods, STARFM [11], FSDAF [12] and Fit-FC [13]. Using the SRF-adjusted bands as input to the STF methods, the spectral quality of the predicted images is improved, as is proved by the obtained results, without modifying the methods. In particular, the results obtained show that STF methods are more effective in predicting heterogeneous landscapes when adjusted bands are used.

The rest of this article is organized as follows. Section II provides a comprehensive description of the data, methods, and materials utilized in this study. The experimental results are presented in Section III. Section IV presents a discussion of the results and conclusion are presented in Section V. The mathematical theory is included in the Annex A.

II. Data and Materials

Fig. 2 presents the flowchart of the SRF-adjustment strategy used to improve the spectral quality of images estimated using three STF methods. $S3^{i,\prime}_{t1}$ and $S3^{i,\prime}_{t2}$ indicate the downscaled S3 bands at t_1 and t_2 with an overlap with $S2_{t1}$, where $i = 1, ..., n$ and n is the number of S3 bands that overlap with one S2 band. Images pairs S2 and S3 were collected on different dates (https://doi.org/10.5281/zenodo.7970846). The base date, denoted t_1, represents the date when both the S2 and S3 images were available. The date t_2 corresponds to the date of the target image. In this study, S2-S3 bands were used, focusing specifically on those with multiple overlaps (see Table I), S2 bands that overlap with only one S3 are not considered in this work. A description of these datasets is included in Section II-A.

The first step of the SRF-adjustment strategy involves calculating the coefficients of the linear RM between the wide band of S2 and the narrow bands of S3 at the base date t_1. The spatially downscaled S3 bands are used along with S2 bands as an input to generate the linear RM coefficients. The linear fitting coefficients are presented in the flowchart as a_i, where $i = 1, 2, ..., n$ and n is the number of original S3 bands that overlap with the S2 band. These fitting coefficients obtained from the inputs at t_1 are then used to generate an adjusted S3 band

TABLE I
DESCRIPTION OF THE SENTINEL-2 AND SENTINEL-3 OLCI BANDS PRESENTED IN FIG. 1

Specification	Band	Sentinel-2		Sentinel-3	
		MSI band	Wavelength range (nm)	OLCI band	Wavelength range (nm)
Band width	Blue	B02	458–523	Oa4 Oa5	485–495 505–515
	Red	B04	650–680	Oa8 Oa9	660–670 670–677
	NIR	B08	780–885	Oa16 Oa17 Oa18	771–786 855–875 880–890
Spatial resolution		10 m		300 m	
Temporal resolution		5 days		1.5 days	

from the original S3 bands at the base date. The same coefficients are then used to generate an adjusted S3 band from the original S3 bands that overlap with the S2 band on the prediction date t_2. Further details of the adjustment process can be found in Section II-B. Once the S3 bands have been adjusted to S2 at t_1 and t_2, they are used as input to the STF methods, along with the S2 band at t_1. Adjusted S3 bands replace the original S3 bands in both t_1 and t_2 in the fusion process. A brief description of the STF methods considered in this work is included in Section II-C, and Section II-D outlines the metrics used to evaluate the quality of the predicted images.

A. Data and Study Areas Description

Table I summarizes the spectral, spatial, and temporal characteristics of the S2 and S3 imagery. As previously mentioned, only bands with more than one overlapping band (blue, red, and NIR) have been considered. S2 is a multispectral sensor with a high spatial resolution of 10–60 m, depending on the band, making it well-suited for vegetation monitoring. However, its temporal resolution is relatively low for monitoring natural disasters as an example, with images being captured every five days. S3 is known for its high temporal resolution of 1.5 days and can cover most of the Earth. However, it obtains data with a spatial resolution of approximately 300 m.

In this study, Sentinel products, specifically S2 level 1C and S3 Ocean and Land Colour Instrument (OLCI) level 1B data, were utilized. S2 level-1C processing involves a range of radiometric and geometric corrections. These corrections encompass ortho-rectification, the addition of radiometric offsets, and spatial registration onto a global reference system with subpixel accuracy. Similarly, S3 OLCI level 1B data processing includes calibration to geophysical units of Top Of Atmosphere (TOA) radiances, georeferencing onto the Earth's surface, and spatial resampling onto a uniformly spaced grid. Various preprocessing steps were performed to generate this dataset. First, S2 and S3 products were atmospherically corrected using iCOR atmospheric correction algorithm [30], [31], which relies on MODTRAN5 look-up-tables (LUTs) [32], designed for each specific product. This correction provided bottom of atmosphere (BOA) reflectances for both products. Next, the S3 images were re-projected to match the S2 projection, and finally, they

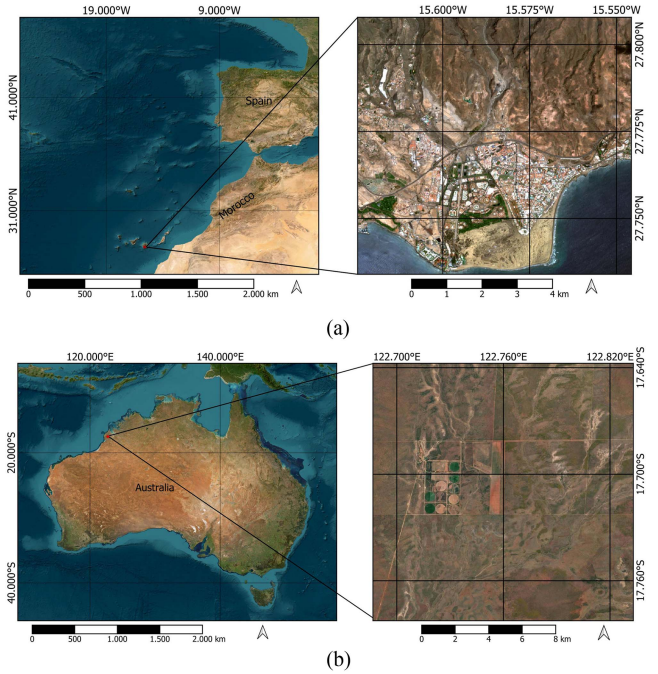


Fig. 3. Geographic location of (a) Maspalomas, Gran Canarias, Spain. (b) Waterbank site, North Australia.

were coregistered to reduce the impact of geometric errors. The coregistration of the S2/S3 data was conducted in two steps. Initially, a single pair of S2/S3 imagery was coregistered using ENVI 5.6.1 software. Subsequently, each of these coregistered images was employed to coregister the rest of the dates of each time series using AROSICS software [33]. In the end, the S3 images were spatially downscaled to match the spatial resolution of S2. Nearest neighbor algorithm was used for downscaling to preserve the original spectral information in S3. At the conclusion of the preprocessing, the range of both types of images is normalized to a scale of 0 to 1.

To evaluate the performance of our method, two sites have been considered. The location of these sites is shown in Fig. 3. The first site is located in a natural reserve in the south of Gran Canaria Island called Maspalomas (27.760°N, 15.587°E) in Spain. The image of this site covers an area of 9 km × 9 km (900 × 900 Sentinel-2 pixels). In addition to the vulnerability of its ecosystem, the site is also known for frequent sandstorms throughout the year. The second site, located in Waterbank (17.706°S, 122.7259°E), is a rural site in North Australia. It exhibits rapid temporal changes due to forest fires. This location is well known for the high frequency of early season fires, which leads to different phenological changes in vegetation throughout the year. The data cover an area of 18 km × 18 km (1800 × 1800 Sentinel-2 pixels). For both sites, 40 different pairs were used in each location from January 2019 to December 2020. We chose the closest prior date to the target date t_2 as the base date t_1. Details of the data set can be found at https://doi.org/10.5281/zenodo.7970846.

The reason for selecting the two sites mentioned above was to analyze the performance of the STF considered methods, when

the adjusted bands are used as input. Sites exhibit scenarios with a high level of complexity due to their pronounced degrees of change and heterogeneity, posing a challenge for STF methods. The first scenario will allow us to test the performance of the adjusted bands in the highly heterogeneous landscape of the Maspalomas site. The Waterbank site will be used for the second scenario due to its rapid changes caused by fire. The importance of spectral information provided by multiple bands appears mainly in such cases, where the heterogeneity of the landscape is characterized by the spectral data, and spatial change vary along with the spectral changes.

B. SRF-Adjustment of Spectrally Overlapped Bands for S2 and S3 Images

Unlike the band adjustment methods commonly found in the literature, our SRF-adjustment approach involves using multiple S3 bands that overlap with a single S2 band. The adjusted band generated with our approach shows a similarity in the spectral characteristics to those of the S2 band. Consequently, these adjusted S3 bands can be seamlessly incorporated as input into the STF process.

The theoretical basis for the spectral adjustment between the narrow S2 (MSI) spectral bands and the wide S3 (OLCI) spectral bands was developed in Appendix A.

A mathematical relationship between the wide and narrow bands was demonstrated in Appendix A, as well as the contribution of the SRF to the adjustment of the spectral bands (A.27). This contribution is provided by the coefficients α^i in (A.25). On the basis of this, we propose a linear RM to generate adjusted bands combining information from the narrow bands that overlap with the one HSLT resolution sensor band. Taking into account the S2 image at the base date t_1 ($S2_{t_1}$) and the spatially downscaled S3 images at the base date ($S3'_{t_1}$) and the prediction date t_2 ($S3'_{t_1}$) for the STF, the relationship between the coarse-resolution and fine-resolution sensor bands can be expressed as a linear RM defined by the following equation:

$$F_{t_1} = \sum_{i=1}^{N} a_{i_1} C_{t_1}^{i,\prime} \tag{II.1}$$

where F_{t_1} is the S2 band on the base date (t_1); N is the number of S3 bands that overlap with F_{t_1}; and $C_{t_1}^{i,\prime}$ is the spatially downscaled S3 band to the spatial resolution of the S2 sensor on the date base (t_1), and a_{i_1} is the regression coefficient calculated using the least squares estimation. Similarly, we have the following equation:

$$F_{t_2} = \sum_{i=1}^{N} a_{i_2} C_{t_2}^{i,\prime} \tag{II.2}$$

with t_2 corresponding to the prediction date t_2. Since the linear regression approach simulates the contribution of the SRFs in the adjustment, we assume that the regression coefficients are constant over time because the SRFs are independent of the date. Thus, $a_{i_1} = a_{i_2} = a_i$, the coefficients obtained from the base date t_1 are used to calculate the adjusted bands S3 at the predicted date t_2. The difference of F between t_1 and t_2 could be written as

$$F_{t_2} - F_{t_1} = \sum_{i=1}^{N} a_i \left(C_{t_2}^{i,\prime} - C_{t_1}^{i,\prime} \right). \tag{II.3}$$

This equation is similar to (A.27), with

$$a_i = \frac{\alpha^i}{m}. \tag{II.4}$$

With m being the number of the narrow overlapped bands. The linear regression coefficients represent the contribution of each S3 band to the adjustment to harmonize the images from the two sensors (S2 and S3).

