Remote Sensing of Environment 295 (2023) 113709

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Unpaired spatio-temporal fusion of image patches (*USTFIP*) from cloud covered images

Harkaitz Goyena[a,b], Unai Pérez-Goya[a,b], Manuel Montesino-SanMartin[a,b], Ana F. Militino[a,b], Qunming Wang[c], Peter M. Atkinson[d,e], M. Dolores Ugarte[a,b,*]

[a] *Department of Statistics, Computer Science and Mathematics, Public University of Navarre, Pamplona, Spain*
[b] *InaMat^2 Institute, Pamplona, Spain*
[c] *College of Surveying and Geo-Informatics, Tongji University, Shanghai, China*
[d] *Lancaster Environment Centre, Lancaster University, Bailrigg, Lancaster LA1 4YW, UK*
[e] *Geography and Environmental Science, University of Southampton, Highfield, Southampton SO17 1BJ, UK*

ARTICLE INFO

Edited by Marie Weiss

Keywords:
Clouds
Fit-FC
Parallel computing
Satellite imagery
Spatio-temporal image fusion

ABSTRACT

Spatio-temporal image fusion aims to increase the frequency and resolution of multispectral satellite sensor images in a cost-effective manner. However, practical constraints on input data requirements and computational cost prevent a wider adoption of these methods in real case-studies. We propose an ensemble of strategies to eliminate the need for cloud-free matching pairs of satellite sensor images. The new methodology called Unpaired Spatio-Temporal Fusion of Image Patches (*USTFIP*) is tested in situations where classical requirements are progressively difficult to meet. Overall, the study shows that *USTFIP* reduces the root mean square error by 2-to-13% relative to the state-of-the-art *Fit-FC* fusion method, due to an efficient use of the available information. Implementation of *USTFIP* through parallel computing saves up to 40% of the computational time required for *Fit-FC*.

1. Introduction

Satellite sensor images provide valuable information in many research areas such as ecology (Pettorelli et al., 2018), agriculture (Weiss et al., 2020), urban studies (Zhao and Wentz, 2020), and economics (Donaldson and Storeygard, 2016). Many of these studies rely on data from publicly available sources like the Landsat (Arvidson et al., 2006; Roy et al., 2014), MODIS (ORNL DAAC, 2017) and Sentinel programs (Drusch et al., 2012; Donlon et al., 2012). The constellations of satellites from each program need to compromise between swath-width and revisiting time leading to trade-offs between spatial and temporal resolutions. For example, the optical bands from Landsat-8 and MODIS have an approximate spatial resolution of 30 m and 500 m, and their revisit frequencies are 16-days and daily, respectively (Roy et al., 2014; ORNL DAAC, 2017). Similarly, Sentinel-2 and Sentinel-3 scan the Earth's surface every 10 and 2 days at 10 m or 20 m and 300 m resolution, respectively (Drusch et al., 2012; Donlon et al., 2012). Many applications can benefit from denser spatial and temporal information to support more detailed analyses (Fritz et al., 2015). A cost-effective manner to achieve this goal is by blending images from complementary programs through spatio-temporal image fusion (*STIF*) methods. *STIF* techniques obtain finer spatial resolution images for a target date based on the coarse-resolution counterpart and fine-coarse image pairs from other dates (Ghamisi et al., 2019).

The high demand for more detailed images and the complexity of the problem being addressed by *STIF* has resulted into a myriad of methods. *STIF* techniques are generally grouped into three major categories depending on the principles applied during the fusion process (Chen et al., 2015; Belgiu and Stein, 2019): reconstruction-based, unmixing-based, and learning-based methods. Reconstruction-based techniques predict fine spatial resolution pixels from a weighted sum of similar pixels around spatio-temporal neighborhoods. One of the first reconstruction-based methods (Belgiu and Stein, 2019) is the spatial and temporal adaptive reflectance fusion model (*STARFM*) (Gao et al., 2006). Over the years, this method has been updated to better deal with spatial heterogeneity (Zhu et al., 2010) and abrupt temporal changes (Zhao et al., 2018). *STARFM*-like methods are very popular and they have been used in crop productivity assessments (Dong et al., 2016), forest monitoring (Walker et al., 2012) and land cover

[*] Corresponding author at: Department of Statistics, Computer Science and Mathematics, Public University of Navarre, Pamplona, Spain.
E-mail address: lola@unavarra.es (M.D. Ugarte).

https://doi.org/10.1016/j.rse.2023.113709
Received 23 January 2023; Received in revised form 27 June 2023; Accepted 2 July 2023
Available online 20 July 2023

mapping (Senf et al., 2015). Unmixing-based methods assume that a coarse resolution pixel is a linear combination of spectral end-members and their abundances. The spatial and temporal data fusion algorithm (*STDFA*) (Wu et al., 2012) is one of the most widespread unmixing-based methods. The *STDFA* was improved recently to better handle inconsistencies between sensors (Wu et al., 2016). *STDFA*-like methods have been used in practice to generate more frequent and spatially denser data products like the normalized difference vegetation index (*NDVI*) (Wu et al., 2018), leaf area index (*LAI*) (Wu et al., 2015b), and land surface temperature (Wu et al., 2015a). Finally, learning-based techniques capture the spatio-temporal patterns between fine and coarse images through empirical relationships. Learning-based methods are increasingly gaining attention through application of artificial neural networks, including convolutional neural networks (Song et al., 2018; Tan et al., 2018), and support vector regression (Moosavi et al., 2015). Some of these methods have been used recently, achieving promising results (Song et al., 2022; Chen et al., 2021). Additionally, hybrid methods combine theories and techniques from several categories. In this category, the *Fit-FC* (Wang and Atkinson, 2018) and the flexible spatiotemporal data fusion (*FSDAF*) (Zhu et al., 2016) are two of the most robust techniques (Liu et al., 2019a; Zhou et al., 2021). In recent years, *FSDAF* has been improved to better account for temporal shifts (Guo et al., 2020) and spatially complex areas (Li et al., 2020a). The *FSDAF* was applied for fire detection (Borini Alves et al., 2018), crop yield estimation (Meng et al., 2018) and leaf area index derivation (Zhai et al., 2020).

Despite the above-mentioned applications, there are operational barriers for the widespread adoption of *STIF* methods in real-case studies. The most relevant limitations are the strict input data requirements and the long processing times. The use of cloud-free images is a common requirement amongst *STIF* methods. However, depending on the location and time of year, cloud-free images can be difficult to find (Ju and Roy, 2008). In addition, methods normally need pairs of images from different sensors captured on the same dates. When satellites have long revisiting times, matching pairs might be difficult to find or non-existent. Both issues are generally solved by expanding the time window for searching for candidates (Wang and Atkinson, 2018). Going further back in time increases the chances of finding a pair of images satisfying these requirements. However, feeding *STIF* methods with a temporally distant image can negatively affect their performance (Zhu et al., 2010, 2016; Xie et al., 2018) since images further apart normally resemble less closely the image to be predicted. Several solutions have been proposed recently to tackle each of the above problems separately. For example, some *STIF* methods like SaTellite dAta IntegRation (*STAIR*) (Luo et al., 2018, 2020) and Improved Flexible Spatiotemporal DAta Fusion (*IFSDAF*) (Liu et al., 2019b) are designed specifically to work with cloudy images. To avoid requiring matching image pairs, Wu et al. (2020) developed a new add-in for traditional fusion methods that achieved promising results. However, further investigation is required to devise fusion methods that can be applied for general purposes, even with cloud-covered data, and without requiring matching image pairs. Another practical issue that has received less attention is the computational cost. Comparative studies reveal that methods invest several minutes or hours to blend information for a single image (Liu et al., 2019a; Wang et al., 2020a). These computational times are not affordable in real-case studies where the trend is to involve increasingly large datasets. This is an important aspect that deserves further attention (Zhu et al., 2018).

In this paper, we propose a solution to simplify the inputs to a series of cloud-contaminated fine-scale images and a clear coarse-resolution image of the target date. The task of selecting the best data from a baseline time series of cloudy images demands the use of a formal methodology to prevent misclassified clouds and the necessity for matching coarse-resolution images. On the other hand, neither cokriging nor spatio-temporal methods appear to be suitable solutions for this particular problem. Cokriging has limitations due to the cross covariance matrix requirements. The utilization of stochastic spatio-temporal methods on satellite imagery presents practical challenges due to the data dimensionality (Das and Ghosh, 2020; Addink and Stein, 1999). As a result, there is a need to develop a new approach. First, cloudy fine-resolution images are turned into their coarse-resolution counterparts through upscaling. Then, we find local optimal information patches comparing the upscaled images with the target coarse image for selecting valid fragments from part-cloud images that the fusion algorithm can ingest (Chen et al., 2020). The term patches refers to small rectangular or square sub-regions extracted from a larger image. These methods work in tandem with *Fit-FC* (Wang and Atkinson, 2018), an efficient and effective alternative among *STIF* methods (Liu et al., 2019a; Zhou et al., 2021). As an attempt to reduce its computational time, the combination of methods is encapsulated into a single procedure and programmed to work in parallel. We refer to this method as Unpaired Spatio-Temporal Fusion of Image Patches (*USTFIP*).

The remainder of the paper is organized as follows: Section 2 introduces the proposed methodological approach. Section 3 describes the experimental set-up. Section 4 reports the results and Section 5 further discusses the proposed approach. Concluding remarks are summarized in Section 6.

2. Materials and methods

In the following, we simplify the mathematical expressions to a single band since the process is analogous in all bands. Let $C_{t_0}(\mathbf{S})$ be a coarse-resolution image (e.g. from MODIS), captured on a target date t_0. Pixels are located at $\mathbf{S} = \{S_1, \ldots, S_i, \ldots, S_N\}$, where S_i is a 2-D coordinate vector, and N is the total number of coarse-resolution pixels. Let $F_{\mathbf{T}}(\mathbf{s})$ be a time-series of fine-resolution images (e.g. from Landsat) captured on dates $\mathbf{T} = \{t_1, \ldots, t_k, \ldots, t_K\}$ (with $t_0 \notin \mathbf{T}$). For clarity, we refer to $\mathbf{T}$ as the baseline period. The fine-resolution pixels are located at $\mathbf{s} = \{s_1, \ldots, s_j, \ldots, s_M\}$ and M is the total number of pixels. The aim of the proposed spatio-temporal fusion method is to predict F_{t_0} based on C_{t_0} and $F_{\mathbf{T}}$, assuming that (1) $F_{\mathbf{T}}$ is a set of images with missing information due to the presence of clouds and (2) some of the coarse imagery $C_{\mathbf{T}}$ matching $F_{\mathbf{T}}$ is non-existent or unavailable.

2.1. Unpaired spatio-temporal fusion of image patches

The *USTFIP* method is summarized in a three-step procedure (see Fig. 1). The first step, referred to as Coarse Harmonization (*CH*), transforms the inputs ($F_{\mathbf{T}}$ and C_{t_0}) into a consistent spatio-spectral set of coarse images ($\hat{C}_{\mathbf{T}}$ and $\hat{C}_{t_0}$). The *CH* avoids requiring matching coarse-scale scenes corresponding to all the fine resolution images from the baseline $\mathbf{T}$. The second step, called Locally Optimal Prediction (*LOP*), transfers the temporal relationships between $\hat{C}_{\mathbf{T}}$ and $\hat{C}_{t_0}$ to the fine resolution. Relationships are established between sub-regions of images, which enables considering clear-sky fragments from partially cloudy scenes (see Fig. 1). *LOP* provides a first prediction of the fine-scale image on the target date ($\hat{F}_{t_0}^{LOP}$). The third and final step, named Spatial Filtering (*SF*), ensures that the output is spatially coherent by applying spatial weights to $\hat{F}_{t_0}^{LOP}$. *SF* mitigates some errors due to simplifications in the *LOP* process. Further details about each of the steps can be found in Sections 2.2–2.4.

2.2. Coarse Harmonization (CH)

The aim of *CH* is to obtain coarse-scale images with matching spatial and spectral resolutions for both the baseline ($\hat{C}_{\mathbf{T}}$), and the target date ($\hat{C}_{t_0}$). To this end, *CH* aggregates the fine-scale images to match the resolution of the coarse image (see (1) in Fig. 1) and

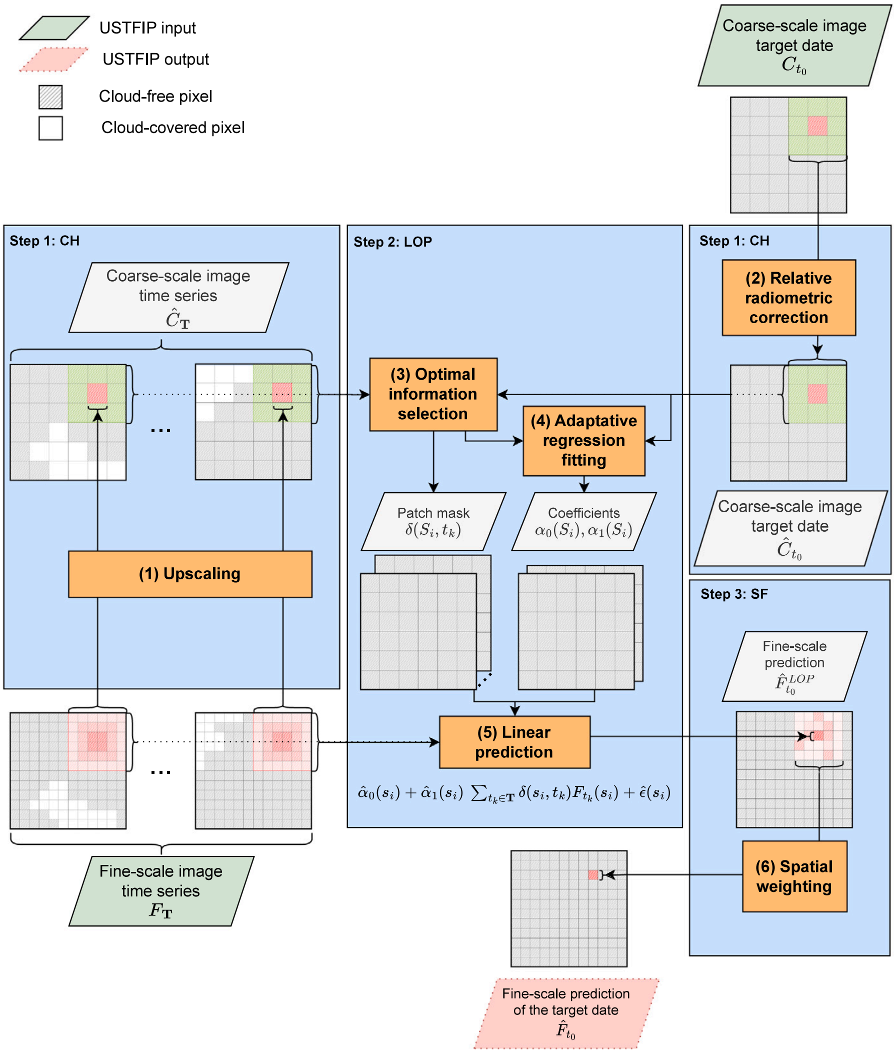
USTFIP input
USTFIP output
Cloud-free pixel
Cloud-covered pixel
Coarse-scale image target date C_{t_0}
Step 1: CH
Coarse-scale image time series $\hat{C}_{\mathbf{T}}$
...
(1) Upscaling
...
Fine-scale image time series $F_{\mathbf{T}}$
Step 2: LOP
(3) Optimal information selection
(4) Adaptive regression fitting
Patch mask $\delta(S_i, t_k)$
Coefficients $\alpha_0(S_i), \alpha_1(S_i)$
(5) Linear prediction
$\hat{\alpha}_0(s_i) + \hat{\alpha}_1(s_i) \sum_{t_k \in \mathbf{T}} \delta(s_i, t_k) F_{t_k}(s_i) + \hat{\epsilon}(s_i)$
Step 1: CH
(2) Relative radiometric correction
Coarse-scale image target date $\hat{C}_{t_0}$
Step 3: SF
Fine-scale prediction $\hat{F}^{LOP}_{t_0}$
(6) Spatial weighting
Fine-scale prediction of the target date $\hat{F}_{t_0}$

Fig. 1. Methodology of the Unpaired Spatio-Temporal Fusion of Image Patches (*USTFIP*). Step 1 corresponds to the Coarse Harmonization (*CH*), Step 2 is the Locally Optimal Prediction (*LOP*), and Step 3 corresponds to the Spatial Filtering (*SF*).

compensates for the bandwidth differences between sensors (see (2) in Fig. 1).

