Integrating spatio-temporal-spectral information for downscaling Sentinel-3 OLCI images

Yijie Tang [a], Qunming Wang [a,*], Xiaohua Tong [a], Peter M. Atkinson [b, c]

[a] College of Surveying and Geo-Informatics, Tongji University, 1239 Siping Road, Shanghai 200092, China

[b] Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, UK

[c] Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK

*Corresponding author. Email: wqm11111@126.com

Abstract: Sentinel-3 is a newly launched satellite implemented by the European Space Agency (ESA) for global observation. The Ocean and Land Colour Imager (OLCI) sensor onboard Sentinel-3 provides 21 band images with a fine spectral resolution and is of great value for ocean, land and atmospheric monitoring. The two platforms (Sentinel-3A and -3B) can provide OLCI images at an almost daily temporal resolution. The coarse spatial resolution of the 21 band OLCI images (i.e., 300 m), however, limits greatly their utility for local, precise monitoring. Sentinel-2, another satellite provided by ESA, carries the Multispectral Imager (MSI) sensor which can supply much finer spatial resolution (e.g., 10 m and 20 m) images. This paper introduces a new fusion framework integrating spatio-temporal-spectral information for downscaling Sentinel-3 OLCI images, which has two parts. Based on bands with similar wavelengths (i.e., bands 2, 3, 4 and 8a for Sentinel-2 and bands Oa4, Oa6, Oa8 and Oa17 for Sentinel-3), the four Sentinel-3 bands are first downscaled to the spatial resolution of Sentinel-2 images by applying spatio-temporal fusion to Sentinel-2 MSI and Sentinel-3 OLCI images. Then, to take full advantage of all 21 available OLCI bands of the Sentinel-3 images, the extended image pair-based spatio-spectral fusion (EIPSSF) method is proposed in this paper to downscale the other 17 bands. EIPSSF is performed based on the new concept of the extended image pair (EIP) and by exploiting existing spatio-temporal fusion approaches. The framework consisting of spatio-temporal and

spatio-spectral fusion is entirely general, which provides a practical solution for comprehensive downscaling of Sentinel-3 OLCI images for fine spatial, temporal and spectral resolution monitoring.

Keywords: Sentinel-3; Sentinel-2; Downscaling; Image fusion.

1. Introduction

Sentinel-3 is a new Earth observation mission of the Global Monitoring for Environment and Security (GMES) program implemented by ESA (Berger et al., 2012; Drinkwater and Helge, 2007; Seitz et al., 2010). It is designed mainly to provide long-term monitoring of the land, ocean and atmosphere (Berger and Aschbacher, 2012; Donlon et al., 2012; Verhoef and Bach, 2012). For Sentinel-3, two constellations (Sentinel-3A and Sentinel-3B launched in February 2016 and April 2018, respectively) provide observations jointly at the global scale (Kravitz et al., 2020). For both the Sentinel-3A and Sentinel-3B satellites, a Sea and Land Surface Temperature Radiometer (SLSTR), Synthetic Aperture Radar Altimeter (SRAL) and Ocean and Land Colour Imager (OLCI) are provided onboard for maritime, land, atmospheric and climate change monitoring (Guzinski and Nieto, 2019; Zhou et al., 2020). The images produced by the OLCI sensor consist of 21 spectral channels ranging from about 400 nm to 1020 nm (i.e., from the visible to near-infrared wavelengths, see Fig. A1 in the Appendix) with a fine spectral resolution. Furthermore, given the complementarity of Sentinel-3A and Sentinel-3B, the temporal resolution of OLCI data reaches <1.4 days, which greatly facilitates frequent monitoring (Malenovsk ý et al., 2012; Nieke et al., 2012; Giannini et al., 2021). Owing to the global spatial coverage and fine spectral and temporal resolutions, the Sentinel-3 OLCI images have been employed widely for monitoring water clarity (Shen et al., 2020), retrieval of chlorophyll-a (Kravitz et al., 2020; Pahlevan et al., 2020) and inversion of inherent optical properties (Xue et al., 2019).

However, the coarse spatial resolution of 300 m limits the applications of OLCI images at the local scale, particularly for heterogeneous landscapes.

One approach to increase the spatial resolution of Sentinel-3 OLCI images is to blend them with fine spatial resolution images acquired by other satellites. Sentinel-2, another mission of ESA, can provide images with a much finer spatial resolution (i.e., ranging from 10 m to 60 m). Similarly to Sentinel-3, Sentinel-2 is also composed of two platforms, Sentinel-2A and -2B, launched in June 2015 and March 2017, respectively (Drusch et al., 2012; Du et al., 2016; Lefebvre et al., 2016; Xu and Somers, 2021). The Multispectral Imager (MSI) onboard the two platforms can provide observations with 13 spectral bands (Ansper and Alikas, 2018; Du et al., 2016; Hagolle et al., 2015; Wang et al., 2021b). The temporal resolution of the MSI, however, is up to 5 days, even though the images from both Sentinel-2A and -2B are considered. Moreover, the number of available images will in practice be much smaller due to cloud and shadow contamination. Alternatively, the daily temporal resolution of Sentinel-3 can maximize the number of effective observations across time for areas affected by cloud and shadow. Thus, it is of great interest to make full use of the fine spatial resolution of Sentinel-2 MSI images and the fine temporal resolution of Sentinel-3 OLCI images to create images with not only fine spatial but also fine temporal resolutions. It is acknowledged that spatio-temporal fusion is a technique developed for this goal (Gao et al., 2006). For spatio-temporal fusion of images from different satellite sensors, bands with similar spectral ranges are required. Among the 13 bands of Sentinel-2 images, the three 10 m MSI bands (i.e., bands 2, 3 and 4 spanning from 458-523 nm, 543-578 nm and 650-680 nm, respectively) and the 20 m MSI band 8a spanning from 855-875 nm have similar wavelengths with the four bands of Sentinel-3 images (i.e., Oa4, Oa6, Oa8 and Oa17). Therefore, spatio-temporal fusion can be applied to downscale the four bands of Sentinel-3 images to the spatial resolution of MSI images.

Over the last decade, a number of spatio-temporal fusion methods have been proposed (Belgiu and Stein, 2019; Wu et al., 2015; Zhou et al., 2021; Zhu et al., 2018). Spatio-temporal fusion methods are commonly divided into three main categories: spatial weighting-based, spatial unmixing-based and hybrid methods (Zhu et al., 2018). Typical spatial weighting-based methods include the spatial and temporal adaptive reflectance

fusion model (STARFM) (Gao et al., 2006), spatial temporal adaptive algorithm for mapping reflectance change (Hilker et al., 2009), the enhanced spatial and temporal adaptive reflectance fusion model (Zhu et al., 2010), the Fit-FC method (Wang and Atkinson, 2018) and the virtual image pair-based spatio-temporal fusion (VIPSTF) method. Based on the multisensor multiresolution technique, several spatial unmixing-based methods were developed by adding different constraints to the unmixing model (Amorós-López et al., 2013; Busetto et al., 2008; Gevaert and García-Haro, 2015; Mustafa et al., 2014; Wang et al., 2021a; Wu et al., 2012; Xu et al., 2015; Zurita-Milla et al., 2008; Zurita-Milla et al., 2011). The hybrid methods integrated the mechanisms of spatial weighting and spatial unmixing, including the Flexible Spatiotemporal DAta Fusion (FSDAF) method (Zhu et al., 2016), the improved FSDAF (Liu et al., 2019) and the enhanced FSDAF that incorporates sub-pixel class fraction (Li et al., 2020).

Up to now, spatio-temporal fusion has generally been performed for downscaling coarse spatial resolution Moderate Resolution Imaging Spectroradiometer (MODIS) or MEdium Resolution Imaging Spectrometer (MERIS) images, by fusing with fine spatial resolution images from the Landsat sensors (e.g., Thematic Mapper (TM), Enhanced Thematic Mapper (ETM+) or Operational Land Imager (OLI)) (Chen and Huang, 2015; Gao et al., 2015; Zhang et al., 2015). Fit-FC is one of the very few methods proposed originally for fusing Sentinel-2 MSI with Sentinel-3 OLCI images (Wang and Atkinson, 2018), where the four bands of Sentinel-3 images (i.e., Oa4, Oa6, Oa8 and Oa17) were downscaled to the spatial resolution of Sentinel-2 images. However, when downscaling Sentinel-3 images in practical cases, two issues remain open. First, from the perspective of data, almost no study has been conducted for downscaling real Sentinel-3 images based on spatio-temporal fusion. Wang and Atkinson (2018) performed spatio-temporal fusion using simulated Sentinel-3 images, which were created by degrading bands 2, 3, 4 and 8a of Sentinel-2 images. In practice, large differences may exist between Sentinel-2 and -3 images, resulting in a non-negligible difference between the real and simulated Sentinel-3 images. The performances of existing spatio-temporal fusion methods remain to be validated for real Sentinel-3 images. Second, based on spatio-temporal fusion, only four bands of Sentinel-3 images can be downscaled since they have similar wavelengths with the four bands of Sentinel-2

images. As acknowledged widely, however, all 21 bands of Sentinel-3 OLCI images convey specific semantic information, as presented in Fig. A1 (Donlon et al., 2012). Therefore, there exists a great need for approaches for downscaling the other 17 OLCI bands.

Spatio-spectral fusion can be a solution for downscaling the other 17 OLCI bands, using four fine spatial resolution OLCI bands predicted by spatio-temporal fusion. Various spatio-spectral fusion methods have been exploited over the past decades, including the component substitution-based and multiresolution analysis-based methods. Gram-Schmidt transformation (Laben and Brower, 2000), Intensity-hue-saturation (Tu et al., 2001), principal component analysis (Shettigara, 1992) and Hyperspherical Color Space (Padwick et al., 2010) are typical component substitution approaches. The multiresolution analysis approaches include the high-pass filtering (Chavez et al., 1991), decimated wavelet transform using an additive injection model (Khan et al., 2008), Morphological Half Gradient (Restaino et al., 2016) and smoothing filter-based intensity modulation (Liu, 2000). There are also several deep learning-based methods developed in recent years (Xie et al., 2019; Xiong et al., 2021; Zhang et al., 2019). Several reviews on the methods are available (Amolins et al., 2007; Garzelli, 2016; Javan et al., 2021). The majority of the existing spatio-spectral fusion methods, however, have some shortcomings when introduced to the downscaling of the other 17 OLCI bands. First, most of the spatio-spectral fusion approaches are mainly designed for pan-sharpening, that is, the case with only one fine spatial resolution band. Thus, the application of these methods will fail to fully utilize all four fine spatial resolution bands in downscaling the 17 OLCI bands. Second, these methods are generally suitable for a small zoom factor (e.g., 2 to 4) between the fine and coarse spatial resolution images. The fusion process becomes more challenging when the zoom factor is large (e.g., the zoom factor for downscaling the 300 OLCI bands is more than 10). Third, in traditional spatio-spectral fusion, the spectral ranges of the coarse bands are required to have a great degree of overlap with those of fine spatial resolution bands. Thus, considering the small overlap between the spectral ranges of the OLCI bands, the application of traditional spatio-spectral fusion approaches will be of great challenge. Therefore, it is necessary to develop specific spatio-spectral fusion methods for the task of downscaling the remaining 17 OLCI bands.

In this paper, a fusion framework integrating spatio-temporal-spectral information is proposed to tackle the above issues. The framework aims at downscaling all 21 OLCI bands of real Sentinel-3 images to the spatial resolution of Sentinel-2 images by fusing with Sentinel-2 images *through two steps*. First, the Oa4, Oa6, Oa8 and Oa17 bands of Sentinel-3 images are downscaled by spatio-temporal fusion. Second, the other 17 OLCI bands are downscaled through an extended image pair-based spatio-spectral fusion (EIPSSF) method. In spatio-temporal fusion, the image pair is acquired at a different time from the coarse band image at the prediction time, but they fall in the same wavelength. Inspired by the traditional image pair, the new concept of the extended image pair (EIP), is proposed in this paper for spatio-spectral fusion. The definition of EIP also follows the same basic assumption of consistent spatial extent, acquisition time and spectral range for the fine and coarse images. However, EIP is acquired at the prediction time, and its wavelength is different from the target coarse bands to be downscaled. For downscaling the remaining 17 OLCI bands, four EIPs are required, composed of the coarse-fine images for the Oa4, Oa6, Oa8 and Oa17 bands, where the fine spatial resolution images can be obtained by the pre-spatio-temporal fusion step. Using the EIPSSF-based spatio-spectral fusion method, the other 17 OLCI bands are downscaled by fusing with the four EIPs.

Three main contributions can be summarized for this paper.

1) Real Sentinel-3 data are considered for downscaling. Different from Wang and Atkinson (2018), which used simulated Sentinel-3 images, the employment of real Sentinel-3 images provides an objective and authentic assessment of current spatio-temporal fusion methods and, more generally, downscaling methods.
2) The typical spatio-temporal fusion methods are compared systematically to identify the most accurate method for fusion of Sentinel-2 and -3 images (i.e., downscaling 300 m Oa4, Oa6, Oa8 and Oa17 bands to 10 m or 20 m).
3) EIPSSF is developed to downscale the other 17 OLCI bands of Sentinel-3 images. EIPSSF inherits the core idea of an image pair in spatio-temporal fusion and makes full use of the Sentinel-3 Oa4, Oa6, Oa8, Oa17 bands based on the proposed concept of EIP. The method is developed to deal with the cases

involving large zoom factor and small overlap between the spectral ranges of the fine and coarse spatial resolution bands.

2. Methods

2.1. Processing Sentinel-2 and -3 data

ESA provides open access to Sentinel-2 MSI and Sentinel-3 OLCI images. For the Sentinel-2 MSI images, Level-2A products were used since they provide directly the bottom of atmosphere (BOA) reflectance. For the Sentinel-3 OLCI images, Level-1 products representing top of atmosphere (TOA) reflectance were chosen since the BOA reflectance products were not available. As the fusion of images requires strict consistency in reflectance and spatial extent between the images from two sensors, pre-processing of the Sentinel-2 and -3 data was required before applying any algorithms (Cazzaniga et al., 2019). To our knowledge, however, very few studies have presented the specific steps for processing Sentinel-2 and -3 data prior to fusion. In this paper, we display the detailed data processing in Fig. 1, which includes two main parts, that is, radiometric correction and geometric correction. All steps were performed using ENVI 5.5.3.

For the input Sentinel-3 OLCI Level-1 data products, the pre-processing step involved geometric positioning, to relate the image to geographic coordinates. Then, radiometric correction and geometric correction were performed separately. In the radiometric correction part, the digital number (DN) value was first transformed to TOA reflectance by radiometric calibration based on the Sentinel-3 data with geographic coordinates. To make the Sentinel-3 data comparable with the Sentinel-2 BOA data, the TOA reflectance of the Sentinel-3 data needs to be converted to BOA reflectance by atmosphere correction. Specifically, the MODTRAN-based Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) atmospheric model was applied (Anderson et al., 2002), which is available for Sentinel-3 in ENVI 5.5.3. For geometric

correction, the Sentinel-3 data were first reprojected to the same projection coordinates as the Sentinel-2 data (i.e., the Universal Transverse Mercator (UTM) zone), thus, achieving a consistent coordinate system. Due to possible variation in the spatial resolution caused by the reprojection process, the Sentinel-3 data were then resized to the original spatial resolution of 300 m by adapting the nearest neighbor interpolation. After the above procedures, the Sentinel-2 and -3 data were comparable in terms of reflectance and their coordinates. In image fusion, however, images with the same spatial extent are required. Since the Sentinel-3 data have a larger swath compared to Sentinel-2 data, the Sentinel-3 data were clipped based on the spatial extent of the Sentinel-2 data. Finally, to minimize the impact of registration error on image fusion, manual geometric rectification was applied between the Sentinel-2 and -3 data, producing the final Sentinel-2 MSI and Sentinel-3 OLCI data for the fusion framework integrating spatio-temporal-spectral information.

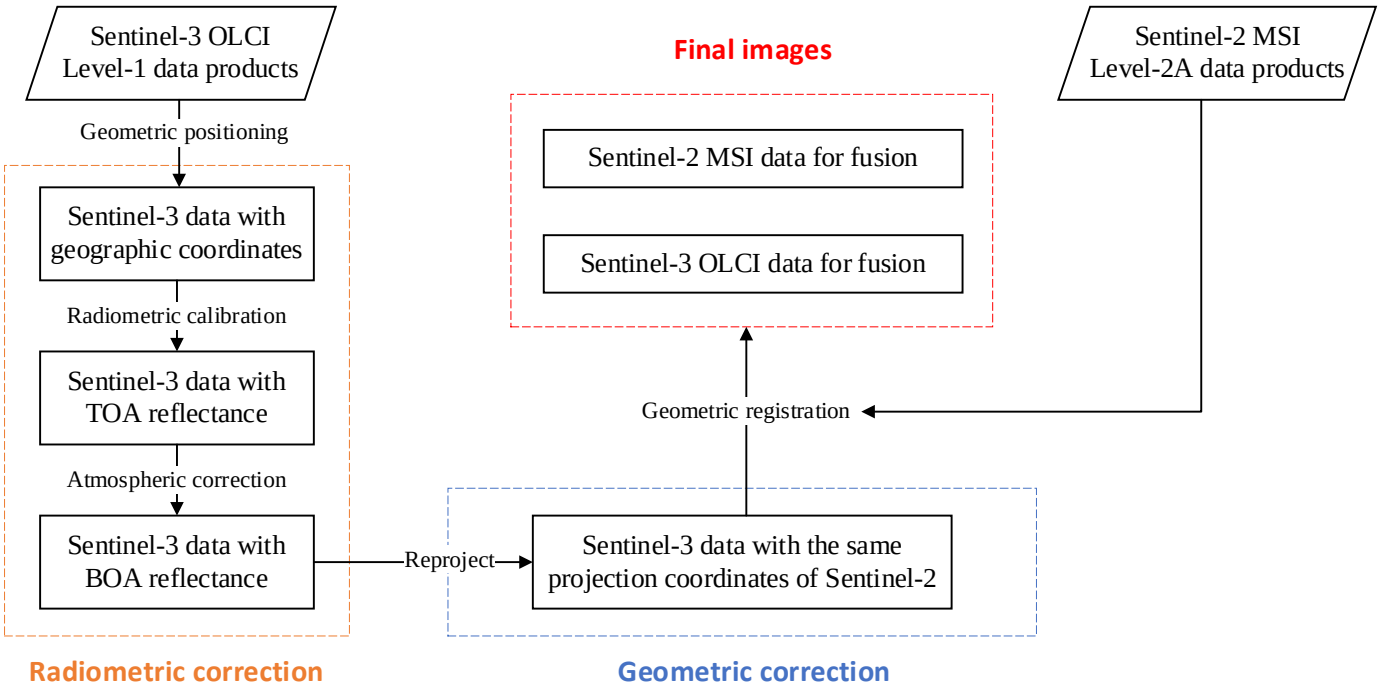


Fig. 1. The processing of Sentinel-2 and -3 data.

