https://doi.org/10.1016/j.rse.2016.10.030

Fusion of Sentinel-2 images

Qunming Wang [a], Wenzhong Shi [b], Zhongbin Li [b], and Peter M. Atkinson [c,d,e] *

[a] Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK

[b] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

[c] Faculty of Science and Technology, Engineering Building, Lancaster University, Lancaster LA1 4YR, UK

[d] School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, BT7 1NN, Northern Ireland, UK

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

*Corresponding author

E-mail: P.M.Atkinson@lancaster.ac.uk

Abstract: Sentinel-2 is a very new programme of the European Space Agency (ESA) that is designed for fine spatial resolution global monitoring. Sentinel-2 images provide four 10 m bands and six 20 m bands. To provide more explicit spatial information, this paper aims to downscale the six 20 m bands to 10 m spatial resolution using the four directly observed 10 m bands. The outcome of this fusion task is the production of 10 Sentinel-2 bands with 10 m spatial resolution. This new fusion problem involves four fine spatial resolution bands, which is different to, and more complex than, the common pan-sharpening fusion problem which involves only one fine band. To address this, we extend the existing two main families of image fusion approaches (i.e., component substitution, CS, and multiresolution analysis, MRA) with two different schemes, a band synthesis scheme and a band selection scheme. Moreover, the recently developed area-to-point regression kriging (ATPRK) approach was also developed and applied for the Sentinel-2 fusion task. Using two Sentinel-2 datasets released online, the three types of approaches (eight CS and MRA-based approaches, and ATPRK) were compared comprehensively in terms of their accuracies to provide recommendations for

the task of fusion of Sentinel-2 images. The downscaled ten-band 10 m Sentinel-2 datasets represent important and promising products for a wide range of applications in remote sensing. They also have potential for blending with the upcoming Sentinel-3 data for fine spatio-temporal resolution monitoring at the global scale.

Keywords: Sentinel-2, image fusion, downscaling, area-to-point regression kriging (ATPRK)

1. Introduction

Sentinel-2 is a wide-swath and fine spatial resolution satellite imaging mission of the European Space Agency (ESA) developed in the framework of the European Union Copernicus programme (Drusch et al., 2012; Hagolle et al., 2015; Segl et al., 2015). According to the primary objectives of the programme, the Sentinel-2 mission is designed for data continuity and enhancement of the Landsat and SPOT missions. The Sentinel-2 satellite covers areas from −56° to 84° latitude, and the data are mainly intended to support global land services, including the monitoring of vegetation, soil and water cover, inland waterways and coastal areas. The Sentinel-2A satellite was launched on 23 June 2015 and is now in operation routinely. The addition of the complementary Sentinel-2B satellite will be launched in mid-2016. The twin satellites will be in the same orbit and 180° apart from each other, thereby increasing the frequency of coverage.

Sentinel-2 images cover 13 spectral bands in the visible, near infrared (NIR) and short wave infrared (SWIR) wavelengths, with four bands at 10 m, six bands at 20 m and three bands at 60 m spatial resolution. Table 1 lists the characteristics of the 13 bands. The sensor covers a field of view of 290 km, a swath much wider than the Landsat sensor (185 km) that has been applied widely for global monitoring over the past decades. Hence, the sensor revisits the same area more frequently (every ten days) with a constant viewing angle. The temporal resolution will be further increased to five days with Sentinel-2B. The fine spatial resolution, global coverage and (relatively) fine temporal resolution make the Sentinel-2 data of great utility for a wide range of applications based on remote sensing.

Table 1 Characteristics of the 13 bands of Sentinel-2 data

Band number	1	2	3	4	5	6	7	8	8a	9	10	11	12
Center (nm)	443	490	560	665	705	740	783	842	865	940	1375	1610	2190
Width (nm)	20	65	35	30	15	15	20	115	20	20	30	90	180
Spatial resolution (m)	60	10	10	10	20	20	20	10	20	60	60	20	20

For Sentinel-2 data, four 10 m and six 20 m bands can be used for land-cover/land-use (LCLU) mapping and change detection. With respect to the three 60 m bands, they are mainly dedicated for atmospheric correction and cloud screening (i.e., 443 nm blue band for aerosol retrieval and cloud detection, 940 nm NIR band for water vapor correction, and 1375 nm SWIR band for cirrus detection) (Drusch et al., 2012; Hagolle et al., 2015). The existence of 10 m bands covering the same scene offers excellent opportunities for downscaling the 20 m bands to 10 m spatial resolution to provide more detailed spatial information. This downscaling issue is termed image fusion in remote sensing. In this paper, for the first time, we fuse the four 10 m and six 20 m bands to produce a complete set of ten-band Sentinel-2 data at 10 m spatial resolution and identify the most accurate image fusion method for this task from the existing approaches.

There are two main families of image fusion approaches, that is, component substitution (CS) and multiresolution analysis (MRA) (Vivone et al., 2015). The CS approach transforms the original multispectral image into a new domain and substitutes one of the components with the fine spatial resolution band (hereafter, fine band), such as the panchromatic (PAN) band in PAN-sharpening. Common CS examples include principal component analysis (PCA) (Shettigara, 1992), intensity-hue-saturation (IHS) (Tu et al., 2001), Brovey transformation (BT) (Gillespie et al., 1987), Gram-Schmidt (GS) transformation (Laben & Brower, 2000), adaptive GS (GSA) (Aiazzi et al., 2007), and partial replacement adaptive component substitution (PRACS) (Choi et al., 2011). In the MRA approach, spatial detail is injected by multiresolution decomposition of the fine band. Algorithms falling into this type are high-pass filtering (HPF) (Chavez et al., 1991), smoothing filter-based intensity modulation (SFIM) (Liu, 2000), decimated wavelet transform using an additive injection model (Indusion) (Khan et al., 2008), *a trous* wavelet transform (ATWT) (Vivone et al., 2014), ATWT using the Model 2 (ATWT-M2) (Ranchin & Wald, 2000) and Model 3 (ATWT-M3) (Ranchin & Wald, 2000), and the generalized Laplacian pyramid with modulation transfer function-matched filter (MTF-GLP) (Aiazzi et al., 2006; Vivone et al., 2015).

Geostatistical approaches based on kriging have also been explored for image fusion, including kriging with external drift (KED) (Sales et al., 2013), downscaling cokriging (DSCK) (Atkinson et al., 2008; Pardo-Iguzquiza et al., 2006, 2011) and the recently developed area-to-point regression kriging (ATPRK) approach (Wang et al., 2015a,b). They have the significant advantage of preserving the spectral properties of the observed coarse images, that is, they are coherent. The geostatistical solutions treat the coarse bands as primary variables and the fine bands as auxiliary variables. DSCK requires cross-semivariogram modeling which involves complex deconvolution and convolution calculation processes. KED simplifies the semivariogram modeling process and makes downscaling easier to automate. However, the size of kriging matrices in KED is larger than that in ATPRK, and KED requires calculation of the kriging weights locally for each fine pixel. By contrast, ATPRK separates "trend" estimation (i.e., the regression part) from residual downscaling, and the kriging weights are calculated only once. ATPRK is, thus, computationally more efficient than KED and more user-friendly than DSCK (Wang et al., 2015a,b). As the advantages of ATPRK over KED and DSCK have been clearly presented both theoretically and experimentally in our previous works (Wang et al., 2015a,b), KED and DSCK were not considered for downscaling Sentinel data in this paper.

The existing CS and MRA approaches were developed originally for image fusion with a single fine band (such as the PAN band in PAN-sharpening). In Sentinel-2, there are *four* 10 m bands treated as such fine bands. Thus, the image fusion task for Sentinel-2 is different from, and more complex than, the conventional PAN-sharpening task. In this paper, the main aim was to extend the CS and MRA approaches to fusion of the Sentinel-2 10 m and 20 m bands. With respect to the advanced ATPRK approach, it can make use of all fine bands straightforwardly by a one-stage multiple regression, and can be applied readily for Sentinel-2 image fusion. The objectives of this paper were, thus, as follows.

1) To extend the existing CS and MRA approaches to fuse Sentinel-2 10 m and 20 m bands. This was achieved by producing a single band from the four fine bands, and an effective scheme was identified.
2) To apply ATPRK for fusion of Sentinel-2 data. ATPRK was used straightforwardly for the fusion problem by using all fine bands simultaneously, but also extended by using single band-based schemes.
3) To compare the three types of image fusion approaches (i.e., CS, MRA and ATPRK) and to identify the most accurate image fusion method for the Sentinel-2 fusion task.

The remainder of this paper is organized as follows. Section 2 first introduces the studied data, followed by the principles of the ATPRK approach and the scheme for extending CS, MRA and ATPRK for Sentinel-2 image fusion. The experimental results are provided in Section 3, in which the three types of image fusion approaches are compared. Section 4 further discusses the Sentinel-2 image fusion issue and the results, and Section 5 finally concludes the paper.

2. Methods

2.1. Data

The first Sentinel-2 products acquired on 18 August 2015 are released online. They are Level-1C products and provide geo-coded top of atmosphere reflectance with a sub-pixel multispectral registration (Drusch et al., 2012). Each product is a tile of 109 km by 109 km area in the UTM/WGS84 projection.

In this paper, we analyzed three datasets. The first one covers a scene in Verona, Italy, the second one covers a scene in Treviso, Italy, and the third one covers a scene in Malmo, Sweden. For each dataset, we selected an area with a spatial extent of 24 km by 24 km. The 20 m bands, thus, contain 1200 by 1200 pixels and the 10 m bands contain 2400 by 2400 pixels. Fig. 1 shows the three Sentinel-2 images. The three study areas are covered mainly by a mix of vegetation and urban fabric. As shown in the three sub-areas in Fig. 1(j)-Fig. 1(l), the 20 m images are visually more ambiguous than the 10 m images, and many textures evident in the 10 m images cannot be resolved in the 20 m images. This motivates downscaling the 20 m Sentinel-2 bands. The geometric registration between the 10 m and 20 m bands is critical in image fusion and the registration errors will potentially affect the accuracy greatly. For 20 m VNIR bands 5, 6, 7 and 8a of the Level-1C products, the correlation coefficient (CC) between them and the 10 m VNIR bands can reach 0.99, suggesting that the registration of the Level-1C products is highly reliable. Thus, no further registration was performed.

