

Article

Reconstruction of MODIS Spectral Reflectance under Cloudy-Sky Condition

Bo Gao^{1,2,3,4}, Huili Gong^{1,2,3,4,*}, Tianxing Wang⁵ and Li Jia⁵

¹ Beijing Laboratory of Water Resources Security, Capital Normal University, Beijing 100048, China; gaobo@irsa.ac.cn

² Base of the State Key Laboratory of Urban Environmental Process and Digital Modeling, Capital Normal University, Beijing 100048, China

³ Key Laboratory of 3D Information Acquisition and Application, Ministry of Education, Capital Normal University, Beijing 100048, China

⁴ College of Resources Environment and Tourism, Capital Normal University, Beijing 100048, China

⁵ State Key Laboratory of Remote Sensing Science, The Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences, Beijing 100101, China; wangtx@radi.ac.cn (T.W.); jiali@radi.ac.cn (L.J.)

* Correspondence: gonghl@263.net; Tel.: +86-10-6890-2280

Academic Editors: Dongdong Wang, Richard Müller and Prasad S. Thenkabail

Received: 27 June 2016; Accepted: 29 August 2016; Published: 9 September 2016

Abstract: Clouds usually cause invalid observations for sensors aboard satellites, which corrupts the spatio-temporal continuity of land surface parameters retrieved from remote sensing data (e.g., MODerate Resolution Imaging Spectroradiometer (MODIS) data) and prevents the fusing of multi-source remote sensing data in the field of quantitative remote sensing. Based on the requirements of spatio-temporal continuity and the necessity of methods to restore bad pixels, primarily resulting from image processing, this study developed a novel method to derive the spectral reflectance for MODIS band of cloudy pixels in the visual–near infrared (VIS–NIR) spectral channel based on the Bidirectional Reflectance Distribution Function (BRDF) and multi-spatio-temporal observations. The proposed method first constructs the spatial distribution of land surface reflectance based on the corresponding BRDF and the solar-viewing geometry; then, a geographically weighted regression (GWR) is introduced to individually derive the spectral surface reflectance for MODIS band of cloudy pixels. A validation of the proposed method shows that a total root-mean-square error (RMSE) of less than 6% and a total R^2 of more than 90% are detected, which indicates considerably better precision than those exhibited by other existing methods. Further validation of the retrieved white-sky albedo based on the spectral reflectance for MODIS band of cloudy pixels confirms an RMSE of 3.6% and a bias of 2.2%, demonstrating very high accuracy of the proposed method.

Keywords: MODIS; reflectance; BRDF; cloudy pixel; GWR

1. Introduction

A large amount of cloudy pixels in remote sensing data corrupts the accuracy of the retrieved parameters based on spectral reflectance for the MODerate Resolution Imaging Spectroradiometer (MODIS) band, such as land cover classification, Normalized Difference Vegetation Index (NDVI), and albedo, and obstructs the fusing of multi-source data. Thus, retrieval of reflectance of pixels under clouds is particularly necessary for most remote sensing land applications.

To retrieve data from bad pixels, multiple methods using image processing have been proposed. These existing methods can be divided into three categories: (1) spatial-based methods; (2) spectral-based methods; and (3) temporal-based methods [1].

Specifically, a few representative spatial-based methods include interpolation based on various mathematical methods (e.g., bilinear, nearest neighbor, and cubic convolution), globe fitting, and Within-Class Globe Fitting (WCGF). Without spatial heterogeneity, spatial-based methods may only be effective when restoring cloudy pixels in a very small area, and their accuracy is unsatisfactory.

Owing to the poor applicability and uncertain accuracy of spatial-based methods, several spectral-based methods have been proposed. Wang et al. were able to retrieve the Aqua MODIS band 6 data using the relationship between MODIS bands 6 and 7, which makes skillful use of the similarity between Aqua MODIS and Terra MODIS [2]. Subsequently, many methods have been proposed on the basis of Wang and coworkers' work, e.g., Histogram Matching and Local Least Squares Fitting (HMLLSF) and Within-Class Local Fitting (WCLF) [3,4]. Furthermore, to optimize the use of multispectral information, Gladcova et al. proposed a new quantitative image-restoration method that considers the correlation among MODIS bands 1–7 [5]. Li et al. retrieved Aqua MODIS band 6 data by fitting the relationship between Aqua MODIS band 6 and the other six bands using robust m-estimator multi-regression, which effectively optimizes the driving data samples [6]. Shen et al. restored Aqua MODIS band 6 data using adaptive spectrum-weighted sparse Bayesian dictionary learning with the other six bands [7]. These methods made the best of other available spectral information and observations at neighboring locations around the target pixels in a corrupted spectral band. However, all spectral channels within the shortwave spectrum will be polluted by clouds when it is cloudy, which means there will not be sufficient complementary information to reconstruct a cloudy pixel; therefore, these methods are not applicable to cloudy pixels.

To retrieve bad pixels with clouds, a temporal-based method has been proposed that uses consecutive observations of the same spectral band at the same location to reconstruct bad pixels when complementary spectral information is lacking. Lin et al. proposed replacing the missing image using optimal information from the same region in a referenced image when the time interval between the missing image and the referenced image is short enough [1]. The method appeared to demonstrate better results than a method proposed by the joint U.S. Geological Survey (USGS) Landsat team to reconstruct data gaps in SLC-off images (i.e., images taken when the Scan Line Corrector (SLC) is not available) [8]; however, the result of this method depends greatly on the selected sampling, and therefore its accuracy is uncertain. In addition, Li et al. proposed multitemporal dictionary learning algorithms and expanded on their K-SVD and Bayesian counterparts to reconstruct bad pixels [9]. These methods reconstruct remote sensing data based on mathematical methods, which may result in the incredible accuracy.