Adjusting the resolution is achieved through an average spatial aggregation. The aggregation can include adjacency effects that can be mitigated applying the convolution of the Point Spread Function (*PSF*), usually expressed as a Gaussian filter. The *PSF* represents the physical blurring effect caused by the movement of the sensor, electronics, atmospheric conditions and re-sampling (Huang et al., 2002; Kaiser and Schneider, 2008). The overall upscaling process can be mimicked through the following convolution

$$\hat{C}_{t_k}(S_i) = h(S_i) * F_{t_k}(S_i), \quad k = 1, \ldots, K \text{ and } i = 1, \ldots, N, \tag{1}$$

where $h(S_i)$ is a Gaussian filter with standard deviation 1 defined in a 3×3 window around S_i and normalized to sum to 1. Some of the reflectances $F_{t_k}(S_i)$ may be missing, likely due to the presence of clouds, and then $\hat{C}_{t_k}(S_i)$ remains missing. This rule may cause a substantial loss of information when clouds are scattered. However, the purpose of this rule is twofold: (1) It avoids biased estimates of the coarse-resolution pixel, since pixels around clouds are more likely to suffer the effect of undetected clouds or cloud shadows, (2) It skips areas around clouds where there might be misclassified shadows or cloud-shadows.

Remaining differences between $\hat{C}_{t_k}$ and C_{t_0} can be attributed to radiometric discrepancies between sensors. For example, the blue band in MODIS corresponds to 459 – 479 nm while in Landsat-5 and 7 it ranges around 450–520 nm and 441–514 nm, respectively (see Table 1). Generally, the consistency is high, but the small spectral differences can translate into bias. Applying relative radiometric correction methods (Chen et al., 2005) can avoid the propagation of such bias into the spatio-temporal fusion. The linear regression method is a simple and effective technique (Paolini et al., 2006) to re-scale the bandwidths between sensors. The aim of the relative radiometric correction is to transform the reflectances of C_{t_0} so they resemble those on the upscaled fine-resolution images $\hat{C}_{t_k}$, $t_k \in \mathbf{T}$. For every pixel S_i we define a $w_{RC} \times w_{RC}$ (5×5 by default) moving window surrounding S_i, and estimate the $\beta_0(S_i)$ and $\beta_1(S_i)$ coefficients of the local linear regression

$$\hat{C}_{\mathbf{T}}(S_i) = \beta_0(S_i) + \beta_1(S_i) C_{\mathbf{T}}(S_i) + \epsilon(S_i), \quad i = 1, \ldots, N, \tag{2}$$

where $\epsilon(S_i)$ is a $N(0, \sigma^2)$ random error, $\hat{C}_{\mathbf{T}}(S_i) = (\hat{C}_{t_1}(S_i), \ldots, \hat{C}_{t_k}(S_i))$ are the upscaled images obtained in Eq. (1), and some of the coarse images $C_{\mathbf{T}}(S_i) = (C_{t_1}(S_i), \ldots, C_{t_k}(S_i))$ are assumed to be known. When a coarse image from the baseline is not available, we simply remove its corresponding observations from the regression models. When the coarse-resolution sensor is MODIS, the estimated coefficients $\hat{\beta}_0(S_i)$ and $\hat{\beta}_1(S_i)$ can be obtained through ordinary least squares, due to its daily revisit time, and are used to provide the corrected coarse-scale image on the target date t_0 given by

$$\hat{C}_{t_0}(S_i) = \hat{\beta}_0(S_i) + \hat{\beta}_1(S_i) C_{t_0}(S_i). \tag{3}$$

When Sentinel-2 and Landsat −8 represent the fine and coarse images, estimating the coefficients by least squares may not feasible due to their long revisit times. In that case, $\beta_0(S_i)$ and $\beta_1(S_i)$ can be replaced with the values from a continental-scale assessment (Zhang et al., 2018).

Performing each of the previous steps separately for each band makes parallelizing the *CH* process straightforward. This is because the layers corresponding to each band can be processed by a different thread.

2.3. Locally Optimal Prediction (LOP)

The aim of the Locally Optimal Prediction (LOP) step is to obtain a first prediction of the fine-resolution image for the target date by capturing the change between the upscaled fine-scale images from the baseline $\hat{C}_{t_k}$, $t_k \in \mathbf{T}$, and the radiometrically corrected image for the target date $\hat{C}_{t_0}(\mathbf{S})$. For this, our method automatically finds the most suitable information within $\hat{C}_{\mathbf{T}}$ to be fused with $\hat{C}_{t_0}$ (see (3) in Fig. 1). Then, *USTFIP* fits a linear regression between patches of $\hat{C}_{t_0}$ and optimal information in $\hat{C}_{\mathbf{T}}$ (see (4) in Fig. 1).

High quality input data are acknowledged as a key factor in achieving accurate results in spatio-temporal fusion. Our method measures the quality using the similarity between $\hat{C}_{t_0}(\mathbf{S})$ and the upscaled images for the baseline $\hat{C}_{t_k}$, $t_k \in \mathbf{T}$. The largest correlation is a widespread strategy used to find the most similar data. Note that this strategy will help achieving a better fit for the linear regression, since correlation measures the degree to which two datasets are linearly related. For each pixel, we compute the linear correlation coefficient $r_{t_k}(S_i)$ as follows

$$r_{t_k}(S_i) = \frac{cov(\hat{C}_{t_k}(S_i), \hat{C}_{t_0}(S_i))}{\sqrt{var(\hat{C}_{t_k}(S_i))\; var(\hat{C}_{t_0}(S_i))}},\ t_k \in \mathbf{T}, \tag{4}$$

where $cov(\hat{C}_{t_k}(S_i), \hat{C}_{t_0}(S_i))$ is the covariance between the reflectances inside the $w_{LR} \times w_{LR}$ window for a date t_k in the baseline $\mathbf{T}$, and the reflectances in the same window for the target date t_0. Both $var(\hat{C}_{t_k}(S_i))$ and $var(\hat{C}_{t_0}(S_i))$ are the variances of the reflectances inside the $w_{LR} \times w_{LR}$ window for t_k and t_0, respectively. To compute the sample correlation coefficient, there must be no missing values inside the window. To select the maximum correlated date, we need that for each pixel S_i there is a window without missing values for at least one date in the baseline. This requirement ensures that we are able to capture temporal changes between the most similar windows while also avoiding any pixels close to clouds, as these are known to be more prone to errors. If there is any pixel without a correlation value for any date, we need to expand the baseline $\mathbf{T}$ and repeat the Coarse Harmonization (Step 1) for the new images. Using the results from Eq. (4), the optimal information can be determined as follows

$$\delta(S_i, t_k) = \begin{cases} 1 & \text{if } r_{t_k}(S_i) = max\{r_{\mathbf{T}}(S_i)\}. \\ 0 & \text{otherwise}. \end{cases} \tag{5}$$

where δ is a mask indicating which t_k is the optimal observation for pixel S_i. Then, *USTFIP* captures the change between the baseline and target time frames using the same moving window as in Eq. (4) to estimate the $\alpha_0(S_i)$ and $\alpha_1(S_i)$ coefficients in the following local linear regression

$$\hat{C}_{t_0}(S_i) = \alpha_0(S_i) + \alpha_1(S_i) \sum_{k=1}^{K} \delta(S_i, t_k) \hat{C}_{t_k}(S_i) + \epsilon(S_i) \quad i = 1, \ldots, N, \tag{6}$$

where $\epsilon(S_i)$ is the white noise process. The coefficients can be estimated by ordinary least squares. The selection mask $\delta(S_i, t_k)$ ensures that only pixels with the largest correlation are involved in the linear regression between $\hat{C}_{t_k}(S_i)$ and $\hat{C}_{t_0}(S_i)$. To obtain a fine-scale prediction, USTFIP downscales the estimated coarse-resolution coefficients $\hat{\alpha}_0(\mathbf{S})$ and $\hat{\alpha}_1(\mathbf{S})$, the regression residuals $\hat{\epsilon}(\mathbf{S})$ and the selection mask $\delta(\mathbf{S}, \mathbf{T})$ to the fine-resolution, to obtain $\hat{\alpha}_0(\mathbf{s})$, $\hat{\alpha}_1(\mathbf{s})$, $\hat{\epsilon}(\mathbf{s})$ and $\delta(\mathbf{s}, \mathbf{T})$. USTFIP uses the nearest neighbor method to resample $\hat{\alpha}_0(\mathbf{S})$, $\hat{\alpha}_1(\mathbf{S})$ and $\delta(\mathbf{S}, \mathbf{T})$, and $\hat{\epsilon}(\mathbf{S})$ is resampled through a bicubic interpolation. Assuming that the coarse-resolution change obtained in Eq. (6) can be applied over the fine-resolution images, we obtain a first estimate of the fine-resolution image for the target date $\hat{F}^{LOP}_{t_0}(\mathbf{s})$, namely

$$\hat{F}^{LOP}_{t_0}(s_i) = \hat{\alpha}_0(s_i) + \hat{\alpha}_1(s_i) \sum_{k=1}^{K} \delta(s_i, t_k) F_{t_k}(s_i) + \hat{\epsilon}(s_i). \tag{7}$$

Performing all of the steps in the *LOP* step separately for each band enables direct parallelization. Similar to the *CH* process, we can process the information corresponding to each band in a separate thread without the need for additional steps.

2.4. Spatial Filtering (SF)

Simplifications made during the transition from Eqs. (6) to (7) lead to errors in $\hat{F}^{LOP}_{t_0}$. Blocky artifacts may appear in $\hat{F}^{LOP}_{t_0}$ as a result of the scale-invariant assumption in Eq. (7). Additionally, the bicubic interpolation of $\hat{\epsilon}$ can cause over-smoothed residuals (see Wang and Atkinson (2018)). A spatial weighting filter can mitigate both issues by canceling-out contrasting errors in similar land-use classes. USTFIP defines the spatial filter like that in Wang and Atkinson (2018). Thus, the second prediction of the fine-scale pixels $\hat{F}_{t_0}(s_i)$ is the weighted sum of n neighboring pixels that are spectrally similar to s_i

$$\hat{F}_{t_0}(s_i) = \sum_{i=1}^{n} W_i \hat{F}^{LOP}_{t_0}(s_i), \tag{8}$$

where W_i is the inverse distance weight for each of the n most similar pixels, and similarity is defined as the inverse of the reflectance difference between the central pixel s_i and those within the extent of a $w \times w$ window around it. Weights are subject to the condition $\sum_{i=1}^{n} W_i = 1$. The main difference with the *Fit-FC* method is that similarity is measured in the optimal patch from $F_{\mathbf{T}}$ that is involved in Eq. (7) through the selection mask δ.

Performing *SF* jointly for all bands hinders direct parallelization of this step. Therefore, additional steps are required to compute it in parallel. To accomplish this, *USTFIP* divides the first prediction image, $\hat{F}^{LOP}_{t_0}$, into equal-sized sub-images or chunks and builds a buffer of size w around each of them to ensure filtering can be performed on all pixels within the chunk. These buffered chunks can then be processed by different threads.

3. Experimental analysis

The performance of the proposed *USTFIP* method is illustrated with real data in situations where the classical data requirements are progressively difficult to meet in terms of cloud coverage and finding matching pairs. More specifically, Section 3.1 describes the datasets used for the experiment and Section 3.2 describes the experimental scenarios used for the quality assessment. We tested *USTFIP* against the *Fit-FC*, *STARFM* and *FSDAF* methods. Since *Fit-FC* achieves better accuracy than *STARFM* and *FSDAF* in all the experimental scenarios (see Tables A.6, A.7 A.8 and A.9), in the remainder of this article we focus on comparing *USTFIP* and *Fit-FC*.

Table 1
Band wavelengths of sensors onboard the Landsat-5, Landsat-7, and MODIS Terra/Aqua satellites.

Name	Wavelengths (nm)		
	Landsat-5 (TM)	Landsat-7 (ETM)	MODIS
Red	630–690	631–692	620–670
Green	520–600	519–601	545–565
Blue	450–520	441–514	459–479
Near Infrared (NIR)	760–900	772–898	841–876
Shortwave Infrared (SWIR) 1	1550–1750	1566–1651	1628–1652
Shortwave Infrared (SWIR) 2	2080–2350	2107–2294	2015–2155
Spatial resolution	30 m	30 m	500 m

Table 2
Band wavelengths of Sentinel-2 and Sentinel-3 images.

Name	Sentinel-2 (MSI)		Sentinel-3 (OLCI)	
	Wavelength	Spatial resolution	Wavelength	Spatial resolution
Red	650–680 nm	10 m	660–670 nm	300 m
Green	543–578 nm	10 m	555–565 nm	300 m
Blue	458–523 nm	10 m	485–495 nm	300 m
Near Infrared (NIR)	855–875 nm	20 m	855–875 nm	300 m

3.1. Datasets

The experimental datasets consist of cloud-free satellite sensor images for four different regions. The first two datasets correspond to Landsat and MODIS surface reflectance images for two widely studied areas (Emelyanova et al., 2013; Zhu et al., 2016; Cheng et al., 2017; Zhao et al., 2018; Li et al., 2020b; Wang and Atkinson, 2018; Liu et al., 2019a): the Coleambally Irrigation Area (Coleambally henceforth) and the Gwydir catchment (Gwydir henceforth). The last two datasets correspond to Sentinel-2 and Sentinel-3 bottom-of-atmosphere images for two square regions in North Dakota, America (Tang et al., 2021).

The Coleambally region intersects with the paths 92/93 and row 84 of the tilling system of Landsat-7 and the MODIS tiles H: 29/V: 12 and H: 30/V: 12 (top row in Fig. 2). Gwydir is covered by the paths 91/92 and rows 80/81 from the tiling system of Landsat-5 and the MODIS tiles H: 30/V: 11 and H: 31/V: 11 (bottom row in Fig. 2). The size of the images from Coleambally and Gwydir are 1400 × 1391 pixels and 2399 × 2400 pixels, covering areas of approximately 1753 km^2 and 5182 km^2, respectively. The period of analysis spans around October 7th, 2001 to May 3rd, 2002 in Coleambally and March 16th, 2004 to March 3rd, 2005 in Gwydir, with a total number of 17 and 14 images, respectively (see Table 3). In Coleambally, the time of analysis is centered around the austral summer to encapsulate the development of summer crops, while in Gwydir it focuses on flooding events on December, 2004. Table 1 shows the bands included in these images.

