A comparison of gap-filling approaches for Landsat-7 satellite data

Gaohong Yin, Gregoire Mariethoz, Ying Sun & Matthew F. McCabe

[+] View supplementary material ⎋

📅 Published online: 10 Aug 2017.

✎ Submit your article to this journal ⎋

🔍 View related articles ⎋

CrossMark View Crossmark data ⎋

Taylor & Francis
Taylor & Francis Group

A comparison of gap-filling approaches for Landsat-7 satellite data

Gaohong Yin[a,b], Gregoire Mariethoz[c], Ying Sun[d] and Matthew F. McCabe[a]

[a]Water Desalination and Reuse Centre, Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; [b]Department of Civil and Environmental Engineering, University of Maryland, College Park, MD, USA; [c]Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland; [d]Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

ABSTRACT

The purpose of this study is to assess the relative performance of four different gap-filling approaches across a range of land-surface conditions, including both homogeneous and heterogeneous areas as well as in scenes with abrupt changes in landscape elements. The techniques considered in this study include: (1) Kriging and co-Kriging; (2) geostatistical neighbourhood similar pixel interpolator (GNSPI); (3) a weighted linear regression (WLR) algorithm; and (4) the direct sampling (DS) method. To examine the impact of image availability and the influence of temporal distance on the selection of input training data (i.e. time separating the training data from the gap-filled target image), input images acquired within the same season (temporally close) as well as in different seasons (temporally far) to the target image were examined, as was the case of using information only within the target image itself. Root mean square error (RMSE), mean spectral angle (MSA), and coefficient of determination (R^2) were used as the evaluation metrics to assess the prediction results. In addition, the overall accuracy (OA) and kappa coefficient (kappa) were used to assess a land-cover classification based on the gap-filled images. Results show that all of the gap-filling approaches provide satisfactory results for the homogeneous case, with $R^2 > 0.93$ for bands 1 and 2 in all cases and $R^2 > 0.80$ for bands 3 and 4 in most cases. For the heterogeneous example, GNSPI performs the best, with $R^2 > 0.85$ for all tested cases. WLR and GNSPI exhibit equivalent accuracy when a temporally close input image is used (i.e. WLR and GNSPI both have an R^2 equal to 0.89 for band 1). For the case of abrupt changes in scene elements or in the absence of ancillary data, the DS approach outperforms the other tested methods.

1. Introduction

The Landsat Program is a global land-imaging mission managed through a joint effort of the United States Geological Survey (USGS) and the National Aeronautics and Space Administration (NASA), with the objective of providing globally continuous, high-resolution,

---

CONTACT Gaohong Yin ✉ gaohong.yin@kaust.edu.sa ▦ Department of Civil and Environmental Engineering,0147E Engineering Laboratory Building, A. James Clark of Engineering School, University of Maryland, College Park, MD, 20742, USA

⊕ Supplemental data for this article can be accessed here.

multispectral data for scientific and educational use. As part of a four-decade-long continuity mission of launched satellites, Landsat data have been employed across a wide range of Earth science activities, including hydrological applications (Ershadi et al. 2013), vegetation dynamics (Houborg et al. 2015), land-cover assessments (Lo and Choi 2004), and climate change impacts (Bormann, McCabe, and Evans 2012). With its high geometric and radiometric accuracy, the satellite has been instrumental in deriving long-term records of land-surface variables, changes, and feature descriptions. Unfortunately, the Scan Line Corrector (SLC) on Landsat 7's Enhanced Thematic Mapper Plus (ETM+) sensor failed on 31 May 2003, affecting the continuity and fidelity of these globally important high-resolution satellite records (USGS 2003). The SLC is an electromechanical device that compensates for the forward motion of the satellite such that the whole image can be acquired in parallel scans without overlap. After the SLC stopped functioning, the scan track followed a zigzag pattern, causing overlap at specific locations and leaving gaps in the retrieved imagery (USGS 2003). The gaps in the central region of the image are near-contiguous and only one or two pixels wide. However, the gaps become larger at the edge of the imagery, reaching more than 10 pixels width (Romero-Sanchez et al. 2015). As a result, the utility of Landsat 7 data was significantly impacted due to the imposition of these spatial discontinuities.

Considering the extensive application of Landsat 7 data and its retained geometric and radiometric fidelity, a number of approaches have been proposed to fill these gaps (Chen et al. 2011; Maxwell, Schmidt, and Storey 2007; Pringle, Schmidt, and Muir 2009; Scaramuzza, Micijevic, and Chander 2004a; Zeng, Shen, and Zhang 2013; Zhu, Liu, and Chen 2012; Yin, Mariethoz, and McCabe 2017). From the perspective of using input data to inform the filling of gaps, the techniques that have been proposed can be categorized into three broad categories: (1) single-source approaches, using non-gap locations within the target images; (2) multi-source approaches, using Landsat 7 imagery acquired at different dates; and (3) multi-source approaches, using data from other remote sensors. To ensure consistency in terminology, the Landsat image with gaps to be filled is hereafter called the target image and the image used as ancillary data for gap filling is called the input image. Single-source approaches (also referred to as non-reference methods) usually exploit resampling algorithms such as nearest neighbourhood, bilinear, and cubic interpolation to fill the gaps (Zeng, Shen, and Zhang 2013). One non-reference method proposed by Zhang et al. (Zhang, Li, and Travis 2007) for gap filling is ordinary Kriging. Kriging is a geostatistical interpolation technique, with the fundamental idea being that pixels spatially close to gaps are more similar than those that are spatially distant (Zhang, Li, and Travis 2007). Therefore, when predicting gap pixels, closer neighbourhood pixels are given more weight in the filling process. To understand the spatial correlation between pixels, variogram models are adopted. Studies have shown that the ordinary Kriging approach is suitable when high levels of per-pixel accuracy may not be required (Zhang, Li, and Travis 2007). That is, the interpolation results are less accurate for small-scale applications, which demand detailed information. Generally, the ordinary Kriging technique is computationally complex with relatively slow processing speeds (Hossain et al. 2015).

Multi-source techniques that employ ancillary data from Landsat 7 imagery overlapping the same area (but acquired at different dates) are the most widely adopted strategies among the proposed methods. The local linear histogram matching (LLHM) technique (Scaramuzza, Micijevic, and Chander 2004b; Scaramuzza, Micijevic, and

Chander 2004a) was proposed soon after the SLC failure. The LLHM method consists of two steps: Phase 1, using one SLC-on Landsat 7 imagery acquired before 2003; and Phase 2, using multiple SLC-off scenes to predict the gap pixels in the target image. The idea behind Phases 1 and 2 is to find a linear transformation between the input images and the target image. The gain and bias in the linear function are simply calculated by the mean and standard deviation of the target and input images. The LLHM method is a simple and fast approach that performs satisfactorily when the input images are of high quality (i.e. the input image shows almost no spectral difference to the target image). However, one drawback is that obvious striping tends to appear at the interface areas of two different land covers. Additionally, the LLHM method yields poor results in the presence of clouds, snow, Sun glint, or abrupt land-cover change. Therefore, the implementation of the LLHM method has generally been limited to relatively homogeneous and consistent areas with accessible high-quality input images (Scaramuzza, Micijevic, and Chander 2004a; Scaramuzza, Micijevic, and Chander 2004b).

Considering the limitations of the LLHM method, Maxwell et al. (Maxwell, Schmidt, and Storey 2007) proposed a multi-scale segmentation approach to fill the gaps. The multi-scale segmentation approach divides one SLC-on image into individual landscape units with different scale parameters, with segmented maps then generated at each scale level. By co-registering the target image to the segmented maps from the smallest scale to the largest scale, the values of gap pixels are set to the same value as the nearest original valid pixels in the same unit. This multi-scale segmentation approach emphasizes the similarity of pixel values within the same unit. It allows for the filling of gaps with coincident spectral values, but on the other hand, can cause a smoothing effect, i.e. the loss of spectral heterogeneity. Therefore, the multi-scale segmentation approach cannot generally be implemented in applications requiring per-pixel accuracy. Pringle et al. (Pringle, Schmidt, and Muir 2009) proposed using the ordinary co-Kriging technique to predict missing pixels. The principle of co-Kriging is the same as that of Kriging, with the difference being that co-Kriging employs input images as secondary data for gap filling (Pardo-Iguzquiza, Atkinson, and Chica-Olmo 2010). The spatial covariance model of a target and an input image, and the cross-covariance model of the target and the input image are all needed for the missing pixel interpolation. It has been demonstrated that ordinary co-Kriging can perform well in various regions, but it requires an assumption of a linear relationship with the input image, and also a stationarity assumption (statistical homogeneity) of both input and target images. Moreover, similar to Kriging, it suffers from low accuracy at the pixel level and relatively slow computation speed (Boucher, Seto, and Journel 2006; Pringle, Schmidt, and Muir 2009).

In an attempt to improve the prediction accuracy, Chen et al. (2011) developed the Neighbourhood Similar Pixel Interpolator (NSPI). Instead of only considering spatial correlation, NSPI takes both spatial and temporal changes into consideration and has been shown to recover the missing pixels more accurately, especially for heterogeneous regions (Chen et al. 2011). The NSPI is a deterministic linear interpolator, which cannot provide statistical uncertainty for each prediction, potentially influencing the robustness of its application. To overcome this limitation, Zhu, Liu, and Chen (2012) improved the approach by integrating it with the theory of geostatistics. The improved geostatistical neighbourhood similar pixel interpolator (GNSPI) models the temporal trend by using a

deterministic approach and predicts the residuals using ordinary Kriging. The GNSPI method offers more accurate predictions than NSPI for heterogeneous areas, besides providing prediction uncertainty. Studies have shown that GNSPI predicts small objects more accurately and reduces the striping phenomenon compared with NSPI and Kriging (Zhu, Liu, and Chen 2012). The concept and implementation of GNSPI are considerably more complex than the other approaches, with slower computing speed than LLHM (Hossain et al. 2015). Another limitation of GNSPI is that it carries with it some properties that are inherent to the Kriging framework, such as considering the interpolated variable as stationary (often not the case for remote-sensing images), along with the fact that Kriging is a smooth interpolant, meaning that it cannot create sharp connected features such as roads or rivers (Journel and Zhang 2006; Romero-Sanchez et al. 2015).

Considering the range of limitations outlined above, Zeng et al. (Zeng, Shen, and Zhang 2013) proposed an integrated recovery method combining the multi-temporal approach and the non-reference method to fill the Landsat 7 gaps. In this technique, a weighted linear regression (WLR) algorithm is first implemented to fill the un-scanned pixels, and then a Laplacian prior regularization method (LPRM), a non-reference method, is used to fill the gap pixels that cannot be predicted by WLR. The integrated method has been demonstrated to predict missing pixels accurately, especially at the sharp edges, which helps retain ground object shapes. However, it may fail when the gaps are too large (Zeng, Shen, and Zhang 2013). The computational speed is relatively slow, although faster than GNSPI (Hossain et al. 2015).

The direct sampling (DS) method (Mariethoz and Renard 2010) is a multiple-point geostatistical approach (Mariethoz and Caers 2014; Tang et al. 2013) that was first proposed for reconstructing geological structures. The idea of the DS is conceptually simple, in that it attempts to find a pattern that is similar to the target pattern, with the pattern search being guided by the covariate information. It has previously been applied for filling gaps in satellite-based Earth observations (Mariethoz, McCabe, and Renard 2012) as well as for downscaling applications (Jha et al. 2015; Jha et al. 2013; Tang, Atkinson, and Zhang 2015). In a recent study, Yin et al. (Yin, Mariethoz, and McCabe 2017) used the DS method to fill Landsat 7 data gaps across a number of different landscapes with varying levels of heterogeneity.