C. STF Methods Description

Three of the most widely known STF models are used to test the effectiveness of the use of adjusted bands in improving the fusion results: STARFM [11], FSDAF [12], and Fit-FC [13]. STARFM is a weight function-based method, FSDAF is a hybrid method, and Fit-FC is one of the few methods proposed to fuse Sentinel-2 with Sentinel-3 OLCI images.

1) STARFM: One of the first methods for image fusion was STARFM. It is a reconstruction-based method that has been widely used in STF [11]. STARFM requires at least one pair of fine-coarse images on a prior date t_1 and a coarse image on the predicted date t_2. The algorithm assumes that both coarse and fine sensors capture the same reflectance with bias and that the temporal changes of all the classes between the coarse images are uniform. If the coarse pixel is homogeneous, the algorithm assumes that the spectral change is consistent between t_1 and t_2, thus the spectral changes are added directly to the fine image at t_1. The land cover in the fine resolution image is considered when the coarse pixel contains mixed land cover types. A moving window is used to calculate a weight function for the contribution of the spatial difference, temporal difference, and spectral difference of neighboring pixels similar to the pixel in the center. The final step consists of calculating the surface reflectance of the predicted image by incorporating the high- and the coarse-resolution images through the weight function calculated in the previous steps.

2) FSDAF: Flexible Spatiotemporal Data Fusion (FSDAF) [12] is a hybrid method that combines spectral unmixing, spatial interpolation, and similar neighboring pixel smoothing to improve prediction accuracy. It is based on spatial unmixing and spatial prediction by thin plate spline (TPS) interpolator. FSDAF requires a fine-coarse pair image on the base date (t_1) and a coarse image on the target date (t_2), as well as a spectral cluster map. First, FSDAF generates a classification map of the fine image using an unsupervised algorithm and then estimates the temporal changes of the two coarse images. A temporal prediction at t_2 is generated from the changes at the class level from the previous step. Next, the coarse image at t_2 is downscaled to match the pixel size of the fine image to obtain another prediction called spatial prediction with a TPS interpolator. The TPS prediction is then used to distribute the residuals of the temporal prediction within a moving window.

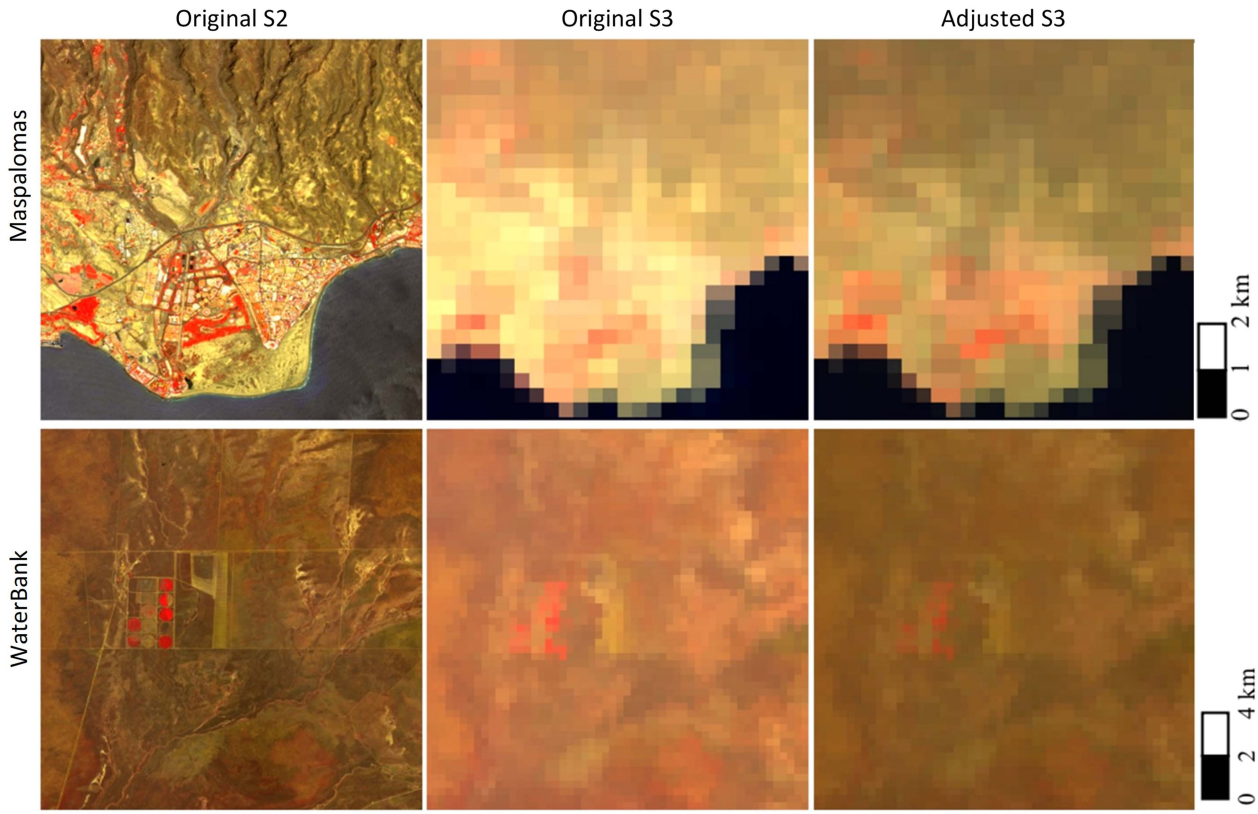


Fig. 4. NIR-Red-Blue composites of S2, S3 (Oa17, Oa8, Oa4) and adjusted S3 images for Maspalomas (Spain) on Apr 30, 2020 and Waterbank (North Australia) on Nov 24, 2020.

Finally, the final prediction is estimated using the information in the neighborhood.

3) Fit-FC: Wang and Atkinson [13] presented a linear RM (Fit-FC) to initially fuse Sentinel-2 and Sentinel-3 OLCI images, with the possibility of applying it to fuse MODIS and Landsat. Fit-FC is based on three main steps to fused coarse and fine images and requires only one pair of coarse-fine images at the base date. The first step is the fitting of the RM between the two coarse images at the base and the predicted time using a moving window to calculate the coefficients of the RM for the center coarse pixel; the models generated at this point are applied to the fine image at t_1 within the center coarse pixel. This step generates a blocky artefact problem; a spatial filtering (SF) was introduced as a second step to deal with spectrally similar pixels used to define a weighted function of each pixel in a local window. The final step Residual compensation (RC) consists of downscaling the residual coarse to match the size of the fine image using bicubic interpolation, then spectrally similar neighboring pixels are used to avoid the smoothing effect of the bicubic interpolation, and finally the updated residual is added back to predicted image generated in the SF prediction to preserve the spectral information.

D. Accuracy Assessment

We evaluated the results using three of the most widely used metrics in the literature. The root mean square error (RMSE) compares the difference in digital pixel values between the original image and the predicted image; the lower the value, the better the accuracy [34]. The structural similarity index (SSIM) is a visual assessment index used to measure spatial similarities between the original image and the predicted image; therefore, a higher value is an indication of good similarity [35]. The spectral angle mapper (SAM) measures spectral similarity by calculating the difference in spectral angle between the real and predicted images, a lower SAM value indicates a lower spectral distortion [36].

III. Results

The fusion outputs of the three STF algorithms using both the original and adjusted bands were visually and qualitatively compared to the ground-truth images (S2 images at prediction time t_2). As mentioned above, this study focused on the bands in the spectral ranges corresponding to Blue, Red, and NIR, as in these ranges, there is more than one S3 band overlapping with one S2 band (see Table I and Fig. 1). To illustrate all the results, we used a false-color composite view (NIR, red, and blue).

NIR-Red-Blue color composites using the original and adjusted bands for the two sites are shown in Fig. 4. The color composition of the original S3 bands shows significant differences in spectral characteristics compared to the ground-truth S2 image. However, it can be observed that the color compositions of the adjusted bands are closer, from a spectral point of view, to the

