remote sensing

MDPI

Article

A Bayesian Data Fusion Approach to Spatio-Temporal Fusion of Remotely Sensed Images

Jie Xue [1], Yee Leung [1,2,3,*] and Tung Fung [1,2,4]

1 Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong, China; jiexue@link.cuhk.edu.hk (J.X.); tungfung@cuhk.edu.hk (T.F.)
2 Institute of Future Cities, The Chinese University of Hong Kong, Hong Kong, China
3 Big Data Decision Analytic Center, The Chinese University of Hong Kong, Hong Kong, China
4 Institute of Environment, Energy and Sustainability, The Chinese University of Hong Kong, Hong Kong, China
* Correspondence: yeeleung@cuhk.edu.hk; Tel.: +852-3943-6473

Received: 24 September 2017; Accepted: 12 December 2017; Published: 13 December 2017

Abstract: Remote sensing provides rich sources of data for the monitoring of land surface dynamics. However, single-sensor systems are constrained from providing spatially high-resolution images with high revisit frequency due to the inherent sensor design limitation. To obtain images high in both spatial and temporal resolutions, a number of image fusion algorithms, such as spatial and temporal adaptive reflectance fusion model (STARFM) and enhanced STARFM (ESTARFM), have been recently developed. To capitalize on information available in a fusion process, we propose a Bayesian data fusion approach that incorporates the temporal correlation information in the image time series and casts the fusion problem as an estimation problem in which the fused image is obtained by the Maximum A Posterior (MAP) estimator. The proposed approach provides a formal framework for the fusion of remotely sensed images with a rigorous statistical basis; it imposes no requirements on the number of input image pairs; and it is suitable for heterogeneous landscapes. The approach is empirically tested with both simulated and real-life acquired Landsat and Moderate Resolution Imaging Spectroradiometer (MODIS) images. Experimental results demonstrate that the proposed method outperforms STARFM and ESTARFM, especially for heterogeneous landscapes. It produces surface reflectances highly correlated with those of the reference Landsat images. It gives spatio-temporal fusion of remotely sensed images a solid theoretical and empirical foundation that may be extended to solve more complicated image fusion problems.

Keywords: Bayesian data fusion; Landsat; MODIS; spatio-temporal image fusion; time series

1. Introduction

Remote sensing provides crucial data sources for the monitoring of land surface dynamics such as vegetation phenology and land-cover changes. Effective terrestrial monitoring requires remotely sensed images high in both spatial and temporal resolutions. However, technological limitations pose challenges [1] for sensor designs, and trade-offs have to be made to balance spatial details with the spatial extent and revisit frequency. For instance, high spatial resolution (henceforth referred to as "high-resolution") images obtained by the Thematic Mapper (TM) sensor or Enhanced Thematic Mapper Plus (ETM+) sensor on Landsat have a spatial resolution of approximately 30 m and have been shown to be useful in applications such as monitoring of land-cover changes. However, Landsat satellites have a revisit cycle of 16 days, and around 35% of the images are contaminated by cloud cover [2]. Together with other poor atmospheric conditions, this limits the applications of Landsat data. On the contrary, a low spatial resolution (henceforth referred to as "low-resolution") sensor, such as the Moderate Resolution Image Spectroradiometer (MODIS) on the Aqua and Terra satellites, has a

daily revisit period but a relatively low spatial resolution ranging from 250 m to 1000 m, limiting its effectiveness in the monitoring of ecosystem dynamics in heterogeneous landscapes.

To capitalize on the strengths of individual sensors, it is of great advantage to fuse multi-sensor data to generate images high in both spatial and temporal resolutions for the monitoring of dense land surface dynamics, keeping in mind that images acquired by multi-sensors may have different acquisition times, different resolutions (e.g., spectral and radiometric), and different bands (e.g., the number of bands, the central wavelengths, and the bandwidths). Driven by the great potential of applications in various areas, the fusion of multi-sensor data has attracted much attention in recent years [3]. A number of approaches for the spatio-temporal fusion of images have been developed [4].

The spatial and temporal adaptive reflectance fusion model (STARFM) is one of the first fusion algorithms that has been widely used for synthesizing Landsat and MODIS imageries [5]. STARFM does not explicitly treat the issue that sensor difference can change for different land-cover types. An improved STARFM was proposed to address this issue by using linear regression models [6]. Roy et al. [7] proposed a semi-physical fusion approach that explicitly handles the directional dependence of reflectance described by the bidirectional reflectance distribution function (BRDF). STARFM is developed on the assumption of homogeneous pixels, and thus may not be suitable for heterogeneous land surfaces [8]. An enhanced STARFM, termed ESTARFM [9], was developed to improve the prediction of surface reflectance in heterogeneous landscapes, followed by a modified version of ESTARFM that combines additional land-cover data [10]. STARFM and ESTARFM are shown to be useful in capturing reflectance changes due to changes in vegetation phenology [11–14], which is a key element of seasonal patterns of water and carbon exchanges between land surfaces and the atmosphere [15–17]. However, STARFM and ESTARFM may have problems in mapping disturbance events when land-cover changes are transient and not recorded in at least one of the baseline high-resolution (e.g., Landsat) images [8]. An improved method based on STARFM, named spatial temporal adaptive algorithm for mapping reflectance changes (STAARCH), was proposed to detect the disturbance and reflectance changes over vegetated land surfaces by tasseled cap transformation [18]. Similarly, another approach was developed for near real-time monitoring of forest disturbances through the fusion of MODIS and Landsat images [19]. However, this algorithm was developed specifically for forest disturbances. It may not be suitable for other kinds of land-cover changes [20,21]. In contrast to STARFM, both STAARCH and ESTARFM require at least two high-resolution images; that is, one before and another after the target image [22]. However, it is difficult to meet this need for areas with frequently cloudy days [23]. Furthermore, the assumption in ESTARFM that land surface reflectance changes linearly over time with a single rate between two acquired Landsat images may not hold in some cases [8].

Besides the abovementioned algorithms, several methods based on sparse representation and dictionary learning have been proposed [24–26]. Although these methods can well predict pixels with land-cover changes, they do not accurately maintain the shape of objects [27]. In addition, the use of strong assumptions on the dictionaries and sparse coding coefficients may not hold in some cases [26].

Another popular branch of image fusion algorithms relies on the unmixing techniques due to their ability to reconstruct images with high spectral fidelity [28–32]. An advantage of the unmixing-based methods is that they do not require high- and low-resolution images to have corresponding spectral bands, and they can downscale extra spectral bands in low-resolution images to increase the spectral resolution of the high-resolution images [30]. In recent years, several approaches have been developed by combining the unmixing algorithm and the STARFM method or other techniques for improved fusion performance [27,29,33–35].

Despite that the approaches mentioned above are well developed and widely used, studies on the effectiveness and applicability of applying the Bayesian estimation theory to the spatio-temporal image fusion problems are scanty [36–38]. In these studies [36–38], linear regression is used to reflect the temporal dynamics, which, however, may not hold in a variety of situations. Moreover, there may be no regression-like trends in some cases. In addition, the Bayesian method can handle uncertainties

of input images in a probabilistic manner, which has shown its effectiveness in spatial–spectral fusion of remotely sensed images [39–44] and superresolution of remote-sensing images [45–47]. It is reasonable to explore the potential of Bayesian approach in spatio-temporal fusion due to its satisfactory performance in the spatial–spectral fusion of remotely sensed images. Although a few Bayesian algorithms have been developed [36–38], there lacks a solid theoretical Bayesian framework with rigorous statistical procedures for the spatio-temporal fusion of remotely sensed images. This paper is aimed to develop a Bayesian data fusion framework which provides a formal construct to systematically obtain the targeted high-resolution image on the basis of a solid probabilistic foundation that enables us to efficiently handle uncertainties by applying the statistical estimation tools developed in our model. Our Bayesian fusion approach to the spatio-temporal fusion of remotely sensed images makes use of the advantage of multivariate arguments in statistics to handle temporal dynamics in a more flexible way rather than just by linear regression. Specifically, we use the joint distribution to embody the covariate information that implicitly expresses the temporal changes of images. This approach imposes no requirements on the number of high-resolution images in the input and can generate high-resolution-like images in homogeneous or relatively heterogeneous areas. Our proposed Bayesian approach will be empirically compared with STARFM and ESTARFM [5,9] through quantitative assessment.

The remainder of the paper is organized as follows. We first provide the theoretical formulation of our Bayesian estimation approach in Section 2. Specific considerations for parameter estimations are described in Section 3. In Section 4, we demonstrate the effectiveness of our proposed approach with experimental studies, including the characteristics of the datasets, the quantitative assessment metrics, and the comparison with STARFM and ESTARFM. Section 5 provides further discussion on the results in Section 4 and limitations of this work. We then conclude our paper with a summary in Section 6.

2. Theoretical Basis of Bayesian Data Fusion

Suppose the high- and low-resolution images acquired at the same dates are comparable with each other after radiometric calibration and geometric rectification, and they share similar spectral bands. Denote the set of all available high-resolution images by $\mathcal{H} = \{\mathbf{H}^{(s)}: s \in \mathrm{T_H}\}$ and the existing low-resolution images by $\mathcal{L} = \{\mathbf{L}^{(t)}, t \in \mathrm{T_L}\}$, where the superscripts s and t represent the acquisition dates of the high- and low-resolution images, respectively. In particular, it is reasonable to assume that the domain of $\mathcal{H}$ is a subset of that of $\mathcal{L}$, namely, $\mathrm{T_H} \subseteq \mathrm{T_L}$ since the high-resolution images generally have lower revisit frequency than the low-resolution images. The objective of this paper is to obtain the image sequence $\mathcal{F} = \{\mathbf{F}^{(r)}, r \in \mathrm{T_F}\}$, which has the same spatial resolution with $\mathcal{H}$ and the temporal frequency with $\mathcal{L}$, i.e., $\mathrm{T_F} = \mathrm{T_L}$, by fusing the corresponding images in $\mathcal{H}$ and $\mathcal{L}$.