Despite the varying prediction accuracies provided by many of these different methods, applications to date have tended to avoid gap-filling assessment for cases exhibiting abrupt changes in the underlying scene. However, abrupt scene changes caused by deforestation, bushfires, tsunamis, or other natural or anthropogenic response are not rare in land-cover change detection application (Wulder et al. 2008). The profile-based interpolator (PBI) *k*-nearest-neighbour method was proposed as a new technique for filling non-stationary data, and it has been demonstrated to predict abrupt changes with good accuracy (Malambo and Heatwole 2016). Furthermore, it does not require cloud-free scenes and can predict the missing values of any date at any location, if a long enough sequence of images is available. However, the method can recreate some unwanted data, such as cloud shadows in the image (Malambo and Heatwole 2016).

In the present study, we focus on an inter-comparison of four approaches that have demonstrated their utility for gap filling. They are (1) Kriging and co-Kriging (Pringle, Schmidt, and Muir 2009; Zhang, Li, and Travis 2007); (2) geostatistical neighbourhood similar pixel interpolator (GSNPI) (Zhu, Liu, and Chen 2012); (3) the WLR algorithm

integrated with LPRM (Zeng, Shen, and Zhang 2013); and (4) the DS method (Mariethoz and Renard 2010; Yin, Mariethoz, and McCabe 2017). For brevity, the WLR integrated with LPRM is referred to as the WLR algorithm. In evaluating these approaches, regions comprising homogeneous areas, heterogeneous areas as well as regions with abrupt scene changes are assessed and inter-compared. In real gap-filling applications, it is not uncommon that temporally close input images are unavailable, or even that there is no covariate information. Therefore, cases of using no input image, using a temporally far input image, and using a temporally close input image are also examined to compare the reliance of the tested approaches on the input imagery.

2. Materials and methods

2.1. Study area and data

To thoroughly evaluate the gap-filling approaches, three regions with starkly different features are examined, comprising a homogeneous area (R1), a heterogeneous area (R2), and an abrupt change area (R3) resulting from the occurrence of a tsunami. R1 (Path 38, Row 38) is located in the Sonoran Desert in Northwest Mexico, east of the Gulf of California, around 31.89°N, 114.33°W. R2 (Path 43, Row 34) is located in central California, USA, around 37.97°N, 121.51°W, and contains crops, open water, and urban features. Both these study sites were chosen to maintain consistency with a previous gap-filling investigation (Yin, Mariethoz, and McCabe 2017). R3 is located on the west coast of Aceh, Indonesia (Path 131, Row 57). The area is selected because significant abrupt changes occurred in the region as a result of the Indian Ocean earthquake and the subsequent tsunami on 26 December 2004. A 400 × 400 pixel (12 km × 12 km) focus area is selected for each region. To evaluate any multispectral information content, Landsat bands 1–4 are used in the study. A description of the study regions is provided in Table 1.

The target images of R1 and R2 were acquired on 19 June 2002 and 24 July 2002, respectively. To assess the gap-filling results, gaps in the target images were imposed manually by using a gap mask acquired on 20 March 2010, since these images occur prior to the SLC failure. Doing this allows for a thorough statistical assessment of the approaches to be undertaken, as the original (and SLC unaffected) imagery can be used as a reference. Two input images, which overlap the same area as the target image but

Table 1. Description of the study regions used across the homogeneous (R1), heterogeneous (R2), and abrupt change (R3) case studies.

Region	Path/Row	Area (km × km)	Location	Land-cover feature
R1	38/38	12 km × 12 km	Sonoran Desert in Northwest Mexico, east of Gulf of California around 31.89°N, 114.33°W	Homogeneous landscape with surface covered by sand
R2	43/34	12 km × 12 km	Central California, USA around 37.97°N, 121.51°W	Heterogeneous landscape including vegetation cover, bare soil, open water, and impervious surface
R3	131/57	12 km × 12 km	West coast of Aceh, Indonesia, around 4.83°N, 95.45°E	Abrupt scene changes caused by the Indian Ocean earthquake and the subsequent tsunami on 26 December 2004

at different dates, are used for gap filling. These include T1 (3 June 2002 for R1 and 8 July 2002 for R2), which is two weeks temporally distant, and T2 (11 February 2002 for R1 and 2 March 2002 for R2), which is more than four months distant from the target image. Considering that T1 is temporally close to the target image and T2 is acquired from a completely different season, it can be expected that the land-cover conditions at T1 are more similar to the target image than T2. This is shown in Figure S1, where the reflectance values of band 4 for T1 and T2 are compared to the target image. For both R1 (homogeneous) and R2 (heterogeneous), input images at T1 provide a higher correlation coefficient. It is therefore assumed that by using a temporally close input image, T1 can generally provide more accurate results than T2.

True colour images of R1 and R2 are displayed in Figure 1. To analyse the spatial features of the studied regions, the reflectance histograms and statistical indices of R1 and R2 across the four examined bands are shown in Figure 2. From the provided statistical indices, it is clear that the relatively heterogeneous area (R2) provides larger standard deviation and skewness values overall. The skewness is also reflected in the histograms of bands 1–3 of R2, with the values of band 4 spread over a large range. On the other hand, the standard deviation and skewness of the homogeneous area (R1) are naturally smaller. Two obvious peaks can be observed from the histograms of bands 1 and 2 of R1, as the lower segment of R1 has relatively smaller reflectance values than the upper region.

A target image demonstrating an abrupt change in surface features (R3) was acquired for 14 January 2005, with gaps reflecting those actually observed in the imagery. A reference image acquired on 22 January 2005 by Landsat 5 was used for qualitatively assessing the gap-filled results. For R3, the input image was acquired on 13 December
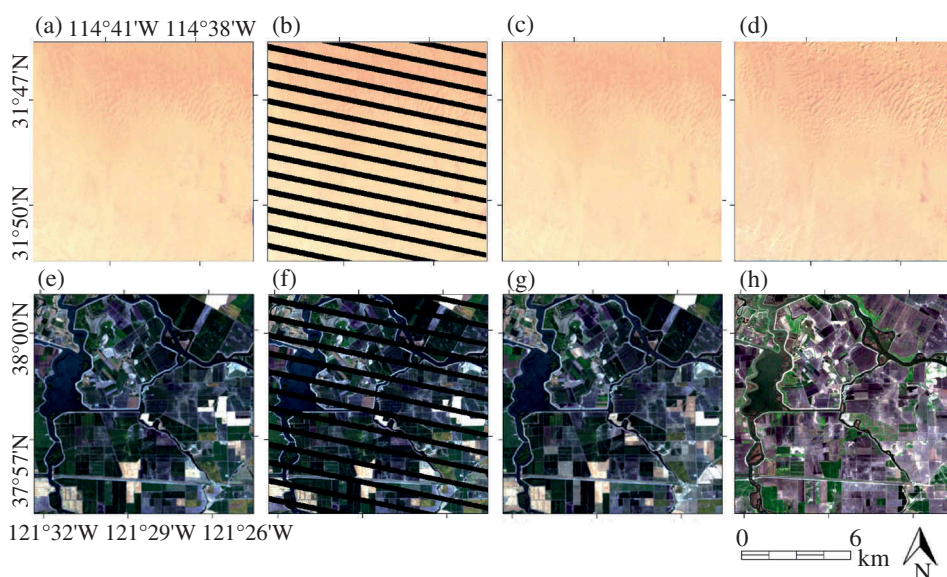

Figure 1. Landsat 7 images used for gap filling, with (*a–d*) R1 representing the homogeneous case, and (*e–h*) R2 representing the heterogeneous case. (*a,e*) are the unaltered images acquired on 19 June 2002 and 24 July 2002, respectively; (*b,f*) are the target images acquired by imposing an artificial gap mask derived from the 20 March 2010 imagery; (*c,g*) are the T1 input images (temporally close) acquired on 3 June 2002 and 8 July 2002, respectively, and (*d,h*) are the T2 input images (temporally far) acquired on 11 February 2002 and 2 March 2002, respectively.
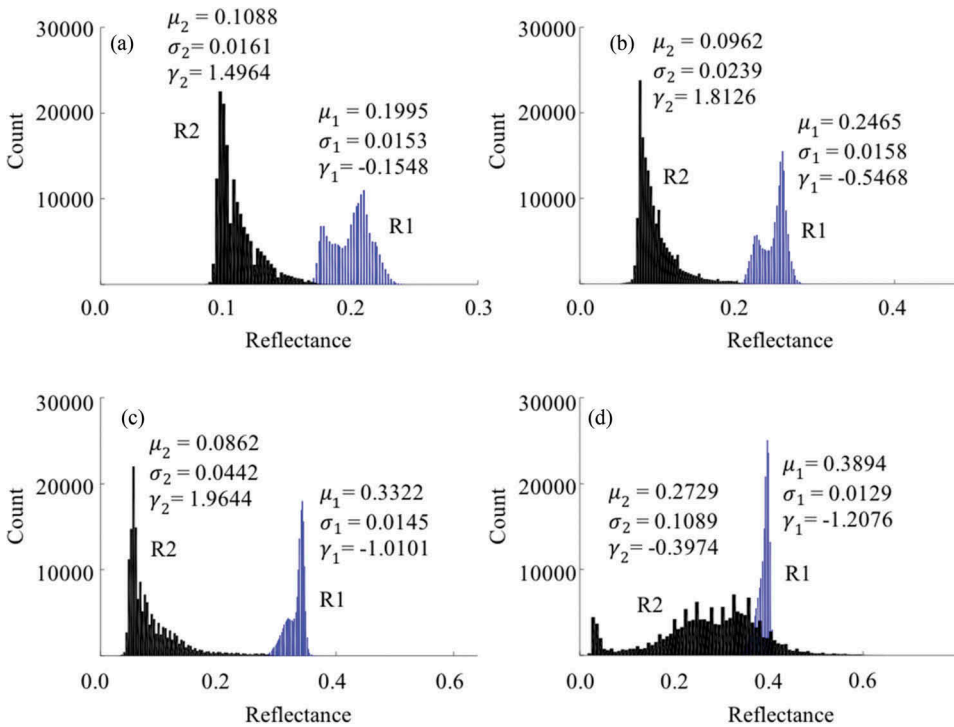
Figure 2. Reflectance histograms and statistical indices of the actual images for R1 (homogeneous area) and R2 (heterogeneous area), with (a–d) representing bands 1–4, respectively. 'μ', 'σ', and 'γ' represent, respectively, the mean, standard deviation, and skewness of data with the subscripts '1' and '2' corresponding to R1 and R2.

2004, some weeks before the tsunami. As a result, the landscape features of the input image and the target image present stark differences. A Landsat 7 image acquired on 19 May 2004 is selected as the time series image for the requirement of the GNSPI approach. True colour images of R3 are displayed in Figure 3, with details on all of the Landsat imageries utilized herein listed in Table 2.