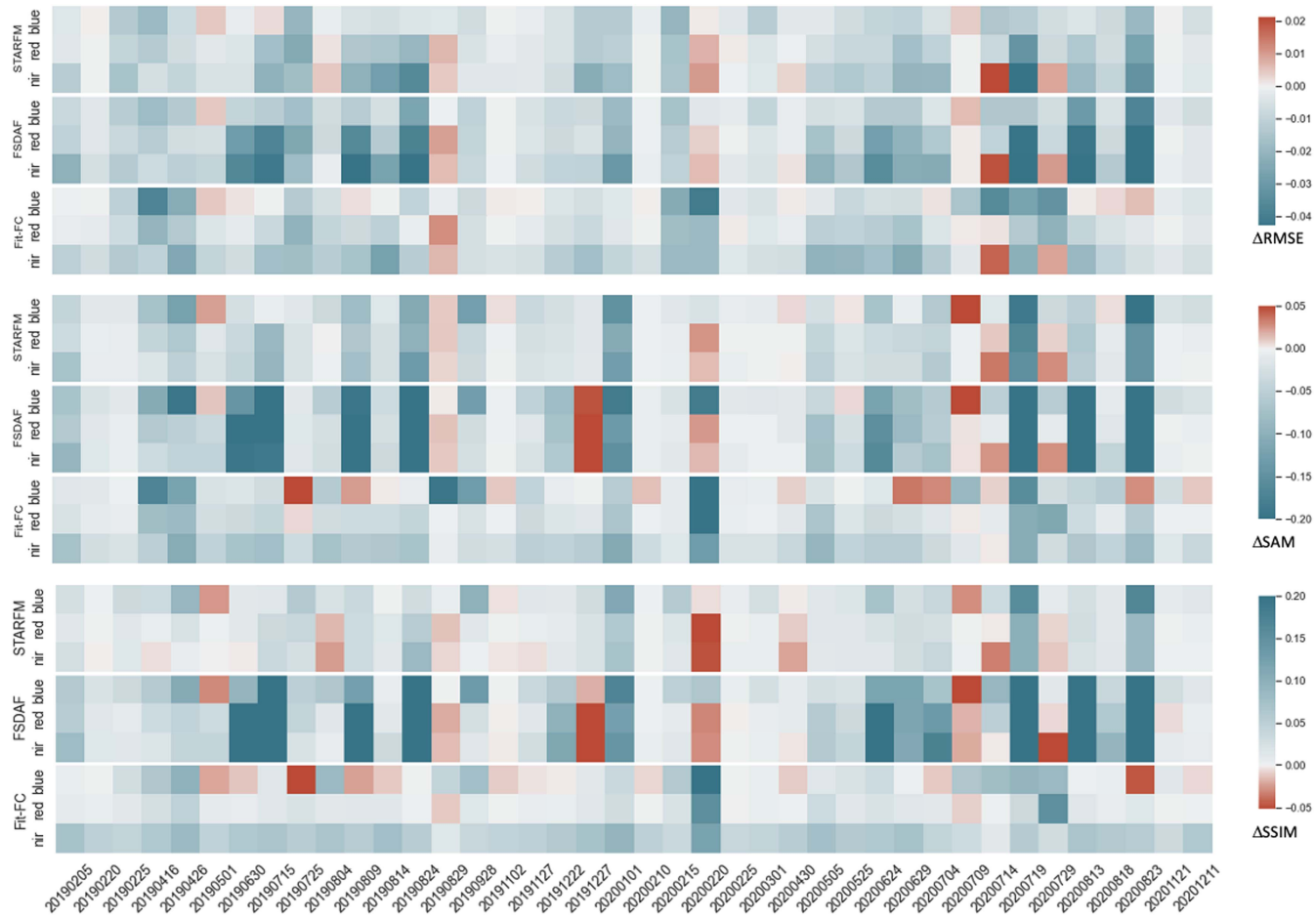


Fig. 5. Differences between the metrics obtained when evaluating the quality of fused images using as input the adjusted bands and the original bands for the Maspalomas site.

S2 images. This demonstrates that the proposed methodology was able to generate adjusted bands with spectral characteristics closer to the ground truth. The SRF-adjustment was performed to all available image pairs.

To demonstrate the impact of using adjusted bands in the fusion process on the quality of the fusion results, the outcomes (prediction t_2) generated by the three aforementioned STF methods were compared. Two sets of inputs were used for the comparison: 1) the S2 bands combined with the adjusted S3 bands, and 2) the original S2 and S3 bands. For all STF methods, the default settings proposed in the original work were used. The evaluation was conducted on all the available image pairs for the blue, red, and NIR spectral bands for each site using the metrics mentioned above.

A. *Scenario 1: Maspalomas Site*

We examined how using adjusted bands instead of the original bands impacts the performance of different STF methods. In this study, we computed the metrics (see Section II-D) by comparing the resulting fused images with the ground truth. The ground truth is an S2 image taken on the target date, denoted as t_2 in Fig. 2. We performed this analysis in two cases: first, when the input for the STF methods is an S3 band obtained using the proposed method, and second, when the input is any of the original S3 bands within the same spectral range. To visualize the impact and for the sake of simplicity, we calculated the differences (Δ) between the metric values obtained in these two cases. These differences have been visualized as heatmaps.

Fig. 5 presents the obtained heatmap for the first scenario (Maspalomas site). Since lower values of RMSE and SAM indicate better performance, negative difference values indicate that the fused images with the adjusted bands have better quality than the ones generated with the original bands. On the contrary, positive difference values indicate better performance when the original bands are used for fusion. Blue colors in the three metrics, Fig. 5, indicate that the fusion with adjusted bands produced better results than the fusion with the original bands, while red means that better results were obtained from the fusion using the original S3 bands. In most cases, the results show that using adjusted bands as input in STF methods leads to a better prediction. Thus, the use of adjusted bands leads to a decrease of around 0.1 in SAM and an increase of approximately 0.1 in SSIM. A significant improvement was recorded on 30 June 2019 and 17 July 2020, with an increase in SSIM of 0.2 and a decrease in RMSE and SAM of 0.04 and 0.2, respectively. However, on 20

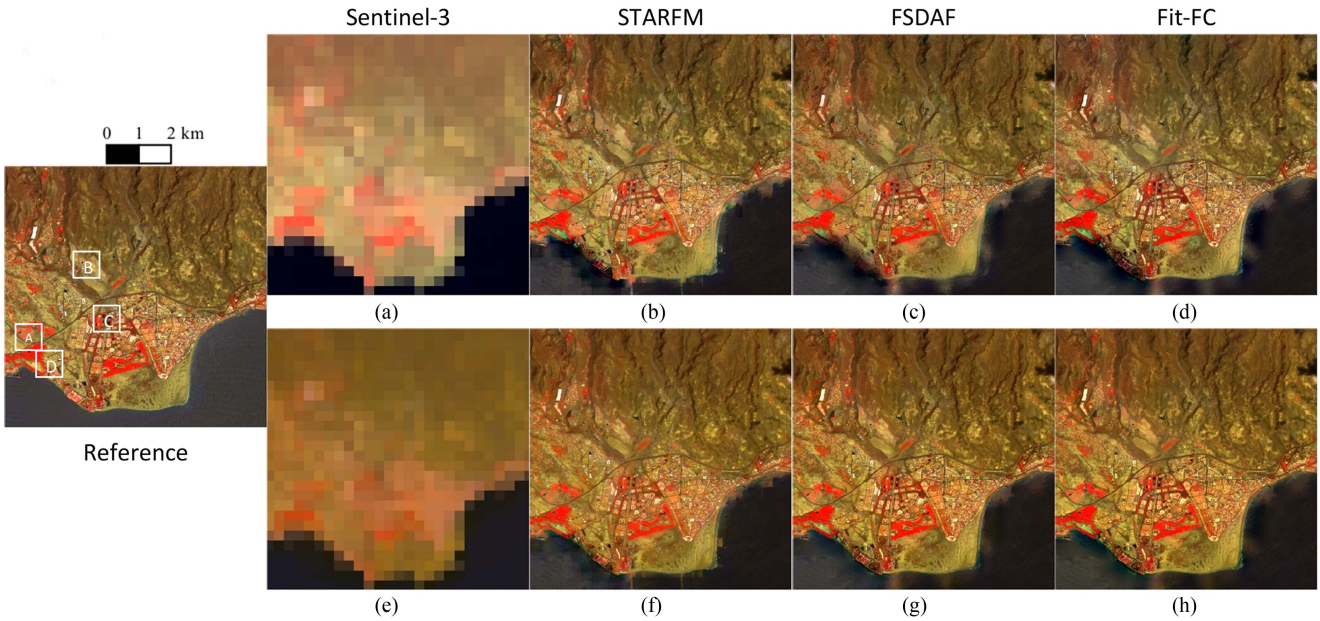


Fig. 6. Color composition of ground truth (S2), original S3 and images predicted by different STF algorithm. (a) S3 original bands; (b), (c) and (d) are the predicted images using original S3 bands by STARFM, FSDAF and Fit-FC, respectively. (e) Sentinel-3 adjusted bands; (f), (g) and (h) are the predicted images using adjusted S3 bands by STARFM, FSDAF and Fit-FC, respectively.

TABLE II
AVERAGE OF THE QUALITY METRICS FOR THE WHOLE DATASET FOR THE MASPALOMAS SITE

	STARFM			FSDAF			Fit-FC		
	RMSE↓	SSIM↑	SAM↓	RMSE↓	SSIM↑	SAM↓	RMSE↓	SSIM↑	SAM↓
Oa4	0.024	0.934	0.537	0.026	0.926	0.604	0.023	0.927	0.610
Oa5	0.025	0.932	0.551	0.026	0.925	0.621	0.023	0.923	0.633
A-blue	**0.021**	**0.945**	**0.468**	**0.022**	**0.943**	**0.475**	**0.021**	**0.940**	**0.485**
Oa8	0.028	0.922	0.694	0.030	0.915	0.767	0.030	0.902	0.840
Oa9	0.028	0.922	0.695	0.031	0.915	0.768	0.030	0.902	0.843
A-red	**0.022**	**0.939**	**0.592**	**0.023**	**0.938**	**0.605**	**0.023**	**0.920**	**0.678**
Oa16	0.031	0.922	0.723	0.033	0.916	0.786	0.035	0.899	0.910
Oa17	0.031	0.921	0.728	0.034	0.914	0.794	0.036	0.897	0.921
Oa18	0.031	0.921	0.730	0.034	0.914	0.795	0.036	0.897	0.922
A-NIR	**0.027**	**0.931**	**0.661**	**0.027**	**0.932**	**0.675**	**0.029**	**0.909**	**0.790**

A-blue = adjusted blue. A-red = adjusted red. A-NIR = adjusted NIR.

February 2020, STF methods generally generated better images using the original bands than the adjusted bands.

The results of the quantitative evaluation of the Maspalomas site are summarized in Table II. The three considered metrics have been calculated for each STF method using, as the ground truth, the S2 bands, and comparing them with the resulting fused images when the individual original S3 bands (Oa4, Oa5, Oa8, Oa9, Oa16, Oa17, and Oa18) are used. In addition, we assessed the performance using the adjusted bands (A-blue, A-red, A-NIR) that combine the information from the S3 bands overlapping with the spectral range of blue, red, and NIR in the S2 product. The fusion results of the adjusted bands yielded the best mean results for all STF methods in the three adjusted bands. The evaluation results demonstrate that the proposed adjusted bands made the STF methods predict the fused images with the lowest RMSE and SSIM and the highest SAM.

The predicted images obtained from the different fusion methods on June 30, 2019, using the original S3 bands (first row) and the adjusted S3 bands (second row) are shown in Fig. 6. The white squares correspond to the zoom-in area presented in Fig. 7. The first column in Fig. 6(a) and (e) shows both the S3 image captured on May 1st, 2019, and their adjusted band images, respectively. It can be observed that the proposed methodology was able to slightly improve the spatial quality of the fused images.

Generally speaking, despite the high scale factor (ranging from around 300 to 10) between the coarse and fine images, it can be observed that the STF methods were able to predict images with better spatial resolution than the original S3 images. However, in general, the algorithms failed to produce accurate results in the coastal area.

In Fig. 7 the fusion results for the zoom-in of the four locations mentioned above are shown. For illustrative purposes, we chose three prediction dates based on the performance of the adjusted bands compared to the performance of the original bands on the fusion (June 30, 2019; February 15 and February 20, 2020). The first row (Reference) presents the ground-truth S2 images. For each STF method, the results of the fusion using the original S3 bands are shown in the left columns (2nd, 4th, and 6th), whereas the results of the fusion using the adjusted S3 bands are displayed in the right columns (3rd, 5th, and 7th). In general, the results of the fusion with the original S3 bands are slightly poor; some details have been lost, and the original color is not preserved. However, the images generated with the adjusted bands show a good recovery of the texture and colors that are more similar to the original S2 images. It was found that fusion has considerably improved in the three STF methods when using the adjusted bands in the fusion results on June 30, 2019.