We assume that each high-resolution image (i.e., $\mathbf{H}^{(s)}$, $s \in \mathrm{T_H}$) has N pixels per band, and each low-resolution image (i.e., $\mathbf{L}^{(t)}$, $t \in \mathrm{T_L}$) has M pixels per band. They have K corresponding spectral bands in this study. The three-dimensional (data cube) images $\mathbf{H}^{(s)}$, $\mathbf{L}^{(t)}$, and $\mathbf{F}^{(r)}$ are represented by the one-dimensional statistical random vectors $\mathbf{x}^{(s)}$, $\mathbf{y}^{(t)}$, and $\mathbf{z}^{(r)}$, composed of the pixels of the images $\mathbf{H}^{(s)}$, $\mathbf{L}^{(t)}$, and $\mathbf{F}^{(r)}$ in the band-interleaved-by-pixel lexicographical order. Their respective expressions are:

$$\mathbf{x}^{(s)} = [x_{1,1}^{(s)}, \ldots, x_{K,1}^{(s)}, x_{1,2}^{(s)}, \ldots, x_{K,2}^{(s)}, \ldots, x_{1,N}^{(s)}, \ldots, x_{K,N}^{(s)}]^{\mathrm{T}} = [(\mathbf{x}_1^{(s)})^{\mathrm{T}}, (\mathbf{x}_2^{(s)})^{\mathrm{T}}, \ldots, (\mathbf{x}_N^{(s)})^{\mathrm{T}}]^{\mathrm{T}} \quad (1)$$

$$\mathbf{y}^{(t)} = [y_{1,1}^{(t)}, \ldots, y_{K,1}^{(t)}, y_{1,2}^{(t)}, \ldots, y_{K,2}^{(t)}, \ldots, y_{1,M}^{(t)}, \ldots, y_{K,M}^{(t)}]^{\mathrm{T}} = [(\mathbf{y}_1^{(t)})^{\mathrm{T}}, (\mathbf{y}_2^{(t)})^{\mathrm{T}}, \ldots, (\mathbf{y}_M^{(t)})^{\mathrm{T}}]^{\mathrm{T}}, \quad (2)$$

$$\mathbf{z}^{(r)} = [z_{1,1}^{(r)}, \ldots, z_{K,1}^{(r)}, z_{1,2}^{(r)}, \ldots, z_{K,2}^{(r)}, \ldots, z_{1,N}^{(r)}, \ldots, y_{K,N}^{(r)}]^{\mathrm{T}} = [(\mathbf{z}_1^{(r)})^{\mathrm{T}}, (\mathbf{z}_2^{(r)})^{\mathrm{T}}, \ldots, (\mathbf{z}_N^{(r)})^{\mathrm{T}}]^{\mathrm{T}}, \quad (3)$$

where $\mathbf{x}_i^{(s)} = [x_{1,i}^{(s)}, \ldots, x_{k,i}^{(s)}, \ldots, x_{K,i}^{(s)}]^{\mathrm{T}}$ is the vector of pixels for all K bands at pixel location i of the high-resolution image on date s, $i = 1, \ldots, N$, within which $x_{k,i}^{(s)}$ is the pixel value of band k at location i

on date s; $\mathbf{y}_j^{(t)} = [y_{1,j}^{(t)}, ..., y_{k,j}^{(t)}, ..., y_{K,j}^{(t)}]^{\mathrm{T}}$, $j = 1, \ldots, M$, and $\mathbf{z}_i^{(r)} = [z_{1,i}^{(r)}, ..., z_{k,i}^{(r)}, ..., z_{K,i}^{(r)}]^{\mathrm{T}}$, $i = 1, \ldots, N$, are defined similarly.

To be specific, we assume the target high-resolution image is missing on date t_0 and available on dates t_k, $k = 1, \ldots, S$, that is, $t_0 \notin \mathrm{T_H}$ and $t_k \in \mathrm{T_H}$, while the low-resolution images are available on both t_0 and t_k. Our objective is to estimate $\mathbf{z}^{(t_0)}$ given the corresponding $\mathbf{y}^{(t_0)}$ at the target date t_0 and the other image pairs $\mathbf{x}^{(t_k)}$ and $\mathbf{y}^{(t_k)}$. In this section, we propose a general framework for Bayesian data fusion depicted in Figure 1. Part A of Figure 1 models the relationship between the low-resolution and high-resolution images at the target date t_0. Part B shows the use of multivariate joint distribution to model the evolution of the images and their temporal correlations. Finally, the targeted high-resolution image is estimated by the Maximum A Posterior (MAP) estimation as shown in part C.

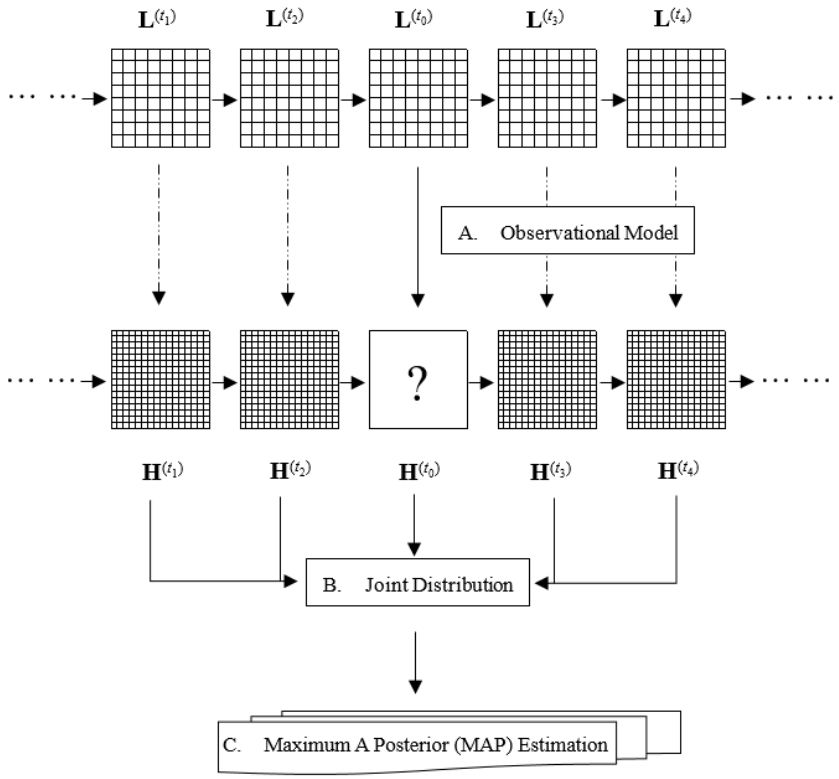

Figure 1. The general framework of Bayesian data fusion using quadruple pairs. It should be noted that the framework can be extended to the use of any number of image-pairs, even just one pair.

In what follows, we give a detail description of the components of the general framework.

2.1. The Observation Model

The basic concept of Bayesian data fusion comes from the idea that variables of interest cannot be directly observed. Thus, it is necessary to model the relationship between variables available and variables of interest, which are unknown. The first step of our approach is to build a model to describe the relationship between the low-resolution image and the high-resolution image at the target date t_0. We call it the observational model, which can be written in a matrix form as [48,49]:

$$\mathbf{y}^{(t_0)} = \mathbf{DBz}^{(t_0)} + \mathbf{e}^{(t_0)} = \mathbf{Wz}^{(t_0)} + \mathbf{e}^{(t_0)}, \tag{4}$$

where $\mathbf{W} = \mathbf{DB}$ is the $KM \times KN$ transformation matrix, $\mathbf{D}$ is the $KM \times KN$ down-sampling matrix, $\mathbf{B}$ is the $KN \times KN$ blurring matrix, and $\mathbf{e}^{(t_0)}$ is the noise. The effects of blurring and down-sampling of the target high-resolution image are combined into the point spread function (PSF) $\mathbf{W}$, which

maps the pixels of the high-resolution image to the pixels of the low-resolution image [38,50]. Each low-resolution pixel is assumed to be the convolution of the high-resolution pixels with the PSFs expressed as the rows in $\mathbf{W}$ plus noise $\mathbf{e}^{(t_0)}$.

In this study, $\mathbf{W}$ is assumed to be a rectangular PSF model, similar to the approach in [51]. This approach proves to be effective. The PSF for each low-resolution pixel is uniform and also non-overlapping among neighboring low-resolution pixels. Specifically, it is defined with an equal weight of $1/w^2$ for the high-resolution pixels contained within the same low-resolution pixel, where $w = n/m$ is the resolution ratio, and $n \times n = N$, $m \times m = M$. Note that N and M are perfect squares. So the form of $\mathbf{W}$ can be expressed by the matrix operation

$$\mathbf{W}_{MK\times NK} = 1/w^2 \cdot \mathbf{H}_{m\times n} \otimes \mathbf{G}_{mK\times nK}, \tag{5}$$

where,

$$\mathbf{H}_{m\times n} = \mathbf{I}_{m\times m} \otimes \mathbf{1}_{1\times w},$$

$$\mathbf{G}_{mK\times nK} = \mathbf{I}_{m\times m} \otimes \mathbf{1}_{1\times w} \otimes \mathbf{I}_{K\times K}$$

The items $\mathbf{I}_{m\times m}$ and $\mathbf{I}_{K\times K}$ are respectively the m-dimensional and K-dimensional unit matrix. The item $\mathbf{1}_{1\times w}$ is a w-dimensional row vector with unit values, and $\otimes$ is the Kronecker product of matrices.

The random vector $\mathbf{e}$ is assumed to be a zero-mean Gaussian random vector with a covariance matrix $\mathbf{C}_{\mathbf{e}^{(t_0)}}$. The probability density function for $\mathbf{e}^{(t_0)}$ is expressed as:

$$p(\mathbf{e}^{(t_0)}) = 1/\sqrt{(2\pi)^{MK}|\mathbf{C}_{\mathbf{e}^{(t_0)}}|}\exp\{-\frac{1}{2}\mathbf{e}^{(t_0)\mathrm{T}}\mathbf{C}^{-1}_{\mathbf{e}^{(t_0)}}\mathbf{e}^{(t_0)}\}. \tag{6}$$

In most applications, it is reasonable to assume that the noise is independent and identically distributed from band-to-band and pixel-to-pixel, which means that the covariance matrix is diagonal with $\mathbf{C}_{\mathbf{e}^{(t_0)}} = \sigma_e^2\mathbf{I}$. It is reasonable to assume that the sensor noise is independent of both $\mathbf{z}^{(t_0)}$ and $\mathbf{x}^{(t_k)}$, $t_k \in \mathrm{T_H} \subseteq \mathrm{T_L}$, $k = 1, \ldots, S$, because it comes from the acquisition of $\mathbf{y}^{(t_0)}$.

This observational model incorporates knowledge about the relationships between the low-resolution and high-resolution images at the target date and the variability caused by the noise. Without other information, this model-based approach for estimating $\mathbf{z}^{(t_0)}$ is ill-posed because the dimension of the unknowns is greater than the dimension of the equations. Thus, it is necessary to include additional information about the desirable high-resolution image, which is provided by the existing high-resolution images in $\mathcal{H}$, into the calibration process.

2.2. The Joint Distribution

The temporal correlation of time-series images depends on the dynamics of the land surface. It is relatively low for situations with rapid phenology changes or sudden land-cover changes due to, for example, forest fires or floods. Previous regression approaches [36–38] use a linear relationship to represent the temporal evolution of the image sequences, which may not hold for cases with rapid phenology changes or sudden land-cover changes. In this paper, we treat the temporal evolution of images as a stochastic process and employ the notion of multivariate joint Gaussian distribution to model it. This approach can be applied to cases with nonlinear temporal evolution and is thus more general than the regression approach. To take advantage of the temporal correlation of image sequences, we assume $\mathbf{x}^{(t_k)}$ (the same with $\mathbf{z}^{(t_k)}$), $t_k \in \mathrm{T_H} \subseteq \mathrm{T_L}$, $k = 1, \ldots, S$, and $\mathbf{z}^{(t_0)}$, $t_0 \in \mathrm{T_L}$ but $t_0 \notin \mathrm{T_H}$, are jointly Gaussian distributed, and define an NKS-dimensional vector $\mathbf{X}$ by cross-combining the S high-resolution image vectors:

$$\mathbf{X} = [(\mathbf{x}_1^{(t_1)})^{\mathrm{T}}, (\mathbf{x}_1^{(t_2)})^{\mathrm{T}}, \ldots, (\mathbf{x}_1^{(t_S)})^{\mathrm{T}}, (\mathbf{x}_2^{(t_1)})^{\mathrm{T}}, \ldots, (\mathbf{x}_2^{(t_S)})^{\mathrm{T}}, \ldots, (\mathbf{x}_N^{(t_1)})^{\mathrm{T}}, \ldots, (\mathbf{x}_N^{(t_S)})^{\mathrm{T}}]^{\mathrm{T}} = [\mathbf{X}_1^{\mathrm{T}}, \mathbf{X}_2^{\mathrm{T}}, \ldots, \mathbf{X}_N^{\mathrm{T}}]^{\mathrm{T}}$$