2.2. Gap-filling methods

2.2.1. Kriging and co-Kriging

Kriging and co-Kriging are two common geostatistical techniques used for point interpolation (Chiles and Delfiner 1999; Pringle, Schmidt, and Muir 2009; Zhang, Li, and Travis 2007). The ordinary Kriging estimator is written as follows (Equation (1)):

$$Z^*(x_0, t_1, v) = \sum_{i=1}^{n} \lambda_i Z(x_i, t_1, v), \tag{1}$$

where λ_i is the ordinary Kriging weight, n is the number of neighbourhood pixels $,\sum_{i=1}^{n} \lambda_i = 1$, $Z(x_i, t_1, v)$ is the value of i-th neighbourhood pixel at location x_i in band v for the target image acquired at t_1, and $Z^*(x_0, t_1, v)$ is the predicted value of the
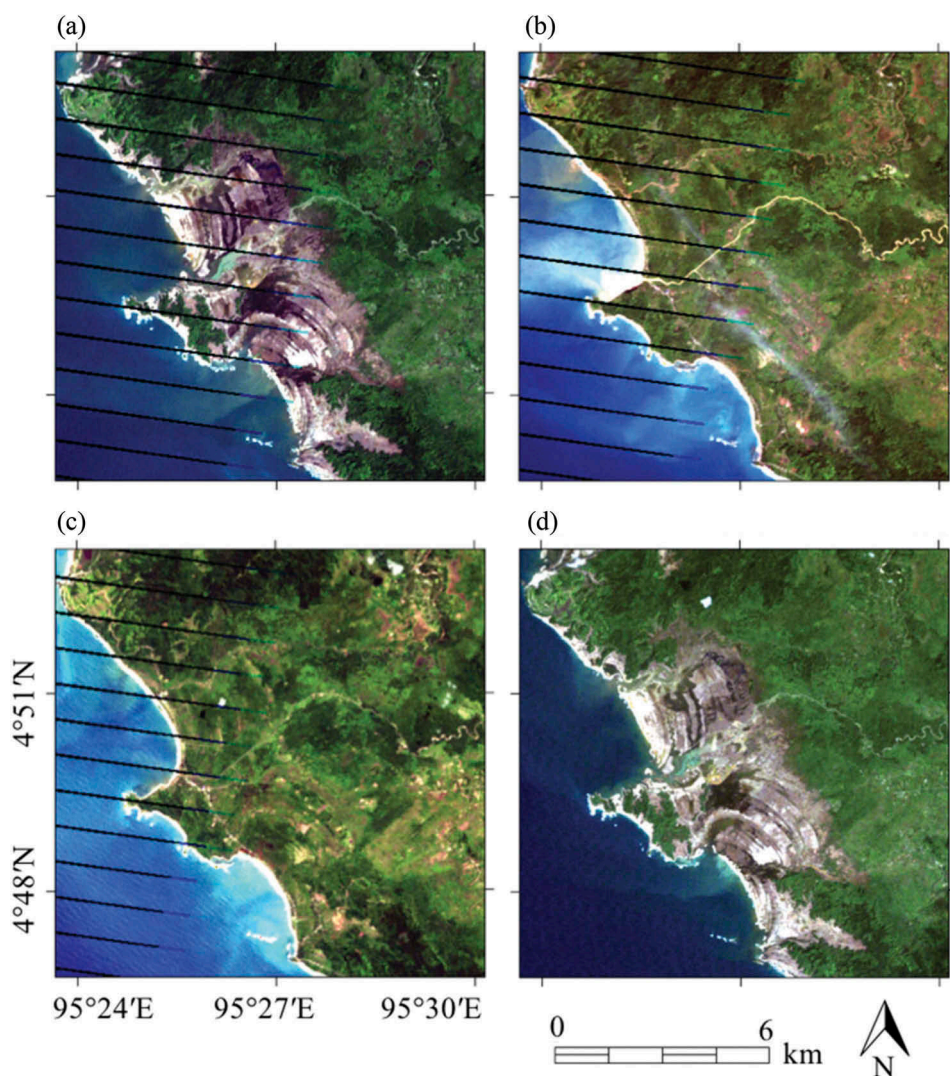
Figure 3. Landsat imagery used for examining the impact of abrupt changes (R3), with (a) the Landsat 7 ETM+ target image acquired on 14 January 2005; (b) the ETM+ input image acquired on 13 December 2004; (c) the ETM+ time series image acquired on 19 May 2004; and (d) the Landsat 5 TM reference image acquired on 22 January 2005.

target pixel at location x_0 in band v at t_1. The ordinary Kriging estimator is unbiased and gives minimized estimation variance. The weights can be calculated from the spatial covariance of data values.

The principle for ordinary co-Kriging is similar to that of ordinary Kriging, but secondary information (i.e. the input image acquired at t_2) is incorporated to predict the target pixels. The weights assigned to the primary (i.e. the reflectance of the target image) and secondary (i.e. the reflectance of the input image) variables can be calculated from the spatial covariance models of the primary and secondary attributes and the cross-covariance model of the primary and secondary attributes.

Table 2. List of Landsat imageries utilized in this study.

Region	Purpose	Date	Sensor
R1	Target image	19 June 2002	ETM SLC-on
	Input image (T1)	3 June 2002	ETM SLC-on
	Input image (T2)	11 February 2002	ETM SLC-on
	Gap mask	20 March 2010	ETM SLC-off
R2	Target image	24 July 2002	ETM SLC-on
	Input image (T1)	8 July 2002	ETM SLC-on
	Input image (T2)	2 March 2002	ETM SLC-on
	Gap mask	20 March 2010	ETM SLC-off
R3	Target image	14 January 2005	ETM SLC-off
	Input image	13 December 2004	ETM SLC-off
	Time series image	19 May 2004	ETM SLC-off
	Reference image	22 January 2005	TM

2.2.2. *Geostatistical neighbourhood similar pixel interpolator*

GNSPI is a geostatistical approach integrating deterministic interpolation and geostatistical principles (Zhu, Liu, and Chen 2012). The technique uses physical or empirical models to first remove the trend of data and then adopts ordinary Kriging to predict the residual. Apart from the target image and one input image, at least one time series image is recommended for the neighbourhood pixel selection. Application of the GNSPI as implemented herein follows the procedure outlined in Zhu, Liu, and Chen (2012). The GNSPI code is written in Interactive Data Language (IDL), and can be accessed online (Zhu 2013). The parameters used in this study are assigned as follows: sample size of the sample pixel is 20, the maximum window size is 12, the estimated number of classes is 4, the number of images in the time series excluding the input image is 1 (using zero is also discussed in Section 3.4), the range of minimum and maximum values of the image is [0, 1], and the block size is 400.

2.2.3. *Weighted linear regression*

The WLR-based multi-temporal recovery method (Zeng, Shen, and Zhang 2013) assumes a linear relationship between multi-temporal images covering the same area, with the linear relationship calculated from locally similar pixels:

$$Z(x_i, t_1, v) = aZ(x_i, t_2, v) + b, \tag{2}$$

where a and b are the regression coefficients. An adaptive determination procedure is proposed to select these similar pixels (Zeng, Shen, and Zhang 2013). With the selected locally similar pixels, the regression coefficients a and b can be calculated using the weighted least-squares method, and thus the pixels in the target image can be predicted using Equation (2). When the gap widths are large, the WLR method may not be able to fill all missing pixels, and the LPRM, a non-reference recovery method, is adopted to fill the remaining gap pixels. The WLR software is available online (SENDIMAGE 2016) and the parameters used in the study are specified as follows: the number of similar pixels is 30, the regularization parameter is 0.01, and the maximum number of iteration is 600.

2.2.4. *DS method*

The DS method (Mariethoz, Renard, and Straubhaar 2010) is a multiple-point geostatistical method that fills gaps by directly sampling patterns either from an input image or

from the known parts of the target image (in this case no input image is required). In the case of no input image, the already-known elements of the target image serve as the input image to provide information for gap filling. The main concept is to find one pixel y in the input image whose neighbourhood N_y is similar to the neighbourhood N_x of the pixel x in the target image. The criterion used to assess the similarity between N_x and N_y is the distance $d(N_x, N_y)$. In the instance of finding a pixel y in the input image whose neighbourhood distance is lower than a given threshold, the reflectance value of y is assigned to pixel x in the target image. In this study, a Euclidian distance is used for univariate simulation:

$$d(N_x, N_y) = \frac{1}{\eta} \sqrt{\sum_{i=1}^{n} [Z(x_i, t_1, v) - Z(y_i, t_2, v)]^2} \in [0, 1],$$ (3)

where η is a normalization factor ensuring the distance is constrained in the interval [0,1]. When multivariate simulations are performed, the distance should be able to evaluate the similarity between multivariate neighbourhoods. Therefore, a weighted Euclidian distance can be used (Mariethoz, Renard, and Straubhaar 2010). In this study, both univariate and bivariate simulations of the DS method are conducted. For the univariate case, no input image is used and the variable considered in the simulation is the reflectance of the target image. When bivariate simulation is performed, the reflectance value of the observed portion of the target image is considered as Variable 1, and the reflectance value of another input image is regarded as Variable 2. Variable 1 and Variable 2 are given the same weight. A simplified version of DS written in Matlab is available online (Mariethoz 2010), but the simulations conducted here are based on a C++ implementation, which is available on request for research purposes. The parameters used in this study are as follows: number of realizations is 10, maximum number of neighbours is 20, threshold is 0.01, and the maximum fraction of the input image to scan is 0.7.

2.3. Experimental setting

The experimental setting for R1 (homogeneous area) and R2 (heterogeneous area) includes three cases without the use of input images ('Kriging-T0', 'WLR-T0', and 'DS-T0'), four cases using the T1 (temporally close) input imagery ('Co-Kriging-T1', 'GNSPI-T1', 'WLR-T1', and 'DS-T1'), and four cases using the T2 (temporally far) input imagery ('Co-Kriging-T2', 'GNSPI-T2', 'WLR-T2', and 'DS-T2'). For the abrupt changes case (R3), only one input image is used. As such, the studies comprise three cases without using this input image and four cases using the input image, with case names annotated in the same way as for the R1 and R2 simulations. As noted in the application of the GNSPI approach (Section 2.2.2), it is recommended to use at least one time series image acquired from a season that is different from the target image, for neighbourhood pixels selection. Every time one input image is used, the other functions as the time series image. That is, when T1 is used as the input image, T2 then serves as the time series image, and vice versa.

The gap-filling results are assessed from both qualitative and quantitative perspectives for R1 and R2. For R3, as there is no gap-unaffected image and just a reference image is used, the results are evaluated solely from a qualitative perspective (i.e. visual assessment of the filled images without using statistical metrics). Root mean square error

(RMSE), mean spectral angle (MSA), and coefficient of determination (R^2) are used as statistical indices to evaluate the filled results, with their expressions shown in Equation (4)–(6).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Z^*(x_i, t, v) - Z(x_i, t, v)\right)^2}, \tag{4}$$

$$MSA = \frac{180}{\pi}\frac{\left(\sum_{i=1}^{N} \arccos\left(\sum_{v=1}^{M} Z^*(x_i, t, v)Z(x_i, t, v)/\sqrt{\sum_{v=1}^{M} Z^{*2}(x_i, t, v)\sum_{v=1}^{M} Z^2(x_i, t, v)}\right)\right)}{N}, \tag{5}$$

$$R^2 = \left(\frac{\sum_{i=1}^{N}\left(Z^*(x_i, t, v) - \overline{Z^*}(x_i, t, v)\right)\left(Z(x_i, t, v) - \overline{Z}(x_i, t, v)\right)}{\sqrt{\sum_{i=1}^{N}\left(Z^*(x_i, t, v) - \overline{Z^*}(x, t, v)\right)^2 \sum_{i=1}^{N}\left(Z(x_i, t, v) - \overline{Z}(x_i, t, v)\right)^2}}\right), \tag{6}$$

where N is the number of pixels located in the gaps, M is the number of examined bands ($M = 4$ here), and $\overline{Z^*}(x, t, v)$ and $\bar{Z}(x, t, v)$ are the mean of the predicted values and actual values, respectively. Smaller RMSE and MSA values generally represent a more accurate prediction, with RMSE assessing the prediction accuracy of a single band and MSA the joint accuracy provided by all of the examined bands. A higher R^2 represents a higher consistency between the predicted values and the actual values.

