

Article

An Unmixing-Based Bayesian Model for Spatio-Temporal Satellite Image Fusion in Heterogeneous Landscapes

Jie Xue ¹ , Yee Leung ^{2,*} and Tung Fung ³

¹ Department of Geography and Resource Management, the Chinese University of Hong Kong, Hong Kong; jiexue@link.cuhk.edu.hk

² Department of Geography and Resource Management, Institute of Future Cities, the Chinese University of Hong Kong, Hong Kong

³ Department of Geography and Resource Management, Institute of Future Cities, Institute of Environment, Energy and Sustainability, the Chinese University of Hong Kong, Hong Kong; tungfung@cuhk.edu.hk

* Correspondence: yeeleung@cuhk.edu.hk; Tel.: +852-3943-6473

Received: 14 January 2019; Accepted: 30 January 2019; Published: 6 February 2019



Abstract: Studies of land surface dynamics in heterogeneous landscapes often require satellite images with a high resolution, both in time and space. However, the design of satellite sensors often inherently limits the availability of such images. Images with high spatial resolution tend to have relatively low temporal resolution, and vice versa. Therefore, fusion of the two types of images provides a useful way to generate data high in both spatial and temporal resolutions. A Bayesian data fusion framework can produce the target high-resolution image based on a rigorous statistical foundation. However, existing Bayesian data fusion algorithms, such as STBDF (spatio-temporal Bayesian data fusion) -I and -II, do not fully incorporate the mixed information contained in low-spatial-resolution pixels, which in turn might limit their fusion ability in heterogeneous landscapes. To enhance the capability of existing STBDF models in handling heterogeneous areas, this study proposes two improved Bayesian data fusion approaches, coined ISTBDF-I and ISTBDF-II, which incorporate an unmixing-based algorithm into the existing STBDF framework. The performance of the proposed algorithms is visually and quantitatively compared with STBDF-II using simulated data and real satellite images. Experimental results show that the proposed algorithms generate improved spatio-temporal-resolution images over STBDF-II, especially in heterogeneous areas. They shed light on the way to further enhance our fusion capability.

Keywords: bayesian estimation framework; remote sensing; high spatiotemporal resolution; image fusion; spatial unmixing; heterogeneous landscape

1. Introduction

Satellites provide spatially continuous observations that are crucial to the monitoring of the earth's surface in various fields, including ecology, hydrology, environmental sciences, etc. The high-resolution information of land surfaces in space and time is of vital importance in many applications. In particular, satellite observations with high spatial resolution and high frequency of coverage are useful in characterizing heterogeneous landscapes. However, currently available satellite sensors, limited by their designs, often need to make tradeoffs between spatial details and revisit frequency. Taking the widely used Landsat and Moderate Resolution Image Spectroradiometer (MODIS) as an example, Landsat provides remote sensing images at comparatively high spatial resolution (i.e., 30 m), useful for characterizing heterogeneous landscapes. However, the use of Landsat may be limited by the relatively low temporal resolution (i.e., 16 days), especially in frequently cloudy areas. MODIS, on the other hand,

has a 1- to 2-day revisit period, suitable for more frequent monitoring. Nevertheless, its relatively low spatial resolution ranging from 250m to 1 km limits its use for detailed characterization of land covers, especially in heterogeneous landscapes. For simplicity, high spatial and low temporal resolution is henceforth referred to as “HSLTR”, low spatial and high temporal resolution is henceforth referred to as “LSHTR”.

Spatio-temporal data fusion techniques facilitate the temporal characterization of heterogeneous landscapes by offering an efficient integration of images from multiple sources [1]. Recently, many algorithms have been proposed to produce spatially and temporally high-resolution images. These fusion approaches can be grouped into four categories: STARFM-based, unmixing-based, learning-based, and Bayesian-based methods.

One of the pioneering fusion approaches is the spatial and temporal adaptive reflectance fusion model (STARFM) [2]. This algorithm is suitable for homogeneous surfaces and has been used for the classification of land surfaces with gradual changes (e.g., phenology changes). Spatial temporal adaptive algorithm for mapping reflectance changes (STAARCH), developed on the basis of STARFM, shows improved ability in detecting land cover changes, especially for forest disturbance events [3,4]. Another enhanced version of STARFM (ESTARFM) was proposed to improve the fusion accuracy in heterogeneous regions [5]. A disadvantage of STAARCH and ESTARFM is that the two methods both require two pairs of images as input [6], which may limit their use in days that are frequently cloudy (e.g., rainy season) [7]. Several algorithms using only one image pair as input have been developed [8] and shown to outperform STARFM in detecting land-cover changes [9,10]. In addition, ESTARFM assumes that the change of surface reflectance follows a constant rate. This problem is also found in the spatial and temporal non-local, filter-based fusion model (STNLFFM) [11], which was especially developed for heterogeneous landscapes and is shown to have improved performance over STARFM and ESTARFM.

The second class of fusion approaches are unmixing-based methods. Spectral unmixing is the process of decomposing the spectrum of a mixed pixel to a combination of the spectra of different pure endmembers. The percentage of each endmember in the mixed pixel is indicated by the corresponding abundance [12,13]. In the case of unmixing-based spatio-temporal fusion, the number of endmembers and their corresponding abundances are derived from the classification map of HSLTR images and the spectral signatures of the endmembers are unmixed from the LSHTR images. Zhukov et al. [14] were perhaps one of the first to develop the unmixing-based algorithm for satellite image fusion. The linear unmixing method with constraints has then been applied to constrain the reflectance values in spatio-temporal fusion [15–18]. The main advantage of the unmixing-based methods is their flexibility in dealing with the spectral difference between HSLTR and LSHTR images [19]. However, they may not capture the intra-class spatial variability in a local or entire region and may also lead to unrealistic endmember spectra. To overcome the limitations and to increase the intra-class spatial variability, several improved algorithms with respect to the traditional unmixing method have recently been developed to get better endmember spectra by employing different clustering methods or by unmixing the LSHTR images on both the input and target dates [20–26].

Another class of methods are learning-based algorithms. Based on the training of the dictionary pair to characterize the similarity between the HSLTR and LSHTR images, a sparse-representation-based spatio-temporal reflectance fusion model (SPSTFM) was developed to introduce a general image super-resolution method into the spatio-temporal fusion of remotely sensed images [27]. This algorithm attempts to capture land cover and phenology changes. Although it can improve the prediction of pixels with land-cover changes, they do not accurately maintain the shapes of objects, especially when the scale difference between HSLTR and LSHTR images is large [21]. Recently, several methods have been developed to address the limitation of assumptions made in dictionary learning and sparse coding [28–30].

Bayesian estimation methods have aroused great interest in various fields, especially in image processing. For example, Bayesian methods have been widely applied to perform

image super-resolution to obtain a single high-spatial-resolution image by combining a set of low-spatial-resolution images of the same scene [31–34]. They have also been employed in multispectral or hyperspectral image fusion to optimally weigh the spatial and spectral information [35–38]. However, there are only a few studies on the use of the Bayesian theory for the spatio-temporal fusion of satellite data [39–41]. In those studies, linear regression is usually assumed to be able to reflect temporal dynamics. Nevertheless, in reality, there may even be no regression-like trends in some situations. Instead of the regression method, Xue et al. [42] proposed a Bayesian data fusion approach in a more flexible way that capitalizes on the advantage of multivariate arguments, including the joint distribution that embodies the covariate information, in expressing implicitly the change of reflectance.

Some studies have demonstrated that the aforementioned four approaches can improve the spatial and temporal resolution of remotely sensed images in specific applications (e.g., vegetation indices) [3,43–45]. Nevertheless, the applicability of each is limited to a certain extent because of the decrease in their accuracy in heterogeneous regions, their inability to detect rapid land-cover changes, or the requirement of the number of HSLTR images. For example, a key assumption of STARFM is the existence of homogeneous LSHTR pixels within the search window, which may not hold over heterogeneous landscapes [46]. ESTARFM requires at least two HSLTR images as inputs, which may limit its applicability in areas with frequent cloudy days [46]. The unmixing-based approaches emphasize the LSHTR image at the target date, which may overlook the temporal changes and result in LSHTR-like images without the full use of spatial heterogeneity of the HSLTR image in the input [47]. Recently, a number of synthetic approaches have been developed to combine the strengths of the unmixing-based algorithms and STARFM, or other techniques, to model non-linear temporal variability and to extract spatial details and variability [21,47–51]. These ensemble approaches tend to produce better fusion results than any single algorithm.

In this paper, we propose two approaches that integrate Bayesian methods and unmixing-based methods to capitalize on the strengths of both. The advantages of the Bayesian methods are in their flexibility in employing various statistical estimation tools for the fusion process and their ability in capturing uncertainties in a probabilistic and statistical manner. The Bayesian data fusion approach in Xue et al. [42] treats the fusion problem as an estimation problem in which the fused image is estimated by a first-order observational model, and then the correlation and temporal evolution information from the image time series are combined in the form of a joint distribution. This method produces comparable or even better results than STARFM and ESTARFM based on their experimental studies [43]. However, due to the difficulty in deriving the theoretical expression of the real point spread function (PSF) for images from different sensors, an approximate average-weighted matrix for the PSF in the first-order observational model is employed. In addition, the mean spectra of the joint distribution are estimated by direct sampling, which may depart from the real situations. These methods may limit the effectiveness of the model, especially over heterogeneous areas. To improve the model, we develop here two improved Bayesian data fusion approaches. They are expected to fully extract useful information from LSHTR pixels with mixed land covers through the help of the unmixing-based algorithms rather than through direct averaging or resampling. The main objectives of this study are: (1) to improve the performance of the current Bayesian data fusion methods in more complex and heterogeneous regions via the incorporation of an unmixing-based algorithm, (2) to conceptually and experimentally test and compare the proposed approaches with the previous Bayesian data fusion method using simulated data and real-life Landsat and MODIS images, and (3) to show the robustness of the proposed algorithms through sensitivity analysis on the impact of the parameters. To fulfill the objectives, the proposed approaches predict the HSLTR image at the target date using the LSHTR image at the target date and two pairs of HSLTR and LSHTR images acquired at the nearest dates before and after the target date. Experiments with both simulated and real satellite images will be performed.

In the remainder of this paper, we first give a detailed description of the theoretical basis of the two improved Bayesian data fusion approaches in Section 2. The experimental design and analysis results are discussed in Sections 3 and 4, respectively. We then conclude our discussion in Section 5.