The spectral artefacts appearing in the fusion using the original bands, presented as green in zones A and C for STARFM,

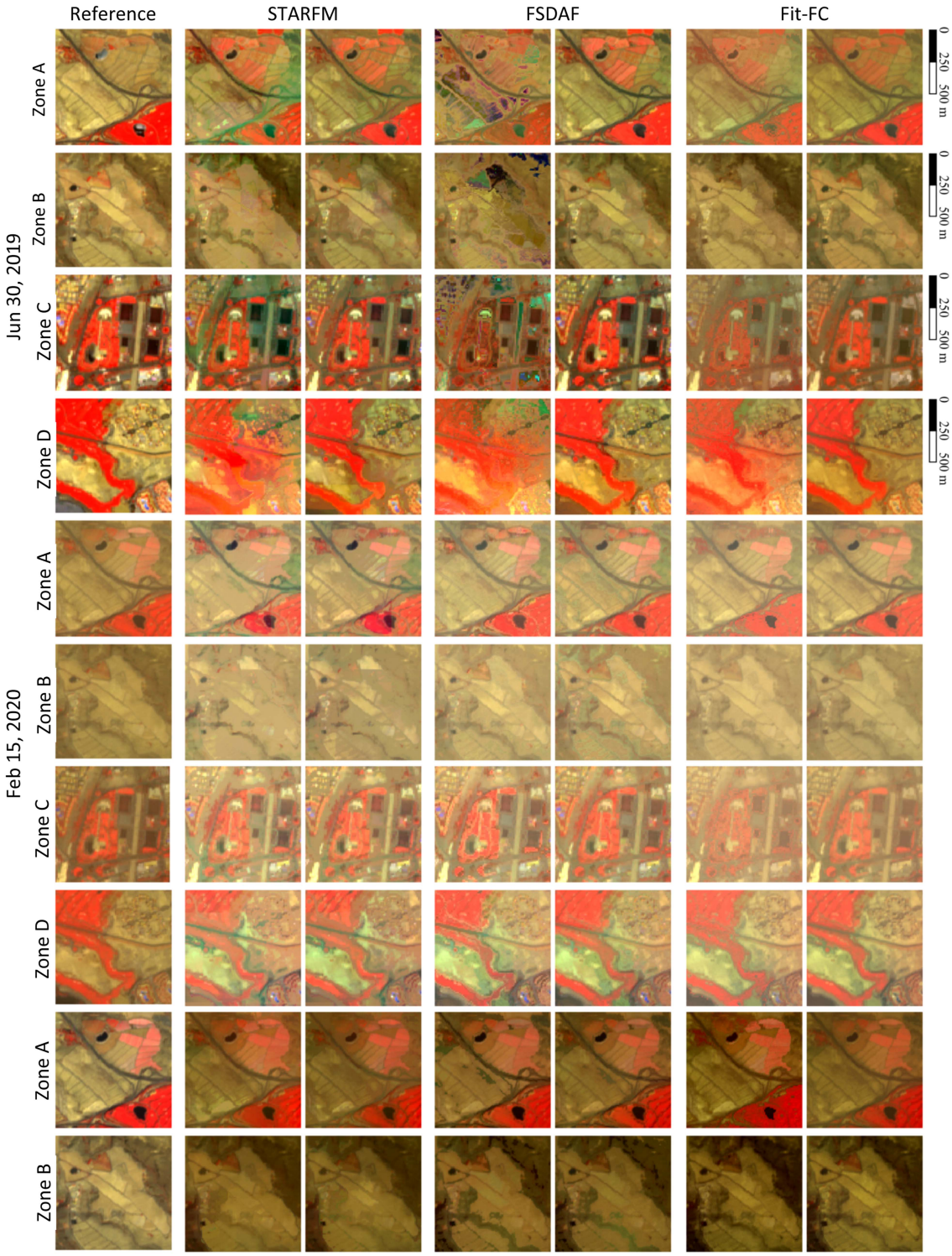


Fig. 7. Visual comparison in Maspalomas site for three prediction dates for the four zoom-in displayed in Fig. 6 (1km x 1km).

incorrect pixel values in FSDAF, and noise in Fit-FC, are corrected and disappeared in the fusion using the adjusted bands for the three dates. The blurring effect shown in zone D, generated by all STF methods, is also improved in the fusion of the adjusted bands. Although all fusion methods produce acceptable prediction results for the dates on February 15, 2020, and February 20, 2020, there are some imperfections in depicting edges and buildings. Notably, spectral noise is observed in the red region when the original S3 bands are used as input for fusion. The details are better defined, and it is easier to distinguish different land covers in the fusion of the adjusted bands. It can be observed that for the three STF methods, the spectral and spatial quality of

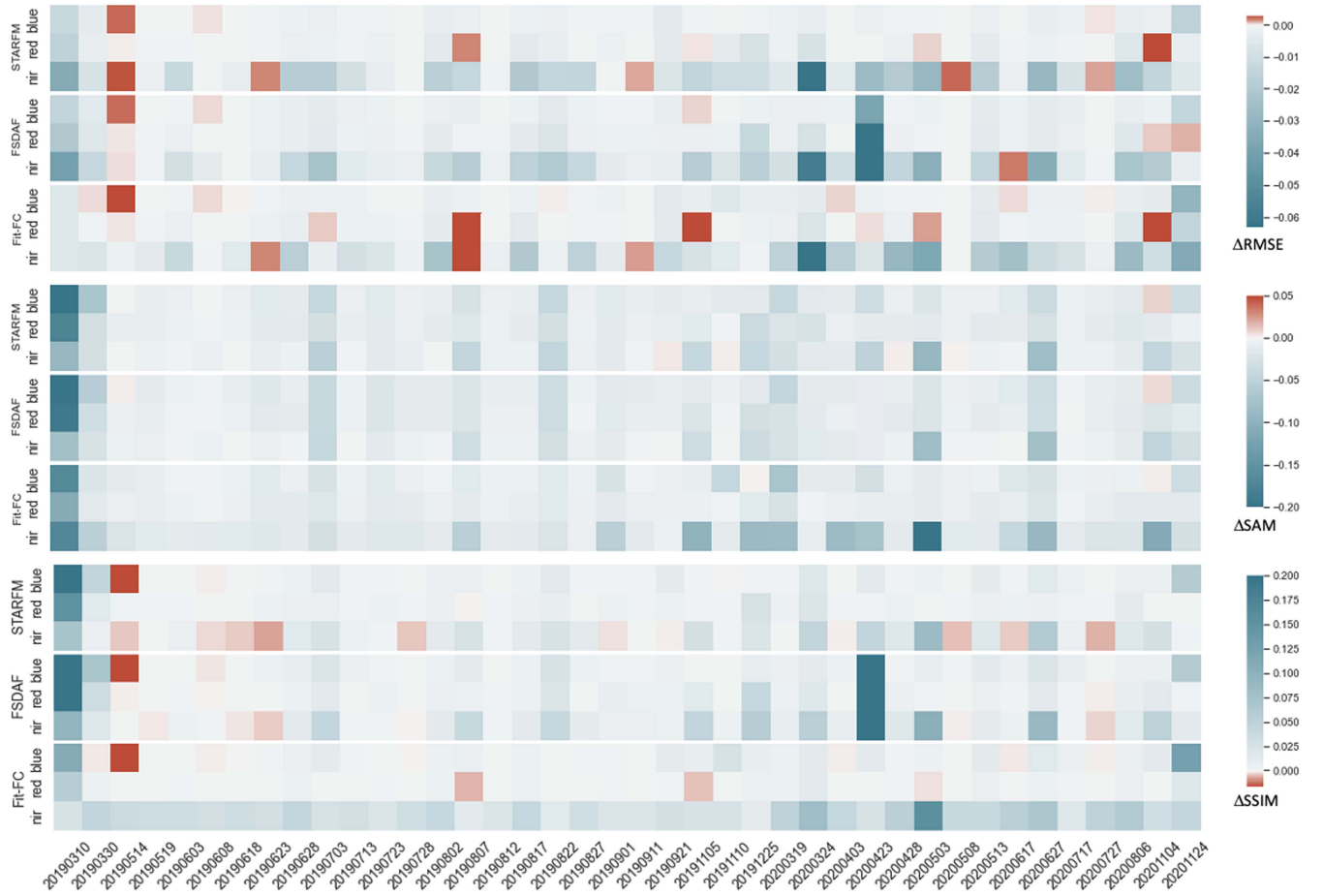


Fig. 8. Differences between the metrics obtained when evaluating the quality of fused images using as input the adjusted bands and the original bands for the Waterbank site.

the images obtained using the adjusted bands is higher than those obtained using the original bands at this site. In particular, the prediction of a heterogeneous landscape is better when adjusted bands are used in the fusion images estimated by the three methods. Thus, the definition of buildings and roads, in zone C zoomed-in area, is better when adjusted bands are used, and in general, the spatial details such as the green areas are also better preserved.

B. Scenario 2: Waterbank Site

Similarly to Fig. 5, Fig. 8 shows the difference in the metrics between two sets of images: fused images with original bands as an input and fused images with adjusted bands as input for the 40 pairs used at the Waterbank site. From the figure, we can see that the use of the adjusted bands in the fusion improved the metrics for the three STF methods in most of the cases. The SAM values for the fusion using the adjusted band was consistently lower than when the original S3 bands were used in the fusion (up to 0.2), indicating that the adjusted bands preserve better the spectral information, and therefore the fusion images are spectrally more similar to S2 at the target date t_2. SSIM and RMSE also show similar improvements, except for the blue band for the day May 14th 2019. The results of this day showed that better metrics were obtained using the original bands. The reason behind this exception could be the difference in the spectral characteristics of S3 bands between the prior and the target day due to saturation of some pixels.