Naturally, the six 20 m bands can be fused with the four 10 m bands to produce ten-band 10 m Sentinel-2 images. In this case, however, no reference at 10 m can be used to examine the downscaling methods objectively. Thus, synthetic datasets (i.e., spatially degraded datasets) were used for reliable assessment and objective comparison between the image fusion approaches. More precisely, convolved with the pre-determined point spread function (PSF), the available 20 m bands and 10 m bands were first upscaled to 40 m and 20 m, respectively. Image fusion was then applied to the observed six 40 m coarse bands, treating the four 20 m bands as fine bands. The produced six-band 20 m fusion results were compared to the original 20 m bands (i.e., the perfect reference) for objective evaluation. This is a scheme used commonly in experimental studies to evaluate downscaling approaches (Atkinson, 2009).

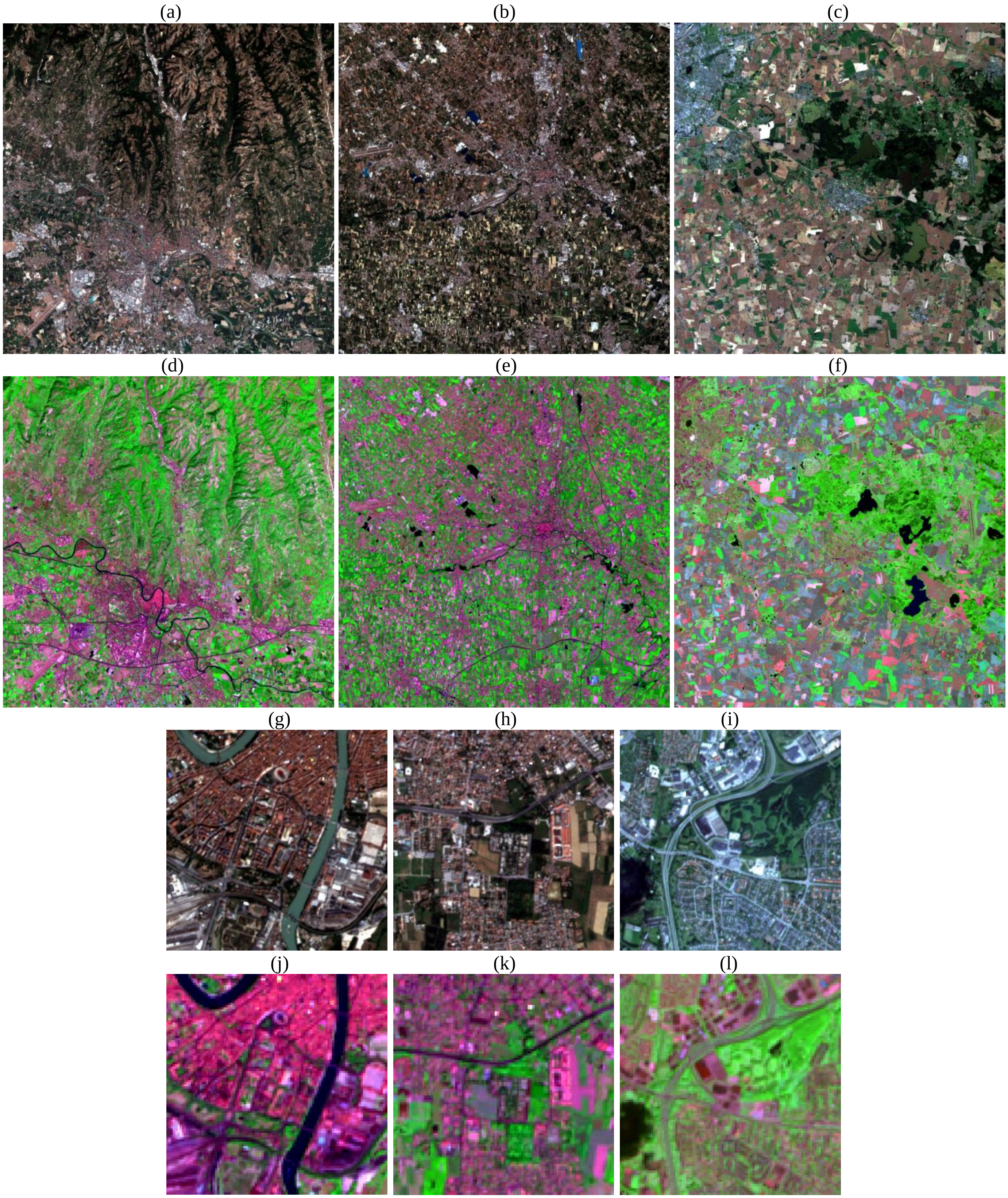


Fig. 1. The three studied Sentinel-2 images. (a)-(c) are 10 m data (2400 by 2400 pixels, bands 4, 3 and 2 as RGB) for Verona, Treviso and Malmo, respectively. (d)-(f) are 20 m data (1200 by 1200 pixels, bands 12, 8a and 5 as RGB) for Verona, Treviso and Malmo, respectively. (g)-(i) are sub-areas with 200 by 200 pixels in (a)-(c). (j)-(l) are three corresponding sub-areas with 100 by 100 pixels in (d)-(f).

2.2. ATPRK

ATPRK was first proposed for fusion of 500 m MODIS bands 3-7 with 250 m bands 1-2 in our previous research (Wang et al., 2015a). The approach consists of regression-based overall "trend" estimation (trend is a term used in geostatistics to refer to the spatially varying mean of a spatial process) and area-to-point (ATPK)-based residual downscaling (where residual refer to the variation remaining after removal of the trend) (Atkinson, 2013; Kerry et al., 2012; Kyriakidis, 2004; Kyriakidis & Yoo, 2005). ATPRK can be viewed as an extension of either regression kriging (Hengl et al., 2004, 2007) or ATPK (Wang et al., 2015a). It is a fast and user-friendly approach. Inheriting the advantages of ATPK, ATPRK accounts *explicitly* for size of support (pixel), spatial correlation, and PSF of the sensor. Importantly, it has the appealing advantage that it can precisely preserve the spectral properties of the original coarse data, that is, it is perfectly coherent. The proof of perfect coherence of ATPRK can be found in Appendix A in Wang et al. (2015a), which is based on the perfect coherence of ATPK (see Pages 267-269 in Kyriakidis (2004)). The coherence characteristic of ATPK and ATPRK is not affected by the specific form of PSF (Kyriakidis, 2004; Wang et al., 2015a).

The principle of ATPRK is briefly introduced in this section. Suppose $Z_V^l(\mathbf{x}_i)$ is the random vector (i.e., brightness value) of pixel V centered at $\mathbf{x}_i$ (i=1,…,M, where M is the number of pixels) in coarse band l (l=1,…,6), and $Z_v^k(\mathbf{x}_j)$ is the random vector of pixel v centered at $\mathbf{x}_j$ (j=1,…, MF^2, where F=2 is the spatial resolution ratio between the coarse and fine spatial resolution bands) in fine band k (k=1,…,4). ATPRK aims to predict the target variable $Z_v^l(\mathbf{x})$ for all fine pixels in all six coarse bands.

Let $\hat{Z}_{v1}^l(\mathbf{x})$ and $\hat{Z}_{v2}^l(\mathbf{x})$ be the predictions of the regression and the ATPK parts, respectively. The ATPRK prediction is given by

$$\hat{Z}_v^l(\mathbf{x}) = \hat{Z}_{v1}^l(\mathbf{x}) + \hat{Z}_{v2}^l(\mathbf{x}) . \quad (1)$$

At a specific location $\mathbf{x}_0$, the regression prediction is calculated as a linear combination of the four fine pixels in the corresponding four fine bands

$$\hat{Z}_{v1}^{l}(\mathbf{x}_0)=\sum_{k=0}^{4}a_k^l Z_v^k(\mathbf{x}_0),\ Z_v^0(\mathbf{x}_0)=1\,. \tag{2}$$

Based on the assumption of scale-invariance, the coefficients $\{a_k^l \mid k=0,...,4\}$ in (2) are calculated according to the relationship between the observed coarse band l and the upscaled bands Z_V^k (k=1,…,4) (created by convolving the fine band with the pre-determined PSF) from the original four fine bands

$$Z_V^l(\mathbf{x})=\sum_{k=0}^{4}a_k^l Z_V^k(\mathbf{x})+R_V^l(\mathbf{x}),\ Z_V^0(\mathbf{x})=1\ \forall\mathbf{x} \tag{3}$$

where $R_V^l(\mathbf{x})$ is a residual term and the coefficients are estimated by ordinary least squares.