In addition to the above-mentioned metrics, we validate the gap-filling results by classifying them and then comparing with a classification of the reference image. Two statistical indices, which are commonly used for evaluating land-cover classification accuracy, are adopted as reference metrics for assessing the prediction accuracy from a large-scale perspective. They are the overall accuracy (OA) and the kappa coefficient (*kappa*) (Congalton 1991). After acquiring the gap-filling results, only pixels located in the gaps of the actual image and the predicted results are classified using a Support Vector Machine (SVM) supervised classifier. The classification results of the actual image are then used as reference data to evaluate whether the predicted results can provide similar classification, although the classification of the actual image is not exactly the ground truth (Zhu, Liu, and Chen 2012). The OA is the proportion of pixels that are correctly classified, whereas *kappa* measures the agreement between the classification and ground truth pixels. Higher OA and *kappa* represent more accurate results from the overall prediction perspective. Owing to the homogeneous land-surface features of R1, the OA and *kappa* indices are only applied to the results of R2 (heterogeneous area). For more details on these two indices, the reader is referred to Congalton (1991).

3. Results and discussion

The gap-filling results for R1 (homogeneous area), R2 (heterogeneous area), and R3 (abrupt changes) are discussed separately in sections 3.1, 3.2, and 3.3, respectively. For each region, the results simulated by different approaches are compared and the influence of temporal distance between the target and the input imagery on the prediction accuracy using different methods is examined. In section 3.4, the impact of

using seasonal time series imagery for GNSPI, as well as the number of input images used in DS, is briefly discussed.

3.1. *Application to a homogeneous area*

The realizations for R1 across all simulation cases, together with a randomly selected focus sub-area, are presented in Figure 4. From a visual perspective, all simulation cases fill the gaps adequately, although a closer inspection reveals subtle differences. For instance, in the zoomed actual image, an acutely angled feature (bounded with red lines) can be observed due to reflectance differences. The angled shape cannot be observed in any of the T0 simulations (i.e. no input image used). However, when an input image is used for filling gaps, the feature can be reconstructed in all of the examined approaches. Differences between the realizations generated using T1 (temporally close) and T2 (temporally far) input images cannot be distinguished from a qualitative perspective.

Given the inherent subjectivity of visually comparing the results, quantitative assessment metrics are introduced for further assessment. The error histograms of band 4 are displayed in Figure 5, with the mean and standard deviation of error values presented for each simulation case. The reflectance error is calculated by subtracting the actual values from the predicted values for each missing pixel. From Figure 5, all simulation cases produce symmetric error distributions around zero, suggesting that the simulations are unbiased. The standard deviation values using the T1 (temporally close) input image are the smallest for all approaches, indicated by a narrower distribution of errors. The T0 cases generally provide the largest standard deviation of errors. Comparing the results given by the different approaches, Kriging, WLR, and DS give similar error distributions when no input image is used (i.e. T0 cases). When using the T1 (temporally close) input image, the error distributions of GNSPI, WLR, and DS are all close in terms of standard deviation, with WLR slightly smaller. When a temporally far (T2) input image is used, the error distributions of GNSPI and co-Kriging are the narrowest and the broadest, respectively.

Combining the error histograms with the RMSE and MSA values listed in Table 3, when no input image is used (T0 cases), the prediction results of Kriging, DS, and WLR appear to be at the same accuracy level, whereas the result provided by WLR is slightly worse. When the T1 (temporally close) input image is used, the performances of GNSPI, WLR and DS are similar, with WLR providing slightly smaller RMSE and MSA values. The co-Kriging approach produces smaller RMSE values for band 1 and band 2, whereas the values become larger for bands 3 and 4. When the temporally far (T2) input image is used, the results given by GNSPI, WLR, and DS do not vary significantly, with DS presenting the smallest RMSE values for bands 1–2, GNSPI the smallest for bands 3–4, and WLR the smallest MSA value. Co-Kriging again provides small RMSE values for bands 1–2 and large values for bands 3–4. An important point to note is that although the values differ amongst the studied approaches, the differences are not large.

The R^2 values of all four tested bands (Table 4), together with the scatter plots of band 4 (Figure S2), are used for further evaluation. The level of consistency between the predicted and actual values provided by the scatter plots and R^2 is coincident with the prediction accuracy reflected by the error histograms and RMSEs. For the non-reference
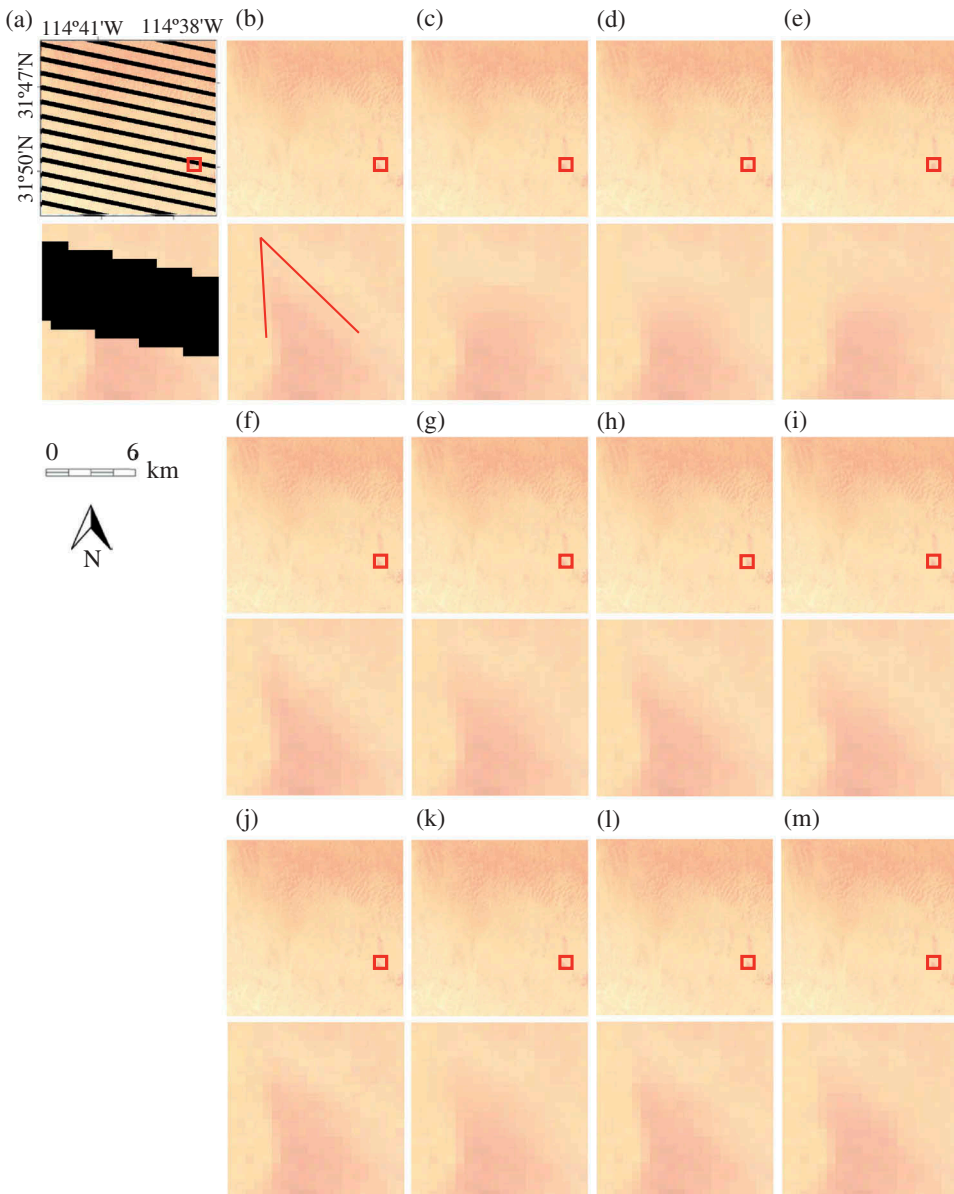
Figure 4. Realizations of R1 (homogeneous area). (*a*) is the target image with gaps and (*b*) is the actual image used to validate the prediction accuracy of the gap-filling results. (*c–e*) represent the non-reference simulation (T0) using Kriging, WLR, and DS, respectively. (*f–i*) are the realizations of GNSPI, co-Kriging, WLR, and DS, respectively, with temporally close input images (T1) implemented. (*j–m*) are the corresponding realizations of the GNSPI, co-Kriging, WLR, and DS using temporally far input images (T2). The area marked with a red square is magnified to the lower row of each pair.

cases (T0), the scatter is wider than for the cases using input images (T1 and T2) for WLR and DS. The R^2 values provided by Kriging, WLR, and DS are close for all four of the examined bands. When the T1 (temporally close) input image is used, co-Kriging gives
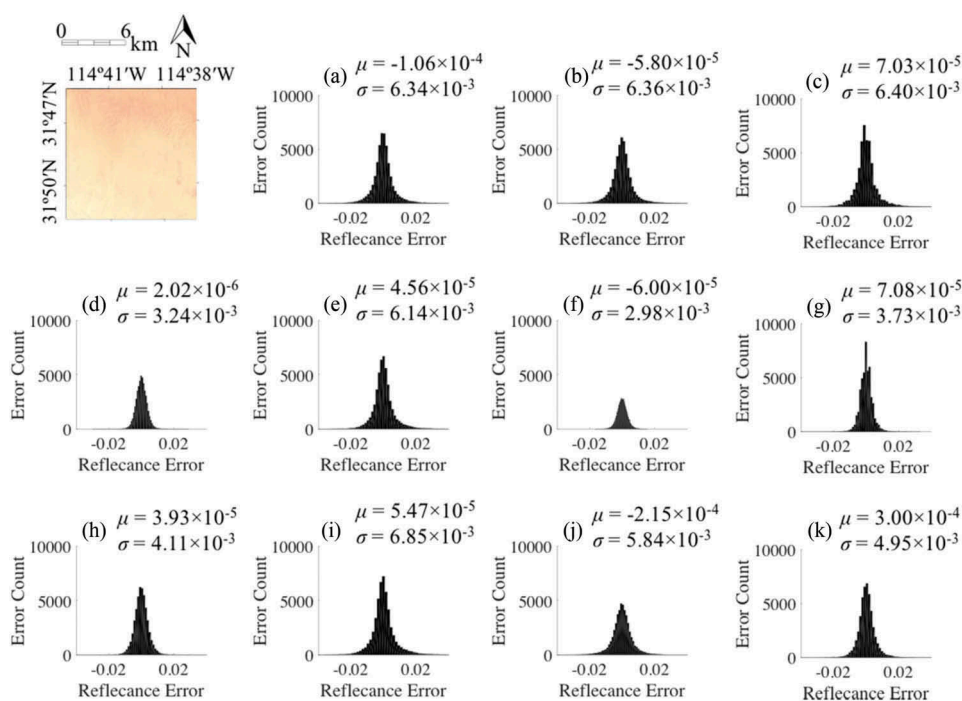
Figure 5. Reflectance error histograms of band 4 produced by each of the studied approaches under different input image conditions for R1 (homogeneous area), with the x-axis representing the reflectance error and the y-axis representing the error count. (a–c) correspond to the cases 'Kriging-T0', 'WLR-T0', and 'DS-T0'; (d–g) correspond to the cases 'GNSPI-T1', 'Cokriging-T1', 'WLR-T1', and 'DS-T1' using temporally close input images (T1); (h–k) correspond to cases 'GNSPI-T2', 'Cokriging-T2', 'WLR-T2', and 'DS-T2' using temporally far input images (T2). μ and σ represent the mean and standard deviation of errors, respectively.

high R^2 for bands 1–2, but relatively low consistency for bands 3–4. GNSPI, WLR, and DS all provide good agreement and high R^2, with WLR giving a slightly higher level of agreement across all four bands. When the T2 (temporally far) input image is used, the agreement provided by GNSPI is higher for bands 3–4, with DS giving a higher R^2 for bands 1–2. The results provided by WLR also show high consistency, with the R^2 values for bands 3–4 relatively lower than GNSPI and DS, but still higher than co-Kriging. Comparing the R^2 provided by GNSPI and WLR in Table 4 with the study of Romero-Sanchez et al. (2015), the R^2 provided in the study is obviously higher because of the homogeneous land-cover features. The R^2 values provided by Kriging and co-Kriging in Table 4 are slightly lower than the results provided by Pringle, Schmidt, and Muir (2009), but the difference does not vary too much. The application to the homogeneous landscape should have produced higher prediction accuracy; however, less ancillary data is implemented in the study compared with the study by Pringle, Schmidt, and Muir (2009). Therefore, the influences of the two factors are compensated and the R^2 values provided by the two studies are consistent.