2. Methodology

2.1. Spatio-Temporal Bayesian Data Fusion Method (STBDF)

The Bayesian data fusion method (STBDF -II) proposed in Xue et al. [42] is employed in this study. In Xue et al. [42], the algorithm is implemented via the following three steps. First, the linear observational model is employed to establish the relationship between the LSHTR and HSLTR images at the target date:

$$\mathbf{y}^{(t_0)} = \mathbf{W}\mathbf{z}^{(t_0)} + \mathbf{e}^{(t_0)} \tag{1}$$

where $\mathbf{y}^{(t_0)}$ is the LSHTR image at time t_0 , $\mathbf{z}^{(t_0)}$ is the corresponding HSLTR image to be predicted, and $\mathbf{e}^{(t_0)}$ is the noise that is assumed to be a zero-mean Gaussian random vector with a covariance matrix $\mathbf{C}_{\mathbf{e}^{(t_0)}}$, and \mathbf{W} is approximated as the rectangular detector model in which each LSHTR pixel is assumed to be the average of all the HSLTR pixels within it [52]. Then, the conditional probability density function is derived as:

$$p(\mathbf{y}^{(t_0)}|\mathbf{z}^{(t_0)}) = \frac{1}{\sqrt{(2\pi)^{MK}|\mathbf{C}_{\mathbf{e}^{(t_0)}}|}} \exp\left\{-\frac{1}{2}(\mathbf{y}^{(t_0)} - \mathbf{W}\mathbf{z}^{(t_0)})^T \mathbf{C}_{\mathbf{e}^{(t_0)}}^{-1} (\mathbf{y}^{(t_0)} - \mathbf{W}\mathbf{z}^{(t_0)})\right\} \tag{2}$$

Second, a multivariate joint Gaussian distribution is used to model the temporal changes of the HSLTR images and their correlation information. Then, the conditional probability density function is also Gaussian, which can be expressed as:

$$p(\mathbf{z}^{(t_0)}|\mathbf{x}^{(t_1)}, \dots, \mathbf{x}^{(t_s)}) = p(\mathbf{z}^{(t_0)}|\mathbf{X}) = \frac{1}{\sqrt{(2\pi)^{NK}|\mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}}|}} \exp\left\{-\frac{1}{2}(\mathbf{z}^{(t_0)} - \boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}})^T \mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}}^{-1} (\mathbf{z}^{(t_0)} - \boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}})\right\} \tag{3}$$

where $\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}}$ is the conditional expectation of $\mathbf{z}^{(t_0)}$ given the input HSLTR-image time series $\mathbf{X} = [(\mathbf{x}^{(t_1)})^T, \dots, (\mathbf{x}^{(t_s)})^T]^T$, and $\mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}}$ is the conditional covariance matrix of $\mathbf{z}^{(t_0)}$ given \mathbf{X} . They can be calculated using the joint statistics:

$$\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}} = E\{\mathbf{z}^{(t_0)}\} + \mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{X}} \mathbf{C}_{\mathbf{X},\mathbf{X}}^{-1} (\mathbf{X} - E\{\mathbf{X}\}) \tag{4}$$

$$\mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}} = \mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{z}^{(t_0)}} - \mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{X}} \mathbf{C}_{\mathbf{X},\mathbf{X}}^{-1} \mathbf{C}_{\mathbf{X},\mathbf{z}^{(t_0)}}^T \tag{5}$$

where $E\{\mathbf{X}\}$ is the mean of \mathbf{X} and $E\{\mathbf{z}^{(t_0)}\}$ is the mean of $\mathbf{z}^{(t_0)}$, which can be estimated by the bilinear interpolation of the LSHTR images together with the high-pass frequencies of the HSLTR images correspondingly. The matrices $\mathbf{C}_{\mathbf{X},\mathbf{X}}$, $\mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{X}}$, and $\mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{z}^{(t_0)}}$ are the cross-covariance matrices, which can be predicted from the joint covariance of different classes in the LSHTR images.

The last step is the Maximum A Posterior (MAP) estimation. The HSLTR image $\mathbf{z}^{(t_0)}$ is derived by maximizing its posterior probability given \mathbf{X} and $\mathbf{y}^{(t_0)}$. It can be given as:

$$\hat{\mathbf{z}}^{(t_0)} = \underset{\mathbf{z}^{(t_0)}}{\operatorname{argmax}} p(\mathbf{z}^{(t_0)}|\mathbf{y}^{(t_0)}, \mathbf{X}) = \underset{\mathbf{z}^{(t_0)}}{\operatorname{argmax}} p(\mathbf{y}^{(t_0)}|\mathbf{z}^{(t_0)}) \cdot p(\mathbf{z}^{(t_0)}|\mathbf{X}) \tag{6}$$

Substituting (2) and (3) into (6), the optimal estimator of $\mathbf{z}^{(t_0)}$ is given by:

$$\hat{\mathbf{z}}^{(t_0)} = \boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}} + \mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}} \mathbf{W}^T [\mathbf{W}\mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}} \mathbf{W}^T + \mathbf{C}_{\mathbf{e}^{(t_0)}}]^{-1} [\mathbf{y}^{(t_0)} - \mathbf{W}\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}}] \tag{7}$$

For a more detailed description of the STBDF-I and -II, please refer to Xue et al. [42].

2.2. Two Improved Bayesian Data Fusion Approaches

In heterogeneous areas, pixels in the LSHTR images can contain highly mixed land-covers. In most cases, the proportions of land-covers within one LSHTR pixel are different. Under this situation, the STBDF-I and -II only employ the equal-weight matrix in the linear observational model and resamples the LSHTR image to obtain the mean estimation of the HSLTR image. This may be inappropriate and may lead to inaccurate results because they may overlook spatial details in the LSHTR images. To better utilize the mixed spatial information in the LSHTR images, unmixing-based algorithms can be instrumental. However, using only the unmixing-based algorithm may result in the homogeneous spot in the fused images because of the average reflectance of one class in the local or entire region. Therefore, in this study, we propose two improved Bayesian data fusion approaches to rectify the shortcomings of STBDF-I and -II by introducing an unmixing-based algorithm to the estimation procedure depicted in Figure 1.

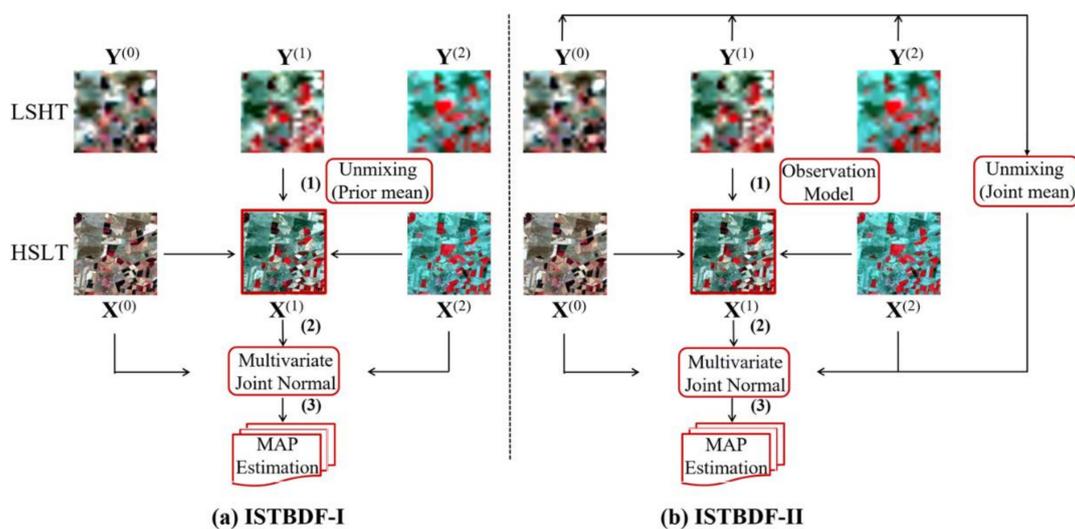


Figure 1. Theoretical workflow of the proposed ISTBDF-I (a) and ISTBDF-II (b).

2.2.1. Spatial and Temporal Reflectance Unmixing Model (STRUM)

Unmixing-based spatio-temporal fusion can be summarized in four steps: (1) perform clustering on the HSLTR images to generate a classification map at high spatial resolution and define the endmembers, (2) derive the abundance of each endmember for each LSHTR pixel, (3) unmix the LSHTR image by using a moving window, and (4) assign the endmember spectra of each class to the HSLTR image. Two important problems arise from such an unmixing approach: one is the collinearity problem and the other is the plausible production of unrealistic spectra. A number of strategies have been proposed to alleviate these two problems [16–22,25,47,50,53,54]. Among them, STRUM is an effective approach that was recently proposed [50]. In that study, the first problem is solved by discarding the endmembers with proportions lower than 1% in at least one of the LSHTR pixels of the moving window. This has been shown to be able to reduce the residual error. The second problem is solved by introducing the Bayesian estimation theory with the ability to incorporate the prior spectra information. The prior spectra can be selected from the LSHTR images. Because of these strengths, STRUM is employed in the current study.

2.2.2. Improved Spatio-Temporal Bayesian Data Fusion Method I (ISTBDF-I)

In STBDF-I and -II [42], the weighted matrix \mathbf{W} in the linear observational model is simply taken as an equal-weight matrix, which means each LSHTR pixel in $y^{(t_0)}$ is the average of all HSLTR pixels

$\mathbf{z}^{(t_0)}$ within it. Although this may work well in some situations, it is actually a rough estimation that should comprise at least two parts: one image filtering matrix representing the blur PSF and one image down-sampling matrix. However, it is difficult to obtain the theoretical expression of \mathbf{W} because it may vary among different sensor images. Thus, the first approach is to employ the unmixing method to downscale LSHTR pixels and derive the mean spectral signature of endmembers, which will be treated as prior spectra. To make it more suitable for heterogeneous landscapes, the evolutionary temporal information expressed in the joint Gaussian distribution is combined with the prior information under the Bayesian framework to avoid the homogeneous spot in the unmixing-based algorithms.

1. Prior Distribution

To better utilize the spatial information in $\mathbf{y}^{(t_0)}$, the pixels in $\mathbf{y}^{(t_0)}$ are unmixed to get the mean spectra of each class assigned as the prior mean $\tilde{\boldsymbol{\mu}}_{\mathbf{z}^{(t_0)}}$ of pixels in the HSLTR image $\mathbf{z}^{(t_0)}$. The unmixing is performed by solving the linear mixing model in the n -by- n moving window with size w for the LSHTR image:

$$\mathbf{y}^{(t_0)} = \mathbf{V}\mathbf{s}^{(t_0)} + \mathbf{e}^{(t_0)} \quad (8)$$

where \mathbf{V} is the abundance map derived from the classification map with K clusters of HSLTR images, and $\mathbf{s}^{(t_0)}$ are the spectral values of the endmembers in $\mathbf{z}^{(t_0)}$. To avoid having unrealistic values of the endmembers, $\mathbf{s}^{(t_0)}$ is assumed to follow a prior Gaussian distribution with mean $\boldsymbol{\mu}_{\mathbf{s}^{(t_0)}}$ and covariance matrix $\mathbf{C}_{\mathbf{s}^{(t_0)}}$. Its probability density function is:

$$p(\mathbf{s}^{(t_0)}) = \frac{1}{\sqrt{(2\pi)^{NK} |\mathbf{C}_{\mathbf{s}^{(t_0)}}|}} \exp\left\{-\frac{1}{2}(\mathbf{s}^{(t_0)} - \boldsymbol{\mu}_{\mathbf{s}^{(t_0)}})^T \mathbf{C}_{\mathbf{s}^{(t_0)}}^{-1} (\mathbf{s}^{(t_0)} - \boldsymbol{\mu}_{\mathbf{s}^{(t_0)}})\right\} \quad (9)$$

The term $\mathbf{e}^{(t_0)}$ is the observational error introduced during the acquisition of the LSHTR image, which is assumed to be a Gaussian random vector with zero mean and covariance matrix $\mathbf{C}_{\mathbf{e}^{(t_0)}}$, i.e.,

$$p(\mathbf{e}^{(t_0)}) = \frac{1}{\sqrt{(2\pi)^{MK} |\mathbf{C}_{\mathbf{e}^{(t_0)}}|}} \exp\left\{-\frac{1}{2}\mathbf{e}^{(t_0)T} \mathbf{C}_{\mathbf{e}^{(t_0)}}^{-1} \mathbf{e}^{(t_0)}\right\} \quad (10)$$

Then, we derive:

$$p(\mathbf{y}^{(t_0)}|\mathbf{s}^{(t_0)}) = \frac{1}{\sqrt{(2\pi)^{MK} |\mathbf{C}_{\mathbf{e}^{(t_0)}}|}} \exp\left\{-\frac{1}{2}(\mathbf{y}^{(t_0)} - \mathbf{V}\mathbf{s}^{(t_0)})^T \mathbf{C}_{\mathbf{e}^{(t_0)}}^{-1} (\mathbf{y}^{(t_0)} - \mathbf{V}\mathbf{s}^{(t_0)})\right\} \quad (11)$$