After regression modeling, ATPK is performed in the second-stage to downscale the coarse residuals $R_V^l(\mathbf{x})$ in (3) to the desired fine spatial resolution. ATPK-based residual downscaling retains the spectral information in the original coarse data. The fine residual at location $\mathbf{x}_0$ is calculated as

$$\hat{Z}_{v2}^{l}(\mathbf{x}_0)=\sum_{i=1}^{N}\lambda_i R_V^l(\mathbf{x}_i),\ \text{s.t.}\ \sum_{i=1}^{N}\lambda_i=1 \tag{4}$$

in which λ_i is the weight for the ith coarse residual centered at $\mathbf{x}_i$ and N is the number of neighboring coarse pixels. The weights $\{\lambda_i \mid i=1,...,N\}$ are calculated according to the kriging matrix

$$\begin{bmatrix}\gamma_{VV}^l(\mathbf{x}_1,\mathbf{x}_1) & \cdots & \gamma_{VV}^l(\mathbf{x}_1,\mathbf{x}_N) & 1\\ \vdots & \ddots & \vdots & \vdots\\ \gamma_{VV}^l(\mathbf{x}_N,\mathbf{x}_1) & \cdots & \gamma_{VV}^l(\mathbf{x}_N,\mathbf{x}_N) & 1\\ 1 & \cdots & 1 & 0\end{bmatrix}\begin{bmatrix}\lambda_1\\ \vdots\\ \lambda_N\\ \theta\end{bmatrix}=\begin{bmatrix}\gamma_{vV}^l(\mathbf{x}_0,\mathbf{x}_1)\\ \vdots\\ \gamma_{vV}^l(\mathbf{x}_0,\mathbf{x}_N)\\ 1\end{bmatrix} \tag{5}$$

where $\gamma_{VV}^l(\mathbf{x}_i,\mathbf{x}_j)$ is the coarse-to-coarse semivariogram between coarse pixels centered at $\mathbf{x}_i$ and $\mathbf{x}_j$ in band l, $\gamma_{vV}^l(\mathbf{x}_0,\mathbf{x}_j)$ is the fine-to-coarse semivariogram between fine and coarse pixels centered at $\mathbf{x}_0$ and $\mathbf{x}_j$ in band l, and θ is the Lagrange multiplier. Let $\mathbf{s}$ be the Euclidean distance between the centroids of any two pixels and $h_V^l(\mathbf{s})$ be the PSF of the sensor. $\gamma_{VV}^l(\mathbf{s})$ and $\gamma_{vV}^l(\mathbf{s})$ are calculated by convoluting the fine-to-fine semivariogram $\gamma_{vv}^l(\mathbf{s})$ with the PSF $h_V^l(\mathbf{s})$ as follows

$$\gamma_{vV}^l(\mathbf{s})=\gamma_{vv}^l(\mathbf{s})*h_V^l(\mathbf{s}) \tag{6}$$

$$\gamma_{VV}^{l}(\mathbf{s}) = \gamma_{vv}^{l}(\mathbf{s}) * h_{V}^{l}(\mathbf{s}) * h_{V}^{l}(-\mathbf{s}) \tag{7}$$

where * is the convolution operator. $\gamma_{vv}^{l}(\mathbf{s})$ is estimated by deconvolution of the coarse semivariogram calculated from the coarse residual image $R_{V}^{l}(\mathbf{x})$. Details on the deconvolution approach can be found in Wang et al. (2015a,b). The sensor PSF is assumed to be the Gaussian filter in (8), which is commonly used in remote sensing

$$h_{V}(\mathbf{x}) = \begin{cases} \dfrac{1}{2\pi\sigma^{2}} \exp\left[-\left(\dfrac{x_{1}^{2} + x_{2}^{2}}{2\sigma^{2}} \right) \right], & \text{if } \mathbf{x} \in V(\mathbf{x}) \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

in which σ is the standard deviation (width of the Gaussian PSF), x_1 and x_2 are coordinates of location $\mathbf{x}$ (i.e., $\mathbf{x} = \{x_1, x_2\}$), $V(\mathbf{x})$ is the spatial neighborhood of the pixel centered at **x**. In this paper, the width of the Gaussian PSF was set to half of the pixel size.

2.3. Image fusion for Sentinel-2

In this paper, for Sentinel-2 image fusion, we consider four CS methods, including adaptive BT (BTA) (Gillespie et al., 1987, Aiazzi et al., 2007), GSA (Aiazzi et al., 2007), context-adaptive GSA (GSA-CA) (Aiazzi et al., 2009) and PRACS (Choi et al., 2011)). We also consider four MRA methods, including ATWT-M3 (Ranchin & Wald, 2000), MTF-GLP (Aiazzi et al., 2006; Vivone et al., 2015), MTF-GLP with context-based decision (MTF-GLP-CBD) (Alparone et al., 2007), and MTF-GLP with high-pass modulation (MTF-GLP-HPM) (Aiazzi et al., 2003).

CS and MRA use a single fine band (e.g., PAN band) to sharpen the coarse bands. This means that a single band needs to be extracted from the four fine bands in Sentinel-2 data. This issue in image fusion has been conceptualized as “hyper-sharpening” in the very recent literature, which originally means fusion of a fine spatial resolution multispectral image with a coarse hyperspectral image (Selva et al., 2015). For accommodation of fine spatial resolution information from multiple fine bands, two schemes were proposed in

Selva et al. (2015), that is, the synthesized band scheme and the selected band scheme. The selected band scheme selects a fine band from the fine band set for each coarse band, which is determined as the one with the largest correlation with the visited coarse band.

The synthesized band scheme synthesizes a single fine band from the fine band set (i.e., fine multispectral image), such as averaging all fine multispectral bands (Selva et al., 2015). However, the synthesized band scheme based on the simple averaging process fails to consider the relation between the visited coarse band and the four fine bands. In this paper, to fully account for the information in the four fine bands, the synthesized band for each coarse band is determined adaptively as a linear combination of the four fine bands. The weights are calculated according to the multiple regression model built between the visited coarse band and the four fine bands. This is a process similar to that in (2) and (3) in ATPRK.

For the fusion of Sentinel-2 images, this paper extends the three types of image fusion approaches (i.e., CS, MRA and ATPRK) using the synthesized and selected band schemes. Based on the multiple regression model in (3), the use of the synthesized band in ATPRK amounts to the use of all four fine bands directly. The performances of the three types of approaches coupled with the two band extraction schemes are illustrated in the following experiments.

3. Experiments

3.1 Experiment on the spatially degraded datasets

Four indices were used for quantitative evaluation, including the CC, universal image quality index (UIQI) (Wang & Bovik, 2002), relative global-dimensional synthesis error (ERGAS) (Ranchin & Wald, 2000), and spectral angle mapper (SAM). For CC and UIQI, they were first calculated for each band, and then the values for all bands were averaged. For SAM, values for the spectra of all pixels were first calculated and then averaged. As mentioned in Wald et al. (1997), in image fusion, it is important that any synthetic image, once degraded to its original spatial resolution, should be as close as possible to the original image. Thus, we also used CC, UIQI, ERGAS and SAM to evaluate the consistency. The fused image was upscaled to the original coarse spatial resolution and the observed coarse image was used as reference. In the upscaling process, exactly the same PSF (i.e., the Gaussian filter in (8)) used for simulating the 40 m and 20 m images and creating Z_V^k (k=1,…,4) in the regression modeling in (3) should be used. That is, the PSF should be consistent.

Figs 2-4 show the downscaling results for three sub-areas of the three datasets. Note that for each method, the band extraction scheme leading to the greatest accuracy is shown. For MRA and ATPRK, the results are those obtained with the synthesized bands. For CS, the results are obtained with the selected bands (except BTA for the Malmo data and PRACS for the Treviso data). The selected bands for the three datasets are displayed in Table 2. For the coarse bands, the selected fine bands generally have the closest wavelengths with them. For example, for the VNIR bands 6 (733-748nm), 7 (765-785nm) and 8a (855-875nm), the spectrally closest band 8 (785-900nm) is selected.

Table 2 The selected fine band (bands 2, 3, 4 or 8) for each coarse band (bands 5, 6, 7, 8a, 11 or 12) of the three datasets

	Band 5	Band 6	Band 7	Band 8a	Band 11	Band 12
Verona	Band 4	Band 8	Band 8	Band 8	Band 4	Band 4
Treviso	Band 4	Band 8	Band 8	Band 8	Band 4	Band 4
Malmo	Band 3	Band 8	Band 8	Band 8	Band 3	Band 4

It can be seen that the BTA result in Fig. 4(d), GSA result in Fig. 4(e) and GSA-CA result in Fig. 4(f) lead to obvious spectral distortion. This is clearly illustrated by the results for road restoration. PRACS and ATWT-M3 produce over-smooth results, in which the texture (e.g., the elongated features in urban areas) cannot be observed very clearly, see Fig. 2(g), Fig. 2(h), Fig. 3(h) and Fig. 4(h). The MTF-GLP-CBD results tend to be superior to MTF-GLP and MTF-GLP-HPM, and are closest to ATPRK results. Tables 3-5 list the quantitative assessment for the three datasets at the target fine spatial resolution. Three observations can be made from the comparisons.

First, for the CS methods, greater accuracies are produced when the selected band scheme is used. For example, for the Verona data, the CC values of GSA and GSA-CA increase by 0.007 and 0.003 using the selected band scheme. Moreover, the GSA and GSA-CA methods are more accurate than BTA and PRACS.

Second, for the MRA and ATPRK methods, the synthesized band scheme is able to produce greater accuracies than the synthesized band scheme. MRA (except ATWT-M3) with the synthesized band is superior to CS with the selected band. Furthermore, the general rank of the four MRA methods in terms of accuracy (from most to least accurate) are MTF-GLP-CBD, MTF-GLP, MTF-GLP-HPM and ATWT-M3. The advantages of the former three methods over ATWT-M3 are obvious. More precisely, for all three datasets, the CC values of the three MRA methods are at least 0.01 larger than that of ATWT-M3, and the gains in UIQI are even larger (e.g., over 0.02 for the Treviso data).

Third, given the same scheme of band extraction for all methods, ATPRK produces the greatest accuracy amongst the three types of approaches. Take the synthesized band scheme as an example, ATPRK produces the largest CC (0.9932, 0.9916 and 0.9963 for Verona, Treviso and Malmo, respectively) and UIQI (0.9931, 0.9914 and 0.9963 for Verona, Treviso and Malmo, respectively) and smallest ERGAS (1.5374, 1.7618 and 1.3456 for Verona, Treviso and Malmo, respectively) and SAM (0.0217, 0.0237 and 0.0198 for Verona, Treviso and Malmo, respectively) for all three datasets.

Tables 6-8 present the consistency of the downscaling methods for the three datasets at coarse spatial resolution. Consistency was used here because it was demonstrated to be able to give reliable assessment of the relative performance of image fusion and to be superior to the commonly used quality no reference (QNR) metrics (Palsson et al., 2016). The three MRA methods, including MTF-GLP-CBD, MTF-GLP, MTF-GLP-HPM, outperform all four CS methods. It is worth noting that ATPRK achieves the ideal CC and UIQI and almost ideal ERGAS and SAM for all three datasets, demonstrating that it can perfectly preserve the spectral properties of the original coarse data.