Then based on the conditional property of the multivariate Gaussian distribution, the conditional probability density function is also Gaussian which can be expressed as [52]:

$$p(\mathbf{z}^{(t_0)}|\mathbf{x}^{(t_1)},...,\mathbf{x}^{(t_s)}) = p(\mathbf{z}^{(t_0)}|\mathbf{X}) = 1/\sqrt{(2\pi)^{NK}|\mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}}|}\exp\{-\tfrac{1}{2}(\mathbf{z}^{(t_0)}-\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}})^{\mathrm{T}}\mathbf{C}^{-1}_{\mathbf{z}^{(t_0)}|\mathbf{X}}(\mathbf{z}^{(t_0)}-\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}})\}, \quad (7)$$

where $\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}}$ is the conditional expectation of $\mathbf{z}^{(t_0)}$ given the existing high-resolution sequences $\mathbf{X}$, 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 by the joint statistics

$$\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}} = \mathrm{E}\{\mathbf{z}^{(t_0)}\} + \mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{X}}\mathbf{C}^{-1}_{\mathbf{X},\mathbf{X}}(\mathbf{X}-\mathrm{E}\{\mathbf{X}\}), \quad (8)$$

$$\mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}} = \mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{z}^{(t_0)}} - \mathbf{C}_{\mathbf{z}^{(t_0)},\mathbf{X}}\mathbf{C}^{-1}_{\mathbf{X},\mathbf{X}}\mathbf{C}^{\mathrm{T}}_{\mathbf{z}^{(t_0)},\mathbf{X}}, \quad (9)$$

where $\mathrm{E}\{\mathbf{X}\}$ is the mean of $\mathbf{X}$ and $\mathrm{E}\{\mathbf{z}^{(t_0)}\}$ is the mean of $\mathbf{z}^{(t_0)}$. 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 of the corresponding multivariate random vectors, respectively.

2.3. The MAP Estimation

Here the goal of the MAP estimation of the Bayesian framework is to obtain the desirable high-resolution image $\mathbf{z}^{(t_0)}$ at the target date by maximizing its conditional probability relative to the existing high-resolution sequences $\mathbf{X}$ and the observed low-resolution image $\mathbf{y}^{(t_0)}$ at the target date. It can be given as:

$$\hat{\mathbf{z}}^{(t_0)} = \underset{\mathbf{z}^{(t_0)}}{\mathrm{argmax}} p(\mathbf{z}^{(t_0)}|\mathbf{y}^{(t_0)},\mathbf{X}), \quad (10)$$

where $p(\mathbf{z}^{(t_0)}|\mathbf{y}^{(t_0)},\mathbf{X})$ is the conditional probability density function of $\mathbf{z}^{(t_0)}$ given $\mathbf{y}^{(t_0)}$ and $\mathbf{X}$, and the optimal $\hat{\mathbf{z}}^{(t_0)}$ is the MAP estimate of $\mathbf{z}^{(t_0)}$ that maximizes $p(\mathbf{z}^{(t_0)}|\mathbf{y}^{(t_0)},\mathbf{X})$. According to the Bayes rule, we have

$$p(\mathbf{z}^{(t_0)}|\mathbf{y}^{(t_0)},\mathbf{X}) = (p(\mathbf{y}^{(t_0)},\mathbf{X}|\mathbf{z}^{(t_0)})\cdot p(\mathbf{z}^{(t_0)}))/p(\mathbf{y}^{(t_0)},\mathbf{X}) = p(\mathbf{y}^{(t_0)}|\mathbf{z}^{(t_0)})\cdot p(\mathbf{X}|\mathbf{z}^{(t_0)})\cdot p(\mathbf{z}^{(t_0)})/p(\mathbf{y}^{(t_0)},\mathbf{X}).$$

Again using the Bayes rule on $p(\mathbf{X}|\mathbf{z}^{(t_0)})$ we yield $p(\mathbf{X}|\mathbf{z}^{(t_0)}) = (p(\mathbf{z}^{(t_0)}|\mathbf{X})\cdot p(\mathbf{X}))/p(\mathbf{z}^{(t_0)})$. Substituting them into (10), the MAP estimator can be rewritten as:

$$\hat{\mathbf{z}}^{(t_0)} = \underset{\mathbf{z}^{(t_0)}}{\mathrm{argmax}} p(\mathbf{y}^{(t_0)}|\mathbf{z}^{(t_0)})\cdot p(\mathbf{z}^{(t_0)}|\mathbf{X}). \quad (11)$$

Based on the observational model in (4) and the probability density function of noise in (6), we obtain

$$p(\mathbf{y}^{(t_0)}|\mathbf{z}^{(t_0)}) = 1/\sqrt{(2\pi)^{MK}|\mathbf{C}_{\mathbf{e}^{(t_0)}}|}\exp\{-\frac{1}{2}(\mathbf{y}^{(t_0)}-\mathbf{W}\mathbf{z}^{(t_0)})^{\mathrm{T}}\mathbf{C}^{-1}_{\mathbf{e}^{(t_0)}}(\mathbf{y}^{(t_0)}-\mathbf{W}\mathbf{z}^{(t_0)})\}. \quad (12)$$

Given that the conditional probability density functions in (7) and (12) are Gaussian distribution, obviously their product, $p(\mathbf{y}^{(t_0)}|\mathbf{z}^{(t_0)})\cdot p(\mathbf{z}^{(t_0)}|\mathbf{X})$, is also a Gaussian distribution with mean

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

and covariance

$$\hat{\mathbf{C}} = [\mathbf{W}^{\mathrm{T}}\mathbf{C}^{-1}_{\mathbf{e}^{(t_0)}}\mathbf{W} + \mathbf{C}^{-1}_{\mathbf{z}^{(t_0)}|\mathbf{X}}]^{-1}.$$

Lastly, the optimal estimator of the desirable high-resolution image $\hat{\mathbf{z}}^{(t_0)}$ is given by

$$\hat{\mathbf{z}}^{(t_0)} = [\mathbf{W}^{\mathrm{T}}\mathbf{C}^{-1}_{\mathbf{e}^{(t_0)}}\mathbf{W} + \mathbf{C}^{-1}_{\mathbf{z}^{(t_0)}|\mathbf{X}}]^{-1}[\mathbf{W}^{\mathrm{T}}\mathbf{C}^{-1}_{\mathbf{e}^{(t_0)}}\mathbf{y}^{(t_0)} + \mathbf{C}^{-1}_{\mathbf{z}^{(t_0)}|\mathbf{X}}\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}}]. \quad (13)$$

In order to reduce the memory space and computational cost and to avoid the problem of the inversion of the noise covariance matrix in (6) going to zero, we apply the matrix inversion lemma [52] to simplify (13) to obtain:

$$\hat{\mathbf{z}}^{(t_0)} = \boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}} + \mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}}\mathbf{W}^{\mathrm{T}}[\mathbf{W}\mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}}\mathbf{W}^{\mathrm{T}} + \mathbf{C}_{\mathbf{e}^{(t_0)}}]^{-1}[\mathbf{y}^{(t_0)} - \mathbf{W}\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}}]. \tag{14}$$

3. Implementation Considerations

In terms of the temporal correlation of image sequences, we have weakened the assumption of the joint Gaussian distribution for the whole high-resolution image vectors to N joint Gaussian distributions for each high-resolution pixel series. This simplification can greatly reduce the calibration of the algorithm. Therefore, the conditional mean vector $\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}}$ and covariance matrix $\mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}}$ can be estimated for each individual pixel [40] as:

$$\boldsymbol{\mu}_{\mathbf{z}^{(t_0)}|\mathbf{X}} = [\boldsymbol{\mu}^{\mathrm{T}}_{\mathbf{z}_1^{(t_0)}|\mathbf{X}_1}, \boldsymbol{\mu}^{\mathrm{T}}_{\mathbf{z}_2^{(t_0)}|\mathbf{X}_2}, ..., \boldsymbol{\mu}^{\mathrm{T}}_{\mathbf{z}_N^{(t_0)}|\mathbf{X}_N}]^{\mathrm{T}}, \tag{15}$$

$$\mathbf{C}_{\mathbf{z}^{(t_0)}|\mathbf{X}} = \begin{bmatrix} \mathbf{C}_{\mathbf{z}_1^{(t_0)}|\mathbf{X}_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\mathbf{z}_2^{(t_0)}|\mathbf{X}_2} & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{C}_{\mathbf{z}_N^{(t_0)}|\mathbf{X}_N} \end{bmatrix}, \tag{16}$$

where

$$\boldsymbol{\mu}_{\mathbf{z}_i^{(t_0)}|\mathbf{X}_i} = \mathrm{E}\{\mathbf{z}_i^{(t_0)}\} + \mathbf{C}_{\mathbf{z}_i^{(t_0)},\mathbf{X}_i}\mathbf{C}^{-1}_{\mathbf{X}_i,\mathbf{X}_i}(\mathbf{X}_i - \mathrm{E}\{\mathbf{X}_i\}), \tag{17}$$

$$\mathbf{C}_{\mathbf{z}_i^{(t_0)}|\mathbf{X}_i} = \mathbf{C}_{\mathbf{z}_i^{(t_0)},\mathbf{z}_i^{(t_0)}} - \mathbf{C}_{\mathbf{z}_i^{(t_0)},\mathbf{X}_i}\mathbf{C}^{-1}_{\mathbf{X}_i,\mathbf{X}_i}\mathbf{C}^{\mathrm{T}}_{\mathbf{z}_i^{(t_0)},\mathbf{X}_i}, \tag{18}$$

and the pre-defined $\mathbf{X}_i = [(\mathbf{x}_i^{(t_1)})^{\mathrm{T}}, (\mathbf{x}_i^{(t_2)})^{\mathrm{T}}, ..., (\mathbf{x}_i^{(t_S)})^{\mathrm{T}}]^{\mathrm{T}}$, $i = 1, \ldots, N$.

3.1. Statistical Parameters: Mean Estimates

Equations (17) and (18) give the calculation of the conditional mean and covariance for each pixel of the desirable image $\mathbf{z}^{(t_0)}$, within which we need to estimate the mean vectors $\mathrm{E}\{\mathbf{z}_i^{(t_0)}\}$ and $\mathrm{E}\{\mathbf{X}_i\}$ for both known and unknown high-resolution images at each pixel. In this application, we adopt two estimation approaches to the construction of the mean vector as shown below.

3.1.1. Interpolated Observation Estimation (IOE)

The first one, we named as STBDF-I (Spatio-Temporal Bayesian Data Fusion I), simply uses the low-resolution images $\mathbf{y}$ to obtain the raw estimates of $\mathrm{E}\{\mathbf{z}^{(t_0)}\}$ and $\mathrm{E}\{\mathbf{X}\}$. We propose to use the bilinear interpolated $\mathbf{y}^{(t_0)}$ to predict $\mathrm{E}\{\mathbf{z}^{(t_0)}\}$ that may be described as:

$$\mathrm{E}\{\mathbf{z}^{(t_0)}\} = \mathcal{B}(\mathbf{y}^{(t_0)}), \tag{19}$$

where the operator $\mathcal{B}(\cdot)$ represents the bilinear interpolation. Similarly, we also use the $\mathbf{y}^{(t_k)}$ after bilinear interpolation, $k = 1, \ldots, S$, to estimate $\mathrm{E}\{\mathbf{X}\}$ correspondingly. Note that $\mathrm{E}\{\mathbf{z}_i^{(t_0)}\}$, $i = 1, \ldots, N$, represents $\mathrm{E}\{\mathbf{z}^{(t_0)}\}$ at each pixel.