Based on the Bayes rule, the posterior distribution $p(\mathbf{s}^{(t_0)}|\mathbf{y}^{(t_0)})$ also follows a Gaussian distribution, where the posterior mean can be derived as the MAP estimator of the endmember spectra:

$$\hat{\mathbf{s}}^{(t_0)} = [\mathbf{V}^T \mathbf{C}_{\mathbf{e}^{(t_0)}}^{-1} \mathbf{V} + \mathbf{C}_{\mathbf{s}^{(t_0)}}^{-1}]^{-1} [\mathbf{V}^T \mathbf{C}_{\mathbf{e}^{(t_0)}}^{-1} \mathbf{y}^{(t_0)} + \mathbf{C}_{\mathbf{s}^{(t_0)}}^{-1} \boldsymbol{\mu}_{\mathbf{s}^{(t_0)}}] \quad (12)$$

In Equation (12), $\boldsymbol{\mu}_{\mathbf{s}^{(t_0)}}$ is selected among the LSHTR pixels of the moving window with the highest abundance levels for each endmember. For simplicity, we define $\mathbf{C}_{\mathbf{e}^{(t_0)}}$ and $\mathbf{C}_{\mathbf{s}^{(t_0)}}$ as spherical covariance matrices with variances $\sigma_{\mathbf{e}^{(t_0)}}^2$ and $\sigma_{\mathbf{s}^{(t_0)}}^2$, respectively, which are free parameters that can be used to control the relative importance of the prior endmember spectra and the unmixed endmember spectra from the LSHTR image. The prior covariance matrix $\tilde{\mathbf{C}}_{\mathbf{z}^{(t_0)}}$ of pixels in $\mathbf{z}^{(t_0)}$ is used to balance the effectiveness of information from the prior part and the joint distribution part together with the joint covariance matrix. It is similarly defined as $\mathbf{C}_{\mathbf{e}^{(t_0)}}$ and $\mathbf{C}_{\mathbf{s}^{(t_0)}}$ with variance $\tilde{\sigma}_{\mathbf{z}^{(t_0)}}^2$. For each endmember in the unmixing procedure, if there are more than 80% pixels within a window with abundance levels smaller than 0.01, we will discard it for the unmixing only in this window, while in

Gevaert and García-Haro [50] they discarded endmembers with abundance levels not exceeding 0.1 (0.05 in Amorós-López et al. [22]) in at least one neighboring LSHTR pixel. As a result, the value of the LSHTR pixel with the largest abundance level for the discarded endmember is used to replace the spectra of the discarded class endmember in the central LSHTR pixel of each moving window when the posterior mean $\hat{\mathbf{s}}^{(t_0)}$ of the endmember spectra is designed. Finally, the prior mean $\tilde{\boldsymbol{\mu}}_{\mathbf{z}^{(t_0)}}$ of $\mathbf{z}^{(t_0)}$ is generated based on the classification map.

2. Joint Distribution

In addition to the prior information obtained from the LSHTR image at the target date, the multivariate joint Gaussian distribution is employed to model the time dependency and evolution of pixels in the HSLTR images, that is, \mathbf{X} and $\mathbf{z}^{(t_0)}$ are jointly Gaussian-distributed. In contrast to the STBDF-I and -II [42], here we generate the conditional probability density function of \mathbf{X} given $\mathbf{z}^{(t_0)}$ alternatively in the form of $p(\mathbf{X}|\mathbf{z}^{(t_0)})$, which is also a Gaussian distribution:

$$p(\mathbf{x}^{(t_1)}, \dots, \mathbf{x}^{(t_s)}|\mathbf{z}^{(t_0)}) = p(\mathbf{X}|\mathbf{z}^{(t_0)}) = \frac{1}{\sqrt{(2\pi)^{NK}|\mathbf{C}_{\mathbf{X}|\mathbf{z}^{(t_0)}}|}} \exp\left\{-\frac{1}{2}(\mathbf{z}^{(t_0)} - \boldsymbol{\mu}_{\mathbf{X}|\mathbf{z}^{(t_0)}})^T \mathbf{C}_{\mathbf{X}|\mathbf{z}^{(t_0)}}^{-1} (\mathbf{z}^{(t_0)} - \boldsymbol{\mu}_{\mathbf{X}|\mathbf{z}^{(t_0)}})\right\} \quad (13)$$

and

$$\boldsymbol{\mu}_{\mathbf{X}|\mathbf{z}^{(t_0)}} = E\{\mathbf{X}\} + \mathbf{C}_{\mathbf{X},\mathbf{z}^{(t_0)}} \mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{z}^{(t_0)}}^{-1} (\mathbf{z}^{(t_0)} - E\{\mathbf{z}^{(t_0)}\}) \quad (14)$$

$$\mathbf{C}_{\mathbf{X}|\mathbf{z}^{(t_0)}} = \mathbf{C}_{\mathbf{X},\mathbf{X}} - \mathbf{C}_{\mathbf{X},\mathbf{z}^{(t_0)}} \mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{z}^{(t_0)}}^{-1} \mathbf{C}_{\mathbf{X},\mathbf{z}^{(t_0)}}^T \quad (15)$$

where $E\{\mathbf{X}\}$ and $E\{\mathbf{z}^{(t_0)}\}$ are the means of \mathbf{X} and $\mathbf{z}^{(t_0)}$, respectively, and these matrices $\mathbf{C}_{\mathbf{X},\mathbf{X}}$, $\mathbf{C}_{\mathbf{X},\mathbf{z}^{(t_0)}}$, and $\mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{z}^{(t_0)}}$ are the cross-covariance matrices of the HSLTR images, which can be similarly estimated according to Xue et al. [42]. Denote $\mathbf{A} = \mathbf{C}_{\mathbf{X},\mathbf{z}^{(t_0)}} \mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{z}^{(t_0)}}^{-1}$ and $\mathbf{B} = E\{\mathbf{X}\} - \mathbf{C}_{\mathbf{X},\mathbf{z}^{(t_0)}} \mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{z}^{(t_0)}}^{-1} E\{\mathbf{z}^{(t_0)}\}$, then:

$$\boldsymbol{\mu}_{\mathbf{X}|\mathbf{z}^{(t_0)}} = \mathbf{A}\mathbf{z}^{(t_0)} + \mathbf{B} \quad (16)$$

$$\mathbf{C}_{\mathbf{X}|\mathbf{z}^{(t_0)}} = \mathbf{C}_{\mathbf{X},\mathbf{X}} - \mathbf{A}\mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{z}^{(t_0)}}^T \quad (17)$$

3. MAP Estimation

In this study, we need to provide another form of the objective function of the MAP estimation based on the Bayes rule because the linear mixing model gives the relationship of pixels in $\mathbf{y}^{(t_0)}$ and the prior spectral signature for each endmember of $\mathbf{z}^{(t_0)}$, which are not the pixel values in $\mathbf{z}^{(t_0)}$. This is different from the format of a joint distribution, which deals with the relationship of pixels directly. Therefore, the objective function takes on the following expression:

$$\hat{\mathbf{z}}^{(t_0)} = \underset{\mathbf{z}^{(t_0)}}{\operatorname{argmax}} p(\mathbf{z}^{(t_0)}|\mathbf{y}^{(t_0)}, \mathbf{X}) = \underset{\mathbf{z}^{(t_0)}}{\operatorname{argmax}} p(\mathbf{X}|\mathbf{z}^{(t_0)}) \cdot p(\mathbf{z}^{(t_0)}|\mathbf{y}^{(t_0)}) \quad (18)$$

Finally, the solution is obtained according to the Bayes' theorem [55]:

$$\hat{\mathbf{z}}^{(t_0)} = [\tilde{\mathbf{C}}_{\mathbf{z}^{(t_0)}}^{-1} + \mathbf{A}^T \mathbf{C}_{\mathbf{X}|\mathbf{z}^{(t_0)}}^{-1} \mathbf{A}]^{-1} [\tilde{\mathbf{C}}_{\mathbf{z}^{(t_0)}}^{-1} \tilde{\boldsymbol{\mu}}_{\mathbf{z}^{(t_0)}} + \mathbf{A}^T \mathbf{C}_{\mathbf{X}|\mathbf{z}^{(t_0)}}^{-1} (\mathbf{X} - \mathbf{B})] \quad (19)$$

$$\hat{\mathbf{z}}^{(t_0)} = \tilde{\boldsymbol{\mu}}_{\mathbf{z}^{(t_0)}} + \tilde{\mathbf{C}}_{\mathbf{z}^{(t_0)}} \mathbf{A}^T [\mathbf{A} \tilde{\mathbf{C}}_{\mathbf{z}^{(t_0)}} \mathbf{A}^T + \mathbf{C}_{\mathbf{X}|\mathbf{z}^{(t_0)}}]^{-1} [\mathbf{X} - \mathbf{B} - \mathbf{A} \tilde{\boldsymbol{\mu}}_{\mathbf{z}^{(t_0)}}] \quad (20)$$

where

$$\boldsymbol{\mu}_{\mathbf{X}|\mathbf{z}^{(t_0)}} = [\boldsymbol{\mu}_{\mathbf{X}_1|\mathbf{z}_1^{(t_0)}}^T, \boldsymbol{\mu}_{\mathbf{X}_2|\mathbf{z}_2^{(t_0)}}^T, \dots, \boldsymbol{\mu}_{\mathbf{X}_N|\mathbf{z}_N^{(t_0)}}^T]^T \quad (21)$$

$$C_{X|Z^{(t_0)}} = \begin{bmatrix} C_{X_1|Z_1^{(t_0)}} & 0 & \dots & 0 \\ 0 & C_{X_2|Z_2^{(t_0)}} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & C_{X_N|Z_N^{(t_0)}} \end{bmatrix} \tag{22}$$

The pixels at different locations in the HSLTR image to be predicted are independently estimated given the before-and-after images in the time-series. Thus, in ISTBDF-I, the conditional mean vector $\mu_{z^{(t_0)}|X}$ and covariance matrix $C_{z^{(t_0)}|X}$ can be estimated for each individual pixel as:

$$\mu_{X_n|z_n^{(t_0)}} = E\{X_n\} + C_{X_n,z_n^{(t_0)}} C_{z_n^{(t_0)},z_n^{(t_0)}}^{-1} (z_n^{(t_0)} - E\{z_n^{(t_0)}\}) = A_n z_n^{(t_0)} + B_n \tag{23}$$

$$C_{X_n|z_n^{(t_0)}} = C_{X_n,X_n} - C_{X_n,z_n^{(t_0)}} C_{z_n^{(t_0)},z_n^{(t_0)}}^{-1} C_{z_n^{(t_0)},X_n}^T = C_{X_n,X_n} - A_n C_{X_n,z_n^{(t_0)}}^T \tag{24}$$

where $A_n = C_{X_n,z_n^{(t_0)}} C_{z_n^{(t_0)},z_n^{(t_0)}}^{-1}$, $B_n = E\{X_n\} - C_{X_n,z_n^{(t_0)}} C_{z_n^{(t_0)},z_n^{(t_0)}}^{-1} E\{z_n^{(t_0)}\}$ and $X_n = [(x_n^{(t_1)})^T, (x_n^{(t_2)})^T, \dots, (x_n^{(t_s)})^T]^T, n = 1, \dots, N$.

