

Spatio-temporal fusion for daily Sentinel-2 images

Qunming Wang^a, Peter M. Atkinson^{a,b,c,*}

^a Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK

^b Geography and Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK

^c School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, BT7 1NN, Northern Ireland, UK

*Corresponding author. E-mail: pma@lancaster.ac.uk

Abstract: Sentinel-2 and Sentinel-3 are two newly launched satellites for global monitoring. The Sentinel-2 Multispectral Imager (MSI) and Sentinel-3 Ocean and Land Colour Instrument (OLCI) sensors have very different spatial and temporal resolutions (Sentinel-2 MSI sensor 10 m, 20 m and 60 m, 10 days, albeit 5 days with 2 sensors, conditional upon clear skies; Sentinel-3 OLCI sensor 300 m, <1.4 days with 2 sensors). For local monitoring (e.g., the growing cycle of plants) one either has the desired spatial or temporal resolution, but not both. In this paper, spatio-temporal fusion is considered to fuse Sentinel-2 with Sentinel-3 images to create nearly daily Sentinel-2 images. A challenging issue in spatio-temporal fusion is that there can be very few cloud-free fine spatial resolution images temporally close to the prediction time, or even available, strong temporal (i.e., seasonal) changes may exist. To this end, a three-step method consisting of regression model fitting (RM fitting), spatial filtering (SF) and residual compensation (RC) is proposed, which is abbreviated as Fit-FC. The Fit-FC method can be performed using only one Sentinel-3–Sentinel-2 pair and is advantageous for cases involving strong temporal changes (i.e., mathematically, the correlation between the two Sentinel-3 images is small). The effectiveness of the method was validated using two datasets. The created nearly daily Sentinel-2 time-series images have great potential for timely monitoring of highly dynamic environmental, agricultural or ecological phenomena.

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

26 **1. Introduction**

27

28 Sentinel-2 is a new programme of the European Space Agency (ESA) for fine spatial resolution global
29 monitoring ([Drusch et al., 2012](#); [Hagolle et al., 2015](#); [Segl et al., 2015](#)). The Sentinel-2A and -2B satellites
30 were launched on 23 June 2015 and 7 March 2017, respectively. The twin satellites are in the same orbit and
31 180 ° apart from each other and they are now releasing data routinely. The Sentinel-2 Multispectral Imager
32 (MSI) provides 13 spectral bands in the visible, near infrared (NIR) and short wave infrared (SWIR)
33 wavelengths, with four bands at 10 m (centered at 490 nm, 560 nm, 665 nm and 842 nm), six bands at 20 m
34 (centered at 705 nm, 740 nm, 783 nm 865 nm, 1610 nm, and 2190 nm) and three bands at 60 m spatial
35 resolution (centered at 443 nm, 940 nm and 1375 nm) ([Drusch et al., 2012](#); [Du et al., 2016](#); [Hagolle et al., 2015](#);
36 [Wang et al., 2016](#)). The Sentinel-2 data can be used to support global land services including monitoring
37 vegetation, soil and water cover, etc.. Such data are receiving increasing attention in remote sensing studies
38 and applications ([Fernández-Manso et al., 2016](#); [Immitzer et al., 2016](#); [Novelli et al., 2016](#); [Storey et al., 2016](#);
39 [Van der Werff and Van der Meer, 2016](#)). The Sentinel-2A or -2B satellite can revisit the same area every 10
40 days (5 days with the twin satellites together). Due to cloud and shadow contamination, however, it generally
41 requires more than 5 days (e.g., probably several months) to acquire a cloud-free Sentinel-2 image for specific
42 areas. The temporally sparse Sentinel-2 observations, especially for areas that can be easily covered by clouds,
43 are not sufficient for monitoring rapid changes such as growing cycle of plants.

44 Sentinel-3, another very new programme of the ESA, is a satellite imaging mission designed for global
45 monitoring for environment and security (GMES) to ensure frequent and near real-time measurements to
46 ocean, land, and atmospheric services ([Berger and Aschbacher, 2012](#); [Donlon et al., 2012](#); [Verhoef and Bach,](#)
47 [2012](#)). The Sentinel-3A satellite was launched on 16 February 2016. The instrument of the satellite includes a
48 Sea and Land Surface Temperature Radiometer (SLSTR), a Synthetic Aperture Radar Altimeter (SRAL) and
49 an Ocean and Land Colour Imager (OLCI). The OLCI sensor delivers 21-band wide-swath optical images at a
50 temporal resolution of less than 2.8 days (will be increased to less than 1.4 days after the launch of the twin

51 satellite Sentinel-3B). Compared to Sentinel-2 MSI, Sentinel-3 OLCI can provide data more frequently for
 52 timely monitoring. However, the Sentinel-3 OLC images are at a much coarser spatial resolution of 300 m.
 53 Such a spatial resolution is too coarse to provide sufficient detail for local areas of interest.

54 There is a great need for data that have simultaneously the spatial resolution of Sentinel-2 (10 m) and
 55 temporal resolution of Sentinel-3 (i.e., nearly daily Sentinel-2 time-series) to provide more informative data
 56 and support a wider range of monitoring applications, particularly for areas where the amount of available
 57 effective Sentinel-2 observations is limited due to cloud contamination. The daily Sentinel-2 images have
 58 great value for dynamic monitoring of rapid changes on the Earth's surface at a required fine spatial resolution,
 59 such as timely crop monitoring (Gao et al., 2017). Both Sentinel-2 MSI and Sentinel-3 OLCI data are freely
 60 available to users and have global coverage. Furthermore, the two sensors have the similar wavelengths for
 61 four bands (i.e., blue, green, red and NIR bands), as shown in Table 1. In our previous study Wang et al. (2016),
 62 an accurate method based on area-to-point regression kriging (ATPRK) (Wang et al, 2015) was used to fuse
 63 the 20 m Sentinel-2 8a band with 10 m bands 2, 3, 4 and 8 to produce 10 m Sentinel-2 8a. This provides an
 64 excellent opportunity for spatio-temporal fusion of 10 m Sentinel-2 MSI and 300 m Sentinel-3 OLCI data to
 65 create 10 m, daily Sentinel-2 images. With this process, the number of cloud-free Sentinel-2 images, as well as
 66 the temporal resolution, can be maximized.

67

68 Table 1 The corresponding bands for Sentinel-2 MSI and Sentinel-3 OLCI images

Sentinel-2			Sentinel-3		
Band number	Wavelength (nm)	Spatial resolution (m)	Band number	Wavelength (nm)	Spatial resolution (m)
2 (Blue)	458-523	10	Oa4 (Blue)	437-447	300
3 (Green)	543-578	10	Oa6 (Green)	555-565	300
4 (Red)	650-680	10	Oa8 (Red)	660-670	300
8a (NIR)	855-875	20	Oa17 (NIR)	855-875	300

69

70 Spatio-temporal fusion approaches have been developed for blending fine spatial resolution, but coarse
 71 temporal resolution Landsat and fine temporal resolution, but coarse spatial resolution Moderate Resolution
 72 Imaging Spectroradiometer (MODIS) or MEdium Resolution Imaging Spectrometer (MERIS) images to
 73 create fine spatio-temporal resolution images (Gao et al., 2015; Zhang et al., 2015; Chen et al., 2015). The

74 implementation requires at least one coarse-fine spatial resolution image pair (e.g., MODIS-Landsat image
75 pair acquired on the same day) or one fine spatial resolution image (hereafter called fine image) that is
76 temporally close to the prediction day. In recent years, several spatio-temporal fusion approaches have been
77 developed. The spatial and temporal adaptive reflectance fusion model (STARFM) is one of the earliest and
78 most widely used spatio-temporal fusion approaches (Gao et al., 2006). Appreciating its simple
79 implementation, it has been used to support various applications, such as forest monitoring, crop monitoring
80 (Gao et al., 2015; Gao et al., 2017), leaf area index (LAI) monitoring (Dong et al., 2016; Houborg et al., 2016),
81 land surface temperature (LST) monitoring (Shen et al., 2016) and gross primary productivity (GPP)
82 monitoring (Singh, 2011). STARFM is performed based on the availability of at least one image-pair. It
83 assumes that the temporal changes of all classes within a coarse pixel are uniform, which is more suitable for
84 homogeneous landscape dominated by pure coarse pixels. To enhance the performance of STARFM for
85 heterogeneous landscapes dominated by mixed pixels, an enhanced STARFM (ESTARFM) method was
86 developed (Zhu et al., 2010). Based on the availability of two coarse-fine image pairs, ESTARFM estimates
87 the temporal change rate of each class separately and assumes the change rates to be stable during a period
88 (Emelyanova et al., 2013). STARFM was also extended for timely monitoring of forest disturbance based on a
89 version termed spatial temporal adaptive algorithm for mapping reflectance change (STAARCH) (Hilker et al.,
90 2009). Based on the mechanism of machine learning, some learning-based methods were proposed, including
91 sparse representation (Huang and Song, 2012; Song and Huang, 2013), extreme learning machine (Liu et al.,
92 2016), artificial neural network and support vector regression (Moosavi et al., 2015), and deep learning (Das
93 and Ghosh, 2016). This type of method learns the relationship between the available coarse-fine image pairs,
94 which is used to guide the prediction of fine images from coarse images on other days.

95 Alternatively, spatial unmixing is a type of spatio-temporal fusion approach that can be performed using one
96 fine image. More precisely, it requires a fine spatial resolution thematic map that can be derived by
97 interpretation of the available fine spatial resolution data (Amorós-López et al., 2011,2013; Gevaert et al.,
98 2015; Zurita-Milla et al., 2008) or from other sources including an aerial image (Mustafa et al., 2014), or

99 land-use database (Zurita-Milla et al., 2009). Different from spectral unmixing which estimates for which the
100 class proportions within coarse pixels and where the class endmembers (spectra) are known, spatial unmixing
101 estimates the class endmembers within coarse pixels and the class proportions are known (calculated by
102 upscaling the fine spatial resolution thematic map) (Busetto et al., 2008; Maselli, 2001; Zhukov et al., 1999).
103 Spatial unmixing assumes that no land-cover/land-use changes occur during the period of interest and the class
104 proportions are constant for coarse images at different times. This approach was used to create 30 m
105 Landsat-like time-series from 300 m MERIS images using a 30 m thematic map obtained by classification of
106 an available Landsat image (Zurita-Milla et al., 2008) or a fine spatial resolution land-use database LGN5
107 (Zurita-Milla et al., 2009). Wu et al. (2012) extended spatial unmixing to cases with one coarse-fine image pair
108 available and proposed a surface reflectance calculation model (SRCM). SRCM performs unmixing separately
109 for two coarse images and estimates the temporal changes of each endmember spectra and finally adds the
110 changes to the known fine image. Similarly to the idea of SRCM, Gevaert et al. (2015) performed unmixing
111 directly for the residual image (defined as the difference between two coarse images) to estimate the changes
112 of endmember spectra. Huang and Zhang (2014) developed an unmixing-based spatio-temporal reflectance
113 fusion model (U-STFM) using two coarse-fine image pairs. The spatial unmixing approach can also be
114 combined with STARFM and some hybrid methods were developed (Xu et al., 2015; Xie et al., 2016; Zhu et
115 al., 2016).