From these homogenous case results, it is clear that all algorithms provide sufficiently accurate results to be suitable gap-filling approaches. The input image characteristics

Table 3. Root mean square error (RMSE) of R1 (homogeneous) gap-filling results simulated by Kriging, co-Kriging, GNSPI, DS, and WLR under different input image conditions for all four bands. The mean spectral angle (MSA) for each simulation case is provided with a unit of degree (°), which represents the overall prediction accuracy of the four bands. GNSPI* represents the case without using a time series image, as discussed in Section 3.4.

Test Case	Approach	RMSE				MSA (°)
		B1	B2	B3	B4	
T0	Kriging	0.0026	0.0037	0.0058	0.0063	0.3383
	WLR	0.0030	0.0039	0.0059	0.0064	0.4061
	DS	0.0027	0.0038	0.0058	0.0064	0.3790
T1	Co-Kriging	0.0025	0.0026	0.0051	0.0061	0.4016
	GNSPI	0.0027	0.0026	0.0033	0.0032	0.4010
	GNSPI*	0.0027	0.0026	0.0033	0.0032	0.3977
	WLR	0.0023	0.0023	0.0030	0.0030	0.3421
	DS	0.0023	0.0026	0.0036	0.0037	0.3480
T2	Co-Kriging	0.0034	0.0035	0.0060	0.0069	0.4833
	GNSPI	0.0037	0.0037	0.0041	0.0041	0.4658
	GNSPI*	0.0036	0.0040	0.0050	0.0054	0.4482
	WLR	0.0027	0.0034	0.0051	0.0058	0.3714
	DS	0.0025	0.0031	0.0045	0.0050	0.3724

Table 4. R^2 of R1 (homogeneous area) gap-filling results simulated by Kriging, co-Kriging, GNSPI, DS, and WLR under different input image conditions for all four bands. GNSPI* represents the case without using a time series image of the GNSPI approach, as discussed in Section 3.4.

Test Case	Approach	R^2			
		B1	B2	B3	B4
T0	Kriging	0.9704	0.9445	0.8386	0.7523
	WLR	0.9623	0.9390	0.8317	0.7525
	DS	0.9690	0.9412	0.8368	0.7471
T1	Co-Kriging	0.9744	0.9726	0.8735	0.7673
	GNSPI	0.9700	0.9719	0.9459	0.9356
	GNSPI*	0.9698	0.9715	0.9460	0.9357
	WLR	0.9784	0.9776	0.9548	0.9454
	DS	0.9769	0.9725	0.9379	0.9149
T2	Co-Kriging	0.9519	0.9498	0.8241	0.7106
	GNSPI	0.9427	0.9440	0.9173	0.8960
	GNSPI*	0.9458	0.9359	0.8808	0.8220
	WLR	0.9680	0.9519	0.8769	0.8011
	DS	0.9730	0.9613	0.9022	0.8490

have an evident influence on the prediction accuracy for all algorithms, of which temporally closer input images generally provide a more accurate gap-filled image. This result is anticipated, because temporally close images usually have similar land-cover characteristics to the target image, and thus also have higher spectral similarity than temporally far images. The performances of Kriging, DS, and WLR in filling gaps without using input images are similar and are all improved when an input image is available in the T1 and T2 cases. The reduced accuracy in the T0 cases can be explained by the fact that some spatial patterns absent in the non-gap locations of the target image can be found in the input images. When an input image is used, the accuracies provided by GNSPI, WLR, and DS are at the same level, although with slight differences. That is, when using a temporally close (T1) input image, WLR offers the most satisfactory result, and when a temporally far (T2) input image is used, GNSPI performs better for bands 3–4, whereas DS presents better performance for bands 1–2. The findings indicate

that WLR prediction accuracy is more sensitive to the spectral similarity between the input and target images, as the linear relationship assumption made by WLR has a greater chance to be invalid. Co-Kriging usually provides good prediction for bands 1–2, but is clearly less accurate for bands 3–4. Given the band characteristics of Landsat 7, we can conclude that DS performs better in monitoring man-made features, whereas GNSPI is better for vegetation monitoring, at least for the analysis undertaken here. Further investigations on these land-cover-specific responses will need to be undertaken to verify the generality of this conclusion.

3.2. Application to a heterogeneous area

To explore the gap-filling performance of the studied methods, the same approaches described above were repeated over a heterogeneous landscape with relatively large gaps. The selected scene contains several small patches of farmland, along with narrow and meandering roads. The range of gap-filling realizations that was explored, along with the corresponding magnified sub-areas, is presented in Figure 6. Analysing the inter-comparison results, it is clear that when no input image is used (T0), the continuity of the road features (bright colour in the magnified images of Figure 6) is not well preserved for any of the tested methods (i.e. Kriging, WLR, and DS). Examining these magnified images shows that DS provides a more visually consistent result compared with Kriging and WLR, as the general contour of the road is maintained. For the case of using a temporally close input image (T1), the filled images appear to be visually consistent from the large-scale perspective for all four algorithms. At the magnified scale however, it is clear that the narrow road contour produced by co-Kriging appears truncated, with the GNSPI, WLR, and DS simulations providing a high-quality prediction at the per-pixel level. In considering a temporally far input image (T2), co-Kriging is still unable to reconstruct the continuous features of the road. The other approaches provide satisfactory results, with the DS technique providing the best preservation of road features. GNSPI and WLR present some truncation in the road features (marked with a red circle in Figure 6), and due to the truncation, different object shapes are generated, as shown in the left circle marked in GNSPI.

To examine the simulation results more closely, the same quantitative metrics implemented in R1 are used, with the error histograms displayed in Figure 7 and the RMSE and MSA values listed in Table 5. For the case of no input image (T0), the Kriging error histogram shows a larger error range and standard deviation of errors than either DS or WLR. The prediction accuracy of DS and WLR is comparable, although DS has a smaller RMSE value for all four bands as well as a smaller MSA. When the temporally close input image (T1) is used, the simulation results provided by GNSPI, WLR, and DS are found to have similar accuracy, with GNSPI offering the smallest RMSE values. DS provides satisfactory results, although with an error histogram less tightly bound and the RMSEs slightly larger than those of GNSPI and WLR. The performance of co-Kriging is the poorest amongst all of the approaches. When the temporally far (T2) input image is used, co-Kriging again gives the poorest result, with a dispersive error distribution and the largest RMSE. The gap-filling results generated by GNSPI are the most accurate among the four methods, whereas the performances of WLR and DS are of similar
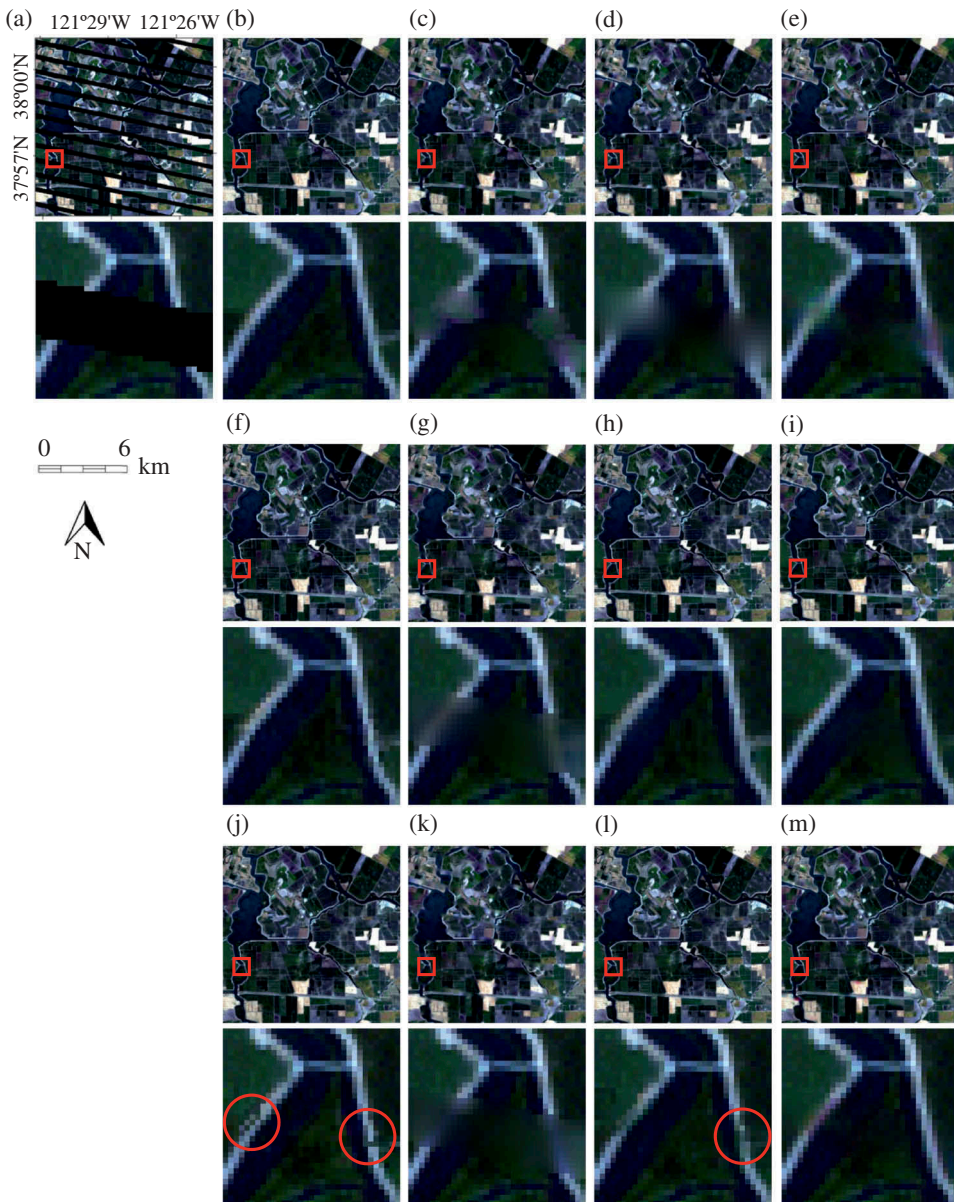
Figure 6. Realizations of R2 (heterogeneous area). (*a*) is the target image with gaps and (*b*) is the actual image used to validate the prediction accuracy of the gap-filling results. (*c–e*) represent the non-reference simulation (T0) using Kriging, WLR, and DS, respectively. (*f–i*) are the realizations of GNSPI, co-Kriging, WLR, and DS with temporally close input images (T1) implemented, respectively. (*j–m*) are the corresponding realizations of the GNSPI, co-Kriging, WLR, and DS using temporally far input images (T2). The area marked with a red square is magnified to the lower row of each pair.

accuracy, with WLR providing smaller RMSE values for bands 1–3, and DS providing smaller RMSE values for band 4 and MSA.