The aforementioned methods deal with bad pixels using computer graphics/image processing methods; however, they have several disadvantages from the viewpoint of quantitative remote sensing: (1) since these methods are based on mathematical and empirical methods, a physical infrastructure is lacking; (2) because of the lack of a physical infrastructure, the applicable range of the methods is very narrow; and (3) quantitative remote sensing evaluations are lacking for these methods, leading to uncertainty in the accuracy of the methods. These disadvantages are the reasons why these methods have not yet been widely used to retrieve spectral reflectance for sensor's band of cloudy pixels in quantitative remote sensing research. Regrettably, only a few methods (e.g., bilinear and histogram matching) are widely used for resampling and interpolation in quantitative remote sensing.

Therefore, determining a method to retrieve the spectral surface reflectance for the sensor's band of cloudy pixels with sufficient accuracy and quality to be used to retrieve high-level land surface parameters is of urgent priority. However, few works reconstruct the spectral reflectance for sensor's band of cloudy pixels. In reality, quantitative remote sensing research involves multiple theories and technology which can deal with cloudy pixels with higher accuracy than that of the methods in the field of image processing. First, the Bidirectional Reflectance Distribution Function (BRDF) can describe the land-surface reflective anisotropy very accurately [10,11]. In addition, the land cover is usually stable over short periods (e.g., a week or a month). If the land surface is stable, its BRDF is relatively unchanged in its spatial distribution in the reflective direction and shifts only in magnitude

within a short period [12]. Based on this, Shuai et al. fused Landsat 5 reflectance at 30-m resolution and MODIS BRDF at 500-m resolution to retrieve Landsat albedo at 30-m resolution in a snow-free area [13]. Franch et al. derived the Landsat surface albedo and daily MODIS land surface albedo based on Vermote–Justice–Bréon (VJB) method [14,15]. In addition, some early companion studies made use of the fusion of multi-source data to improve the spatio-temporal resolution of the land surface albedo [16,17]. Based on these studies, the reflectance data can be used to derive the albedo data using BRDF parameters, and in theory, the spectral reflectance for sensor’s band of cloudy pixels can be retrieved by fitting a relative of the spectral reflectance for sensor’s band and the corresponding BRDF from the cloud-free and neighboring pixels. The method proposed in this study retrieves spectral reflectance for MODIS band based on a multitude of driving data that have a spatio-temporal mismatch with respect to the target data; however, such driving data do not exist to derive the albedo. Therefore, a scheme to decrease the error from the spatial mismatches between the driving data and target data is required. The proposed method introduces Tobler’s First Law (TFL), which defines all attribute values on a geographic surface as being related to each other, with closer values being more strongly related than farther values. All geographical studies are confronted with the problem of spatial heterogeneity, which has been identified in many studies [18]. Subsequently, numerous local or location-based spatial analysis methods have been developed to address this problem over the past two decades, including the geographically weighted regression (GWR) method [19]. Theoretically, GWR can solve the problem of spatial heterogeneity while retrieving land surface parameters. In addition, several studies have used GWR to address spatial heterogeneity while retrieving precipitation data [20,21]. Accordingly, this study introduces GWR to address errors arising from spatial heterogeneity while fitting spatial data.

With the aim of retrieving spectral reflectance for MODIS the band of cloudy pixels with a higher accuracy, this study made use of multi-temporal data extracted from MODIS observations (e.g., spectral reflectance for MODIS band and solar-viewing geometry) and the MODIS BRDF (MCD43A1 V005) to retrieve MODIS surface reflectance (MOD09GA) of cloudy pixels based on GWR. Compared with other existing methods, the proposed method offers several improvements, i.e., the method: (1) applies an algorithm that retrieves land surface albedo to retrieve the spectral reflectance for MODIS band of cloudy pixels, which improves the accuracy and applicability of the method compared with existing methods; (2) constructs sample reflectance pairs from the observed data and the data calculated with the BRDF to eliminate errors from the anisotropic bidirectional reflectance distribution of a land surface; (3) introduces GWR and optimizes the samples of the driving data to eliminate the error arising from spatial heterogeneity; and (4) includes a more comprehensive evaluation scheme with the criterion of quantitative remote sensing to ensure that the results have sufficient precision. An experiment was conducted in the Heihe River Basin to validate the method. Section 2 introduces the theoretical basis and the details of the procedure to derive the Terra MODIS data. Section 3 introduces the study area and the data. In Section 4, the results of applying the proposed method to the study area are described. Section 5 proposes a comprehensive evaluation scheme and discusses the accuracy of the proposed method by comparing it with other methods (e.g., WCGF, Lin, and VJB). Section 6 discusses the limitations of the proposed method and future improvements to enhance it.

2. Theoretical Basis and Method

2.1. BRDF

The anisotropic reflectance of the land surface can be described by a kernel-driven BRDF model of the official algorithm of MCD43A1, as introduced in in previous studies [10,22,23]:

$$BRDF(\theta_i, \theta_v, \varphi_v, \varphi_i, t) = f_{iso} + f_{vol}k_{vol}(\theta_i, \theta_v, \varphi_v, \varphi_i) + f_{geo}k_{geo}(\theta_i, \theta_v, \varphi_v, \varphi_i) \quad (1)$$

where f_{vol} is the coefficient of the scattering kernel k_{vol} for the solar-viewing geometry $(\theta_i, \theta_v, \varphi_v, \varphi_i)$, f_{geo} is the coefficient of the geometrical optical kernel k_{geo} , and f_{iso} is the coefficient of isotropy.