3.1.2. Sharpened Observation Estimation (SOE)

The interpolated low-resolution images can account for the global fluctuations of, but may be smoother than, the actual high-resolution images. The second approach coined the sharpened

observation estimation (we call STBDF-II), is expected to be an improved version of the interpolated observation estimation. The principle of this approach is to add the high-pass frequencies of the high-resolution image to the interpolated low-resolution image on the same date to account for the local spatial details. First, the high-resolution image is decomposed into a low-frequency image and a high-frequency image using a low-pass or high-pass filter (e.g., Gaussian filter and Laplacian filter) [53,54]. Second, the high-frequency image is combined with the interpolated low-resolution image on the same date to estimate the mean of the high-resolution image. The resulting image is expected to obtain local details provided by the high-resolution images, whereas the global fluctuations are provided by the low-resolution image at the target date.

For any t_k, where $t_k \in \mathrm{T_H} \subseteq \mathrm{T_L} = \mathrm{T_F}$, $k = 1, \ldots, S$, the dates on which both the low-resolution and the high-resolution images are available, the estimation of the mean vector is given by

$$\mathrm{E}\{\mathbf{x}^{(t_k)}\} = \mathcal{S}(\mathbf{x}^{(t_k)}, \mathcal{B}(\mathbf{y}^{(t_k)})), \tag{20}$$

where $\mathcal{S}(\mathrm{L}_1, \mathrm{L}_2)$ is the sharpening operator that extracts the high frequencies from the image L_1 and adds them to the image L_2. We can obtain $\mathrm{E}\{\mathbf{X}\}$ from the estimated mean of $\mathbf{x}^{(t_k)}$.

However, for any $t_0 \in \mathrm{T_L} = \mathrm{T_F}$ but $t_0 \notin \mathrm{T_H}$, only the low-resolution image on that date is available. The corresponding high-resolution image is the target (unknown) image, from which it is impossible to extract high frequencies. Under this situation, we assume that there are no big disturbances in the image sequences. The spatial details of the high-resolution images experience relatively slow variations across time. It is thus reasonable to take advantage of the spatial details from the nearest high-resolution images.

For convenience, we only derive $\mathrm{E}\{\mathbf{z}^{(t_0)}\}$ for the case in which the nearest before-and-after high-resolution images of date t_0 are available. Suppose the two nearest high-resolution images are acquired at dates t_k and t_{k+1}. The estimation of the mean vector at t_0 ($t_k < t_0 < t_{k+1}$) is given by

$$\mathrm{E}\{\mathbf{z}^{(t_0)}\} = \mathrm{W}_k \mathcal{S}(\mathbf{x}^{(t_k)}, \mathcal{B}(\mathbf{y}^{(t_0)})) + \mathrm{W}_{k+1} \mathcal{S}(\mathbf{x}^{(t_{k+1})}, \mathcal{B}(\mathbf{y}^{(t_0)})), \tag{21}$$

where W_k and W_{k+1} are the weights used to determine the percentage of the estimated mean vector from each date. They are calculated based on the temporal correlations between the nearest high-resolution images and the target image. Since the high-resolution image $\mathbf{z}^{(t_0)}$ is not available, we use the correlation coefficients between the corresponding low-resolution images. They are given by

$$\mathrm{W}_k = \frac{\mathcal{C}(\mathbf{y}^{(t_k)}, \mathbf{y}^{(t_0)})}{\mathcal{C}(\mathbf{y}^{(t_k)}, \mathbf{y}^{(t_0)}) + \mathcal{C}(\mathbf{y}^{(t_{k+1})}, \mathbf{y}^{(t_0)})}, \tag{22}$$

$$\mathrm{W}_{k+1} = \frac{\mathcal{C}(\mathbf{y}^{(t_{k+1})}, \mathbf{y}^{(t_0)})}{\mathcal{C}(\mathbf{y}^{(t_k)}, \mathbf{y}^{(t_0)}) + \mathcal{C}(\mathbf{y}^{(t_{k+1})}, \mathbf{y}^{(t_0)})}, \tag{23}$$

where the operator $\mathcal{C}(\mathrm{M}_1, \mathrm{M}_1)$ represents the correlation coefficient between image M_1 and image M_2. It should be noted that $\mathrm{W}_k + \mathrm{W}_{k+1} = 1$ is satisfied. For the case in which only one nearest high-resolution image of date t_0 is available, we can have $\mathrm{W}_k = 1$ and $\mathrm{W}_{k+1} = 0$.

3.2. Statistical Parameters: Covariance Estimates

The unexplained temporal correlation and variability from the mean vector and the local fluctuations may be estimated by the covariance matrix. In Formulae (17) and (18), besides the mean vector, the other group of statistical parameters to be estimated are the cross-covariance matrices $\mathbf{C}_{\mathbf{z}_i^{(t_0)}, \mathbf{z}_i^{(t_0)}}$, $\mathbf{C}_{\mathbf{z}_i^{(t_0)}, \mathbf{X}_i}$, and $\mathbf{C}_{\mathbf{X}_i, \mathbf{X}_i}$. For simplification, we define a new joint random vector

$\mathbf{U}_i = [\mathbf{X}_i^{\mathrm{T}}, (\mathbf{z}_i^{(t_0)})^{\mathrm{T}}]^{\mathrm{T}}$. Then the joint covariance matrix of $\mathbf{U}_i$ is a combination of those cross-covariance matrices expressed as

$$\mathbf{C}_{\mathbf{U}_i} = \begin{bmatrix} \mathbf{C}_{\mathbf{X}_i,\mathbf{X}_i} & \mathbf{C}^{\mathrm{T}}_{\mathbf{z}_i^{(t_0)},\mathbf{X}_i} \\ \mathbf{C}_{\mathbf{z}_i^{(t_0)},\mathbf{X}_i} & \mathbf{C}_{\mathbf{z}_i^{(t_0)},\mathbf{z}_i^{(t_0)}} \end{bmatrix}.$$

The problem then boils down to estimating $\mathbf{C}_{\mathbf{U}_i}$.

It is difficult to directly derive the joint covariance from the high-resolution-image sequences. The joint covariance from the observed low-resolution images is employed as a substitute for that of the high-resolution images.

We form the joint random vector $\mathbf{V}_j = [(\mathbf{y}_j^{(t_1)})^{\mathrm{T}}, (\mathbf{y}_j^{(t_2)})^{\mathrm{T}}, \ldots, (\mathbf{y}_j^{(t_S)})^{\mathrm{T}}, (\mathbf{y}_j^{(t_0)})^{\mathrm{T}}]$, $j = 1, \ldots, M$, at low-resolution with respect to $\mathbf{U}_i$ and group these vectors into D clusters using the k-means clustering method, which classifies pixels with the same temporal dynamics to the same cluster. D clusters mean that there are D types of temporal evolution for the multi-temporal MODIS images. It should be noted that other unsupervised classification algorithms such as iterative self-organizing data analysis technique (ISODATA) and Gaussian mixture models may be equally used [28,55]. For each cluster, we calculate the covariance that represents the time dependency and correlation among the image sequences, and the mean vector that indicates the cluster centroid. Then, in order to estimate each $\mathbf{C}_{\mathbf{U}_i}$, we assign $\mathbf{U}_i$ to a cluster and set $\mathbf{C}_{\mathbf{U}_i}$ to be the cluster covariance based on the Euclidean distance between $\mathbf{U}_i$ and the cluster centroid. We use the interpolated $\mathbf{y}^{(t_0)}$ to substitute $\mathbf{z}^{(t_0)}$ in the assignment because $\mathbf{z}_i^{(t_0)}$ is unknown. With these joint covariance and mean vectors estimated, the MAP estimate can be established. It should be noted that by using IOE and SOE for mean vector estimation, we have two Bayesian data fusion approaches, STBDF-I and STBDF-II, respectively.

4. Empirical Applications and Algorithm Evaluation

In this section, our two proposed algorithms are evaluated through the comparison with the well-known STARFM and ESTARFM algorithms. The high-resolution image in our algorithms is from the Landsat-7 ETM+ data while the low-resolution image is from the MODIS data. To estimate the Landsat image on the target date (say t_0), the inputs are from the MODIS image on t_0 together with both the Landsat and MODIS images from the two nearest dates before and after t_0. It is important to note that though we use two Landsat–MODIS pairs as inputs, one pair or more than two pairs of images can also be handled by our algorithms. We construct three tests to evaluate our algorithms: one test with simulated MODIS images and two tests with real MODIS images. The description of them will be detailed in the following subsections.

Before fusion, the Landsat images are radiometrically calibrated and geometrically rectified. The MODIS images are re-projected onto the WGS84 coordinate system to make them consistent with the Landsat images using the MODIS Reprojection Tools (MRT). Similar to STARFM and ESTARFM algorithms, here we use three bands: green, red, and near-infrared (NIR), commonly used for land-cover classifications and detections of vegetation phenology changes. Their spectral ranges are highlighted in Table 1 [3]. In this study, we independently test our algorithms band by band, though a more general test by putting all bands together and incorporating spectral correlations can be achieved with our algorithms.

Table 1. Landsat Enhanced Thematic Mapper Plus (ETM+) and Moderate Resolution Imaging Spectroradiometer (MODIS) Bandwidth.

Band Name	Landsat		MODIS	
	Band Number	Bandwidth	Band Number	Bandwidth
Blue	1	450–520 nm	3	459–479 nm
Green	2	530–610 nm	4	545–565 nm
Red	3	630–690 nm	1	620–670 nm
Near Infrared	4	780–900 nm	2	841–876 nm
Middle Infrared	5	1550–1750 nm	6	1628–1652 nm
Middle Infrared	7	2090–2350 nm	7	2105–2155 nm

4.1. Metrics for Performance Assessment

We employ four assessment metrics, namely the average absolute difference (AAD), root-mean-square error (RMSE), correlation coefficient (CC), and the Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS), to assess the performance of the algorithms, where the AAD, RMSE, and CC are commonly used metrics. The metric ERGAS measures the similarity between the fused and the reference images. A lower ERGAS value indicates a higher fusion quality. ERGAS differs from the other metrics by treating all three bands together. Its form is as follows [33]:

$$\text{ERGAS} = 100\frac{h}{l}\sqrt{\frac{1}{N_{\text{ban}}}\sum_{k=1}^{N_{\text{ban}}}(RMSE_k/M_k)^2}, \tag{24}$$

where h is the spatial resolution of the Landsat image, l is the spatial resolution of the MODIS image, N_{ban} is the number of spectral bands (3 for this experiment), and M_k is the mean value of band k.

Besides the quantitative metrics, we employ visual assessments and scatter plots to provide an intuitive understanding of the fusion quality.