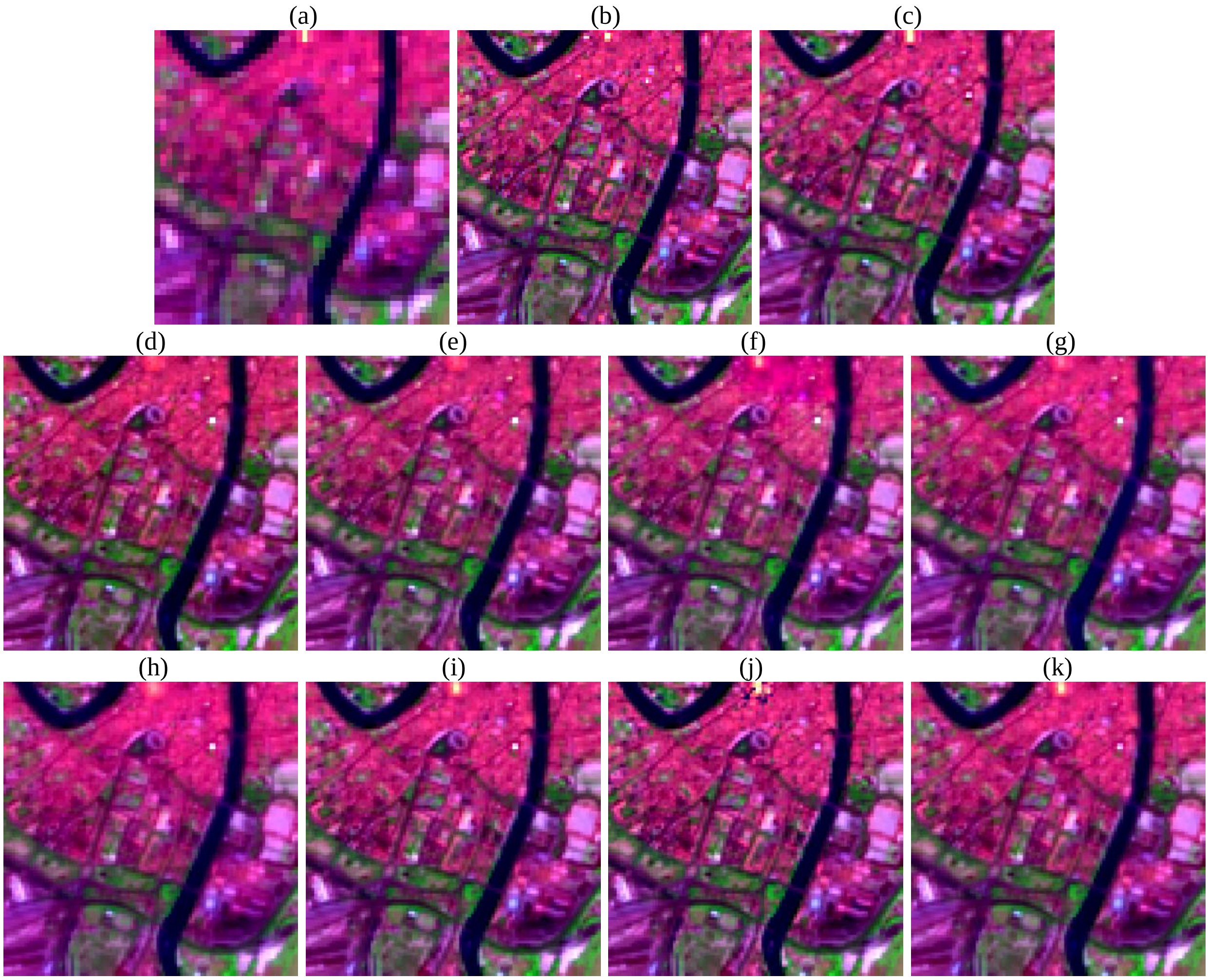


Fig. 2. Downscaling results (20 m) for the sub-area in Verona (bands 12, 8a and 5 as RGB). (a) 40 m coarse image. (b) 20 m reference image. (c) ATPRK. (d) BTA. (e) GSA. (f) GSA-CA. (g) PRACS. (h) ATWT-M3. (i) MTF-GLP. (j) MTF-GLP-CBD. (k) MTF-GLP-HPM. For each method in (h)-(k), the result for the case with greatest accuracy is shown.

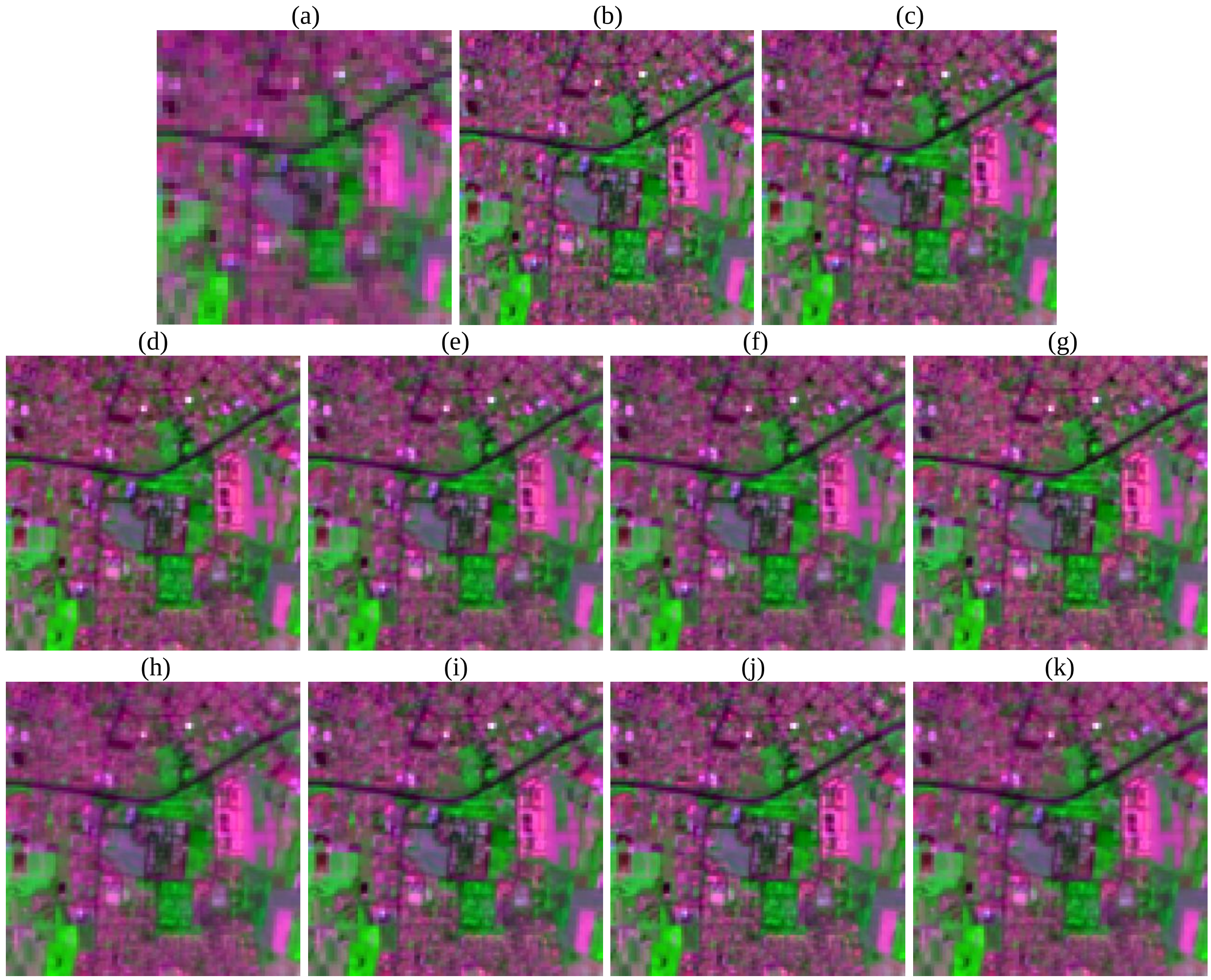


Fig. 3. Downscaling results (20 m) for the sub-area in Treviso (bands 12, 8a and 5 as RGB). (a) 40 m coarse image. (b) 20 m reference image. (c) ATPRK. (d) BTA. (e) GSA. (f) GSA-CA. (g) PRACS. (h) ATWT-M3. (i) MTF-GLP. (j) MTF-GLP-CBD. (k) MTF-GLP-HPM. For each method in (h)-(k), the result for the case with greatest accuracy is shown.