2.2. The fusion framework integrating spatio-temporal-spectral information

Based on the acquired Sentinel-2 MSI and Sentinel-3 OLCI images, a fusion framework integrating spatio-temporal-spectral information is proposed to downscale all 21 bands of Sentinel-3 images to the spatial resolution of Sentinel-2 images. More precisely, the target fine spatial resolution was defined as 20 m in this paper. The finer spatial resolution of 10 m was not considered as the zoom factor of 30 in this case necessarily involves greater uncertainty and may be too large for meaningful downscaling. The proposed fusion framework integrating spatio-temporal-spectral information can be divided into two separate steps, spatio-temporal fusion and spatio-spectral fusion, as shown in Fig. 2. In the first step, spatio-temporal fusion methods are applied to fuse the four corresponding bands of Sentinel-2 (i.e., MSI bands 2, 3, 4 and 8a) and Sentinel-3 images (OLCI bands Oa4, Oa6, Oa8 and Oa17). For the Sentinel-2 images, the 10 m bands 2, 3 and 4 are upscaled to 20 m in advance to match the spatial resolution of band 8a. For the Sentinel-3 images, only bands Oa4, Oa6, Oa8 and Oa17 are involved in spatio-temporal fusion since they have similar spectral ranges with the four bands of the Sentinel-2 images. In the second step, the novel EIPSSF method is developed to downscale the other 17 bands (i.e., bands Oa1 to Oa21, excluding bands Oa4, Oa6, Oa8 and Oa17) of the Sentinel-3 images. The details of the two parts are further illustrated in Sections 2.2.1 and 2.2.2.

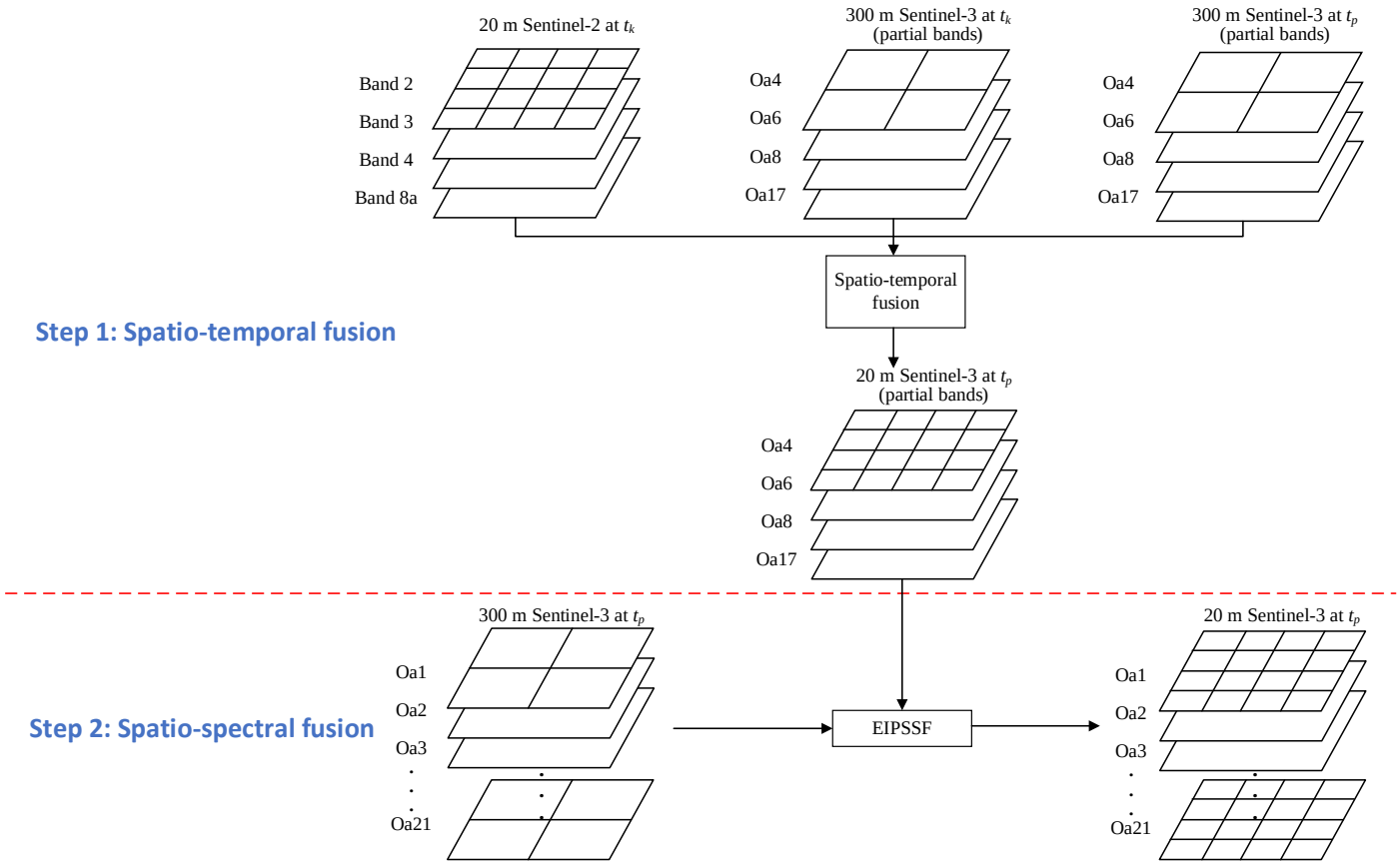


Fig. 2. Flowchart illustrating the proposed fusion framework integrating spatio-temporal-spectral information.

2.2.1. Spatio-temporal fusion

Suppose the acquisition times of the known Sentinel-2 image and coarse Sentinel-3 image to be downscaled (i.e., the prediction time) are t_k and t_p, respectively. In spatio-temporal fusion, the 300 m Oa4, Oa6, Oa8 and Oa17 bands of Sentinel-3 images at t_p are downscaled to 20 m by fusing with 20 m Sentinel-2 images acquired at t_k and 300 m Sentinel-3 images acquired at t_k and t_p. In this paper, seven typical spatio-temporal fusion methods, the unmixing-based data fusion (UBDF) algorithm (Zurita-Milla et al., 2008), the spatial-temporal data fusion approach (STDFA) (Wu et al., 2012), spatial unmixing-based VIPSTF (VIPSTF-SU), FSDAF, STARFM, Fit-FC and spatial weighting-based VIPSTF (VIPSTF-SW), are

considered. This section introduces the mechanisms of the seven methods, categorized by the type of the spatio-temporal fusion methods. Generally, the framework of spatio-temporal fusion can be represented as

$$\begin{aligned}\mathbf{F}_p &= \mathbf{F}_k + \Delta\mathbf{F} \\ &= \mathbf{F}_k + f(\Delta\mathbf{C}) \\ &= \mathbf{F}_k + f(\mathbf{C}_p - \mathbf{C}_k)\end{aligned} \tag{1}$$

where $\mathbf{F}_k$ and $\mathbf{F}_p$ are the fine spatial resolution images at the known and prediction times, respectively. $\mathbf{C}_k$ and $\mathbf{C}_p$ are the coarse spatial resolution images at the known and prediction times, respectively. $\Delta\mathbf{F}$ represents a fine spatial resolution level increment estimated by applying different downscaling operations f to $\Delta\mathbf{C}$, which refers to the increment from $\mathbf{C}_k$ to $\mathbf{C}_p$. The difference between various spatio-temporal fusion methods mainly lies in the estimation of $\Delta\mathbf{F}$.

1) Spatial weighting-based methods (STARFM, Fit-FC and VIPSTF-SW)

The basic principle of spatial weighting-based methods is to estimate the reflectance of each fine spatial resolution pixel by applying a weighting function to spatially neighboring pixels. In this category of method, for clarity, $\Delta\mathbf{F}$ in Eq. (1) is represented as $\Delta\mathbf{F}_{\mathrm{SW}}$, which is estimated by applying different spatial weighting-based operations f to $\Delta\mathbf{C}$. For STARFM, $\Delta\mathbf{F}_{\mathrm{SW}}$ is calculated by applying a weighting function to spatially neighboring, spectrally similar pixels, which takes consideration of the temporal difference between the images at the known and prediction times. In Fit-FC, a local fitting model is first applied to enhance the correlation between the coarse images at the known and prediction times, and $\Delta\mathbf{F}_{\mathrm{SW}}$ is further estimated through a spatial filtering and a residual compensation process. VIPSTF-SW is implemented by employing the spatial weighting strategy of STARFM to the VIPSTF framework. For VIPSTF-SW, a linear transformation is first applied to the fine and coarse images at the known time to produce a virtual image pair, as expressed in Eqs. (2) and (3)

$$\mathbf{F}_{\mathrm{VIP}} = \sum_{i=1}^{n} a_i \mathbf{F}_i + b \tag{2}$$

$$\mathbf{C}_{\mathrm{VIP}} = \sum_{i=1}^{n} a_i \mathbf{C}_i + b \tag{3}$$

where a_i is the transformation coefficient for the *i*th fine spatial resolution image, *n* is the number of known image pairs and b is a constant. $\mathbf{F}_i$ and $\mathbf{C}_i$ are the *i*th fine and coarse spatial resolution images, respectively. $\mathbf{F}_{\mathrm{VIP}}$ and $\mathbf{C}_{\mathrm{VIP}}$ are the virtual fine and coarse images, respectively. Based on the assumption and derivation in Wang et al. (2020), the transformation coefficients (i.e., a_i and *b*) can be obtained through a linear regression model constructed between the coarse images at the known and prediction times.

Application of the virtual image pair reduces the difference (in feature space) between the images at the known and prediction times, and the difference between the coarse images can be updated to

$$\Delta\mathbf{C}' = \mathbf{C}_p - \mathbf{C}_{\mathrm{VIP}}. \tag{4}$$

Then, the spatial weighting strategy is adopted to predict the fine spatial resolution level increment $\Delta\mathbf{F}_{\mathrm{SW}}$ based on the coarse spatial resolution level increment $\Delta\mathbf{C}'$

$$\Delta F_{\mathrm{sw}}(x_0, y_0) = \sum_{i=1}^{n_s} \lambda_i \Delta C'(x_i, y_i) \tag{5}$$

where $\Delta C'(x_i, y_i)$ is the coarse increment for the *i*th similar pixel located at (x_i, y_i) surrounding the pixel located at (x_0, y_0), n_s is the number of similar pixels. Moreover, λ_i is the weight for the *i*th similar pixel, which is characterized by the inverse of its spatial distance (i.e., Euclidean distance) to the center pixel.

2) Spatial unmixing-based methods (UBDF, STDFA and VIPSTF-SU)

Spatial unmixing-based methods estimate the reflectance for each class of the fine spatial resolution image by applying an unmixing model to the coarse spatial resolution image. Alternatively, $\Delta\mathbf{F}$ in Eq. (1) is represented as $\Delta\mathbf{F}_{\mathrm{SU}}$, which is estimated by different spatial unmixing-based methods. For UBDF, $\mathbf{F}_k$ and $\Delta\mathbf{F}_{\mathrm{SU}}$ are considered as a whole to provide the prediction. Specifically, the fine spatial resolution image at the known time is used to acquire a land cover map. $\mathbf{F}_p$ is predicted by an unmixing model, which decomposes

the target coarse image directly to predict the class reflectance, with the coarse proportion image produced by upscaling the known fine spatial resolution land cover map. For STDFA, $\Delta\mathbf{F}_{SU}$ is calculated alternatively by decomposing the coarse increment $\Delta\mathbf{C}$. VIPSTF-SU is derived from the VIPSTF framework by employing the spatial unmixing model in STDFA.

3) Hybrid methods (FSDAF)

FSDAF is a hybrid method combining the strengths of both the spatial unmixing- and the spatial weighting-based methods. The fine spatial resolution level temporal change is first estimated using the spatial unmixing model in STDFA. Then, residuals calculated by thin plate spline interpolation are distributed to the fine spatial resolution pixels based on the spatial weighting scheme in STARFM.

2.2.2. Spatio-spectral fusion

In the second step of the fusion framework integrating spatio-temporal-spectral information, the EIPSSF method is proposed to downscale the other 17 OLCI bands of Sentinel-3 images. Different from spatio-temporal fusion, EIPSSF is performed based on the new concept of the EIP, which differs from the image pair in spatio-temporal fusion. To present clearly the definition of EIP, the general concepts of spatio-temporal fusion and spatio-spectral fusion are shown in Fig. 3. In spatio-temporal fusion, suppose that we have coarse spatial resolution images $C(t_1, B_i)$ to $C(t_n, B_i)$ with the spectrum B_i acquired from t_1 to t_n, and fine spatial resolution images $F(t_1, B_i)$ and $F(t_n, B_i)$ with the spectrum B_i acquired at t_1 and t_n. The absent fine spatial resolution images between $F(t_1, B_i)$ and $F(t_n, B_i)$ need to be predicted. Here, the fine and coarse images acquired on the same date (e.g., $F(t_1, B_i)$ and $C(t_1, B_i)$) can be regarded as an image pair, which are consistent in their spatial extent, acquisition time and spectral range. Moreover, the image pairs in spatio-temporal fusion are acquired at different times from the coarse resolution image to be downscaled.

In spatio-spectral fusion, however, all the images are acquired at the same prediction time t_p. Suppose that we have coarse spatial resolution images $C(t_p, B_1)$ to $C(t_p, B_s)$ acquired on the same date t_p, but with different spectra ranging from B_1 to B_s, and fine spatial resolution images $F(t_p, B_1)$ and $F(t_p, B_s)$ acquired on the same date t_p but with different spectra B_1 and B_s. In this case, $C(t_p, B_1)$ and $F(t_p, B_1)$ is defined as an EIP. The missing fine spatial resolution images between $F(t_p, B_1)$ and $F(t_p, B_s)$ need to be predicted based on the EIPs. In spatio-spectral fusion, it is acknowledged that the corresponding coarse and fine band images (e.g., $C(t_1, B_i)$ and $F(t_1, B_i)$) are also in accordance with the definition of an image pair in terms of consistency in space, time and spectrum. Importantly, the EIP in spatio-spectral fusion differs from the image pair in spatio-temporal fusion in two aspects. First, the EIP has a different spectrum from the coarse image to be downscaled (i.e., $C(t_p, B_i)$), while the spectra of both types of images are consistent in spatio-temporal fusion. Second, the EIP is acquired at the prediction time in spatio-temporal fusion.

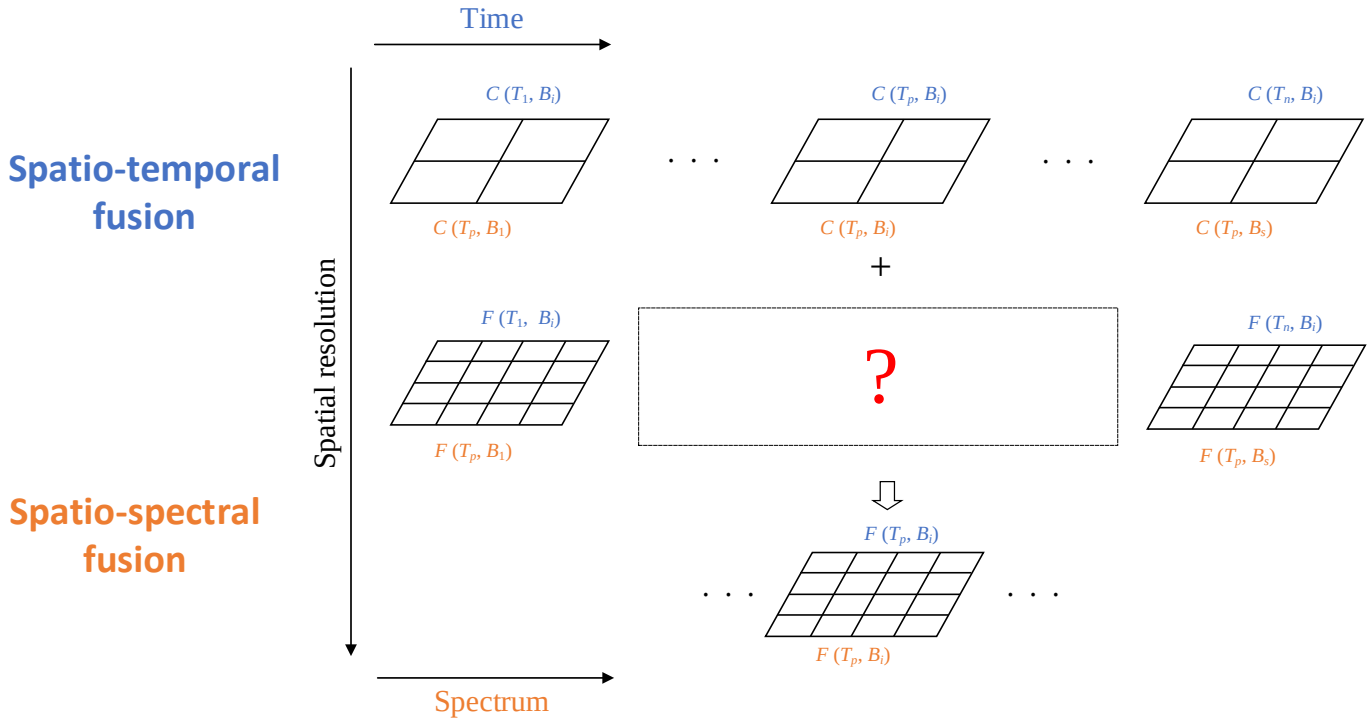


Fig. 3. Difference between image pairs in spatio-temporal fusion and spatio-spectral fusion.

With the use of the EIP, existing spatio-temporal fusion methods can be transferred to spatio-spectral fusion. Specifically, by applying different spatio-temporal methods to spatio-spectral fusion, different EIPSSF methods (e.g., UBDF-based EIPSSF or FSDAF-based EIPSSF) can be produced. Thus, the proposed EIPSSF provides a general, modular framework for spatio-spectral fusion, which is theoretically applicable to almost all spatio-temporal fusion methods. It is noted that there is a prerequisite for EIPSSF, that is, the spectral distance between EIPs and the coarse bands should be limited to a certain range to ensure the accuracy of fusion. To further explain the principle of EIPSSF, we specify the mechanism using the VIPSTF-SW method in spatio-temporal fusion as an example, producing the VIPSTF-SW-based EIPSSF method. After the 20 m Oa4, Oa6, Oa8, Oa17 bands of Sentinel-3 images have been predicted by spatio-temporal fusion in Section 2.2.1, the other 17 OLCI bands are downscaled based on EIPSSF separately. When downscaling a specific band image within the 17 OLCI bands, the virtual 20 m band image $\mathbf{B}_{\mathrm{VIP}}^{20\mathrm{m}}$ and the virtual 300 m band image $\mathbf{B}_{\mathrm{VIP}}^{300\mathrm{m}}$ are created by applying a linear transformation to the four known 20 m and 300 m band images (i.e., four EIPs), respectively

$$\mathbf{B}_{\mathrm{VIP}}^{20\mathrm{m}} = \sum_{i=1}^{s} a_i \mathbf{B}_i^{20\mathrm{m}} + b \tag{6}$$

$$\mathbf{B}_{\mathrm{VIP}}^{300\mathrm{m}} = \sum_{i=1}^{s} a_i \mathbf{B}_i^{300\mathrm{m}} + b \tag{7}$$

where $\mathbf{B}_i^{20\mathrm{m}}$ and $\mathbf{B}_i^{300\mathrm{m}}$ are the *i*th 20 m and 300 m images, respectively, among the Oa4, Oa6, Oa8 and Oa17 bands, and *s* is the number of EIPs (e.g., $s = 4$ in this case). a_i and b are the transformation coefficients estimated based on the regression model constructed between the 300 m band to be predicted and the four 300 m images. Then, the 20 m downscaling result $\mathbf{B}_p^{20\mathrm{m}}$ of band *p* can be predicted by

$$\begin{aligned} \mathbf{B}_p^{20\mathrm{m}} &= \mathbf{B}_{\mathrm{VIP}}^{20\mathrm{m}} + \Delta\mathbf{B}^{20\mathrm{m}} \\ &= \mathbf{B}_{\mathrm{VIP}}^{20\mathrm{m}} + f(\Delta\mathbf{B}'^{300\mathrm{m}}) \\ &= \mathbf{B}_{\mathrm{VIP}}^{20\mathrm{m}} + f(\mathbf{B}_p^{300\mathrm{m}} - \mathbf{B}_{\mathrm{VIP}}^{300\mathrm{m}}) \end{aligned} \tag{8}$$

where $\mathbf{B}_p^{300\mathrm{m}}$ is the 300 m image for band *p* and $\Delta\mathbf{B}'^{300\mathrm{m}}$ is the difference between $\mathbf{B}_p^{300\mathrm{m}}$ and $\mathbf{B}_{\mathrm{VIP}}^{300\mathrm{m}}$. $\Delta\mathbf{B}^{20\mathrm{m}}$ is estimated by applying the algorithm f (i.e., the same spatial weighting strategy applied in VIPSTF-SW) to $\Delta\mathbf{B}'^{300\mathrm{m}}$.