Considering a stable land cover in a short phenological period (e.g., one month), a BRDF can describe the land surface reflectance for any solar-viewing geometry in a sufficiently short period during which the land cover is stable and

$$\rho(\theta_i, \theta_v, \phi_v, \phi_i, \lambda) = R(\theta_i, \theta_v, \phi_v, \phi_i, \lambda, t) \quad (2)$$

where ρ is the land surface reflectance observed by the sensor, R is the land surface reflectance calculated by the corresponding BRDF at the solar-viewing geometry $(\theta_i, \theta_v, \phi_v, \phi_i)$ of ρ , λ is the spectral band, and t indicates that the BRDF is derived from the land surface reflectance within a period t (e.g., MODIS's period is 8 days).

2.2. Geographically Weighted Regression (GWR)

According to TFL, closer objects are more closely related than farther objects. A local statistical approach known as GWR describes TFL based on mathematical methods very well and includes the spatial coordinates of the variables in the analysis, which may result in a more appropriate basis to investigate their spatially varying relationships. Therefore, GWR is used to explore the linear relationship between the spectral reflectance data for a sensor's band. According to a simple linear model

$$y_i(x_1, \dots, x_n) = a_0 + \sum_{l=1, \dots, n} a_l x_{il} + \sigma_i \quad (3)$$

where y is the dependent variable of the model, x is the independent variable of the model, i is the number of samples, a represents the coefficients of the model, l is the number of coefficients, and σ is the error term.

Combined with TFL, the distance between the target pixel and the sample pixel needs to be considered when fitting the model parameters. Equations (4) and (5) show the fitting parameters with the geographical weights, W_G ,

$$\hat{a} = (X^t W_G X)^{-1} X^t W_G Y \quad (4)$$

$$W_G(i, j) = \exp(-(d_{i,j}/b)^2) \quad (5)$$

where (i, j) indicates the coordinates of the sample, $d_{i,j}$ is the distance between the sample pixel and the bad pixel that needs to be retrieved, and b is the bandwidth.

2.3. Method

Based on Sections 2.1 and 2.2, Terra MODIS surface reflectance (MOD09GA) of cloudy pixels in the Heihe River Basin was retrieved. First, we prepared the data, the MODIS BRDF derived from the kernel-driven model based on good-quality MODIS surface reflectance with no clouds and the MODIS cloud state data corresponding to the MODIS surface reflectance. Second, as a sample, we chose for each target cloudy pixel that needed to be retrieved on day j , the corresponding spectral reflectance for MODIS band observed by the satellite on consecutive days (e.g., $j - 1$, j , and/or $j + 1$) without clouds with the same land cover as that of the target pixel. Third, we calculated the spectral reflectance for MODIS band corresponding to the sample based on the corresponding BRDF data (MCD43A1) for the solar-viewing geometry of the sample, and we used each pair of samples (e.g., the chosen observed spectral reflectance for MODIS band and those calculated with the BRDF) to construct a linear coefficient. Fourth, we fitted the coefficients using GWR based on all the samples, which effectively eliminated the error arising from the spatial heterogeneity of the samples. The first step can be ignored because the MODIS product includes these data (e.g., the MCD43A1 of the MODIS BRDF and the 1-km State Quality Assessment Descriptions of the MODIS surface reflectance). And, the process of the proposed method is shown in Figure 1.

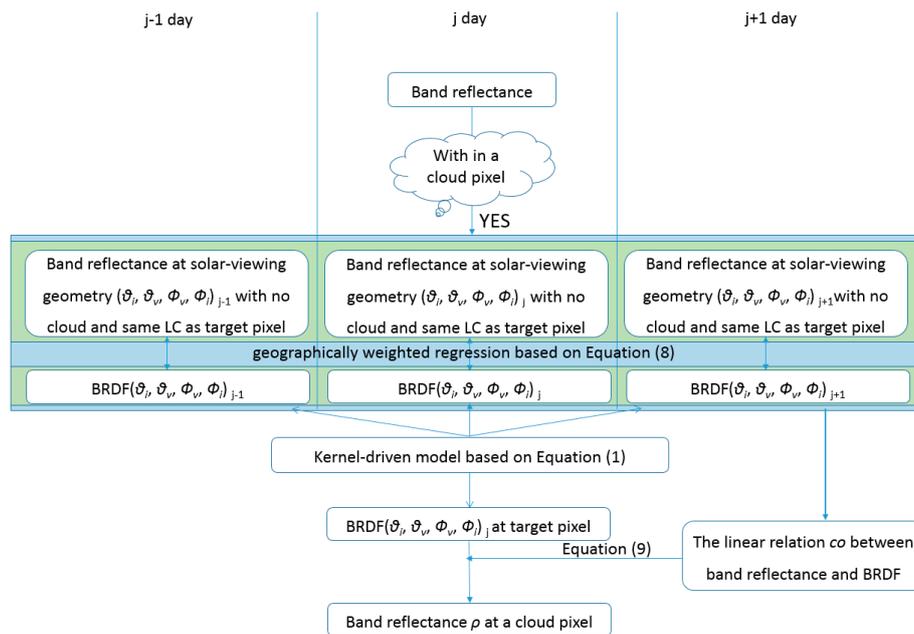


Figure 1. A flowchart of the proposed method.