2.2.3. Improved Spatio-Temporal Bayesian Data Fusion Method II (ISTBDF-II)

In STBDF-I and -II [42], the mean vector of the multivariate joint Gaussian distribution, which is used to measure the correlation and change tendency of the HSLTR images, is estimated by combining the bilinear interpolation of the corresponding LSHTR images to provide global fluctuations together with the high-pass frequencies of the nearest high-resolution images to account for the local spatial details. The global pattern of the HSLTR image on the prediction date is approximated by interpolating the LSHTR image at the same time, which may lead to a smooth estimation and overlook the information in the mixed pixels of the LSHTR image when the study region is complex and heterogeneous. Even though the local details obtained from the HSLTR images before-and-after the image at the target date are introduced to replace the unknown spatial details of the image to be predicted, they are not the same as an extension, which depends on temporal changes. Thus, to improve the performance of the existing Bayesian method in more complex heterogeneous areas, our second approach is developed by incorporating the pixel unmixing algorithm to estimate the mean vector of the joint distribution rather than by resampling with high frequencies. This is the second contribution of our proposed method in this study. Similar to STBDF-I and -II [42], the MAP estimator of the HSLTR image at the target date can be expressed as:

$$\hat{z}^{(t_0)} = \mu_{z^{(t_0)}|X} + C_{z^{(t_0)}|X} W^T [W C_{z^{(t_0)}|X} W^T + C_{e^{(t_0)}}]^{-1} [y^{(t_0)} - W \mu_{z^{(t_0)}|X}] \tag{25}$$

where

$$\mu_{z^{(t_0)}|X} = E\{z^{(t_0)}\} + C_{z^{(t_0)},X} C_{X,X}^{-1} (X - E\{X\}) \tag{26}$$

$$C_{z^{(t_0)}|X} = C_{z^{(t_0)},z^{(t_0)}} - C_{z^{(t_0)},X} C_{X,X}^{-1} C_{z^{(t_0)},X}^T \tag{27}$$

The aforementioned $\mu_{z^{(t_0)}}$ obtained by the unmixing algorithm in Equation (12) is used to estimate the mean vectors $E\{z^{(t_0)}\}$ in Equation (26), and $E\{X\}$ is estimated in the same way. The covariance matrix $C_{z^{(t_0)}|X}$ is similarly estimated as that in STBDF-I and -II. In addition to the joint distribution, the other parts that are the first-order observational model and the objective MAP function remain unchanged in ISTBDF-II.

3. Experimental Tests

We first describe the data, simulated and real-life satellite data, in our experiments. Similar to STBDF-I and -II, ISTBDF-I and -II are able to work on one pair of HSLTR and LSHTR images as inputs. Here in this study we used two pairs of images. Specifically, the algorithms predict the HSLTR images on the target date using the LSHTR image of the target date and two pairs of HSLTR and LSHTR images acquired at the nearest dates before and after the target date.

We then introduce the metrics used to quantitatively evaluate the performance of our methods in comparison with STBDF-I and -II. We only need to compare with STBDF-II because it has been shown to perform better than the other methods in Xue et al. [42].

3.1. Data and Experimental Design

The two proposed algorithms ISTBDF-I and -II were tested on both simulated and real-life satellite images. For illustration, simulated data were used to avoid the interference of radiometric and geometric inconsistencies between the HSLTR and LSHTR images obtained from satellite sensors. ISTBDF-I and -II are developed to produce improved fusion results over heterogeneous areas in which pixels in the LSHTR images often contain mixed land covers. It was expected that the improvement of ISTBDF-I and -II over STBDF-II becomes substantial as the heterogeneity of the target landscape increases. Given this hypothesis, three cases with increasing heterogeneity were designed.

In Simulated Experiment I, a water pool (circle) was placed in the middle of a vegetated land, as shown in Figure 2. This case contains three subcases, all with fixed shapes for water and vegetation covers. In the three subcases, we assumed that the reflectance of water remained unchanged over the observation period. Table 1 gives the phenology dynamics, the reflectance of vegetation changed over time but at different rates in the three subcases. Most of the LSHTR pixels in Experiment I were homogeneous and the heterogeneous pixels were distributed mainly around the edge of the water pool. The HSLTR images in the three subcases had 150×150 pixels. The reflectance value for each pixel was filled according to its land cover type and timing, as shown in Table 1. It should be noted that the reflectance changed over time. For example, the HSLTR images for subcase 1 are shown in Figure 2a–c. A stochastic perturbation was then added to the images using random values generated by a Gaussian distribution with mean 0 and standard deviation 0.001. The images were then downscaled to a lower resolution of 10×10 pixels. Figure 2d–f presents LSHTR images in subcase 1 as an example.

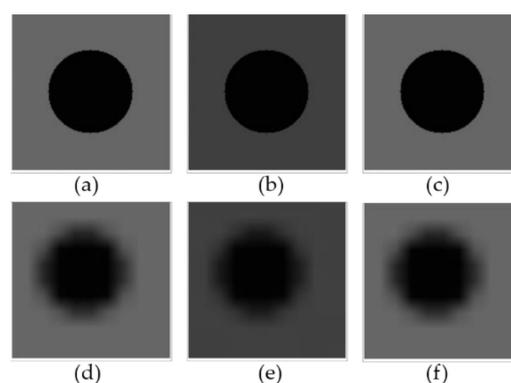


Figure 2. (a–c) Simulated HSLTR images for times t_0 , t_1 , and t_2 , respectively; and (d–f) LSHTR images for times t_0 , t_1 , and t_2 , respectively.

Simulated Experiment II was constructed in a fashion similar to that of Experiment I except that a soil cover was added to increase the heterogeneity of the landscape. The reflectance of the water body (small circle) and the soil (surrounding) were assumed to remain unchanged over the observation period, while the reflectance of vegetation (big circle) varied at different rates. The reflectance value

for each land cover type at different points in time are shown in Table 2. Their shapes were kept unchanged. Figure 3 shows subcase 1 as an example.

Table 1. Temporal Changes of the Reflectance for Three Simulated Subcases in Experiment I.

	Subcase 1		Subcase 2		Subcase 3	
	Water	Vegetation	Water	Vegetation	Water	Vegetation
t_0	0.01	0.4	0.03	0.02	0.06	0.2
t_1	0.01	0.25	0.03	0.15	0.06	0.15
t_2	0.01	0.4	0.03	0.35	0.06	0.1

Table 2. Temporal Changes of Reflectance for Three Simulated Subcases in Experiment II.

	Subcase 1			Subcase 2			Subcase 3		
	Soil	Water	Vegetation	Soil	Water	Vegetation	Soil	Water	Vegetation
t_0	0.3	0.01	0.4	0.2	0.03	0.02	0.15	0.06	0.2
t_1	0.3	0.01	0.25	0.2	0.03	0.15	0.15	0.06	0.15
t_2	0.3	0.01	0.4	0.2	0.03	0.35	0.15	0.06	0.1

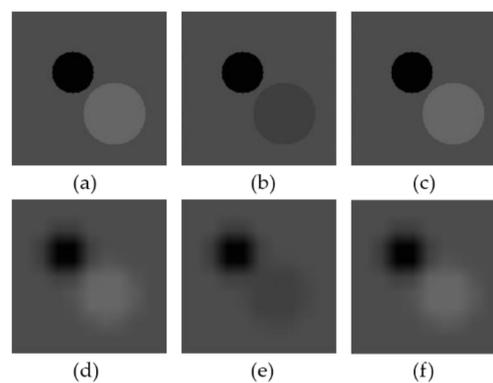


Figure 3. (a–c) Simulated HSLTR images for times t_0 , t_1 , and t_2 , respectively; (d–f) LSHTTR images for times t_0 , t_1 , and t_2 , respectively.

To increase the heterogeneity and to be close to real cases, Simulated Experiment III employed images generated from satellite images. Specifically, we used real Landsat data as the HSLTR images. The LSHTTR images were simulated by aggregating pixels in the Landsat images, termed simulated-LSHTTR images. This way of constructing experiments was similar to most studies on the spatio-temporal fusion of satellite images [2,5]. To match our motivation, a heterogeneous landscape that covered an area of 10 km \times 10 km in southern New South Wales, Australia (145.0674°E, 34.0034°S) was used in this study [46]. For the study area, we acquired Landsat images on three dates: 3 December 2001, 4 January 2002, and 11 January 2002. Each Landsat image contains 200 \times 200 pixels with a 25-m resolution that were extracted from the original image. We focus on NIR, red, and green bands. The dominant land covers of the study area are green crops, including irrigated rice croplands, agricultural drylands, and woodlands. The parcels are generally small with irregular shapes, highlighting the heterogeneity of the landscape. Compared to December 3, 2001, it can be observed that the phenology on 4 and 11 January 2002, are closer to each other. Examining the temporal dynamics, we can see that the phenology of irrigated croplands changed over a single summer growing season, while the surrounding drylands and woodlands were less variable throughout [46]. The NIR–red–green composites of the subset of Landsat images and the corresponding simulated-LSHTTR images are shown in Figure 4.

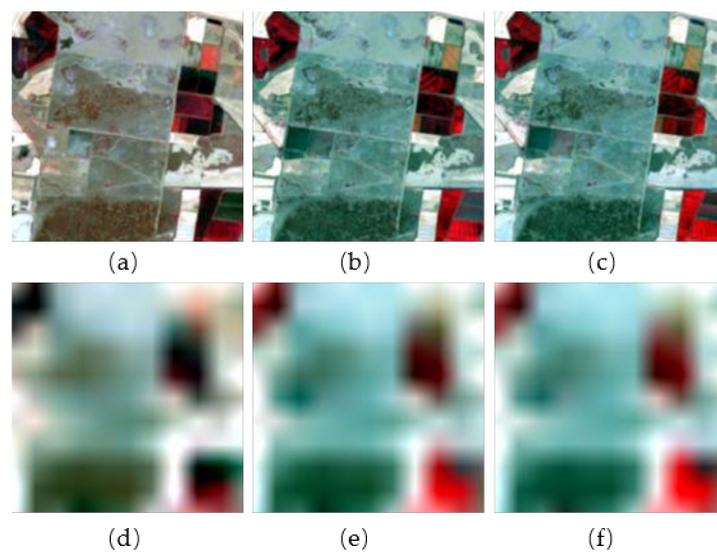


Figure 4. NIR–red–green composites of the Landsat imagery (top row) and the simulated-LSHTR imagery (bottom row) taken on (a), (d) 3 December 2001; (b), (e) 4 January 2002; and (c), (f) 11 January 2002.

In addition to the three simulated experiments, we conducted a real-case experiment using satellite images. We used the same set of Landsat images as Simulated Experiment III. However, we employed a larger subset of 400×400 pixels extracted from the full images. Furthermore, for LSHTR images, we used real MODIS images containing 20×20 pixels with a 500-m resolution instead of the simulated ones from Landsat images. The NIR–red–green composites of Landsat and MODIS images for real-case experiments are shown in Figure 5.

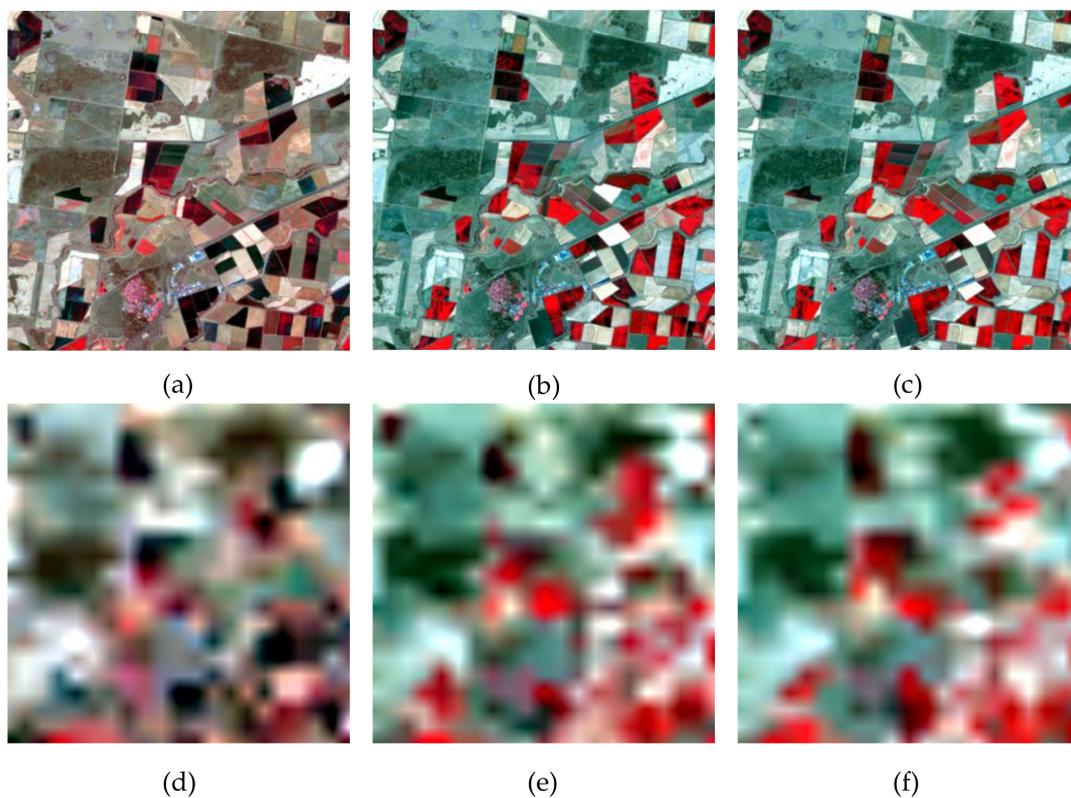


Figure 5. NIR–red–green composites of the Landsat imagery (top row) and MODIS imagery (bottom row) taken on (a), (d) December 3, 2001; (b), (e) 4 January 2002; and (c), (f) 11 January 2002 [46].