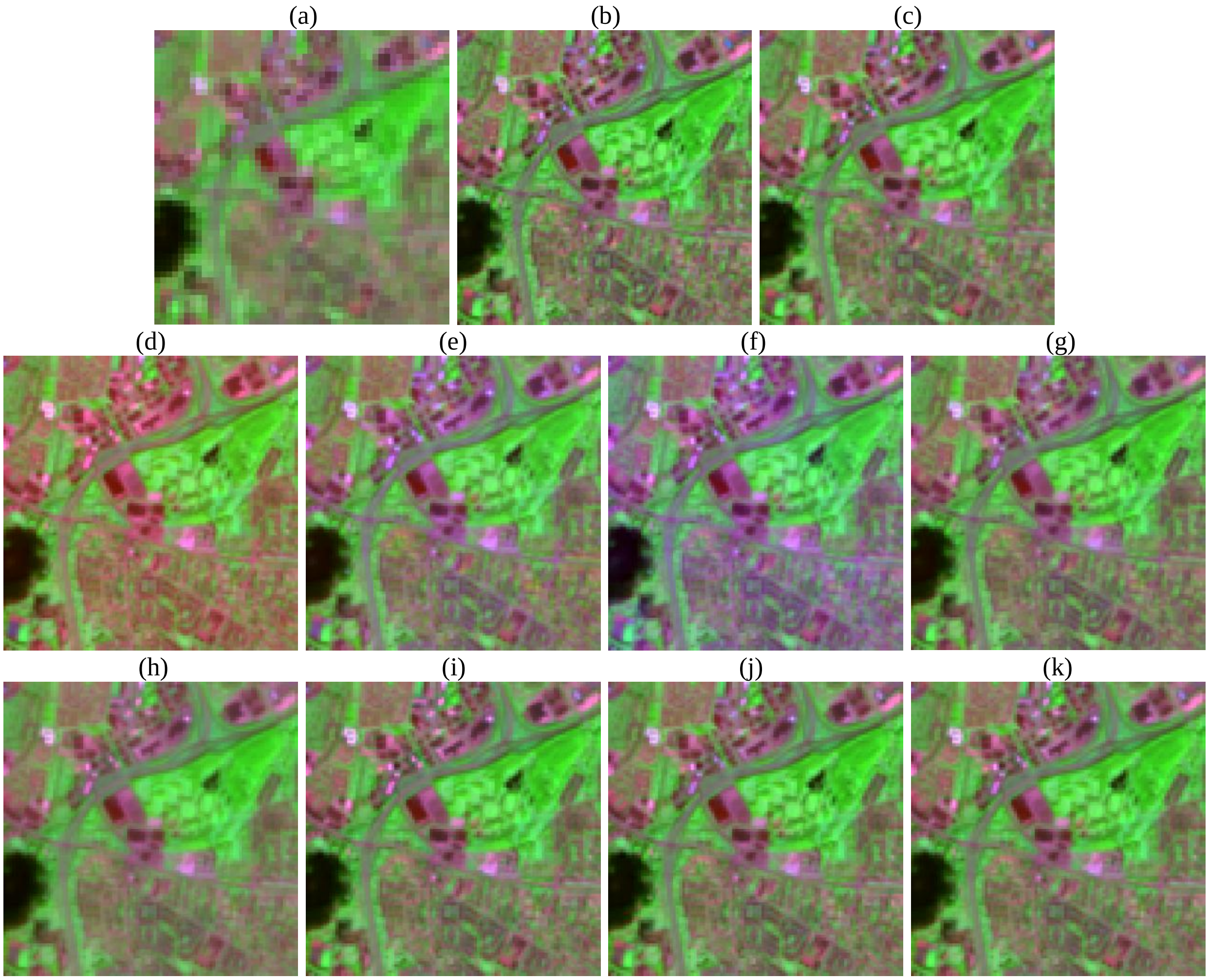


Fig. 4. Downscaling results (20 m) for the sub-area in Malmo (bands 12, 8a and 5 as RGB). (a) 40 m coarse image. (b) 20 m reference image. (c) ATPRK. (d) BTA. (e) GSA. (f) GSA-CA. (g) PRACS. (h) ATWT-M3. (i) MTF-GLP. (j) MTF-GLP-CBD. (k) MTF-GLP-HPM. For each method in (h)-(k), the result for the case with greatest accuracy is shown.

Table 3 Quantitative assessment of the downscaling methods at fine spatial resolution of 20 m for the Verona data

		CC	ERGAS	UIQI	SAM
Ideal		1	0	1	0
BTA	Synthesized	0.9807	3.0684	0.9702	0.0396
	Selected	0.9818	2.8730	0.9731	0.0375
GSA	Synthesized	0.9784	2.9111	0.9781	0.0417
	Selected	0.9852	2.3126	0.9843	0.0297
GSA-CA	Synthesized	0.9824	2.6451	0.9823	0.0366
	Selected	0.9857	2.3010	0.9852	0.0315
PRACS	Synthesized	0.9796	2.8179	0.9764	0.0378
	Selected	0.9823	2.6974	0.9774	0.0328
ATWT-M3	Synthesized	0.9761	3.1499	0.9677	0.0376
	Selected	0.9723	3.4039	0.9633	0.0406
MTF-GLP	Synthesized	0.9883	2.0504	0.9873	0.0269
	Selected	0.9857	2.2670	0.9843	0.0295
MTF-GLP-CBD	Synthesized	0.9898	1.9321	0.9894	0.0268
	Selected	0.9867	2.2199	0.9863	0.0336
MTF-GLP-HPM	Synthesized	0.9874	2.1975	0.9856	0.0286
	Selected	0.9867	2.1716	0.9854	0.0281
ATPRK	Synthesized	0.9932	1.5374	0.9931	0.0217
	Selected	0.9915	1.7034	0.9913	0.0232

Table 4 Quantitative assessment of the downscaling methods at fine spatial resolution of 20 m for the Treviso data

		CC	ERGAS	UIQI	SAM
Ideal		1	0	1	0
BTA	Synthesized	0.9806	3.2863	0.9660	0.0391
	Selected	0.9810	3.1430	0.9683	0.0388
GSA	Synthesized	0.9796	2.8244	0.9793	0.0485
	Selected	0.9833	2.4932	0.9823	0.0317
GSA-CA	Synthesized	0.9845	2.4514	0.9843	0.0333
	Selected	0.9843	2.4138	0.9835	0.0311
PRACS	Synthesized	0.9807	2.8661	0.9754	0.0365
	Selected	0.9792	3.0492	0.9722	0.0380
ATWT-M3	Synthesized	0.9716	3.5460	0.9597	0.0412
	Selected	0.9679	3.7639	0.9556	0.0440
MTF-GLP	Synthesized	0.9854	2.3448	0.9838	0.0289
	Selected	0.9829	2.5023	0.9810	0.0307
MTF-GLP-CBD	Synthesized	0.9870	2.2090	0.9865	0.0285
	Selected	0.9850	2.3474	0.9843	0.0307
MTF-GLP-HPM	Synthesized	0.9842	2.5561	0.9816	0.0314
	Selected	0.9838	2.4353	0.9820	0.0303
ATPRK	Synthesized	0.9916	1.7618	0.9914	0.0237
	Selected	0.9897	1.9124	0.9895	0.0256

Table 5 Quantitative assessment of the downscaling methods at fine spatial resolution of 20 m for the Malmo data

		CC	ERGAS	UIQI	SAM
Ideal		1	0	1	0
BTA	Synthesized	0.9862	2.9574	0.9821	0.0388
	Selected	0.9853	2.8504	0.9816	0.0346
GSA	Synthesized	0.9841	3.0043	0.9840	0.0527
	Selected	0.9884	2.3572	0.9882	0.0346
GSA-CA	Synthesized	0.9801	3.4851	0.9790	0.0716
	Selected	0.9824	2.9505	0.9821	0.0588
PRACS	Synthesized	0.9909	2.2179	0.9896	0.0335
	Selected	0.9906	2.2059	0.9892	0.0307
ATWT-M3	Synthesized	0.9844	2.8906	0.9812	0.0350
	Selected	0.9822	3.0349	0.9787	0.0363

MTF-GLP	Synthesized	0.9927	1.9103	0.9925	0.0277
	Selected	0.9918	1.9742	0.9916	0.0286
MTF-GLP-CBD	Synthesized	0.9923	1.9598	0.9922	0.0280
	Selected	0.9916	2.0121	0.9915	0.0296
MTF-GLP-HPM	Synthesized	0.9931	1.8444	0.9929	0.0259
	Selected	0.9924	1.9077	0.9921	0.0263
ATPRK	Synthesized	0.9963	1.3456	0.9963	0.0198
	Selected	0.9959	1.3983	0.9958	0.0208

Table 6 Quantitative assessment of the downscaling methods at coarse spatial resolution of 40 m for the Verona data

		CC	ERGAS	UIQI	SAM
Ideal		1	0	1	0
BTA	Synthesized	0.9885	2.2276	0.9808	0.0287
	Selected	0.9914	1.9589	0.9839	0.0248
GSA	Synthesized	0.9866	2.0717	0.9865	0.0309
	Selected	0.9956	1.1216	0.9955	0.0131
GSA-CA	Synthesized	0.9898	1.8708	0.9897	0.0260
	Selected	0.9959	1.1521	0.9958	0.0159
PRACS	Synthesized	0.9893	1.8453	0.9880	0.0254
	Selected	0.9957	1.2202	0.9942	0.0147
ATWT-M3	Synthesized	0.9922	1.6580	0.9897	0.0193
	Selected	0.9909	1.8093	0.9881	0.0216
MTF-GLP	Synthesized	0.9986	0.6792	0.9985	0.0083
	Selected	0.9981	0.7884	0.9979	0.0105
MTF-GLP-CBD	Synthesized	0.9990	0.5704	0.9990	0.0074
	Selected	0.9984	0.7190	0.9984	0.0118
MTF-GLP-HPM	Synthesized	0.9984	0.7688	0.9981	0.0101
	Selected	0.9983	0.7333	0.9982	0.0097
ATPRK	Synthesized	0.9999	0.1518	0.9999	0.0020
	Selected	0.9999	0.1843	0.9999	0.0026

Table 7 Quantitative assessment of the downscaling methods at coarse spatial resolution of 40 m for the Treviso data

		CC	ERGAS	UIQI	SAM
Ideal		1	0	1	0
BTA	Synthesized	0.9905	2.2026	0.9802	0.0260
	Selected	0.9931	2.0278	0.9821	0.0243
GSA	Synthesized	0.9899	1.7648	0.9899	0.0377
	Selected	0.9960	1.1059	0.9959	0.0141
GSA-CA	Synthesized	0.9945	1.3135	0.9944	0.0189
	Selected	0.9967	1.0021	0.9966	0.0129
PRACS	Synthesized	0.9933	1.5464	0.9912	0.0205
	Selected	0.9959	1.2962	0.9937	0.0187
ATWT-M3	Synthesized	0.9906	1.8251	0.9868	0.0212
	Selected	0.9894	1.9424	0.9855	0.0228
MTF-GLP	Synthesized	0.9982	0.7438	0.9981	0.0089
	Selected	0.9978	0.8090	0.9976	0.0100
MTF-GLP-CBD	Synthesized	0.9986	0.6447	0.9986	0.0081
	Selected	0.9983	0.7072	0.9982	0.0094
MTF-GLP-HPM	Synthesized	0.9980	0.8554	0.9976	0.0105
	Selected	0.9979	0.7818	0.9977	0.0098
ATPRK	Synthesized	0.9998	0.2296	0.9998	0.0031
	Selected	0.9998	0.2571	0.9998	0.0039

Table 8 Quantitative assessment of the downscaling methods at coarse spatial resolution of 40 m for the Malmo data