2.3.1. Constructing a Linear Relationship between the Samples of Spectral Reflectance for MODIS Band and the Corresponding BRDF

Based on the previous section, the land surface reflectance should be equal to the corresponding BRDF; however, in reality, this relationship does not hold for MODIS BRDF data and MODIS surface reflectance. The MODIS BRDF was derived from the MODIS surface reflectance with no clouds on the basis of a kernel-driven model over a 16-day period. Therefore, both the observed reflectance and the reflectance calculated from BRDF cannot correspond completely in the time dimension; instead, MODIS BRDF describes the average status of the land surface reflectance within the no-cloud pixels over a 16-day period in which the land use type is unchanged. Unchanged land use means that the BRDF shape of the land surface is unchanged, and the BRDF for a single location shows only a shift in its magnitude over the 16-day period. Therefore, a linear relation, co , is constructed based on Equation (6) [17]:

$$\rho_{obs}(\theta_i, \theta_v, \phi_v, \phi_i, \lambda_{nb}) = co \times R_{BRDF}(\theta_i, \theta_v, \phi_v, \phi_i, \lambda_{nb}, t) \tag{6}$$

where ρ_{obs} is the land surface reflectance observed by MODIS, R_{BRDF} is the land surface reflectance calculated by the MODIS BRDF for the solar-viewing geometry of the corresponding reflectance observed by MODIS ρ_{obs} , λ_{nb} is the MODIS band, and t indicates that the MODIS BRDF is derived from the MODIS land surface reflectance within a period t (16 days).

When BRDF is known, co becomes key to retrieving the land surface reflectance, and the land surface reflectance at the cloudy pixels can be retrieved based on Equation (7). Land surface reflectance with no clouds and the same land cover as the cloudy pixel, which needs to be retrieved, can be used to derive co ; however, how to choose the samples is key to the retrieval precision; therefore, a multitude of sample pixels are chosen. Then, a linear relation is constructed for each sample pixel based on Equation (7):

$$\rho_{(x,y,T)}(\theta_i, \theta_v, \phi_v, \phi_i, \lambda, t) = co \times R_{(x,y,T)}(\theta_i, \theta_v, \phi_v, \phi_i, \lambda, t) \tag{7}$$

where (x, y, T) belongs to a set of pixels with no clouds and with the same land cover as that of the target pixel that needs to be retrieved, T indicates that ρ and R are close in the time dimension and are within the same period t (16 days), $\rho_{(x,y,T)}$ is ρ_{obs} with no clouds and the same land cover as that of the target pixel, and $R_{(x,y,T)}$ is R_{BRDF} with no clouds and the same land cover as the target pixel.

2.3.2. Retrieving the Spectral Reflectance for MODIS Band of Cloudy Pixels Based on GWR

Based on Equation (7), co can be calculated while spectral reflectance for MODIS band with the solar-viewing geometry and the corresponding BRDF are known. However, there are multiple spatio-temporal samples including band surface reflectances and corresponding BRDFs in (x, y, T) , and the distances between each sample pixel and the target pixel are different, which introduces an error due to spatial heterogeneity. Therefore, co can be fitted using multiple spatio-temporal samples with GWR in order to eliminate this error allowing co to be calculated according to Equation (4):

$$co = (R_{(x,y,T)}^t W_G R_{(x,y,T)})^{-1} R_{(x,y,T)}^t W_G \rho_{(x,y,T)} \quad (8)$$

where (x, y, T) indicates the pixels with no clouds and the same land cover as that of the target pixel in the continuous scene observed by MODIS on days $j - 1$, j , and $j + 1$, and W_G is the geographical weight between the target pixel and the sample pixel referencing to Equation (5).

Then, co can be acquired, and the spectral reflectance for MODIS band of the cloudy pixels can be retrieved by multiplying the corresponding BRDF with co :

$$\rho_{tar}(\theta_i, \theta_v, \phi_v, \phi_i, \lambda) = co \times R_{tar}(\theta_i, \theta_v, \phi_v, \phi_i, \lambda) \quad (9)$$

where the subscript tar indicates the target pixels and ρ_{tar} is the retrieved spectral reflectance of the cloudy pixel.

3. Study Area and Data

The Heihe River Basin is an inland river basin in China and has great elevation difference and temperature difference between day and night. It contains various types of ecosystems and land cover, making it an ideal natural watershed for comprehensive research in hydrology, phenology, and climate (Figure 2). Hence, this region was selected as the study area because numerous scientific experiments have been conducted in the region, providing sufficient remote sensing data and ground measurements for our study [24,25]. Moreover, our team has also conducted considerable research and we have a deeper understanding of this region.

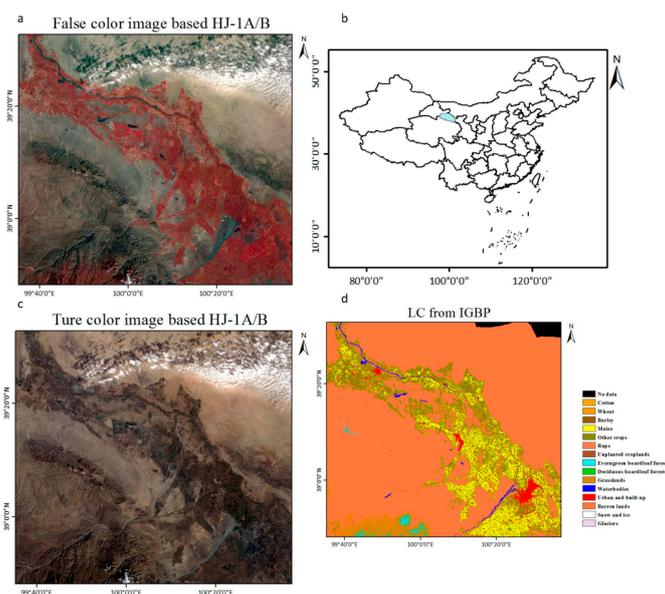


Figure 2. The Heihe River Basin study area is situated in Northwest China: (a) a false color image from 25 September 2012; (b) a map of China highlighting the location of the river basin; (c) a true color image from 25 September 2012; and (d) the land cover of the region in 2012.