4.2. Test with Simulated Images

In the first test, we use the real Landsat images from satellites and the simulated MODIS images that are resampled from Landsat images. In this way, we expect to eliminate the influence of the mismatching factors such as bandwidth, acquisition time, and geometric inconsistencies between the Landsat ETM+ and the MODIS sensors. We employ the Landsat-7 ETM+ images with a 30-m resolution from the Canadian Boreal Ecosystem-Atmosphere Study (BOREAS) provided by [5] and also used in [9]. We use a subset of 400 × 400 pixels extracted from the original image. Three Landsat images acquired on 24 May, 11 July, and 12 August in 2001 are used. They are from the summer growing season, characterized by large vegetation phenology changes. For each image, we use three spectral bands: green, red, and NIR. We simulated the low-resolution image, termed the simulated MODIS, by aggregating every 17 × 17 pixels in the Landsat image without overlapping to degrade the resolution from 30 m to 500 m. Figure 2 shows the images using NIR–red–green as red–green–blue composite for both the Landsat and simulated MODIS images. This study area has relatively simple land-cover distribution with large land-cover patches. Most temporal changes are due to phenology changes during the period.

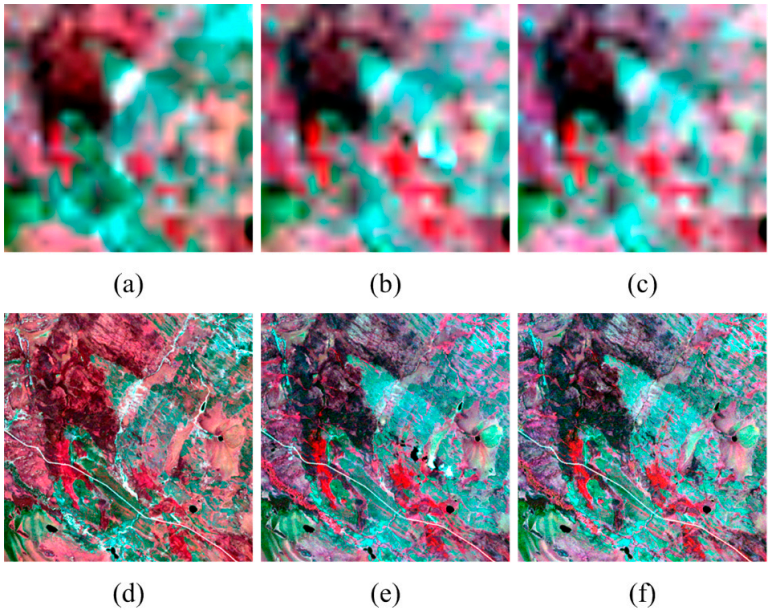

Figure 2. Near-infrared (NIR)–green–blue composites of the simulated-MODIS imagery (**upper row**) and Landsat imagery (**lower row**) taken on (**a**,**d**) 24 May 2001, (**b**,**e**) 11 July 2001, and (**c**,**f**) 12 August 2001, respectively.

The Landsat image on 11 July 2001 is set as the target image for prediction. The other Landsat and simulated MODIS images are used as inputs for STARFM, ESTARFM, and our proposed STBDF-I and STBDF-II algorithms, respectively. The predicted high-resolution images are compared with the reference Landsat image to evaluate the performance of the algorithms.

Figure 3 shows the reference image on 11 July 2001 and the predicted ones by STARFM, ESTARFM, STBDF-I, and STBDF-II, respectively. It is clear that all four algorithms have good performance in capturing the main reflectance changes over time. STARFM performs almost equally well with ESTARFM. STBDF-I and -II are better in capturing the reflectance changes and fine spatial details of small parcels, and in identifying objects with sharp edges and bright colors. The images predicted by STBDF-I and -II seem to be more similar to the reference image and maintain more spatial details than those obtained by STARFM and ESTARFM. Their differences are highlighted in the zoomed-in images shown in Figure 4. Clear shapes and distinct textures can be found in the results of our algorithms (Figure 4b,c), while the shapes are blurry and some spatial details are missing in the results of STARFM and ESTARFM (Figure 4d,e); the shapes are even difficult to identify in Figure 4d. For the second set of the zoomed-in images (Figure 4f–j), the color of the images obtained from our algorithms (Figure 4g,h) are closer to the reference image (Figure 4f) than the results of STARFM and ESTARFM (Figure 4i,j), respectively, which are distorted. The red color in the result of STARFM (Figure 4i) turns out to be a little bit pink and the green color in the result of ESTARFM (Figure 4j) looks greener than in the reference image (Figure 4f).

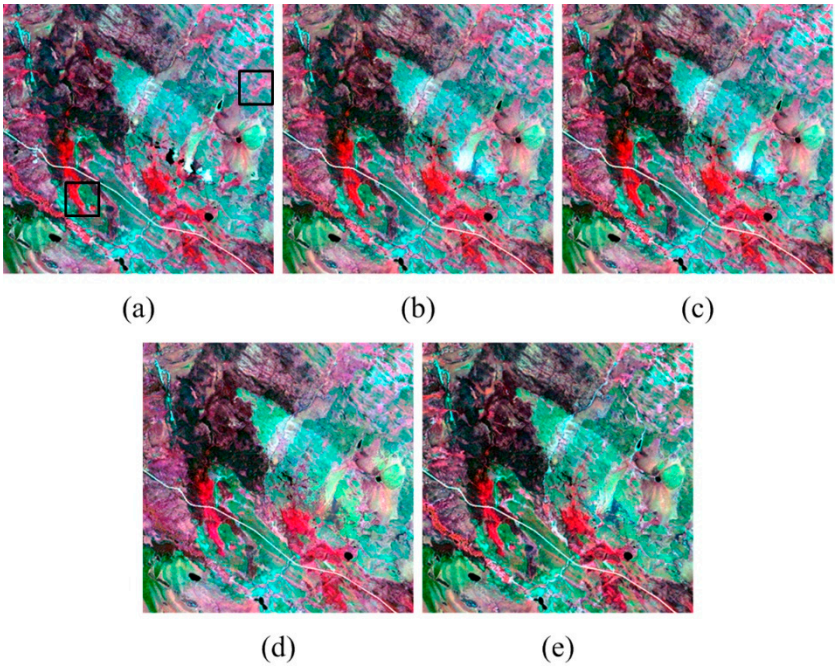

Figure 3. The (**a**) actual image acquired on 11 July 2001 and its predictions by (**b**) STBDF-I (Spatio-Temporal Bayesian Data Fusion I), (**c**) STBDF-II, (**d**) STARFM (Spatial and Temporal Adaptive Reflectance Fusion Model), and (**e**) ESTARFM (enhanced STARFM).

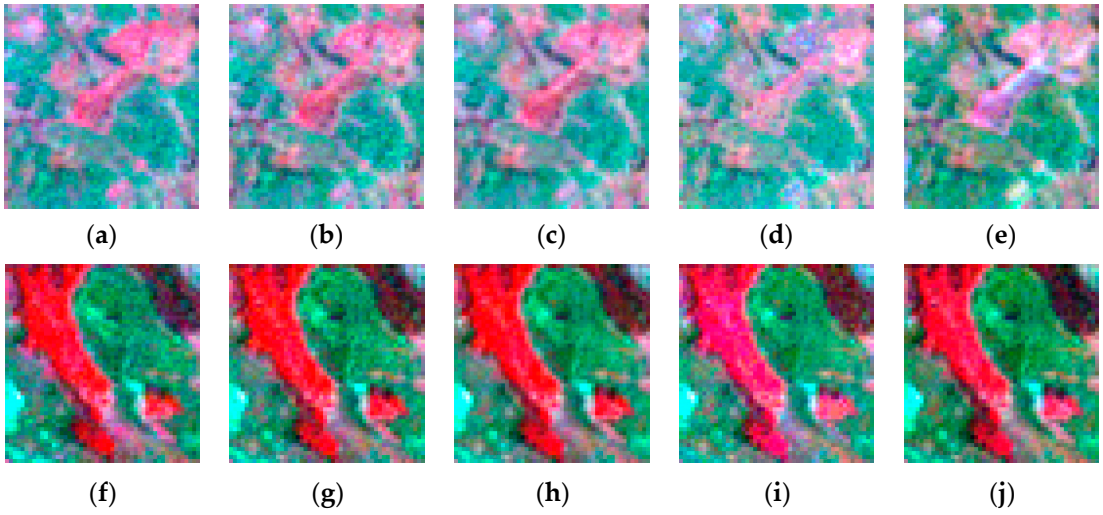

Figure 4. Zoomed-in images of the upper-right box (**top row**) and lower-left box (**bottom row**) marked in Figure 3a corresponding to the images in Figure 3a–e.

In order to quantitatively evaluate these approaches, AAD, RMSE, CC, and ERGAS are computed and shown in Table 2 for the three bands. The values of AAD, RMSE, and CC are comparable between STARFM and ESTARFM. For the green band, STARFM performs slightly better than ESTARFM, while for red and NIR bands, ESTARFM shows better performance. In terms of ERGAS, which evaluates the overall performance by putting the three bands together, the prediction error of ESTARFM is smaller than that of STARFM, in agreement with [9]. In comparison, our proposed STBDF-I and -II both exhibit improved prediction than STARFM and ESTARFM in terms of the three bands and the four metrics. Especially, STBDF-II outperforms STBDF-I, implying that adding information from the high-frequency image of the nearest Landsat images is capable of enhancing the fusion result.

In summary, the predicted Landsat images obtained from STBDF-I and -II show better performance compared to the reference image than those from STARFM and ESTARFM, and STBDF-II gives the most accurate result in terms of both visual analysis and quantitative assessment.

Table 2. Quantitative metrics of the fusion results of the simulated images.

	AAD			RMSE			CC			ERGAS
Approaches	**Green**	**Red**	**NIR**	**Green**	**Red**	**NIR**	**Green**	**Red**	**NIR**	**Overall**
STARFM	0.0033	0.0038	0.0111	0.0084	0.0088	0.0182	0.6679	0.7513	0.9001	1.0551
ESTARFM	0.0038	0.0037	0.0111	0.0086	0.0085	0.0164	0.6670	0.7705	0.9252	1.0312
STBDF-I	0.0032	0.0037	0.0089	0.0077	0.0083	0.0143	0.7344	0.7867	0.9382	0.9718
STBDF-II	0.0029	0.0034	0.0089	0.0075	0.0081	0.0143	0.7475	0.7995	0.9386	0.9461

AAD, Average Absolute Difference; RMSE, Root-Mean-Square Error; CC, Correlation Coefficient; ERGAS, Erreur Relative Globale Adimensionnelle de Synthèse.

4.3. Test with Satellite Images

It is difficult to eliminate the influence of mismatching factors such as bandwidth, acquisition time, and geometric inconsistencies between two different sensors (i.e., Landsat and MODIS). With these in mind, we construct two test cases using real MODIS images. Since STBDF-II outperforms STBDF-I in the test using simulated MODIS images, we focus on STBDF-II in the following two tests. The first test employs the data source same as the test discussed in the previous subsection while the second test focuses on the Panyu area of Guangzhou city in China, where the landscape has more heterogeneous characteristics and small patches, which generally results in more mixed pixels in the MODIS images [9].

4.3.1. Experiment on a Homogeneous Region

In this experiment, we use the same data source as the previous test. The difference is that we employ three real Landsat–MODIS image-pairs acquired on 24 May, 11 July, and 12 August in 2001, respectively. Another difference from the previous test is that we use a subset of 800 × 800 pixels from the original images in [5]. We focus on three bands: green, red, and NIR. The NIR–red–green composites of both Landsat and MODIS images are shown in Figure 5. Similar to the previous test, the objective here is to predict the high-resolution image on 11 July 2001 using all other images as inputs.