3. Experiments

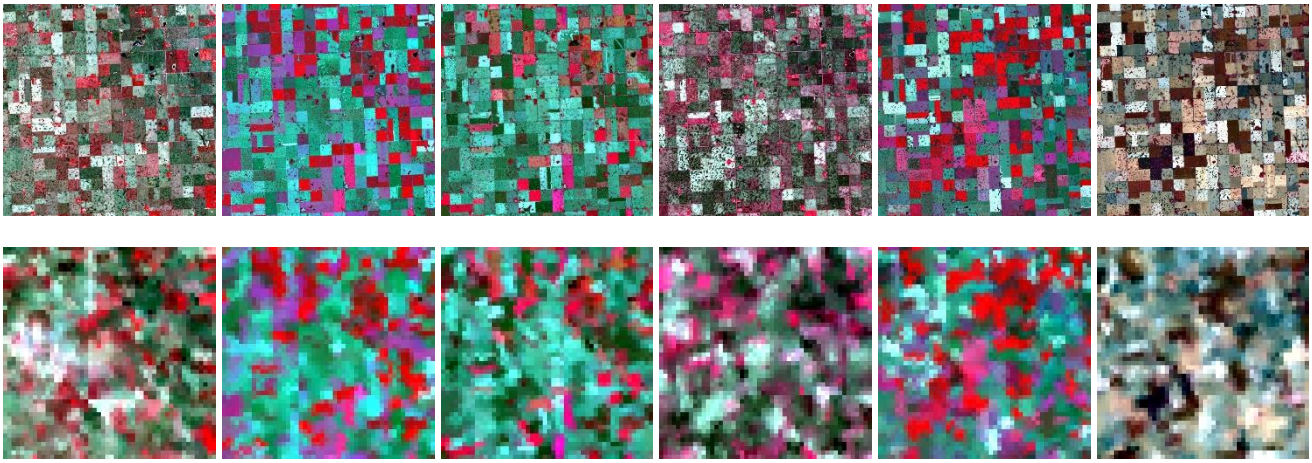
3.1. Data and experimental setup

Three datasets were utilized to examine the performance of the proposed fusion framework integrating spatio-temporal-spectral information for downscaling Sentinel-3 images. For Sites 1 and 2, Sentinel-2 and -3 time-series images covering two 15 km by 15 km sites in the State of North Dakota, America were utilized. They were clipped from the Sentinel-2 and -3 time-series images covering the same 109.5 km by 109.5 km area, which were processed according to the steps in Section 2.1. Site 3 covers an area of 15 km by 15 km in Angers, France. The acquisition times of Sentinel-2 and -3 images for three sites were listed in Table 1. For Sites 1 and 2, the acquisition dates of the Sentinel-2 and -3 time-series images range from 6 June 2019 to 10 October 2020, presenting a relative uniform distribution. Note that the images from October 2019 to February 2020 are absent owing to the influence of the snow cover. For both sites, the Sentinel-3 image acquired on a certain date was chosen for downscaling, with the images acquired on other dates known. In the experiments, the Sentinel-3 images acquired on 16 September 2019 and 20 August 2019 were selected for downscaling for Sites 1 and 2, respectively. For Site 3, the known and prediction dates are 2 June 2020 and 6 August 2020, respectively. Ultimately, all 21 bands of Sentinel-3 images were downscaled to the spatial resolution of Sentinel-2 images by applying the proposed fusion framework integrating spatio-temporal-spectral information. The partial Sentinel-2 and -3 time-series images for Sites 1, 2 and 3 are presented in Figs. 4, 5 and

6, respectively. Both Sites 1 and 2 are mainly covered by vegetation (e.g., crops), while Site 3 presents a more complex landscape, consisting of vegetation, urban and river.

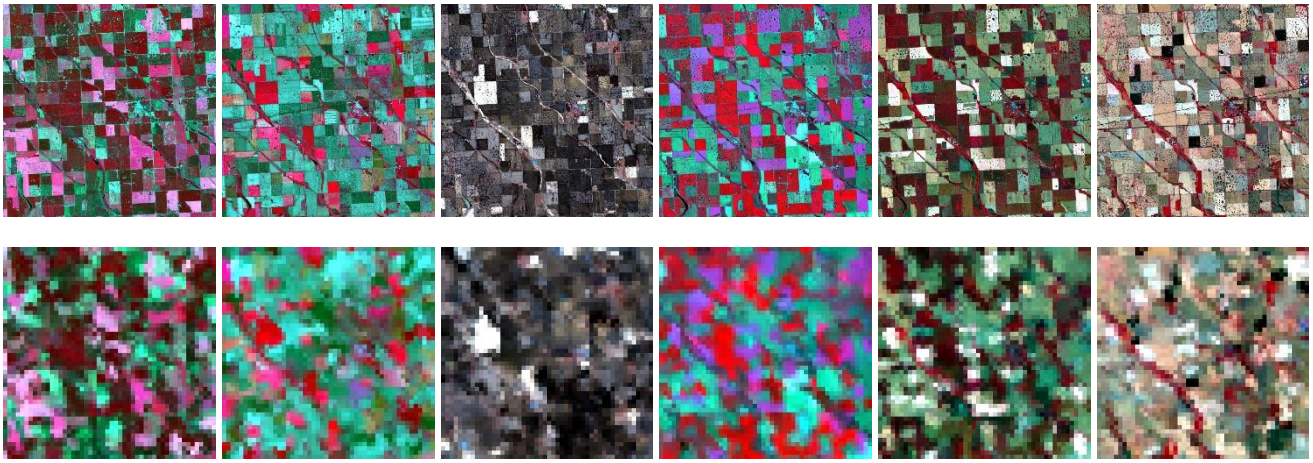
Table 1 Acquisition dates of the Sentinel-2 and -3 images

	Site 1	Site 2	Site 3
Known dates	2019.6.6	2019.6.6	2020.6.2
	2019.7.18	2019.7.18	
	2019.9.16	2019.8.20	
	2020.3.27	2020.3.27	
	2020.6.7	2020.6.7	
	2020.8.11	2020.8.11	
	2020.8.24	2020.8.24	
	2020.9.10	2020.9.10	
	2020.9.25	2020.9.25	
	2020.10.10	2020.10.10	
Prediction date	2019.8.20	2019.9.16	2020.8.6



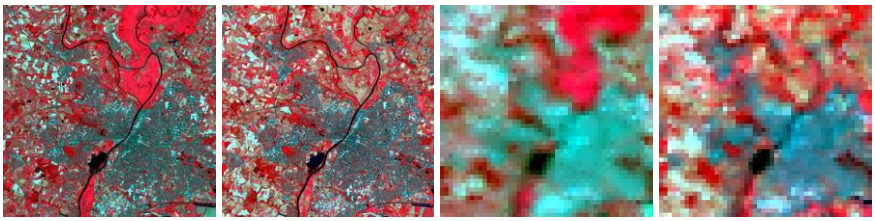
(a) (b) (c) (d) (e) (f)

Fig. 4. Partial Sentinel-2 (first line) and Sentinel-3 (second line) BOA reflectance images for Site 1 (8a, 4, 3 bands for Sentinel-2 and Oa17, Oa8, Oa6 for Sentinel-3 as RGB, respectively). (a) 6 June 2019. (b) 20 August 2019. (c) 16 September 2019. (d) 7 June 2020. (e) 24 August 2020. (f) 25 September 2020.



(a) (b) (c) (d) (e) (f)

Fig. 5. Partial Sentinel-2 (first line) and Sentinel-3 (second line) BOA reflectance images for Site 2 (8a, 4, 3 bands for Sentinel-2 and Oa17, Oa8, Oa6 for Sentinel-3 as RGB, respectively). (a) 18 July 2019. (b) 16 September 2019. (c) 27 March 2020. (d) 11 August 2020. (e) 10 September 2020. (f) 10 October 2020.



(a) (b) (c) (d)

Fig. 6. Sentinel-2 and Sentinel-3 BOA reflectance images for Site 3 (8a, 4, 3 bands for Sentinel-2 and Oa17, Oa8, Oa6 for Sentinel-3 as RGB, respectively). (a) Sentinel-2 on 2 June 2020. (b) Sentinel-2 on 6 August 2020. (c) Sentinel-3 on 2 June 2020. (d) Sentinel-3 on 6 August 2020.

Generally, the experiments are divided into two parts, spatio-temporal fusion and spatio-spectral fusion, as illustrated in Sections 3.2 and 3.3, respectively. Section 3.2 provides downscaling results for bands Oa4, Oa6, Oa8 and Oa17 of Sentinel-3 for Sites 1 and 2 based on different spatio-temporal fusion methods (i.e., UBDF, STDFA, VIPSTF-SU, FSDAF, STARFM, Fit-FC and VIPSTF-SW), and also the applicability of the results for land cover mapping. Section 3.3 provides downscaling results for the other 17 bands of Sentinel-3 by applying the EIPSSF-based spatio-spectral fusion method. Section 3.4 presents the application of the entire fusion framework to more complex landscapes in Site 3.

3.2. Spatio-temporal fusion for downscaling OLCI bands Oa4, Oa6, Oa8 and Oa17

3.2.1. Experiment on Site 1

Amongst the 11 Sentinel-2 and -3 image pairs acquired for Site 1, the Sentinel-3 image acquired on 16 September 2019 was selected to be downscaled using the different spatio-temporal fusion methods. The Sentinel-2 and -3 image pairs acquired on other dates were chosen as inputs, in turn, together with the Sentinel-3 image acquired on 16 September 2019. Consequently, 10 predictions for downscaling bands Oa4, Oa6, Oa8 and Oa17 were obtained with 10 image pairs acquired on different dates as input. The predictions of the seven methods with the use of the image pair acquired on 20 August 2019 are displayed in Fig. 7 for visual observation. It is obvious that the predictions for UBDF, VIPSTF-SU, Fit-FC and VIPSTF-SW are closer to the reference image visually. For exmaple, the green block in the middle-upper part of the subarea is wrongly predicted as blue in the predictions of the other methods. Moreover, in the predictions of UBDF and VIPSTF-SU, the block effect emerges to some extent. Fit-FC produces a smoothing effect. With respect to VIPSTF-SW, its prediction is visually more accurate as it is closest to the reference in terms of the recovery of both spatial detail and spectral information. To present the difference between the seven methods more clearly, the errors between the downscaling result and the reference images for bands Oa4, Oa6, Oa8 and Oa17 are shown in Fig. 8. The white represents the prediction with no error, while the blue and red represent the largest negative and positive errors, respectively. It is noted that the error images of UBDF, STDFA and VIPSTF-SU are mainly covered by red and blue for all four bands. The difference between the results of seven methods tend to be more obvious in bands Oa8 and Oa17. Specifically, the error images of UBDF, STDFA, VIPSTF-SU, FSDAF and STARFM are generally dominated by blue and red pixels, while those of Fit-FC and VIPSTF-SW are mainly covered by white, light red and light blue pixels. Therefore, the predictions of Fit-FC and VIPSTF-SW are considered to be closer to the reference.

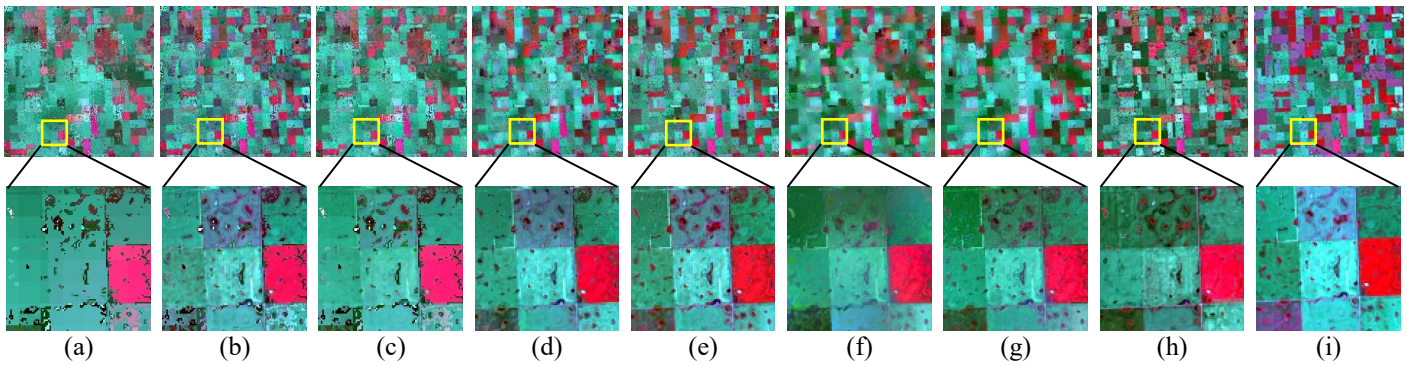


Fig. 7. Results of different spatio-temporal fusion methods for downcaling 300 m bands Oa4, Oa6, Oa8 and Oa17 to 20 m for Site 1 (prediction time on 16 September 2019; image pair on 20 August 2019 as input) (bands Oa17, Oa8 and Oa6 as RGB). (a) UBDF. (b) STDFA. (c) VIPSTF-SU. (d) FSDAF. (e) STARFM. (f) Fit-FC. (g) VIPSTF-SW. (h) Reference. (i) Input Sentinel-2 image.

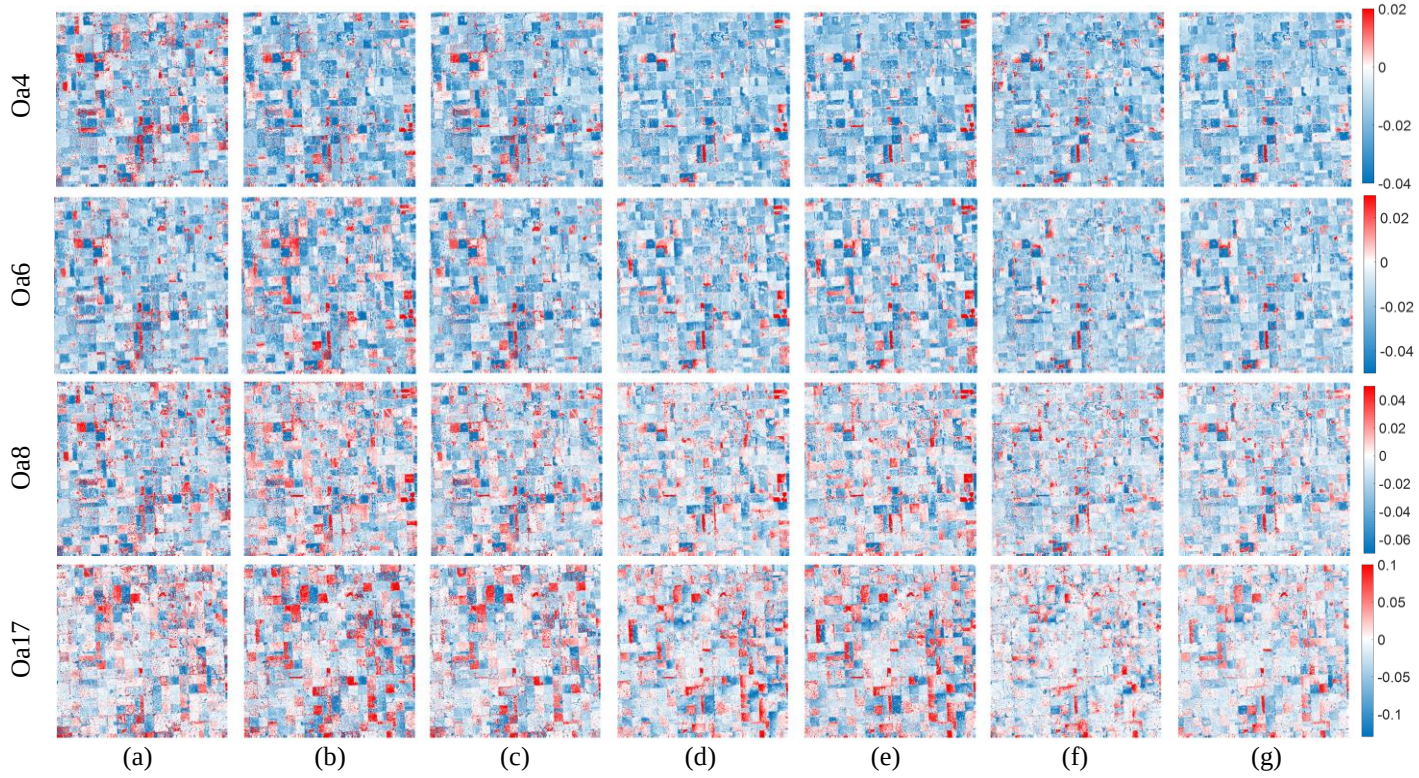


Fig. 8. Error images of different methods for Site 1. (a) UBDF. (b) STDFA. (c) VIPSTF-SU. (d) FSDAF. (e) STARFM. (f) Fit-FC. (g) VIPSTF-SW.

Quantative evaluation was conducted based on the indices of the root mean square error (RMSE) and the corelation coefficient (CC), as displayed in Table 2. It is noted that for the predictions of each band, VIPSTF-SW produces consistently the smallest RMSE and the largest CC, indicating that the VIPSTF-SW

has the greatest accuracy, which is in accordance with the conclusion based on visual assessment. More precisely, the mean CCs for UBDF, STDFA and VIPSTF-SU are 0.5713, 0.6388 and 0.6417, respectively, which are obviously smaller than for Fit-FC, FSDAF and STARFM. Fit-FC, FSDAF and STARFM produce mean CCs of 0.7741, 0.7606 and 0.7458, which are 0.0380, 0.0515 and 0.0663 smaller than for VIPSTF-SW, respectively. Overall, VIPSTF-SW is found to be more accurate than the other six methods according to both visual and quantitative assessment.