TABLE III
AVERAGE OF THE QUALITY METRICS FOR THE WHOLE DATASET FOR THE WATERBANK SITE

	STARFM			FSDAF			Fit-FC		
	RMSE↓	SSIM↑	SAM↓	RMSE↓	SSIM↑	SAM↓	RMSE↓	SSIM↑	SAM↓
Oa4	0.010	0.964	0.126	0.011	0.939	0.125	0.010	0.967	0.116
Oa5	0.011	0.960	0.133	0.012	0.934	0.131	0.011	0.965	0.122
A-Blue	**0.009**	**0.973**	**0.111**	**0.009**	**0.972**	**0.110**	**0.009**	**0.974**	**0.105**
Oa8	0.018	0.942	0.124	0.021	0.920	0.120	0.018	0.953	0.112
Oa9	0.018	0.941	0.126	0.021	0.918	0.123	0.018	0.952	0.113
A-Red	**0.016**	**0.950**	**0.109**	**0.017**	**0.951**	**0.106**	**0.016**	**0.956**	**0.101**
Oa16	0.039	0.926	0.083	0.045	0.903	0.080	0.040	0.911	0.088
Oa17	0.045	0.921	0.089	0.049	0.896	0.085	0.045	0.906	0.101
Oa18	0.046	0.920	0.090	0.050	0.894	0.086	0.046	0.904	0.105
A-NIR	**0.034**	**0.934**	**0.071**	**0.035**	**0.935**	**0.070**	**0.034**	**0.943**	**0.066**

A-blue = adjusted blue. A-red = adjusted red. A-NIR = adjusted NIR.

The average results of the quantitative evaluation of the Waterbank site are presented in Table III similarly to Table II. It can be seen that the adjusted bands improve the spatial and spectral

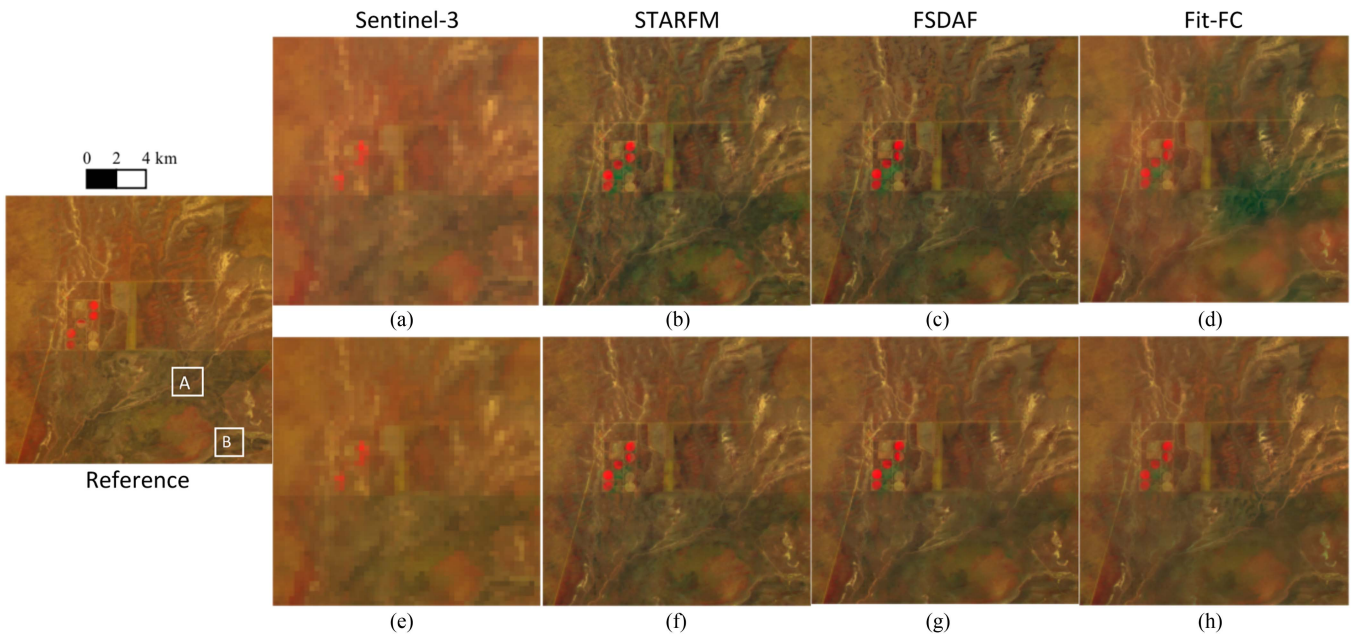


Fig. 9. Ground truth and images predicted by different STF algorithm. (a) Sentinel-3 original bands, (e) Sentinel-3 adjusted bands. (b), (c) and (d) are the predicted images using original S3 bands by STARFM, FSDAF and Fit-FC, respectively. (f), (g) and (h)are the predicted images using adjusted S3 bands by STARFM, FSDAF and Fit-FC, respectively.

accuracy of the predicted images in the three spectral bands. The three adjusted bands made the STF method predict images with lower RMSE and SAM and higher SSIM.

The results for the entire study area in the Waterbank (Australia) site are presented in Fig. 9. This site is known for the rapid changes caused by forest fires and changes in vegetation throughout the year. As in the previous scenario, the figure shows that the images generated by the adjusted bands are spectrally closer to the reference S2 than those generated by the original S3 bands. In the first row of the figure, we can see that the green color is more intense than the reference, while in the second row, the green color is similar to the reference image. A closer examination of the two set of predicted images (using the adjusted and the original bands) for the regions inside the white boxes is presented in Fig. 10. The first row (Prior) presents S2 images at the base date t_1. The second row (Target) presents the ground-truth S2 images (t_2). Similarly to Fig. 7, and for each STF method, the results of the fusion using the original S3 bands are shown in the left columns (3rd, 5th, and 7th), whereas the results of the fusion using the adjusted S3 bands are displayed in the right columns (4th, 6th, and 8th). The three dates presented in the zoomed-in experienced significant and rapid changes between the prior and target dates. A comparison between the fused images and the original bands revealed significant artefacts in the fusion that were reduced by using the adjusted bands. The predicted images using the adjusted bands were found to be closer to the reference images (see zones A and B on December 25, 2019). On 3 May 2020, the area experienced a fire between the prior and target dates, which caused the burned areas to appear in black in zone B in the Fig. 10. The prediction of this rapid change presents a challenge for most of the STF methods in the literature. The three STF methods did not predict this change and spectral artefacts were high in this area. While the fused images obtained with adjusted bands led to a slight improvement in spatial details in those zones, a more significant enhancement was observed in the spectral characteristics of the area through the reduction of green color in the burned region.

From the zoomed-in area, we can see in the bottom corners of the fusion results of each STF method using the original bands, that the predicted images by the three STF methods present poor spectral quality, and the texture information is barely visible. In the results provided by the FSDAF method, the spatial information and the contours of the road can be seen in the fusion, but the predicted Fit-FC images have almost no spectral information. Our proposed bands improved and outperformed the original bands in terms of spectral information. For example, in zone A of the Fit-FC results, our proposed bands showed a significant improvement.

In summary, the adjusted bands better preserve the spectral information, which improves the spectral quality of the fusion. The images generated with the adjusted bands are closer to the target images in terms of both spectral information and spatial texture. These results are consistent with the findings of the quantitative evaluation.

IV. Discussion

To our knowledge, there are no criteria for selecting the best narrow spectral band to perform a STF when many narrow spectral bands overlap with a wider band. In general, only one of the overlapped bands is used during this process, neglecting the spatial and spectral information contained in the rest of the bands. This may lead to increase the spectral differences between the sensors. This spectral difference is significant enough to be considered to ensure accuracy in sensor measurements [37], [38].

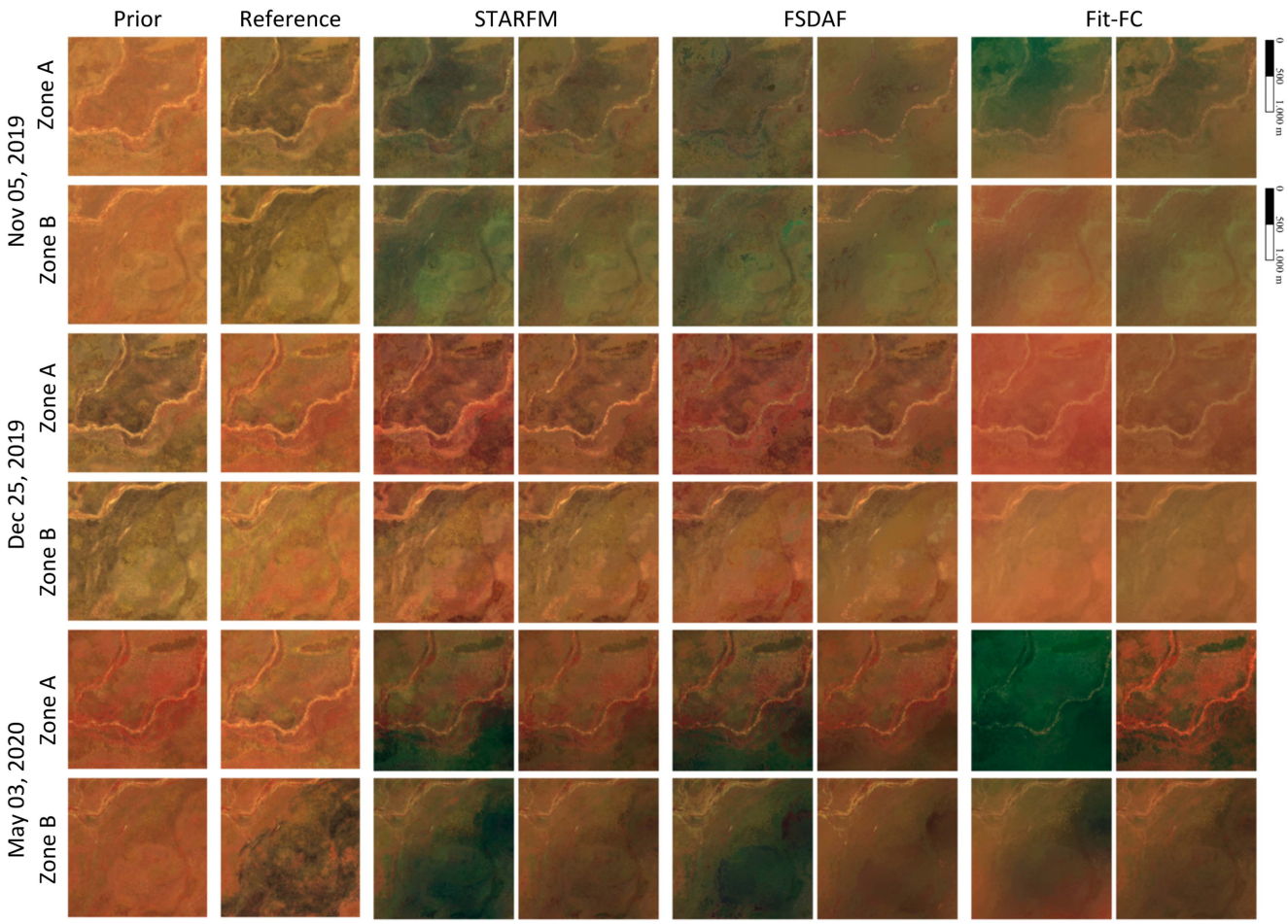


Fig. 10. Illustrative comparison in Waterbank site for three prediction dates for the two zoom-in area mentioned in fig 9 (200 x 200 S2 pixels).