In terms of the consistency between the actual and predicted reflectance values, as shown in Table 6 and Figure S3, the DS and GNSPI methods provide the strongest
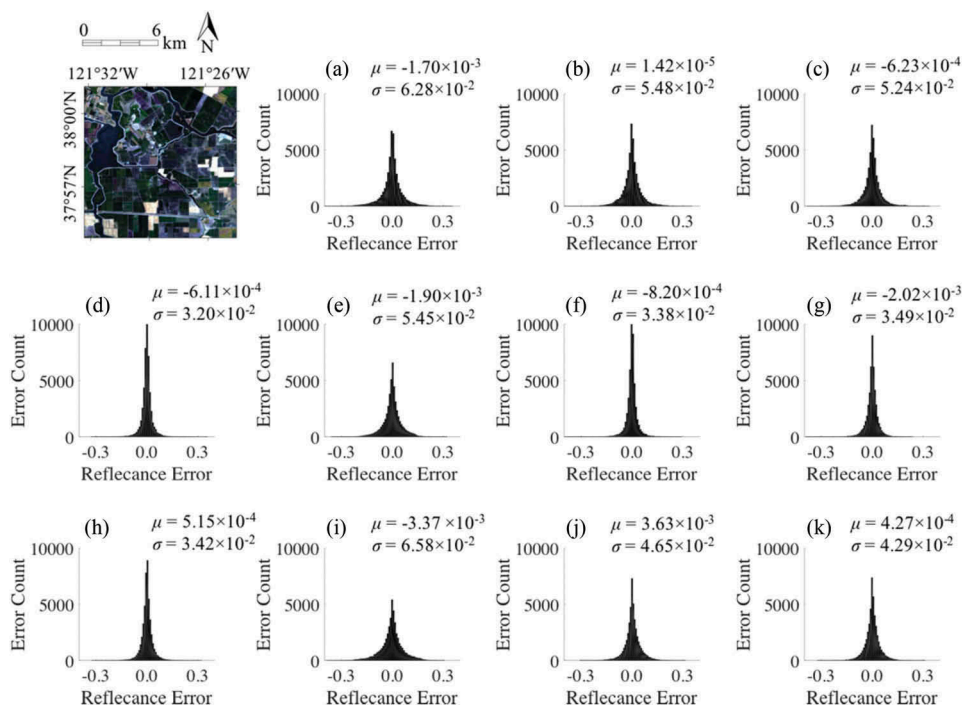
Figure 7. Reflectance error histograms of band 4 produced by each of the studied approaches under different input image conditions for R2 (heterogeneous area), with the x-axis representing the reflectance error and the y-axis representing the error count. (a–c) correspond to cases 'Kriging-T0', 'WLR-T0', and 'DS-T0'; (d–g) correspond to cases 'GNSPI-T1', 'co-Kriging-T1', 'WLR-T1', and 'DS-T1' using temporally close input images (T1); (h–k) correspond to cases 'GNSPI-T2', 'co-Kriging-T2', 'WLR-T2', and 'DS-T2' using temporally far input images (T2). μ and σ represent the mean and standard deviation of errors, respectively.

Table 5. Root mean square error (RMSE) of R2 (heterogeneous) gap-filling results simulated by Kriging, co-Kriging, GNSPI, DS, and WLR under different input image conditions for all four bands. The mean spectral angle (MSA) for each simulation case is provided with a unit of degree (°), which represents the overall prediction accuracy of the four bands. GNSPI* represents the case without using a time series image, as discussed in Section 3.4.

		RMSE				
Test Case	Approach	B1	B2	B3	B4	MSA (°)
T0	Kriging	0.0102	0.0146	0.0231	0.0628	**5.9339**
	WLR	0.0103	0.0135	0.0228	0.0548	**5.4498**
	DS	0.0090	0.0117	0.0200	0.0524	**5.1778**
T1	Co-Kriging	0.0105	0.0128	0.0214	0.0545	**5.4216**
	GNSPI	0.0054	0.0066	0.0110	0.0320	**2.6518**
	GNSPI*	0.0054	0.0068	0.0113	0.0322	**2.7499**
	WLR	0.0053	0.0067	0.0115	0.0338	**2.6437**
	DS	0.0055	0.0072	0.0118	0.0349	**2.9249**
T2	Co-Kriging	0.0099	0.0153	0.0261	0.0659	**6.6517**
	GNSPI	0.0062	0.0078	0.0136	0.0342	**3.0573**
	GNSPI*	0.0072	0.0090	0.0166	0.0441	**3.9393**
	WLR	0.0070	0.0091	0.0171	0.0466	**4.1083**
	DS	0.0076	0.0097	0.0179	0.0429	**4.0960**

consistency in non-reference applications and with input image conditions, respectively. For the WLR algorithm, the points lie closer to the 1:1 line when the temporally close input image (T1) is used, but diverge when a temporally far input image (T2) or no input image (T0) is employed. Kriging and co-Kriging present less consistent results across all of the studied cases. The R^2 values provided by GNSPI and WLR using T1 or T2 as the input images are slightly higher than the R^2 values in the study of Romero-Sanchez et al. (2015), whereas for co-Kriging, the R^2 values in the study are lower than the corresponding values provided by Pringle, Schmidt, and Muir (2009). However, the differences are all not significant, and are mainly due to the different study regions.

The OA and *kappa* values are commonly used metrics for evaluating the accuracy of land-surface classification, which is one the most important applications of Landsat imagery. From Table 7, the OA and *kappa* statistical metrics reflect the findings detailed in the sections above. Comparing the values provided by GNSPI and WLR in Table 7 to the study of Romero-Sanchez et al. (2015), the results are consistent with a slight difference owing to the different study regions. Training samples are visually selected from the reference image, and four classes are defined, including water, crops type 1, crops type 2, and bare ground. The sample numbers for each class are 656, 713, 639, and 616, respectively. Recall that higher OA and *kappa* values represent more accurate results from an overall prediction perspective (see Section 2.3). When no input image is used (T0), the classification result given by DS provides the highest OA and *kappa*, whereas the corresponding values given by WLR are smaller than Kriging. When using a temporally far input image (T2), the OA and *kappa* values given using WLR and DS are lower than for GNSPI, with DS slightly higher than WLR. The values generated by using co-Kriging are the lowest. When the temporally close input image (T1) is implemented, the values provided by all four approaches are improved. The OA and *kappa* values from GNSPI, WLR, and DS are all similar, with WLR slightly higher and DS slightly lower.

The influence of the distance to the gap edge on the prediction accuracy can be relevant for selecting the gap-filling approach. Figures S4–S6 display scatter plots of the predicted versus reference band 4 values for pixels that are either close to or far from the gap boundary. Pixels close to the boundary are defined as gap pixels that are adjacent to the

Table 6. R^2 of R2 (heterogeneous area) gap-filling results simulated by Kriging, co-Kriging, GNSPI, DS, and WLR under different input image conditions for all four bands. GNSPI* represents the case without using a time series image of the GNSPI approach, as discussed in Section 3.4.

Test Case	Approach	R^2			
		B1	B2	B3	B4
T0	Kriging	0.6102	0.6271	0.7282	0.6902
	WLR	0.6221	0.6936	0.7415	0.7647
	DS	0.6845	0.7550	0.7921	0.7804
T1	Co-Kriging	0.5725	0.7065	0.7617	0.7630
	GNSPI	0.8886	0.9213	0.9372	0.9181
	GNSPI*	0.8860	0.9180	0.9335	0.9173
	WLR	0.8904	0.9190	0.9319	0.9085
	DS	0.8843	0.9063	0.9284	0.9031
T2	Co-Kriging	0.6265	0.5920	0.6507	0.6552
	GNSPI	0.8516	0.8909	0.9044	0.9068
	GNSPI*	0.8000	0.8560	0.8588	0.8446
	WLR	0.8115	0.8520	0.8489	0.8273
	DS	0.7812	0.8363	0.8385	0.8530

Table 7. Accuracy assessment of R2 (heterogeneous area) land-cover classification from all the gap-filled results.

Test Case	Approach	Overall Accuracy	Kappa Value
T0	Kriging	0.7205	0.6014
	WLR	0.7015	0.5763
	DS	0.7786	0.6791
T1	Co-Kriging	0.7332	0.6200
	GNSPI	0.8721	0.8127
	WLR	0.8811	0.8230
	DS	0.8682	0.8053
T2	Co-Kriging	0.6907	0.5669
	GNSPI	0.8654	0.8022
	WLR	0.7885	0.6979
	DS	0.8076	0.7211

known pixels, and pixels far from the boundary are defined as those that are at least four pixels away from the gap edges. For all studied cases, the prediction accuracy of pixels far from the boundary is clearly lower than for pixels close to the boundary. However, the reduction of R^2 is the most significant for Kriging and co-Kriging. The R^2 values of pixels close to the boundary provided by Kriging or co-Kriging are similar to those obtained by the other methods, but the accuracy becomes much lower than the other approaches for pixels far from the boundary. The accuracy difference between pixels close to and far from the boundary is also obvious with WLR, especially when there is no input image (T0) or when using a temporally far input image (T2). This accuracy reduction may be explained by the fact that WLR may fail to fill all gap pixels and then a regularization method is implemented (as explained in Section 3.1.2). GNSPI consistently provides the highest R^2 for all cases where input images are used.

Overall, GNSPI consistently provides stable and satisfactory results regardless of the temporal distance of the input image being used, although it occasionally fails to preserve the shapes of narrow object such as roads precisely. The prediction accuracies provided by DS and WLR are comparable, especially when a temporally far input image is used for filling gaps. When no auxiliary input image is available for filling the gaps, the performance of DS is better than that of WLR. For all of the tested cases, Kriging and co-Kriging always give poorer results compared with the other approaches, with gap traces always present.

3.3. Application to areas with abrupt changes

Filling gaps located in regions where abrupt scene changes may have occurred is a situation that most gap-filling approaches avoid, due to the inherently challenging nature of the process. However, abrupt changes caused by natural disasters such as tsunamis, earthquakes, or bushfires occur frequently and represent an important topic for the improved monitoring of changes in land-cover features. In the R3 examples, the abrupt scene change is caused by a tsunami and the target image used is an actual image. In this case, there are no observed values of gap pixels available for quantitative assessment. Here, a temporally close image acquired by Landsat 5 is adopted as the reference image for qualitative assessment.

These realizations, together with the magnified sub-area, are presented in Figure 8. Simulations with and without the use of an input image are annotated as 'T1' and 'T0', respectively. The actual (Landsat 7) and reference (Landsat 5) images are also displayed in Figure 8. The filled images are all visually consistent from a large-scale perspective. However, when the realizations are zoomed in, the differences become apparent. For the non-reference (T0) cases, the edges of the gaps are very clear in the magnified images of Kriging, but much less obvious in the realization given by WLR and DS. Observing the same location of the magnified reference image, which is one-week temporally shifted from the target image, it is assumed that the truncation that appears in the Kriging result should not be present in the actual image. When one input image acquired before the tsunami is used for filling the gaps (T1), a striping effect can be observed in the realizations given by GNSPI, co-Kriging, and WLR. This effect is not produced in the DS-T1 simulation. Comparing WLR and DS under conditions of T0 and T1, it is clear that the results of DS are improved from T0 (no input image used) to T1 (using one input image) cases, whereas WLR presents an opposite change (i.e. WLR-T0 performs better than WLR-T1). The difference is likely a consequence of the underlying rationale behind the two approaches. For WLR, the approach assumes a linear pattern of temporal change for each pixel. When abrupt changes occur, this assumed pattern no longer matches the reality, and the approach fails. When no input image is used (T0), LPRM is conducted instead of WLR and it does not enforce such an assumption, resulting in improved predictions. For DS, the approach fills gaps without relying on interpolation, instead acquiring data for prediction from the input image directly. When no input image is used (T0), it uses the non-gap locations in the target image to fill gaps, which contain the spatial information only. When one input image is available (T1), both temporal and spatial information are used for filling gaps and the prediction accuracy is improved.