		CC	ERGAS	UIQI	SAM
Ideal		1	0	1	0
BTA	Synthesized	0.9911	2.3194	0.9881	0.0299
	Selected	0.9909	2.1611	0.9882	0.0239
GSA	Synthesized	0.9892	2.4027	0.9891	0.0462
	Selected	0.9940	1.5933	0.9940	0.0245
GSA-CA	Synthesized	0.9848	3.0103	0.9831	0.0644
	Selected	0.9876	2.3700	0.9871	0.0509
PRACS	Synthesized	0.9966	1.3174	0.9960	0.0203
	Selected	0.9978	1.0619	0.9973	0.0158
ATWT-M3	Synthesized	0.9948	1.5974	0.9937	0.0184
	Selected	0.9940	1.6809	0.9929	0.0195
MTF-GLP	Synthesized	0.9989	0.7026	0.9989	0.0113
	Selected	0.9988	0.7458	0.9987	0.0119
MTF-GLP-CBD	Synthesized	0.9991	0.6547	0.9991	0.0086
	Selected	0.9989	0.6982	0.9989	0.0100
MTF-GLP-HPM	Synthesized	0.9990	0.6754	0.9990	0.0088
	Selected	0.9988	0.7200	0.9988	0.0096
ATPRK	Synthesized	1	0.1518	1	0.0019
	Selected	0.9999	0.1705	0.9999	0.0025

3.2 Experiment on real datasets

From the test in section 3.1, it is evident that ATPRK with the synthesized band can produce more accurate results than the other methods. In this experiment, we fused the original 20 m band 10 m data using the ATPRK with the synthesized band approach. The 10 m downscaling results for the three datasets are shown in Figs 5-7, where each figure show results for three sub-areas. We visually compare the 10 m results with the original 20 m images. It can be seen clearly that by borrowing the 10 m information from the four-band 10 m data, more explicit information can be presented. The small patches, boundaries of classes and textural information (e.g., elongated features) are shown more clearly in the 10 m downscaling results. Moreover, the spectral information of the 20 m data is accurately retained.

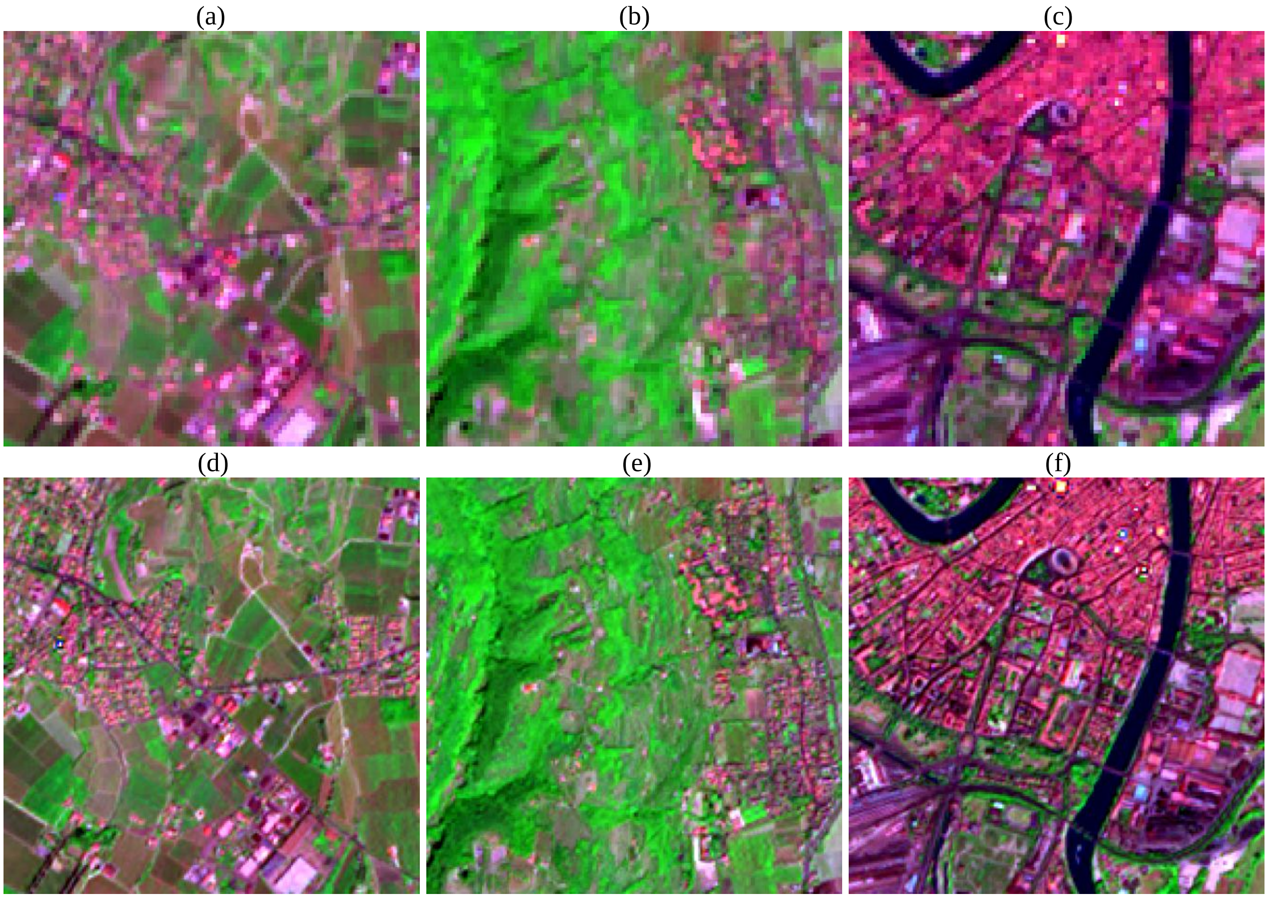
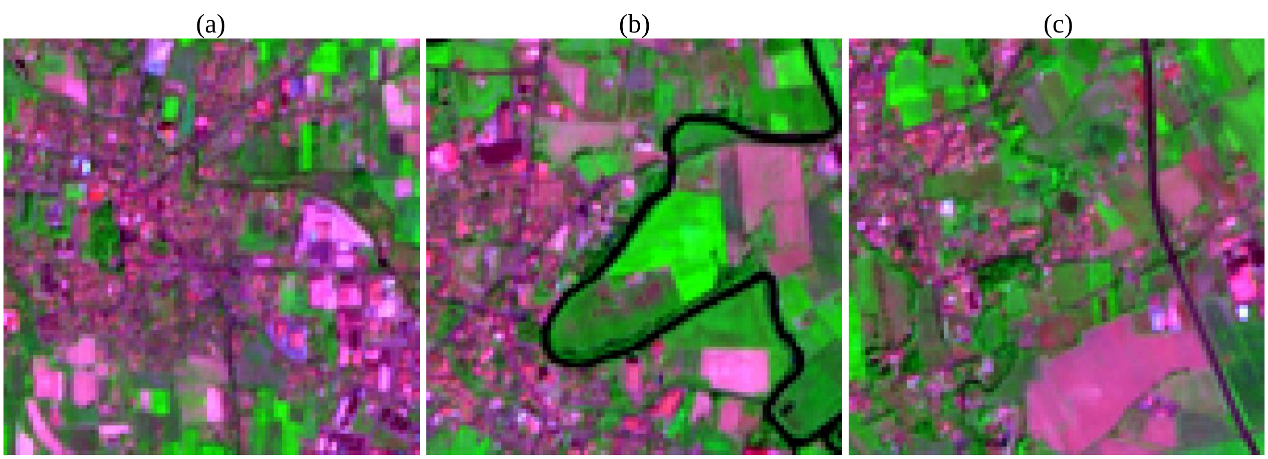


Fig. 5. Downscaling results (10 m) for three sub-areas in Verona (bands 12, 8a and 5 as RGB). (a)-(c) The 20 m data. (d)-(f) The 10 m results.



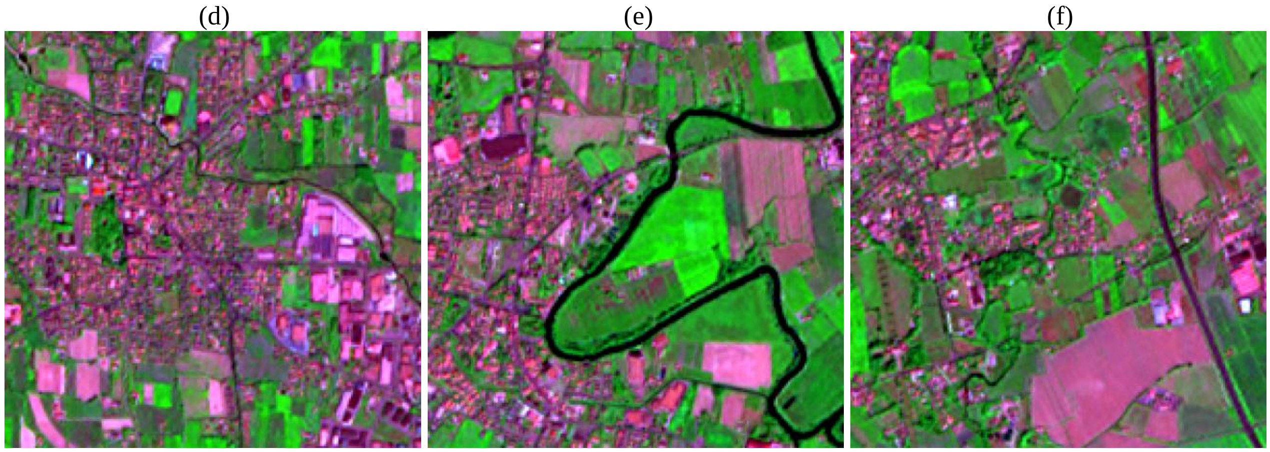


Fig. 6. Downscaling results (10 m) for three sub-areas in Treviso (bands 12, 8a and 5 as RGB). (a)-(c) The 20 m data. (d)-(f) The 10 m results.