In addition, two MODIS datasets were used in the experiment. One is the MODIS land surface reflectance (MOD09GA), which has a 1-day temporal resolution and a 500-m spatial resolution; the other is the BRDF of MODIS (MCD43A1), which has an 8-day temporal resolution and a 500-m spatial resolution and is used to construct the linear relationship between the observed reflectance and that the reflectance calculated with the BRDF.

Based on the International Geosphere-Biosphere Program (IGBP) [26,27], HJ-1A/B [28] land cover data are divided into 15 classes (i.e., cotton, wheat, barley, maize, other crops, rape, unplanted crops, evergreen broadleaf forests, deciduous broadleaf forests, grasslands, waterbodies, urban/built-up, barren lands, snow/ice, and glaciers), and neglected land cover types are not present in the Heihe River Basin. For spatial matching with MODIS data, HJ-1A/B land cover data are upscaled from 30 m to 500 m depending on the main class at 500-m MODIS pixel, and the HJ-1A/B land cover data can indicate the distinguishing land covers of the MODIS pixels very well.

4. Results

Based on the proposed method, an experiment was conducted in the Heihe River Basin. The proposed method was used for the cloudy pixels in the TERRA MODIS MOD09GA data. Moreover, we retrieved the Terra MODIS band 7 reflectance with clouds on the Heihe River Basin for 14–25 September 2012 (Figure 3).

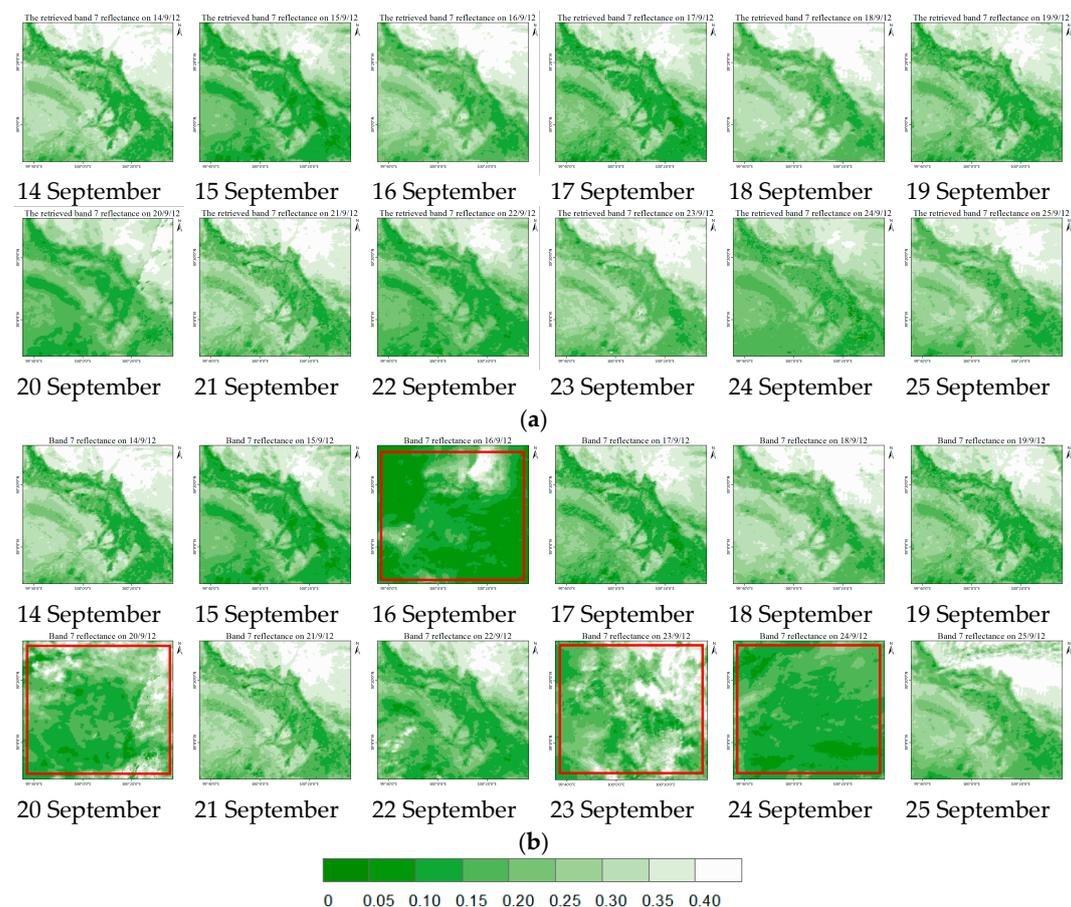


Figure 3. The Band 7 (a) retrieved reflectance and (b) observed reflectance for 12–25 September 2012.

The retrieved band 7 reflectances show similar spatial distributions from 14 September to 25 September 2012, which means that the dataset has good quality and continuity through time, and the retrieved reflectance identifies the variation in the spatial distribution of the land surface

according to the seasonal variations in the phenology. By contrast, the band 7 reflectance observed by MODIS displayed different spatial distributions owing to clouds (e.g., the data on 16, 20, 22, 23, 24, and 25 September 2012), which means that the effective observation is discontinuous. In particular, the observations on 16, 20, 23, and 24 September were completely covered with clouds; however, the corresponding retrieved data show a reasonable spatial and temporal distribution, as shown in Figure 4 (the retrieved band 1–7 reflectance on 23 September 2012 was randomly selected as an example).