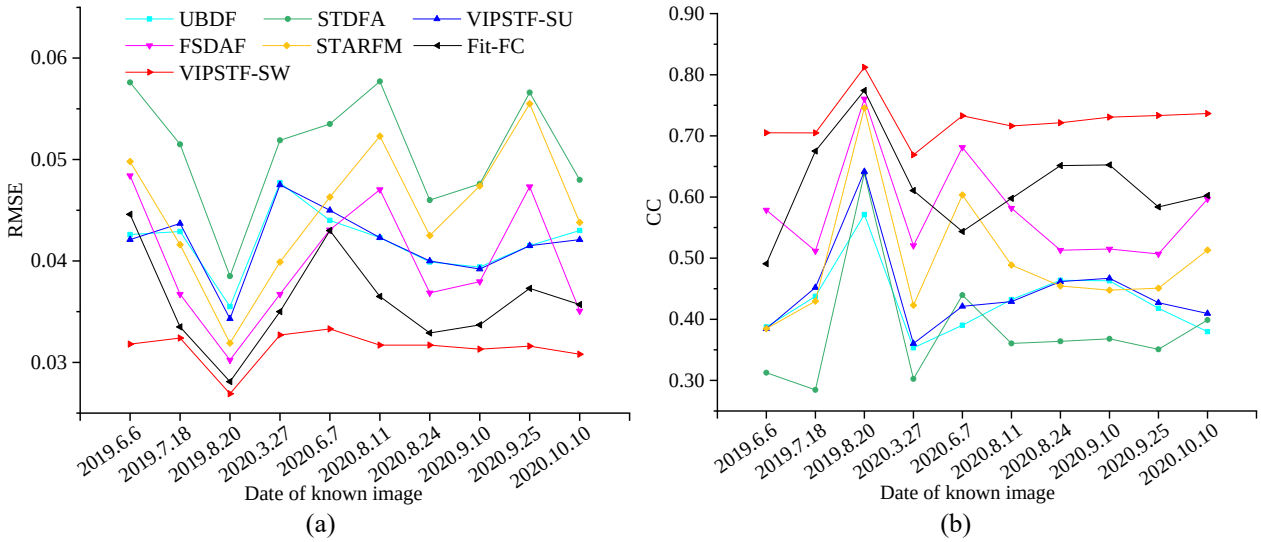


Fig. 9. Accuracies of spatio-temporal fusion based on image pairs collected at different times for Site 1. (a) Mean RMSE of four bands. (b) Mean CC of four bands.

Table 2 Accuracies of different spatio-temporal fusion methods for Site 1 (prediction time on 16 September 2019; image pair on 20 August 2019 as input)

		UBDF	STDFA	VIPSTF-SU	FSDAF	STARFM	Fit-FC	VIPSTF-SW
RMSE	Oa4	0.0207	0.0217	0.0205	0.0189	0.0194	0.0184	**0.0178**
	Oa6	0.0245	0.0265	0.0240	0.0217	0.0227	0.0211	**0.0207**
	Oa8	0.0340	0.0330	0.0310	0.0263	0.0283	0.0259	**0.0249**
	Oa17	0.0626	0.0728	0.0618	0.0540	0.0573	0.0472	**0.0444**
	Mean	0.0355	0.0385	0.0343	0.0302	0.0319	0.0281	**0.0269**
CC	Oa4	0.5885	0.6980	0.6956	0.8108	0.7977	0.7912	**0.8248**
	Oa6	0.4680	0.5040	0.5109	0.6615	0.6316	0.6985	**0.7560**
	Oa8	0.5524	0.6631	0.6508	0.7730	0.7575	0.7790	**0.8109**
	Oa17	0.6762	0.6903	0.7097	0.7969	0.7965	0.8278	**0.8566**
	Mean	0.5713	0.6388	0.6417	0.7606	0.7458	0.7741	**0.8121**

Fig. 9 shows the RMSEs and the CCs of the seven methods using the Sentinel-2 and -3 image pair acquired on different dates (i.e., 6 June 2019 to 10 October 2020, except 16 September 2019, 10 cases in all). For all predictions based on 10 different dates, VIPSTF-SW produces consistently the smallest mean RMSE and the largest mean CC. Specifically, the mean RMSEs and the mean CCs for VIPSTF-SW range from 0.0269 to 0.0422 and 0.6680 to 0.8121, respectively. Also, the accuracy of VIPSTF-SW is the most stable among all the methods (see the CC from 27 March 2020 to 10 October 2020), as VIPSTF-SW is less sensitive to the temporal change of landscapes owing to the construction of the vitual image pair (Wang et al., 2020). Amongst all seven methods, the accuracy of Fit-FC is similar in trend and is the closest to that of VIPSTF-SW. Moreover, FSDAF and STARFM produce less accurate predictions than VIPSTF-SW and Fit-FC, and produce more accurate predictions than UBDF, STDFA and VIPSTF-SU in most cases. To provide a more intuitive presentation of the accuracies for each band produced by the different methods, the bias, RMSE and CC for each case are summarized by blocks with different colors, as shown in Fig. 10. Specifically, each block represents the accuracy evaluation index for one of the four bands of one method, using one image pair of the 10 cases. Thus, for each method, there are 40 blocks for an index, and there are 21 groups of blocks in all. For bias, prediction with zero bias is displayed as white while the prediction with the largest positive bias and negative bias are presented as red and blue, respectively. It is noted that the results for STDFA, FSDAF and STARFM present obvious dark blue blocks in a number of cases, indicating a large bias. Compared with UBDF and VIPSTF-SU, Fit-FC and VIPSTF-SW provide more accurate results in most cases. For the RMSE, darker red represents a larger value. It can be seen that the RMSEs of different methods do not differ as greatly as the bias values and the difference is more noticeable for the results of band Oa17. Amongst the seven methods, the results for VIPSTF-SW present the lightest color in all 40 blocks, especially for the 10 blocks of band Oa17. Moreover, the CCs of different methods vary from 0.17 to 0.85 and the result is shown in dark blue when the CC is large. Obviously, the color in VIPSTF-SW is the darkest in all cases, which indicates that VIPSTF-SW produces the largest CC amongst the seven methods. Therefore, VIPSTF-SW produces consistently the greatest accuracy for all four bands when using different image pairs.

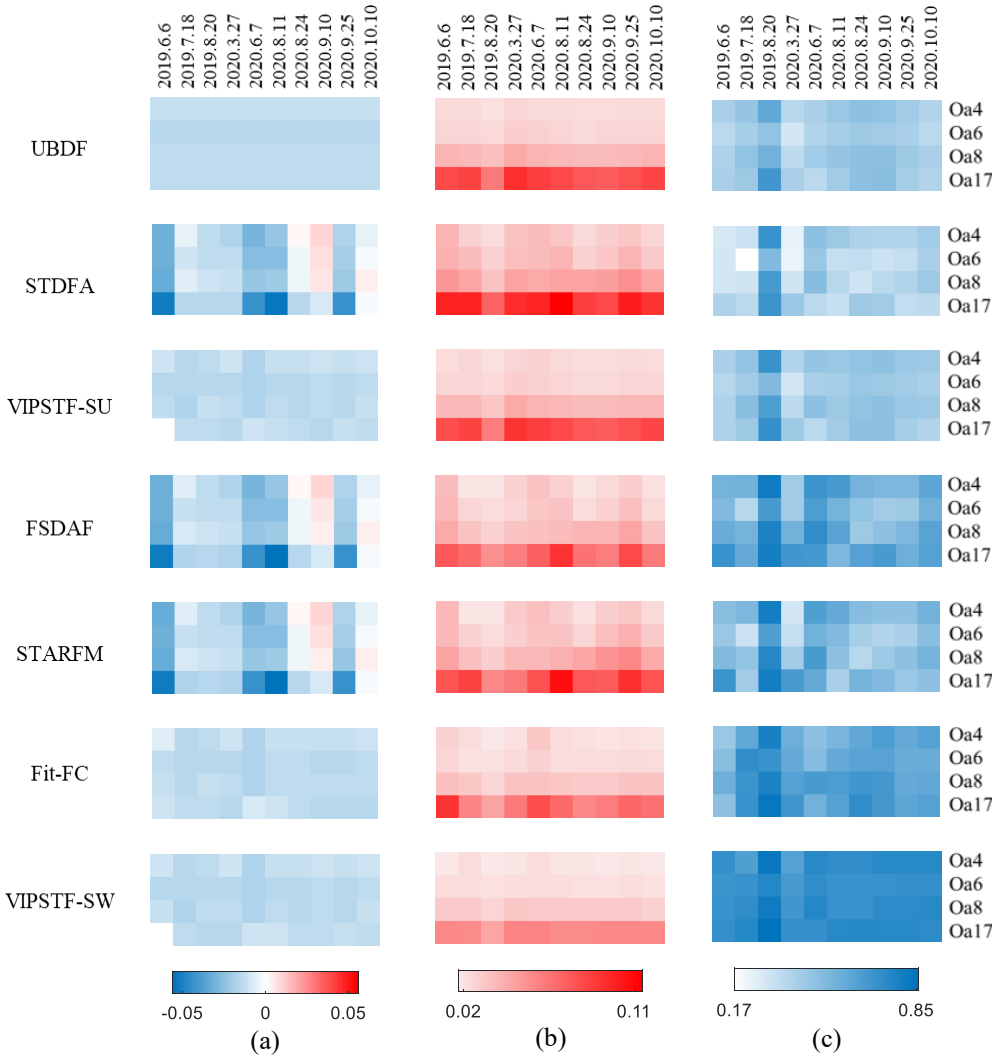


Fig. 10. Accuracies of spatio-temporal fusion for bands Oa4, Oa6, Oa8 and Oa17 based on image pairs collected at different times for Site 1. (a) Bias. (b) RMSE. (c) CC.

3.2.2. Experiment on Site 2

Spatio-temporal fusion was implemented for Site 2 to downscale bands Oa4, Oa6, Oa8 and Oa17 of the Sentinel-3 image acquired on 20 August 2019. The Sentinel-2 and -3 image pairs collected on the other 10 dates were used. The results of using the image pair acquired on 18 July 2019 as input are shown in Fig. 11. For the UBDF, STDFA and VIPSTF-SU results, small fragments exist noticeably, which are substantially different from the reference. Although the FSDAF and STARFM predictions are cleaner compared to the above three methods, the colors are obviously different from the reference image, indicating great spectral

distortion. It can be noted that the Fit-FC and VIPSTF-SW predictions have a more similar color to the reference image (see the red blocks in the subarea). Moreover, the prediction of VIPSTF-SW presents more spatial detail, such as for roads. For quantitative evaluation, the accuracies of the seven methods for Site 2 are displayed in Table 3. It is seen that Fit-FC and VIPSTF-SW produce smaller RMSE and larger CC values than the other five methods. Moreover, VIPSTF-SW produces the largest mean CC of 0.6819, which is 0.0066, 0.0548 and 0.0303 larger than that of Fit-FC, STARFM and FSDAF, respectively. UBDF, STDFA and VIPSTF-SU produce much less accurate predictions, with the mean CCs below 0.52 and mean RMSEs above 0.05 generally.

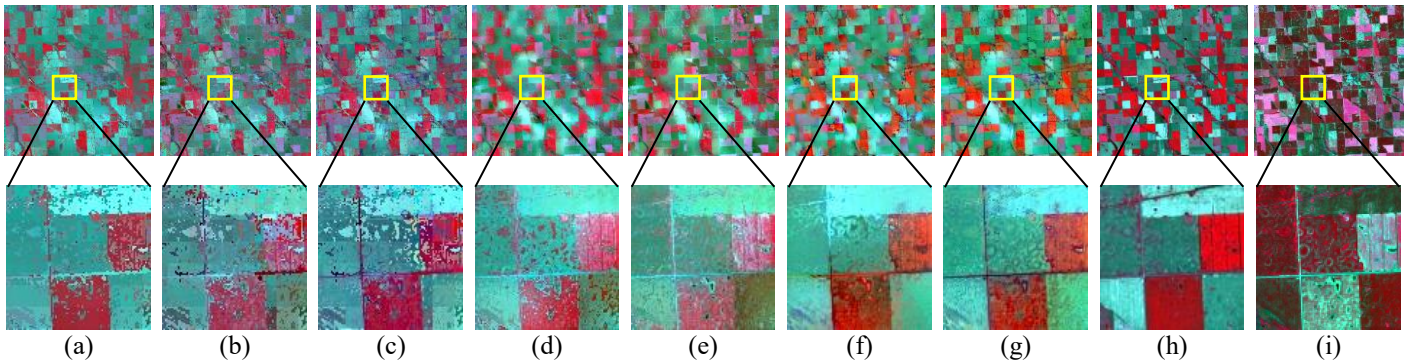


Fig. 11. Results of different spatio-temporal fusion methods for downcaling the 300 m bands Oa4, Oa6, Oa8 and Oa17 to 20 m for Site 2 (prediction time on 20 August 2019; image pair on 18 July 2019 as input) (bands Oa17, Oa8 and Oa6 as RGB). (a) UBDF. (b) STDFA. (c) VIPSTF-SU. (d) FSDAF. (e) STARFM. (f) Fit-FC. (g) VIPSTF-SW. (h) Reference. (i) Input Sentinel-2 image.

Table 3 Accuracies of different spatio-temporal fusion methods for Site 2 (prediction time on 20 August 2019; image pair on 18 July 2019 as input)

		UBDF	STDFA	VIPSTF-SU	FSDAF	STARFM	Fit-FC	VIPSTF-SW
RMSE	Oa4	0.0237	0.0274	0.0240	0.0214	0.0220	**0.0199**	0.0209
	Oa6	0.0276	0.0335	0.0278	0.0244	0.0251	**0.0211**	0.0217
	Oa8	0.0492	0.0551	0.0498	**0.0413**	0.0419	0.0434	0.0420
	Oa17	0.0993	0.1038	0.0969	0.0665	0.0682	0.0682	**0.0624**
	Mean	0.0500	0.0550	0.0496	0.0384	0.0393	0.0382	**0.0368**
CC	Oa4	0.4744	0.4176	0.4528	0.6017	0.5611	**0.6072**	0.5377
	Oa6	0.4725	0.3843	0.4668	0.5760	0.5409	0.6954	**0.7056**
	Oa8	0.5455	0.4798	0.5500	0.6620	0.6523	0.6480	**0.6744**
	Oa17	0.5481	0.5729	0.5759	0.7668	0.7540	0.7508	**0.8097**
	Mean	0.5101	0.4637	0.5114	0.6516	0.6271	0.6753	**0.6819**

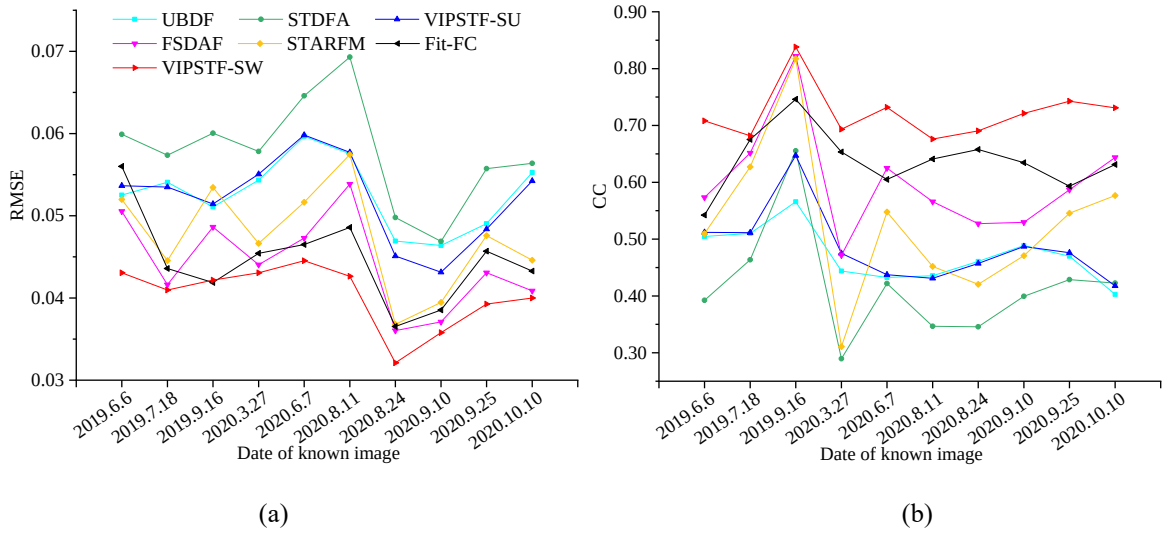


Fig. 12. Accuracies of spatio-temporal fusion based on image pairs collected at different times for Site 2. (a) Mean RMSE of four bands. (b) Mean CC of four bands.

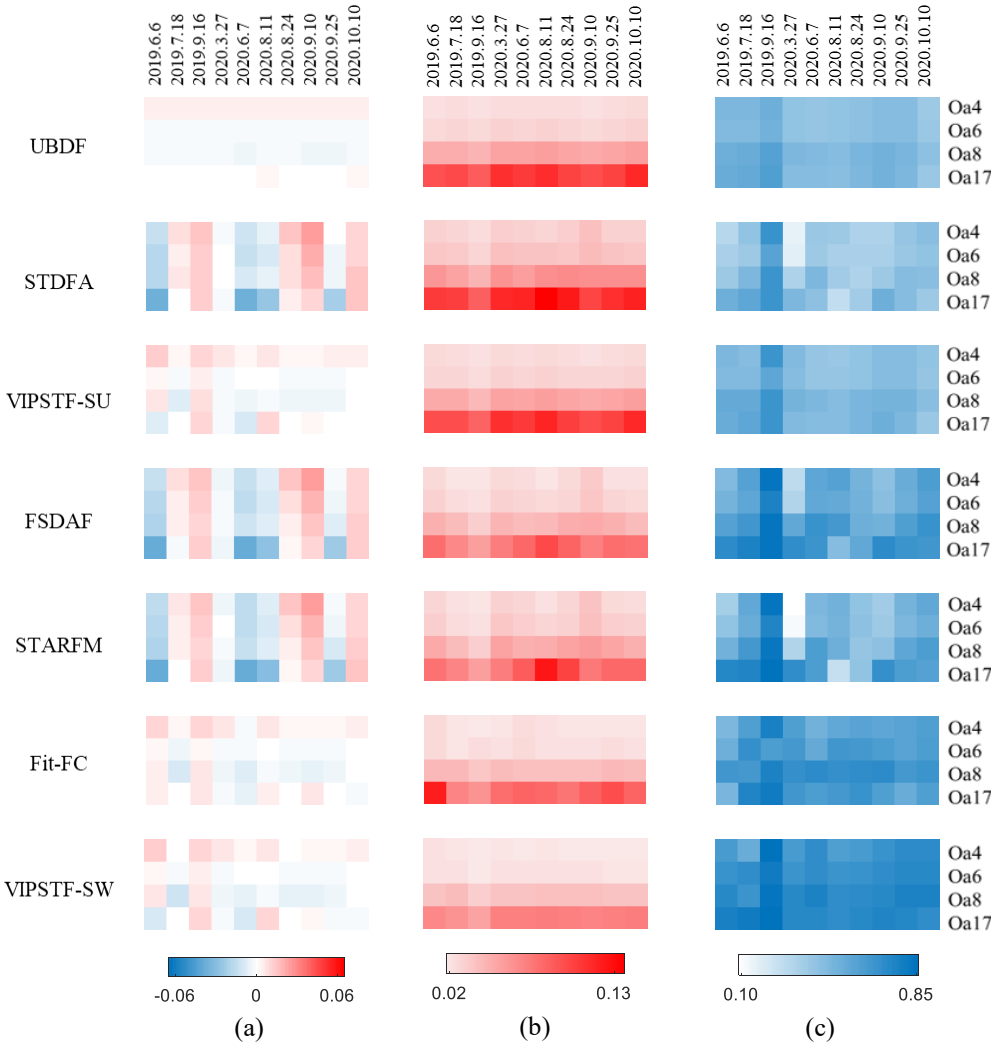


Fig. 13. Accuracies of spatio-temporal fusion for bands Oa4, Oa6, Oa8 and Oa17 based on image pairs collected at different times for Site 2. (a) Bias. (b) RMSE. (c) CC.