3.2. Metrics for Model Assessment

In addition to the direct interpretation of the fusion quality through visual comparison, five quantitative metrics were employed to evaluate the performance of the proposed ISTBDF-I and -II together with the original STBDF-II. The absolute average difference (AAD) was used to measure the overall bias of the fused image. The root mean square error (RMSE) was employed to measure the difference between the fused and reference image. The peak signal-to-noise ratio (PSNR) was adopted to evaluate the spatial reconstruction quality of the fused image. The correlation coefficient (CC) was used to measure the linear relationship between the fused and reference images. The universal image quality index (UIQI) was employed to model the image distortion in three aspects: loss of correlation, luminance distortion, and contrast distortion [56]. Furthermore, the Erreur Relative Global Adimensionnelle de Synthèse (ERGAS) was employed to measure the similarity between the fused and the reference image by combining all three bands together [47,50].

4. Experimental Results and Interpretations

In this section, our proposed fusion algorithms ISTBDF-I and -II are compared with STBDF-II using the aforementioned simulated images and real-life satellite images as inputs. The fused HSLTR images from the three fusion methods are visually compared to the reference image. The quantitative assessment is also made using the metrics introduced in Section 3.2.

4.1. Experiments with Simulated Images

For subcase 1 in Simulated Experiment I, the high-resolution images at t_0 (Figure 2a) and t_2 (Figure 2c), and LSHTR images at t_0 , t_1 , and t_2 (Figure 2d–f) were fused to generate the HSLTR image at t_1 . Fused images (Figure 6b–d) obtained by all three algorithms are similar to the reference image (Figure 6a), which is the same for the other two subcases of the experiment. As shown in Table 3, they also exhibit relatively good performance from the quantitative perspective. In terms of all metrics used, ISTBDF-I and -II outperform STBDF-II in all three subcases while ISTBDF-I exhibits a slightly better skill than ISTBDF-II. It should be pointed out that the improvement of the two proposed methods over STBDF-II is not substantial, which may be due to the homogeneity over most areas.

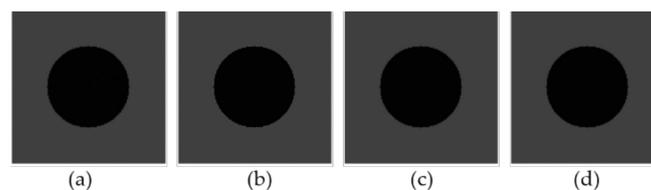


Figure 6. The (a) reference image at t_1 and its predictions using (b) STBDF-II, (c) ISTBDF-I, and (d) ISTBDF-II for subcase 1 in Simulated Experiment I.

Table 3. Quantitative Metrics of the Fusion Results of Simulated Experiment I.

Subcase	AAD				RMSE			PSNR	
	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I (10 ⁻⁴)	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II
1	0.0019	0.0008	0.0009	0.0067	9.9288	0.0011	31.5955	48.1588	47.1330
2	0.0011	0.0008	0.0009	0.0019	9.9287	0.0011	37.9822	43.8126	43.0206
3	0.0010	0.0008	0.0010	0.0012	9.9290	0.0012	42.0637	43.8123	42.0265
Subcase	CC				UIQI		ERGAS		
	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	
1	0.9979	1.0000	0.9999	0.9978	1.0000	0.9999	0.2269	0.0337	0.0379
2	0.9993	0.9998	0.9998	0.9992	0.9998	0.9998	0.1051	0.0537	0.0589
3	0.9995	0.9996	0.9995	0.9995	0.9996	0.9995	0.0623	0.0510	0.0626

Similar to Simulated Experiment I, the HSLTR image at time t_1 for the three subcases of Simulated Experiment II were predicted. Furthermore, subcase 1 was taken as an illustration, as shown in Figure 7b–d. Through visual comparison of the fused images obtained using the three algorithms with the reference image, it is clear that they all capture the spatial details of vegetation phenology changes, implying that all three algorithms can generally take advantage of the temporal information in the LSHTR images to generate the target HSLTR image. ISTBDF-I and -II produce images with clearer boundaries and sharper features between water, soil, and vegetation (Figure 7c,d). Meanwhile the image predicted by STBDF-II have the “LSHTR-image boundary” effect and a few gray patches around them (Figure 7b), which is not observed in the reference image (Figure 7a). This may be caused by the large difference in values of the vegetation reflectances between the image at the target date (0.25) and the input images before and after (0.4). Another reason is that the reflectance of the LSHTR pixels varies widely from their surrounding LSHTR pixels and the resampled data or the equal-weight matrix used in STBDF-II method leads to the uniform spectral reflectance within the LSHTR pixels. This example highlights the importance of extracting adequate information from mixed LSHTR pixels.

Quantitative comparisons, shown in Table 4, are consistent with the visual comparisons. These metrics indicate that ISTBDF-I and -II provide more accurate predictions than STBDF-II, while ISTBDF-II exhibits the highest prediction capability with the smallest AAD, RMSE, and ERGAS, and the highest PSNR, CC, and UIQI. The improvement of ISTBDF-I and -II over STBDF-II generally becomes larger as the heterogeneity of the landscape increases from Experiment I to II. For instance, in terms of ERGAS, the accuracy of ISTBDF-II over STBDF-II increases by 83%, 44%, and 0% for the three subcases in Experiment I. Meanwhile this increase grows to 89%, 59%, and 52% as it comes to Experiment II. These results from Experiments I and II are consistent with our hypothesis that ISTBDF-I and -II have the advantage over STBDF-II in dealing with heterogeneous landscapes.

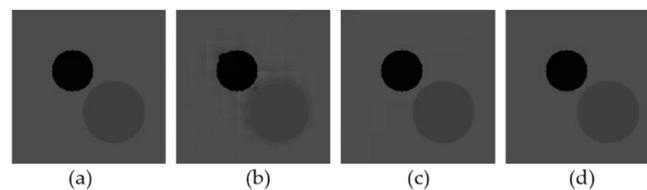


Figure 7. The (a) reference image at t_1 and its predictions using (b) STBDF-II, (c) ISTBDF-I, and (d) ISTBDF-II for subcase 1 in Simulated Experiment II.

Table 4. Quantitative Metrics of the Fusion Results in Simulated Experiment II.

Subcase	AAD			RMSE			PSNR		
	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II
1	0.0052	0.0023	0.0012	0.0134	0.0031	0.0015	27.1189	39.9119	46.1953
2	0.0014	0.0017	0.0011	0.0032	0.0022	0.0013	35.9727	39.5011	43.6164
3	0.0075	0.0079	0.0057	0.0178	0.0149	0.0086	18.7488	20.3180	25.0782
Subcase	CC			UIQI			ERGAS		
	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II
1	0.9827	0.9994	0.9998	0.9813	0.9994	0.9998	0.3222	0.0739	0.0358
2	0.9969	0.9992	0.9995	0.9968	0.9991	0.9995	0.1176	0.0783	0.0488
3	0.9228	0.9502	0.9502	0.7856	0.8498	0.9146	0.9860	0.8231	0.4758

For Simulated Experiment III, the Landsat image for 4 January 2002 (Figure 4b) was predicted by fusing two image-pairs from December 3, 2001 (Figure 4a,d), and 11 January 2002 (Figure 4c,f), together with the simulated-LSHTR image on January 04, 2002 (Figure 4e). The visual comparison between the predicted and reference images is shown in Figure 8. We clearly see that all three algorithms could capture many key features of the spatial pattern well in the reference image. A closer examination shows that the images predicted by ISTBDF-I (Figure 8c) and ISTBDF-II (Figure 8d) seem

to be more similar to the reference image (Figure 8a) than that obtained using STBDF-II (Figure 8b). This can be found by the color and spatial textures of the red polygons in the predicted images. The quantitative comparison in terms of the selected metrics is shown in Table 5. Both ISTBDF-I and ISTBDF-II show improved predictions than STBDF-II in terms of all six metrics, and ISTBDF-I performs better than ISTBDF-II. The relatively substantial improvement highlights the effectiveness of introducing an unmixing procedure into the Bayesian fusion algorithms in enhancing fusion results in heterogeneous landscapes.

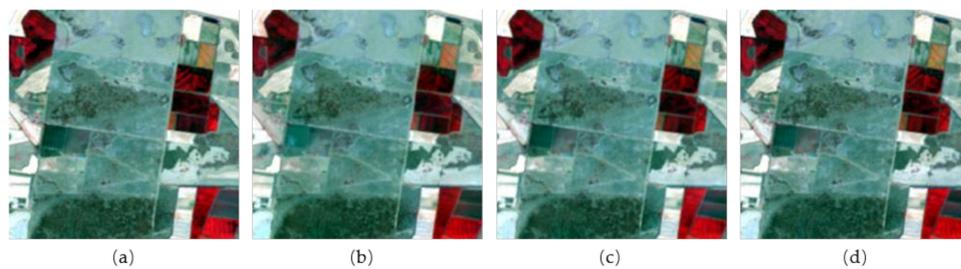


Figure 8. The (a) reference image at t_1 and its predictions using (b) STBDF-II, (c) ISTBDF-I, and (d) ISTBDF-II for Simulated Experiment III.

Table 5. Quantitative Metrics of the Fusion Results in Simulated Experiment III.

	AAD		RMSE				PSNR		
	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II
NIR	0.0113	0.0098	0.0106	0.0182	0.0151	0.0163	31.2201	32.8820	32.2143
Red	0.0090	0.0073	0.0083	0.0136	0.0109	0.0121	29.2283	31.1726	30.2273
Green	0.0064	0.0051	0.0058	0.0095	0.0073	0.0082	29.9955	32.2588	31.2124
Subcase	CC				UIQI		ERGAS		
	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II
NIR	0.9540	0.9686	0.9633	0.9513	0.9682	0.9629	0.4635	0.3689	0.4095
Red	0.9792	0.9856	0.9821	0.9758	0.9856	0.9820			
Green	0.9715	0.9823	0.9775	0.9680	0.9823	0.9774			

4.2. Experiments with Satellite Images

Two image pairs from 3 December 2001 (Figure 5a,d) and 11 January 2002 (Figure 5c,f), together with the MODIS image on 4 January 2002 (Figure 5e) were used to estimate the Landsat image on January 4, 2002 (Figure 5c), which will be compared with the reference image (Figure 5c). The optimal parameter sets were first chosen for our fusion algorithms for prediction.