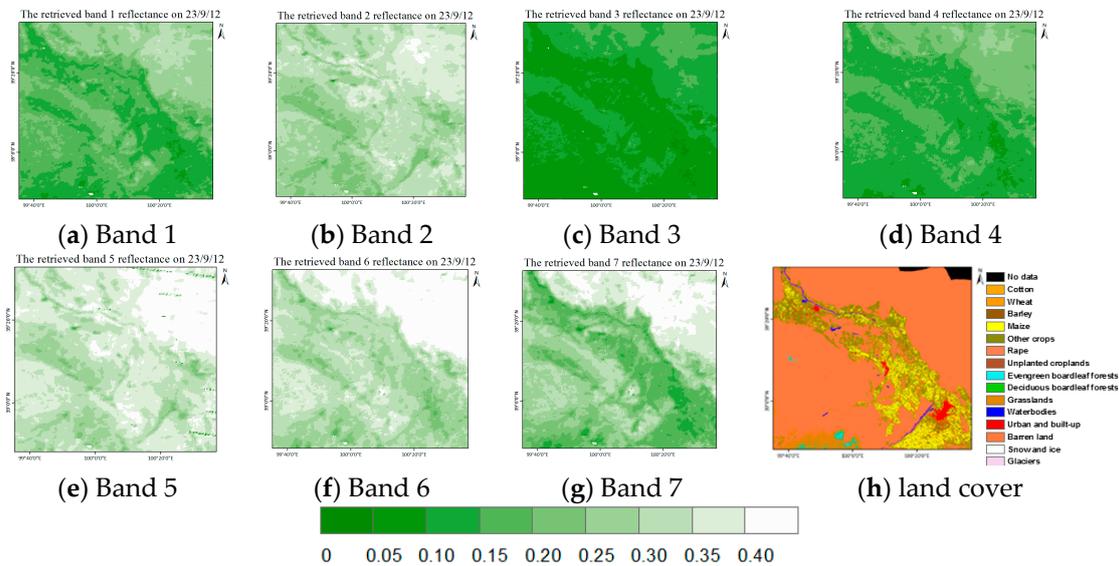
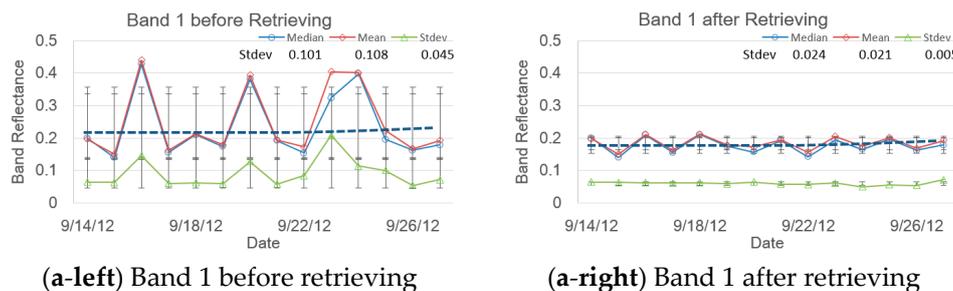


Figure 4. (a–g) The spatial distribution of the retrieved spectral reflectance for MODIS band 1–7 on 23 September 2012; and (h) the corresponding land cover data.

Based on Figure 3, the band 7 observation on 23 September 2012 was masked by clouds completely; therefore, the band 1–6 reflectance should have also been completely masked due to the optical reflection characteristics of clouds. Figure 4 shows that the retrieved spectral reflectance for MODIS band (Figure 3a–g) had a reasonable spatial distribution compared with the land cover in Figure 3h. Therefore, the retrieved reflectance appears to be valid and close to the true value.

To further analyze the improved results of the proposed method, two time series from 14 September to 27 September 2012 (the observed and the retrieved values) were constructed and compared. Based on Figure 3, the spectral reflectance for MODIS band included clouds at different levels on all days, resulting in a dramatic fluctuation in the statistical indices (e.g., the median, mean, and standard deviation) through time. However, the reflectance values of bands 1, 2, 3, 4, 6, and 7 are expected to be stable in September, judging from the phenological rule. Therefore, a trend of slow growth will appear as the chlorophyll in the vegetation as well as the soil moisture decrease, as shown by the blue trendline in Figure 5.



(a-left) Band 1 before retrieving

(a-right) Band 1 after retrieving

Figure 5. Cont.