Both locations in North Dakota cover two different 15 km by 15 km sites clipped from the Sentinel-2 Multispectral Imager (MSI) and Sentinel-3 Ocean and Land Color Imager (OLCI) time-series images covering the same 109.5 km by 109.5 km area (see Fig. 3). Fine images for both regions are 750 × 750 pixels, with a spatial resolution of 20 m, independent of the original resolution. The period of analysis spans around June 6th, 2019 to October 10th, 2020 in both sites, with a total number of 11 images (see Table 3). Table 2 shows the bands included in these images.

Before fusion, images from each dataset were coregistered and radiometrically corrected. Misalignment between satellite sensors has been recognized as an important source of error (Gevaert and García-Haro, 2015; Tang et al., 2020; Wang et al., 2020b; Zhou et al., 2021). A sharper description of the exact preprocessing steps can be found in Emelyanova et al. (2013) for Coleambally and Gwydir and in Tang et al. (2021) for the North Dakota datasets. According to Zhou et al. (2021), *Fit-FC* is highly sensitive to bias between sensors and correction algorithms from different satellite programs. Therefore, images were also radiometrically normalized assuming a linear relationship between the coarse and fine scenes (MODIS and Landsat, and Sentinel-3 and Sentinel-2) captured on the same dates.

Table 3
Time-series of cloud-free imagery in Coleambally, Gwydir and North Dakota.

No.	Coleambally	Gwydir	North Dakota
1	2001-10-07	2004-04-16	2019-06-06
2	2001-10-16	2004-05-02	2019-07-18
3	2001-11-01	2004-07-05	2019-08-20
4	2001-11-08	2004-08-06	2019-09-16
5	2001-11-24	2004-08-22	2020-03-27
6	2001-12-03	2004-10-25	2020-06-07
7	2002-01-04	2004-11-26	2020-08-11
8	2002-01-11	2004-12-12	2020-08-24
9	2002-02-12	2004-12-28	2020-09-10
10	2002-02-21	2005-01-13	2020-09-25
11	2002-03-09	2005-01-29	2020-10-10
12	2002-03-16	2005-02-14	–
13	2002-04-01	2005-03-02	–
14	2002-04-10	2005-04-03	–
15	2002-04-17	–	–
16	2002-04-26	–	–
17	2002-05-03	–	–

The experiments simulated the presence of clouds by masking the scenes in Table 3 artificially. For this, the cloud masks were extracted from the quality bands of other fine scenes, either from Landsat or Sentinel-2. The fraction of cloud-covered pixels ranged between 4–73% and 1–86% in Coleambally and Gwydir, respectively, and between 15.5–88.2% for North Dakota. Information about the coverage, shape and distribution of clouds can be found in Appendix A. Simulating the presence of clouds offers several advantages over using actual cloud-covered images. Notably, it provides greater flexibility and allows for greater control over the missing data. Additionally, it enables the perfect detection of clouds and cloud-shadows, a topic which falls outside the scope of this manuscript.

3.2. Experimental scenarios

The assessment comprised several scenarios representing increasingly challenging situations for finding matching pairs of cloud-free images. That is, for a given date of prediction (target), it is necessary to go further back in time to locate a clear fine image and its coarse counterpart (reference image, henceforth). The images between the target and reference images are assumed partially cloud-covered according to Figs. A.7, A.8 and A.9 in Appendix A.

Each scenario involves several experiments changing the date of the target image from image no. 7 onward, in sequence along the time-series in Table 3. The starting point at image no. 7 ensures that all scenarios predict the same images which facilitates their comparison.

Overall, the assessment comprised 55 experiments in Coleambally (11 targets × 5 scenarios), 40 experiments (8 targets × 5 scenarios) in Gwydir, and 25 experiments in the North Dakota regions (5 targets × 5 scenarios). In each experiment, *Fit-FC* predicted the fine-scale image on the target date using the fine-coarse pair of images from the reference date plus the coarse image from the target date. In contrast, *USTFIP* considered the fine images available from the reference to the target and the coarse image from the target date. For example, in scenario 1, the first experiment in Coleambally targeted the prediction of image no. 7 assuming that the clean reference image is image no. 5 and image no. 6 is contaminated by clouds (47%). Subsequent scenarios increase the temporal separation between the reference and target image, expanding the baseline period and increasing the number of cloud-covered scenes. Thus, in the first experiment in scenario 2 in Coleambally, the reference

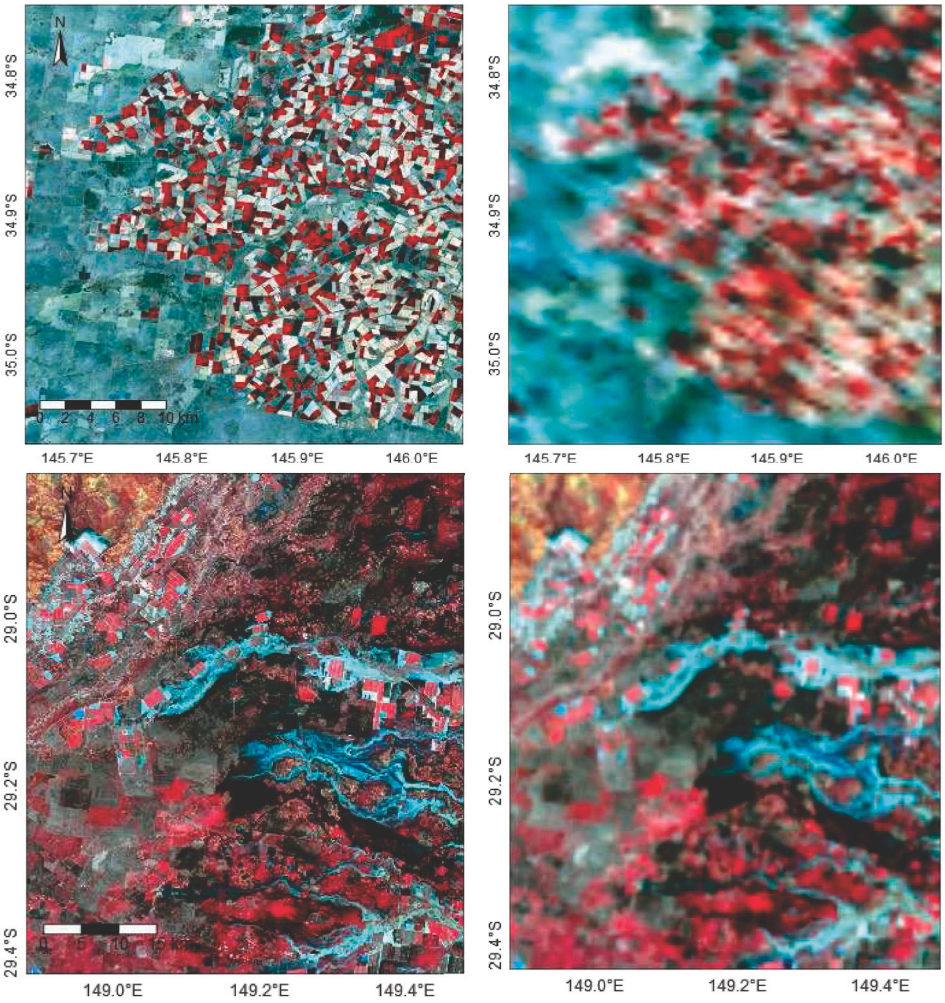

Fig. 2. False color representation (NIR, red and green as RGB) of the fine-scale (left) and coarse-scale (right) imagery at Coleambally (top) and Gwydir (bottom). The images correspond to images no. 7 and 8 of the corresponding time-series, captured on January 4th 2002 and December 12th, 2004, respectively.

and target correspond to images no. 4 and 7, while no. 5 and 6 are cloudy (22% and 47%, respectively). The scenarios are named according to the number of cloudy scenes that exist between the reference and target images. Therefore, scenario 1 groups experiments where there is a single cloud-covered image between the reference and target images. Scenario 2 corresponds to experiments where there are two cloud-covered scenes between the cloud-free reference and target images. Scenarios 3 to 5 group experiments with 3 to 5 cloud-covered images between the cloud-free reference and target. The study examined up to five scenarios, which in the North Dakota sites represents almost a year between the reference and target. In every experiment, *Fit-FC* used a 5 × 5 neighborhood for the regression model fitting (*RM*) stage and 31 × 31 pixel window and 20 similar pixels for the spatial filtering (*SF*) and residual compensation (*RC*) steps. We also used the same set of parameters as in Liu et al. (2019a) for the *STARFM* and *FSDAF*. *USTFIP* also used a 31 × 31 pixel window and 20 similar pixels for the spatial filtering.

We measured the computational time spent for each method to carry out the spatio-temporal fusion and contrasted the outputs against the actual fine-scale image. The evaluation metrics chosen for the experiment are the Root Mean Square Error (*RMSE*) and the Relative Root Mean Square Error Difference (*RRMSED*) to evaluate the spectral accuracy, and the Robert's Edge (*Edge*) to evaluate the spatial accuracy.

The RMSE for a fusion method m is given by

$$RMSE_m = \sqrt{\frac{\sum_i^n (F_{t_0}(s_i) - \hat{F}_{t_0}^m(s_i))^2}{n}} \tag{9}$$

where $\hat{F}_{t_0}^m$ is the prediction of the target fine image given by method m.

The RRMSED is given by

$$RRMSED = \frac{RMSE_{FFC} - RMSE_{USTFIP}}{RMSE_{FFC}} \tag{10}$$

where $RMSE_{FFC}$ and $RMSE_{USTFIP}$ are the RMSE obtained for the Fit-FC and USTFIP methods, respectively.

The spatial accuracy is evaluated based on the normalized difference of the Robert's Edge spatial feature, between the output and the reference images given by:

$$S_m(s_i) = (S_{t_0}(s_i) - \hat{S}_{t_0}^m(s_i))/(S_{t_0}(s_i) + \hat{S}_{t_0}^m(s_i)), \quad s_i \in \mathbf{s} \tag{11}$$

where S_{t_0} is the spatial feature obtained from F_{t_0} and $\hat{S}_{t_0}^m$ is the spatial feature obtained from $\hat{F}_{t_0}^m$. Then, we average the $S_m(s_i)$ values corresponding to pixels with $\hat{S}_{t_0}^m(s_i)$ values greater than the 90th percentile,

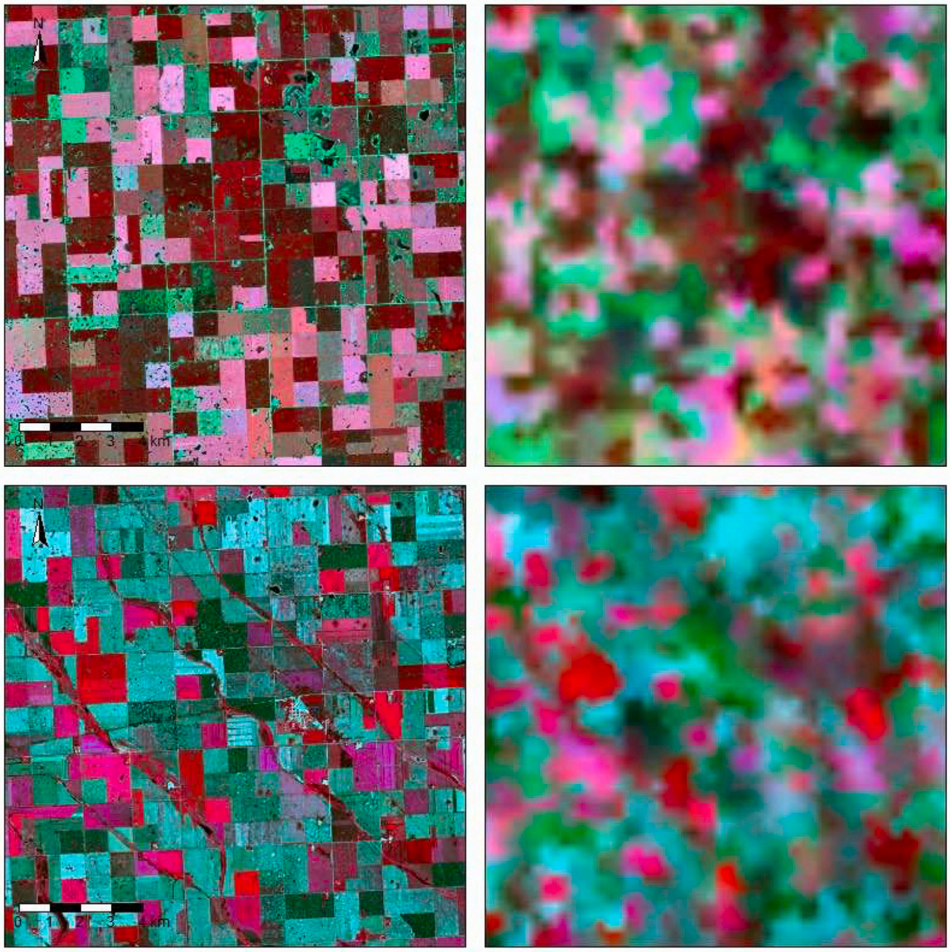

Fig. 3. False color representation (NIR, red and green as RGB) of the fine-scale (left) and coarse-scale (right) imagery at both regions in North Dakota. The images correspond to images no. 3 and 4 of the corresponding time-series, captured on August 20th 2019 and September 16th, 2019, respectively.

to obtain a single value for the spatial accuracy, denoted by $Edge_m$. A spatial accuracy of 0 represents a feature that is the same in both the real and the output images, negative values indicate over-smoothing in the feature for the prediction, and positive values indicate over-sharpening. For a sharper description of the Edge feature see Zhu et al. (2022). These metrics are calculated for each band and averaged across the bands described in Tables 1 and 2. The analysis was carried out on a computer with an Intel(R) Core(TM) i7-6700 @3.40 GHz processor and four cores.

4. Results

4.1. Visual inspection

The visual inspection focuses on the 7th image from Coleambally (Fig. 4), the 8th image of the time-series from Gwydir (Fig. 5) and the 7th image from both North Dakota regions. The target image from Coleambally corresponds to January 4th, 2002, a time in the season when crops experience rapid vegetative growth. In Gwydir, the target scene is from December 12th, 2004, a date in which the area was flooded. In the North Dakota regions, the target image corresponds to August 24th, 2020, the date that should be the most affected by the huge temporal gaps around the 4th image. Figures display a general overview of the location and a smaller area for detailed analysis. In each sub-group, the top and bottom rows correspond to predictions from *USTFIP* and *Fit-FC*, respectively. Images are represented in false-color. Coleambally and the North Dakota regions use *NIR*, red, and green as *RGB*, which highlights the presence of vegetation and water. Images from Gwydir use *SWIR*, *NIR*, and green as *RGB* to emphasize the flooding. The predictions are labeled using the date of the reference image which is increasingly distant from the target date. Note that, in addition to the reference image, *USTFIP* also takes advantage of the cloudy scenes between the reference and target dates.

The detailed area in Fig. 4 shows that, in general, *USTFIP* makes predictions closer to the actual image than *Fit-FC*. This is especially evident in the white field at the top and the dark areas in the central parts of the image (yellow rectangles). As the clear-sky input image separates from the target image, the similarity between the real and predicted images decreases. Colors and boundaries degrade more quickly with *Fit-FC* than with *USTFIP*.