The RMSEs and CCs for the predictions using different image pairs are shown in Fig. 12. Clearly, VIPSTF-SW produces consistently the most accurate results, with the smallest mean RMSEs and the largest mean CCs in all cases. Amongst the other six methods, the accuracy of Fit-FC is the closest to that for VIPSTF-SW and both are more accurate than FSDAF, STARFM, UBDF, STDFA and VIPSTF-SU. The biases, RMSEs and CCs for the four bands produced by different methods and image pairs are summarized as blocks in Fig. 13. For STDFA, FSDAF and STARFM, the biases appear as dark blue and red, indicating large errors. Generally, the biases of Fit-FC and VIPSTF-SW are smaller. Moreover, for VIPSTF-SW, the color appears to be the lightest in the RMSE results and the darkest in the CC results, which demonstrates that VIPSTF-SW provides the most accurate predictions.

3.2.3. Land cover mapping based on the spatio-temporal fusion results

To examine the performance of different spatio-temporal fusion methods more comprehensively, land cover classification was also conducted based on different fusion results, as shown in Fig. 14. Specifically, the images were classified into two classes (vegetation and non-vegetation) with *k*-means-based unsupervised classification. It is noted in Fig. 14 that there are obvious speckle artifacts in the classification results of UBDF, STDFA, VIPSTF-SU, FSDAF and STARFM. In comparison with other methods, Fit-FC and VIPSTF-SW produce more satisfactory results, with much less speckle artifacts and greater similarity to the reference. Furthermore, influenced by the smooth effect, Fit-FC fails to reproduce the detailed boundaries of vegetation and several small patches. Thus, the prediction of VIPSTF-SW has the greatest classification accuracy. Quantative evaluation for the classification accuracy was conducted based on the indices of overall accuracy (OA), as listed in Table 4. Checking the classification accuracy for all seven methods, VIPSTF-SW produces the largest OA of 0.9280, with an increase of 0.0091 to 0.0563 compared to Fit-FC and UBDF.

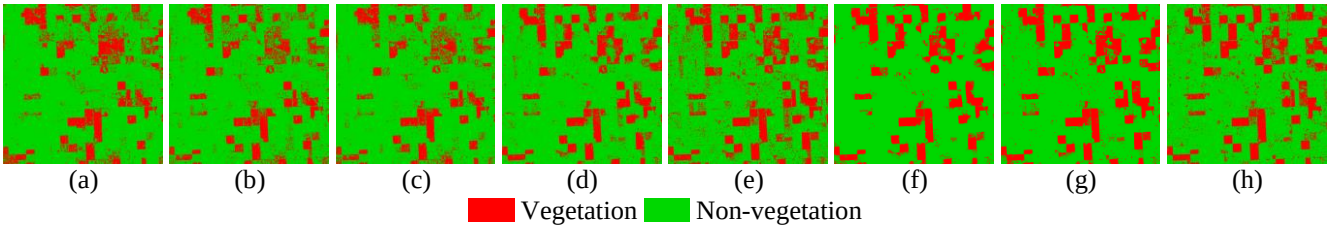


Fig. 14. Land cover mapping based on different spatio-temporal fusion results for Site 1. (a) UBDF. (b) STDFA. (c) VIPSTF-SU. (d) FSDAF. (e) STARFM. (f) Fit-FC. (g) VIPSTF-SW. (h) Reference.

Table 4 Classification accuracy (in terms of OA) for different spatio-temporal fusion results

UBDF	STDFA	VIPSTF-SU	FSDAF	STARFM	Fit-FC	VIPSTF-SW
0.8717	0.8807	0.8782	0.9134	0.9069	0.9189	**0.9280**

3.3. Spatio-spectral fusion for downscaling the other 17 OLCI bands

Spatio-spectral fusion was performed to downscale the other 17 OLCI bands of the Sentinel-3 images at the corresponding prediction times defined in Section 3.2 (i.e., 16 September 2019 for Site 1 and 20 August 2019 for Site 2) to the same target spatial resolution of 20 m. The method identified as the most accurate in the spatio-temporal fusion part, VIPSTF-SW, was applied to spatio-spectral fusion. That is, EIPSSF was implemented using VIPSTF-SW. For VIPSTF-SW-based EIPSSF, 30 similar pixels were selected within each local window with 65 by 65 Sentinel-2 pixels at the spatial resolution of 20 m. Considering the fact that the performance for downscaling the 17 bands cannot be evaluated quantitatively because of the lack of 20 m reference images, a simulation test was performed to validate the feasibility of the VIPSTF-SW-based EIPSSF method. Specifically, one of the 300 m bands Oa4, Oa6, Oa8 and Oa17 was downscaled to 20 m separately, using the 20 m results of the other three bands obtained by spatio-temporal fusion and the four 300 m bands of Sentinel-3 (i.e., three EIPs coupled with a 300 m coarse band to be downscaled). Moreover, the real 20 m image for the target band (i.e., the corresponding band of Sentinel-2 with the same spectral range) was applied for objective evaluation. For example, when downscaling the 300 m band Oa4, the EIPs composed of 20 m and 300 m images of bands Oa6, Oa8 and Oa17, together with the 300 m band Oa4 are used as input, and the band

2 of the Sentinel-2 image (corresponds to band Oa4 of Sentinel-3) serves as reference. For the pre-spatio-spectral fusion step, the known image pairs on 20 August and 16 September 2019 were selected for Sites 1 and 2, respectively. Quantitative evaluation for Sites 1 and 2 is shown in Table 3. In terms of the RMSE, it can be noted that the predictions for band Oa17 are less accurate than for the other three bands, with RMSEs of 0.0550 and 0.0696 for Sites 1 and 2, respectively. For the other three bands, the RMSEs range from 0.0164 to 0.0265 and 0.0177 to 0.0299 for Sites 1 and 2, respectively. Moreover, the CCs range from 0.7420 to 0.8347 and 0.7799 to 0.8775 for Sites 1 and 2, respectively. The CC for Site 2 is generally above 0.80 and the largest CC (i.e., the CC for band Oa8) even reaches 0.8775. Thus, the results suggest that the EIPSSF scheme can produce downscaling results with satisfactory accuracy.

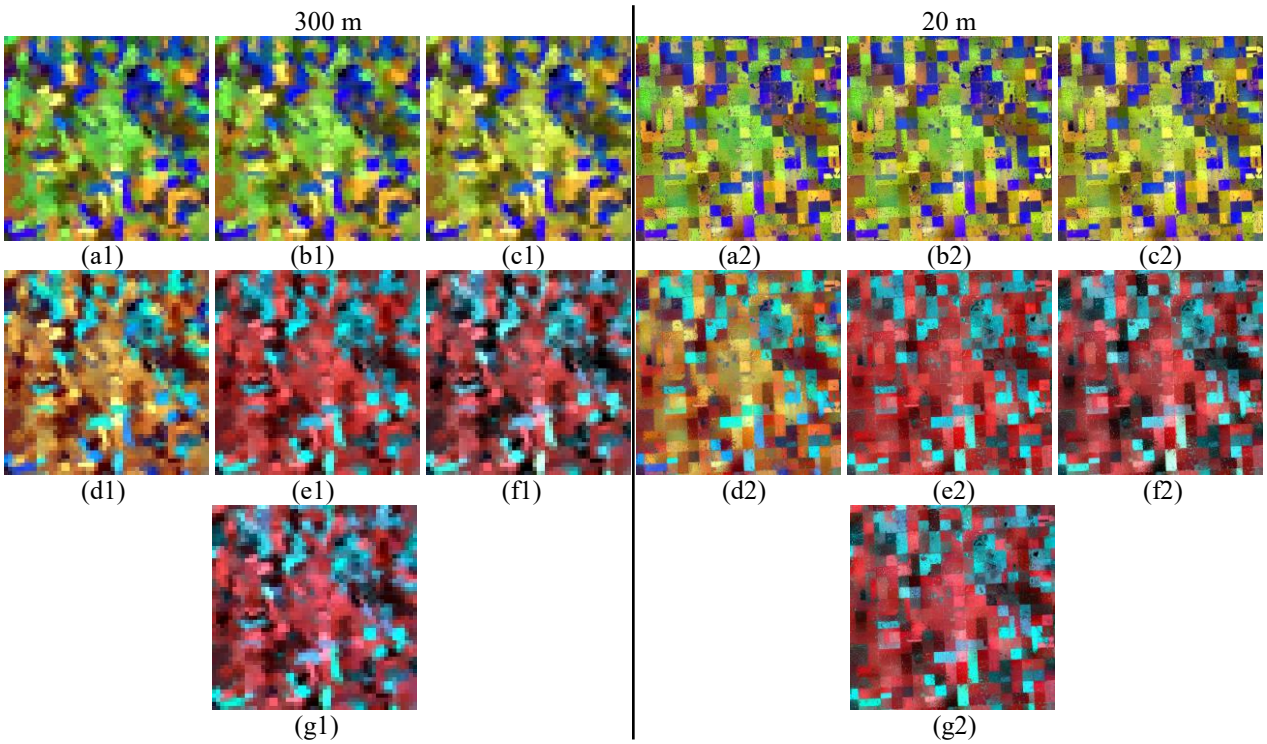


Fig. 15. Downscaling results for all 21 OLCI bands of Sentinel-3 images for Site 1. (a1)-(g1) are original 300 m bands of Sentinel-3 images. (a2)-(g2) are the 20 m downscaling results. (a1) and (a2) are displayed using Oa1, Oa8, Oa15 as RGB. (b1) and (b2) are displayed using Oa2, Oa9, Oa16 as RGB. (c1) and (c2) are displayed using Oa3, Oa10, Oa17 as RGB. (d1) and (d2) are displayed using Oa4, Oa11, Oa18 as RGB. (e1) and (e2) are displayed using Oa5, Oa12, Oa19 as RGB. (f1) and (f2) are displayed using Oa6, Oa13, Oa20 as RGB. (g1) and (g2) are displayed using Oa7, Oa14, Oa21 as RGB.

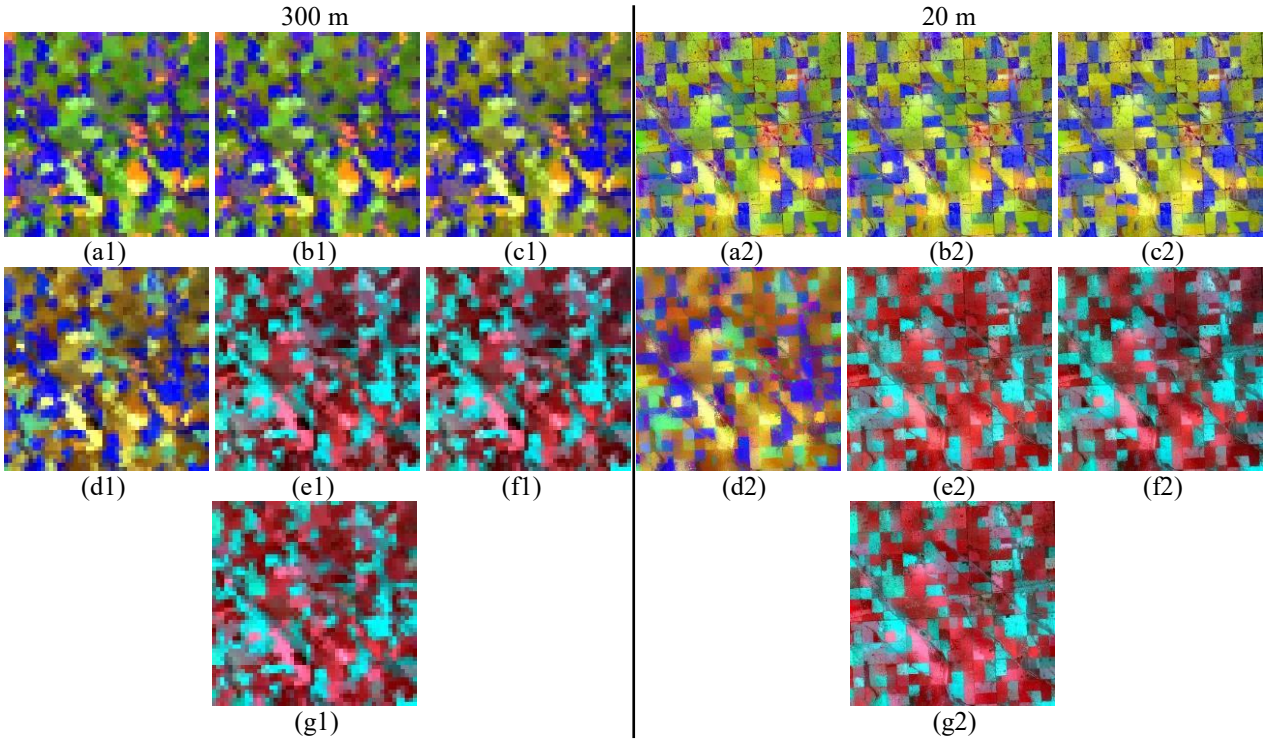


Fig. 16. Downscaling results for all 21 OLCI bands of Sentinel-3 images for Site 1. (a1)-(g1) are original 300 m bands of Sentinel-3 images. (a2)-(g2) are the 20 m downscaling results. (a1) and (a2) are displayed using Oa1, Oa8, Oa15 as RGB. (b1) and (b2) are displayed using Oa2, Oa9, Oa16 as RGB. (c1) and (c2) are displayed using Oa3, Oa10, Oa17 as RGB. (d1) and (d2) are displayed using Oa4, Oa11, Oa18 as RGB. (e1) and (e2) are displayed using Oa5, Oa12, Oa19 as RGB. (f1) and (f2) are displayed using Oa6, Oa13, Oa20 as RGB. (g1) and (g2) are displayed using Oa7, Oa14, Oa21 as RGB.

Table 5 Accuracies of EIPSSF-based spatio-spectral fusion for downscaling OLCI bands Oa4, Oa6, Oa8 and Oa17. When downscaling each band, the coarse-fine EIPs of the other three bands are used as input.

		Oa4	Oa6	Oa8	Oa17
RMSE	Site 1	0.0164	0.0213	0.0265	0.0550
	Site 2	0.0177	0.0196	0.0299	0.0696
CC	Site 1	0.8347	0.7420	0.7929	0.7551
	Site 2	0.8096	0.8020	0.8775	0.7799

Figs. 15 and 16 exhibit the 300 m Sentinel-3 images and the corresponding 20 m downscaling results for Sites 1 and 2, respectively. The spatio-spectral fusion step was implemented by using the VIPSTF-SW-based predictions in Fig. 7 and Fig. 11 as the EIPs for Sites 1 and 2, respectively. The results are displayed using every three bands as RGB, where the bands for each composite were selected according to the principle of the maximum spectral distance to enhance the contrast, that is, the bands furthest apart from each other were

chosen as RGB (e.g., bands Oa1, Oa8 and Oa15 as RGB). Seven different composite images were produced by using different combinations amongst the 21 bands. Compared with the original 300 m Sentinel-3 images, more spatial detail is reproduced obviously in the downscaling results for both Sites 1 and 2 (e.g., the boundaries of the ground objects appear to be much clearer). With the implementation of the proposed fusion framework integrating spatio-temporal-spectral information, all 21 OLCI bands were downscaled satisfactorily.

The ablation analysis was also conducted to show the performances of spatio-spectral fusion where bands 2, 3, 4 and 8a of MSI at the prediction time are available in the input EIPs. That is, no spatio-temporal fusion results are used in the spatio-spectral fusion part. The accuracy for this case is shown in Table 6. Obviously, EIPSSF is again demonstrated to be a satisfactory solution for spatio-spectral fusion, where the CCs are generally above 0.90. Moreover, the accuracy of spatio-spectral fusion in this case is obviously greater than the case where bands 2, 3, 4 and 8a of MSI at the prediction time are unavailable. This reveals that if the bands 2, 3, 4 and 8a of MSI at the prediction time are available, they are certainly preferable choice in EIPSSF, rather than predictions based on pre-spatio-temporal fusion.

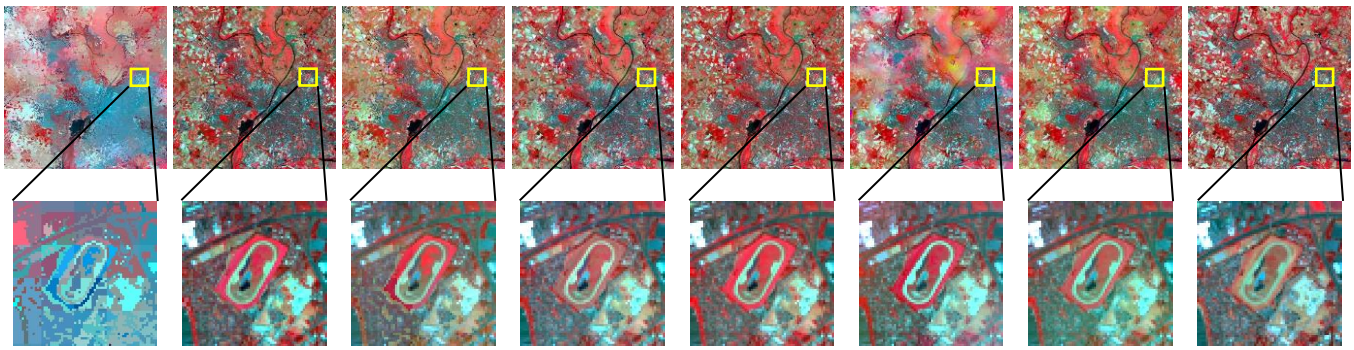
Table 6 Accuracies of the ablation analysis (i.e., bands 2, 3, 4 and 8a of MSI at the prediction time are available) for EIPSSF-based spatio-spectral fusion

		Oa4	Oa6	Oa8	Oa17
RMSE	Site 1	0.0041	0.0041	0.0105	0.0288
	Site 2	0.0070	0.0070	0.0164	0.0489
CC	Site 1	0.9821	0.9884	0.9790	0.9451
	Site 2	0.9863	0.9881	0.9811	0.8941

3.4. Application of the fusion framework to the urban area (Site 3)

To further examine the applicability of the fusion framework, the case study for a more complex landscape in an urban area was also conducted, using the Sentinel-2 and -3 images in Site 3. Compared to Sites 1 and 2, the distribution of land cover in Site 3 presents stronger heterogeneity, bringing greater challenge to spatio-temporal and spatio-spectral fusion. The spatio-temporal fusion results for Site 3 are shown in Fig. 17.