116 For spatio-temporal fusion in practice, one challenging problem is that sometimes very few fine images
117 (Sentinel-2 image in this paper) that are temporally close to the prediction time are available for use, due to
118 cloud and shadow contamination. Another problem is that even where one fine image is available, strong
119 temporal changes may have occurred from the time of the fine image to prediction. This means that the
120 observations at two times may be very different and do not have a strong correlation. This is exacerbated for
121 the fusion of 10 m Sentinel-2 MSI and 300 m Sentinel-3 OLCI images, which involves a large zoom factor of
122 30 (double of that from 500 m MODIS to 30 m Landsat spatial resolution) and a number of mixed pixels. In
123 this case, the available fine image on one day may be very different to the ideal prediction on another day.

124 Thus, how to make full use of the available fine image is a critical issue. The U-STFM (Huang and Zhang,
125 2014) and flexible spatiotemporal data fusion (FSDAF) (Zhu et al., 2016) methods were developed to deal
126 with strong temporal changes. However, U-STFM requires at least two coarse-fine image pairs. Although
127 FSDAF requires only one image pair, its performance may sometimes be compromised by the unmixing
128 process where a global, linear unmixing model is considered.

129 In this paper, to cope with the abovementioned problems, we propose a new method for fusion of Sentinel-2
130 MSI and Sentinel-3 OLCI images. The new method consists of three stages, including regression model fitting
131 (RM fitting, hereafter called RM), spatial filtering (SF) and residual compensation (RC). RM aims to increase
132 the correlation between the observations acquired at two times. The SF step removes the blocky artifacts in the
133 RM prediction. The residuals from RM are finally compensated by the RC step to make full use of the coarse
134 spectral information. For simplicity, we abbreviate the new method as the Fit-FC method. This method can be
135 performed with the availability of only one coarse-fine spatial resolution image (i.e., Sentinel-3–Sentinel-2)
136 pair, and it is a promising choice for cases where strong temporal changes occurred from the time of the
137 available image pair to prediction.

138

139

140 2. Methods

141

142 For spatio-temporal fusion of Sentinel-2 MSI and Sentinel-3 OLCI images, only the blue, green, red and
143 NIR bands for Sentinel-2 (bands 2, 3, 4 and 8a) and Sentinel-3 (Oa4, Oa6, Oa8 and Oa17) were considered, as
144 shown in Table 1. The 20 m Sentinel-2 8a band needs to be downscaled to 10 m in advance. This is achieved
145 by fusion of this band with four 10 m Sentinel-2 bands (bands 2, 3, 4 and 8) acquired at the same time.
146 According to our previous work (Wang et al., 2016), the ATPRK method (Wang et al., 2015) can produce
147 more accurate results than eight benchmark methods and, thus, ATPRK was employed.

148 For simplicity of the problem description, suppose there are two 300 m Sentinel-3 images at time t_1 and t_2 ,
 149 and the 10 m Sentinel-2 image at t_1 is available. The ultimate task of spatio-temporal fusion is to predict the 10
 150 m Sentinel-2 image at t_2 based on the three available images. Let $C_1(\mathbf{X}_i, l_b)$ and $C_2(\mathbf{X}_i, l_b)$ be the reflectance
 151 of the Sentinel-3 pixel centered at location \mathbf{X}_i in band l_b ($b=1, 2, 3$, and 4 represents the blue, green, red and
 152 NIR bands, respectively) of the Sentinel-3 OLCI image at t_1 and t_2 , and $F_1(\mathbf{x}_j, l_b)$ be the reflectance of the
 153 Sentinel-2 pixel centered at location \mathbf{x}_j in band l_b of the Sentinel-2 MSI image at t_1 . The objective is to
 154 predict target variables $F_2(\mathbf{x}_j, l_b)$ for all Sentinel-2 pixels in all four bands. The Fit-FC method consisting of
 155 RM, SF and RC is detailed as follows.

156

157 2.1. Regression model fitting (RM)

158

159 In the first stage, RM is used to relate the observations acquired at two times and minimize their differences.
 160 Based on the local spatial variation of land cover, a local linear regression model is considered. Specifically,
 161 for the two Sentinel-3 images, a moving window is used to fit the regression model for each center Sentinel-3
 162 pixel in each band.

163

$$C_2(\mathbf{X}, l_b) = a(\mathbf{X}, l_b)C_1(\mathbf{X}, l_b) + b(\mathbf{X}, l_b) + R(\mathbf{X}, l_b) \quad (1)$$

164

164 where $a(\mathbf{X}, l_b)$ and $b(\mathbf{X}, l_b)$ are two coefficients for the pixel centered at \mathbf{X} in band l_b , and $R(\mathbf{X}, l_b)$ is the 300
 165 m coarse residual. Since $C_1(\mathbf{X}, l_b)$ and $C_2(\mathbf{X}, l_b)$ for all Sentinel-3 pixels in the local window are known, the
 166 two coefficients can be estimated using the least square method.

167

167 The models constructed at 300 m coarse resolution are applied to 10 m fine spatial resolution Sentinel-2
 168 images. For any fine pixel in band l_b , say a fine pixel centered at \mathbf{x}_0 , the RM prediction at t_2 , denoted as

169

$\hat{F}_{\text{RM}}(\mathbf{x}_0, l_b)$, is calculated as

170

$$\hat{F}_{\text{RM}}(\mathbf{x}_0, l_b) = a(\mathbf{X}_0, l_b)F_1(\mathbf{x}_0, l_b) + b(\mathbf{X}_0, l_b) \quad (2)$$

171 where \mathbf{X}_0 is the center of the coarse pixel in which the fine pixel centered at \mathbf{x}_0 falls. The two regression
172 coefficients $a(\mathbf{X}, l_b)$ and $b(\mathbf{X}, l_b)$ estimated from (1) vary spatially at 300 m (i.e., at a spatial resolution of 300
173 m) and the 10 m RM prediction in (2) is a linear transformation of each block of 30 by 30 Sentinel-2 pixels at
174 t_1 . The inconsistency between the spatial resolutions of the regression coefficients and the Sentinel-2 image at
175 t_1 means that the RM prediction is dominated by blocky artifacts with a spatial size of 30 by 30 pixels, which
176 corresponds with a Sentinel-3 pixel. The blocky artifacts are more obvious when the spatial and temporal
177 variation in the two 300 m Sentinel-3 images is large (i.e., spatially and temporally non-stationary), where the
178 two regression coefficients can vary greatly at different locations.

180 2.2. Spatial filtering (SF)

181
182 To deal with the blocky artifacts problem in the RM prediction, SF is considered in the second step. A
183 natural solution is to use the neighboring 10 m pixels in a local window for filtering and to remove the blocky
184 artifacts. However, we cannot use all neighboring pixels in the local window for SF. This is because the
185 neighboring pixels are mostly not for the same class as the center pixel and may have very different
186 reflectances. Using all of them will decrease the difference in spectra between classes, thereby leading to
187 over-smooth predictions. Actually, the blocky artifacts in the RM prediction are produced due to the
188 inconsistency in reflectance for pixels that belong to the same class and should have similar spectra. Thus, for
189 SF, it is a more reasonable choice to select out these neighboring pixels and use them to reduce the difference
190 in spectra for the same class. Since there already exist blocky artifacts in the RM prediction, it is inappropriate
191 to use the contaminated image to search for spectrally similar neighboring pixels. Alternatively, the four bands
192 of the Sentinel-2 image at t_1 are used, based on the assumption of stable land cover *boundaries* (or very few
193 changes in boundaries) occurring from t_1 to t_2 . This assumption allows changes to the class of an entire object,
194 as long as the boundary of the object is retained (i.e., no abrupt changes). This is a strategy used in STARFM

195 (Gao et al., 2006). The spectral difference between a neighboring pixel at \mathbf{x}_i and the center pixel at \mathbf{x}_0 is
 196 calculated as

$$197 \quad D = \sqrt{\sum_{b=1}^4 [F_1(\mathbf{x}_i, l_b) - F_1(\mathbf{x}_0, l_b)]^2} / 4. \quad (3)$$

198 In the local window with w by w fine pixels, the first n pixels with smallest D (including the center pixel itself)
 199 are identified as spectrally similar neighbors and are selected for SF. For each 10 m pixel at t_2 , the SF
 200 prediction is determined as the linear combination of the RM prediction of spectrally similar neighboring
 201 pixels

$$202 \quad \hat{F}_{\text{SF}}(\mathbf{x}_0, l_b) = \sum_{i=1}^n W_i \hat{F}_{\text{RM}}(\mathbf{x}_i, l_b) \quad (4)$$

203 where W_i is a weight determined according to the spatial distance between the neighboring and center pixels.

204 Based on spatial dependence, spatially closer pixels are more likely to have similar reflectance to the center
 205 pixel and, thus, receive larger weights

$$206 \quad W_i = (1/d_i) / \sum_{i=1}^n (1/d_i) \quad (5)$$

$$207 \quad d_i = 1 + \sqrt{|\mathbf{x}_i - \mathbf{x}_0|^2} / (w/2). \quad (6)$$

208 The distance d_i needs to be constrained to an appropriate range to exert reasonable influence on the weight W_i .

209 Thus, 1 and $w/2$ are used in (6) and d_i ranges from 1 to $1 + \sqrt{2}$ correspondingly. With the filtering scheme in
 210 (4), the spectral difference for the same class in the RM prediction can be reduced, which can alleviate the
 211 blocky artifacts.

212

213 2.3. Residual compensation (RC)

214

215 In the first RM step, there exist residuals in the regression models. Without incorporation of the residuals,
216 the uncertainty is propagated to the SF process and leads to spectral distortion. To further enhance the SF
217 prediction, we take full advantage of the spectral information in the 300 m Sentinel-3 image at t_2 in the third
218 step, RC.

219 The 300 m coarse residual $R(\mathbf{X}, l_b)$ needs to be downscaled to 10 m to match the target spatial resolution
220 first. This can be achieved using the simple bicubic interpolation. The produced 10 m residual, denoted as
221 $r(\mathbf{x}, l_b)$, can be added straightforwardly to the 10 m SF predictions. This scheme, however, will produce
222 over-smooth results due to the smoothing effect in bicubic interpolation. The SF process enables the pixels
223 with similar spectra at t_1 in the local window to have similar spectra in the intermediate SF prediction for t_2 .
224 Based on the assumption of stable land cover *boundaries* from t_1 to t_2 , in the final prediction for t_2 , these pixels
225 are expected to have similar spectra. Thus, it is reasonable to assume that these pixels should have similar
226 residuals.

227 The 10 m residuals are used mainly to preserve the 300 m spectral information at t_2 , and they cannot provide
228 much useful spatial detail at 10 m. To avoid the smoothing effect from bicubic interpolation in the final
229 prediction, a possible solution is to average the 10 m residuals for the same class in the local window or within
230 an object. For this process, a 10 m thematic map at t_2 that contains class/object information is required, which
231 can be approximately estimated as a segmentation prediction of the Sentinel-2 image at t_1 . Each class always
232 covers a number of objects. For this solution, however, the number of classes is often related closely to the size
233 of objects and needs to be set appropriately according to the study scene. On the one hand, the number of
234 classes needs to be as large as possible (i.e., the size of objects needs to be as small as possible) to retain more
235 spatial variation. On the other hand, the number of classes needs to be as small as possible (i.e., the size of
236 objects needs to be as large as possible) to ensure the removal of the smoothing effect in the 10 m bicubic
237 interpolation-derived residuals. Moreover, this solution is more suitable for areas where the intra-class
238 variation of spectra is small.