The impression from the visual inspection is supported by quantitative assessment for the overall region. The *RMSE* from *Fit-FC* increases in 0.0086 units (from 0.0389 to 0.0475) as the reference moves from November 24th to October 7th. Instead, the error from *USTFIP* increased by 0.0075 units (from 0.0376 to 0.0452) in the same situation. The detailed area demonstrates that using additional information from

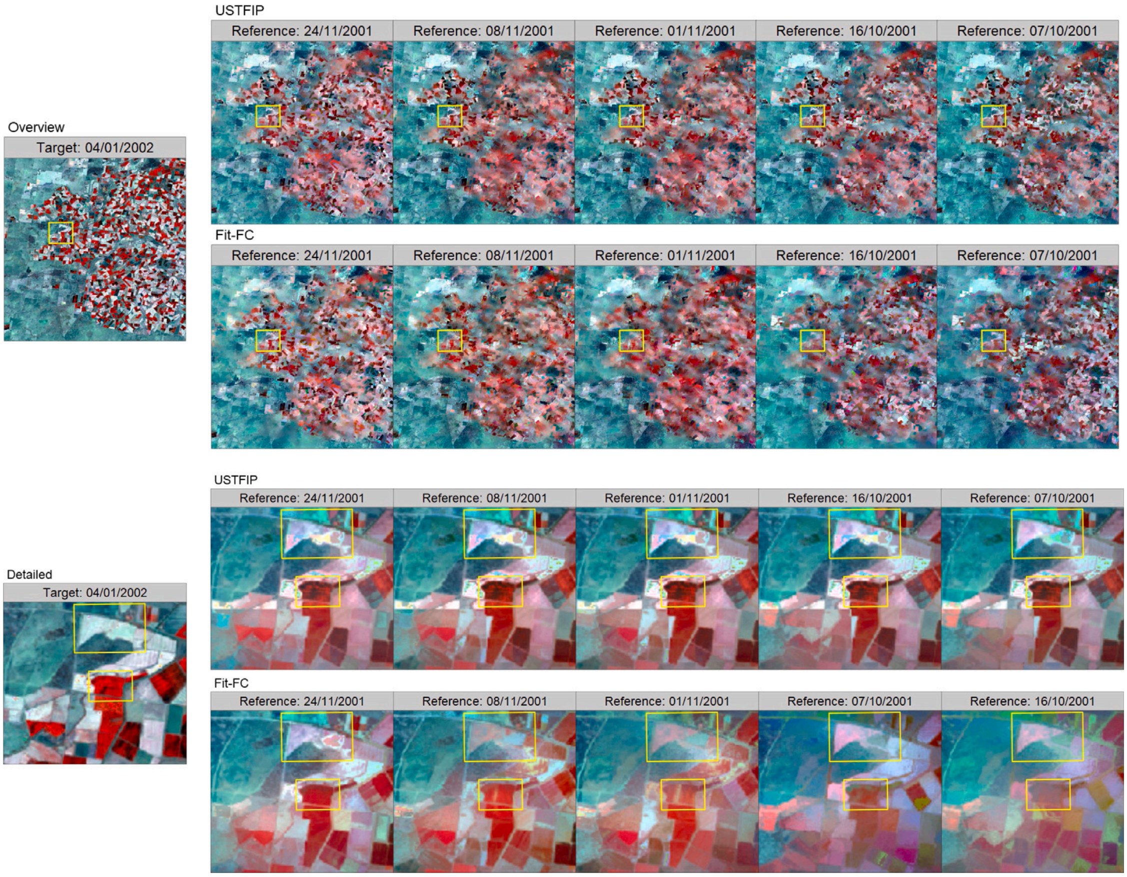

Fig. 4. Comparison of the target image and the predictions from *USTFIP* and *Fit-FC* for the different scenarios for Coleambally on January 4th, 2002. On the far left, an overview of the Landsat-7 target image and below, a detail of the yellow region within the overview image. From left to right, predictions based on reference images captured on increasingly separated dates. The first two rows show the entire region. Last two rows are zoomed from the specific yellow area, where new yellow squares are added for visual inspection. Representations use a false-color palette (NIR, red and green as *RGB*).

cloudy images captured closer in time makes the *USTFIP* predictions exhibit a slower decline in contrast and a more gradual increase in blurriness compared to *Fit-FC*.

The results of the experiment in Gwydir can be seen in Fig. 5. The flooding event makes the target image very different from the previous ones, as reflected by the small correlation between them (always lower than 0.2519). The images from *USTFIP* and *Fit-FC* are very similar to each other regardless of the separation between the baseline and target. The predictions represent well the colors of the real image, but the boundaries and shapes of the flooded area differ from the original (see rightmost yellow rectangle in the last two rows from Fig. 5). The poor prediction of shapes is expected since fine-scale images from previous moments do not provide evidence about the location of flooded terrain . This suggests that using recent data to predict transient abrupt changes does not bring an improvement in prediction power. In line with visual perception, the *RMSE* between the predictions and the actual image remain nearly constant for the different experiments (0.0320–0.0323 for *Fit-FC* and 0.0317–0.0320 for *USTFIP*). Nevertheless, *USTFIP* performs moderately more accurately than *Fit-FC* in non-flooded areas.

The results of the experiments in both North Dakota regions can be seen in Fig. 6. The images from *USTFIP* and *Fit-FC* are very similar to each other when the gap between the baseline and target is a single image, however, *USTFIP* more accurately preserves the shapes in the original image, as evidenced by the difference in the Edge feature presented in Table 5. In addition, the *USTFIP* predictions deprecate at a slower rate than the ones from *Fit-FC*, due to the use of intermediate images, which makes the predictions much less sensitive to the large temporal gap between images 3 to 5.

4.2. *Quantitative assessment*

Tables 4 and 5 summarize the performance of the *USTFIP* and *Fit-FC* methods as the temporal gap between the reference and target images increases. Table 4 shows the average results of the eleven and eight experiments in each experimental scenario in Coleambally and Gwydir, respectively. As expected, the reference-target resemblance decreases with time in both locations. In Coleambally, the correlation between these images decreases from 0.75 to 0.40 as their temporal separation increases from 28 to 86 days. Similarly, in Gwydir, the correlations decrease from 0.54 to 0.28 as time separation increases from 44 to 158 days. In general, images from Coleambally have a slightly larger correlation with the target image than those from Gwydir. This suggests that subsequent images are more alike in Coleambally. Yet, the larger decrease in correlation in a shorter time-span in Coleambally implies that images change more quickly. Table 5 shows the average results of the five experiments in each experimental scenario in both North Dakota sites. Images from these sites have much smaller correlations, probably due to the long temporal gaps between reference and target images.

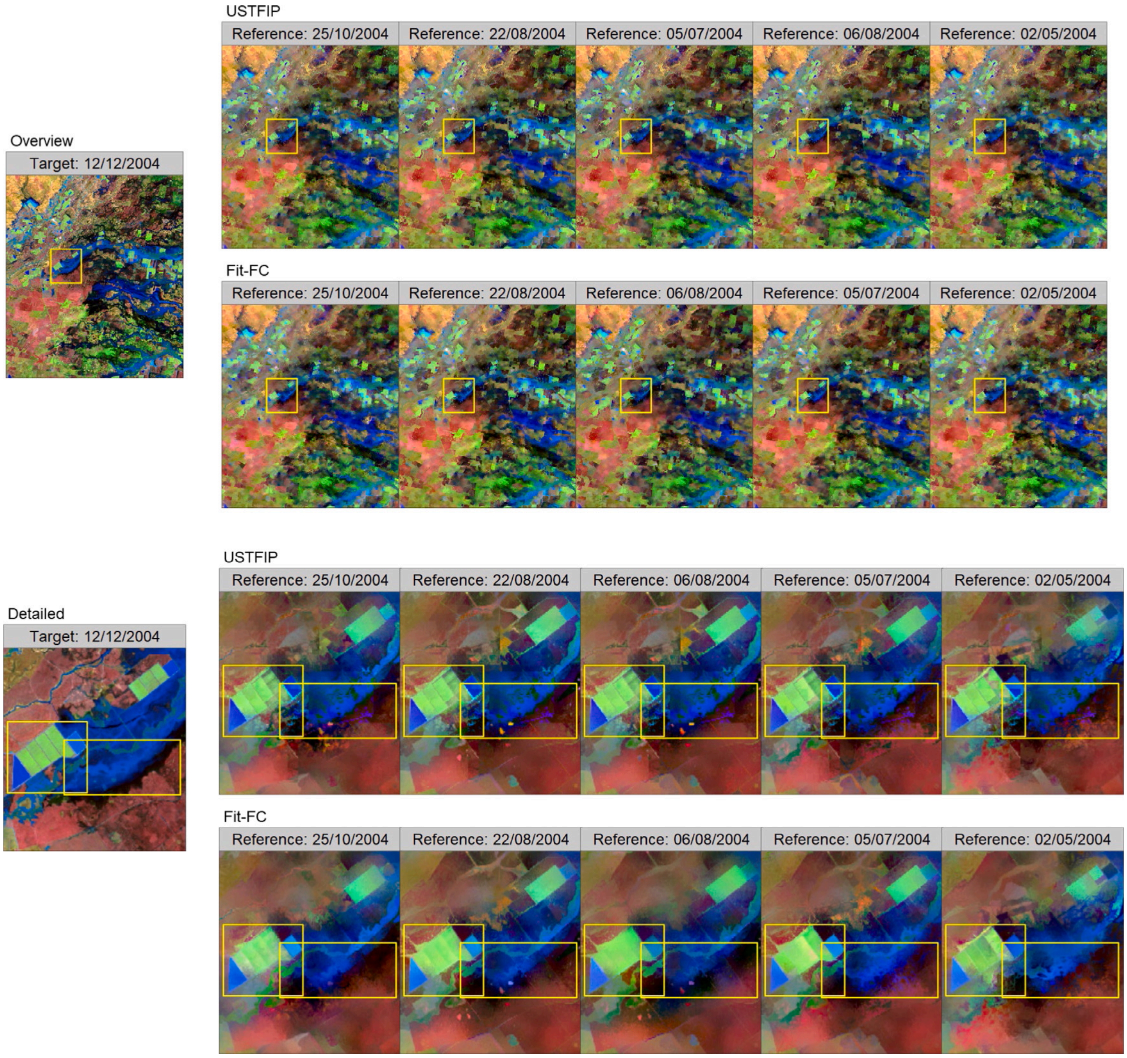

Fig. 5. Comparison of the target image and the predictions from *USTFIP* and *Fit-FC* for the different scenarios for Gwydir on December 12th, 2004. On the far left, an overview of the Landsat-5 target image and below, a detail of the yellow region within the overview image. From left to right, predictions based on reference images captured on increasingly separated dates. The first two rows show the entire region. Last two rows are zoomed from the specific yellow area, where new yellow squares are added for visual inspection. Representations use a false-color palette (SWIR, NIR and green as *RGB*).

The disparity between the reference and target images negatively affects the prediction accuracy of *Fit-FC*. In Coleambally, the *RMSE* increases from 0.0281 to 0.0340 as the correlation declines from 0.75 to 0.40. Similarly, in Gwydir the *RMSE* increases from 0.0235 to 0.0273 when the correlation declines from 0.54 to 0.28. The prediction error of *USTFIP* also depends on the baseline-target similarity, but to a lesser extent. In both Coleambally and Gwydir, the *RMSE* escalates with smaller correlation (from 0.0269 to 0.0293 in Coleambally and from 0.0229 up to 0.0238 in Gwydir), but the growth tends to stagnate more rapidly. The results from the North Dakota regions show that *USTFIP* is much less sensitive to small correlations and large temporal gaps, particularly when referring to spatial accuracy. The Robert's Edge feature is preserved much better by *USTFIP* than by *Fit-FC* (see Table 5).

In this algorithm, other aspects such as the distribution of clouds in space and time may play a critical role. For example, predicting the image from January 4th, 2002, in Coleambally results in a slightly larger *RMSE*, 0.0456 vs. 0.0452 from scenario $N = 4$ to $N = 5$. The fact that the image from October 16th has a clear-sky fragment at the bottom right corner, where the remaining images are contaminated by clouds, might be an advantage over having a closer clear-sky image. In general, the prediction accuracy of *USTFIP* improves when the number of available observations grows because this increases the chances of finding a suitable observation in LOP (see Tables A.6, A.7, A.8 and A.9 in Appendix A).

As Table 4 reveals, the *USTFIP* framework increases the accuracy from the *Fit-FC* method in all scenarios. Increases in accuracy are around 2-to-13% depending on the location and temporal gap. At short distances (scenario 1), *USTFIP* surpasses *Fit-FC* with a decrease in the *RMSE* of 5% in Coleambally and 2% in Gwydir and both North Dakota regions. In Coleambally and Gwydir, the difference in performance between *USTFIP* and *Fit-FC* increases as the reference-target separation increases. In scenario 5, the decrease in *RMSE* reaches 13% for Coleambally and Gwydir and 8% in both North Dakota regions. The larger gap between the methods as scenarios evolve confirms that the predictions from *USTFIP* deprecate at a slower rate than for *Fit-FC*. Note that *USTFIP* predictions take advantage of the increasingly available partial information between the baseline and target, so they rely less and less on the reference image. Individual experiments reveal

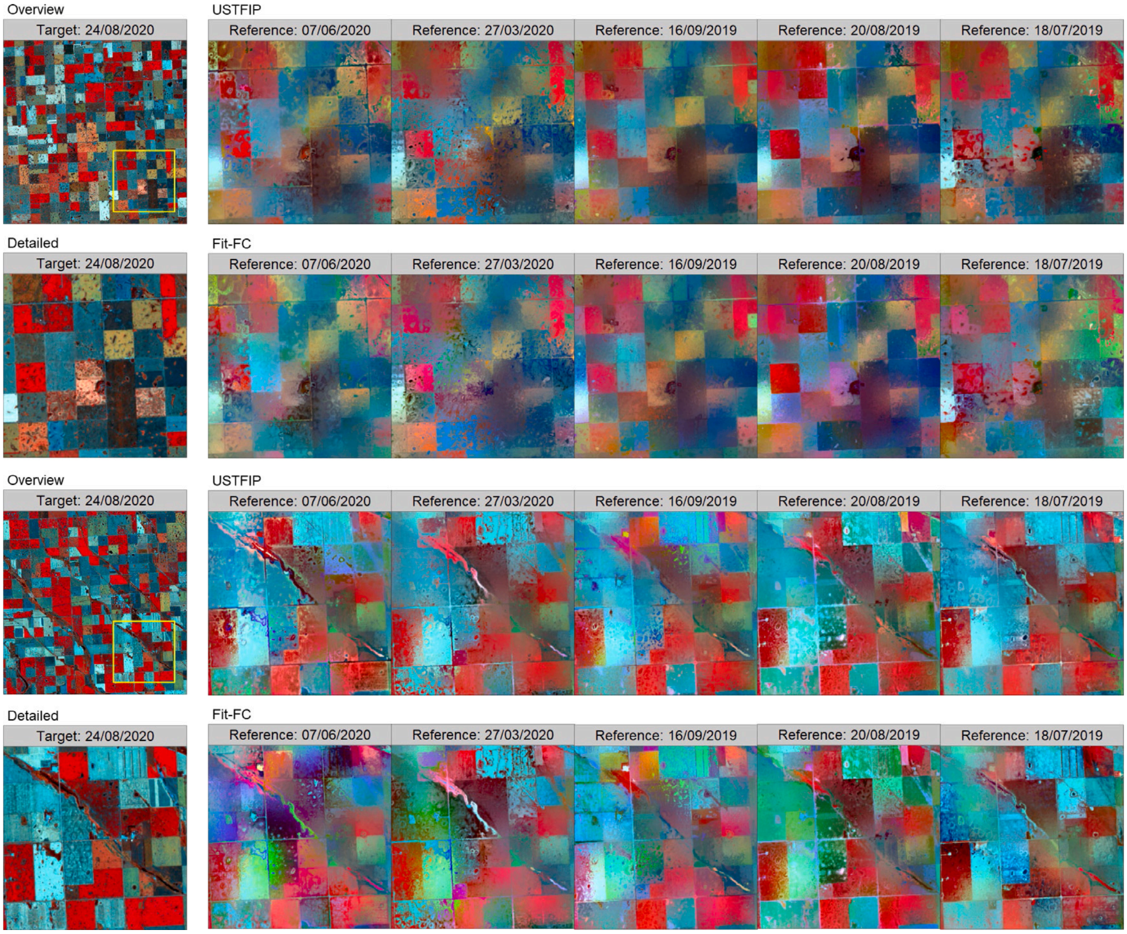

Fig. 6. Comparison of the target image and the predictions from *USTFIP* and *Fit-FC* for the different scenarios for both sites in North Dakota on August 24th, 2020. On the left, the original Sentinel-2 target images, both for the full region and a zoomed-in view of the yellow squares. From left to right, predictions for the zoomed region based on reference images captured on increasingly separated dates. The first two rows show the first region. Last two rows show the results for the second region. Representations use a false-color palette (NIR, red and green as *RGB*).