Although this aspect has been addressed in pansharpening, considering the influence of SRF on the fusion process [22], to our knowledge, our study is the first to extend it to the STF case, considering more than one overlapped band. Inspired by this seminal work, a new preprocessing strategy was proposed to generate adjusted bands from multiple narrow bands that include as much spectral and spatial information as possible. The adjusted bands would be used as input for the fusion methods without the need of modifying the STF methods.

In order to assess the efficacy of the proposed SRF-adjustement strategy, different experiments were carried out, including: 1) multiple narrow bands with different bandwidths; 2) different types of STF methods; and 3) the evaluation of diverse sites that presented various challenging scenarios for the STF. All results demonstrated that the adjusted bands allow the STF methods to estimate fused images with higher quality, retain more spatial details, and reduce spectral distortion, especially in heterogeneous landscapes.

The first contribution of this work is that the results proved that the proposed methodology was capable of generating adjusted bands with spectral characteristics closer to the ground truth than the original S3 bands. The spectral characteristics in the NIR-Red-Blue compositions are closer to the ground truth, as a consequence of including information from several narrow bands through the use of band adjustment. The reason behind this difference could be explained by how sensors react and capture light from a certain wavelength. The wider spectral band of the sensor (S2) has the ability to detect a higher amount of energy reflected from Earth than the narrow spectral bands of the sensor (S3); these differences generate discrepancies between the reflectance values provided by the sensors to be compared and/or combined. Adjustments of the S3 bands to the S2 bands reduce these discrepancies by obtaining regression coefficients to establish a relationship between the S2 and S3 bands. These coefficients are then used to transfer the original S3 bands to match the S2 bands.

It is important to note that some studies in the field (e.g., [13], [39]) have preferred working with simulated S3 data over real S3 data for the STF process. Wang and Atkinson [13] used simulated S3 bands derived from S2 bands to assess the effectiveness of the Fit-FC method in reducing uncertainties introduced by the substantial disparities in spatial scale and spectral characteristics between S2 and S3 sensors, which posed a significant challenge. Only a few studies in the literature, apart from ours, have utilized actual S2-S3 data in the fusion task (e.g., [40]).

From our study, the second finding that emerges is that the SRF should be considered in the STF, including all the spectral information provided by different narrow bands to reduce the

effect of the sensors differences in the physical properties. In addition, and on the basis of the fusion results obtained, both quantitative and qualitative, it has been shown that the inclusion of more spectral information as input for the STF methods mostly mitigates not only spectral distortions but also spatial differences. Moreover, the use of only one spectral band (when multiple spectral bands exist) in the fusion despite the difference in the spectral characteristics reduces the quality of the fusion regardless of the STF methods. The results obtained in this article support the idea of the importance of spectral information in the STF process discussed in previous works [21], [41].

Third, the results of the first scenario showed a significant improvement in the fusion results using adjusted bands; these findings are consistent with previous studies that demonstrated the sensitivity of land covers to SRF [37]. The STF methods used in this study assume that both coarse and fine images capture the same surface reflectance; in other words, the spectral differences between the coarse-fine images are not considered in the fusion. This assumption may not be accurate, considering that the sensors have different designs (SRF, different bandwidth, imaging time, etc.), which makes the task of fusion even more challenging. In addition to this assumption, the STF methods used in this work have limitations in predicting small spatial details when the size of this object is smaller than the coarse pixel and a homogeneous pixel cannot be found within a window. The strategy of the moving-window used in the three STF methods sometimes leads to inaccurate estimation in a highly heterogeneous landscape, since it is assumed that the reflectance change of the coarse-resolution pixel equals the change of the fine-resolution pixels within it. However, this assumption is not always true considering the reasons mentioned earlier. Adjusted bands significantly reduce the effect of differences in SRF as shown in Fig. 4, leading to an improvement in prediction in heterogeneous landscapes. SRF-adjusted spectral bands in the fusion process provided ground truth-like images with improved identification of spatial structures, as buildings or roads, which are much better identified, independently of the heterogeneity of the land covers.

The rapid change in land cover between reference and target dates presents a challenge for most STF methods in the literature [2], [17]. The spectral changes and diversity in the phenology, especially in objects smaller than the coarse resolution pixel, make it difficult for the STF methods to predict the changes, thus sometimes generating incorrect values as shown in Fig. 10. The results obtained in the second scenario, which involved fusing images with high and rapid phenological changes, provide strong evidence supporting the utilization of additional spectral information to enhance the quality of the image fusion. Despite the fact that the use of the adjusted bands in the fusion did not provide a significant improvement of the spatial details, the spectral similarity between the fused and the target date has been improved, as shown in the prediction metrics. This spectral correction of the input bands reduces the spectral artefacts and differences, which improves the overall spectral accuracy.

The visual results have been supported by the numerical comparison of the quality of the fused images using the original S3 bands and the adjusted bands.

With a few exceptions, the metric values obtained indicate that, regardless of the spatial and spectral characteristics of the source images, the spectral ranges, and the three methods used, the substitution of the original spectral bands by the adjusted bands provides fused images much closer to the reference (S2 at the prediction date). Furthermore, when the average values of the entire data set have been obtained, the differences in the fusion indices with the original and adjusted reach an improvement of more than 37%, in particular for some metrics and STF methods. Thus, for example, for STARFM and FSDAF, the improvement of the SAM metric is 16% and 23% in the blue range, 15% and 21% in the red range, both for the Maspalomas site and 21% and 19% in the NIR for the Waterbank site. A higher improvement was recorded for the Fit-FC method with 23% for the blue range, 19% in the red range in the Maspalomas site, and 37% in the NIR range in the Waterbank site. The improvement in Fit-FC performance could be explained by its sensitivity to radiometric errors and spectral inconsistencies between the two sensors. With different percentages of improvement, these patterns repeat for the different metrics, which seems to mean that the effect of the use of the adjusted bands is not the same for different STF methods. However, what can be ensured from the analysis of the obtained results is that the use of adjusted bands provided more homogeneous results (less variability) in a whole dataset, regardless of the STF method considered for all spectral ranges and for all metrics evaluated.

Finally, it should be noted that the results obtained are susceptible to being improved, either through the formulation of the algorithm, considering either the temporal variance or the spectral difference, which need to be investigated in the future. Furthermore, in our study, we tested the effectiveness of including SRF in the fusion for S2–S3 images; we call for more studies for other sensors, since the phenomenon of band overlapping is common in most sensors. It is strongly recommended that differences in the SRF between sensors be further examined in data fusion to reduce their impact on prediction results.

V. Conclusion

In this study, we explored the potential of bands adjustment using the SRF to minimize spectral differences between sensors. To achieve this, we proposed a new preprocessing strategy based on linear regression, which combines multiple narrow bands with an overlapping with one wide band. This process generated an adjusted band that closely matches the spectral characteristics of the wide band, which was then used as input for three of most used STF methods instead of the original S3 bands. The proposed methodology was tested at two different sites, using real data from Sentinel-2 and Sentinel-3 satellites. These sites presented diverse and challenging scenarios for STF methods. Our findings demonstrated that the proposed methodology significantly enhances the spectral quality of the prediction of S2-like images and improves the overall accuracy of the STF methods. A final conclusion of the proposed preprocessing step is that the inclusion with all the spectral available information provided by multiple narrow bands can reduce the spectral and spatial distortions resulting in a remarkable improvement of the STF task.

APPENDIX A
THEORETICAL BASIS

The sensor SRF represents the probability that a given sensor will detect a photon at a given frequency. In this work, two SRFs should be considered: the SRF corresponding to the HSLT, designed as $R_h(\nu)$ and the SRF of the ith band of the LSHT, designed as $R_{l,i}(\nu)$, with $i = 1, 2, ...n$. n being the number of bands of the LSHT sensor. The probability that a photon will be detected by the HSLT sensor can be defined as the probability of the event h

$$P(h) = \int R_h(\nu)\, d\nu. \tag{A.1}$$

Similarly, the probability of the detection of a photon by the LSHT sensor for a band i can be defined as the probability of the event l_i (A.2).

$$P(l_i) = \int R_{l,i}(\nu)\, d\nu. \tag{A.2}$$

The probability of events (h) and l_i can be geometrically understood as the area below their SRF.

Taking into account n_h and $n_{l,i}$ are the number of photons detected by HSLT and LSHT$_i$. Total photons detected simultaneously by the HSLT and LSHT sensors $n_{h,l,i}$ could be defined as

$$n_{h,l,i} = P(h \,|\, l_i) \cdot n_{l,i}. \tag{A.3}$$

Thus, given the number n_h of photons detected by the HSLT sensor, we can predict the number $n'_{l,i}$ of photons that the LSHT sensor should detect [22]. In term of image fusion, (A.4) provides the the number of HSLT photons, which contain spatial details

$$n'_{l,i} = \frac{P(l_i \,|\, h)}{P(h \,|\, l_i)} \cdot n_h. \tag{A.4}$$

The relationship between the digital value and the TOA reflectance can be defined as follows [42]:

$$DN = \frac{\rho \cdot E_s \cdot \cos\theta}{\text{Gain} \cdot d^2 \cdot \pi} \tag{A.5}$$

where ρ is the reflectance of the top of the atmosphere (TOA), E_s is the solar irradiance at TOA, θ is the solar zenith angle, (Gain) is the band-specific rescaling gain factor and (d) is the solar-earth astronomical distance.

The spectral irradiance in the sensor aperture is an average of the solar irradiance over the SRF range. The DN value can be approximated by integrating the solar radiance weighted by the product of SRF and TOA reflectance [24], [43]

$$L_t = \frac{\cos\theta_{l_t}}{\pi \cdot d^2 \cdot \text{Gain}_{l_t}} \int_{\lambda_{\min}}^{\lambda_{\max}} R_l(\lambda)\rho_{l_t}(\lambda)E_t(\lambda)d\lambda \tag{A.6}$$

where L_t is the DN value of the LSHT sensor on the time t, $R_l(\lambda)$ the spectral response of LSHT band at wavelength λ, ρ_{l_t} is the TOA surface reflectance at λ on t and $E_t(\lambda)$ is the TOA solar irradiance on t at λ. Similarly, the DN value of the HSLT sensor at t

$$H_t = \frac{\cos\theta_{h_t}}{\pi \cdot d^2 \cdot \text{Gain}_{h_t}} \int_{\lambda_{\min}}^{\lambda_{\max}} R_h(\lambda)\rho_{h_t}(\lambda)E_t(\lambda)d\lambda. \tag{A.7}$$

Taking into account the response function of the LSHT sensor between $\lambda_{\min}$ and $\lambda_{\max}$ in the case where there is no overlap between bands of the same sensor, R_l is a linear function between $\lambda_{\min}$ and $\lambda_{\max}$ and could be written as follows:

$$R_l = \sum_{i=1}^{m} R_l^i \tag{A.8}$$

where R_l^i is the sensor response corresponding to band i, with $i = 1, 2, ..., m$, m being the number of LSHT bands in the interval $[\lambda_{\min}, \lambda_{\max}]$.