239 Alternatively, a more generic solution is adopted in this paper, which allows a significant simplification of
 240 the RC process. Specifically, similarly to the SF process, the spectrally similar neighboring pixels are used to
 241 update the bicubic interpolation-derived residuals, $r(\mathbf{x}, l_b)$. For a fine pixel centered at \mathbf{x}_0 , the 10 m bicubic
 242 interpolation-derived residuals of spectrally similar neighboring pixels are weighted and combined linearly,
 243 thus, producing the new 10 m residual

$$244 \quad \hat{F}_{\text{RC}}(\mathbf{x}_0, l_b) = \sum_{i=1}^n W_i r(\mathbf{x}_i, l_b). \quad (7)$$

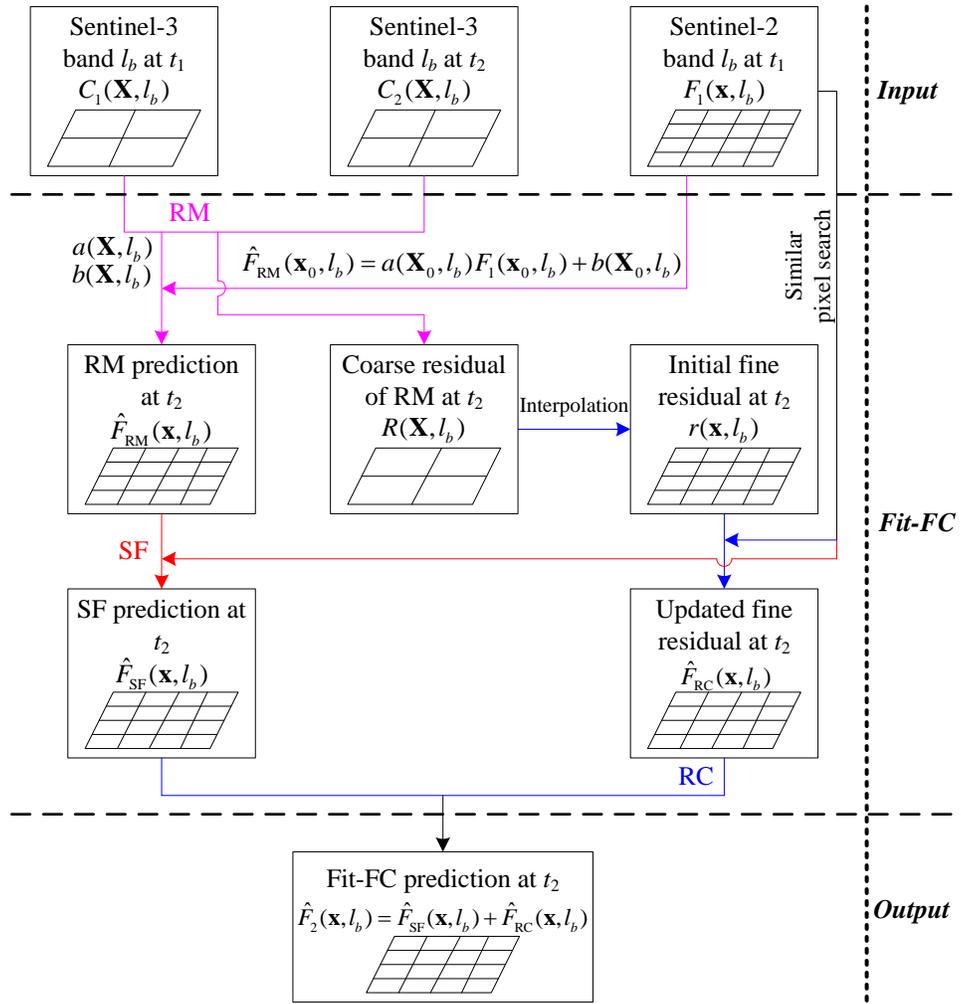
245 In (7), the weight W_i is calculated in the same way as that in (5) and (6). This process enables pixels with
 246 similar spectra at t_1 in the local window to have similar residuals.

247 Finally, the updated 10 m residual is added back to the SF prediction to preserve the spectral information,
 248 and the Fit-FC prediction is produced as follows

$$249 \quad \hat{F}_2(\mathbf{x}_0, l_b) = \hat{F}_{\text{SF}}(\mathbf{x}_0, l_b) + \hat{F}_{\text{RC}}(\mathbf{x}_0, l_b). \quad (8)$$

250 A flowchart describing the Fit-FC method is given in Fig. 1, where the processes of the three steps are
 251 marked in different colors. It is seen that the 10 m valuable information in the Sentinel-2 image at t_1 is used
 252 throughout all three steps, but in different ways. In RM, according to the linear regression model built between
 253 the Sentinel-3 images at different times, the Sentinel-2 image at t_1 is linearly transformed to produce the RM
 254 prediction for the Sentinel-2 image at t_2 . In SF and RC, by assuming stable land cover boundaries (i.e.,
 255 non-shape changes), the Sentinel-2 image at t_1 is used to search for spectrally similar pixels in every local
 256 window (or provide object information for RC, if required).

257



258

259 Fig. 1 Flowchart of the Fit-FC method. The lines in magenta, red and blue represent the processes of RM, SF and RC, respectively.

260

261

262 **3. Experiments**

263

264 *3.1. Data and experimental setup*

265

266 For the Sentinel-2 images, the original Level-1C products were provided in top of atmosphere reflectance
 267 (TOA) units. The TOA was transformed to the bottom of atmosphere reflectance (BOA) using the *Sen2Cor*
 268 algorithm released by ESA (Muller-Wilm, 2016; Vuolo et al., 2016). The 20 m band 8a was fused with 10 m
 269 bands 2, 3, 4 and 8 to produce 10 m Sentinel-2 8a using the ATPRK method (Wang et al., 2016). For the

270 Sentinel-3 images, the atmospheric correction is currently under methodological development. To concentrate
271 solely on the performance of the spatio-temporal fusion methods and avoid uncertainties introduced by
272 radiometric correction and geometric registration, for the present experimental purpose we simulated 300 m
273 Sentinel-3 images by degrading the four-band 10 m Sentinel-2 (in BOA) images with a zoom factor of 30.
274 Such a setup is used widely in studies on spatio-temporal fusion ([Gevaert et al., 2015](#); [Zhang et al., 2015](#); [Zhu
275 et al., 2016](#)).

276 Two sets of images were used to validate the Fit-FC method. The images are located in Coleambally and
277 Currawarna, Australia. Both have a spatial size of 15 km by 15 km, corresponding to 1500 by 1500 Sentinel-2
278 pixels and 50 by 50 Sentinel-3 pixels. For both sites, the images were acquired on 25 December 2015 (t_1) and
279 21 August 2016 (t_2). Fig. 1(a) and Fig. 1(b) show the Sentinel-2 images (in BOA) acquired on the two days for
280 Coleambally, while Fig. 2(a) and Fig. 2(b) show the Sentinel-2 images (in BOA) for Currawarna. The
281 corresponding Sentinel-3 images for the two sites are shown in Figs. 2(c) and 2(d) and Figs. 3(c) and 3(d).
282 Many spatial details are lost in the 300 m Sentinel-3 images and the 10 m Sentinel-2 images can provide much
283 more abundant spatial information at the local scale. As observed from the images at the two times, the study
284 sites experienced strong temporal changes from 25 December 2015 to 21 August 2016. This necessitates the
285 use of spatio-temporal fusion methods to create temporally more frequent Sentinel-2 images at a spatial
286 resolution of 10 m. The task of spatio-temporal fusion for both sites in the experiments is to predict the 10 m
287 Sentinel-2 image on 21 August 2016, using the 300 m Sentinel-3 images on 25 December 2015 to 21 August
288 2016 and the 10 m Sentinel-2 images on 25 December 2015 as inputs. The available real Sentinel-2 image on
289 21 August 2016 was used to evaluate the predictions objectively.

290 Three spatio-temporal fusion methods, STARFM ([Gao et al., 2006](#)), spatial unmixing ([Zurita-Milla et al.,
291 2008](#)), and FSDAF ([Zhu et al., 2016](#)), can be implemented using the available input images (i.e., one
292 coarse-fine image pair and one coarse image at the prediction time). Thus, they were considered as benchmark
293 methods and compared with the Fit-FC method. Through trial-and-error tests, 30 spectrally similar pixels were
294 found in the moving window with 30 by 30 fine pixels (at 10 m resolution) for STARFM and Fit-FC. The

295 number of classes was set to six for SU and FSDAF. For quantitative assessment, three indices were used: the
 296 root mean square error (RMSE), correlation coefficient (CC), and universal image quality index (UIQI) (Wang
 297 and Bovik, 2002). The ideal values of RMSE, CC and UIQI are 0, 1 and 1, respectively. The predictions are
 298 more accurate when the RMSE value is smaller and the CC and UIQI values are larger.

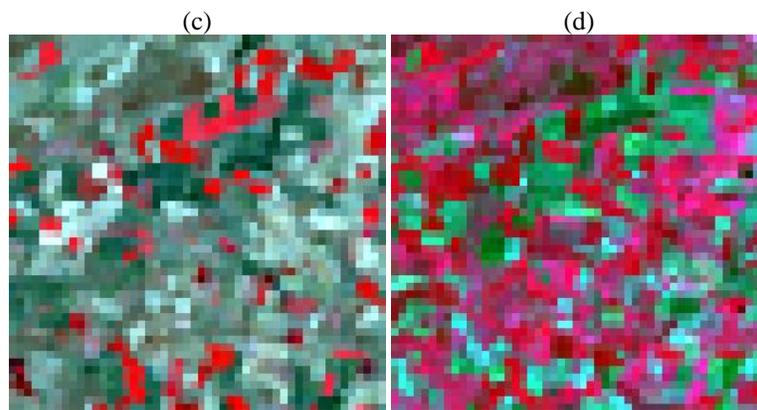
299

300



301

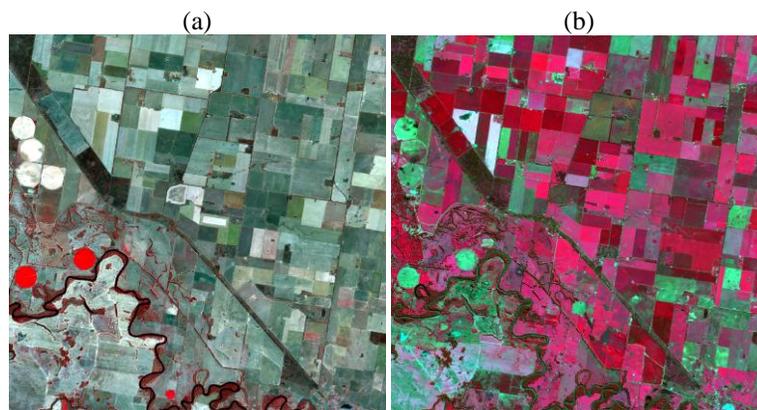
302



303

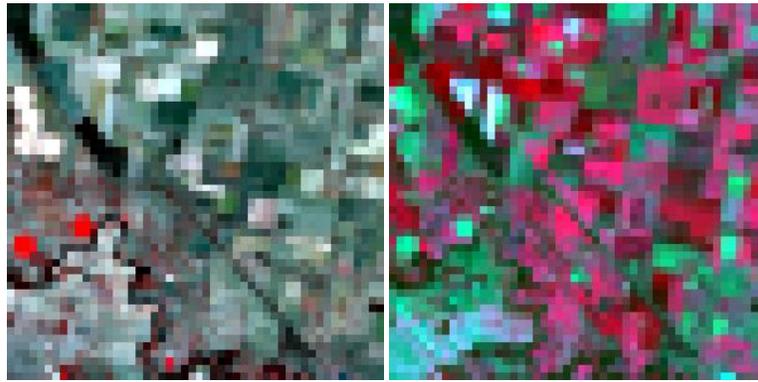
304 Fig. 2 Images for Coleambally (15 km by 15 km). (a) and (b) are Sentinel-2 images (NIR, red, and green bands as RGB) acquired on
 305 25 December 2015 (t_1) and 21 August 2016 (t_2), respectively. (c) and (d) are the corresponding Sentinel-3 images.