4.2.1. Optimization of Parameters

A sensitivity analysis of ISTBDF-I and -II to two important parameters, i.e., the number of clusters (K) and the window size (w), were first conducted. In this study, K was set as 4, 8, 12, and 16 by the k-means method, while w changed from 11 to 21 with step size 2. Regarding the unmixing algorithm used in this study, another important parameter is the standard error ratio $\sigma_{\text{ratio}} = \sigma_{\mathbf{e}(t_0)} / \sigma_{\mathbf{s}(t_0)}$, where $\sigma_{\mathbf{e}(t_0)}$ is the variance of the prior endmember spectra, and $\sigma_{\mathbf{s}(t_0)}$ is the variance of the unmixed endmember spectra. This is used to weigh the unmixing endmember relative to the prior endmember, which was set to vary from 2 to 30 with step size 4. The optimal parameter set was determined based on ERGAS. This was done for ISTBDF-I and -II, separately.

The relationship between the parameters and ERGAS is illustrated in Figure 9. The results (Figure 9a,b) indicate that a small number of clusters (i.e., $K = 4$ or 8) generates better fusion results indicated by smaller ERGAS values, and reduces the sensitivity of the models to the variations of neighborhood size w for both ISTBDF-I and -II. Eight clusters for ISTBDF-I and four for ISTBDF-II were

preferred in this study. Employing a moving window can improve the fusion quality by introducing more spatial details. However, a higher window size may reduce the spectral variability of the fused image and lead to a smoother fused image. To keep spectral variability while at the same time preserving fusion quality, the optimal window sizes were set to 19 and 17, respectively. Furthermore, it is observed that the inclusion of a prior spectral information improves the fusion results. For instance, a lower ERGAS is observed when a higher error ratio is set, which means assigning a higher variance to prior information than that of measurements (Figure 9c,d). Based on the parameter sensitivity analyses, the optimal error ratio was set as 26 for both methods in this study (Figure 9c,d).

Besides optimal parameters, our algorithms need the classification map of the target Landsat image. It is found that the MODIS image on 4 January 2002, has a higher correlation with the image on 11 January 2002, than that on December 03, 2002. Based on the assumption that MODIS and Landsat data share a similar temporal evolution, the classification map on 11 January 2002, was used to represent the map on 4 January 2002. This assumption was further validated through the similarity between the two classification maps, which was tested using three indices [57]: the overall accuracy (oa), kappa statistic, and the average of the user's and producer's accuracy (aup). The classification is more accurate as the three indices are closer to 1. As shown in Table 6, regardless of the eight clusters for ISTBDF-I or four clusters for ISTBDF-II, the relatively high values of these indices give us the confidence in using the simulated classification map based on the Landsat image from 11 January 2002.

Furthermore, the computational complexity based on the running time were also compared among them. The three algorithms were written in MATLAB software (version R2018b by MathWorks, Inc, USA) and run on a computer with Intel(R) Core(TM) i7-7500 processor and the 8GB random access memory (RAM). The running time for STBDF-II, ISTBDF-I, and ISTBDF-II with their optimal parameters are around 10 seconds, 120 seconds, and 200 seconds for this experiment with data of size of 400×400 per band. All of them are quite computationally efficient, though ISTBDF-I and -II need more running time than STBDF-II.

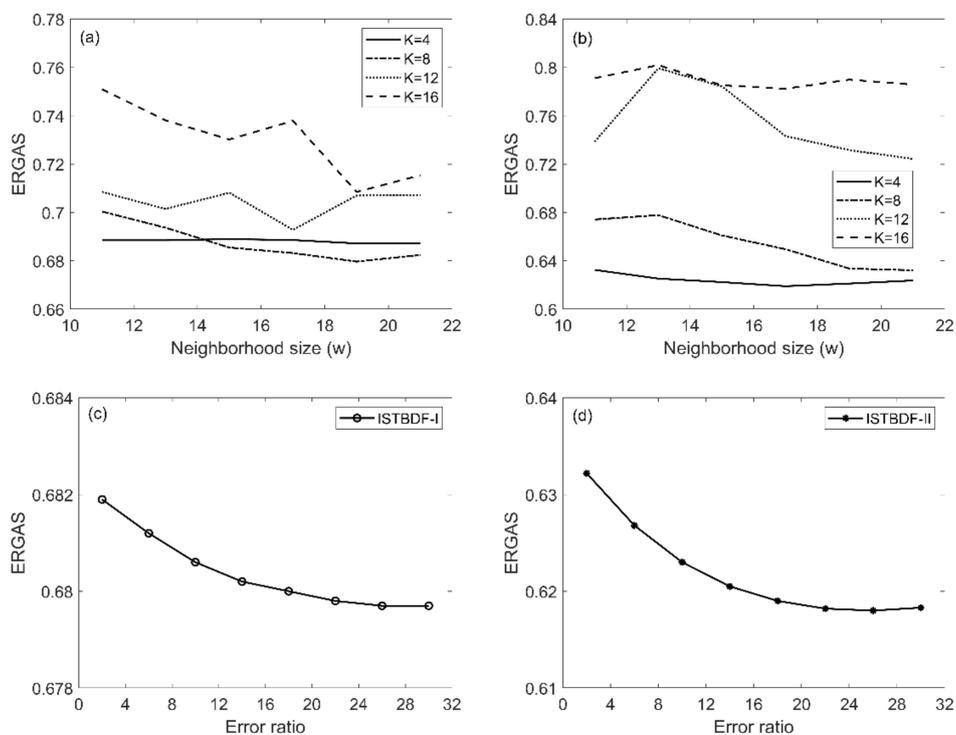


Figure 9. Changes of the ERGAS index with respect to moving window size (w) and number of clusters (K) for (a) ISTBDF-I and (b) ISTBDF-II, both with a fixed error ratio 26, and the changes of the ERGAS index with respect to the variation of the error ratio for (c) ISTBDF-I with a fixed $w = 19$, $K = 8$ and (d) ISTBDF-II with a fixed $w = 17$, $K = 4$.

Table 6. Classification Accuracy for ISTBDF-I with Eight Clusters and ISTBDF-II with Four Clusters in Terms of Overall Accuracy, Kappa Statistic, and Average of User’s and Producer’s Accuracy.

	oa	kappa		aup						
ISTBDF-I	0.6932	0.6349	0.5851	0.8568	0.7417	0.7979	0.7397	0.6478	0.7480	0.6600
ISTBDF-II	0.8328	0.7648	0.8149		0.8645		0.9052		0.8207	

4.2.2. Comparison of Fusion Results and Discussion

Figure 10 shows the reference Landsat image on 4 January 2002, and three corresponding fused images predicted by STBDF-II, ISTBDF-I, and ISTBDF-II with their optimal parameters, respectively. By visual comparison, the three predicted images are in general similar to the reference Landsat image. This indicates that all three methods capture the main features of the reference image in terms of visual comparison and it implies that all three methods could successfully predict the temporal change of croplands from 3 December to 11 January 2002. However, the fused image of STBDF-II [Figure 10b] appears to be hazy and blurry, and it seems to be covered by a layer of fog with color distortion. Compared with STBDF-II, ISTBDF-I and -II show improved skill in predicting the temporal change of reflectance and capturing the texture of small objects with clearer boundaries. Figure 11 further highlights the contrast of the three algorithms using zoomed-in images. The red “stain” in the middle of the STBDF-II-predicted image (Figure 11b) is not observed in the reference image (Figure 11a) and the other two predicted images (Figure 11c,d). ISTBDF-II seems to provide slightly better prediction than ISTBDF-I (Figure 11c,d). In addition to visual comparison, we show the scatter plots of the reference and predicted reflectance for pixels corresponding to the red “stain” in red dots in Figure 12. For both ISTBDF-I and -II, the red dots are generally closer to the “1:1” line than STBDF-II, suggesting better predictions. In particular, STBDF-II underestimates the reflectance of the red “stain” in the red band (Figure 12d), which is not seen in ISTBDF-I and -II (Figure 12e,f). RMSE between STBDF-II predicted and reference reflectance for the red “stain” in red band is 0.7863, which drops to 0.4385 and 0.3453 for ISTBDF-I and -II, respectively.

Figure 12 depicts the scatter plots between the reflectance of the reference image and the fused images generated by the three methods, respectively. It can be clearly observed that the stripes of scattered points generated by ISTBDF-II are narrowest and closest to the diagonal 1:1 line for all three bands, followed by ISTBDF-I and the STBDF-II. The result suggests that the reflectance of ISTBDF-II is closest to the reflectance of the reference image, which is consistent with Figure 10. In addition, the quantitative indices given in Table 7 indicate results similar to the visual comparisons and the scatter plots. For all three bands, the fused images of ISTBDF-I and -II have smaller AAD and RMSE, and higher PSNR, CC, and UIQI than STBDF-II, implying a more accurate prediction than STBDF-II. ISTBDF-II shows the highest prediction skill in terms of these indices. The ERGAS index draws a conclusion similar to STBDF-II (0.7279), ISTBDF-I (0.6797), and ISTBDF-II (0.6180). To conclude, even though various patches in Landsat are smaller than one MODIS pixel in the heterogeneous landscape, ISTBDF-I and -II show convincingly better skill in capturing these fine spatial details through converting the reflectance changes of the mixed MODIS pixel into the corresponding Landsat pixels.

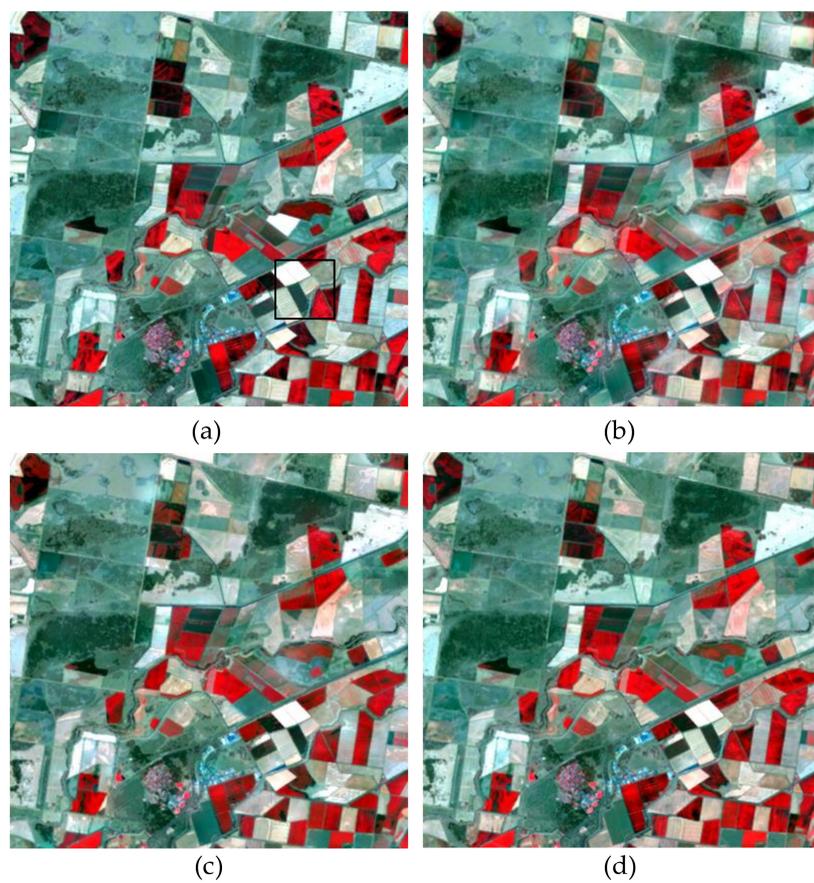


Figure 10. The (a) reference image acquired on 4 January 2002, and its predictions using (b) STBDF-II, (c) ISTBDF-I, and (d) ISTBDF-II.