Overall, the predictions of STDFA, VIPSTF-SU, FSDAF, STARFM and VIPSTF-SW are closer to the reference image visually. Checking the sub-area, the FSDAF and VIPSTF-SW results present more similar color to the reference image (see the vegetation). The quantitative evaluation results for the seven methods are displayed in Table 7. It is noted that VIPSTF-SW produces the smallest mean RMSE of 0.0388, which is 0.0025 greater than for FSDAF. As the CC for FSDAF is only 0.0049 larger than for VIPSTF-SW, the performance of FSDAF and VIPSTF-SW can be regarded to be very similar. Again, VIPSTF-SW was chosen to perform the spatio-spectral fusion for downscaling the other 17 bands of OLCI, and the results are shown in Fig. 18. Obviously, the texture appears to be clearer and the boundaries of the river and roads are recovered satisfactorily in the downscaling results. Thus, the proposed fusion framework is also applicable to urban areas with complex spatial structure.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 17. Results of different spatio-temporal fusion methods for downcaling 300 m bands Oa4, Oa6, Oa8 and Oa17 to 20 m for Site 3 (prediction time on 6 August 2020; image pair on 2 June 2020 as input) (bands Oa17, Oa8 and Oa6 as RGB). (a) UBDF. (b) STDFA. (c) VIPSTF-SU. (d) FSDAF. (e) STARFM. (f) Fit-FC. (g) VIPSTF-SW. (h) Reference.

Table 7 Accuracies of different spatio-temporal fusion methods for Site 3 (prediction time on 6 August 2020; image pair on 2 June 2020 as input)

		UBDF	STDFA	VIPSTF-SU	FSDAF	STARFM	Fit-FC	VIPSTF-SW
RMSE	Oa4	0.0403	0.0356	0.0288	0.0334	0.0359	0.0316	**0.0284**
	Oa6	0.0434	0.0362	0.0315	0.0328	0.0359	0.0324	**0.0309**
	Oa8	0.0573	0.0542	0.0489	0.0478	0.0537	0.0474	**0.0460**
	Oa17	0.0854	0.0631	0.0529	0.0514	0.0609	0.0558	**0.0500**
	Mean	0.0566	0.0473	0.0405	0.0414	0.0466	0.0418	**0.0388**
CC	Oa4	0.3278	0.6867	0.6908	**0.7186**	0.6965	0.6412	0.7039
	Oa6	0.3086	0.6437	0.6484	**0.6815**	0.6540	0.6457	0.6630
	Oa8	0.3150	0.5374	0.5215	**0.5885**	0.5476	0.5770	0.5804
	Oa17	0.6235	0.7636	0.8095	0.8108	0.7960	0.8006	**0.8323**
	Mean	0.3938	0.6578	0.6676	**0.6999**	0.6735	0.6661	0.6949

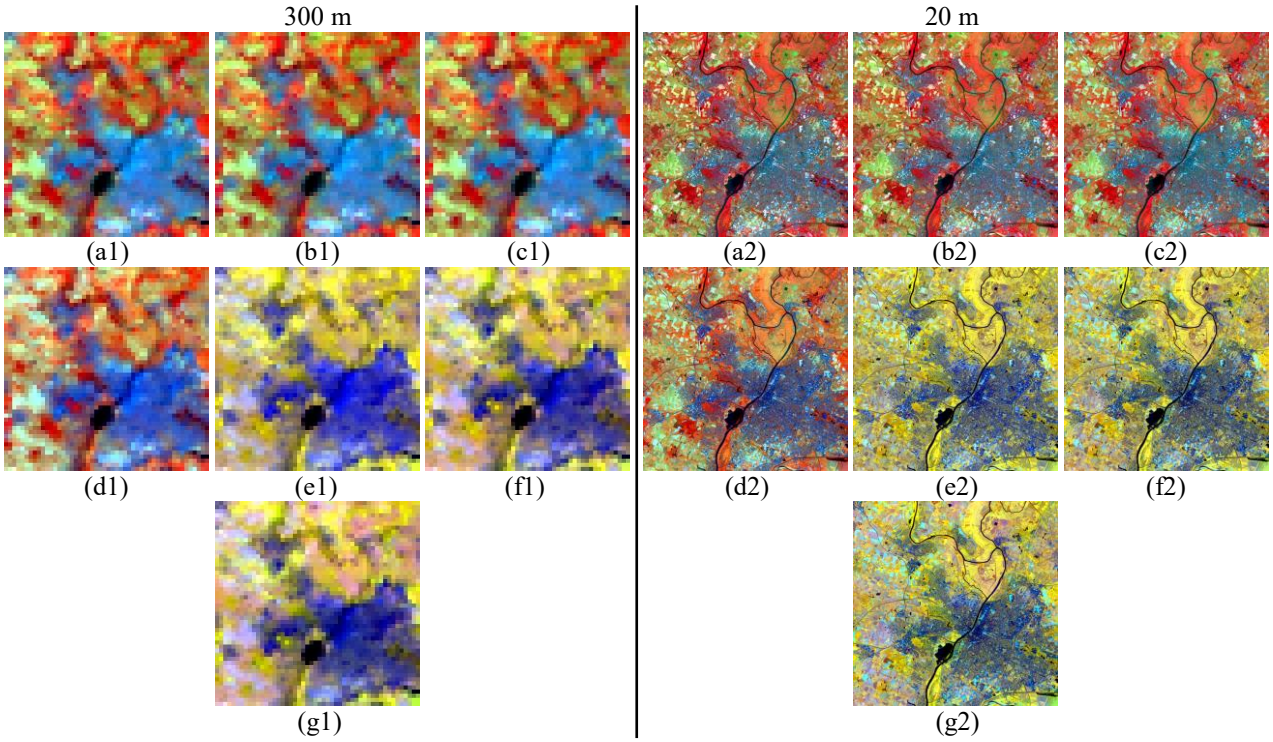


Fig. 18. Downscaling results for all 21 OLCI bands of Sentinel-3 images for Site 3. (a1)-(g1) are original 300 m bands of Sentinel-3 images. (a2)-(g2) are the 20 m downscaling results. (a1) and (a2) are displayed using Oa15, Oa8, Oa1 as RGB. (b1) and (b2) are displayed using Oa16, Oa9, Oa2 as RGB. (c1) and (c2) are displayed using Oa17, Oa10, Oa3 as RGB. (d1) and (d2) are displayed using Oa18, Oa11, Oa4, as RGB. (e1) and (e2) are displayed using Oa19, Oa12, Oa5 as RGB. (f1) and (f2) are displayed using Oa20, Oa13, Oa6 as RGB. (g1) and (g2) are displayed using Oa21, Oa14, Oa7 as RGB.

4. Discussion

4.1. The feasibility of downscaling Sentinel-3 images using spatio-spectral fusion

In this paper, the spatio-temporal fusion method (i.e., VIPSTF-SW) was applied to implement spatio-spectral fusion, which transfers the fusion problem from the temporal domain to the spectral domain. In spatio-temporal fusion, it is validated that the prediction accuracy is greater when the correlation between the images at the known and prediction times is larger (Tang et al., 2020). Generally, high-quality images acquired at times close to the prediction time are involved in the spatio-temporal fusion to ensure a high accuracy. On

the contrary, when a large difference exists between the images at the known and prediction times, the uncertainty will undermine the practical value of spatio-temporal fusion. Similarly, to ensure the feasibility of the spatio-spectral fusion scheme in this paper, a requirement is presented when downscaling a specific coarse band, that is, the spectral distance between the EIPs and the bands for fusion should be limited within a certain range. This can ensure a sufficiently large correlation between the known bands and the bands for downscaling. To investigate the distribution of the data (i.e., BOA reflectance) of the Sentinel-3 images, the box plot for the 21 bands of the Sentinel-3 image acquired on 20 August 2019 for Site 2 are shown in Fig. 19. Generally, the range of the reflectance changes continuously from bands Oa1 to Oa20, with a gradual, increasing trend. Moreover, the box plots for the spectrally adjacent bands appear to be similar (see, e.g., bands Oa8, Oa9 and Oa10).

To further explore the correlation between the bands of Sentinel-3 images, the absolute CCs between every two bands were calculated. To allow a clear presentation of the 231 absolute CCs, we summarize them in color blocks, as shown in Fig. 20. Specifically, the absolute CC is larger when the color is darker. It is obvious that the dark color is distributed mainly amongst bands Oa1 to Oa11, and amongst bands Oa12 to Oa21, indicating a large correlation between these bands. Conversely, relatively smaller correlations lie mainly between bands Oa1 to Oa11 and Oa12 to Oa21. This phenomenon is caused mainly by the difference in spectral range of the bands, as a gap can be observed noticeably between the data of bands Oa11 and Oa12 in Fig. 19. Thus, the majority of the bands of Sentinel-3 images is suitable for spatio-spectral fusion. In EIPSSF, however, the EIPs of all four bands (i.e., bands Oa4, Oa8, Oa6 and Oa17) are involved in the downscaling of the other 17 bands, including the bands with small correlations. Fortunately, the linear transformation strategy in the VIPSTF-SW-based EIPSSF method can reduce the negative influence of these weakly correlated bands effectively. In VIPSTF-SW-based EIPSSF, the virtual EIP is produced by assigning different coefficients to the images of bands Oa4, Oa8, Oa6 and Oa17. In Fig. 21, the values of normalized $|a|$ for 17 bands are exhibited, which represents the contribution of the four bands in the production of the virtual EIP. Specifically,

the normalized $|a_i|$ for each band i is calculated as

$$\text{normalized } |a_i| = \frac{|a_i|}{|a_4|+|a_6|+|a_8|+|a_{17}|},\ i = 4,\ 6,\ 8 \text{ or } 17. \tag{9}$$

When downscaling a specific band amongst the 17 bands, a_4, a_6, a_8 and a_{17} are the coefficients for band Oa4, Oa6, Oa8 and Oa17, respectively.

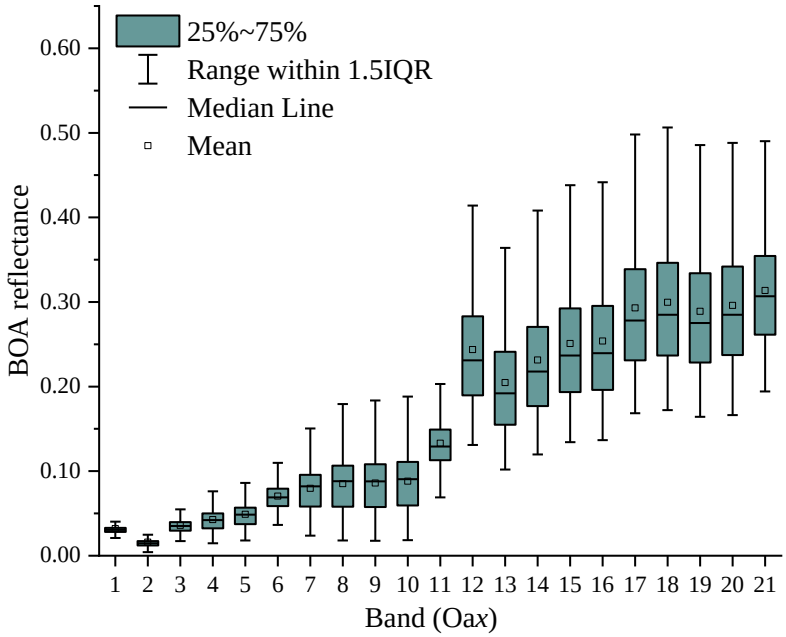


Fig. 19. Box plot of the BOA reflectance of 21 bands for the Sentinel-3 images acquired on 20 August 2019 for Site 2. For each box plot, the interquartile range (IQR) between the first and third quartiles is presented, and the 1.5 IQR is indicated by the whiskers. The upper and lower quartiles are indicated by the top and bottom boundaries of the box, respectively. The square and the line within the box corresponds to the mean and median, respectively.

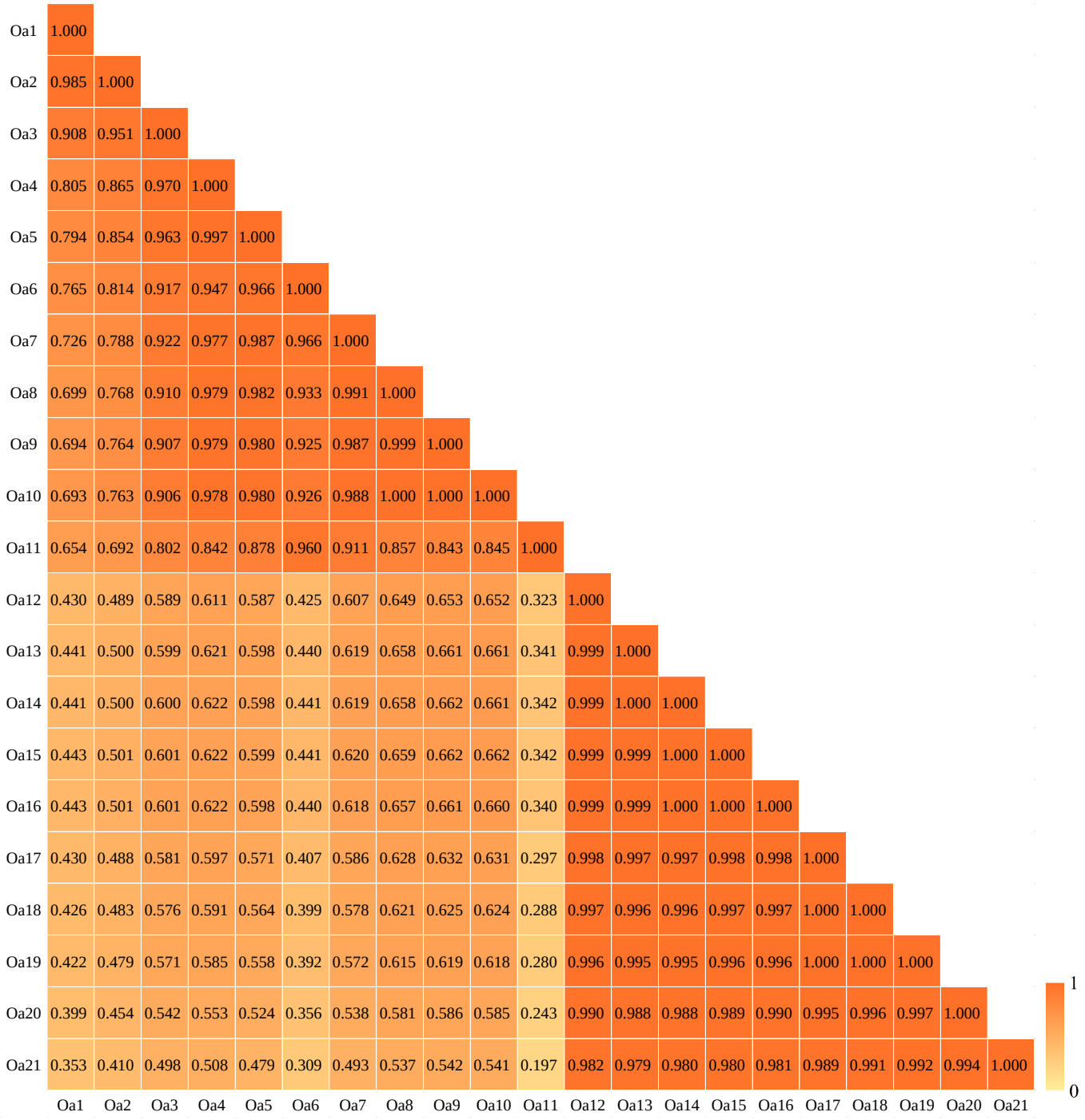


Fig. 20. Absolute CC between OLCI bands of the Sentinel-3 image acquired on 20 August 2019 for Site 2.

Generally, for each band Oa4, Oa6, Oa8 and Oa17, the largest normalized $|a|$ in downscaling lies around the coarse bands spectrally closest to it. Thus, when creating the virtual EIP for downscaling a coarse band, the spectrally closest known bands will be assigned the largest weights, while the other bands will be given

smaller weights. For example, the largest and the smallest normalized $|a|$ for downscaling band Oa10 lie in band Oa8 and Oa17, respectively. This mechanism takes full advantage of the known band images with large correlations, and reduces the negative influence of the bands with small correlations. Therefore, the feasibility of EIPSSF can be ensured.

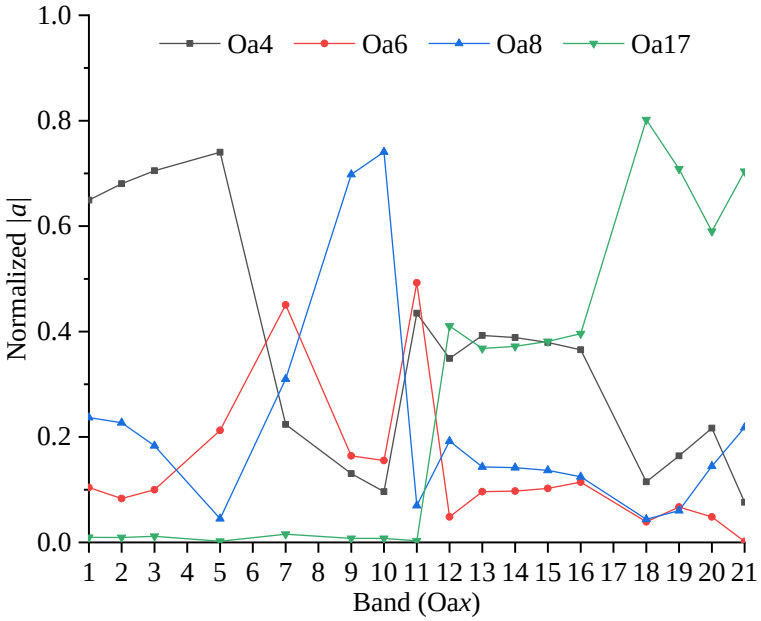


Fig. 21. The normalized $|a|$ of the four known bands (i.e., bands Oa4, Oa6, Oa8 and Oa17) for downscaling the remaining 17 bands of the Sentinel-3 images acquired on 20 August 2019 for Site 2.