306



307

308



309

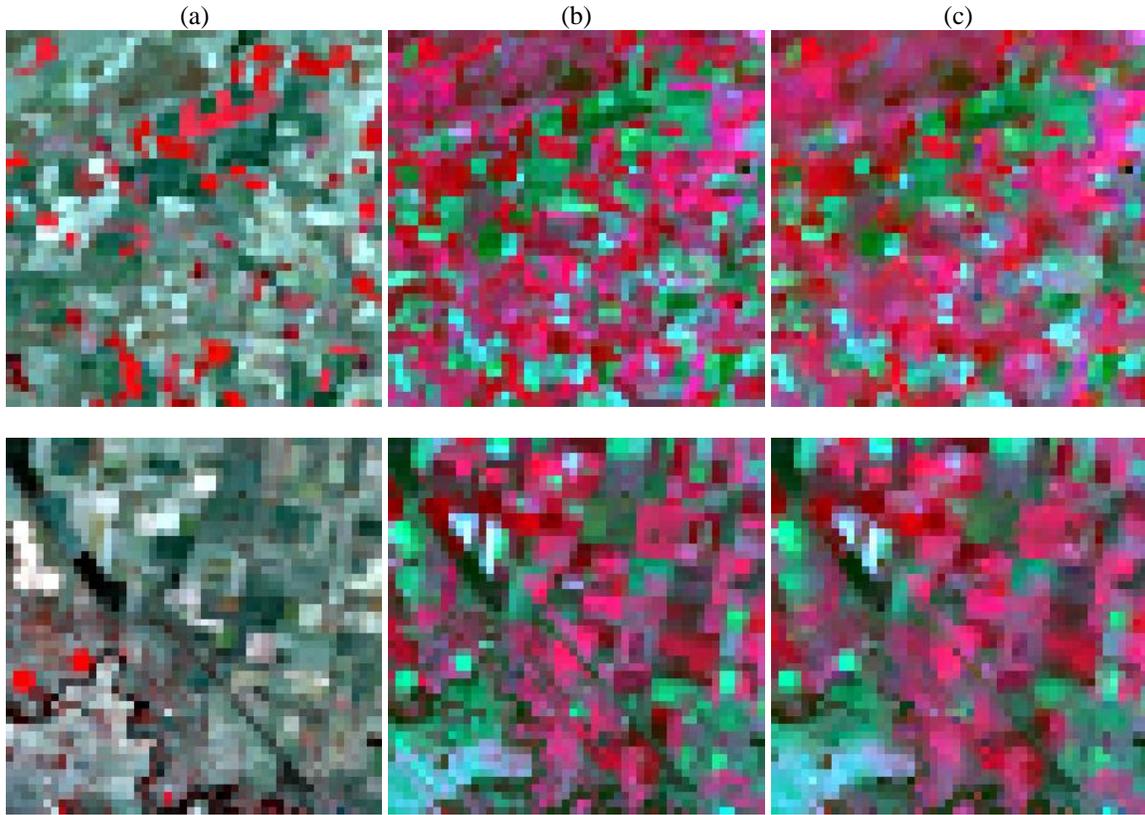
310 Fig. 3 Images for Currawarna (15 km by 15 km). (a) and (b) are Sentinel-2 images (NIR, red, and green bands as RGB) acquired on
 311 25 December 2015 (t_1) and 21 August 2016 (t_2), respectively. (c) and (d) are the corresponding Sentinel-3 images.

312

313 3.2. Implementation

314

315 Fig. 4 shows the interim Sentinel-3 images produced in the RM process for both sites (i.e., a local, linear
 316 transformation of the Sentinel-3 image at t_1 , according to the regression model constructed in (1)). It is seen
 317 that the Sentinel-3 images at t_1 and t_2 are very different. By RM, however, the produced interim Sentinel-3
 318 images are visually highly similar to the image at t_2 . Table 2 is the quantitative evaluation of the original
 319 Sentinel-3 (or Sentinel-2) images at t_1 and the interim Sentinel-3 (or Sentinel-2, i.e., the RM predictions)
 320 images, with the Sentinel-3 (or Sentinel-2) images at t_2 as reference. The Sentinel-3 or Sentinel-2 images at t_1
 321 and t_2 have only a very small correlation with each other. The mean CCs or UIQIs between the Sentinel-3
 322 images for both sites are below 0.10. The linearly transformed Sentinel-3 images have a much larger
 323 correlation with the Sentinel-3 image at t_2 , and the mean CCs or UIQIs for two sites are around 0.90. For
 324 Sentinel-2 images, the mean CCs increase from below 0.15 to around 0.70 after RM, and the corresponding
 325 mean RMSEs decrease from around 0.12 to below 0.04. This means that the regression model built at fine
 326 spatial resolution (i.e., Equation (2)) based on the model at coarse spatial resolution (i.e., Equation (1)) is
 327 satisfactory and the RM process can much better relate the observation at t_1 to that at t_2 . Starting with the
 328 interim Sentinel-3 images and the corresponding 10 m RM predictions (i.e., the interim coarse-fine image pair)
 329 in the RM process, it is expected to have more reliable predictions of Sentinel-2 images at t_2 .



333 Fig. 4 The interim Sentinel-3 images produced from the RM process (NIR, red, and green bands as RGB). (a) and (b) are Sentinel-3
 334 images at t_1 and t_2 . (c) is the interim Sentinel-3 image. Lines 1 and 2 are for Coleambally and Currawarna, respectively.

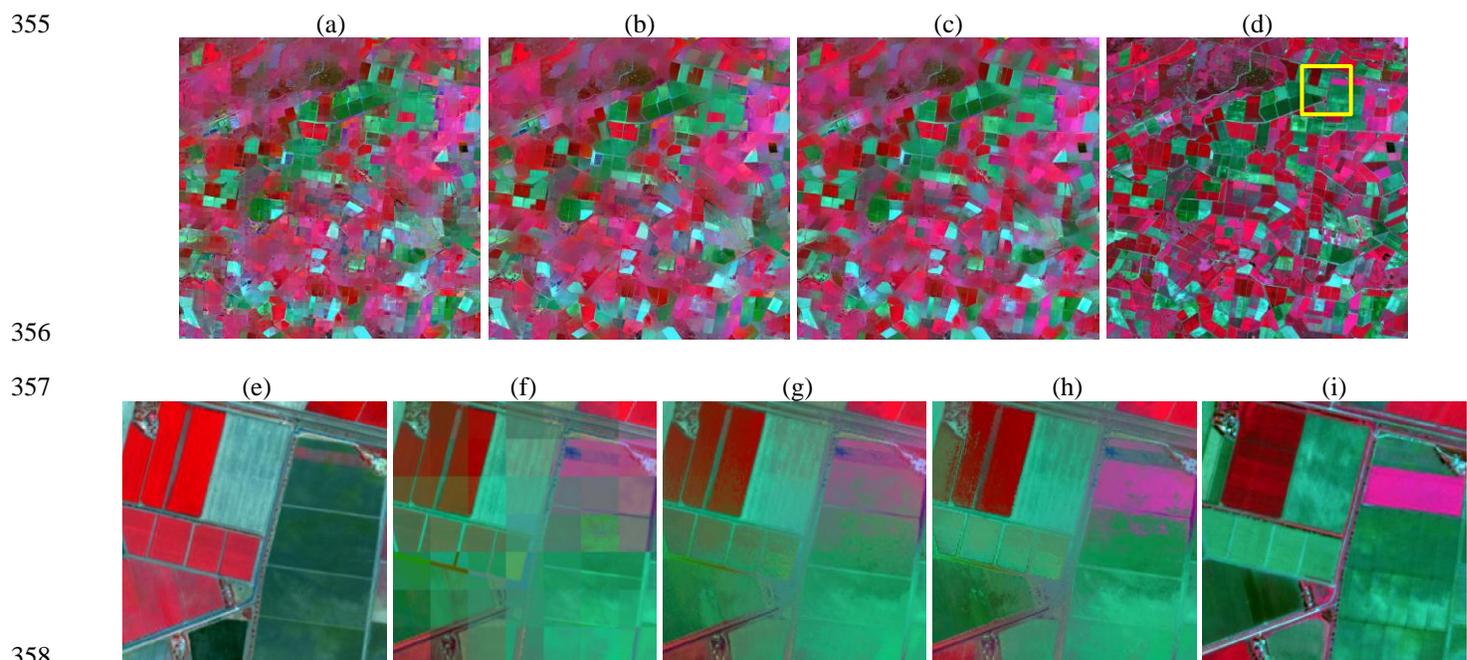
335

336 Table 2 Relation between the original image at t_1 or RM prediction with the image at t_2 (t_2 as reference), where the bold values mean
 337 the most accurate result in each term

		Ideal	Coleambally				Currawarna			
			Sentinel-3		Sentinel-2		Sentinel-3		Sentinel-2	
			t_1	RM	t_1	RM	t_1	RM	t_1	RM
RMSE	Blue	0	0.0581	0.0055	0.0639	0.0124	0.0692	0.0033	0.0736	0.0089
	Green	0	0.0683	0.0059	0.0750	0.0136	0.0843	0.0056	0.0898	0.0141
	Red	0	0.1259	0.0115	0.1352	0.0242	0.1594	0.0070	0.1656	0.0179
	NIR	0	0.1410	0.0511	0.1740	0.1057	0.1125	0.0368	0.1413	0.0952
	Mean	0	0.0983	0.0185	0.1120	0.0390	0.1063	0.0132	0.1176	0.0340
CC	Blue	1	-0.0031	0.8970	0.0380	0.6840	0.1278	0.9192	0.1477	0.6430
	Green	1	0.1747	0.8895	0.1644	0.6571	0.1786	0.9328	0.2208	0.7193
	Red	1	-0.0508	0.8959	-0.0060	0.6980	0.1264	0.9304	0.1377	0.6948
	NIR	1	0.0608	0.8976	0.0596	0.6905	-0.0684	0.9133	0.0453	0.6459
	Mean	1	0.0454	0.8950	0.0640	0.6824	0.0911	0.9239	0.1379	0.6758
UIQI	Blue	1	-0.0014	0.8847	0.0181	0.6597	0.0293	0.9102	0.0350	0.6213
	Green	1	0.0824	0.8754	0.0850	0.6206	0.0863	0.9247	0.1072	0.7058
	Red	1	-0.0205	0.8815	-0.0025	0.6727	0.0303	0.9232	0.0332	0.6803
	NIR	1	0.0554	0.8826	0.0540	0.6724	-0.0566	0.9038	0.0374	0.6336
	Mean	1	0.0290	0.8811	0.0387	0.6563	0.0223	0.9154	0.0532	0.6602