Table 4
Averages of the quality metrics for each scenario, where *GapN* is the number of cloud-covered scenes between the reference (clear-sky) and the target image. *Gap* represents the number of days between the reference and the target date, CC_{rt} is the correlation coefficient between the reference and target images. *Clouds* represents the average cloud coverage during the baseline period. *RMSE* is the root mean squared error of the prediction either from *Fit-FC* (*FFC*) or *USTFIP*. Edge is the mean of the normalized difference in the Robert's Edge feature and t is the computational time dedicated to fuse the images for Coleambally (1400 × 1391 pixels) and Gwydir (2399 × 2400 pixels).

Coleambally										
GapN	Gap	CC_{rt}	Clouds	$RMSE_{FFC}$	$RMSE_{USTFIP}$	RRMSED	$Edge_{FFC}$	$Edge_{USTFIP}$	t_{FFC}	t_{USTFIP}
no	days	–	%	–	–	%	–	–	min	min
1	28	0.75	17	0.0281	0.0266	5.07	−0.1438	−0.1121	1.97	1.69
2	42	0.66	22	0.0302	0.0279	7.10	−0.1659	−0.1234	1.87	1.64
3	57	0.58	25	0.0320	0.0286	10.24	−0.1557	−0.1181	1.89	1.75
4	72	0.50	29	0.0331	0.0290	11.70	−0.1758	−0.1429	1.91	1.86
5	86	0.40	31	0.0340	0.0293	13.33	−0.1909	0.1556	1.93	1.97
Gwydir										
GapN	Gap	CC_{rt}	Clouds	$RMSE_{FFC}$	$RMSE_{USTFIP}$	RRMSED	$Edge_{FFC}$	$Edge_{USTFIP}$	t_{FFC}	t_{USTFIP}
no	days	–	%	–	–	%	–	–	min	min
1	44	0.54	23	0.0235	0.0229	2.05	−0.1278	−0.1281	6.56	3.92
2	68	0.46	32	0.0252	0.0235	6.53	−0.1543	−0.1398	6.48	4.39
3	94	0.38	38	0.0263	0.0237	9.63	−0.1707	−0.1573	6.28	5.07
4	126	0.33	40	0.0271	0.0238	12.18	−0.1849	−0.1699	6.15	5.55
5	158	0.28	42	0.0273	0.0237	13.40	−0.1767	−0.1682	6.23	5.94

Table 5
Averages of the quality metrics for each scenario, where *GapN* is the number of cloud-covered scenes between the reference (clear-sky) and the target image. *Gap* represents the number of days between the reference and the target date, CC_{rt} is the correlation coefficient between the reference and target images. *Clouds* represents the average cloud coverage during the baseline period. *RMSE* is the root mean squared error of the prediction either from *Fit-FC* (*FFC*) or *USTFIP*. Edge is the mean of the normalized difference in the Robert's Edge feature and t is the computational time (in seconds) dedicated to fuse the (750 × 750 pixel) images for both sites in North Dakota.

Site 1										
GapN	Gap	CC_{rt}	Clouds	$RMSE_{FFC}$	$RMSE_{USTFIP}$	RRMSED	$Edge_{FFC}$	$Edge_{USTFIP}$	t_{FFC}	t_{USTFIP}
no	days	–	%	–	–	%	–	–	seconds	seconds
1	61	0.24	27	0.0456	0.0447	1.73	−0.1652	−0.1260	32.81	54.42
2	133	0.10	36	0.0469	0.0460	2.46	−0.1917	−0.1407	32.43	55.29
3	207	0.07	39	0.0471	0.0459	3.09	−0.1777	−0.1328	31.53	56.00
4	285	0.12	40	0.0472	0.0464	2.48	−0.2042	−0.1685	32.01	56.32
5	359	0.12	39	0.0517	0.0472	7.88	−0.2261	−0.1859	30.53	54.90
Site 2										
GapN	Gap	CC_{rt}	Clouds	$RMSE_{FFC}$	$RMSE_{USTFIP}$	RRMSED	$Edge_{FFC}$	$Edge_{USTFIP}$	t_{FFC}	t_{USTFIP}
no	days	–	%	–	–	%	–	–	seconds	seconds
1	61	0.24	27	0.0427	0.0418	2.77	−0.1472	−0.0595	28.00	45.89
2	133	0.10	36	0.0437	0.0427	3.16	−0.1513	−0.0706	27.91	46.65
3	207	0.08	39	0.0438	0.0429	2.27	−0.1677	−0.1048	27.96	47.19
4	285	0.16	40	0.0440	0.0431	2.06	−0.1726	−0.0963	27.93	47.49
5	359	0.15	39	0.0487	0.0436	7.73	−0.1916	−0.1220	28.14	47.83

that *USTFIP* does not provide more accurate results than *Fit-FC* in all situations. As Tables A.6, A.7, A.8 and A.9 in Appendix A show, *Fit-FC* gives more accurate predictions than *USTFIP* for the pixels with a single available observation in the baseline. As a consequence, in 7.3% and 17.5% of the experiments in Coleambally and Gwydir, *Fit-FC* predicts more accurately than *USTFIP*. In these situations, the *RRMSED* is 0.79% (0.47-to-1.02%) in Coleambally and 1.98% (0.13-to-7.74%) in Gwydir. These results transfer well to a band-by-band analysis (see Appendix B), where again USTFIP outperforms the rest of the methods.

USTFIP runs faster than *Fit-FC* in most scenarios even when considering the generation of coarse-scale images, reducing the computational time by up to 40% (Table 4). The reduction is accomplished through the parallel application of linear model fitting, filtering and compensation to each band or image sub-region. Note that the computational cost of the *Fit-FC* remains approximately constant, around 2 and 6 min in all experiments in Coleambally and Gwydir. Instead, *USTFIP* running times increase with the number of fine-scale input images, from 1.69 to 1.97 min in Coleambally and 3.92 to 5.94 min in Gwydir. Longer baselines entail a greater number of calculations in *USTFIP* to generate synthetic coarse-scale images and select the optimal information within a greater pool of data. Naturally, the number of pixels of the image is another factor to consider, since the additional steps required for the parallel processing can make it counterproductive for small images, such as those of the North Dakota regions (750 × 750 pixels). On average, the time saved by *USTFIP* is 7% in Coleambally (1400 × 1391 pixels) and 21% in Gwydir (2399 × 2400).

5. Discussion

STIF methods impose strict input data requirements that make them less functional for real-case studies. In this manuscript, we proposed combining two strategies to relax these demands. The first involves optimal information selection (Xie et al., 2018) in sub-areas of images, which eliminates the need for clear-sky images and, additionally, makes use of partly covered scenes. The second one is the generation of synthetic coarse images (Wu et al., 2020) which avoids the necessity of matching pairs of fine-coarse images. Here, both strategies were coupled with the *Fit-FC* method (Wang and Atkinson, 2018), creating an overall framework referred to as the Unpaired Spatio-Temporal Fusion of Image Patches (*USTFIP*). The *Fit-FC* and *USTFIP* methods were tested and compared using two widely analyzed MODIS-Landsat time-series datasets (Emelyanova et al., 2013) and two Sentinel-2 and 3 datasets (Tang et al., 2021). In successive simulation scenarios, the experiments assumed that a clear pair of coarse-fine images is available further apart from the time of prediction as a consequence of frequent clouds or the absence of overlapping images between sensors.

Results showed that the strategies not only make *USTFIP* more adapted to real conditions, but also led to increased accuracy. The assimilation of partially cloudy images enables consideration of data closer to the time of prediction. These data are sometimes more informative than the data from further away in time which increases fusion accuracy. In our simulation experiments, *USTFIP* achieved on average a smaller *RMSE* than *Fit-FC* in all scenarios. Moreover, relying on the latest observations, if available and adequate, makes the framework within *USTFIP* more robust to challenging conditions than traditional *STIF* methods. The slower growth of *RMSE* from *USTFIP* with distance between the latest clear-sky image and the target image suggests that our proposal is especially advantageous in cloud prone seasons or regions. Using the latest observations available is also important in progressively and rapidly changing landscapes. Similarity between images, a key contributor to prediction accuracy, declines quickly in this type of environment. Thus, *USTFIP* is likely to be preferable over the rest of the benchmark methods. Finally, in line with the conclusions from Wu et al. (2020), our findings challenge the idea that matching pairs of images are essential in spatio-temporal fusion. The benefits of using the most recent information can offset the errors from the generation of synthetic coarse-scale images. Our experiments demonstrate how parallel computing techniques can better exploit the available computational resources to reduce running times. Here, the time reductions achieved through parallel computing counteracts the higher computational demands from achieving flexibility in the input requirements. The generation of coarse scale images and the selection of optimal information require further calculations than using a single clear-sky image which may trigger estimation errors in the fusion model. This is more evident in the less restrictive scenario (scenario 1). However, even in this case, our method outperforms Fit-FC on average (see Tables 4 and 5).

Matching pixel locations is a crucial step for achieving accurate predictions using linear regression models. Extending the *CH* step to include a coregistration step, could greatly improve the prediction when input images are not already coregistered. Furthermore, as the method was originally developed for MODIS-Landsat fusion, the PSF filter was defined as a Gaussian filter (Huang et al., 2002; Kaiser and Schneider, 2008) with specific window size and standard deviation that provided the best balance between avoiding areas obstructed by clouds and minimizing data loss. Remarkably, we found that this PSF filter also performed well when downscaling Sentinel-3 images (see Tables 4 and 5). Since a primary objective of the *USTFIP* method is to reduce input requirements by eliminating the need for matching pairs of images, it is specifically designed to execute the *LOP* step from upscaled images. Interestingly, our previous experiments showed that even when the original coarse imagery is available, the predictions derived from upscaled images are superior in most cases. Our method may also have problems retrieving abrupt changes, as illustrated by the flooded area from Fig. 5. One way to overcome this potential problem is by considering a baseline period containing information about these changes. In cases where these changes are periodic, extending the baseline to include the same period from previous years may facilitate the fusion.

We also assessed the impact of false negatives on predictions by simulating false negatives in the cloud masks. Our results suggested that they have little-to-no effect on the second predictions. In fact, to mitigate this issue, our method avoids areas around clouds in the *CH* step, and to some extent when selecting optimal information in the *LOP* step. To have an effect on the second prediction, the false negatives would need to be isolated from the predicted clouds while still being the most similar to the observation from the target date.

The above considerations highlight the general suitability of our method. Nevertheless, the method may be better suited for certain types of problems than for others, depending on the requirements of the particular problem. Firstly, while our method is well-suited for applications that do not need a dense temporal resolution, such as agriculture or geology, it may be less optimal for applications that require denser temporal resolutions. In such cases, it is advisable to explore alternative models that can accommodate the desired temporal resolutions. Despite this, it is important to acknowledge that our method leverages the best available temporal resolution between two series of images, as is usual in *STIF* methods, which is typically predetermined by the satellite programs. Secondly, our method excels in making predictions for various types of image time-series, especially in cloud-prone regions where finding cloud-free images can be challenging. However, in regions with completely cloud-free date ranges, alternative methods that operate under the assumption of cloud-free data may be better suited. In such cases, leveraging the data from the entire time-series might provide minimal additional information compared to the nearest cloud-free image, while requiring the processing of significantly larger amounts of data. Thus, for regions with consistently cloud-free date ranges, methods that specifically consider the cloud-free assumption may offer more efficiency and optimal results.

6. Conclusions

Spatio-temporal fusion methods usually impose severe restrictions on the input data, such as the need for clear-sky images and finding matching pairs of images from the involved sensors. These restrictions, along with long computational times, prevent wider adoption of fusion methods in applied research. As a consequence, a significant amount of valuable data is discarded, which contravenes a fundamental statistical principle of utilizing all available data. Our research explored the adaptation and combination of two strategies to mitigate these restrictions on the inputs; the selection of local optimal information and an information degradation model to generate synthetic coarse-scale images. This innovative combination of methods produces a new methodology for selecting valuable data from time series of images that circumvents misclassified clouds and eliminates the requirement for matching coarse-scale images. Both strategies are integrated with the *Fit-FC* fusion method to create a novel approach termed Unpaired Spatio-Temporal Fusion of Image Patches (*USTFIP*). *USTFIP* is able to apply these component methods in parallel to increase computational efficiency. *USTFIP* and *Fit-FC* were compared as benchmarks in a simulation exercise where scenarios assumed that finding matching pairs of cloud-free images is progressively more difficult, so they are from a date further apart from the prediction date.

Our results demonstrate that *USTFIP* makes the fusion more convenient, accurate, robust and efficient. Depending on the boundary conditions, relaxing the input data requirements reduces the error between 2 and 13%. As the fusion becomes more challenging in terms of cloud frequency and lack of matching pairs, the proposed *USTFIP* preserves accuracy to a larger extent than the original *Fit-FC* method. In the experiments, the *RMSE* from *Fit-FC* increased by 21% and 16% versus the 10% and 4% for *USTFIP* in Coleambally and Gwydir, respectively, as the distance between the reference pair and the target image expands. For the Sentinel-2 time series, the increases were 13% and 14% for *Fit-FC* and 5% and 4% for *USTFIP*. The flexibility achieved through optimal data selection and coarse image generation required additional calculations. However, the results here suggest that advanced programming techniques could not only compensate for the additional computational cost, but also reduce running times relative to *Fit-FC*. In our case, parallelization saved up to 40% of the computational time. Further experiments are needed to quantify the benefits of *USTFIP* to a wide range of case studies.