4.2. Difference between Sentinel-2 and -3 images

An objective evaluation is presented in this paper for the performance of different spatio-temporal fusion methods based on real Sentinel-3 data. It is necessary to compare this evaluation to previous research performed on simulated Sentinel-3 data, which were created by upscaling Sentinel-2 data (Wang and Atkinson, 2018). To quantify the differences between the two studies, the predictions based on simulated Sentinel-3 images are provided in this section. Specifically, two methods, Fit-FC and VIPSTF-SW, were considered. The former was developed for simulated Sentinel-3 data in Wang and Atkinson (2018), while the

latter was identified to be the most accurate method in the experiments in this paper. Fig. 22 shows the CCs for the predictions using real and simulated Sentinel-3 images for the two methods. The dataset for Site 1 was used, and the known images on different dates were considered. It is obvious that the CCs for the predictions using simulated Sentinel-3 images are consistently larger than for the real Sentinel-3 images. For example, when using the image pair acquired on 24 August 2020 as input for the case of simulated Sentinel-3 images, the CC increases by 0.1582 and 0.0785 for Fit-FC and VIPSTF-SW, respectively. Moreover, Fit-FC and VIPSTF-SW have very close performances for simulated Sentinel-3 images, but VIPSTF-SW is more advantageous for the real case. Since the simulated Sentinel-3 images were produced by degrading the Sentinel-2 images, the difference between the predictions using real and simulated Sentinel-3 images is caused mainly by the difference between the Sentinel-2 and -3 images.

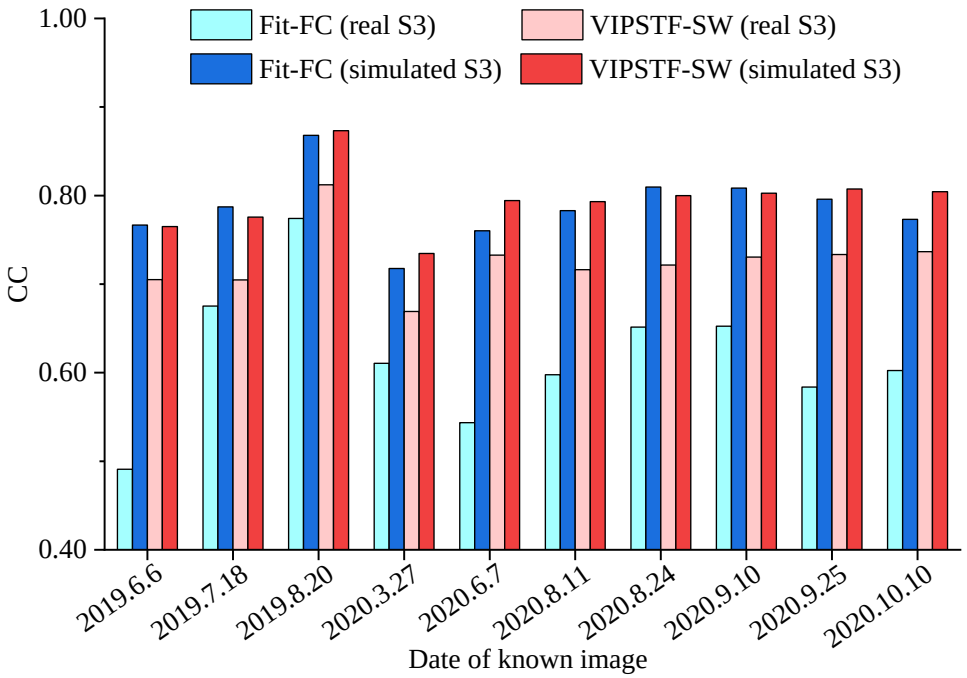


Fig. 22. CC for downscaling bands Oa4, Oa6, Oa8 and Oa17 of the Sentinel-3 images acquired on 16 September 2019 for Site 1 using real and simulated Sentinel-3 images.

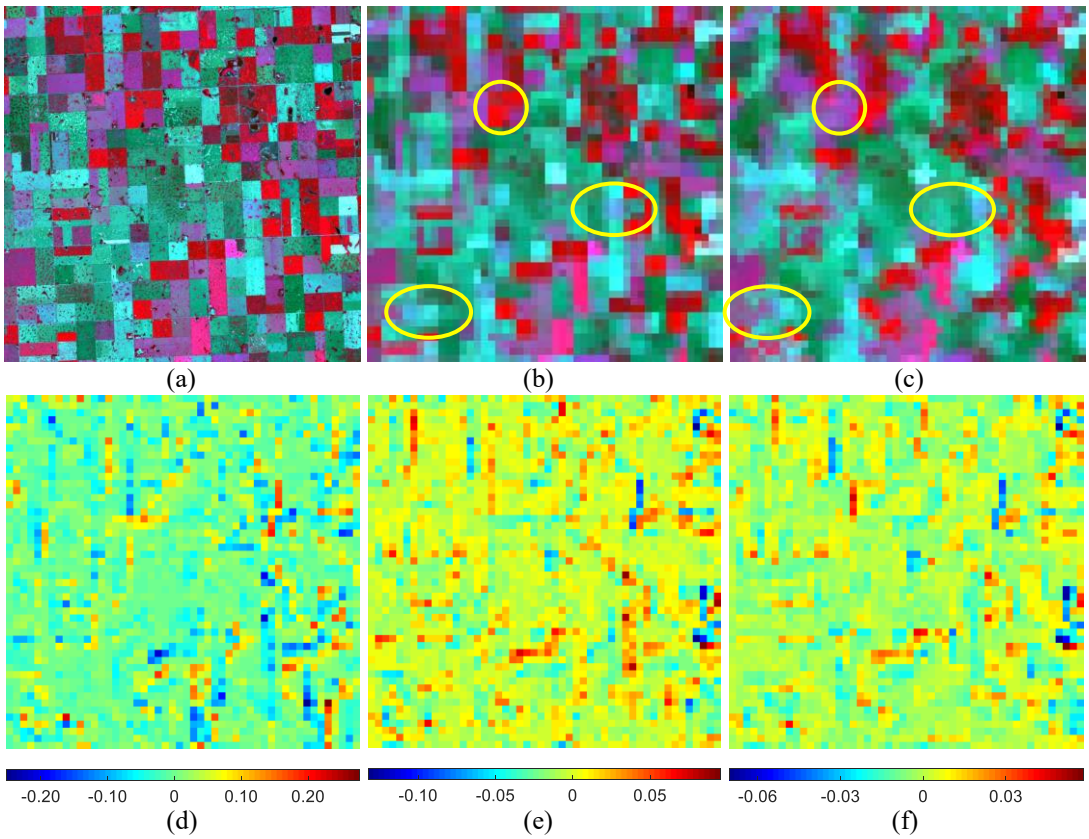


Fig. 23. Comparison between Sentinel-2 and -3 data. (a) Sentinel-2 image acquired on 20 August 2019 for Site 1 (bands 8a, 4 and 3 as RGB). (b) Simulated Sentinel-3 image produced by upscaling (a). (c) Real Sentinel-3 image acquired on 20 August 2019 for Site 1 (bands Oa17, Oa8 and Oa6 as RGB). (d-f) Difference between the real and simulated Sentinel-3 images for (d) Oa17, (e) Oa8 and (f) Oa6.

Table 8 CC between real and simulated Sentinel-3 images for Site 1

	Oa4	Oa6	Oa8	Oa17	Mean
2019.6.6	0.7400	0.7723	0.8126	0.8034	0.7821
2019.7.18	0.7363	0.8160	0.8064	0.8048	0.7909
2019.8.20	0.8028	0.7959	0.8238	0.8364	0.8147
2019.9.16	0.8527	0.8430	0.8501	0.8609	0.8517
2020.3.27	0.8086	0.8173	0.8256	0.8288	0.8201
2020.6.7	0.8142	0.8221	0.8177	0.8046	0.8147
2020.8.11	0.7763	0.8241	0.8127	0.7932	0.8016
2020.8.24	0.8060	0.8183	0.8251	0.8315	0.8202
2020.9.10	0.8548	0.8567	0.8617	0.8479	0.8553
2020.9.25	0.7488	0.7487	0.7468	0.7452	0.7474
2020.10.10	0.8339	0.8347	0.8332	0.8424	0.8360

To further investigate the differences between the Sentinel-2 and -3 images, we chose the images acquired on 16 September 2019 for Site 1 for illustration, as shown in Fig. 23. It is noted that the simulated Sentinel-3

image differs noticeably from the real Sentinel-3 image (see, e.g., the pixels in the yellow circles). As the Sentinel-2 image appears to be in rectangular blocks generally, the simulated Sentinel-3 image also presents a similar blocky distribution, with straight boundaries for objects. For the real Sentinel-3 image, however, the shape of the objects tends to be more irregular, with curving boundaries. The difference images for bands Oa17, Oa8 and Oa6 are also displayed in Fig. 23. Obviously, the main difference lies in the boundaries between land cover classes. Furthermore, the CCs between the real and simulated Sentinel-3 images for Site 1 are provided in Table 8. As can be observed, the CCs for images acquired on different dates range from 0.7363 to 0.8617, with most of the values larger than 0.8000. The difference between the two types of images is caused mainly by acquisition conditions (e.g., Sun-sensor geometry, atmospheric effects, the response function and noise) (Wang et al., 2020). In future research, more mature processing techniques (i.e., radiometric correction and geometric correction, as mentioned in Section 2.2) need to be developed to decrease the differences between Sentinel-2 and -3 images, and more importantly, to increase the accuracy of spatio-temporal fusion.

4.3. The target fine spatial resolution

In this paper, 20 m is selected as the target spatial resolution for downscaling Sentinel-3 images. Actually, among the four bands of Sentinel-2 images for fusion, only the spatial resolution of band 8a is 20 m, while the other three bands have a finer spatial resolution of 10 m. Theoretically, images with finer spatial resolution are deemed to be more advantageous as they can be used for more detailed monitoring. However, the downscaling process can be more challenging as the uncertainty generally increases with the zoom factor. To investigate the influence of the zoom factor on downscaling Sentinel-3 images, we also conducted experiments for downscaling the Oa4, Oa6, Oa8 and Oa17 bands to 10 m, 50 m, 60 m and 100 m, in turn, with the image acquired on 16 September 2019 for Site 1 used as an example, and the image pair acquired on 20 August 2019 used as input. It is noted that when the target spatial resolution is 10 m, band 8a of the Sentinel-2 image needs

to be downscaled to the spatial resolution of 10 m in advance by fusing the 20 m band 8a with the 10 m bands 2, 3 and 4. A previously justified method, area-to-point regression kriging (ATPRK) (Wang et al., 2016), was applied for this purpose. For the other target spatial resolutions (i.e., 50 m, 60 m and 100 m), the 20 m Sentinel-2 images should be upscaled to harmonize the resolution in advance.

The accuracies, or more specifically the prediction precision (in terms of CC), of the different methods are depicted in Fig. 24. Generally, for all methods, the CC increases as the zoom factor decreases, and the predictions at 10 m are the least accurate amongst all predictions. Specifically, the decrease in CC can reach 0.1514 when the zoom factor increases from 3 to 30 for UBDF. Thus, although 10 m predictions are able to provide more spatial detail, the accuracies decrease on the contrary, suggesting that the reliability of the spatial detail at 10 m decreases simultaneously. It is necessary to find a suitable balance between the accuracy of prediction and the target spatial resolution (Wang et al., 2019). In this paper, 20 m was selected as the target spatial resolution since it can satisfy common monitoring requirements. In practice, the target spatial resolution should be basically determined by the requirements of each specific application.

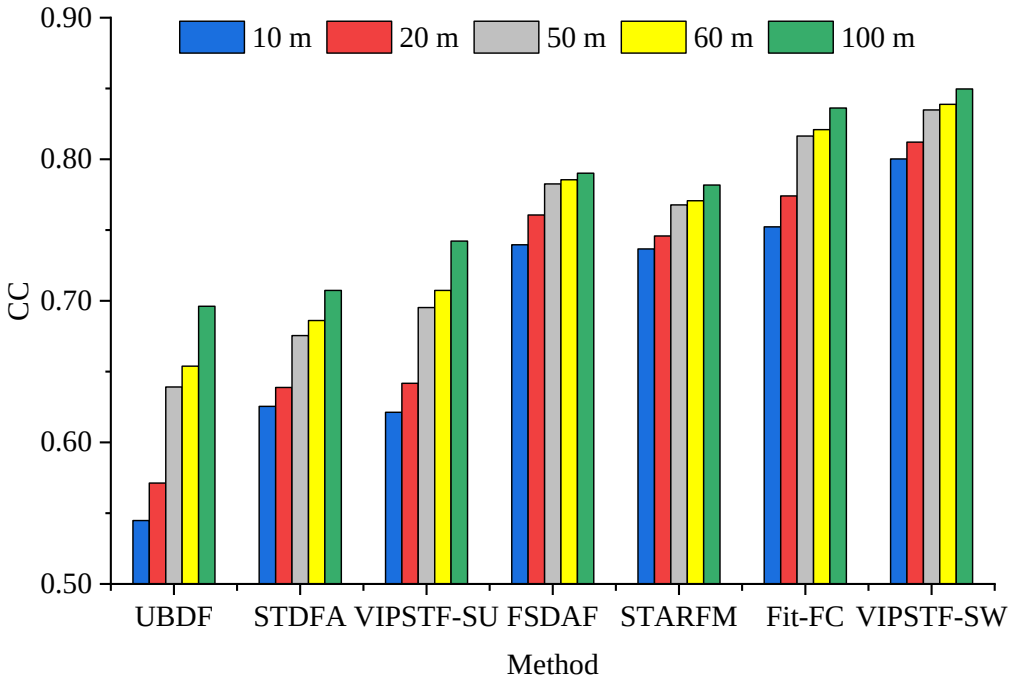


Fig. 24. CC for downcaling Oa4, Oa6, Oa8 and Oa17 bands of the Sentinel-3 image (acquired on 16 September 2019 for Site 1) to 10 m, 20 m, 50 m, 60 m and 100 m.

4.4. The applicability of EIPSSF

In this paper, the EIPSSF method is proposed to downscale the remaining 17 OLCI bands. Based on the principles of various spatio-temporal fusion methods, different EIPSSF methods may be developed. Thus, EIPSSF is a general framework which is applicable to almost all existing spatio-temporal fusion methods. The key of adjusting spatio-temporal fusion methods to EIPSSF is to make full use of the four EIPs. Moreover, different from traditional pan-sharpening, EIPSSF takes full advantage of all fine spatial resolution images as well as their coarse observations. In this paper, a specific form of EIPSSF, VIPSTF-SW-based EIPSSF, is proposed, as the VIPSTF-SW method was shown to be the most accurate choice in spatio-temporal fusion of Sentinel-2 and -3 images. However, it should be stressed that it is also worthwhile to explore more spatio-temporal fusion methods (e.g., Bayesian-based (Li et al., 2013; Shen et al., 2016; Xue et al., 2017) or learning-based (Das and Ghosh, 2016; Liu et al., 2016; Song and Huang, 2012) methods) for the implementation of EIPSSF. In addition, the fusion framework integrating spatio-temporal-spectral information proposed in this paper performs the spatio-temporal and spatio-spectral fusion parts independently. It would also be promising to develop a joint model to consider information of the three aspects (i.e., spatial, temporal and spectral) simultaneously in future research, where the uncertainty in exploring bands 2, 3, 4 and 8a of Sentinel-2 and all 21 Sentinel-3 OLCI bands could be controlled jointly. The potential model may be theoretically superior, but computationally impractical.

4.5. The applicability of the fusion framework integrating spatio-temporal-spectral information

In this paper, the fusion framework integrating spatio-temporal-spectral information was developed to blend two products (i.e., Sentinel-2 and -3) among the Sentinel satellite constellation series. From the perspective of the input data, the implementation of the framework requires images from two sensors, one with fine spatial

resolution, together with another with fine temporal and spectral resolutions. Although this framework is proposed for Sentinel-2 and -3 images, it is also potentially applicable to other sensor systems satisfying the abovementioned requirements. For example, it can be applied to fuse MERIS images (containing 15 bands at the spatial resolution of 300 m and with the revisit period of 2-3 days) with 30 m Landsat TM/ETM+/OLI images to create 15 bands at 30 m. Apart from the MERIS-Landsat sensor system, several other combinations (e.g., MERIS-Sentinel-2 and Sentinel-3-Landsat) are also amenable to this framework. Thus, the fusion framework integrating spatio-temporal-spectral information provides a general solution to fuse images with different spatial, temporal and spectral resolutions.

It should be noted that this framework is designed for fusion of optical data, but may not work for other categories of remote sensing data (e.g., Synthetic Aperture Radar (SAR)). Furthermore, in this paper, Sentinel-2 and -3 images covering three sites were selected to examine the fusion framework integrating spatio-temporal-spectral information. Since this framework is not limited to the spatial pattern of images, it will be worthwhile to test images for more sites from Sentinel-2 and -3 satellites or other optical sensor systems, to validate more widely the effectiveness of the fusion framework in future research.

4.6. Quantitative evaluation of EIPSSF for downscaling the remaining 17 OLCI bands

In the EIPSSF-based spatio-spectral fusion process, only visual comparison between the original Sentinel-3 images and the downscaling results was performed for evaluation. Apart from the visual evaluation, two possible quantitative evaluation strategies could also be considered in future research, but both carrying uncertainties. First, hyperspectral data with finer spatial resolution than 20 m (e.g., aerial data) could be acquired and used as reference for evaluating the downscaling results. The spectral range of the hyperspectral data, however, may not correspond to that of the OLCI bands of Sentinel-3 exactly. Thus, to match the spectral range between the hyperspectral data and the OLCI bands, the convolution operation should first be applied to the hyperspectral data. Specifically, the convolution function can be applied by assigning different weights to

the bands by referring to the spectral response function of Sentinel-3 images. It should be pointed out, however, that the spectral response function is generally defined for spectrally continuous (in a mathematical sense) bands. Although hyperspectral images have a very fine spectral resolution, they are still spectrally discrete. It is a key issue to appropriately define the spectral response function in transformation between the two categories of images. Moreover, the difference in the platform and the acqusition condition between the hyperspectral data and the Sentinel-3 images are also important problems.

Second, quantitative evaluation might be conducted by performing spatio-spectral fusion at a spatial resolution coarser than the original Sentinel-3 images. Specifically, the 300 m Sentinel-3 images can first be degraded to 4.5 km with a scale factor of 15. Correspondingly, the four 20 m Sentinel-2 bands can be degraded to 300 m, which can match the spatial resolution of the observed 300 m Sentinel-3 bands. Based on the 4.5 km degraded Sentinel-3 images and the 300 m degraded Sentinel-2 images, the 300 m images of the 17 OLCI bands can then be predicted by EIPSSF. In this case, the orginal 300 m Sentinel-3 bands can be used to evaluate the accuracy quantitatively. Although the reference image is known perfectly, the practical meaning remains to be considered for this strategy. In geography, it is acknowledged widely that the information presented in images with different spatial resolutions varies greatly owing to the scale effect (Quattrochi and Goodchild, 1997). For example, a method suitable for downscaling from 4.5 km to 300 m may not be a good choice for downscaling the 300 m images to 20 m. In future research, more unsupervised approaches (i.e., without the need of reference) should be investigated for evaluating the downscaling predictions, such as evaluation based on spatial texture.