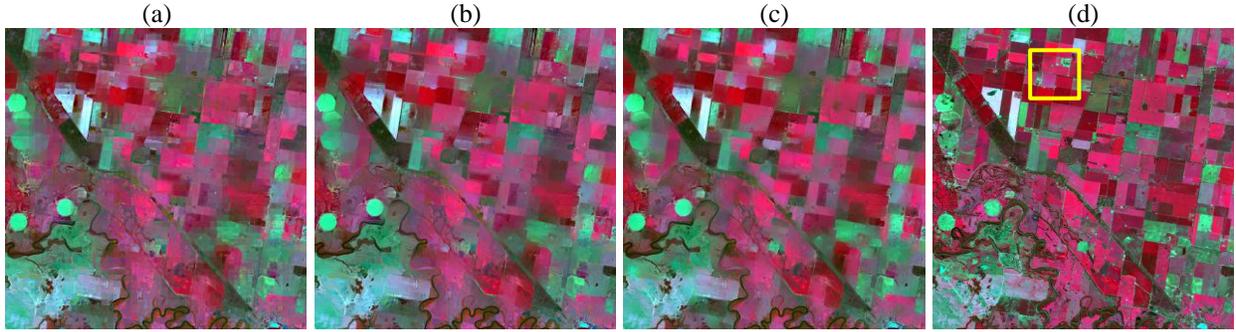
339 Figs. 5 and 6 show the 10 m predictions of the three different stages of Fit-FC for the two sites, where the
 340 results of two sub-areas are zoomed to facilitate visual comparison. Noticeably, the RM predictions contain a
 341 number of blocky artifacts. The further SF process can eliminate the blocky artifacts and produce visually
 342 continuous results. Without the consideration of the residuals in the RM process, however, the SF predictions
 343 suffer from spectral distortion when compared to the reference. With the residuals compensated for the SF
 344 predictions in the third RC step, more spectral information is preserved and the results are closer to the
 345 reference. The benefits of RC can be observed obviously by checking the restoration of the magenta pixels in
 346 Fig. 5(h) and the green pixels in Fig. 6(h).

347 Table 3 gives the accuracy for the predictions at each step and the accuracy gains from RM to SF and SF to
 348 Fit-FC. For all four bands, the RMSEs decrease gradually and the CCs and UIQIs increase gradually from the
 349 first to the third steps, suggesting more accurate results are produced by considering all three steps. For
 350 Coleambally, from RM to SF, the mean RMSE decreases by 0.0018 and the mean CC and UIQI increase by
 351 0.0272 and 0.0162, respectively. From SF to Fit-FC, the mean CC and UIQI increase by 0.0769 and 0.0947,
 352 respectively. For Currawarna, the mean CC and UIQI increase by 0.0272 and 0.0183 from RM to SF and by
 353 0.0576 and 0.0686 from SF to Fit-FC. As for the RMSE, it decreases by 0.0045 from RM to Fit-FC. The results
 354 reveal that all three steps are indispensable in the Fit-FC method.



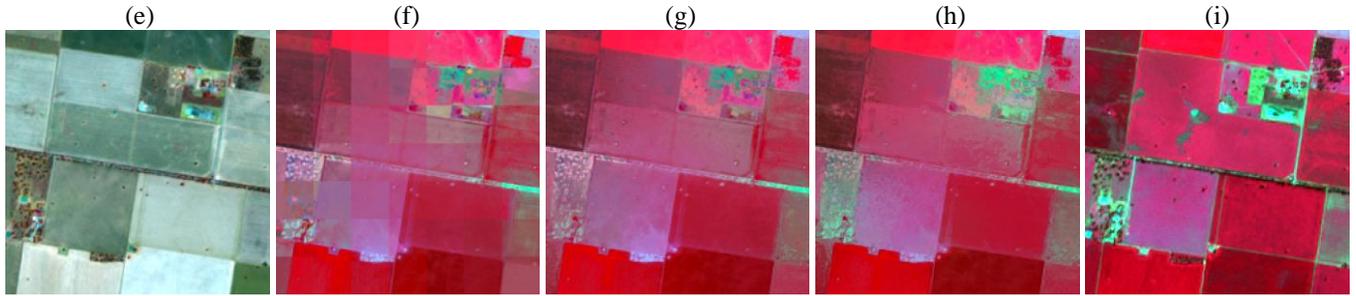
359 Fig. 5 Results of the three different stages of Fit-FC for Coleambally (NIR, red, and green bands as RGB) at t_2 (21 August 2016). (a)
 360 RM. (b) SF. (c) Fit-FC. (d) Reference at t_2 (21 August 2016). (e) The sub-area (marked in yellow in (d)) for the Sentinel-2 image at
 361 t_1 (25 December 2015). (f)-(i) are the corresponding results at t_2 (21 August 2016) for the sub-area in (a)-(d).

362



363

364



365

366 Fig. 6 Results of the three different stages of Fit-FC for Currawarna (NIR, red, and green bands as RGB) at t_2 (21 August 2016). (a)
 367 RM. (b) SF. (c) Fit-FC. (d) Reference at t_2 (21 August 2016). (e) The sub-area (marked in yellow in (d)) for the Sentinel-2 image at
 368 t_1 (25 December 2015). (f)-(i) are the corresponding results at t_2 (21 August 2016) for the sub-area in (a)-(d).

369

370

Table 3 Accuracy for three different stages of Fit-FC (the bold values mean the most accurate result in each term)

		Ideal	Coleambally			Currawarna		
			RM	SF (gain over RM)	Fit-FC (gain over SF)	RM	SF (gain over RM)	Fit-FC (gain over SF)
RMSE	Blue	0	0.0124	0.0119 (0.0005)	0.0105 (0.0014)	0.0089	0.0085 (0.0004)	0.0079 (0.0006)
	Green	0	0.0136	0.0132 (0.0005)	0.0119 (0.0013)	0.0141	0.0134 (0.0007)	0.0122 (0.0012)
	Red	0	0.0242	0.0232 (0.0010)	0.0202 (0.0030)	0.0179	0.0171 (0.0008)	0.0156 (0.0015)
	NIR	0	0.1057	0.1005 (0.0052)	0.0857 (0.0148)	0.0952	0.0902 (0.0050)	0.0822 (0.0080)
	Mean	0	0.0390	0.0372 (0.0018)	0.0321 (0.0051)	0.0340	0.0323 (0.0017)	0.0295 (0.0028)
CC	Blue	1	0.6840	0.7137 (0.0297)	0.7852 (0.0715)	0.6430	0.6718 (0.0288)	0.7310 (0.0592)
	Green	1	0.6571	0.6800 (0.0229)	0.7533 (0.0733)	0.7193	0.7443 (0.0250)	0.7952 (0.0509)
	Red	1	0.6980	0.7243 (0.0263)	0.8007 (0.0764)	0.6948	0.7182 (0.0234)	0.7734 (0.0552)
	NIR	1	0.6905	0.7205 (0.0300)	0.8068 (0.0863)	0.6459	0.6776 (0.0317)	0.7427 (0.0651)
	Mean	1	0.6824	0.7096 (0.0272)	0.7865 (0.0769)	0.6758	0.7030 (0.0272)	0.7606 (0.0576)
UIQI	Blue	1	0.6597	0.6774 (0.0177)	0.7656 (0.0882)	0.6213	0.6379 (0.0166)	0.7101 (0.0722)
	Green	1	0.6206	0.6312 (0.0106)	0.7235 (0.0923)	0.7058	0.7229 (0.0171)	0.7848 (0.0619)
	Red	1	0.6727	0.6888 (0.0161)	0.7839 (0.0951)	0.6803	0.6966 (0.0163)	0.7619 (0.0653)
	NIR	1	0.6724	0.6927 (0.0203)	0.7959 (0.1032)	0.6336	0.6567 (0.0231)	0.7316 (0.0749)
	Mean	1	0.6563	0.6725 (0.0162)	0.7672 (0.0947)	0.6602	0.6785 (0.0183)	0.7471 (0.0686)

371

3.3. Comparison with other methods

The results of the STARFM, SU and FSDAF methods for the two sites are shown in Fig. 7(a)-(c) and Fig. 8(a)-(c). The results of two sub-areas are zoomed for convenience of visual comparison. Due to the strong temporal changes and existence of a large number of mixed pixels in the 300 m Sentinel-3 images (see Figs. 2(c) and 2(d) and Figs. 3(c) and 3(d)), there is spectral distortion in the STARFM predictions where the color appears noticeably different from the reference. For example, for the rectangular object in the top right of the area in Fig. 7(f), they should be in cyan but are inaccurately predicted as mixed red and green. With respect to the SU predictions, they contain blocky artifacts as the unmixing process is implemented in units of 300 m coarse pixels. This method cannot reproduce spatial variation within each object, especially for large-sized objects. FSDAF can reproduce more spatial variation for objects than SU and the spectral information can be more accurately preserved. However, the spatial variation is over-estimated in some cases (see the prediction for the red and green objects in Fig. 7(h)) and the spectral distortion is still obvious for some pixels (see the prediction for the green object in Fig. 8(h)). Compared to the three benchmark methods, the Fit-FC method performs satisfactorily in preserving both spatial and spectral information and the predictions are closer to the reference. This can be demonstrated by the restoration of the cyan and red objects in Fig. 7(i) and green objects in Fig. 8(i).

Quantitative assessment for the four methods is listed in Table 4. As the relation between the observations at two times is weak (see Table 2), for both sites, the STARFM and SU methods produce mean CCs of around 0.55 and 0.45, respectively. FSDAF increases the accuracy obviously and the CCs of the predictions are close to 0.70. For Fit-FC, the mean RMSEs for Coleambally is 0.0321, with a decrease of 0.0116, 0.0254 and 0.0031 when compared to STARFM, SU and FSDAF. For Currawarna, the mean RMSE of Fit-FC is 0.0112, 0.0218 and 0.0015 smaller than that of STARFM, SU and FSDAF, respectively. The mean CC and UIQI of Fit-FC for Coleambally is 0.7865 and 0.7672, with gains of 0.0951 and 0.0822 over that of FSDAF. For Currawarna, the mean CC and UIQI of Fit-FC are 0.0866 and 0.0826 larger than that of FSDAF.