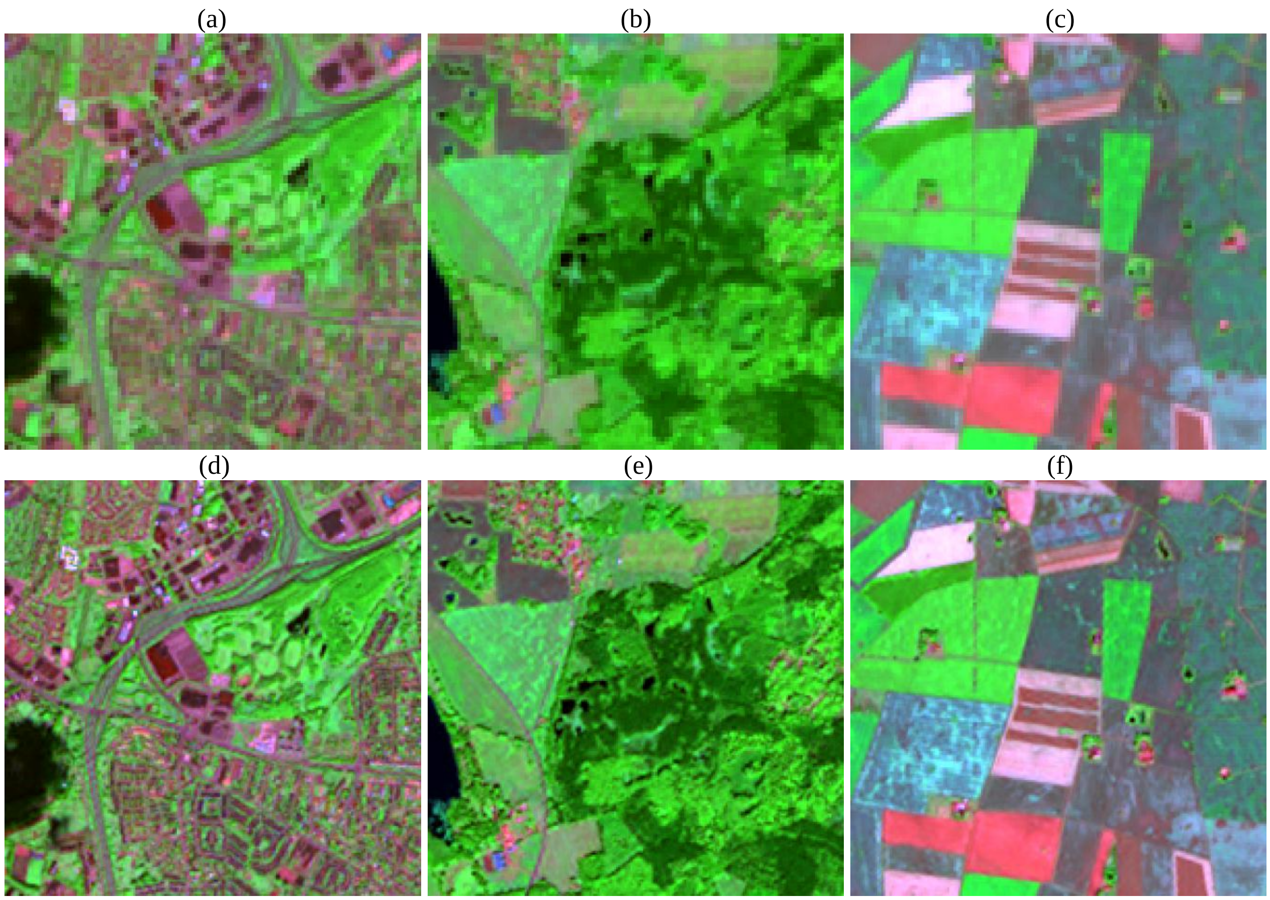


Fig. 7. Downscaling results (10 m) for three sub-areas in Malmo (bands 12, 8a and 5 as RGB). (a)-(c) The 20 m data. (d)-(f) The 10 m results.

4. Discussion

Sentinel-2 provides six bands (5, 6, 7, 8a, 11 and 12) with a spatial resolution of 20 m, and only four bands (2, 3, 4 and 8) at the finer spatial resolution of 10 m. It is, therefore, a natural task to take full advantage of the spatial information in the four 10 m bands to downscale the six 20 m bands to the finer spatial resolution, to provide users with more detailed information in the six bands, and create a complete set of 10 bands with a fine spatial resolution of 10 m. This paper achieves this important objective by extending eight existing image fusion approaches (i.e., CS and MRA), and the recently developed ATPRK approach, to the specific Sentinel-2 image fusion issue.

Fusing the six Sentinel-2 20 m bands with the four 10 m bands is different from the conventional PAN-sharpening problem where a single PAN image is available as the reference for several coarse multispectral bands. This paper, thus, extends eight CS and MRA approaches to make them suitable for Sentinel-2 image fusion via two schemes, the synthesized band and selected band schemes. The extended versions were tested in the experiments and their performances were also compared to the advanced ATPRK approach. It was shown that the bands 5, 6, 7, 8a, 11 and 12 downscaled to 10 m were visually clearer than the 20 m bands, and greater spatial detail (e.g., small patches and elongated features) was reproduced. This will certainly increase the utility of these six bands in a greater number and range of applications. For example, as shown in Fig. 5(f), Fig. 6(d) and Fig. 7(d), the urban fabric can be observed much more clearly after image fusion. The encouraging results will motivate the utility of these six downscaled bands in urban mapping.

Experimental comparison between the three types of approaches revealed the advantages of ATPRK for the fusion of Sentinel-2 images: ATPRK reproduces accurate spatial textures and preserves perfectly the spectral properties of the original coarse data. The quantitative assessment illustrated that ATPRK consistently produces greater accuracies than all of the CS and MRA approaches. In addition, for ATPRK, the synthesized band scheme can produce more accurate fusion results than the selected band scheme, suggesting that the former band extraction scheme is a preferable choice for ATPRK-based Sentinel-2 image fusion.

This paper provides the first guidance (including the option of the band extraction scheme and the image fusion approach) for fusion of Sentinel-2 images. Applying the image fusion approaches to freely available multi-temporal Sentinel-2 images, ten-band 10 m time-series products will be produced. Appreciating the wide swath and frequent revisit capabilities, such 10 m products will show great potential for dynamic LCLU monitoring and change detection at the global scale. For example, bands 3 (centered at 560 nm) and 11 (centered at 1610 nm) of Sentinel-2 can be used to calculate the Normalized Difference Snow Index (NDSI) and Modified Normalized Difference Water Index (MDNWI) (Xu, 2006), which can extract snow-covered areas and water bodies (e.g., inland water in urban areas) on the Earth, respectively. In the original Sentinel-2 data, however, they are at different spatial resolutions (10 m for bands 3 and 20 m for band 11). The fused ten-band 10 m Sentinel-2 dataset offers compatibility to calculate the NDSI and MDNWI at 10 m, a spatial resolution finer than that of the original band 11.

The ten-band 10 m time-series products will offer excellent opportunities for blending with the forthcoming Sentinel-3 data (Berger & Aschbacher, 2012; Donlon et al., 2012; Verhoef & Bach, 2012). The Sentinel-3 mission aims to provide fine temporal resolution (<2 days) data for timely monitoring. The 21 bands for the task (i.e., the Ocean and Land Colour Instrument (OLCI) bands), however, are provided at a coarse spatial resolution of 300 m. The ten-band 10 m Sentinel-2 data can be blended with the Sentinel-3 data to generate time-series data at both fine temporal resolution (<2 days) and spatial resolution (10 m). Specifically, bands 5, 6, 7 and 8a of Sentinel-2 correspondingly have close wavelength as OLCI bands 11, 12, 16 and 17 of Sentinel-3. With the approaches developed in this paper, Sentinel-2 will be able to provide important reference data with a spatial resolution of 10 m for the 300 m OLCI bands 11, 12, 16 and 17 of Sentinel-3. It is worth noting that, if achieveable, the 10 m time-series data produced by blending Sentinel-2 with Sentinel-3 data would have appealing advantages in terms of spatial resolution in comparison with the 30 m data produced by blending Landsat and MODIS data (Gao et al., 2006, 2015; Zhu et al., 2010). This will be an important topic for future research in remote sensing.

For the Sentinel-2 mission, the required signal-to-noise ratio (SNR) for bands 5, 6, 7, 8a, 11 and 12 can be achieved only at a spatial resolution coarser than 10 m (i.e., 20 m in the products). This necessitates the investigation of the relation between the SNR of the 10 m fused images and the original 20 m images. We approximately calculated the SNR of an image based on the ratio of the mean value to the standard deviation of the image. Table 9 lists the reduction in SNR for each band after image fusion for the three datasets, where the ATPRK method with the synthesized band scheme was used. The SNR reduction is a function of the study area and sharpened band of interest. Generally, the reduction is smaller than 10%. The reduction in SNR is a cost of sharpening the 20 m bands.

Table 9 The percentage of SNR reduction after downscaling bands 5, 6, 7, 8a, 11 and 12 based on ATPRK with the synthesized band scheme

	Band 5	Band 6	Band 7	Band 8a	Band 11	Band 12
Verona	10.94%	9.67%	8.88%	8.15%	7.66%	10.52%
Treviso	12.73%	9.62%	9.10%	8.55%	8.87%	13.44%
Malmo	5.14%	6.85%	7.18%	7.01%	3.44%	4.69%

This paper is the first to downscale Sentinel-2 bands from 20 m to 10 m spatial resolution. The proposed approaches were tested with three 24 km by 24 km datasets. It is encouraged to test further cases (e.g., datasets with larger size) in future research, to provide a more comprehensive comparison of the available image fusion approaches. Moreover, in addition to the synthesized and selected bands schemes, it is of great interest to develop new alternatives to make full use of the four 10 m bands of Sentinel-2.

The 30 m Landsat 8 operational land imager (OLI) data have been used widely for global monitoring. By pan-sharpening, the OLI data can be increased to a finer spatial resolution of 15 m (Zhang & Roy, 2016). Such a resolution is coarser than 10 m that can be achieved by the new Sentinel-2 data associated with an image fusion approach. Such a difference in spatial resolution (i.e., 5 m) may be critical for identifying small targets, such as residential buildings and roads. In addition, it would be interesting to conduct a systematic study on pan-sharpening Landsat 8 OLI data based on the three types of image fusion approaches (e.g., CS, MRA and ATPRK) investigated in this paper.