Overall, it is evident that when gaps are located in areas where abrupt scene changes have occurred, and the only input images are acquired before the changes, traditional approaches may fail to satisfy the gap-filling criteria. Although Zhu, Liu, and Chen (2012) emphasize the ability of GNSPI for filling gaps accurately with rapid changes, when the land cover of a large-scale area changes dramatically, GNSPI does not present the best reconstruction, with the assumption that pixels with high spectral similarity should have similar temporal changing pattern perhaps constraining the result in this case. For the DS case, this multiple-point geostatistical method is potentially useful to provide accurate prediction results, even without using any input images.

3.4. *Gap filling with time series images*

As introduced in Section 2.2.2, apart from using just one input image, the recommended application of GNSPI requires at least one time series image in order to select the neighbourhood pixels. To fulfil the ideal use of GNSPI, all of the results discussed above are generated using one time series image in addition to the single input image, which means it uses more auxiliary information than any of the other approaches. In practice, the number of time series image can be set to zero, and the pixels that are classified within the same category as the target pixel in the input image are used as the neighbourhood pixel. The RMSE, MSA, and R^2 results for R1 and R2
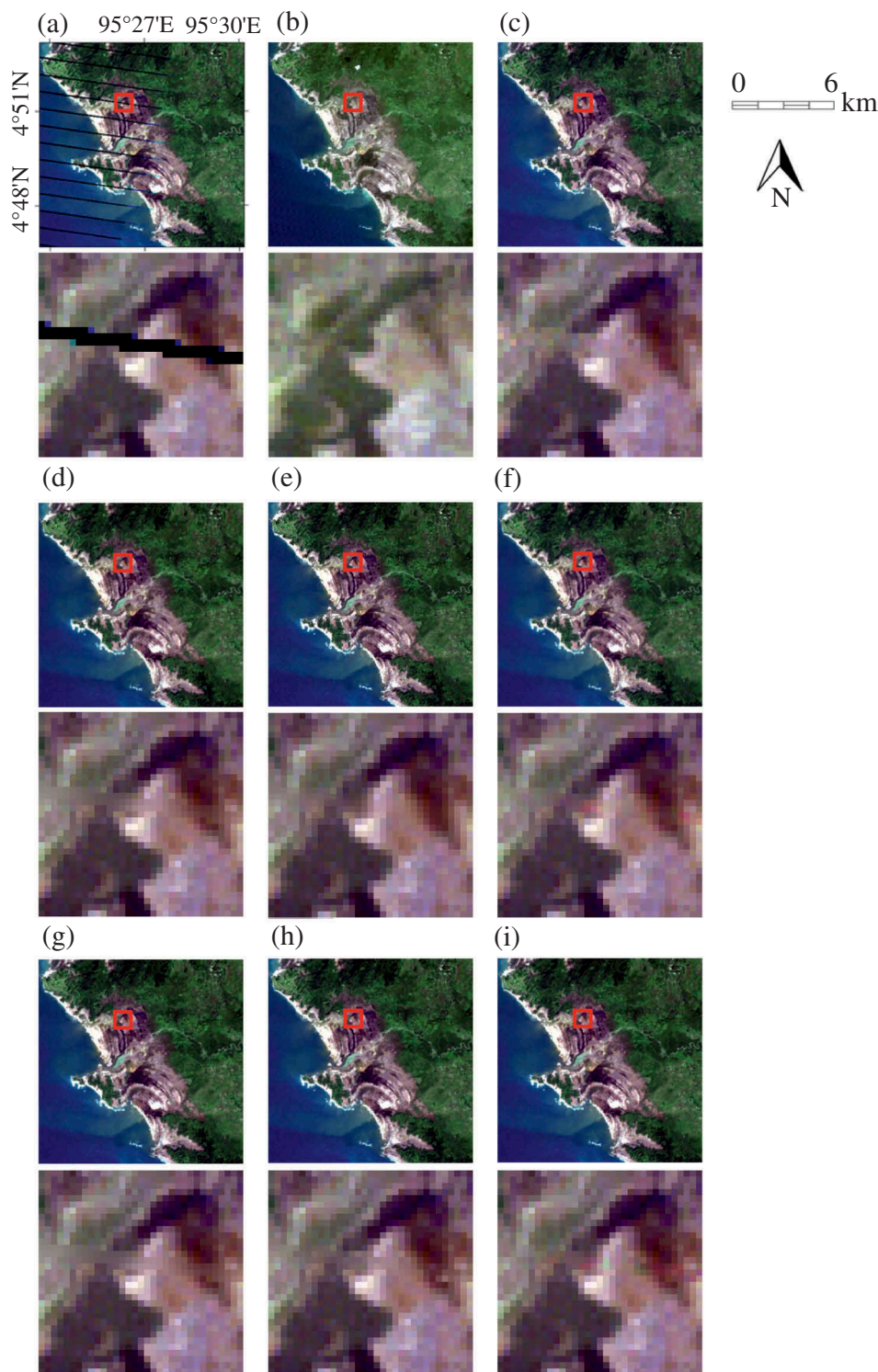
Figure 8. Mean realizations for the case of abrupt change (R3), with (*a*) the target image with gaps and (*b*) the reference image used to qualitatively assess the prediction accuracy of the gap-filling results. (*d–f*) represent the non-reference simulation (T0) using Kriging, WLR, and DS, respectively. (*c,g–i*) are the realizations of GNSPI, co-Kriging, WLR, and DS using the input image acquired before the occurrence of tsunami. The area marked with a red square is magnified to the lower row of each pair.

without using a time series image are listed in Tables 3–6 (notated as 'GNSPI*'). From Table 3–4, it is clear that for homogeneous area (R1), the results given by GNSPI* (without a time series image) and GNSPI (with a time series image) are comparable, with GNSPI* sometimes providing a higher accuracy than GNSPI. As for the heterogeneous area (R2), the case without using a time series image (GNSPI*) always provides a less accurate result (larger RMSE and smaller R^2 values) than the case using a time series image, and the lower accuracy becomes more evident when a temporally far (T2) input image is used for filling gaps. Therefore, we conclude that when filling gaps located in heterogeneous areas and when there is no temporally close input image available, the adoption of time series images can improve the prediction accuracy.

The DS method can also perform gap filling with more than two variables (i.e. the reflectance of the target image and two input images). Such simulations were studied, but not included here. From these results, it was concluded that using two input images acquired at dates that are both temporally close to the target image improves the prediction accuracy. However, if one temporally close and one temporally far input images are used, the prediction accuracy lies between the accuracy provided by using solely the temporally close image (i.e. bivariate simulation with the reflectance of the target image and the temporally close input image) and using solely the temporally far input image. The reason for the decreased accuracy of the tri-variate simulation compared with the bivariate simulation is possibly due to all variables being given the same weights when conducting the DS simulation. It is expected that when appropriate weights are given to each variable (i.e. reflectance of each image), the prediction results can be improved, but further study on this particular aspect is required.

From the comparison of different approaches in a variety of gap-filling situations, the advantages, disadvantages as well as the recommended area of application are summarized in Table 8 as a general guidance for selecting an appropriate approach. Overall, GNSPI can be implemented for general situations including both homogeneous and heterogeneous landscapes, especially when multi-temporal input images are provided. When high-quality input images that are cloud free and spectrally similar to the target image are available, WLR is recommended. DS is useful when there is no accessible input image (e.g. abrupt change) or when the input images are of poor quality (e.g. high cloud cover). In addition to using Landsat 7 imagery as the input images, implementing ancillary data from other sensors is an attractive option for gap filling, with a series of approaches developed to do this (Chen, Tang, and Qiu 2010; Reza and Ali 2008; Roy et al. 2008; Chen, Zhao, and Ye 2012). The potential of applying data from other sensors as input images in these studied approaches can be of interest for further investigation. Since information such as the orbit, bandwidth, and revisit period of different remote sensors varies, finding sensors that fulfil the requirements of similar spatial and radiometric resolution, spectral bands, and short time distance, or developing reliable methods to calibrate the data to satisfy these requirements, is of considerable importance.

4. Conclusions

Given the importance of the Landsat series of satellites in reproducing multi-decadal time series of a variety of land-surface elements and features, a range of gap-filling approaches has been proposed to address the failure of the scan-line corrector. In this study, four methods that have been demonstrated to provide satisfactory results for

Table 8. Summary of the advantages, disadvantages, and recommended application for each of the studied gap-filling methodologies.

Approach	Advantage	Disadvantage	Recommended Application
Kriging & Co-Kriging	Unbiased estimation of gap pixels; can be used without input images	Assumption of stationarity; low accuracy at the pixel level	Homogeneous areas
GNSPI	Accurate and robust estimation	Complicated procedure; needs at least one input image	Both homogeneous and heterogeneous areas, especially when multi-temporal input images are available; vegetation monitoring
WLR	Accurate estimation; preserves the shape of ground features; can be used without input images	May fail when gap widths are large	Both homogeneous and heterogeneous areas, especially when high-quality, cloud-free, and spectrally similar input images are available
DS	Accurate estimation; simple concept; preserves the shape of ground features; can be used without input images	No explicit guidance on the selection of optimum parameters	Both homogeneous and heterogeneous areas, especially when no input images are available or of poor quality (e.g. abrupt scene changes); man-made feature monitoring

filling gaps are examined and compared. The different methods are applied to a homogeneous area (R1), a heterogeneous area (R2), and an area that has undergone an abrupt change (R3) (but where the real data are unknown). Both temporally close and far input images are used for filling the gaps to study the influence of spectral similarity in estimation accuracy. Although temporally close input images do not always guarantee better spectral similarity to the target image, the assumption is reasonable in general. Therefore, the use of T1 (temporally close) and T2 (temporally far) images in the study correspond to using ancillary data that have different levels of similarity to the target conditions. From the results, it is clear that using a temporally close input image can improve the estimation accuracy.

Analysis indicates that Kriging and co-Kriging provide acceptable results for both homogeneous and heterogeneous landscapes at a large-scale level. However, when per-pixel accuracy is required (i.e. for the detailed study of small regions), the results generated by Kriging and co-Kriging are not appropriate. The prediction accuracy provided by the remaining three approaches is comparable, although with case-specific advantages. The GNSPI algorithm provides the highest prediction accuracy across both homogeneous and heterogeneous areas, regardless of the temporal distance of the input image. It can be applied to areas where rapid changes occur over a small scale, but when abrupt changes occur on a large scale (as in the R3 case), the results should be interpreted with care. The WLR algorithm also performs satisfactorily in both homogeneous and heterogeneous cases, especially when a temporally close input image is used, with the gap-filling prediction accuracy comparable with GNSPI.

Our results suggest that the WLR algorithm performs well when temporally close input images are available. With the DS method, the results are marginally poorer than

for GNSPI or WLR when using an input image, although with occasionally higher accuracy (such as in the case of the temporally far input image for filling gaps in R1). The most obvious merit of DS over the other approaches is the prediction of abrupt scene changes. When rapid changes occur due to tsunamis or bushfires, the land surface can rapidly become starkly different and satellite imagery provides one of the primary means of assessing such change. In such cases, GNSPI and WLR, which rely on the pixel interpolation of the values in the input images and target images, become less reliable because the reflectance changes do not follow the assumed temporal pattern. DS does not assume a specific pattern and not only uses the temporal information within the input image, but also uses the spatial information contained in the target image, providing increased consistency in the filled areas.