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

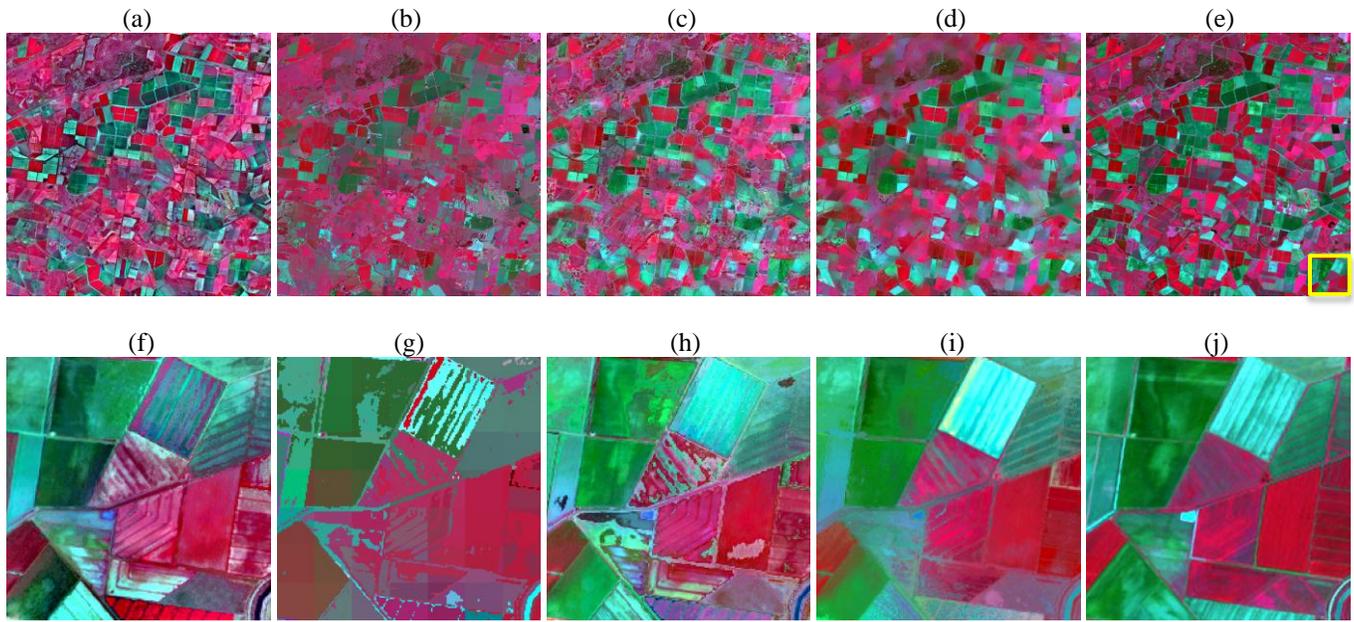


Fig. 7 Results of different spatio-temporal fusion methods for Coleambally (NIR, red, and green bands as RGB). (a) STARFM. (b) Spatial unmixing. (c) FSDAF. (d) The Fit-FC method. (e) Reference. (f)-(j) are the corresponding results for the sub-area (marked in yellow in (e)) in (a)-(e).

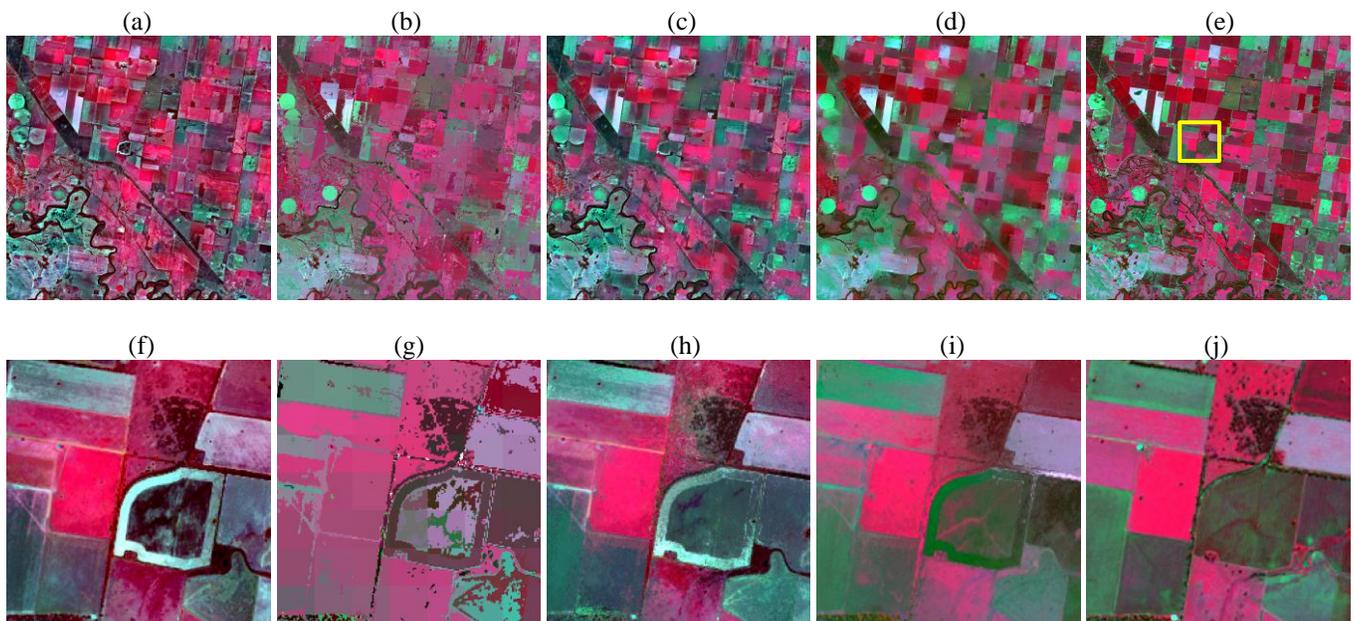


Fig. 8 Results of different spatio-temporal fusion methods for Currawarna (NIR, red, and green bands as RGB). (a) STARFM. (b) Spatial unmixing. (c) FSDAF. (d) The Fit-FC method. (e) Reference. (f)-(j) are the corresponding results for the sub-area (marked in yellow in (e)) in (a)-(e).

414

Table 4 Accuracy for different spatio-temporal fusion methods (the bold values mean the most accurate result in each term)

		Ideal	Coleambally				Currawarna			
			STARFM	SU	FSDAF	Fit-FC	STARFM	SU	FSDAF	Fit-FC
RMSE	Blue	0	0.0217	0.0196	0.0134	0.0105	0.0214	0.0149	0.0118	0.0079
	Green	0	0.0249	0.0212	0.0161	0.0119	0.0258	0.0229	0.0163	0.0122
	Red	0	0.0388	0.0393	0.0249	0.0202	0.0378	0.0344	0.0209	0.0156
	NIR	0	0.0893	0.1500	0.0864	0.0857	0.0779	0.1328	0.0747	0.0822
	Mean	0	0.0437	0.0575	0.0352	0.0321	0.0407	0.0513	0.0310	0.0295
CC	Blue	1	0.4532	0.4356	0.6635	0.7852	0.3834	0.3972	0.5617	0.7310
	Green	1	0.4418	0.4059	0.6035	0.7533	0.5566	0.4949	0.6914	0.7952
	Red	1	0.4925	0.4313	0.6964	0.8007	0.4429	0.3937	0.6543	0.7734
	NIR	1	0.7869	0.5171	0.8023	0.8068	0.7678	0.4775	0.7887	0.7427
	Mean	1	0.5436	0.4475	0.6914	0.7865	0.5377	0.4408	0.6740	0.7606
UIQI	Blue	1	0.4315	0.4309	0.6609	0.7656	0.3068	0.3823	0.5537	0.7101
	Green	1	0.4087	0.4023	0.6035	0.7235	0.5084	0.4842	0.6899	0.7848
	Red	1	0.4806	0.4261	0.6893	0.7839	0.3875	0.3676	0.6528	0.7619
	NIR	1	0.7660	0.5146	0.7865	0.7959	0.7373	0.4741	0.7617	0.7316
	Mean	1	0.5217	0.4435	0.6850	0.7672	0.4850	0.4270	0.6645	0.7471

415

416

417 **4. Discussion**

418

419 In the experiments, each of the used dataset covers a local area of 15 km by 15 km. However, the Fit-FC
420 method can be applied readily to larger areas for larger scale monitoring, as local models are essentially
421 considered in the new method (i.e., a local window is used in all three steps). As an example, the nearly daily
422 Sentinel-2 images have great potential for monitoring phenology. For example, for smallholder croplands, the
423 temporally frequent remote sensing data available to monitor crop growth dynamics and estimate phenology
424 (such as Sentinel-3 or MODIS data) are too coarse to capture the heterogeneity and local variability of
425 production (Duncan et al., 2015). Thus, there are large uncertainties in yield estimation and the identification
426 of yield gaps based on phenological approaches. With a much finer spatial resolution of 10 m, the nearly daily
427 Sentinel-2 images have great potential to enhance yield gap assessments, for example, in smallholder
428 landscapes.

429 Spatio-temporal fusion often requires fine images that are temporally close to the coarse images to be
430 downscaled. Due to data quality issues (e.g., cloud contamination), sometimes very few effective Sentinel-2
431 images are available for fusion. There may exist strong temporal changes from the time of the only effective

432 Sentinel-2 image to prediction, and the only available Sentinel-2 image may have only a small correlation with
433 the ideal prediction of Sentinel-2 image at the prediction time. Fit-FC is proposed to deal with such strong
434 temporal changes in this paper, which can be performed using only one coarse-fine image pair. Different from
435 the classical STARFM method, Fit-FC uses a local regression model to create an interim coarse-fine image
436 pair, which has much greater relation with the observations at the prediction time. This is not the case for
437 STARFM, which starts directly from the time of available Sentinel-2 image and exacerbates this
438 underperformance when the temporal changes are strong. In the two study areas, the relation between the
439 Sentinel-2 images at two times is small (CCs are below 0.15). As a result, the STARFM predictions are not as
440 accurate as those in other literature where the temporal changes are not strong and the relation between the t_1
441 and t_2 images is not weak.

442 In this paper, one fine image is considered for the Fit-FC method. However, multiple fine images can be
443 readily used in Fit-FC, as long as they are available for the study area. Specifically, in the RM step, the
444 multiple fine images can be incorporated by multiple regression. In SF and RC, the multiple fine images can be
445 stacked together for searching spectrally similar pixel (or segmentation, if required in RC). Generally,
446 uncertainties in spatio-temporal fusion tend to decrease when the number of available fine images increases,
447 especially where there exist fine images both before and after the coarse image to be downscaled. On the one
448 hand, the multiple fine images are more realistic for areas that are not easily contaminated by cloud. On the
449 other hand, the multiple Sentinel-2 images can be created from other temporally close data, such as 30 m
450 Landsat images. The Sentinel-2 data and Landsat data have similar wavelengths and the same geographic
451 coordinate system. In our previous work (Wang et al., 2017), 30 m Landsat 8 OLI images were fused with the
452 temporally close 10 m Sentinel-2 MSI images to create 10 m Landsat 8 OLI images. This seems a plausible
453 solution to provide more temporally close Sentinel-2 images for Sentinel-3 images, but more importantly,
454 provide another complementary solution to increase the temporal resolution of the Sentinel-2 images.