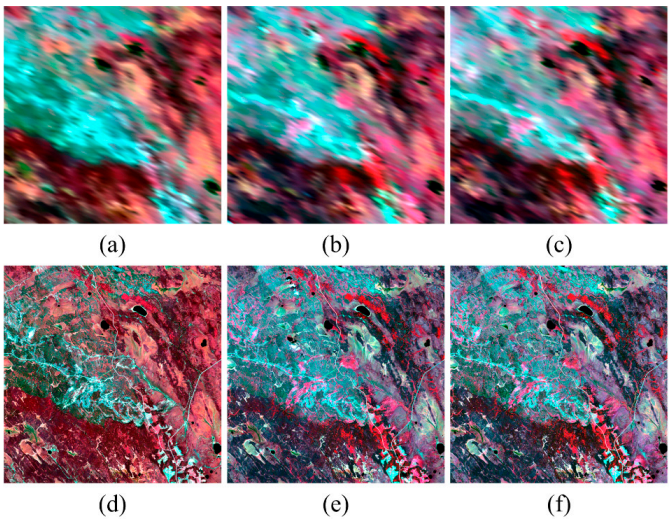

Figure 5. NIR–green–blue composites of the MODIS imagery (**upper row**) and Landsat imagery (**lower row**) taken on (**a**,**d**) 24 May 2001, (**b**,**e**) 11 July 2001, and (**c**,**f**) 12 August 2001, respectively.

In this experiment, we employed STARFM, ESTARFM, and STBDF-II to predict the target Landsat image on 11 July 2001. As shown in Figure 6, all fused images obtained from the three algorithms capture the main spatial pattern and subtle details of the reference Landsat image. The upper-left portion of the predicted image from STARFM seems to have some color distortion, which is greener than in the three other images. This may be due to the relatively poorer performance of STARFM in predicting the red band than the other two methods, consistent with the quantitative results shown in Table 3. The results obtained by STBDF-II are similar to ESTARFM in both color and spatial details. Figure 7 shows the scatter plots of the predicted reflectance against the reference reflectance for each pixel in the NIR, red, and green bands, respectively. In terms of NIR and green bands, it appears that the points from STARFM and ESTARFM are more scattered, while those from STBDF-II are closer to the 1:1 line. For the red band, ESTARFM seems to produce better scatter plots than STARFM and our STBDF-II method. These results are consistent with the quantitative assessment results shown in Table 3. In terms of the synthetic metric ERGAS, STBDF-II (1.0196) outperforms STARFM (1.1493) and ESTARFM (1.0439).

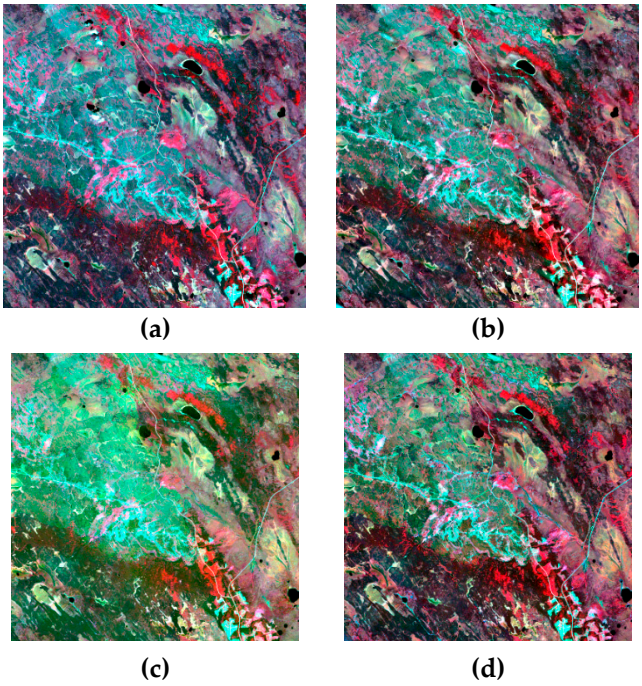

Figure 6. The (**a**) actual image acquired on 11 July 2001 and its predictions by (**b**) STBDF-II; (**c**) STARFM; and (**d**) ESTARFM.

Table 3. Quantitative metrics of the fusion results applied to the Boreal images.

	AAD			RMSE			CC			ERGAS
Approaches	**Green**	**Red**	**NIR**	**Green**	**Red**	**NIR**	**Green**	**Red**	**NIR**	**Overall**
STARFM	0.0036	0.0044	0.0133	0.0076	0.0099	0.0207	0.7818	0.7700	0.9102	1.1493
ESTARFM	0.0041	0.0040	0.0159	0.0081	0.0081	0.0226	0.7575	0.8439	0.8913	1.0439
STBDF-II	0.0035	0.0043	0.0132	0.0075	0.0084	0.0191	0.7939	0.8340	0.9216	1.0196

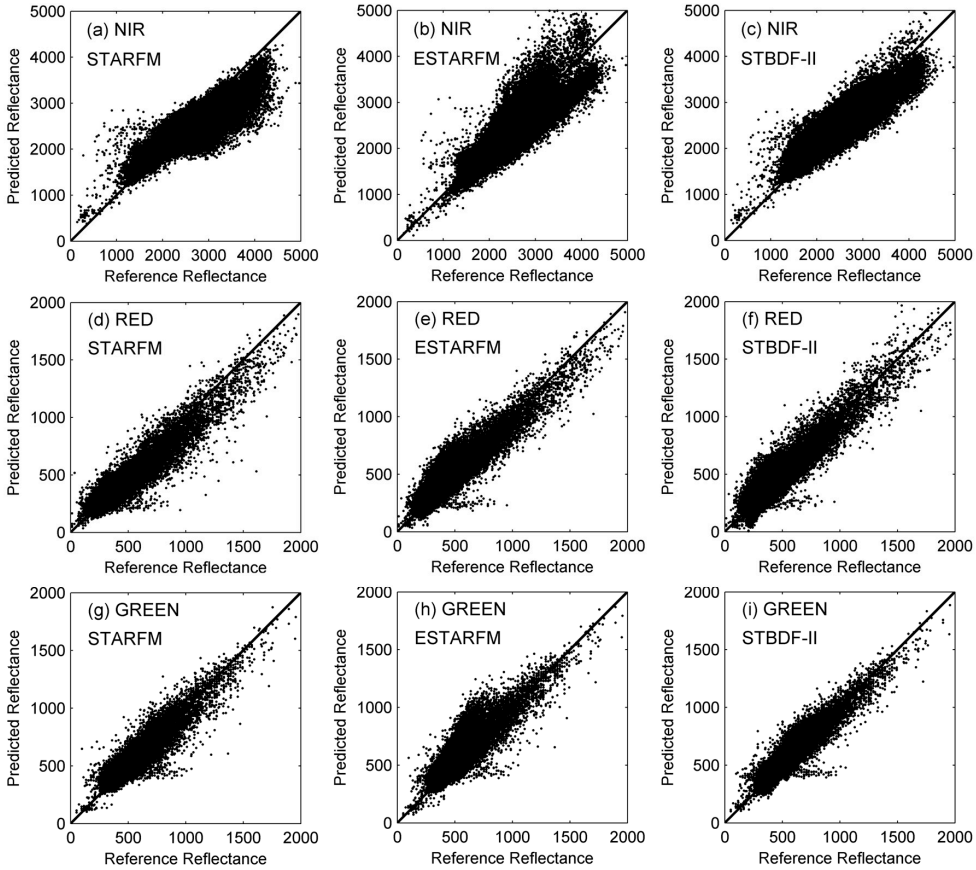

Figure 7. Scatter plots of the predicted reflectance against the actual reflectance. (**a**–**c**) Scatter plots of the predicted reflectance using STARFM, ESTARFM, and STBDF-II against the actual reflectance in the NIR band, respectively; (**d**–**f**) the red band; (**g**–**i**) the green band.

4.3.2. Experiment on a Heterogeneous Region

In this experiment, our satellite images were acquired over the Panyu area in Guangzhou, China. Panyu is an important area for crop production in the Pearl River Delta with complex land-cover types, including cropland, farmland, forest, bare soil, urban land, and water body. In this area, Landsat-7 ETM+ images (30 m) with 300 × 300 pixels and MODS images (1000 m) over the same region were acquired on 29 August 2000, 13 September 2000, and 1 November 2000, respectively (Figure 8). Changes in the vegetation phenology are clearly observed in the Landsat images from 13 September to 1 November 2000 (Figure 8e,f). The high-resolution image on 13 September 2000 is targeted for prediction. All five other images are used as inputs for STARFM, ESTARFM, and STBDF-II, respectively.

Figure 9 depicts the reference Landsat image and the three predicted images from STARFM, ESTARFM, and STBDF-II, respectively. The three algorithms are all capable of predicting the target image by capturing many spatial details. It appears that colors of the image predicted by STBDF-II are visually more similar to the reference image in terms of better hue and saturation than those by STARFM and ESTARFM. Moreover, STBDF-II captures almost all the spatial details and fine textures in the reference image, while STARFM and ESTARFM fail to maintain some details. To highlight the visual difference in Figure 9, we give two sets of zoomed-in images as illustrated in Figure 10 with respect to the two black boxes shown in Figure 9. As shown in the first set of the zoomed-in images (Figure 10a–d), the colors and the shapes of vegetation patches in the image generated by STBDF-II (Figure 10b) is closer to the reference image (Figure 10a) than those by STARFM and ESTARFM (Figure 10c,d). Some pixels produced by STARFM and ESTARFM seem to have abnormal values, leading to a visually brown color in several patches. From the second set of the zoomed-in images

(Figure 10e–h), it is clear that the shapes of the buildings predicted by STBDF-II (Figure 10f) are more similar to those in the reference image (Figure 10e). The buildings (in blue color) in the middle of the predicted images by STARFM and ESTARFM (Figure 10g,h) seem to lose some spatial details and shapes with blurry boundaries. Figure 11 shows the scatter plots of the predicted reflectance against the reference reflectance in the NIR, red, and green bands, respectively. Results of the quantitative measures are summarized in Table 4. STBDF-II achieves the best results, except for the AAD and RMSE values for the red band that are, however, very close to those of the other two algorithms (STARFM produces a slightly better result). Nevertheless, the ERGAS values in Table 4 indicate that the fusion quality of STBDF-II outperforms those of STARFM and ESTARFM from a synthetic perspective.

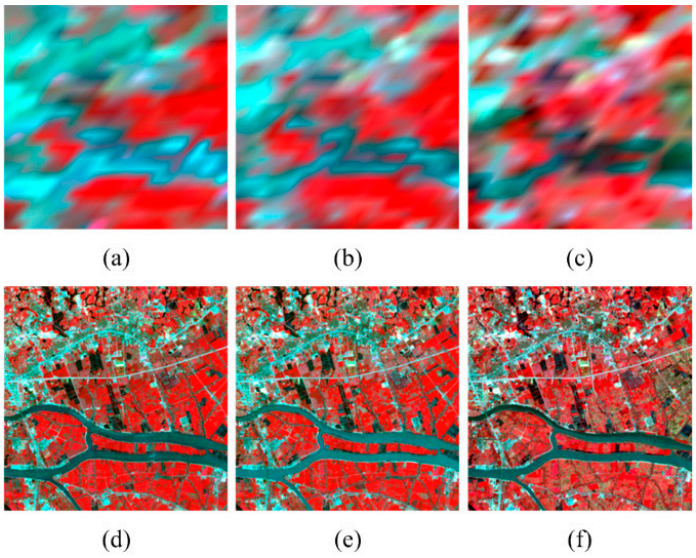

Figure 8. NIR–green–blue composites of MODIS imagery (upper row) and Landsat imagery (lower row) taken on (**a**,**d**) 29 August 2000, (**b**,**e**) 13 September 2000, and (**c**,**f**) 1 November 2000, respectively.