Using (A.8), (A.6) could be written as

$$L_t = \frac{\cos\theta_{l_t}}{\pi \cdot d^2 \cdot \text{Gain}_{l_t}} \sum_{i=1}^{m} \int_{\lambda_{\min}}^{\lambda_{\max}} R_l^i(\lambda)\rho_{l_t}(\lambda)E_t(\lambda)d\lambda. \tag{A.9}$$

Equations (A.7) and (A.9) are converted to the equivalent matrix form to simplify the calculation

$$H_t = \frac{\cos\theta_{h_t}}{\pi \cdot d^2 \cdot \text{Gain}_{h_t}} R_h \rho_{h_t} E_t \tag{A.10}$$

$$L_t = \frac{\cos\theta_{l_t}}{\pi \cdot d^2 \cdot \text{Gain}_{l_t}} \sum_{i=1}^{m} R_l^i \rho_{l_t} E_t. \tag{A.11}$$

Note that

$$L_t^i = \frac{\cos\theta_{l_t}}{\pi \cdot d^2 \cdot \text{Gain}_{l_t}} R_l^i \rho_{l_t} E_t. \tag{A.12}$$

L_t^i is the DN value of the LSHT sensor at t that corresponds to the band i that overlap with the same HSLT sensor band. Similarly, we can obtain the DN value of both sensors for both coarse and fine bands at t_1 corresponding to the base date and t_2 corresponding to the prediction date

$$H_{t_1} = \frac{\cos\theta_{h_{t_1}}}{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_1}}} R_h \rho_{h_{t_1}} E_{t_1} \tag{A.13}$$

$$H_{t_2} = \frac{\cos\theta_{h_{t_2}}}{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_2}}} R_h \rho_{h_{t_2}} E_{t_2} \tag{A.14}$$

$$L_{t_1}^i = \frac{\cos\theta_{l_{t_1}}}{\pi \cdot d^2 \cdot \text{Gain}_{l_{t_1}}} R_l^i \rho_{l_{t_1}} E_{t_1} \tag{A.15}$$

$$L_{t_2}^i = \frac{\cos\theta_{l_{t_2}}}{\pi \cdot d^2 \cdot \text{Gain}_{l_{t_2}}} R_l^i \rho_{l_{t_2}} E_{t_2}. \tag{A.16}$$

In most STF methods, a band of each sensor is fused at a base time t_1, therefore, $L_{t_1}^i$ would be used in fusion. Calculating the temporal changes from t_1 to t_2 for coarse sensors for the two bands and the fine sensor band

$$H_{t_2} - H_{t_1} = \frac{\cos\theta_{h_{t_1}}}{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_1}}} (R_h \rho_{h_{t_2}} E_{t_2} - R_h \rho_{h_{t_1}} E_{t_1}) \tag{A.17}$$

$$L_{t_2}^i - L_{t_1}^i = \frac{\cos\theta_{l_{t_2}}}{\pi \cdot d^2 \cdot \text{Gain}_{l_{t_2}}} (R_l^i \rho_{l_{t_2}} E_{t_2} - R_l^i \rho_{l_{t_1}} E_{t_1}) \tag{A.18}$$

For a given pixel (x, y), we assume that the difference in reflectance between a coarse-resolution pixel and a fine-resolution

pixel is affected by system errors between the different sensors. Thus, the relationship between the coarse and the fine resolution reflectances for a given date is a linear function [44] as

$$F(x, y, t_1) = a_1 * C(x, y, t_1) + b_1 \tag{A.19}$$

where F and C are the reflectances of the fine and coarse resolution sensors of the pixel (x, y) on the date $t1$ and a and b are the coefficients of the linear model. Assuming that the system errors and the land cover do not change from the date t_1 to the date t_2, the model will have the same parameters: $a_1 = a_2 = a$. Similarly to (A.19) and for the date t_2 with the assumption that b_1 is small, we can obtain

$$F(x, y, t_2) - F(x, y, t_1) = a * (C(x, y, t_2) - C(x, y, t_1)) \tag{A.20}$$

and from (A.19) we get

$$\frac{F(x, y, t_2)}{C(x, y, t_2)} = \frac{F(x, y, t_1)}{C(x, y, t_1)} = a. \tag{A.21}$$

Equation (A.21) indicates that there is no change in the land cover over time in the relationship between F and C in the location (x, y). As a result, it is assumed that the ratio of TOA reflectance remains constant over time at various wavelengths since the SRF is constant over time [22]. Similarly, we obtain

$$\frac{\rho_{h_{t_2}}}{\rho_{l_{t_2}}} = \frac{\rho_{h_{t_1}}}{\rho_{l_{t_1}}} = \frac{\rho_h}{\rho_l} \cdot I. \tag{A.22}$$

From (A.17) and (A.22) we can obtain

$$\rho_{l_1} \cdot R_h^{-1} \frac{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_1}}}{\cos\theta_{h_{t_1}}} (H_{t_2} - H_{t_1}) = \rho_{h_1} (\rho_{l_2} E_{t_2} - \rho_{l_1} E_{t_1}) \tag{A.23}$$

$$\rho_{l_1} \cdot R_h^{-1} \frac{\pi \cdot d^2 \cdot \text{Gain}_{h_{t_1}}}{\cos\theta_{h_{t_1}}} (H_{t_2} - H_{t_1}) = \rho_{h_1} (R_l^i)^{-1} \frac{\pi \cdot d^2 \cdot \text{Gain}_{l_{t_1}}}{\cos\theta_{l_{t_1}}} (L^i_{t_2} - L^i_{t_1}). \tag{A.24}$$

Using the equations above, given a coarse-fine pair of images at date t_1 and a coarse image at date t_2, H_{t_2} can be calculated as

$$H_{t_2} = H_{t_1} + \alpha^i \left(L^i_{t_2} - L^i_{t_1}\right) \tag{A.25}$$

and

$$m.H_{t_2} = m.H_{t_1} + \left(\alpha^1 \left(L^1_{t_2} - L^1_{t_1}\right) + \cdots + \alpha^m (L^m_{t_2} - L^m_{t_1})\right) \tag{A.26}$$

we obtain

$$H_{t_2} = H_{t_1} + \frac{1}{m} \sum_{i=1}^{m} \alpha^i (L^i_{t_2} - L^i_{t_1}). \tag{A.27}$$

With

$$\alpha^i = R_h \left(\frac{\rho_h}{\rho_l} I\right) \left(R_l^i\right)^{-1} \frac{\text{Gain}_{l_{t_1}} \cdot \cos\theta_{h_{t_1}}}{\text{Gain}_{h_{t_1}} \cdot \cos\theta_{l_{t_1}}} \tag{A.28}$$

The coefficients α^i are easy to compute, most SRFs are available online in the literature, and the TOA reflectance of the HSLT and LSHT sensors at date t_1 could be calculated from the DN value using (A.5).

Similar equations to (A.25) have been used in the literature with different meanings. Thus, the coefficients α^i correspond to the temporal variance from date t_1 to date t_2 in [45]; in [44], they correspond to the error systems; in [46] they indicate the spatial variance between the coarse and fine resolution sensors; in [47] they represent the high-pass modulation between the two sparsely represented LSHT; while in [48] they define the coefficients corresponding to the spectral difference of the sensors. In this work, these coefficients describe the contribution of the SRFs to the adjustment of the spectral bands.

The relationship between radiance Top-Of-Atmosphere and Bottom-Of-Atmosphere could be simplified as [49], [50]

$$\text{TOA}_\lambda = \text{BOA}_\lambda \tau_\lambda + I^\uparrow_\lambda. \tag{A.29}$$

TOA_λ is the TOA spectral radiance, $I^\uparrow_\lambda$ is the spectral atmospheric upwelling radiance; τ_λ is the atmospheric transmittance. For HSLT and LSHT sensors we can obtain

$$L_{\text{BOA}}.\tau = L_{\text{TOA}} - I^\uparrow \tag{A.30}$$

$$H_{\text{BOA}}.\tau = H_{\text{TOA}} - I^\uparrow \tag{A.31}$$

We have then

$$\frac{H_{\text{BOA}}}{L_{\text{BOA}}} = \frac{H_{\text{TOA}} - I^\uparrow}{L_{\text{TOA}} - I^\uparrow}. \tag{A.32}$$

Under optimal atmospheric conditions and across various land cover types with minimal presence of aerosols or haze, $I^\uparrow$ is smaller than TOA and $I^\uparrow << H_{\text{TOA}} \& L_{\text{TOA}}$ and we have [51] :

$$\frac{H_{\text{BOA}}}{L_{\text{BOA}}} = \frac{H_{\text{TOA}}}{L_{\text{TOA}}} \tag{A.33}$$

This is similar to (A.21), which indicates that the linear band adjustment is also valid at the Bottom of Atmosphere.

ACKNOWLEDGMENT

Meryeme Boumahdi would like to thanks GMV and the Fundación Mujeres por Africa for their support for her doctoral thesis under the Learn Africa program.

REFERENCES

[1] M. Liu, Y. Ke, Q. Yin, X. Chen, and J. Im, "Comparison of five spatio-temporal satellite image fusion models over landscapes with various spatial heterogeneity and temporal variation," *Remote Sens.*, vol. 11, no. 22, 2019, Art. no. 2612.

[2] X. Zhu, F. Cai, J. Tian, and T.K.-A. Williams, "Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 527.

[3] Z. Shao, J. Cai, P. Fu, L. Hu, and T. Liu, "Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product," *Remote Sens. Environ.*, vol. 235, 2019, Art. no. 111425.

[4] F. Gao et al., "Fusing landsat and MODIS data for vegetation monitoring," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 47–60, Sep. 2015.

[5] A. Poortinga et al., "Mapping plantations in Myanmar by fusing Landsat-8, Sentinel-2 and Sentinel-1 data along with systematic error quantification," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 831.