CRediT authorship contribution statement

Harkaitz Goyena: Validation, Formal analysis, Investigation, Data curation and coding, Review & editing. **Unai Pérez-Goya:** Conceptualization, Methodology, Coding, Review & editing. **Manuel Montesino-SanMartin:** Investigation, Data curation and coding, Original draft. **Ana F. Militino:** Conceptualization, Research organization, Coding revision, Review & editing. **Qunming Wang:** Methodology, Writing – review & editing. **Peter M. Atkinson:** Methodology, Conceptualization, Writing – review & editing. **M. Dolores Ugarte:** Conceptualization, Methodology, Research organization, Review & Editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing fiinancial interests or personal relationships that could have influenced the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was started as part of a visit by Unai Pérez-Goya to Lancaster Environment Centre in Lancaster University, under the supervision of Peter M. Atkinson.

We would like to express our gratitude to the editor and four reviewers for their valuable comments, which have greatly contributed to enhancing an earlier version of this manuscript. Open access provided by Universidad Pública de Navarra.

Funding

This research was supported by the Spanish Research Agency and Next Generation EU (PDC2021-120796-I00 project) and by the Spanish Research Agency (PID 2020-113125RB-I00/MCIN/AEI/10.13039/501100011033 project).

This work was supported by the National Natural Science Foundation of China under Grant 42222108.

Appendix A. Simulated cloud masks

The cloud masks for Coleambally and Gwydir were interpreted from the quality band (QA) of Landsat images. These Landsat images were different from those in Table 3. The masks represent actual clouds observed in Coleambally and Gwydir during 2019–2020. Figs. A.7 and A.8 show the fraction of the image covered (white area) and the shape or distribution of the clouds.

The cloud masks for the North Dakota regions represent actual clouds from Sentinel-2 images. Fig. A.9 show the fraction of the image covered (white area) and the shape or distribution of the clouds.

Tables A.6, A.7, A.8 and A.9 show the average quality metrics for each scenario conditional on the number of available observations for each pixel. For example, $GapN = 1$ and $nobs = 2$ are the average metrics across all the pixel locations for which there are two available observations in the baseline for scenario 1.

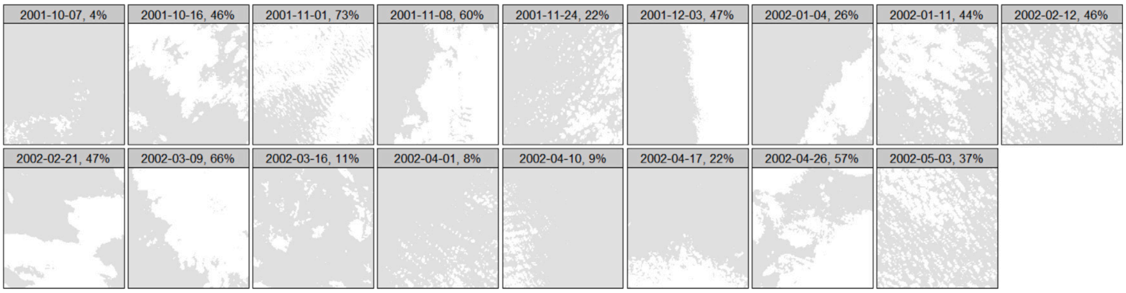

Fig. A.7. Cloud-masks to simulate the presence of clouds in Coleambally. White and gray pixels represent cloud-covered and clear-sky areas respectively. The date of the mask is specified on the top panel together with the fraction of missing pixels.

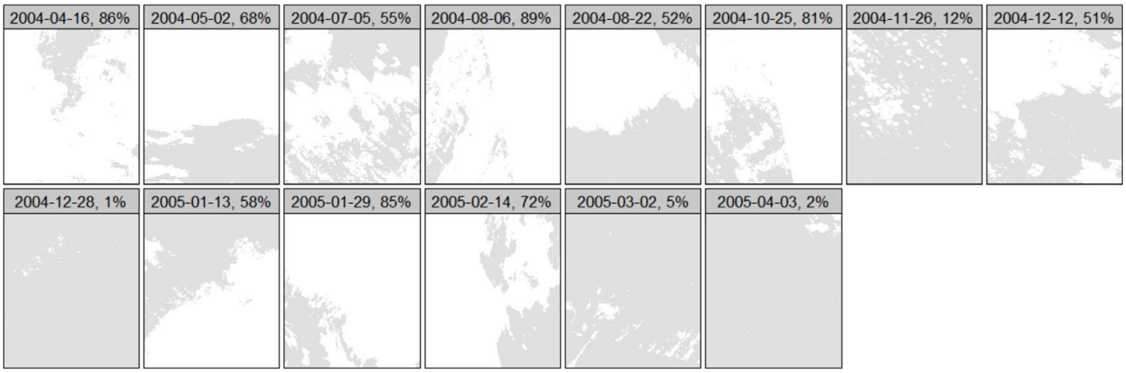

Fig. A.8. Cloud-masks to simulate the presence of clouds in Gwydir. White and gray pixels represent cloud-covered and clear-sky areas respectively. The date of the mask is specified on the top panel together with the fraction of missing pixels.

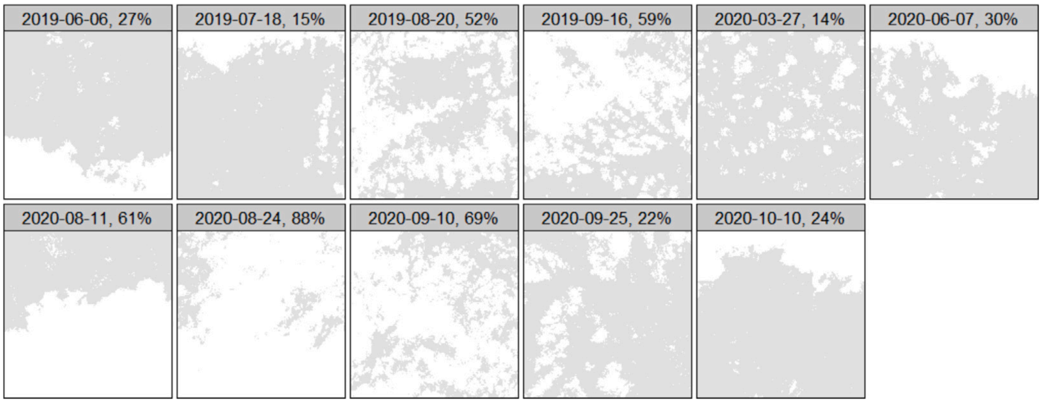

Fig. A.9. Cloud-masks to simulate the presence of clouds in the North Dakota sites. White and gray pixels represent cloud-covered and clear-sky areas respectively. The date of the mask is specified on the top panel together with the fraction of missing pixels.

Table A.6

Averages of the quality metrics for each scenario *GapN* and number of clean observation *nobs* in Coleambally. *RMSE* is the root mean squared error of the prediction either from *FSDAF*, *USTFIP*, *Fit-FC* (*FFC*) or *STARFM*. *Edge* is the average normalized difference in the Robert's Edge spatial feature of the prediction either from *FSDAF*, *USTFIP*, *Fit-FC* (*FFC*) or *STARFM*.

GapN	nobs	Gap	RMSE$_{FSDAF}$	RMSE$_{USTFIP}$	RMSE$_{FFC}$	RMSE$_{STARFM}$	Edge$_{FSDAF}$	Edge$_{USTFIP}$	Edge$_{FFC}$	Edge$_{STARFM}$
no	no	days	–	–	–	–	–	–	–	–
1	1	28	0.0296	0.0272	0.0278	0.0311	−0.2210	−0.1108	−0.1008	−0.2564
	2	28	0.0292	0.0258	0.0276	0.0307	−0.1996	−0.0818	−0.0876	−0.2488
2	1	42	0.0322	0.0290	0.0296	0.0340	−0.2675	−0.1376	−0.1222	−0.2964
	2	42	0.0328	0.0284	0.0306	0.0357	−0.2400	−0.1004	−0.1063	−0.2832
	3	42	0.0315	0.0268	0.0289	0.0340	−0.2244	−0.0842	−0.0872	−0.2759
3	1	57	0.0328	0.0297	0.0289	0.0369	−0.2755	−0.1386	−0.1217	−0.3141
	2	57	0.0358	0.0291	0.0327	0.0399	−0.2742	−0.1166	−0.1070	−0.3103
	3	57	0.0347	0.0280	0.0314	0.0387	−0.2520	−0.0951	−0.0939	−0.3015
	4	57	0.0338	0.0264	0.0297	0.0369	−0.2368	−0.0839	−0.0816	−0.2914
4	1	72	0.0353	0.0294	0.0286	0.0380	−0.2602	−0.1167	−0.0902	−0.3200
	2	72	0.0366	0.0293	0.0320	0.0412	−0.2871	−0.1221	−0.1046	−0.3194
	3	72	0.0372	0.0287	0.0323	0.0422	−0.2778	−0.1069	−0.0993	−0.3234
	4	72	0.0382	0.0289	0.0330	0.0433	−0.2723	−0.0984	−0.0941	−0.3246
	5	72	0.0365	0.0257	0.0300	0.0399	−0.2693	−0.0864	−0.0864	−0.3180
5	1	82	0.0333	0.0304	0.0289	0.0390	−0.3168	−0.1710	−0.1098	−0.3606
	2	86	0.0387	0.0288	0.0316	0.0436	−0.2944	−0.1287	−0.1063	−0.3225
	3	86	0.0397	0.0291	0.0330	0.0454	−0.2957	−0.1179	−0.1044	−0.3338
	4	86	0.0403	0.0285	0.0331	0.0459	−0.2809	−0.1005	−0.0939	−0.3293
	5	86	0.0421	0.0290	0.0342	0.0476	−0.2822	−0.0950	−0.0914	−0.3322
	6	86	0.0383	0.0249	0.0295	0.0417	−0.2796	−0.0876	−0.0813	−0.3172

Table A.7

Averages of the quality metrics for each scenario *GapN* and number of clean observation *nobs* in Gwydir. *RMSE* is the root mean squared error of the prediction either from *FSDAF*, *USTFIP*, *Fit-FC* (*FFC*) or *STARFM*. *Edge* is the average normalized difference in the Robert's Edge spatial feature of the prediction either from *FSDAF*, *USTFIP*, *Fit-FC* (*FFC*) or *STARFM*.

GapN	nobs	Gap	RMSE$_{FSDAF}$	RMSE$_{USTFIP}$	RMSE$_{FFC}$	RMSE$_{STARFM}$	Edge$_{FSDAF}$	Edge$_{USTFIP}$	Edge$_{FFC}$	Edge$_{STARFM}$
no	no	days	–	–	–	–	–	–	–	–
1	1	44	0.0279	0.0248	0.0245	0.0278	−0.3243	−0.1485	−0.1347	−0.2747
	2	44	0.0268	0.0223	0.0231	0.0265	−0.3144	−0.1109	−0.1149	−0.2588
2	1	68	0.0287	0.0244	0.0244	0.0302	−0.3679	−0.1331	−0.1255	−0.3283
	2	68	0.0299	0.0235	0.0253	0.0307	−0.3659	−0.1504	−0.1662	−0.3199
	3	68	0.0286	0.0223	0.0246	0.0291	−0.3174	−0.1097	−0.1333	−0.2840
3	1	96	0.0295	0.0248	0.0248	0.0329	−0.4207	−0.1992	−0.1635	−0.3797
	2	94	0.0317	0.0244	0.0269	0.0330	−0.4052	−0.1736	−0.1842	−0.3448
	3	94	0.0304	0.0226	0.0256	0.0313	−0.3970	−0.1520	−0.1715	−0.3283
	4	94	0.0297	0.0217	0.0251	0.0299	−0.3658	−0.1224	−0.1499	−0.3029
4	1	128	0.0313	0.0251	0.0247	0.0340	−0.4436	−0.2207	−0.2451	−0.4153
	2	126	0.0331	0.0248	0.0276	0.0352	−0.4442	−0.1862	−0.1951	−0.3774
	3	126	0.0322	0.0236	0.0268	0.0338	−0.4357	−0.1753	−0.1933	−0.3633
	4	126	0.0313	0.0223	0.0257	0.0319	−0.4314	−0.1582	−0.1731	−0.3419
	5	126	0.0313	0.0215	0.0256	0.0307	−0.4063	−0.1406	−0.1651	−0.3260
5	1	166	0.0315	0.0258	0.0250	0.0330	−0.4510	−0.2644	−0.2269	−0.4340
	2	158	0.0335	0.0249	0.0282	0.0360	−0.4342	−0.1798	−0.1768	−0.3856
	3	158	0.0324	0.0239	0.0271	0.0349	−0.4463	−0.1837	−0.1942	−0.3801
	4	158	0.0324	0.0229	0.0266	0.0339	−0.4301	−0.1688	−0.1712	−0.3578
	5	158	0.0311	0.0215	0.0258	0.0323	−0.4225	−0.1505	−0.1593	−0.3440
	6	158	0.0294	0.0207	0.0246	0.0303	−0.3898	−0.0843	−0.1697	−0.3164

Table A.8

Averages of the quality metrics for each scenario *GapN* and number of clean observation *nobs* in Site 1 in North Dakota. *RMSE* is the root mean squared error of the prediction either from *FSDAF*, *USTFIP*, *Fit-FC* (*FFC*) or *STARFM*. *Edge* is the average normalized difference in the Robert's Edge spatial feature of the prediction either from *FSDAF*, *USTFIP*, *Fit-FC* (*FFC*) or *STARFM*.

GapN	nobs	Gap	RMSE_{FSDAF}	RMSE_{USTFIP}	RMSE_{FFC}	RMSE_{STARFM}	Edge_{FSDAF}	Edge_{USTFIP}	Edge_{FFC}	Edge_{STARFM}
no	no	days	–	–	–	–	–	–	–	–
1	1	61	0.0487	0.0450	0.0456	0.0514	−0.2728	−0.1227	−0.1609	−0.2921
	2	61	0.0488	0.0447	0.0459	0.0513	−0.2669	−0.1265	−0.1636	−0.2873
2	1	133	0.0501	0.0463	0.0468	0.0536	−0.2414	−0.1473	−0.2039	−0.2951
	2	133	0.0495	0.0461	0.0474	0.0529	−0.2383	−0.1419	−0.1921	−0.2834
	3	133	0.0499	0.0461	0.0466	0.0533	−0.2203	−0.1392	−0.1798	−0.2794
3	1	207	0.0518	0.0470	0.0485	0.0546	−0.2123	−0.1201	−0.1692	−0.2718
	2	207	0.0509	0.0460	0.0472	0.0541	−0.2091	−0.1371	−0.1753	−0.2674
	3	207	0.0512	0.0458	0.0469	0.0545	−0.2108	−0.1347	−0.1720	−0.2622
	4	207	0.0505	0.0447	0.0454	0.0531	−0.2062	−0.1147	−0.1715	−0.2563
4	1	285	0.0516	0.0478	0.0484	0.0533	−0.2090	−0.1719	−0.2035	−0.2656
	2	285	0.0497	0.0465	0.0472	0.0522	−0.2156	−0.1685	−0.2043	−0.2691
	3	285	0.0495	0.0464	0.0473	0.0521	−0.2035	−0.1636	−0.1999	−0.2598
	4	285	0.0493	0.0463	0.0473	0.0517	−0.2013	−0.1833	−0.2054	−0.2566
	5	285	0.0485	0.0450	0.0465	0.0505	−0.1937	−0.1560	−0.1953	−0.2399
5	1	359	0.0549	0.0497	0.0551	0.0555	−0.2346	−0.1625	−0.2321	−0.2500
	2	359	0.0511	0.0477	0.0532	0.0525	−0.2408	−0.1773	−0.2342	−0.2613
	3	359	0.0506	0.0477	0.0524	0.0522	−0.2341	−0.1822	−0.2242	−0.2522
	4	359	0.0498	0.0471	0.0516	0.0515	−0.2305	−0.1891	−0.2290	−0.2497
	5	359	0.0489	0.0462	0.0508	0.0506	−0.2239	−0.1923	−0.2219	−0.2489
	6	359	0.0509	0.0480	0.0497	0.0523	−0.1908	−0.1154	−0.1572	−0.2396

Table A.9

Averages of the quality metrics for each scenario *GapN* and number of clean observation *nobs* in Site 2 in North Dakota. *RMSE* is the root mean squared error of the prediction either from *FSDAF*, *USTFIP*, *Fit-FC* (*FFC*) or *STARFM*. *Edge* is the average normalized difference in the Robert's Edge spatial feature of the prediction either from *FSDAF*, *USTFIP*, *Fit-FC* (*FFC*) or *STARFM*.