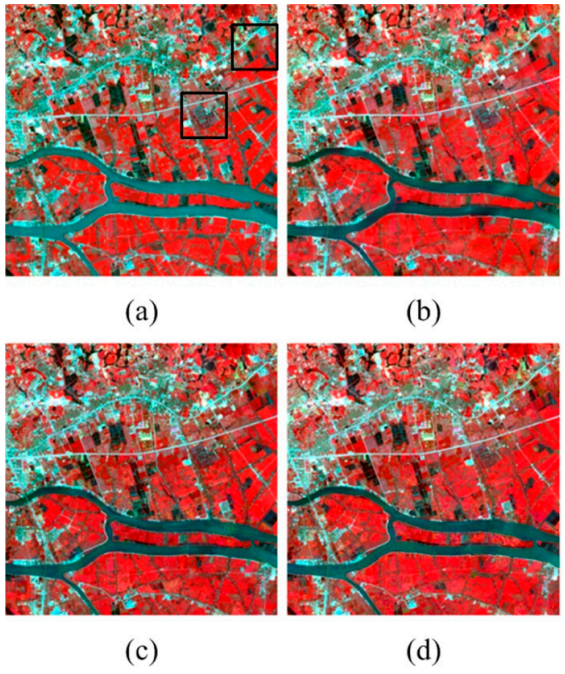

Figure 9. The (**a**) actual image acquired on 13 September 2000 and its predictions by (**b**) STBDF-II, (**c**) STARFM, and (**d**) ESTARFM.

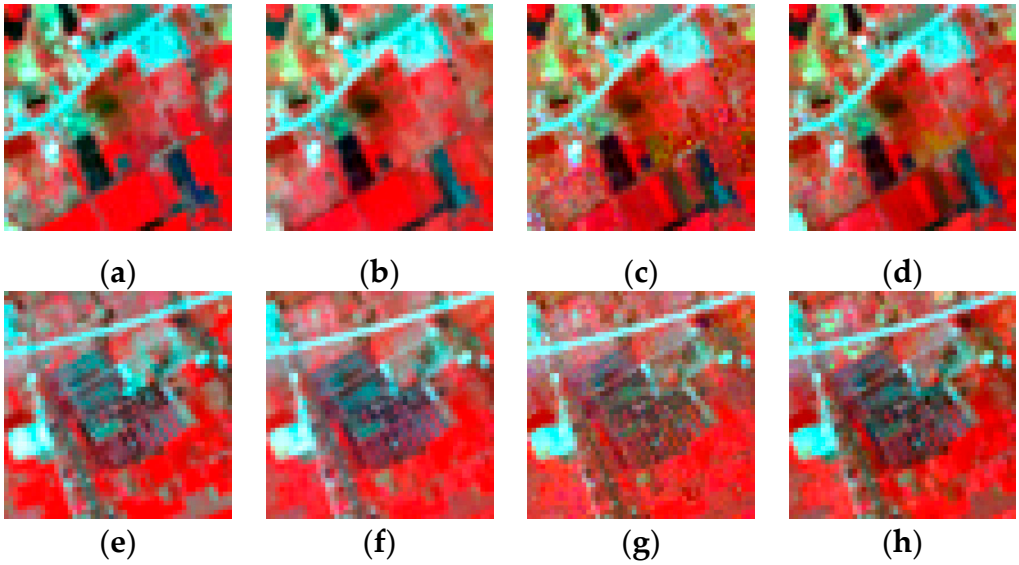

Figure 10. Zoomed-in images of the upper-right box (**top row**) and lower-left box (**bottom row**) marked in Figure 10a corresponding to the images in Figure 10a–d.

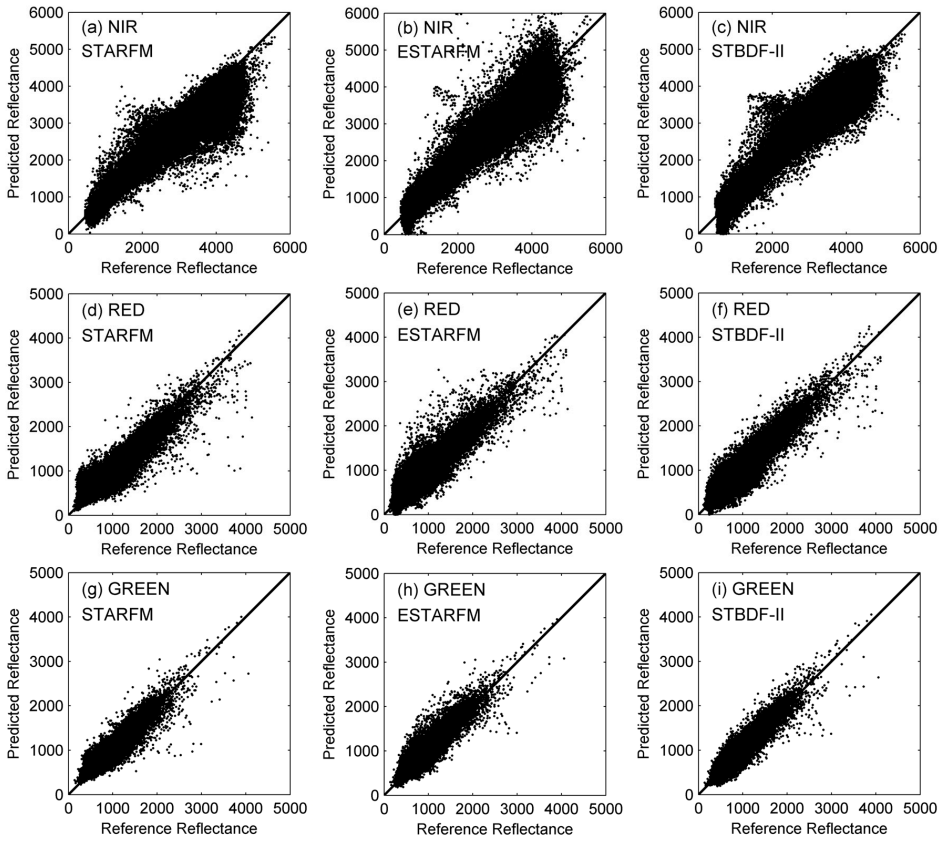

Figure 11. Scatter plots of the predicted reflectance against the actual reflectance. (**a**–**c**) Scatter plots of the predicted reflectance using STARFM, ESTARFM, and STBDF-II against the actual reflectance in the NIR band, respectively; (**d**–**f**) the red band; (**g**–**i**) the green band.

Table 4. Quantitative metrics of the fusion results applied to the Panyu images.

	AAD			RMSE			CC			ERGAS
Approaches	**Green**	**Red**	**NIR**	**Green**	**Red**	**NIR**	**Green**	**Red**	**NIR**	**Overall**
STARFM	0.0109	0.0130	0.0267	0.0140	0.0204	0.0383	0.9057	0.8998	0.9376	1.1442
ESTARFM	0.0104	0.0130	0.0244	0.0136	0.0205	0.0347	0.9157	0.8997	0.9420	1.1220
STBDF-II	0.0104	0.0140	0.0243	0.0133	0.0205	0.0333	0.9208	0.9049	0.9443	1.1087

5. Discussion

The fusion results with simulated data demonstrated that our methods are more accurate than both STARFM and ESTARFM, and that STBDF-II outperforms STBDF-I. The improved performance of STBDF-II over STBDF-I highlights the importance of incorporating more temporal information for parameter estimation. The better performance of STBDF-II over STARFM and ESTARFM is shown in another two tests using acquired real Landsat–MODIS image pairs. STARFM is specifically developed for homogeneous landscapes, while STBDF-II, to some extent, incorporates sub-pixel heterogeneity into fusion process through the employment of high frequencies of high-resolution images, which may partly explain the outperformance of our algorithm over STARFM. ESTARFM assumes a linear trend for temporal changes, which does not hold for the three bands in terms of mean reflectance. However, this nonlinear change is well captured by STBDF-II (not shown), which may be partly due to the use of joint Gaussian distribution in our algorithm to simulate temporal evolution of the image series. Though the improvement of our approach over ESTARFM is not as significant as it is over STARFM, the advantage of our approach over ESTARFM is that we impose no requirements on the number of input image pairs, and it can work on one input image pair. ESTARFM can only fuse images based on two input image pairs.

Our study employed two datasets with different landscape heterogeneities to understand the utility of our algorithm. Following the results section, here we provide more discussions on this issue. To objectively define the sub-pixel heterogeneity, we compute the standard deviation (SD) of reflectance for Landsat pixels within each MODIS pixel. High SD indicates high heterogeneity. Within each MODIS pixel, we can also compute relative average absolute difference (RAAD) to represent the bias of the predicted Landsat image relative to the reference Landsat image. The changing trend of RAAD with respect to SD would shed light on the impact of heterogeneity on the performance of our algorithm at MODIS pixel scales (Figure 12). Reasonably, we observe that the homogeneous Boreal data generally shows lower SD than the heterogeneous Panyu data. In red and green bands, SDs are well concentrated below 0.2 for Boreal data, while most pixels in Panyu data have SDs larger than 0.2. In the NIR band, SDs approximately range from 0 to 0.1 for Boreal data, and from 0.02 to 0.14 for Panyu data. For the two datasets with different heterogeneities, the RAAD of most pixels from the three bands is well constrained at low values, suggesting the skill of our algorithm for landscapes with different complexities. Another interesting aspect is that the lower boundary of RAAD generally shows a slowly increasing trend with respect to SD, suggesting that increased heterogeneity would increase the difficulty of accurate fusion. However, the trend of increase is quite flat (Figure 12). The varying upper boundary may indicate the impacts from other factors.

The assumptions for estimating the mean vector and covariance matrix of the joint Gaussian distribution of the high-resolution image series may have impact on the fusion accuracy. Here we conduct several experiments to explore this impact quantitatively in terms of CC and ERGAS. For the mean estimate, we assume the target high-resolution image is available and add the high frequencies of it to the interpolated low-resolution image to estimate the mean vector, which is named as case ideal-I. For the covariance estimate, we assign the target high-resolution image instead of the interpolated low-resolution image to clusters to estimate the covariance, named case ideal-II. We use Panyu dataset as an example. Results from the two cases are compared with the control case in which the assumptions are not relaxed, respectively (Tables 5 and 6). As shown in Table 5, the use of high frequencies from

target image for mean vector estimates improves the fusion results in terms of CC and ERGAS, especially for green and red bands. However, we do not find the significant improvements through the use of reference high-resolution image for covariance estimation (Table 6). These results suggest that the fusion accuracy is more sensitive to approximations in mean estimates than in covariance estimates. It should be noted that the mean has a much higher contribution than that of the covariance to the fused image, which may partly explain the sensitivity we see.