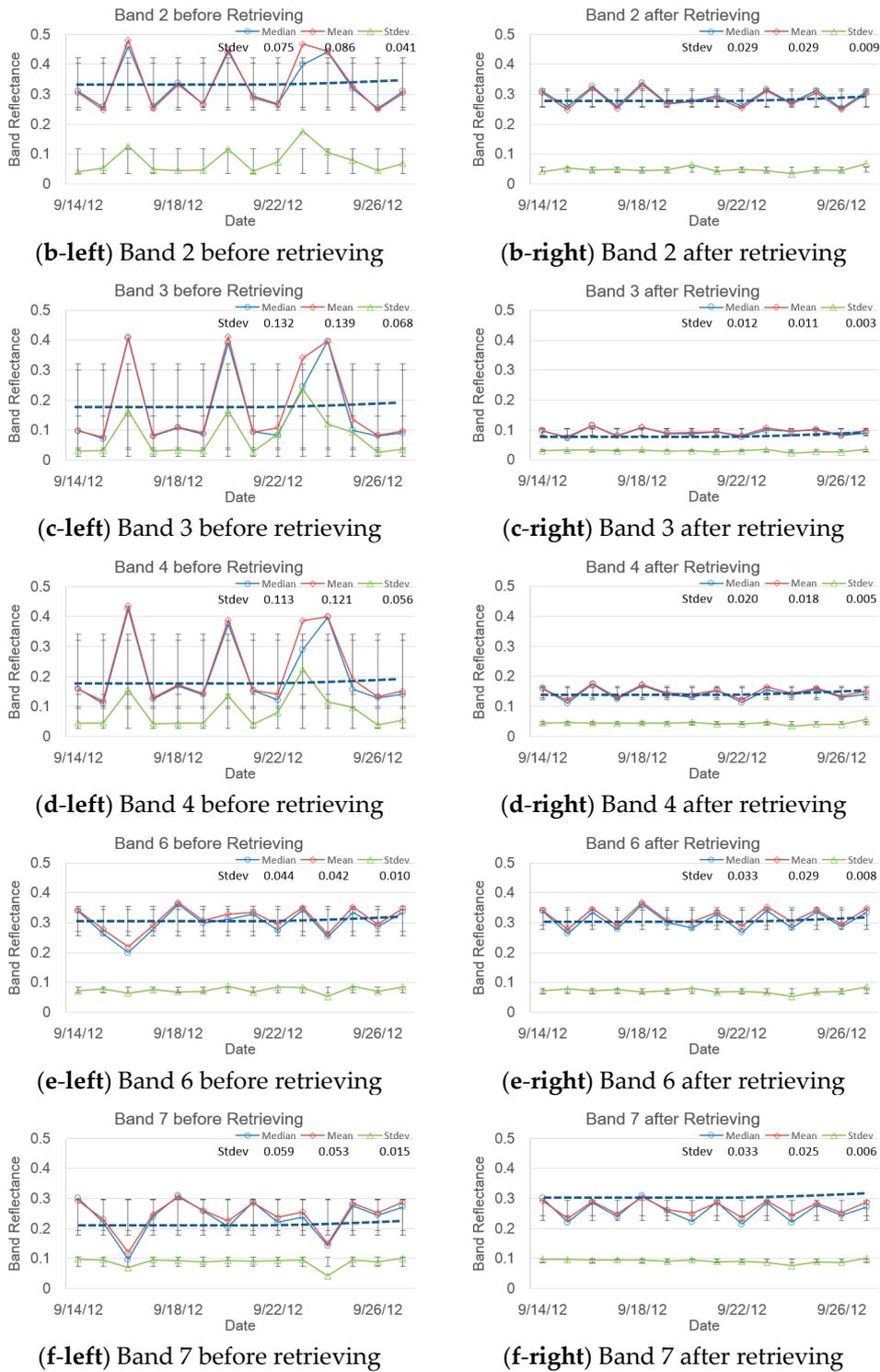


Figure 5. Time series of the median, mean, and standard deviation of the observed spectral reflectance for MODIS bands and the retrieved spectral reflectance for MODIS bands from 14 September to 27 September 2012: **(a–f: left)** the observed spectral reflectance for MODIS bands 1, 2, 3, 4, 6 and 7, **(a–f: right)** the retrieved spectral reflectance for MODIS bands 1, 2, 3, 4, 6 and 7.

In Figure 5, the time series of the statistical indexes of the retrieved data are smoother than those of the observed data and closer to the blue trendline. This is because the observed data are influenced by the clouds, causing singularities in, and greater volatility of, the observations. This shows that the

proposed method is very effective for retrieving the reflectance of cloudy pixels. The time series of the statistical indexes of the retrieved reflectance (e.g., the median, mean, and standard deviation) still show slight fluctuations, possibly because the land surface BRDF varies with varying solar-viewing geometry and the solar-viewing geometry of the spectral reflectance for MODIS band is different on each day [29]. However, all the retrieved spectral reflectance for MODIS band had a reasonable spatial distribution that was the same as that of the land cover data with a slightly fluctuating and smoothly increasing trend, which is appropriate and in accordance with the phenological rule.

5. Validation and Discussion

From Section 4, qualitative analysis shows that the results of the proposed method are rational. However, no improvement over the results of the other methods is shown. Therefore, in Section 5, a complete validating scheme for determining the accuracy and improvement of the proposed method is discussed.

5.1. Validation and the Robustness of the Proposed Method

Although the effectiveness of the proposed method has been demonstrated, a quantitative validation is still necessary to show the extent to which the proposed method improves the results, because it is important that the accuracy of the proposed method should be able to satisfy the precision requirements for quantitative remote sensing.

To validate the precision of the proposed method and to check its robustness, a scheme comprising two parts was drawn: First, acquiring the validation data. For this purpose, three simulated cloudy areas were constructed in the study area (Figure 6) with different ranges (Simulation a (S-a): 15 km × 15 km; Simulation b (S-b): 30 km × 30 km; and Simulation c (S-c): 45 km × 45 km); furthermore, the reflectance at the no-cloud pixels within the simulating cloudy area was selected as validation data on 19 September 2012. Then, four methods, i.e., WCGF as the classic spatial-based method, Lin's method as the classic temporal-based method, VJB method as the semi-empirical method, and the proposed method, were used on the three simulated cloudy areas separately based on the MCD09GA data on 19 September 2012.