455 Although Fit-FC is proposed initially for spatio-temporal fusion of Sentinel-2 and Sentinel-3 images in this
456 paper, it is a generic method suitable for spatio-temporal fusion of other satellite sensor images, such as

457 MODIS/MERIS and Landsat images. For example, for spatio-temporal fusion of the MODIS and Landsat
458 images for Coleambally, the CCs of the STARFM, SU and Fit-FC predictions are 0.4366, 0.4106 and 0.7332,
459 respectively. For Currawarna, the corresponding CCs are 0.4986, 0.3771 and 0.7491, respectively. The
460 advantages of the new method would be more obvious when the temporal changes in the study areas are strong.
461 We also investigated the performance for downscaling the Sentinel-3 image on 14 March 2016, using the
462 Sentinel-3–Sentinel-2 image pair on 25 December 2015. The CC between the Sentinel-2 images on 25
463 December 2015 and 14 March 2016 is 0.2711, obviously larger than the CC between 25 December 2015 and
464 21 August 2016 (see Section 3.2 and Table 2). As a result, the CCs of the STARFM, SU, FSDAF and Fit-FC
465 predictions are 0.8256, 0.6578, 0.8366 and 0.8811, respectively. The accuracy gains of Fit-FC are not as
466 obvious as those for the predictions on 21 August 2016.

467 Apart from reflectance data, Fit-FC is applicable for spatio-temporal fusion issues of other variables, such
468 as Normalized Difference Vegetation Index (NDVI) (Gao et al., 2015; Rao et al., 2015), Normalized
469 Difference Water Index (NDWI) (Jarhani et al., 2014), LST (Weng et al., 2014; Shen et al., 2016),
470 evapotranspiration (Bhattarai et al., 2014; Ke et al., 2017), LAI (Dong et al., 2016; Houborg et al., 2016), and
471 GPP (Singh, 2011). For example, when performing on the NDVI images for the Coleambally dataset, the CC
472 and UIQI of the Fit-FC-based NDVI prediction are 0.8400 and 0.8307, which are 0.0664 and 0.0656 larger
473 than that of the STARFM method. This provides a set of promising avenues for future research.

474 A limitation of Fit-FC is the restoration of abrupt changes (such as the magenta objects in Fig. 5(i)).
475 Actually, abrupt changes have been a long-standing problem in spatio-temporal fusion. The information on
476 abrupt changes is not represented in the available fine image. The problem is greatest where there is a change
477 of land cover boundaries. For example, in the available Sentinel-2 image, a region may be dominated by a
478 large pure patch (e.g., bare soil), but may be broken down into several smaller patches of very different classes
479 (e.g., vegetation, water, and impervious surface) at the prediction time. In this case, the available Sentinel-2
480 image cannot provide much helpful information on the new boundaries. In Fit-FC, it is assumed that there are
481 very few abrupt changes (i.e., stable land cover or object boundaries), an assumption used in existing

482 spatio-temporal fusion methods, including the classical STARFM and SU methods. As mentioned by [Wu et al.](#)
483 [\(2016\)](#), the performance of spatio-temporal fusion can be greatly influenced by the changes of land cover
484 boundaries. A straightforward solution to the issue of abrupt changes is to seek as much auxiliary data as
485 possible to provide related information for the new boundaries at the prediction time. For fusion of Sentinel-2
486 and Sentinel-3 images, such data can be intermediate spatial resolution data such as 30 m Landsat images that
487 are temporally close to the prediction time.

488 It is also worthwhile to develop more alternative schemes for the RM, SF and RC steps in the framework of
489 Fit-FC. The Fit-FC method is substantially different from the regression-based spatio-temporal fusion method
490 developed in [Hazaymeh and Hassan \(2015\)](#), which classifies the entire ratio image between two coarse images
491 into three clusters (i.e., negligible, negative and positive changes) and build a regression model for each cluster.
492 Fit-FC, however, performs RM for each local window (note that the global regression model cannot change
493 the correlation between coarse images), and need not identify the change type. The residuals in the regression
494 models, however, are ignored in [Hazaymeh and Hassan \(2015\)](#). There are also relevant studies for considering
495 residuals, such as those from the unmixing process ([Zhu et al., 2016](#)). It would also be interesting to consider
496 sparse representation for RC ([Huang and Song, 2012](#); [Song and Huang, 2013](#)). All these topics are worthy of
497 investigation for the framework of Fit-FC in future research.

498 In the experiments, we tested the performance of spatio-temporal fusion by creating a Sentinel-2 image for
499 a single day. This paper can be treated as a crucial step towards creating a Sentinel-2 time-series. In future
500 research, the Fit-FC method will be employed to create a daily Sentinel-2 time-series to support practical
501 applications. In applications where Sentinel-2 time-series data are required (e.g., for long-time monitoring of
502 dynamic environmental, agricultural or ecological phenomena), an issue worthy of consideration would be the
503 selection of the fine Sentinel-2 images. A natural solution is to consider the Sentinel-2 image that is temporally
504 close to the Sentinel-3 image to be downscaled or if applicable, Sentinel-2 images that are available both
505 before and after the Sentinel-3 image. It would also be worthy to consider alternative schemes based on a
506 cascade approach ([Demir et al., 2013](#)). For example, in a forecasting sense, the already-created Sentinel-2

507 images (from earlier Sentinel-3 images in the time-series) along with the available Sentinel-2 images can be
508 accumulated to provide the fine image set for downscaling future Sentinel-3 images. This would also work in
509 the same way for hind-casting where accumulated Sentinel-2 images can be used to downscale older
510 Sentinel-3 images. The approach would be particularly suitable for land cover experiencing unidirectional (in
511 time) changes, such as the expansion of urban areas in urbanization and reduction of forest coverage in
512 deforestation. However, it should be stressed that there are uncertainties in the created Sentinel-2 images,
513 which will be propagated and enlarged along the accumulation directions (e.g., past to future in forecasting). It
514 is not clear how much the accumulated uncertainties will affect the final predictions. Moreover, in creating a
515 Sentinel-2 time-series (especially for a very long time), some challenges will be encountered in handling “big
516 data”, such as computational costs and memory issues. All these issues motivate future research.

517

518

519 **5. Conclusion**

520

521 The newly launched Sentinel-2 and Sentinel-3 satellites provide excellent new opportunities for global
522 monitoring. This paper presents a new method, called Fit-FC, for spatio-temporal fusion of Sentinel-2 and
523 Sentinel-3 images to create nearly daily Sentinel-2 images. It includes regression model fitting (RM fitting),
524 spatial filtering (SF) and residual compensation (RC), and all three steps are indispensable. The RM fitting
525 step is used to relate the observations acquired at two times and minimize their differences. SF aims to remove
526 the blocky artifacts in the RM fitting prediction, while RC compensates the residuals from RM fitting to
527 preserve the spectral information. The Fit-FC method can be performed with a minimum number of input
528 images. It was tested using two datasets and compared to the existing STARFM, SU and FSDAF methods. The
529 results show that Fit-FC can produce more accurate results. Fit-FC is particularly relevant for cases involving
530 strong temporal changes, where the correlation between the coarse images is small.

531

532 **References**

533

534 Amorós-López, J., Gómez-Chova, L., Alonso, L., Guanter, L., Moreno, J., Camps-Valls, G., 2011. Regularized multiresolution
535 spatial unmixing for ENVISAT/MERIS and Landsat/TM image fusion. *IEEE Geoscience and Remote Sensing Letters* 8(5),
536 844–848.

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

540 Berger, M., Aschbacher, J., 2012. Preface: The Sentinel missions—new opportunities for science. *Remote Sensing of Environment*
541 120, 1–2.

542 Bhattarai, N., Quackenbush, L. J., Dougherty, M., Marzen, L. J., 2014. A simple Landsat–MODIS fusion approach for monitoring
543 seasonal evapotranspiration at 30 m spatial resolution. *International Journal of Remote Sensing* 36(1), 115–143.

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

546 Das, M., Ghosh, S. K., 2016. Deep-STEP: A deep learning approach for spatiotemporal prediction of remote sensing data. *IEEE*
547 *Geoscience and Remote Sensing Letters* 13, 1984–1988.

548 Demir, B., Bovolo, F., Bruzzone, L., 2013. Classification of time series of multispectral images with limited training data. *IEEE*
549 *Transactions on Image Processing* 22(8), 219–3233.

550 Dong, T., Liu, J., Qian, B., Zhao, T., Jing, Q., Geng, X., Wang, J., Huffman, T., Shang, J., 2016. Estimating winter wheat biomass by
551 assimilating leaf area index derived from fusion of Landsat-8 and MODIS data. *International Journal of Applied Earth*
552 *Observation and Geoinformation* 49, 63–74.

553 Donlon, C. et al., 2012. The global monitoring for environment and security (GMES) Sentinel-3 mission. *Remote Sensing of*
554 *Environment* 120, 37–57.

555 Drusch, M. et al., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of*
556 *Environment* 120, 25–36.

557 Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., Li, X., 2016. Water bodies' mapping from Sentinel-2 imagery with modified
558 normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR Band. *Remote Sensing* 8, 354.

559 Duncan J. M. A., Dash J., Atkinson P. M., 2015. The potential of satellite-observed crop phenology to enhance yield gap
560 assessments in smallholder landscapes. *Frontiers in Environmental Science* 3, 56.

- 561 Emelyanova, I. V., McVicar, T. R., Van Niel, T. G., Li, L. T., van Dijk, A. I. J. M., 2013. Assessing the accuracy of blending
562 Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for
563 algorithm selection. *Remote Sensing of Environment* 133,193–209.
- 564 Fernández-Manso, A., Fernández-Manso, O., Quintano, C., 2016. SENTINEL-2A red-edge spectral indices suitability for
565 discriminating burn severity. *International Journal of Applied Earth Observation and Geoinformation* 50, 170–175.
- 566 Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: Predicting daily
567 Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing* 44(8), 2207–2218.
- 568 Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J. G., Wang, P., Yang, Y., 2015. Fusing Landsat and MODIS data for vegetation
569 monitoring. *IEEE Geoscience and Remote Sensing Magazine* 3, 47–60.
- 570 Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., Mueller, R., Johnson, D. M., Prueger, J. H., 2017.
571 Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment*
572 188, 9–25.
- 573 Gevaert, C. M., García-Haro, F. J., 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data
574 fusion. *Remote Sensing of Environment* 156, 34–44.
- 575 Chen, B., Huang, B., Xu, B., 2015. Comparison of spatiotemporal fusion models: A review. *Remote Sensing*, 1798–1835.
- 576 Hagolle, O. et al., 2015. SPOT-4 (Take 5): Simulation of Sentinel-2 time series on 45 large sites. *Remote Sensing* 7, 12242–12264.
- 577 Hazaymeh, K., Hassan, Q. K., 2015. Spatiotemporal image-fusion model for enhancing the temporal resolution of Landsat-8 surface
578 reflectance images using MODIS images. *Journal of Applied Remote Sensing* 9, 096095-1–096095-14.
- 579 Hilker, T., Wulder, M. A., Coops, N. C., Linke, J., McDermid, J., Masek, J. G., Gao, F., White, J. C., 2009. A new data fusion model
580 for high spatial- and temporal-resolution mapping of forest based on Landsat and MODIS. *Remote Sensing of Environment* 113,
581 1613–1627.
- 582 Houborg, R., McCabe, M. F., Gao, F., 2016. A spatio-temporal enhancement method for medium resolution LAI (STEM-LAI).
583 *International Journal of Applied Earth Observation and Geoinformation* 47, 15–29.
- 584 Huang, B., Song, H., 2012. Spatiotemporal reflectance fusion via sparse representation. *IEEE Transactions on Geoscience and*
585 *Remote Sensing* 50, 3707–3716.
- 586 Huang, B., Zhang, H., 2014. Spatio-temporal reflectance fusion via unmixing: Accounting for both phenological and land cover
587 changes. *International Journal of Remote Sensing* 35, 6213–6233.
- 588 Immitzer, M., Vuolo, F., Atzberger, C., 2016. First experience with Sentinel-2 data for crop and tree species classifications in central
589 Europe. *Remote Sensing* 8, 166.