4.7. The potential of Sentinel-3-downscaled products

This research provides a practical means for processing Sentinel-3 data and for producing Sentinel-3 images with finer spatial resolution (i.e., 20 m in this paper). This research can be of great value cosidering its typical contributions. First, the processing procedure of Sentinel-2 and -3 data presented in this paper can provide

guidance for ESA. Since ESA offers open access to data from the Sentinel missions, the processed data can provide more choices for users when Sentinel data are required for their research. Second, the 20 m Sentinel-3 products can potentially support more applications than the original 300 m data. For example, estimation of local carbon flux, monitoring of vegetation seasonal dynamics and precise characterization of land surface changes place strict requirements on the spatial resolution of images. With the downscaled Sentinel-3 images, all these tasks can be undertaken more reliably. Finally, it is acknowledged that the processing and downscaling approach proposed in this paper is not confined to a certain region. Thus, considering the large number of Sentinel-3 images acquried everyday, the proposed method can be applied to all available Sentinel-3 images. Therefore, daily 20 m Sentinel-3 products can be produced potentially at the global scale, which would provide extremely beneficial support for fine scale monitoring across the entire globe.

5. Conclusion

The Sentinel-3 satellite supported by ESA provides an effective data source for global monitoring of ocean, land and atmosphere. The OLCI sensor onboard the Sentinel-3 satellite provides 21 spectral channels with various significant functions at a coarse spatial resolution of 300 m. To facilitate the application of Sentinel-3 data at a local scale, this paper proposed a fusion framework integrating spatio-temporal-spectral information to downscale the 21 OLCI bands to the 20 m Sentinel-2 spatial resolution through two separate steps, spatio-temporal fusion (by fusing with Sentinel-2 MSI images) and spatio-spectral fusion (based on EIPSSF). The fused images inherit the fine spatial resolution of Sentinel-2, and the fine temporal and spectral resolutions of Sentinel-3. Through the experiments on the three sites, we summarized three main findings.

1) The VIPSTF-SW method is shown to be the most appropriate (through both qualitative and quantitative evaluation) for downscaling the Oa4, Oa6, Oa8 and Oa17 bands of Sentinel-3 OLCI images amongst the seven spatio-temporal fusion methods investigated in this paper.

2) The proposed EIPSSF is a feasible solution to downscale the other 17 bands of Sentinel-3 OLCI images. The 20 m downscaling predictions are visually more pleasant than the orignal 300 m images. The rationale of EIPSSF is based on the large correlation (due to the close spectral distance) between the four EIPs of bands Oa4, Oa6, Oa8 and Oa17 and the other 17 OLCI bands.
3) The proposed spatio-temporal-spectral fusion provides a flexible framework for downscaling the 21 bands of Sentinel-3 OLCI images.

This paper is one of the very few studies for comprehensive downscaling of Sentinel-3 images and will provide important guidance for future studies and applications based on the need for downscaled Sentinel-3 time-series images. The Sentinel-2 and -3 time-series images used in the experiments will be publicly available on https://qunmingwang.github.io/.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant 41971297 and Tongji University under Grant 02502350047.

Appendix A

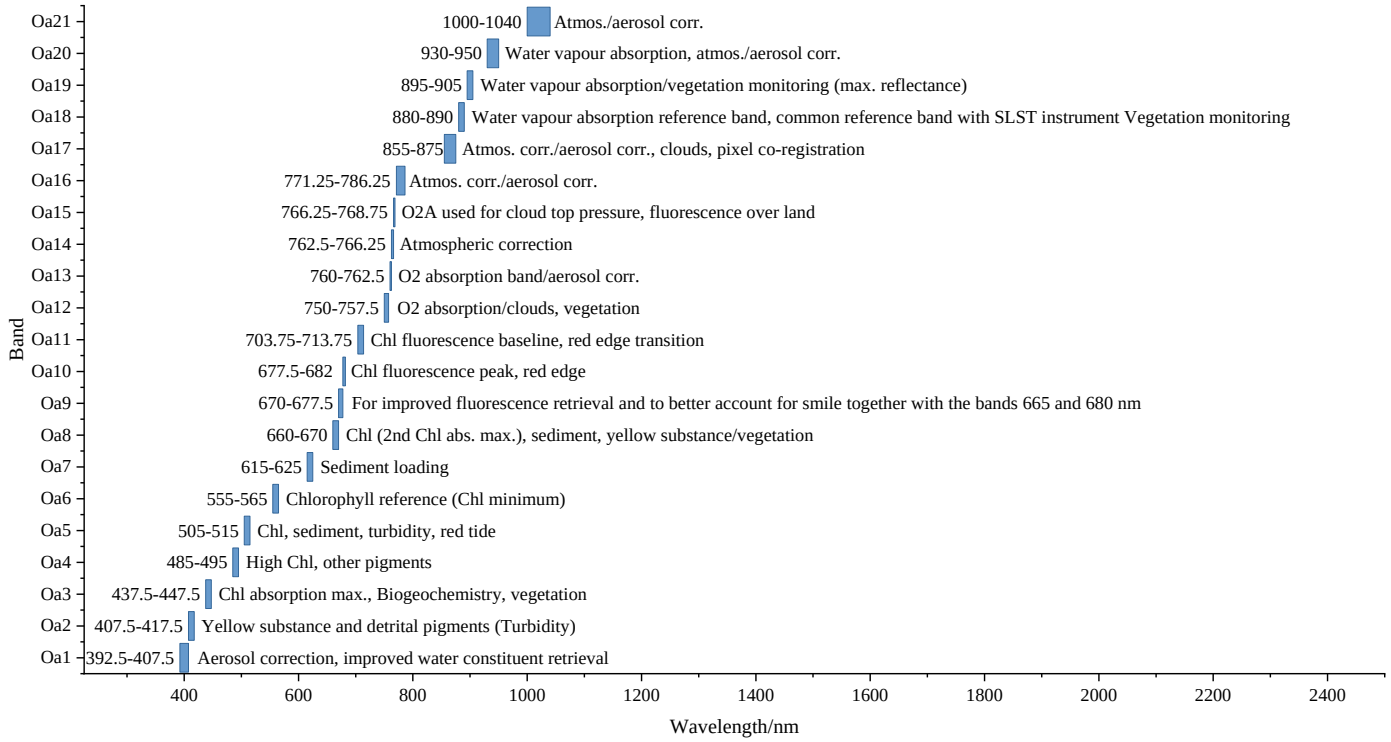


Fig. A1. Band characteristics of the Sentinel-3 OLCI sensor.

References

Amolins, K., Zhang, Y., Dare, P., 2007. Wavelet based image fusion techniques—an introduction, review and comparison. ISPRS Journal of Photogrammetry and Remote Sensing 62(4), 249-263.

Amorós-López, J., Gómez-Chova, L., Alonso, L., Guanter, L., Zurita-Milla, R., Moreno, J., Camps-Valls, G., 2013. Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring. International Journal of Applied Earth Observation and Geoinformation 23, 132-141.

Anderson, G.P., Felde, G.W., Hoke, M.L., Ratkowski, A.J., Cooley, T.W., Chetwynd, J.H., Gardner, J.A., Adler-Golden, S.M., Matthew, M.W., Berk, A., 2002. MODTRAN4-based atmospheric correction algorithm: FLAASH (fast line-of-sight atmospheric analysis of spectral hypercubes). Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII.

Ansper, A., Alikas, K., 2018. Retrieval of Chlorophyll a from Sentinel-2 MSI data for the European Union water framework directive reporting purposes. Remote Sensing 11(1), 64.

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

Berger, M., Aschbacher, J., 2012. Preface: The Sentinel missions—new opportunities for science. Remote Sensing of Environment 120, 1-2.

Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F., Hanssen, R.F., 2012. ESA's Sentinel missions in support of Earth system science. Remote Sensing of Environment 120, 84-90.

Busetto, L., Meroni, M., Colombo, R., 2008. Combining medium and coarse spatial resolution satellite data to improve the estimation of sub-pixel NDVI time series. Remote Sensing of Environment 112(1), 118-131.

Cazzaniga, I., Bresciani, M., Colombo, R., Bella, D.V., Padula, R., Giardino, C., 2019. A comparison of Sentinel-3-OLCI and Sentinel-2-MSI-derived Chlorophyll- a maps for two large Italian lakes. Remote Sensing Letters 10, 978-987.

Chavez Jr., P.S., Sides, S.C., Anderson, J.A., 1991. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. Photogrammetric Engineering and Remote Sensing 57(3), 295–303.

Chen, B., Huang, B., 2015. Comparison of spatiotemporal fusion models: a review. Remote Sensing 7(2), 1798-1835.

Das, M., Ghosh, S.K., 2016. Deep-STEP: a deep Learning approach for spatiotemporal prediction of remote sensing data. IEEE Geoscience and Remote Sensing Letters 13(12), 1984-1988.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., Sciarra, R., 2012. The global monitoring for Environment and Security (GMES) Sentinel-3 mission. Remote Sensing of Environment 120, 37-57.

Drinkwater, M.R., Helge, R., 2007. Sentinel-3: Mission requirements document version 2.0.

Drusch, M., Bello, D.U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment 120, 25-36.

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., Li, X., 2016. Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR Band. Remote Sensing 354, 1-19.

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 Transactions on Geoscience and Remote Sensing 44(8), 2207-2218.

Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., Yang, Y., 2015. Fusing Landsat and MODIS data for vegetation monitoring. IEEE Geoscience and Remote Sensing Magazine 3, 47-60.

Garzelli, A., 2016. A review of image fusion algorithms based on the super-resolution paradigm. Remote Sensing 8(10), 797.

Gevaert, C., García-Haro, F., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. Remote Sensing of Environment 156, 34–44.

Giannini, F., Hunt, B.P.V., Jacoby, D., Costa, M., 2021. Performance of OLCI Sentinel-3A satellite in the Northeast Pacific coastal waters. Remote Sensing of Environment 256, 112317.

Guzinski, R., Nieto, H., 2019. Evaluating the feasibility of using Sentinel-2 and Sentinel-3 satellites for high-resolution evapotranspiration estimations. Remote Sensing of Environment 221, 157-172.

Hagolle, O., Sylvander, S., Huc, M., Claverie, M., Clesse, D., Dechoz, C., Lonjou, V., Poulain, V., 2015. SPOT-4 (Take 5): simulation of Sentinel-2 time series on 45 large sites. Remote Sensing 7(9), 12242-12264.

Hilker, T., Wulder, M.A., Coops, N.C., Linke, J., McDermid, G., Masek, J.G., Gao, F., White, J.C., 2009. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. Remote Sensing of Environment 113(8), 1613-1627.

Javan, F.D., Samadzadegan, F., Mehravar, S., Toosi, A., Khatami, R., Stein, A., 2021. A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery. ISPRS Journal of Photogrammetry and Remote Sensing 171, 101-117.

Khan, M.M., Chanussot, J., Condat, L., Montavert, A., 2008. Indusion: fusion of multispectral and panchromatic images using the induction scaling technique. IEEE Geoscience and Remote Sensing Letters 5(1), 98–102.

Kravitz, J., Matthews, M., Bernard, S., Griffith, D., 2020. Application of Sentinel 3 OLCI for chl-a retrieval over small inland water targets: Successes and challenges. Remote Sensing of Environment 237, 111562.

Laben, C.A., Brower, B.V., 2000. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening. U.S. Patent, 6011875.

Lefebvre, A., Sannier, C., Corpetti, T., 2016. Monitoring urban areas with Sentinel-2A data: application to the update of the copernicus high resolution layer imperviousness degree. Remote Sensing 8(7), 606.

Li, A., Bo, Y., Zhu, Y., Guo, P., Bi, J., He, Y., 2013. Blending multi-resolution satellite sea surface temperature (SST) products using Bayesian maximum entropy method. Remote Sensing of Environment 135, 52-63.

Li, X., Foody, G.M., Boyd, D.S., Ge, Y., Zhang, Y., Du, Y., Ling, F., 2020. SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion. Remote Sensing of Environment 237, 111537.

Liu, J.G., 2000. Smoothing filter-based intensity modulation: a spectral preserve image fusion technique for improving spatial details. International Journal of Remote Sensing 21(18), 3461–3472.

Liu, M., Yang, W., Zhu, X., Chen, J., Chen, X., Yang, L., Helmer, E.H., 2019. An Improved Flexible Spatiotemporal DAta Fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series. Remote Sensing of Environment 227, 74-89.

Liu, X., Deng, C., Wang, S., Huang, G.-B., Zhao, B., Lauren, P., 2016. Fast and accurate spatiotemporal fusion based upon extreme learning machine. IEEE Geoscience and Remote Sensing Letters 13(12), 2039-2043.

Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M.E., García-Santos, G., Fernandes, R., Berger, M., 2012. Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. Remote Sensing of Environment 120, 91-101.

Mustafa, Y.T., Tolpekin, V.A., Stein, A., 2014. Improvement of spatio-temporal growth estimates in heterogeneous forests using Gaussian Bayesian networks. IEEE Transactions on Geoscience and Remote Sensing 52(8), 4980-4991.

Nieke, J., Borde, F., Mavrocordatos, C., Berruti, B., Delclaud, Y., Riti, J.B., Garnier, T., 2012. The Ocean and Land Colour Imager (OLCI) for the Sentinel 3 GMES Mission: status and first test results. Earth Observing Missions and Sensors: Development, Implementation, and Characterization II.

Padwick, C., Deskevich, M., Pacifici, F., Smallwood, S., 2010. WorldView-2 pansharpening. Proceedings of the ASPRS 2010 Annual Conference, San Diego, CA, USA.

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sensing of Environment 240, 111604.

Quattrochi, D.A., Goodchild, M.F., 1997. Scale in remote sensing and GIS. Lewis Publishers, New York.

Restaino, R., Vivone, G., Dalla Mura, M., Chanussot, J., 2016. Fusion of multispectral and panchromatic images based on morphological operators. IEEE Transactions on Image Processing 25, 2882–2895.

Seitz, B., Mavrocordatos, C., Rebhan, H., Nieke, J., Klein, U., Borde, F., Berruti, B., 2010. The sentinel-3 mission overview. IEEE International Geoscience and Remote Sensing Symposium.

Shen, H., Meng, X., Zhang, L., 2016. An integrated framework for the spatio–temporal–spectral fusion of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing 54(12), 7135-7148.

Shettigara, V.K., 1992. A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. Photogrammetric Engineering and Remote Sensing 58(5), 561–567.

Shen, M., Duan, H., Cao, Z., Xue, K., Qi, T., Ma, J., Liu, D., Song, K., Huang, C., Song, X., 2020. Sentinel-3 OLCI observations of water clarity in large lakes in eastern China: implications for SDG 6.3.2 evaluation. Remote Sensing of Environment 247, 111950.

Song, H., Huang, B., 2012. Spatiotemporal reflectance fusion via sparse representation. IEEE Transactions on Geoscience and Remote Sensing 50(10), 3707-3716.

Tang, Y., Wang, Q., Zhang, K., Atkinson, P.M., 2020. Quantifying the effect of registration error on spatio-temporal fusion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 487-503.

Tu, T.-M., Su, S.-C., Shyu, H.-C., Huang, P.S., 2001. A new look at IHS-like image fusion methods. Information Fusion 2(3), 177–186.

Verhoef, W., Bach, H., 2012. Simulation of Sentinel-3 images by four-stream surface–atmosphere radiative transfer modeling in the optical and thermal domains. Remote Sensing of Environment 120, 197-207.

Wang, Q., Shi, W., Li, Z., Atkinson, P.M., 2016. Fusion of Sentinel-2 images. Remote Sensing of Environment 187, 241-252.

Wang, Q., Atkinson, P.M., 2018. Spatio-temporal fusion for daily Sentinel-2 images. Remote Sensing of Environment 204, 31-42.

Wang, Q., Tang, Y., Tong, X., Atkinson, P.M., 2020. Virtual image pair-based spatio-temporal fusion. Remote Sensing of Environment 249, 112009.

Wang, Q., Ding, X., Tong, X., Atkinson, P.M., 2021a. Spatio-temporal spectral unmixing of time-series images. Remote Sensing of Environment 259, 112407.

Wang, Q., Wang, L., Wei, C., Jin, Y., Li, Z., Tong, X. Atkinson, P.M.,2021b. Filling gaps in Landsat ETM+ SLC-off images with Sentinel-2 images. International Journal of Applied Earth Observation and Geoinformation 101, 102365.

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. Journal of Applied Remote Sensing 6(1), 063507.

Wu, P., Shen, H., Zhang, L., Göttsche, F.-M., 2015. Integrated fusion of multi-scale polar-orbiting and geostationary satellite observations for the mapping of high spatial and temporal resolution land surface temperature. Remote Sensing of Environment 156, 169-181.

Xie, W., Lei, J., Cui, Y., Li, Y., Du, Q., 2019. Hyperspectral pansharpening with deep priors. IEEE Transactions on Neural Networks and Learning Systems 31(5), 1529-1543.

Xiong, Z., Guo, Q., Liu, M., Li, A, 2021. Pan-sharpening based on convolutional neural network by using the loss function with no-reference. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 897-906.

Xu, F., Somers, B., 2021. Unmixing-based Sentinel-2 downscaling for urban land cover mapping. ISPRS Journal of Photogrammetry and Remote Sensing 171, 133-154.

Xu, Y., Huang, B., Dr, Y., Cao, K., Guo, C., Meng, D., 2015. Spatial and temporal image fusion via regularized spatial unmixing. IEEE Geoscience and Remote Sensing Letters 12(6), 1362-1366.

Xue, J., Leung, Y., Fung, T., 2017. A Bayesian data fusion approach to spatio-temporal fusion of remotely sensed images. Remote Sensing 9(12), 1310.

Xue, K., Ma, R., Duan, H., Shen, M., Boss, E., Cao, Z., 2019. Inversion of inherent optical properties in optically complex waters using sentinel-3A/OLCI images: a case study using China's three largest freshwater lakes. Remote Sensing of Environment 225, 328-346.

Zhang, H., Huang, B., Zhang, M., Cao, K., Yu, L., 2015. A generalization of spatial and temporal fusion methods for remotely sensed surface parameters. International Journal of Remote Sensing 36, 4411-4445.

Zhang, Y., Liu, C., Sun, M., Ou, Y., 2019. Pan-sharpening using an efficient bidirectional pyramid network. IEEE Transactions on Geoscience and Remote Sensing 57(8), 5549-5563.

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 Sensing of Environment 252, 112130.

Zhou, X., Wang, P., Tansey, K., Ghent, D., Zhang, S., Li, H., Wang, L., 2020. Drought monitoring using the Sentinel-3-based multiyear vegetation temperature condition index in the Guanzhong plain, China. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 129-142.

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 Sensing of Environment 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 Sensing of Environment 172, 165-177.

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 Sensing 10(4), 527.

Zhukov, B., Oertel, D., Lanzl, F., Reinhackel, G., 1999. Unmixing-based multisensor multiresolution image fusion. IEEE Transactions on Geoscience and Remote Sensing 37(3), 1212-1226.

Zurita-Milla, R., Clevers, J., Schaepman, M.E., 2008. Unmixing-based Landsat TM and MERIS FR data fusion. IEEE Geoscience and Remote Sensing Letters 5(3), 453-457.

Zurita-Milla, R., Gomez-Chova, L., Guanter, L., Clevers, J.G.P.W., Camps-Valls, G., 2011. Multitemporal unmixing of Medium-Spatial-Resolution Satellite Images: a case study using MERIS images for land-cover mapping. IEEE Transactions on Geoscience and Remote Sensing 49(11), 4308-4317.