Understanding the relative performance of common gap-filling approaches is important, particularly for land-cover scenarios that expand beyond the more common idealized homogeneous case studies. While the work presented here is by no means an exhaustive assessment, it does provide a framework for future assessments in terms of examining both qualitative and quantitative aspects, besides introducing multiple performance metrics. Further analysis spanning a larger range of land-cover types (the spatial domain) and across dynamic land-cover conditions (the temporal domain) is required to provide additional constraint on the choice of an appropriate technique. However, it is important to recognize that the optimal approach for one set of conditions does not necessarily translate to an alternate set, so prior understanding of the relative performance limits is required. Importantly, although the analysis here has focused on a specific problem set (i.e. Landsat 7 scan line correction failure), the findings are applicable to other gap-filling problems, e.g. gaps resulting from cloud contamination or orbital characteristics.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

References

Bormann, K. J., M. F. McCabe, and J. P. Evans. 2012. "Satellite Based Observations for Seasonal Snow Cover Detection and Characterisation in Australia." *Remote Sensing of Environment* 123: 57–71. doi:10.1016/j.rse.2012.03.003.

Boucher, A., K. C. Seto, and A. G. Journel. 2006. "A Novel Method for Mapping Land Cover Changes: Incorporating Time and Space with Geostatistics." *IEEE Transactions on Geoscience and Remote Sensing* 44 (11): 3427–3435. doi:10.1109/TGRS.2006.879113.

Chen, F., L. Tang, and Q. Qiu. 2010. "Exploitation of CBERS-02B as Auxiliary Data in Recovering the Landsat7 ETM+ SLC-off Image." In *2010 18th International Conference on Geoinformatics, Geoinformatics 2010*. http://www.scopus.com/inward/record.url?eid=2-s2.0-77958053760&partnerID=40&md5=96242f35ce4c7a0006e98c6e4303d60f.

Chen, F., X. Zhao, and H. Ye. 2012. "Making Use of the Landsat 7 SLC-off ETM+ Image through Different Recovering Approaches." In *Data Acquisition Applications*, edited by Z. Karakehayov, Rijeka: InTech.

Chen, J., X. Zhu, J. E. Vogelmann, F. Gao, and S. Jin. 2011. "A Simple and Effective Method for Filling Gaps in Landsat ETM+ SLC-off Images." *Remote Sensing of Environment* 115 (4): 1053–1064. doi:10.1016/j.rse.2010.12.010.

Chiles, J., and P. Delfiner. 1999. "Geostatistics: Modeling Spatial Uncertainty." *Wiley Series in Probability and Statistics. Applied Probability and Statistics Section.*, no. xi: 695. doi:10.1007/s11004-012-9429-y.

Congalton, R. G. 1991. "A Review of Assessing the Accuracy of Classification of Remotely Sensed Data." *Remote Sensing of Environment* 4257 (September): 34–46. doi:10.1016/0034-4257(91)90048-B.

Ershadi, A., M. F. McCabe, J. P. Evans, and J. P. Walker. 2013. "Effects of Spatial Aggregation on the Multi-Scale Estimation of Evapotranspiration." *Remote Sensing of Environment* 131: 51–62. doi:10.1016/j.rse.2012.12.007.

Hossain, M. S., J. S. Bujang, M. H. Zakaria, and M. Hashim. 2015. "Assessment of Landsat 7 Scan Line Corrector-Off Data Gap-Filling Methods for Seagrass Distribution Mapping." *International Journal of Remote Sensing* 36 (4): 1188–1215. Taylor and Francis

Houborg, R., M. F. McCabe, A. Cescatti, F. Gao, M. Schull, and A. Gitelson. 2015. "Joint Leaf Chlorophyll Content and Leaf Area Index Retrieval from Landsat Data Using a Regularized Model Inversion System (REGFLEC)." *Remote Sensing of Environment* 159: 203–221. doi:10.1016/j.rse.2014.12.008.

Jha, S. K., G. Mariethoz, J. Evans, and M. F. McCabe. 2013. "Demonstration of a Geostatistical Approach to Physically Consistent Downscaling of Climate Modeling Simulations." *Water Resources Research* 49 (1): 245–259. doi:10.1029/2012WR012602.

Jha, S. K., G. Mariethoz, J. Evans, M. F. McCabe, and A. Sharma. 2015. A Space and Time Scale-Dependent Nonlinear Geostatistical Approach for Downscaling Daily Precipitation and Temperature. *Water Resources Research* 51 (8): 6244–6261. Blackwell Publishing. doi:10.1002/2014WR016729.

Journel, A., and T. Zhang. 2006. "The Necessity of a Multiple-Point Prior Model." *Mathematical Geology* 38 (5): 591–610. doi:10.1007/s11004-006-9031-2.

Lo, C. P., and J. Choi. 2004. "A Hybrid Approach to Urban Land Use/Cover Mapping Using Landsat 7 Enhanced Thematic Mapper Plus (ETM+) Images." *International Journal of Remote Sensing* 25 (14): 2687–2700. doi:10.1080/01431160310001618428.

Malambo, L., and C. D. Heatwole. 2016. "A Multitemporal Profile-Based Interpolation Method for Gap Filling Nonstationary Data." *IEEE Transactions on Geoscience and Remote Sensing* 54 (1): 252–261. Institute of Electrical and Electronics Engineers doi:10.1109/TGRS.2015.2453955.

Mariethoz, G., and J. Caers. 2014. Multiple-Point Geostatistics: Stochastic Modeling with Training Images. Multiple-Point Geostatistics: Stochastic Modeling with Training Images. doi:10.1002/9781118662953.

Mariethoz, G., M. F. McCabe, and P. Renard. 2012. "Spatiotemporal Reconstruction of Gaps in Multivariate Fields Using the Direct Sampling Approach." *Water Resources Research* 48: 10. doi:10.1029/2012WR012115.

Mariethoz, G., and P. Renard. 2010. "Reconstruction of Incomplete Data Sets or Images Using Direct Sampling." *Mathematical Geosciences* 42 (3): 245–268. doi:10.1007/s11004-010-9270-0.

Mariethoz, G., P. Renard, and J. Straubhaar. 2010. "The Direct Sampling Method to Perform Multiple-Point Geostatistical Simulations." *Water Resources Research* 46: 11. doi:10.1029/2008WR007621.

Mariethoz, G. 2010. "The Direct Sampling Code." http://www.minds.ch/gm/downloads.htm.

Maxwell, S. K., G. L. Schmidt, and J. C. Storey. 2007. "A Multi-Scale Segmentation Approach to Filling Gaps in Landsat ETM+ SLC-off Images." *International Journal of Remote Sensing* 28 (23): 5339–5356. doi:10.1080/01431160601034902.

Pardo-Iguzquiza, E., P. M. Atkinson, and M. Chica-Olmo. 2010. "DSCOKRI: A Library of Computer Programs for Downscaling Cokriging in Support of Remote Sensing Applications." *Computers and Geosciences* 36 (7): 881–894. doi:10.1016/j.cageo.2009.10.006.

Pringle, M. J., M. Schmidt, and J. S. Muir. 2009. "Geostatistical Interpolation of SLC-off Landsat ETM + Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (6): 654–664. doi:10.1016/j.isprsjprs.2009.06.001.

Reza, M. M., and S. N. Ali. 2008. "Using IRS Products to Recover 7ETM+ Defective Images." *American Journal of Applied Sciences* 5 (6): 618–625. doi:10.3844/ajassp.2008.618.625.

Romero-Sanchez, M. E., R. Ponce-Hernandez, S. E. Franklin, and C. A. Aguirre-Salado. 2015. "Comparison of Data Gap-Filling Methods for Landsat ETM+ SLC-off Imagery for Monitoring Forest Degradation in a Semi-Deciduous Tropical Forest in Mexico." *International Journal of Remote Sensing* 36 (11): 2786–2799. Taylor and Francis. doi:10.1080/01431161.2015.1047991.

Roy, D. P., J. Ju, P. Lewis, C. Schaaf, F. Gao, M. Hansen, and E. Lindquist. 2008. "Multi-Temporal MODIS-Landsat Data Fusion for Relative Radiometric Normalization, Gap Filling, and Prediction of Landsat Data." *Remote Sensing of Environment* 112 (6): 3112–3130. doi:10.1016/j.rse.2008.03.009.

Scaramuzza, P., E. Micijevic, and G. Chander. 2004a. "SLC Gap-Filled Products Phase One Methodology." *Landsat Technical Notes*. https://landsat.usgs.gov/documents/SLC_Gap_Fill_Methodology.pdf.

Scaramuzza, P., E. Micijevic, and G. Chander. 2004b. "Phase 2 Gap-Fill Algorithm: SLC-off Gap-Filled Products Gap-Filled Algorithm Methodology." *Landsat Technical Notes*. https://landsat.usgs.gov/sites/default/files/documents/L7SLCGapFilledMethod.pdf

SENDIMAGE. 2016. "Software Tool for Recovering Missing Pixels of Remotely Sensed Image." Accessed November 3. http://sendimage.whu.edu.cn/en/resources/.

Tang, Y., P. M. Atkinson, N. A. Wardrop, and J. Zhang. 2013. Multiple-Point Geostatistical Simulation for Post-Processing a Remotely Sensed Land Cover Classification. *Spatial Statistics* 5 (1): 69–84. Elsevier. doi:10.1016/j.spasta.2013.04.005.

Tang, Y., P. M. Atkinson, and J. Zhang. 2015. "Downscaling Remotely Sensed Imagery Using Area-to-Point Cokriging and Multiple-Point Geostatistical Simulation." *ISPRS Journal of Photogrammetry and Remote Sensing* 101: 174–185. Elsevier. doi:10.1016/j.isprsjprs.2014.12.016.

USGS. 2003. Preliminary Assessment of Landsat 7 ETM+ Data following Scan Line Corrector Malfunction. Accessed 16 July 2003. https://landsat.usgs.gov/sites/default/files/documents/SLC_off_Scientific_Usability.pdf

Wulder, M. A., S. M. Ortlepp, J. C. White, and S. Maxwell. 2008. "Evaluation of Landsat-7 SLC-off Image Products for Forest Change Detection." *Canadian Journal of Remote Sensing* 34 (2): 93–99. doi:10.5589/m08-020.

Yin, G., G. Mariethoz, and M. F. McCabe. 2017. "Gap-Filling of Landsat 7 Imagery Using the Direct Sampling Method." *Remote Sensing* 9 (1): 12. doi:10.3390/rs9010012.

Zeng, C., H. Shen, and L. Zhang. 2013. "Recovering Missing Pixels for Landsat ETM+ SLC-off Imagery Using Multi-Temporal Regression Analysis and a Regularization Method." *Remote Sensing of Environment* 131: 182–194. doi:10.1016/j.rse.2012.12.012.

Zhang, C., W. Li, and D. Travis. 2007. "Gaps-Fill of SLC-off Landsat ETM+ Satellite Image Using a Geostatistical Approach." *International Journal of Remote Sensing* 28 (22): 5103–5122. doi:10.1080/01431160701250416.

Zhu, X., D. Liu, and J. Chen. 2012. "A New Geostatistical Approach for Filling Gaps in Landsat ETM+ SLC-off Images." *Remote Sensing of Environment* 124: 49–60. doi:10.1016/j.rse.2012.04.019.

Zhu, X. 2013. "Open-Source Code." Accessed October 23. http://xiaolinzhu.weebly.com/open-source-code.html.