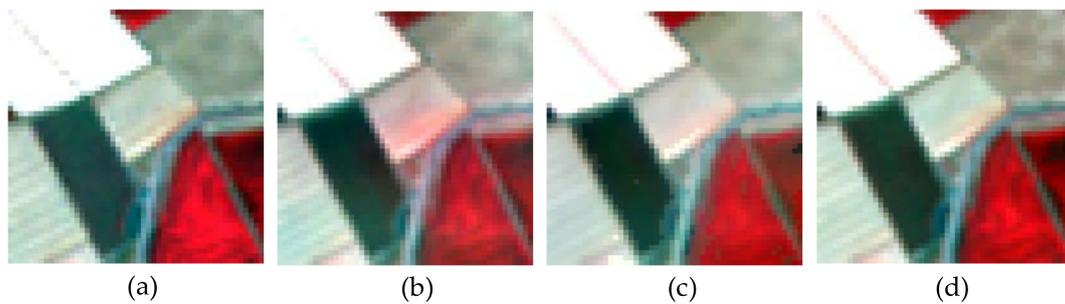


Figure 11. Zoomed-in images of the area marked in Figure 5a corresponding to the images in Figure 8a–d.

Table 7. Quantitative Metrics of the Fusion Results Applied to Satellite Data.

	AAD			RMSE			PSNR		
	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II
NIR	0.0200	0.0178	0.0151	0.0289	0.0266	0.0244	27.2340	27.9362	28.6903
Red	0.0168	0.0152	0.0119	0.0233	0.0212	0.0191	24.5500	25.3725	26.2780
Green	0.0102	0.0100	0.0079	0.0142	0.0139	0.0127	29.1497	29.3766	30.1477
	CC			UIQI			ERGAS		
	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II	STBDF-II	ISTBDF-I	ISTBDF-II
NIR	0.9131	0.9268	0.9431	0.9016	0.9185	0.9304	0.7279	0.6797	0.6180
Red	0.9464	0.9567	0.9643	0.9420	0.9504	0.9614			
Green	0.9439	0.9478	0.9584	0.9398	0.9476	0.9525			

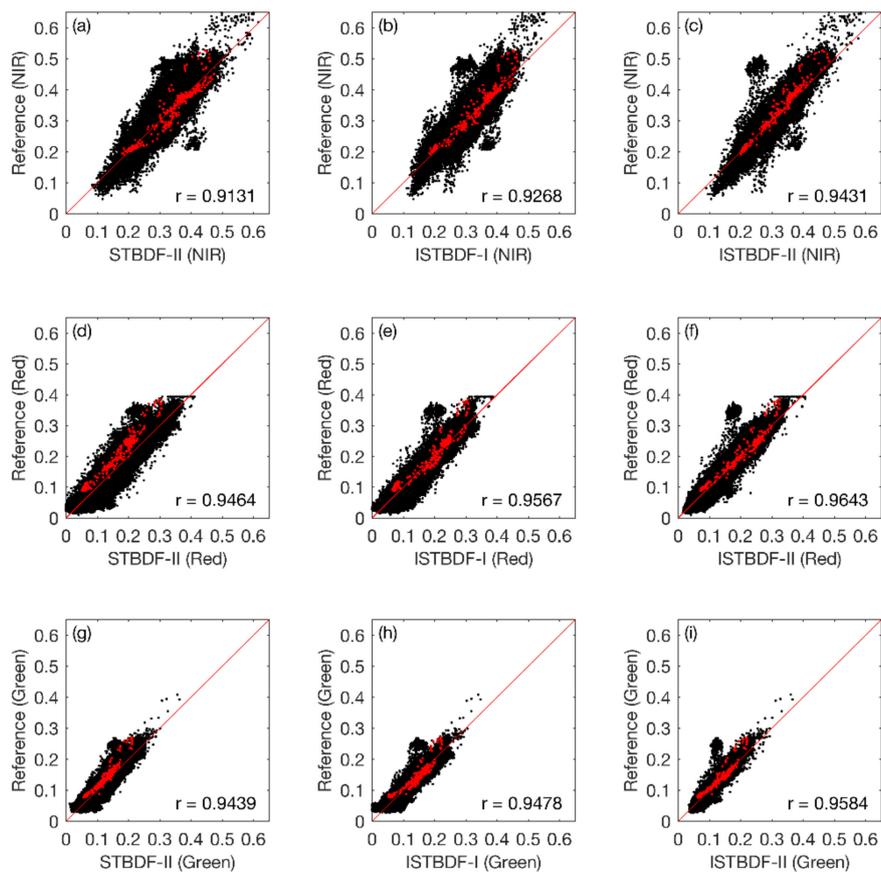


Figure 12. Scatter plots of the reference reflectance and predicted reflectance using STBDF-II (left column), ISTBDF-I (middle column), and ISTBDF-II (right column) for (a–c) the NIR band, (d–f) the red band, and (g–i) the green band. The red dots show pixels corresponding to “red stains” in the predicted image using STBDF-II in Figure 11.

5. Discussion

In this section, we provide more discussion on the strengths and limitations of our new approach.

Compared to the original STBDF-II, the significant improvement made by ISTBDF-I and ISTBDF-II over heterogeneous landscapes can be attributed to the strengths of the unmixing-based algorithm in two different ways. First, in STBDF-II, the linear observational model that is used to build the relationship between LSHTR and HSLTR images employs an equal-weight matrix. This matrix is a rough approximation to the true PSF function, which may vary among different sensors. Although this assumption may be reasonable in some cases, reflectance values of the LSHTR pixel may be beyond the simple averaging of the HSLTR pixels within it, especially for heterogeneous areas. This is enhanced in ISTBDF-I by unmixing the LSHTR image in order to maintain the mixed information within the mixed pixels, which is introduced as the prior estimation of the target HSLTR image to ensure a reasonable estimation of reflectance for small patches. Second, resampling the LSHTR images with the addition of spatial details from the high frequencies of neighboring HSLTR images in STBDF-II is also limited, especially when temporal changes of the study area are relatively large, making the high frequencies of the target HSLTR image differ from those of the neighboring HSLTR images. To overcome this problem, an unmixing-based algorithm is integrated into the whole fusion framework to estimate the mean of the joint distribution by unmixing the LSHTR images instead of resampling. Such a process can guarantee that proper spatial details are included in the fusion process.

However, there are several limitations of ISTBDF-I and ISTBDF-II that motivate directions for further study. First, the assumption that the error and the change tendency are Gaussian may not hold for all situations. Future research can try other distributions, such as a skew normal

distribution, according to different circumstances, and then formulate the corresponding algorithms for spatio-temporal image fusion. Second, the end-members in the linear unmixing algorithm are deterministic, which may be inaccurately generated. Future studies can employ stochastic Gaussian mixture models in which the endmembers are statistically modeled as Gaussian-distributed random vectors parameterized by their mean vector and covariance matrix. Third, the selection of the input image-pairs plays an important role in producing the final fusion result for most algorithms. There are no requirements on the number of image pairs for the ISTBDF-I and ISTBDF-II algorithms. They are suitable for one or more pairs of images. However, in some cloudy areas, it is difficult to acquire two image pairs with high quality simultaneously. The performance of the proposed algorithms with one image pair will be studied in further research. The extent to which the accuracy of the algorithms is affected by the number of image pairs and the similarity of the input image pairs to the target image constitute an interesting problem that will also be investigated in future studies. Fourth, similar to STARFM [2] and ESTARFM [5], land-cover changes, particularly those with transient shape changes, may not be accurately predicted by ISTBDF-I and -II. Future studies should improve this aspect by developing algorithms that can detect land-cover changes such as mapping forest events (e.g., floods and fires). Lastly, like STARFM-based or unmixing-based methods, ISTBDF-I and ISTBDF-II are developed to fuse reflectance data, which is then used to calculate vegetation indices for ecological and hydrological applications. An alternative approach is to first calculate linearly additive vegetation indices, like leaf area index (LAI), and then fuse the indices [58–60]. Earlier studies reported that this alternative approach may produce better fusion results and attributed it to the introduction of less error and its propagation. In addition, this approach has lower computational cost because of the blended single index rather than multiple bands. This provides a new perspective for future studies.

6. Conclusions

To better monitor land surface dynamics, various methods have been proposed for the spatio-temporal fusion of remotely sensed images over the years. Though they are relatively effective, each suffers from certain limitations, particularly in the following three aspects: requirement on the number of input image pairs, ability in handling heterogeneous areas, and ability in monitoring land-cover changes. STBDF-I and -II predict better than STARFM and ESTARFM in a relatively homogeneous landscape, and they have no requirements on the number of input image pairs. However, they do not adequately make use of the mixture information in LSHTR pixels, which leads to their inability to detect the temporal changes of fine patches. The unmixing-based algorithms can extract possible spatial details from the mixed LSHTR pixels but they frequently generate LSHTR-like reflectance. To overcome these limitations, we have proposed, on the basis of STBDF-I and -II, two improved Bayesian data fusion methods with unmixing-based algorithms, coined ISTBDF-I and ISTBDF-II, to generate remote sensing images with high resolution in both space and time. We have compared and evaluated the performance of STBDF-II, ISTBDF-I, and ISTBDF-II using simulated data, as well as Landsat and MODIS data. Compared to the original STBDF-II algorithm, the proposed algorithms generate improved spatio-temporal resolution images, especially over heterogeneous areas.

To conclude, the proposed ISTBDF-I and ISTBDF-II enhance our capability in capturing phenology changes over heterogeneous landscapes by generating reasonably good satellite images that are high in both spatial and temporal resolutions. This is an important addition to the family of spatio-temporal image fusion techniques and can be applied to other remotely sensed images.

Author Contributions: Conceptualization, J.X.; Formal analysis, J.X., Y.L. and T.F.; Methodology, J.X. and Y.L.; Supervision, Y.L. and T.F.; Writing—original draft, J.X.; Writing—review and editing, J.X., Y.L., and T.F.

Funding: This research was funded by the Hong Kong Research Grant Council, grant number CUHK 444411.