[6] G. Kaplan and U. Avdan, "Sentinel-1 and Sentinel-2 data fusion for wetlands mapping: Balikdami, Turkey," *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 729–734, 2018.

[7] Z. Tan, M. Gao, X. Li, and L. Jiang, "A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5601413.

[8] Q. Liu, X. Meng, X. Li, and F. Shao, "Detail injection-based spatiotemporal fusion for remote sensing images with land cover changes," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5401514.

[9] X. Meng, Q. Liu, F. Shao, and S. Li, "Spatio–temporal–spectral collaborative learning for spatio–temporal fusion with land cover changes," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5704116. [Online]. Available: https://api.semanticscholar.org/CorpusID:249966508

[10] G. Y. C. Yang et al., "MSFusion: Multistage for remote sensing image spatiotemporal fusion based on texture transformer and convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4653–4666, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:249332237

[11] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.

[12] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sens. Environ.*, vol. 172, pp. 165–177, 2016.

[13] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, 2018.

[14] C. Xu, X. Du, Z. Yan, J. Zhu, S. Xu, and X. Fan, "VSDF: A variation-based spatiotemporal data fusion method," *Remote Sens. Environ.*, vol. 283, 2022, Art. no. 113309.

[15] M. Belgiu and A. Stein, "Spatiotemporal image fusion in remote sensing," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 818.

[16] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.

[17] P. Ghamisi et al., "Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 1, pp. 6–39, Mar. 2019.

[18] C. Li, L. Liu, J. Wang, C. Zhao, and R. Wang, "Comparison of two methods of the fusion of remote sensing images with fidelity of spectral information," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2004, pp. 2561–2564.

[19] E. Ibarrola-Ulzurrun, C. Gonzalo-Martin, J. Marcello-Ruiz, A. Garcia-Pedrero, and D. Rodriguez-Esparragon, "Fusion of high resolution multispectral imagery in vulnerable coastal and land ecosystems," *Sensors*, vol. 17, no. 2, 2017, Art. no. 228.

[20] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.

[21] S. Liu, J. Zhou, Y. Qiu, J. Chen, X. Zhu, and H. Chen, "The first model: Spatiotemporal fusion incorrporting spectral autocorrelation," *Remote Sens. Environ.*, vol. 279, 2022, Art. no. 113111.

[22] X. Otazu, M. González-Audícana, O. Fors, and J. Núñez, "Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005.

[23] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS + Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.

[24] Z. Cao, S. Chen, F. Gao, and X. Li, "Improving phenological monitoring of winter wheat by considering sensor spectral response in spatiotemporal image fusion," *Phys. Chem. Earth Parts A/B/C*, vol. 116, 2020, Art. no. 102859.

[25] J. Li, Y. Li, R. Cai, L. He, J. Chen, and A. Plaza, "Enhanced spatiotemporal fusion via MODIS-like images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5610517.

[26] G. Chander et al., "Applications of spectral band adjustment factors (SBAF) for cross-calibration," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1267–1281, Mar. 2013.

[27] G. Chander et al., "Use of EO-1 hyperion data to calculate spectral band adjustment factors (SBAF) between the l7 ETM+ and terra MODIS sensors," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 1667–1670.

[28] J. Zhou, Y. Qiu, J. Chen, and X. Chen, "A geometric misregistration resistant data fusion approach for adding red-edge (RE) and short-wave infrared (SWIR) bands to high spatial resolution imagery," *Sci. Remote Sens.*, vol. 4, 2021, Art. no. 100033.

[29] Q. Wang, W. Shi, P. M. Atkinson, and Y. Zhao, "Downscaling MODIS images with area-to-point regression kriging," *Remote Sens. Environ.*, vol. 166, pp. 191–204, 2015.

[30] L. D. Keukelaere et al., "Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm: Validation for coastal and inland waters," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 525–542, 2018.

[31] E. Wolters et al., "iCOR atmospheric correction on Sentinel-3/OLCI over land: Intercomparison with AERONET, RadCalNet, and SYN level-2," *Remote Sens.*, vol. 13, no. 4, 2021, Art. no. 654. [Online]. Available: https://www.mdpi.com/2072-4292/13/4/654

[32] A. Berk et al., "MODTRAN5: 2006 update," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII*. Bellingham, WA, USA: SPIE, 2006, pp. 508–515.

[33] D. Scheffler, A. Hollstein, H. Diedrich, K. Segl, and P. Hostert, "AROSICS: An automated and robust open-source image co-registration software for multi-sensor satellite data," *Remote Sens.*, vol. 9, no. 7, 2017, Art. no. 676.

[34] J. Zhou, C. Kwan, and B. Budavari, "Hyperspectral image super-resolution: A hybrid color mapping approach," *J. Appl. Remote Sens.*, vol. 10, no. 3, 2016, Art. no. 035024.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[36] F. A. Kruse et al., "The spectral image processing system (SIPS)—Interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ.*, vol. 44, no. 2/3, pp. 145–163, 1993.

[37] A. P. Trishchenko, J. Cihlar, and Z. Li, "Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors," *Remote Sens. Environ.*, vol. 81, no. 1, pp. 1–18, 2002.

[38] X. Fan and Y. Liu, "A global study of NDVI difference among moderate-resolution satellite sensors," *ISPRS J. Photogrammetry Remote Sens.*, vol. 121, pp. 177–191, 2016.

[39] N. Mileva, S. Mecklenburg, and F. Gascon, "New tool for spatio-temporal image fusion in remote sensing: A case study approach using Sentinel-2 and Sentinel-3 data," in *Image and Signal Processing for Remote Sensing XXIV*. Bellingham, WA, USA: SPIE, 2018, pp. 198–208.

[40] Y. Tang, Q. Wang, X. Tong, and P. M. Atkinson, "Integrating spatio-temporal-spectral information for downscaling Sentinel-3 OLCI images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 180, pp. 130–150, 2021.

[41] H. Astola, T. Häme, L. Sirro, M. Molinier, and J. Kilpi, "Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region," *Remote Sens. Environ.*, vol. 223, pp. 257–273, 2019.

[42] G. Chander, B. L. Markham, and D. L. Helder, "Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM, and EO-1 ALI sensors," *Remote Sens. Environ.*, vol. 113, no. 5, pp. 893–903, 2009.

[43] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan. 2011.

[44] J. Wang and B. Huang, "A rigorously-weighted spatiotemporal fusion model with uncertainty analysis," *Remote Sens.*, vol. 9, no. 10, 2017, Art. no. 990.

[45] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, 2010.

[46] Q. Weng, P. Fu, and F. Gao, "Generating daily land surface temperature at landsat resolution by fusing landsat and modis data," *Remote Sens. Environ.*, vol. 145, pp. 55–67, 2014.

[47] H. Song and B. Huang, "Spatiotemporal satellite image fusion through one-pair image learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 1883–1896, Apr. 2013.

[48] H. Shen, P. Wu, Y. Liu, T. Ai, Y. Wang, and X. Liu, "A spatial and temporal reflectance fusion model considering sensor observation differences," *Int. J. Remote Sens.*, vol. 34, no. 12, pp. 4367–4383, 2013.

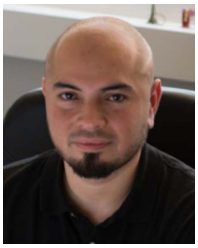
[49] J. R. J. R. Schott, *Remote Sensing: The Image Chain Approach*, 1st ed. New York, NY, USA: Oxford Univ. Press, 1997.

[50] S. Zhou and J. Cheng, "A new bottom-of-atmosphere (BOA) radiance-based hybrid method for estimating clear-sky surface longwave upwelling radiation from MODIS data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5001825.

[51] M. Gilabert, C. Conese, and F. Maselli, "An atmospheric correction method for the automatic retrieval of surface reflectances from TM images," *Int. J. Remote Sens.*, vol. 15, no. 10, pp. 2065–2086, 1994.

Meryeme Boumahdi received the bachelor's degree in mathematics and computer science with the University of Soultan Moulay Sliman, Morocco, in 2015, the master's degree in system information and networking from University Abdelmalek Essaadi, Tangier, Morocco, in 2017. She is currently working toward the Ph.D. degree in software, systems and computing with the School of Computer Engineering, Universidad Politecnica de Madrid, Spain.

Her research interests include image processing and artificial intelligence applied to Earth Observation.



Angel García-Pedrero received degree in computer science, the M.Sc. degree in computer science, and the Ph.D. in advanced computing for science and engineering.

He is currently an Assistant Professor with the Department of Computer Systems Architecture and Technology, Universidad Politécnica de Madrid and a Researcher with Center for Biomedical Technology. His main research interests include data analysis, image understanding, image interpretation (remotely sensed and medical images) and deep learning. More specifically, he examines automatic techniques based on artificial intelligence for image analysis. His research experience includes the publication of more than 60 scientific papers (30 JCR-indexed journal articles and 2 non-indexed journal articles; 32 conference papers; 2 book chapters).

Mario Lillo-Saavedra born in Chile. He received the B.E. degree in electrical engineering from the Universidad de Concepción, Chillán, Chile and the Ph.D. degree in computer sciences from the Universidad Politécnica de Madrid, Madrid, Spain.

Since 2000, he has held the esteemed position of Full Professor within the Faculty of Agricultural Engineering, University of Concepción. His scholarly pursuits pivot around remote sensing applied to agricultural water management, in addition to machine learning and data analytics. His role as a valued member of the Water Research Center for Agriculture and Mining (CHRIAM) underscores his dedication to propelling advancements in water-related research.

Consuelo Gonzalo-Martin received the B.A. degree from the Salamanca University and the Ph.D. degree from the Complutense University of Madrid, Madrid, Spain, both in physics, in 1986 and 1989, respectively.

She is currently a Full Professor with the Department of Architecture and Technology of Computers in the Computer School (Universidad Politécnica de Madrid). In particular, she has worked in the development of different algorithms for optical image fusion and ART and SOM artificial neural networks. Since 2012, she is member of the MIDAS (Data Mining and Simulation) research group at the Center for Biomedical Technology of the UPM, where she is involved in research and development projects for text and image mining in the health care domain. As a result of her research, she has led more than 20 research projects and she has participated in 65. With 46 publications indexed in the JCR (21 Q1, 17 Q2, 5 Q3, and 3 Q4), another 28 indexed in Scopus, and SJR, 5 book chapters. She was award with the mention of Distinguished Visit from the University of Concepción (Chile) in 2015. Her main research areas are image processing, based on pixels and objects, and artificial neural networks for application in remote sensing, medical images and face recognition.