- 590 Jarihani, A. A., McVicar, T. R., Van Niel, T. G., Emelyanova, I. V., Callow, J. N., Johansen, K., 2014. Blending Landsat and
591 MODIS data to generate multispectral indices: A comparison of “Index-then-Blend” and “Blend-then-Index” approaches.
592 *Remote Sensing* 6, 9213–9238.
- 593 Ke, Y., Im, J., Park, S., Gong, H., 2017. Spatiotemporal downscaling approaches for monitoring 8-day 30 m actual
594 evapotranspiration. *ISPRS Journal of Photogrammetry and Remote Sensing* 126, 79–93.
- 595 Liu, X., Deng, C., Wang, S., Huang, G., Zhao, B., Lauren, P., 2016. Fast and accurate spatiotemporal fusion based upon extreme
596 learning machine. *IEEE Geoscience and Remote Sensing Letters* 13, 2039–2043.
- 597 Maselli, F., 2001. Definition of spatially variable spectral endmembers by locally calibrated multivariate regression analyses.
598 *Remote Sensing of Environment* 75, 29–38
- 599 Moosavi, V., Talebi, A., Mokhtari, M. H., Shamsi, S. R. F., Niazi, Y., 2015. A wavelet-artificial intelligence fusion approach
600 (WAIFA) for blending Landsat and MODIS surface temperature. *Remote Sensing of Environment* 169, 243–254.
- 601 Muller-Wilm, U., 2016. Sentinel-2 MSI—Level-2A Prototype Processor Installation and User Manual. Available online at
602 <http://step.esa.int/thirdparties/sen2cor/2.2.1/S2PAD-VEGA-SUM-0001-2.2.pdf>.
- 603 Mustafa, Y. T., Tolpekin, V. A., Stein, A., 2014. Improvement of spatio-temporal growth estimates in heterogeneous forests using
604 Gaussian bayesian networks. *IEEE Transactions on Geoscience and Remote Sensing* 52(8), 4980–4991.
- 605 Novelli, A., Aguilar, M. A., Nemmaoui, A., Aguilar, F. J., Tarantino, E., 2016. Performance evaluation of object based greenhouse
606 detection from Sentinel-2 MSI and Landsat 8 OLI data: A case study from Almería (Spain). *International Journal of Applied*
607 *Earth Observation and Geoinformation* 52, 403–411.
- 608 Rao, Y., Zhu, X., Chen, J., Wang, J., 2015. An improved method for producing high spatial-resolution NDVI time series datasets
609 with multi-temporal MODIS NDVI data and Landsat TM/ETM+ images. *Remote Sensing* 7, 7865–7891.
- 610 Segl, K., Guanter, L., Gascon, F., Kuester, T., Rogass, C., Mielke, C., 2015. S2eteS: An end-to-end modeling tool for the simulation
611 of Sentinel-2 image products. *IEEE Transactions on Geoscience and Remote Sensing* 53, 5560–5571.
- 612 Shen, H., Huang, L., Zhang, L., Wu, P., Zeng, C., 2016. Long-term and fine-scale satellite monitoring of the urban heat island effect
613 by the fusion of multi-temporal and multi-sensor remote sensed data: A 26-year case study of the city of Wuhan in China.
614 *Remote Sensing of Environment* 172, 109–125.
- 615 Singh, D., 2011. Generation and evaluation of gross primary productivity using Landsat data through blending with MODIS data.
616 *International Journal of Applied Earth Observation and Geoinformation* 13, 59–69.
- 617 Song, H., Huang, B., 2013. Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Transactions on*
618 *Geoscience and Remote Sensing* 51, 1883–1896.

- 619 Storey, J., Roy, D. P., Masek, J., Gascon, F., Dwyer, J., Choate, M., 2016. A note on the temporary misregistration of Landsat-8
620 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sensing of Environment* 186,
621 121–122.
- 622 Van der Werff, H., Van der Meer, F., 2016. Sentinel-2A MSI and Landsat 8 OLI provide data continuity for geological remote
623 sensing. *Remote Sensing* 8, 883.
- 624 Verhoef, W., Bach, H., 2012. Simulation of Sentinel-3 images by four-stream surface–atmosphere radiative transfer modeling in the
625 optical and thermal domains. *Remote Sensing of Environment* 120, 197–207.
- 626 Vuolo, F., Zóltak, M., Pipitone, C., Zappa, L., Wenng, H., Immitzer, M., Weiss, M., Baret, F., Atzberger, C., 2016. Data service
627 platform for Sentinel-2 surface reflectance and value-added products: system use and examples. *Remote Sensing* 8, 938.
- 628 Wang, Q., Shi, W., Atkinson, P. M., Zhao, Y., 2015. Downscaling MODIS images with area-to-point regression kriging. *Remote*
629 *Sensing of Environment* 166, 191–204.
- 630 Wang, Q., Shi, W., Li, Z., Atkinson, P. M., 2016. Fusion of Sentinel-2 images. *Remote Sensing of Environment* 187, 241–252.
- 631 Wang, Q., Blackburn, A., Onojeghu, A., Dash, J., Zhou, L., Zhang, Y., Atkinson, P. M., 2017. Fusion of Landsat 8 OLI and
632 Sentinel-2 MSI data. *IEEE Transactions on Geoscience and Remote Sensing* 55, 3885–3899.
- 633 Weng, Q., Fu, P., Gao, F., 2014. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data.
634 *Remote Sensing of Environment* 145, 55–67.
- 635 Wu, M., Niu, Z., Wang, C., Wu, C., Wang, L., 2012. Use of MODIS and Landsat time series data to generate high-resolution
636 temporal synthetic Landsat data using a spatial and temporal reflectance fusion model. *Journal of Applied Remote Sensing* 6,
637 063507-1–063507-13.
- 638 Wu, M., Wu, C., Huang, W., Niu, Z., Wang, C., Li, W., Hao, P., 2016. An improved high spatial and temporal data fusion approach
639 for combining Landsat and MODIS data to generate daily synthetic Landsat imagery. *Information Fusion* 31, 14-25.
- 640 Xie, D., Zhang, J., Zhu, X., Pan, Y., Liu, H., Yuan, Z., Yun, Y., 2016. An improved STARFM with help of an unmixing-based
641 method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions. *Sensors* 16.
- 642 Xu, Y., Huang, B., Xu, Y., Cao, K., Guo, C., Meng, D., 2015. Spatial and temporal image fusion via regularized spatial unmixing.
643 *IEEE Geoscience and Remote Sensing Letters* 12(6), 1362–1366.
- 644 Zhang, H. K., Huang, B., Zhang, M., Cao, K., Yu, L., 2015. A generalization of spatial and temporal fusion methods for remotely
645 sensed surface parameters. *International Journal of Remote Sensing* 36(17), 4411–4445.
- 646 Zhu, X., Chen, J., Gao, F., Chen, X., Masek, J. G., 2010. An enhanced spatial and temporal adaptive reflectance fusion model for
647 complex heterogeneous regions. *Remote Sensing of Environment* 114, 2610–2623.

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

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

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

654 Zurita-Milla, R., Kaiser, G., Clevers, J. G. P. W., Schneider, W., Schaepman, M. E., 2009. Downscaling time series of MERIS full
655 resolution data to monitor vegetation seasonal dynamics. *Remote Sensing of Environment* 113, 1874–1885.

656

657 LIST OF FIGURE CAPTIONS

658 Fig. 1 Flowchart of the Fit-FC method. The lines in magenta, red and blue represent the processes of RM, SF and RC, respectively.

659 Fig. 2 Images for Coleambally (15 km by 15 km). (a) and (b) are Sentinel-2 images (NIR, red, and green bands as RGB) acquired on
660 25 December 2015 (t_1) and 21 August 2016 (t_2), respectively. (c) and (d) are the corresponding Sentinel-3 images.

661 Fig. 3 Images for Currawarna (15 km by 15 km). (a) and (b) are Sentinel-2 images (NIR, red, and green bands as RGB) acquired on
662 25 December 2015 (t_1) and 21 August 2016 (t_2), respectively. (c) and (d) are the corresponding Sentinel-3 images.

663 Fig. 4 The interim Sentinel-3 images produced from the RM process (NIR, red, and green bands as RGB). (a) and (b) are Sentinel-3
664 images at t_1 and t_2 . (c) is the interim Sentinel-3 image. Lines 1 and 2 are for Coleambally and Currawarna, respectively.

665 Fig. 5 Results of the three different stages of Fit-FC for Coleambally (NIR, red, and green bands as RGB) at t_2 (21 August 2016). (a)
666 RM. (b) SF. (c) Fit-FC. (d) Reference at t_2 (21 August 2016). (e) The sub-area (marked in yellow in (d)) for the Sentinel-2 image at
667 t_1 (25 December 2015). (f)-(i) are the corresponding results at t_2 (21 August 2016) for the sub-area in (a)-(d).

668 Fig. 6 Results of the three different stages of Fit-FC for Currawarna (NIR, red, and green bands as RGB) at t_2 (21 August 2016). (a)
669 RM. (b) SF. (c) Fit-FC. (d) Reference at t_2 (21 August 2016). (e) The sub-area (marked in yellow in (d)) for the Sentinel-2 image at
670 t_1 (25 December 2015). (f)-(i) are the corresponding results at t_2 (21 August 2016) for the sub-area in (a)-(d).

671 Fig. 7 Results of different spatio-temporal fusion methods for Coleambally (NIR, red, and green bands as RGB). (a) STARFM. (b)
672 Spatial unmixing. (c) FSDAF. (d) The Fit-FC method. (e) Reference. (f)-(j) are the corresponding results for the sub-area (marked in
673 yellow in (e)) in (a)-(e).

674 Fig. 8 Results of different spatio-temporal fusion methods for Currawarna (NIR, red, and green bands as RGB). (a) STARFM. (b)
675 Spatial unmixing. (c) FSDAF. (d) The Fit-FC method. (e) Reference. (f)-(j) are the corresponding results for the sub-area (marked in
676 yellow in (e)) in (a)-(e).