Acknowledgments: We would like to thank Irina V. Emelyanova's team for making their dataset available (<https://doi.org/10.4225/08/514B7FD1C798C>). We thank Maofeng Liu for insightful discussion.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Khaleghi, B.; Khamis, A.; Karray, F.O.; Razavi, S.N. Multisensor data fusion: A review of the state-of-the-art. *Inf. Fusion* **2013**, *14*, 28–44. [[CrossRef](#)]
2. Gao, F.; Masek, J.; Schwaller, M.; Hall, F. On the blending of the landsat and modis surface reflectance: Predicting daily landsat surface reflectance. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2207–2218.
3. Hilker, T.; Wulder, M.A.; Coops, N.C.; Seitz, N.; White, J.C.; Gao, F.; Masek, J.G.; Stenhouse, G. Generation of dense time series synthetic landsat data through data blending with modis using a spatial and temporal adaptive reflectance fusion model. *Remote Sens. Environ.* **2009**, *113*, 1988–1999. [[CrossRef](#)]
4. Xin, Q.; Olofsson, P.; Zhu, Z.; Tan, B.; Woodcock, C.E. Toward near real-time monitoring of forest disturbance by fusion of modis and landsat data. *Remote Sens. Environ.* **2013**, *135*, 234–247. [[CrossRef](#)]
5. Zhu, X.; Chen, J.; Gao, F.; Chen, X.; Masek, J.G. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sens. Environ.* **2010**, *114*, 2610–2623. [[CrossRef](#)]
6. Gao, F.; Hilker, T.; Zhu, X.; Anderson, M.; Masek, J.; Wang, P.; Yang, Y. Fusing landsat and modis data for vegetation monitoring. *IEEE Geosci. Remote Sens. Mag.* **2015**, *3*, 47–60. [[CrossRef](#)]
7. Hazaymeh, K.; Hassan, Q.K. Spatiotemporal image-fusion model for enhancing the temporal resolution of landsat-8 surface reflectance images using MODIS images. *J. Appl. Remote Sens.* **2015**, *9*, 096095. [[CrossRef](#)]
8. Kwan, C.; Budavari, B.; Gao, F.; Zhu, X. A hybrid color mapping approach to fusing modis and landsat images for forward prediction. *Remote Sens.* **2018**, *10*, 520. [[CrossRef](#)]
9. Wu, B.; Huang, B.; Cao, K.; Zhuo, G. Improving spatiotemporal reflectance fusion using image inpainting and steering kernel regression techniques. *Int. J. Remote Sens.* **2017**, *38*, 706–727. [[CrossRef](#)]
10. Gaulton, R.; Hilker, T.; Wulder, M.A.; Coops, N.C.; Stenhouse, G. Characterizing stand-replacing disturbance in western alberta grizzly bear habitat, using a satellite-derived high temporal and spatial resolution change sequence. *For. Ecol. Manag.* **2011**, *261*, 865–877. [[CrossRef](#)]
11. Cheng, Q.; Liu, H.; Shen, H.; Wu, P.; Zhang, L. A spatial and temporal nonlocal filter-based data fusion method. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4476–4488. [[CrossRef](#)]
12. Keshava, N.; Mustard, J.F. Spectral unmixing. *IEEE Signal Process. Mag.* **2002**, *19*, 44–57. [[CrossRef](#)]
13. Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 354–379. [[CrossRef](#)]
14. Zhukov, B.; Oertel, D.; Lanzl, F.; Reinhackel, G. Unmixing-based multisensor multiresolution image fusion. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 1212–1226. [[CrossRef](#)]
15. Zurita-Milla, R.; Clevers, J.; Van Gijssel, J.; Schaepman, M. Using meris fused images for land-cover mapping and vegetation status assessment in heterogeneous landscapes. *Int. J. Remote Sens.* **2011**, *32*, 973–991. [[CrossRef](#)]
16. Zurita-Milla, R.; Clevers, J.G.; Schaepman, M.E. Unmixing-based landsat tm and meris fr data fusion. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 453–457. [[CrossRef](#)]
17. Zurita-Milla, R.; Gómez-Chova, L.; Guanter, L.; Clevers, J.G.; Camps-Valls, G. Multitemporal unmixing of medium-spatial-resolution satellite images: A case study using meris images for land-cover mapping. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4308–4317. [[CrossRef](#)]
18. Zurita-Milla, R.; Kaiser, G.; Clevers, J.G.P.W.; Schneider, W.; Schaepman, M.E. Downscaling time series of meris full resolution data to monitor vegetation seasonal dynamics. *Remote Sens. Environ.* **2009**, *113*, 1874–1885. [[CrossRef](#)]
19. Amoros-Lopez, J.; Gomez-Chova, L.; Alonso, L.; Guanter, L.; Moreno, J.; Camps-Valls, G. Regularized multiresolution spatial unmixing for envisat/meris and landsat/TM image fusion. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 844–848. [[CrossRef](#)]
20. Wu, M.; Wu, C.; Huang, W.; Niu, Z.; Wang, C.; Li, W.; Hao, P. An improved high spatial and temporal data fusion approach for combining landsat and modis data to generate daily synthetic landsat imagery. *Inf. Fusion* **2016**, *31*, 14–25. [[CrossRef](#)]
21. Zhu, X.; Helmer, E.H.; Gao, F.; Liu, D.; Chen, J.; Lefsky, M.A. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sens. Environ.* **2016**, *172*, 165–177. [[CrossRef](#)]

22. Amorós-López, J.; Gómez-Chova, L.; Alonso, L.; Guanter, L.; Zurita-Milla, R.; Moreno, J.; Camps-Valls, G. Multitemporal fusion of landsat/tm and envisat/meris for crop monitoring. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *23*, 132–141. [[CrossRef](#)]
23. Doxani, G.; Mitraka, Z.; Gascon, F.; Goryl, P.; Bojkov, B.R. A spectral unmixing model for the integration of multi-sensor imagery: A tool to generate consistent time series data. *Remote Sens.* **2015**, *7*, 14000–14018. [[CrossRef](#)]
24. Zhang, W.; Li, A.; Jin, H.; Bian, J.; Zhang, Z.; Lei, G.; Qin, Z.; Huang, C. An enhanced spatial and temporal data fusion model for fusing landsat and modis surface reflectance to generate high temporal landsat-like data. *Remote Sens.* **2013**, *5*, 5346–5368. [[CrossRef](#)]
25. Huang, B.; Zhang, H. Spatio-temporal reflectance fusion via unmixing: Accounting for both phenological and land-cover changes. *Int. J. Remote Sens.* **2014**, *35*, 6213–6233. [[CrossRef](#)]
26. Ma, J.; Zhang, W.; Marinoni, A.; Gao, L.; Zhang, B. An improved spatial and temporal reflectance unmixing model to synthesize time series of landsat-like images. *Remote Sens.* **2018**, *10*, 1388. [[CrossRef](#)]
27. Huang, B.; Song, H. Spatiotemporal reflectance fusion via sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716. [[CrossRef](#)]
28. Wu, B.; Huang, B.; Zhang, L. An error-bound-regularized sparse coding for spatiotemporal reflectance fusion. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6791–6803. [[CrossRef](#)]
29. Wei, J.; Wang, L.; Liu, P.; Song, W. Spatiotemporal fusion of remote sensing images with structural sparsity and semi-coupled dictionary learning. *Remote Sens.* **2016**, *9*, 21. [[CrossRef](#)]
30. Chen, B.; Huang, B.; Xu, B. A hierarchical spatiotemporal adaptive fusion model using one image pair. *Int. J. Digit. Earth* **2017**, *10*, 639–655. [[CrossRef](#)]
31. Tipping, M.E.; Bishop, C.M. Bayesian image super-resolution. In Proceedings of the NIPS, Vancouver, BC, Canada, 9–14 December 2002; pp. 1279–1286.
32. Pickup, L.C.; Capel, D.P.; Roberts, S.J.; Zisserman, A. Bayesian methods for image super-resolution. *Comput. J.* **2009**, *52*, 101–113. [[CrossRef](#)]
33. Zhang, H.; Zhang, Y.; Li, H.; Huang, T.S. Generative bayesian image super resolution with natural image prior. *IEEE Trans. Image Process.* **2012**, *21*, 4054–4067. [[CrossRef](#)] [[PubMed](#)]
34. Villena, S.; Vega, M.; Babacan, S.D.; Molina, R.; Katsaggelos, A.K. Bayesian combination of sparse and non-sparse priors in image super resolution. *Digit. Signal Process.* **2013**, *23*, 530–541. [[CrossRef](#)]
35. Sharma, R.K.; Leen, T.K.; Pavel, M. Bayesian sensor image fusion using local linear generative models. *Opt. Eng.* **2001**, *40*, 1364–1376.
36. Fasbender, D.; Radoux, J.; Bogaert, P. Bayesian data fusion for adaptable image pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1847–1857. [[CrossRef](#)]
37. Eismann, M.T.; Hardie, R.C. Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 455–465. [[CrossRef](#)]
38. Hardie, R.C.; Eismann, M.T.; Wilson, G.L. Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor. *IEEE Trans. Image Process.* **2004**, *13*, 1174–1184. [[CrossRef](#)] [[PubMed](#)]
39. Fasbender, D.; Obsomer, V.; Bogaert, P.; Defourny, P. Updating scarce high resolution images with time series of coarser images: A bayesian data fusion solution. In *Sensor and Data Fusion*; Milisavljevic, N., Ed.; IntechOpen: Vienna, Austria, 2009.
40. Fasbender, D.; Obsomer, V.; Radoux, J.; Bogaert, P.; Defourny, P. Bayesian Data fusion: Spatial and Temporal Applications. In Proceedings of the 2007 International Workshop on the Analysis of Multi-temporal Remote Sensing Images, Leuven, Belgium, 18–20 July 2007; pp. 1–6.
41. Huang, B.; Zhang, H.; Song, H.; Wang, J.; Song, C. Unified fusion of remote-sensing imagery: Generating simultaneously high-resolution synthetic spatial–temporal–spectral earth observations. *Remote Sens. Lett.* **2013**, *4*, 561–569. [[CrossRef](#)]
42. Xue, J.; Leung, Y.; Fung, T. A bayesian data fusion approach to spatio-temporal fusion of remotely sensed images. *Remote Sens.* **2017**, *9*, 1310. [[CrossRef](#)]
43. Walker, J.; De Beurs, K.; Wynne, R.; Gao, F. Evaluation of landsat and modis data fusion products for analysis of dryland forest phenology. *Remote Sens. Environ.* **2012**, *117*, 381–393. [[CrossRef](#)]
44. Singh, D. Generation and evaluation of gross primary productivity using landsat data through blending with modis data. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 59–69. [[CrossRef](#)]

45. Bhandari, S.; Phinn, S.; Gill, T. Preparing landsat image time series (lits) for monitoring changes in vegetation phenology in Queensland, Australia. *Remote Sens.* **2012**, *4*, 1856–1886. [[CrossRef](#)]
46. Emelyanova, I.V.; McVicar, T.R.; Van Niel, T.G.; Li, L.T.; van Dijk, A.I.J.M. Assessing the accuracy of blending landsat–modis surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. *Remote Sens. Environ.* **2013**, *133*, 193–209. [[CrossRef](#)]
47. Xu, Y.; Huang, B.; Xu, Y.; Cao, K.; Guo, C.; Meng, D. Spatial and temporal image fusion via regularized spatial unmixing. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1362–1366.
48. Bin, C.; Bing, X. A unified spatial-spectral-temporal fusion model using landsat and modis imagery. In Proceedings of the 2014 Third International Workshop on Earth Observation and Remote Sensing Applications (EORSA), Changsha, China, 11–14 June 2014; pp. 256–260.
49. Weng, Q.; Fu, P.; Gao, F. Generating daily land surface temperature at landsat resolution by fusing landsat and modis data. *Remote Sens. Environ.* **2014**, *145*, 55–67. [[CrossRef](#)]
50. Gevaert, C.M.; García-Haro, F.J. A comparison of starfm and an unmixing-based algorithm for landsat and modis data fusion. *Remote Sens. Environ.* **2015**, *156*, 34–44. [[CrossRef](#)]
51. Xie, D.; Zhang, J.; Zhu, X.; Pan, Y.; Liu, H.; Yuan, Z.; Yun, Y. An improved starfm with help of an unmixing-based method to generate high spatial and temporal resolution remote sensing data in complex heterogeneous regions. *Sensors* **2016**, *16*, 207. [[CrossRef](#)]
52. Peng, J.; Liu, Q.; Wang, L.; Liu, Q.; Fan, W.; Lu, M.; Wen, J. Characterizing the pixel footprint of satellite albedo products derived from modis reflectance in the heihe river basin, China. *Remote Sens.* **2015**, *7*, 6886–6907. [[CrossRef](#)]
53. Heinz, D.C.; Cheln, I.C. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 529–545. [[CrossRef](#)]
54. Miao, L.; Qi, H. Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 765–777. [[CrossRef](#)]
55. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.
56. Zhou, W.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [[CrossRef](#)]
57. Liu, C.; Frazier, P.; Kumar, L. Comparative assessment of the measures of thematic classification accuracy. *Remote Sens. Environ.* **2007**, *107*, 606–616. [[CrossRef](#)]
58. Jarihani, A.; McVicar, T.; Van Niel, T.; Emelyanova, I.; Callow, J.; Johansen, K. Blending landsat and modis data to generate multispectral indices: A comparison of “index-then-blend” and “blend-then-index” approaches. *Remote Sens.* **2014**, *6*, 9213–9238. [[CrossRef](#)]
59. Busetto, L.; Meroni, M.; Colombo, R. Combining medium and coarse spatial resolution satellite data to improve the estimation of sub-pixel NDVI time series. *Remote Sens. Environ.* **2008**, *112*, 118–131. [[CrossRef](#)]
60. Rao, Y.; Zhu, X.; Chen, J.; Wang, J. An improved method for producing high spatial-resolution NDVI time series datasets with multi-temporal modis ndvi data and landsat TM/ETM+ images. *Remote Sens.* **2015**, *7*, 7865–7891. [[CrossRef](#)]