GapN	nobs	Gap	RMSE_{FSDAF}	RMSE_{USTFIP}	RMSE_{FFC}	RMSE_{STARFM}	Edge_{FSDAF}	Edge_{USTFIP}	Edge_{FFC}	Edge_{STARFM}
no	no	days	–	–	–	–	–	–	–	–
1	1	61	0.0465	0.0418	0.0430	0.0502	−0.2373	−0.0565	−0.1476	−0.2942
	2	61	0.0463	0.0416	0.0425	0.0499	−0.2270	−0.0591	−0.1485	−0.2952
2	1	133	0.0495	0.0426	0.0438	0.0534	−0.2190	−0.0599	−0.1395	−0.2901
	2	133	0.0487	0.0421	0.0429	0.0528	−0.2119	−0.0618	−0.1408	−0.2861
	3	133	0.0490	0.0431	0.0443	0.0530	−0.1961	−0.0830	−0.1611	−0.2764
3	1	207	0.0478	0.0411	0.0421	0.0520	−0.2200	−0.1016	−0.1621	−0.2939
	2	207	0.0482	0.0431	0.0441	0.0516	−0.2044	−0.0968	−0.1608	−0.2613
	3	207	0.0476	0.0426	0.0434	0.0514	−0.2039	−0.1057	−0.1665	−0.2668
	4	207	0.0495	0.0454	0.0462	0.0529	−0.1819	−0.0934	−0.1491	−0.2358
4	1	285	0.0453	0.0416	0.0423	0.0481	−0.2244	−0.0915	−0.1681	−0.2877
	2	285	0.0458	0.0425	0.0435	0.0481	−0.1967	−0.0843	−0.1689	−0.2512
	3	285	0.0457	0.0429	0.0439	0.0481	−0.1927	−0.1011	−0.1746	−0.2502
	4	285	0.0463	0.0432	0.0439	0.0490	−0.1879	−0.0796	−0.1565	−0.2455
	5	285	0.0500	0.0463	0.0472	0.0526	−0.1581	−0.0729	−0.1320	−0.2259
5	1	359	0.0478	0.0437	0.0461	0.0511	−0.2034	−0.1106	−0.1467	−0.2734
	2	359	0.0465	0.0429	0.0469	0.0485	−0.1961	−0.1170	−0.1810	−0.2383
	3	359	0.0460	0.0430	0.0470	0.0476	−0.1987	−0.1245	−0.1922	−0.2367
	4	359	0.0461	0.0432	0.0479	0.0479	−0.2001	−0.1200	−0.1846	−0.2347
	5	359	0.0481	0.0444	0.0499	0.0497	−0.1872	−0.1121	−0.1757	−0.2244
	6	359	0.0523	0.0470	0.0522	0.0536	−0.1505	−0.1031	−0.1559	−0.1979

Appendix B. Band metrics

Tables B.10, B.11, B.12 and B.13, show the averages of the quality metrics by band in the study regions. The general conclusions from Section 4 are well represented for most of the bands.

Table B.10
Averages of the quality metrics by band for each scenario in Site 1 in North Dakota, where *GapN* is the number of cloud-covered scenes between the reference (clear-sky) and the target image. CC_{rt} is the correlation coefficient between the reference and target images. *RMSE* is the root mean squared error of the prediction either from *Fit-FC* (*FFC*), *USTFIP*, *STARFM* or *FSDAF*. *Edge* is the spatial accuracy regarding the Robert's Edge feature of the prediction either from *Fit-FC* (*FFC*), *USTFIP*, *STARFM* or *FSDAF*.

GapN	Band	CC$_{rt}$	RMSE$_{FFC}$	RMSE$_{USTFIP}$	RMSE$_{STARFM}$	RMSE$_{FSDAF}$	Edge$_{FFC}$	Edge$_{USTFIP}$	Edge$_{STARFM}$	Edge$_{FSDAF}$
no	–	–	–	–	–	–	–	–	–	–
1	Red	0.19	0.0483	0.0486	0.0537	0.0537	−0.1898	−0.2074	−0.2881	−0.3488
	Green	0.32	0.0336	0.0316	0.0367	0.0367	−0.1863	−0.1295	−0.2750	−0.2432
	Blue	0.25	0.0307	0.0283	0.0321	0.0321	−0.1795	−0.0143	−0.3007	−0.2406
	NIR	0.21	0.0697	0.0703	0.0821	0.0821	−0.1053	−0.1525	−0.3027	−0.2597
2	Red	0.00	0.0512	0.0509	0.0565	0.0565	−0.2063	−0.2112	−0.2815	−0.2998
	Green	0.18	0.0343	0.0332	0.0370	0.0370	−0.2177	−0.1631	−0.2513	−0.1947
	Blue	0.12	0.0325	0.0291	0.0340	0.0340	−0.2185	−0.0320	−0.2812	−0.1864
	NIR	0.11	0.0695	0.0708	0.0847	0.0847	−0.1243	−0.1564	−0.3255	−0.2490
3	Red	0.06	0.0505	0.0500	0.0541	0.0541	−0.2023	−0.2093	−0.2282	−0.2380
	Green	0.08	0.0351	0.0337	0.0392	0.0392	−0.1809	−0.1251	−0.2372	−0.2095
	Blue	0.09	0.0318	0.0291	0.0356	0.0356	−0.2083	−0.0254	−0.2468	−0.1486
	NIR	0.04	0.0712	0.0709	0.0876	0.0876	−0.1192	−0.1714	−0.3356	−0.2355
4	Red	0.18	0.0501	0.0496	0.0517	0.0517	−0.2234	−0.2336	−0.2211	−0.2156
	Green	0.12	0.0348	0.0339	0.0372	0.0372	−0.2153	−0.1744	−0.2534	−0.2181
	Blue	0.14	0.0315	0.0289	0.0342	0.0342	−0.2358	−0.0565	−0.2589	−0.1720
	NIR	0.03	0.0725	0.0731	0.0838	0.0838	−0.1424	−0.2094	−0.3004	−0.2065
5	Red	0.16	0.0546	0.0502	0.0531	0.0531	−0.2383	−0.2528	−0.2132	−0.2515
	Green	0.12	0.0379	0.0341	0.0382	0.0382	−0.2424	−0.1875	−0.2560	−0.2336
	Blue	0.11	0.0307	0.0290	0.0331	0.0331	−0.2326	−0.0793	−0.2494	−0.2112
	NIR	0.09	0.0834	0.0756	0.0823	0.0823	−0.1912	−0.2241	−0.2834	−0.2240

Table B.11
Averages of the quality metrics by band for each scenario in Site 2 in North Dakota, where *GapN* is the number of cloud-covered scenes between the reference (clear-sky) and the target image. CC_{rt} is the correlation coefficient between the reference and target images. *RMSE* is the root mean squared error of the prediction either from *Fit-FC* (*FFC*), *USTFIP*, *STARFM* or *FSDAF*. *Edge* is the spatial accuracy regarding the Robert's Edge feature of the prediction either from *Fit-FC* (*FFC*), *USTFIP*, *STARFM* or *FSDAF*.

GapN	Band	CC$_{rt}$	RMSE$_{FFC}$	RMSE$_{USTFIP}$	RMSE$_{STARFM}$	RMSE$_{FSDAF}$	Edge$_{FFC}$	Edge$_{USTFIP}$	Edge$_{STARFM}$	Edge$_{FSDAF}$
no	–	–	–	–	–	–	–	–	–	–
1	Red	0.19	0.0483	0.0486	0.0537	0.0537	−0.1898	−0.2074	−0.2881	−0.3488
	Green	0.32	0.0336	0.0316	0.0367	0.0367	−0.1863	−0.1295	−0.2750	−0.2432
	Blue	0.25	0.0307	0.0283	0.0321	0.0321	−0.1795	−0.0143	−0.3007	−0.2406
	NIR	0.21	0.0697	0.0703	0.0821	0.0821	−0.1053	−0.1525	−0.3027	−0.2597
2	Red	0.00	0.0512	0.0509	0.0565	0.0565	−0.2063	−0.2112	−0.2815	−0.2998
	Green	0.18	0.0343	0.0332	0.0370	0.0370	−0.2177	−0.1631	−0.2513	−0.1947
	Blue	0.12	0.0325	0.0291	0.0340	0.0340	−0.2185	−0.0320	−0.2812	−0.1864
	NIR	0.11	0.0695	0.0708	0.0847	0.0847	−0.1243	−0.1564	−0.3255	−0.2490
3	Red	0.06	0.0505	0.0500	0.0541	0.0541	−0.2023	−0.2093	−0.2282	−0.2380
	Green	0.08	0.0351	0.0337	0.0392	0.0392	−0.1809	−0.1251	−0.2372	−0.2095
	Blue	0.09	0.0318	0.0291	0.0356	0.0356	−0.2083	−0.0254	−0.2468	−0.1486
	NIR	0.04	0.0712	0.0709	0.0876	0.0876	−0.1192	−0.1714	−0.3356	−0.2355
4	Red	0.18	0.0501	0.0496	0.0517	0.0517	−0.2234	−0.2336	−0.2211	−0.2156
	Green	0.12	0.0348	0.0339	0.0372	0.0372	−0.2153	−0.1744	−0.2534	−0.2181
	Blue	0.14	0.0315	0.0289	0.0342	0.0342	−0.2358	−0.0565	−0.2589	−0.1720
	NIR	0.03	0.0725	0.0731	0.0838	0.0838	−0.1424	−0.2094	−0.3004	−0.2065
5	Red	0.16	0.0546	0.0502	0.0531	0.0531	−0.2383	−0.2528	−0.2132	−0.2515
	Green	0.12	0.0379	0.0341	0.0382	0.0382	−0.2424	−0.1875	−0.2560	−0.2336
	Blue	0.11	0.0307	0.0290	0.0331	0.0331	−0.2326	−0.0793	−0.2494	−0.2112
	NIR	0.09	0.0834	0.0756	0.0823	0.0823	−0.1912	−0.2241	−0.2834	−0.2240

Table B.12
Averages of the quality metrics by band for each scenario in Coleambally, where *GapN* is the number of cloud-covered scenes between the reference (clear-sky) and the target image. CC_{rt} is the correlation coefficient between the reference and target images. *RMSE* is the root mean squared error of the prediction either from *Fit-FC* (*FFC*), *USTFIP*, *STARFM* or *FSDAF*. *Edge* is the spatial accuracy regarding the Robert's Edge feature of the prediction either from *Fit-FC* (*FFC*), *USTFIP*, *STARFM* or *FSDAF*.

GapN	Band	CC$_{rt}$	RMSE$_{FFC}$	RMSE$_{USTFIP}$	RMSE$_{STARFM}$	RMSE$_{FSDAF}$	Edge$_{FFC}$	Edge$_{USTFIP}$	Edge$_{STARFM}$	Edge$_{FSDAF}$
no	–	–	–	–	–	–	–	–	–	–
1	Red	0.73	0.0128	0.0117	0.0145	0.0145	−0.0661	−0.0840	−0.2710	−0.2251
	Green	0.69	0.0158	0.0148	0.0184	0.0184	−0.0826	−0.1050	−0.2635	−0.2197
	Blue	0.75	0.0241	0.0226	0.0276	0.0276	−0.0984	−0.1089	−0.2493	−0.2040
	NIR	0.60	0.0420	0.0401	0.0470	0.0470	−0.0833	−0.0554	−0.2495	−0.1931
	SWIR1	0.85	0.0388	0.0369	0.0426	0.0426	−0.0968	−0.0776	−0.2342	−0.1865
	SWIR2	0.87	0.0352	0.0337	0.0373	0.0373	−0.1047	−0.0910	−0.2211	−0.1762
2	Red	0.65	0.0132	0.0121	0.0153	0.0153	−0.0696	−0.0814	−0.2921	−0.2456
	Green	0.60	0.0170	0.0155	0.0206	0.0206	−0.0915	−0.1069	−0.2952	−0.2553
	Blue	0.64	0.0270	0.0243	0.0325	0.0325	−0.1061	−0.1106	−0.2855	−0.2494
	NIR	0.44	0.0450	0.0424	0.0554	0.0554	−0.0966	−0.0668	−0.2905	−0.2277
	SWIR1	0.82	0.0410	0.0381	0.0471	0.0471	−0.1090	−0.0816	−0.2657	−0.2083
	SWIR2	0.83	0.0378	0.0350	0.0412	0.0412	−0.1197	−0.0960	−0.2492	−0.1953
3	Red	0.58	0.0140	0.0123	0.0168	0.0168	−0.0543	−0.0785	−0.3041	−0.2564
	Green	0.51	0.0180	0.0157	0.0227	0.0227	−0.0766	−0.1053	−0.3102	−0.2697
	Blue	0.52	0.0291	0.0249	0.0370	0.0370	−0.0917	−0.1108	−0.3053	−0.2690
	NIR	0.30	0.0472	0.0437	0.0636	0.0636	−0.0866	−0.0715	−0.3125	−0.2419
	SWIR1	0.78	0.0432	0.0388	0.0517	0.0517	−0.1055	−0.0813	−0.2913	−0.2260
	SWIR2	0.80	0.0405	0.0359	0.0455	0.0455	−0.1171	−0.0960	−0.2716	−0.2123
4	Red	0.50	0.0146	0.0124	0.0182	0.0182	−0.0509	−0.0835	−0.3189	−0.2744
	Green	0.41	0.0183	0.0159	0.0243	0.0243	−0.0741	−0.1104	−0.3270	−0.2877
	Blue	0.40	0.0297	0.0250	0.0409	0.0409	−0.0960	−0.1154	−0.3248	−0.2911
	NIR	0.18	0.0470	0.0442	0.0698	0.0698	−0.0871	−0.0771	−0.3369	−0.2645
	SWIR1	0.74	0.0457	0.0398	0.0570	0.0570	−0.1152	−0.0858	−0.3200	−0.2503
	SWIR2	0.74	0.0433	0.0368	0.0507	0.0507	−0.1226	−0.0977	−0.2988	−0.2376
5	Red	0.41	0.0147	0.0124	0.0196	0.0196	−0.0467	−0.0855	−0.3234	−0.2836
	Green	0.31	0.0185	0.0159	0.0260	0.0260	−0.0747	−0.1149	−0.3304	−0.2956
	Blue	0.29	0.0298	0.0250	0.0443	0.0443	−0.0956	−0.1192	−0.3295	−0.2994
	NIR	0.07	0.0470	0.0443	0.0751	0.0751	−0.0908	−0.0816	−0.3491	−0.2754
	SWIR1	0.68	0.0480	0.0405	0.0624	0.0624	−0.1167	−0.0888	−0.3335	−0.2640
	SWIR2	0.67	0.0458	0.0374	0.0561	0.0561	−0.1244	−0.1010	−0.3141	−0.2544

Table B.13
Averages of the quality metrics by band for each scenario in Gwydir, where *GapN* is the number of cloud-covered scenes between the reference (clear-sky) and the target image. CC_{rt} is the correlation coefficient between the reference and target images. *RMSE* is the root mean squared error of the prediction either from *Fit-FC* (*FFC*), *USTFIP*, *STARFM* or *FSDAF*. *Edge* is the spatial accuracy regarding the Robert's Edge feature of the prediction either from *Fit-FC* (*FFC*), *USTFIP*, *STARFM* or *FSDAF*.