5. Conclusion

In this paper, the six 20 m bands of Sentinel-2 were downscaled to 10 m with the aid of the four 10 m bands of the same satellite sensor. The existing CS- and MRA-based image fusion approaches and the recently developed ATPRK approach were all developed for this downscaling problem. To use the fine spatial resolution from the four 10 m bands, two schemes, the synthesized band scheme and the selected band scheme, were considered to obtain a single fine resolution band for matching with each 20 m multispectral image or 20 m band. The generic findings from the three case studies on downscaling Sentinel-2 images are summarized as follows.

1) For ATPRK or a MRA-based method, the synthesized band scheme is able to produce greater accuracies than the selected band scheme (Tables 3-5). For a CS-based method, however, the selected band scheme tends to be more accurate.
2) MRA methods (except ATWT-M3) are superior to CS methods. For the Malmo data, the CC values of MRA methods are above 0.99, but the CC values of CS methods are below 0.99.
3) Given the same scheme of band extraction, ATPRK can produce more accurate results than the CS and MRA approaches (Tables 3-5). That is, ATPRK with the synthesized band is the most accurate method amongst the three groups of approaches. Using the synthesized band, the UIQI of ATPRK is 0.032, 0.008, 0.005 and 0.010 larger than ATWT-M3, MTF-GLP, MTF-GLP-CBD and MTF-GLP-HPM for the Treviso data; the ERGAS of ATPRK is 1.5, 0.6, 0.6 and 0.5 smaller than ATWT-M3, MTF-GLP, MTF-GLP-CBD and MTF-GLP-HPM for the Malmo data.
4) ATPRK has the property of perfect prediction coherence (almost ideal CC, UIQI, ERGAS, and SAM are achieved in Tables 6-8).

Acknowledgment

This work was supported in part by the Research Grants Council of Hong Kong under Grant PolyU 15223015 and 5249/12E, in part by the National Natural Science Foundation of China under Grant 41331175, in part by the Leading talent Project of National Administration of Surveying under grant K.SZ.XX.VTQA, and in part by the Ministry of Science and Technology of China under Grant 2012BAJ15B04. The authors would also like to thank the handling editor and anonymous reviewers for their valuable comments and suggestions which greatly improved the work.

References

Alparone, L., Wald, L., Chanussot, J., Thomas, C., Gamba, P., & Bruce, L. M. (2007). "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data fusion contest," *IEEE Transactions on Geoscience and Remote Sensing*, 45, 3012–3021.

Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., & Selva, M. (2003). "An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas," in *Proceedings of 2nd GRSS/ISPRS Joint Workshop Remote Sensing Data Fusion over urban Areas*, 90–94.

Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., & Selva, M. (2006). MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogrammetric Engineering and Remote Sensing*, 72, 591–596.

Aiazzi, B., Baronti, S., & Selva, M. (2007). Improving component substitution pansharpening through multivariate regression of MS+Pan data. *IEEE Transactions on Geoscience and Remote Sensing*, 45, 3230–3239.

Aiazzi, B., Baronti, S., Lotti, F., & Selva, M. (2009). A comparison between global and context-adaptive pansharpening of multispectral images. *IEEE Geoscience and Remote Sensing Letters*, 6, 302-306.

Atkinson, P. M., Pardo-Igúzquiza, E., & Chica-Olmo, M. (2008). Downscaling cokriging for super-resolution mapping of continua in remotely sensed images. *IEEE Transactions on Geoscience and Remote Sensing*, 46, 573–580.

Atkinson, P. M. (2009). Issues of uncertainty in super-resolution mapping and their implications for the design of an inter-comparison study. *International Journal of Remote Sensing*, 30, 5293–5308.

Atkinson, P. M. (2013). Downscaling in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 22, 106–114.

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

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, 295–303.

Choi, J., Yu, K., & Kim, Y. (2011). A new adaptive component-substitution based satellite image fusion by using partial replacement. *IEEE Transactions on Geoscience and Remote Sensing*, 49, 295–309.

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

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

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, 2207–2218.

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

Gillespie, A. R., Kahle, A. B., & Walker, R. E. (1987). Color enhancement of highly correlated images—II. Channel ratio and "Chromaticity" trans-form techniques. *Remote Sensing of Environment*, 22, 343–365.

Hagolle, O. *et al*. (2015). SPOT-4 (Take 5): Simulation of Sentinel-2 time series on 45 large sites. *Remote Sensing*, 7, 12242–12264.

Hengl, T., Heuvelinkb, G. B. M., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120, 75–93.

Hengl, T., Heuvelinkb, G. B. M., & Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computers & Geosciences*, 33, 1301–1315.

Kerry, R., Goovaerts, P., Rawlins, B. G., & Marchant, B. P. (2012). Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma*, 170, 347–358.

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, 98–102.

Kyriakidis, P. C. (2004). A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, 36, 259–289.

Kyriakidis, P., & Yoo, E.-H. (2005). Geostatistical prediction and simulation of point values from areal data. *Geographical Analysis*, 37, 124–151.

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

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, 3461–3472.

Palsson, F., Sveinsson, J. R., Ulfarsson, M. O., & Benediktsson, J. A. (2016). Quantitative quality evaluation of pansharpened imagery: consistency versus synthesis. *IEEE Transactions on Geoscience and Remote Sensing*, 54, 1247–1259.

Pardo-Igúzquiza, E., Chica-Olmo, M., & Atkinson, P. M. (2006). Downscaling cokriging for image sharpening. *Remote Sensing of Environment*, 102, 86–98.

Pardo-Iguzquiza, E., Rodríguez-Galiano, V. F., Chica-Olmo, M., & Atkinson, P. M. (2011). Image fusion by spatially adaptive filtering using downscaling cokriging. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 337–346.

Ranchin, T., & Wald, L. (2000). Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation. *Photogrammetric Engineering and Remote Sensing*, 66, 49–61.

Sales, M. H. R., Souza, Jr., C. M., & Kyriakidis, P. C. (2013). Fusion of MODIS images using kriging with external drift. *IEEE Transactions on Geoscience and Remote Sensing*, 51, 2250–2259.

Segl, K., Guanter, L., Gascon, F., Kuester, T., Rogass, C., & Mielke, C. (2015). S2eteS: An end-to-end modeling tool for the simulation of Sentinel-2 image products. *IEEE Transactions on Geoscience and Remote Sensing*, 53, 5560–5571.

Selva, M., Aiazzi, B., Butera, F., Chiarantini, L., & Baronti, S. (2015). Hyper-sharpening: A first approach on SIM-GA data. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 8, 3008–3024.

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, 561–567.

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, 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.

Vivone, G., Restaino, R., Dalla Mura, M., Licciardi, G., & Chanussot, J. (2014). Contrast and error-based fusion schemes for multispectral image pan-sharpening. *IEEE Geoscience and Remote Sensing Letters*, 11, 930–934.

Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A., Restaino, R., & Wald, L. (2015). A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53, 2565–2586.

Wald, L., Ranchin, T., & Mangolini, M. (1997). Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing*, 63, 691–699.

Wang, Q., Shi, W., Atkinson, P. M., & Zhao, Y. (2015a). Downscaling MODIS images with area-to-point regression kriging. *Remote Sensing of Environment*, 166, 191–204.

Wang, Q., Shi, W., Atkinson, P. M., & Pardo-Iguzquiza, E. (2015b). A new geostatistical solution to remote sensing image downscaling. *IEEE Transactions on Geoscience and Remote Sensing*, 54, 386–396.

Wang, Z., & Bovik, A. C. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 9, 81–84.

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27, 3025–3033.

Zhang, H., & Roy, D. (2016). Computationally inexpensive Landsat 8 Operational Land Imager (OLI) pansharpening. *Remote Sensing*, 8, 180.

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, 2610–2623.

Fig. 1. The three studied Sentinel-2 images. (a)-(c) are 10 m data (2400 by 2400 pixels, bands 4, 3 and 2 as RGB) for Verona, Treviso and Malmo, respectively. (d)-(f) are 20 m data (1200 by 1200 pixels, bands 12, 8a and 5 as RGB) for Verona, Treviso and Malmo, respectively. (g)-(i) are sub-areas with 200 by 200 pixels in (a)-(c). (j)-(l) are three corresponding sub-areas with 100 by 100 pixels in (d)-(f).

Fig. 2. Downscaling results (20 m) for the sub-area in Verona (bands 12, 8a and 5 as RGB). (a) 40 m coarse image. (b) 20 m reference image. (c) ATPRK. (d) BTA. (e) GSA. (f) GSA-CA. (g) PRACS. (h) ATWT-M3. (i) MTF-GLP. (j) MTF-GLP-CBD. (k) MTF-GLP-HPM. For each method in (h)-(k), the result for the case with greatest accuracy is shown.

Fig. 3. Downscaling results (20 m) for the sub-area in Treviso (bands 12, 8a and 5 as RGB). (a) 40 m coarse image. (b) 20 m reference image. (c) ATPRK. (d) BTA. (e) GSA. (f) GSA-CA. (g) PRACS. (h) ATWT-M3. (i) MTF-GLP. (j) MTF-GLP-CBD. (k) MTF-GLP-HPM. For each method in (h)-(k), the result for the case with greatest accuracy is shown.

Fig. 4. Downscaling results (20 m) for the sub-area in Malmo (bands 12, 8a and 5 as RGB). (a) 40 m coarse image. (b) 20 m reference image. (c) ATPRK. (d) BTA. (e) GSA. (f) GSA-CA. (g) PRACS. (h) ATWT-M3. (i) MTF-GLP. (j) MTF-GLP-CBD. (k) MTF-GLP-HPM. For each method in (h)-(k), the result for the case with greatest accuracy is shown.

Fig. 5. Downscaling results (10 m) for three sub-areas in Verona (bands 12, 8a and 5 as RGB). (a)-(c) The 20 m data. (d)-(f) The 10 m results.

Fig. 6. Downscaling results (10 m) for three sub-areas in Treviso (bands 12, 8a and 5 as RGB). (a)-(c) The 20 m data. (d)-(f) The 10 m results.

Fig. 7. Downscaling results (10 m) for three sub-areas in Malmo (bands 12, 8a and 5 as RGB). (a)-(c) The 20 m data. (d)-(f) The 10 m results.