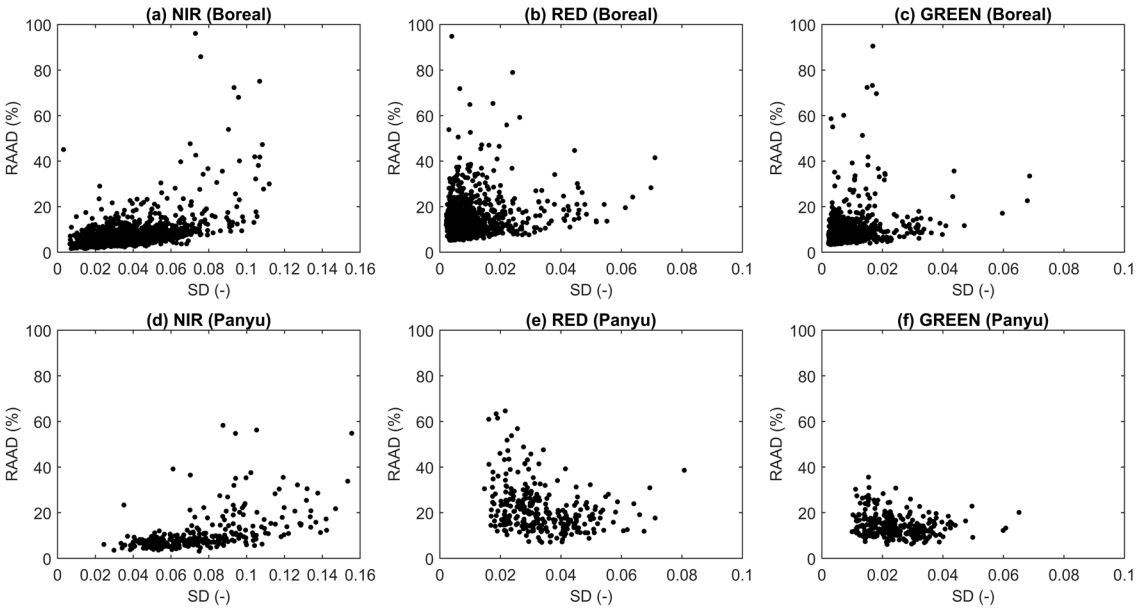

Figure 12. Changes of relative average absolute difference (RAAD) with respect to standard deviation (SD) of reflectance from Landsat pixels within each MODIS pixel for (**a**–**c**) Boreal data and (**d**–**f**) Panyu data.

Table 5. CC and ERGAS of the fusion results applied to the Panyu images for both control and Ideal-I cases.

	CC			ERGAS
Approaches	**Green**	**Red**	**NIR**	**Overall**
Control	0.9208	0.9049	0.9443	1.1087
Ideal-I	0.9776	0.9750	0.9740	0.6764

Table 6. CC and ERGAS of the fusion results applied to the Panyu images for both control and Ideal-II cases.

	CC			ERGAS
Approaches	**Green**	**Red**	**NIR**	**Overall**
Control	0.9208	0.9049	0.9443	1.1087
Ideal-II	0.9208	0.9047	0.9445	1.1085

Although the proposed Bayesian approach enhances our capability in generating remotely sensed images high in both spatial and temporal resolutions, there are still limitations and constraints that need further study. Firstly, the factors related to different sensors (i.e., different acquisition times, bandwidths, spectral response functions, etc.) and the model assumptions (i.e., rectangular PSF, joint Gaussian distribution, etc.) may affect the accuracy of fusion results. Secondly, we have not considered the hyperparameter prior information of the targeted high-resolution image, which might need to be

taken into consideration in future studies. Thirdly, our approach, as well as STARFM and ESTARFM, have not fully incorporated the mixed-class spectra within a low-resolution pixel, which might limit its potential to retrieve the accurate class spectra in high resolution on the target date. Thus, we will combine the unmixing method together with the proposed Bayesian approach in a unified way in the future. Finally, the proposed framework can be extended to incorporate time series of low-resolution images in the fusion process. This will be considered in future studies.

6. Conclusions

In this study, we have developed a Bayesian framework for the fusion of multi-scale remotely sensed images to obtain images high in both spatial and temporal resolutions. We treat fusion as an estimation problem in which the fused image is estimated by first constructing a first-order observational model between the target high-resolution image and the corresponding low-resolution image, and then by incorporating the information of the temporal evolution of images into a joint Gaussian distribution. The Bayesian estimator provides a systematic way to solve the estimation problem by maximizing the posterior probability conditioned on all other available images. Based on this approach, we have proposed two algorithms, namely STBDF-I and STBDF-II. STBDF-II is an improved version of STBDF-I through the incorporation of more information from the high-frequency images of the high-resolution images nearest to the target image to produce better mean estimates.

We have tested the proposed approaches with an experiment using simulated data and two experiments using acquired real Landsat–MODIS image pairs. Through the comparison with two widely used fusion algorithms STARFM and ESTARFM, our algorithms show better performance based on visual analysis and quantitative metrics, especially for NIR and green bands. In terms of a comprehensive metric ERGAS, STBDF-II improves the fused images by approximately 3–11% over STARFM and approximately 1–8% over ESTARFM for the three experiments.

In summary, we have proposed a formal and flexible Bayesian framework for satellite image fusion which gives spatio-temporal fusion of remotely sensed images a solid theoretical and empirical foundation on which rigorous statistical procedures and tests can be formulated for estimations and assessments. It imposes no limitations on the number of input high-resolution images and can generate images high in both spatial and temporal resolutions, regardless of over-homogeneous or relatively heterogeneous areas.

Acknowledgments: This study was supported by the earmarked grant CUHK 444411 of the Hong Kong Research Grant Council. The authors would like to thank F. Gao and X. Zhu for making available the data and their algorithms on the internet for our empirical analysis and comparison. The authors would like to thank Y. Xu for providing the Panyu dataset to us for the empirical analysis.

Author Contributions: Jie Xue, Yee Leung, and Tung Fung conceived the research and designed the experiments. Jie Xue and Yee Leung developed the theoretical framework and wrote the paper. Jie Xue analyzed the data and performed the experiments. Tung Fung, Jie Xue, and Yee Leung interpreted the experimental results. All authors reviewed and approved the manuscript.

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

References

1. Pohl, C.; Van Genderen, J.L. Review article multisensor image fusion in remote sensing: Concepts, methods and applications. *Int. J. Remote Sens.* **1998**, *19*, 823–854. [CrossRef]
2. Ju, J.; Roy, D.P. The availability of cloud-free landsat etm+ data over the conterminous united states and globally. *Remote Sens. Environ.* **2008**, *112*, 1196–1211. [CrossRef]
3. 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]
4. Chen, B.; Huang, B.; Xu, B. Comparison of spatiotemporal fusion models: A review. *Remote Sens.* **2015**, *7*, 1798–1835. [CrossRef]
5. 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.

6. Shen, H.; Wu, P.; Liu, Y.; Ai, T.; Wang, Y.; Liu, X. A spatial and temporal reflectance fusion model considering sensor observation differences. *Int. J. Remote Sens.* **2013**, *34*, 4367–4383. [CrossRef]
7. Roy, D.P.; Ju, J.; Lewis, P.; Schaaf, C.; Gao, F.; Hansen, M.; Lindquist, E. Multi-temporal modis–landsat data fusion for relative radiometric normalization, gap filling, and prediction of landsat data. *Remote Sens. Environ.* **2008**, *112*, 3112–3130. [CrossRef]
8. 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]
9. 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]
10. Fu, D.; Chen, B.; Wang, J.; Zhu, X.; Hilker, T. An improved image fusion approach based on enhanced spatial and temporal the adaptive reflectance fusion model. *Remote Sens.* **2013**, *5*, 6346–6360. [CrossRef]
11. 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]
12. 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]
13. Singh, D. Evaluation of long-term ndvi time series derived from landsat data through blending with modis data. *Atmósfera* **2012**, *25*, 43–63.
14. Hwang, T.; Song, C.; Bolstad, P.V.; Band, L.E. Downscaling real-time vegetation dynamics by fusing multi-temporal modis and landsat ndvi in topographically complex terrain. *Remote Sens. Environ.* **2011**, *115*, 2499–2512. [CrossRef]
15. Gray, J.; Song, C. Mapping leaf area index using spatial, spectral, and temporal information from multiple sensors. *Remote Sens. Environ.* **2012**, *119*, 173–183. [CrossRef]
16. 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]
17. 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]
18. Hilker, T.; Wulder, M.A.; Coops, N.C.; Linke, J.; McDermid, G.; Masek, J.G.; Gao, F.; White, J.C. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on landsat and modis. *Remote Sens. Environ.* **2009**, *113*, 1613–1627. [CrossRef]
19. 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]
20. 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]
21. 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]
22. 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]
23. 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]
24. Huang, B.; Song, H. Spatiotemporal reflectance fusion via sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3707–3716. [CrossRef]
25. 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]
26. 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]
27. 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]
28. 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]

29. 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]
30. 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]
31. 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]
32. 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]
33. 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]
34. 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] [PubMed]
35. 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.
36. 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.; InTech: Rijeka, Croatia, 2009.
37. 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.
38. 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]
39. Zhang, H.; Huang, B. A new look at image fusion methods from a bayesian perspective. *Remote Sens.* **2015**, *7*, 6828–6861. [CrossRef]
40. 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]
41. Fasbender, D.; Radoux, J.; Bogaert, P. Bayesian data fusion for adaptable image pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1847–1857. [CrossRef]
42. Zhang, Y. Wavelet-based bayesian fusion of multispectral and hyperspectral images using gaussian scale mixture model. *Int. J. Image Data Fusion* **2012**, *3*, 23–37. [CrossRef]
43. Joshi, M.; Jalobeanu, A. Map estimation for multiresolution fusion in remotely sensed images using an igmrf prior model. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 1245–1255. [CrossRef]
44. Wei, Q.; Dobigeon, N.; Tourneret, J.-Y. Bayesian fusion of multi-band images. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 1117–1127. [CrossRef]
45. Šroubek, F.; Flusser, J. Resolution enhancement via probabilistic deconvolution of multiple degraded images. *Pattern Recognit. Lett.* **2006**, *27*, 287–293. [CrossRef]
46. Akhtar, N.; Shafait, F.; Mian, A. Bayesian sparse representation for hyperspectral image super resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3631–3640.
47. Zhang, H.; Zhang, L.; Shen, H. A super-resolution reconstruction algorithm for hyperspectral images. *Signal Process.* **2012**, *92*, 2082–2096. [CrossRef]
48. 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]
49. Sharma, R.K.; Leen, T.K.; Pavel, M. Bayesian sensor image fusion using local linear generative models. *Opt. Eng.* **2001**, *40*, 1364–1376.

50. Hardie, R.C.; Barnard, K.J.; Bognar, J.G.; Armstrong, E.E.; Watson, E.A. High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system. *Opt. Eng.* **1998**, *37*, 247–260.
51. 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. [CrossRef]
52. Kay, S.M. *Fundamentals of Statistical Signal Processing: Estimation Theory*; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1993; p. 595.
53. Leung, Y.; Liu, J.; Zhang, J. An improved adaptive intensity–hue–saturation method for the fusion of remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 985–989. [CrossRef]
54. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A. Context-driven fusion of high spatial and spectral resolution images based on oversampled multiresolution analysis. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2300–2312. [CrossRef]
55. Wu, M.; Niu, Z.; Wang, C.; Wu, C.; Wang, L. Use of modis and landsat time series data to generate high-resolution temporal synthetic landsat data using a spatial and temporal reflectance fusion model. *J. Appl. Remote Sens.* **2012**, *6*, 063507.