GapN	Band	CC$_{rt}$	RMSE$_{FFC}$	RMSE$_{USTFIP}$	RMSE$_{STARFM}$	RMSE$_{FSDAF}$	Edge$_{FFC}$	Edge$_{USTFIP}$	Edge$_{STARFM}$	Edge$_{FSDAF}$
no	–	–	–	–	–	–	–	–	–	–
1	Red	0.58	0.0105	0.0100	0.0118	0.0118	−0.1086	−0.1098	−0.2700	−0.2976
	Green	0.56	0.0137	0.0132	0.0155	0.0155	−0.1044	−0.0965	−0.2487	−0.2893
	Blue	0.53	0.0182	0.0175	0.0208	0.0208	−0.1021	−0.0946	−0.2475	−0.2903
	NIR	0.47	0.0321	0.0313	0.0408	0.0408	−0.1727	−0.1799	−0.3231	−0.4037
	SWIR1	0.53	0.0354	0.0349	0.0401	0.0401	−0.1342	−0.1405	−0.2763	−0.3685
	SWIR2	0.57	0.0313	0.0305	0.0349	0.0349	−0.1445	−0.1472	−0.2676	−0.3261
2	Red	0.52	0.0114	0.0102	0.0129	0.0129	−0.1438	−0.1210	−0.3061	−0.3290
	Green	0.53	0.0148	0.0134	0.0166	0.0166	−0.1384	−0.1091	−0.2785	−0.3010
	Blue	0.48	0.0194	0.0178	0.0228	0.0228	−0.1174	−0.1030	−0.2820	−0.3139
	NIR	0.27	0.0349	0.0327	0.0500	0.0500	−0.2008	−0.1983	−0.3881	−0.4590
	SWIR1	0.48	0.0373	0.0353	0.0432	0.0432	−0.1600	−0.1502	−0.3110	−0.3779
	SWIR2	0.49	0.0336	0.0313	0.0386	0.0386	−0.1653	−0.1575	−0.3059	−0.3687
3	Red	0.46	0.0118	0.0103	0.0133	0.0133	−0.1689	−0.1447	−0.3298	−0.3635
	Green	0.50	0.0151	0.0135	0.0170	0.0170	−0.1588	−0.1293	−0.2933	−0.3336
	Blue	0.43	0.0200	0.0180	0.0237	0.0237	−0.1269	−0.1140	−0.2963	−0.3536
	NIR	0.09	0.0369	0.0331	0.0546	0.0546	−0.2198	−0.2125	−0.4205	−0.5037
	SWIR1	0.41	0.0388	0.0357	0.0447	0.0447	−0.1708	−0.1697	−0.3222	−0.4387
	SWIR2	0.40	0.0351	0.0317	0.0405	0.0405	−0.1789	−0.1736	−0.3220	−0.4064
4	Red	0.39	0.0122	0.0104	0.0142	0.0142	−0.1754	−0.1511	−0.3615	−0.4128
	Green	0.44	0.0159	0.0136	0.0181	0.0181	−0.1675	−0.1377	−0.3270	−0.3837
	Blue	0.39	0.0204	0.0180	0.0248	0.0248	−0.1435	−0.1285	−0.3247	−0.3963
	NIR	0.03	0.0381	0.0335	0.0564	0.0564	−0.2360	−0.2289	−0.4379	−0.5159
	SWIR1	0.38	0.0394	0.0357	0.0465	0.0465	−0.1899	−0.1862	−0.3460	−0.4470
	SWIR2	0.34	0.0363	0.0317	0.0434	0.0434	−0.1972	−0.1872	−0.3506	−0.4409
5	Red	0.33	0.0122	0.0103	0.0148	0.0148	−0.1623	−0.1453	−0.3795	−0.4295
	Green	0.40	0.0158	0.0135	0.0187	0.0187	−0.1599	−0.1352	−0.3438	−0.3914
	Blue	0.34	0.0206	0.0178	0.0258	0.0258	−0.1362	−0.1289	−0.3372	−0.3996
	NIR	−0.03	0.0387	0.0336	0.0574	0.0574	−0.2350	−0.2344	−0.4367	−0.5207
	SWIR1	0.35	0.0398	0.0355	0.0465	0.0465	−0.1812	−0.1811	−0.3425	−0.4208
	SWIR2	0.27	0.0369	0.0317	0.0446	0.0446	−0.1854	−0.1846	−0.3553	−0.4344

References

Addink, E., Stein, A., 1999. A comparison of conventional and geostatistical methods to replace clouded pixels in NOAA-AVHRR images. International Journal of Remote Sensing 20 (5), 961–977.

Arvidson, T., Goward, S., Gasch, J., Williams, D., 2006. Landsat-7 long-term acquisition plan. Photogramm. Eng. Remote Sens. 72 (10), 1137–1146.

Belgiu, M., Stein, A., 2019. Spatiotemporal image fusion in remote sensing. Remote Sens. 11 (7), 818.

Borini Alves, D., Montorio Llovería, R., Pérez-Cabello, F., Vlassova, L., 2018. Fusing landsat and MODIS data to retrieve multispectral information from fire-affected areas over tropical savannah environments in the Brazilian Amazon. Int. J. Remote Sens. 39 (22), 7919–7941.

Chen, Y., Cao, R., Chen, J., Zhu, X., Zhou, J., Wang, G., Shen, M., Chen, X., Yang, W., 2020. A new cross-fusion method to automatically determine the optimal input image pairs for NDVI spatiotemporal data fusion. IEEE Trans. Geosci. Remote Sens..

Chen, B., Huang, B., Xu, B., 2015. Comparison of spatiotemporal fusion models: A review. Remote Sens. 7 (2), 1798–1835.

Chen, Y., Shi, K., Ge, Y., Zhou, Y., 2021. Spatiotemporal remote sensing image fusion using multiscale two-stream convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 60, 1–12.

Chen, X., Vierling, L., Deering, D., 2005. A simple and effective radiometric correction method to improve landscape change detection across sensors and across time. Remote Sens. Environ. 98 (1), 63–79.

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.

Das, M., Ghosh, S.K., 2020. Data-driven approaches for spatio-temporal analysis: A survey of the state-of-the-arts. J. Comput. Sci. Tech. 35, 665–696.

Donaldson, D., Storeygard, A., 2016. The view from above: Applications of satellite data in economics. J. Econ. Perspect. 30 (4), 171–198.

Dong, T., Liu, J., Qian, B., Zhao, T., Jing, Q., Geng, X., Wang, J., Huffman, T., Shang, J., 2016. Estimating winter wheat biomass by assimilating leaf area index derived from fusion of landsat-8 and MODIS data. Int. J. Appl. Earth Obs. Geoinf. 49, 63–74.

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.

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.

Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Li, L.T., Van Dijk, A.I., 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.

Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., et al., 2015. Mapping global cropland and field size. Global Change Biol. 21 (5), 1980–1992.

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.

Gevaert, C.M., García-Haro, F.J., 2015. A comparison of STARFM and an unmixing-based algorithm for landsat and MODIS data fusion. Remote Sens. Environ. 156, 34–44.

Ghamisi, P., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., Gloaguen, R., et al., 2019. Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. IEEE Geosci. Remote Sens. Mag. 7 (1), 6–39.

Guo, D., Shi, W., Hao, M., Zhu, X., 2020. FSDAF 2.0: Improving the performance of retrieving land cover changes and preserving spatial details. Remote Sens. Environ. 248, 111973.

Huang, C., Townshend, J.R., Liang, S., Kalluri, S.N., DeFries, R.S., 2002. Impact of sensor's point spread function on land cover characterization: assessment and deconvolution. Remote Sens. Environ. 80 (2), 203–212.

Ju, J., Roy, D.P., 2008. The availability of cloud-free landsat ETM+ data over the conterminous United States and globally. Remote Sens. Environ. 112 (3), 1196–1211.

Kaiser, G., Schneider, W., 2008. Estimation of sensor point spread function by spatial subpixel analysis. Int. J. Remote Sens. 29 (7), 2137–2155.

Li, X., Foody, G.M., Boyd, D.S., Ge, Y., Zhang, Y., Du, Y., Ling, F., 2020a. 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., 2020b. Spatio-temporal fusion for remote sensing data: an overview and new benchmark. Inform. Sci. 63 (140301), 1–140301.

Liu, M., Ke, Y., Yin, Q., Chen, X., Im, J., 2019a. Comparison of five spatio-temporal satellite image fusion models over landscapes with various spatial heterogeneity and temporal variation. Remote Sens. 11 (22), 2612.

Liu, M., Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., Helmer, E.H., 2019b. An improved flexible spatiotemporal data fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. Remote Sens. Environ. 227, 74–89.

Luo, Y., Guan, K., Peng, J., 2018. STAIR: A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product. Remote Sens. Environ. 214, 87–99.

Luo, Y., Guan, K., Peng, J., Wang, S., Huang, Y., 2020. STAIR 2.0: A generic and automatic algorithm to fuse modis, landsat, and sentinel-2 to generate 10 m, daily, and cloud-/gap-free surface reflectance product. Remote Sens. 12 (19), 3209.

Meng, L., Liu, H., Zhang, X., Xu, M., Guo, D., Pan, Y., 2018. Cotton yield estimation model based on fusion image from MODIS and landsat data. In: 2018 7th International Conference on Agro-Geoinformatics (Agro-Geoinformatics). IEEE, pp. 1–5.

Moosavi, V., Talebi, A., Mokhtari, M.H., Shamsi, S.R.F., Niazi, Y., 2015. A wavelet-artificial intelligence fusion approach (WAIFA) for blending Landsat and MODIS surface temperature. Remote Sens. Environ. 169, 243–254.

ORNL DAAC, 2017. MODIS collection 6 land product subsets web service. http://dx.doi.org/10.3334/ORNLDAAC/1557, URL https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1557.

Paolini, L., Grings, F., Sobrino, J.A., Jiménez Muñoz, J.C., Karszenbaum, H., 2006. Radiometric correction effects in Landsat multi-date/multi-sensor change detection studies. Int. J. Remote Sens. 27 (4), 685–704.

Pettorelli, N., Schulte to Bühne, H., Tulloch, A., Dubois, G., Macinnis-Ng, C., Queirós, A.M., Keith, D.A., Wegmann, M., Schrodt, F., Stellmes, M., et al., 2018. Satellite remote sensing of ecosystem functions: opportunities, challenges and way forward. Remote Sens. Ecol. Conserv. 4 (2), 71–93.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., et al., 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172.

Senf, C., Leitão, P.J., Pflugmacher, D., van der Linden, S., Hostert, P., 2015. Mapping land cover in complex mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. Remote Sens. Environ. 156, 527–536.

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.

Song, Y., Zhang, H., Huang, H., Zhang, L., 2022. Remote sensing image spatiotemporal fusion via a generative adversarial network with one prior image pair. IEEE Trans. Geosci. Remote Sens. 60, 1–17.

Tan, Z., Yue, P., Di, L., Tang, J., 2018. Deriving high spatiotemporal remote sensing images using deep convolutional network. Remote Sens. 10 (7), 1066.

Tang, Y., Wang, Q., Tong, X., Atkinson, P.M., 2021. Integrating spatio-temporal-spectral information for downscaling sentinel-3 OLCI images. ISPRS J. Photogramm. Remote Sens. 180, 130–150.

Tang, Y., Wang, Q., Zhang, K., Atkinson, P.M., 2020. Quantifying the effect of registration error on spatio-temporal fusion. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 487–503.

Walker, J., De Beurs, K., Wynne, R., Gao, F., 2012. Evaluation of Landsat and MODIS data fusion products for analysis of dryland forest phenology. Remote Sens. Environ. 117, 381–393.

Wang, Q., Atkinson, P.M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. Remote Sens. Environ. 204, 31–42.

Wang, Q., Tang, Y., Tong, X., Atkinson, P.M., 2020a. Virtual image pair-based spatio-temporal fusion. Remote Sens. Environ. 249, 112009.

Wang, L., Wang, X., Wang, Q., Atkinson, P.M., 2020b. Investigating the influence of registration errors on the patch-based spatio-temporal fusion method. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 6291–6307.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sens. Environ. 236, 111402.

Wu, J., Cheng, Q., Li, H., Li, S., Guan, X., Shen, H., 2020. Spatiotemporal fusion with only two remote sensing images as input. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 6206–6219.

Wu, M., Huang, W., Niu, Z., Wang, C., Li, W., Yu, B., 2018. Validation of synthetic daily landsat NDVI time series data generated by the improved spatial and temporal data fusion approach. Inf. Fusion 40, 34–44.

Wu, M., Li, H., Huang, W., Niu, Z., Wang, C., 2015a. Generating daily high spatial land surface temperatures by combining ASTER and MODIS land surface temperature products for environmental process monitoring. Environ. Sci.: Process. Impacts 17 (8), 1396–1404.

Wu, M., Niu, Z., Wang, C., Wu, C., Wang, L., 2012. Use of MODIS and landsat time series data to generate high-resolution temporal synthetic landsat data using a spatial and temporal reflectance fusion model. J. Appl. Remote Sens. 6 (1), 063507.

Wu, M., Wu, C., Huang, W., Niu, Z., Wang, C., 2015b. High-resolution leaf area index estimation from synthetic landsat data generated by a spatial and temporal data fusion model. Comput. Electron. Agric. 115, 1–11.

Wu, M., Wu, C., Huang, W., Niu, Z., Wang, C., Li, W., Hao, P., 2016. An improved high spatial and temporal data fusion approach for combining landsat and MODIS data to generate daily synthetic landsat imagery. Inf. Fusion 31, 14–25.

Xie, D., Gao, F., Sun, L., Anderson, M., 2018. Improving spatial-temporal data fusion by choosing optimal input image pairs. Remote Sens. 10 (7), 1142.

Zhai, H., Huang, F., Qi, H., 2020. Generating high resolution LAI based on a modified FSDAF model. Remote Sens. 12 (1), 150.

Zhang, H.K., Roy, D.P., Yan, L., Li, Z., Huang, H., Vermote, E., Skakun, S., Roger, J.-C., 2018. Characterization of sentinel-2A and landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. Remote Sens. Environ. 215, 482–494.

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.

Zhao, Q., Wentz, E.A., 2020. Editorial for the special issue:“remote sensing of urban ecology and sustainability”.

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., Cai, F., Tian, J., Williams, T.K.-A., 2018. Spatiotemporal fusion of multisource remote sensing data: literature survey, taxonomy, principles, applications, and future directions. Remote Sens. 10 (4), 527.

Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J.G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. Remote Sens. Environ. 114 (11), 2610–2623.

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

Zhu, X., Zhan, W., Zhou, J., Chen, X., Liang, Z., Xu, S., Chen, J., 2022. A novel framework to assess all-round performances of spatiotemporal fusion models. Remote Sens. Environ. 274, 113002.