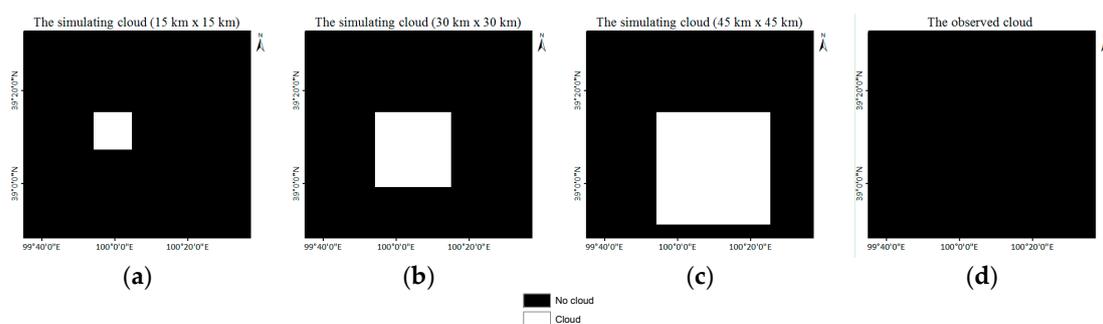


Figure 6. The simulated cloud areas: (a) 15 km × 15 km; (b) 30 km × 30 km; and (c) 45 km × 45 km; and (d) the observed cloud product from MODIS.

Figure 7 shows the results from the four methods based on the simulated cloud areas presented in Figure 6b in order to validate the precision of the proposed method. The results from WCGF show the worst spatial distribution among the four methods, while the other three methods had a similar spatial distribution. The red square encompasses both the simulated cloud region and the no-cloud region, and unreasonable spatial variations in the magnitude appear along the boundary of the simulated cloud region and the no-cloud region when using WCGF. Meanwhile, the spatial distribution of the result obtained from the proposed method and VJB method appears to be more similar to the MODIS reflectance than the result obtained from Lin's method (the red square area in Figure 7). Statistics on the precision of the four methods are shown in Figure 8.

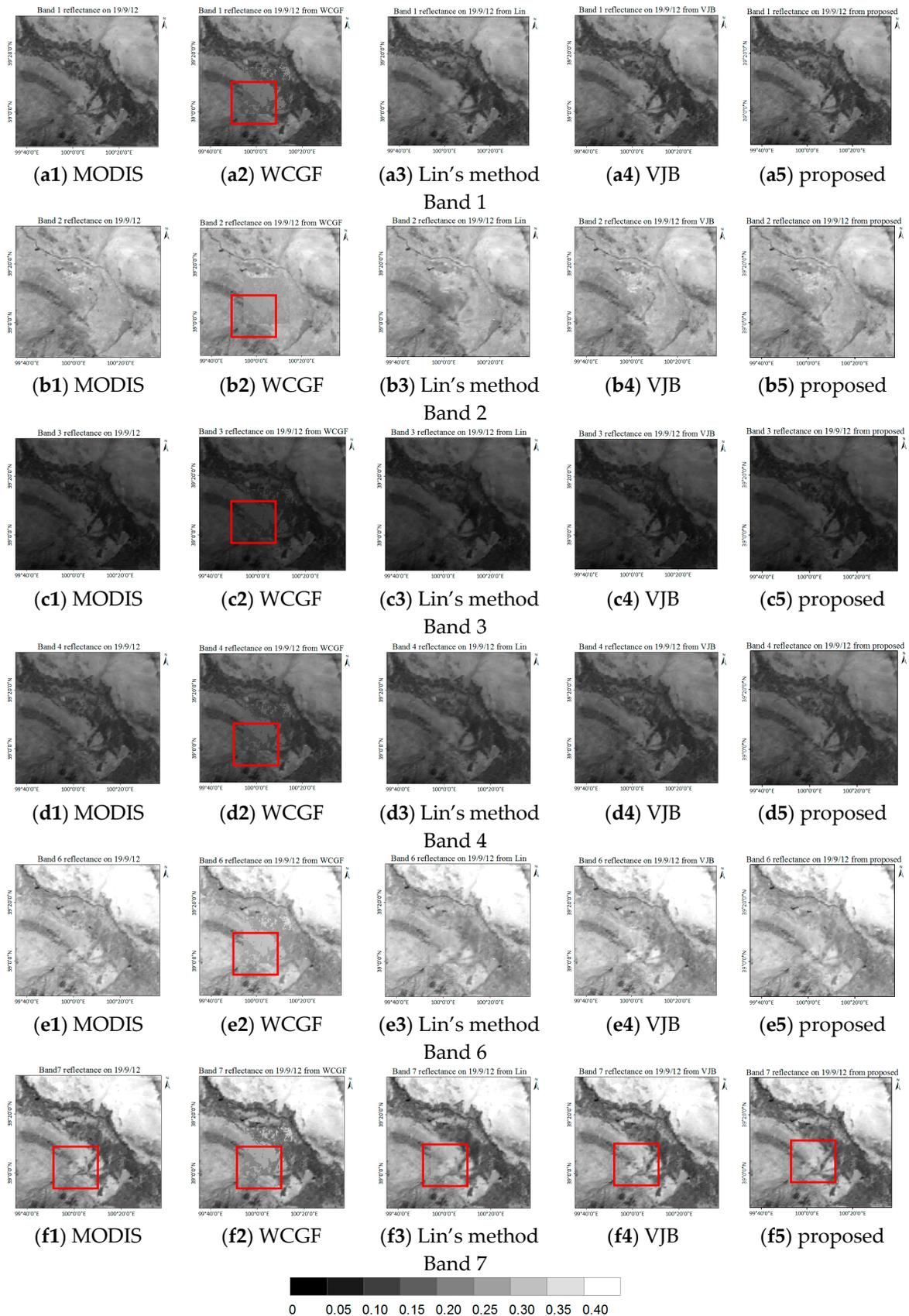


Figure 7. The spectral reflectance for MODIS bands 1, 2, 3, 4, 6, and 7 (a1–f1); and the retrieved spectral reflectance for MODIS bands 1, 2, 3, 4, 6, and 7 for: (a2–f2) WCGF; (a3–f3) Lin's method; (a4–f4) VJB; and (a5–f5) the proposed method in the simulated cloud area from Figure 6b.

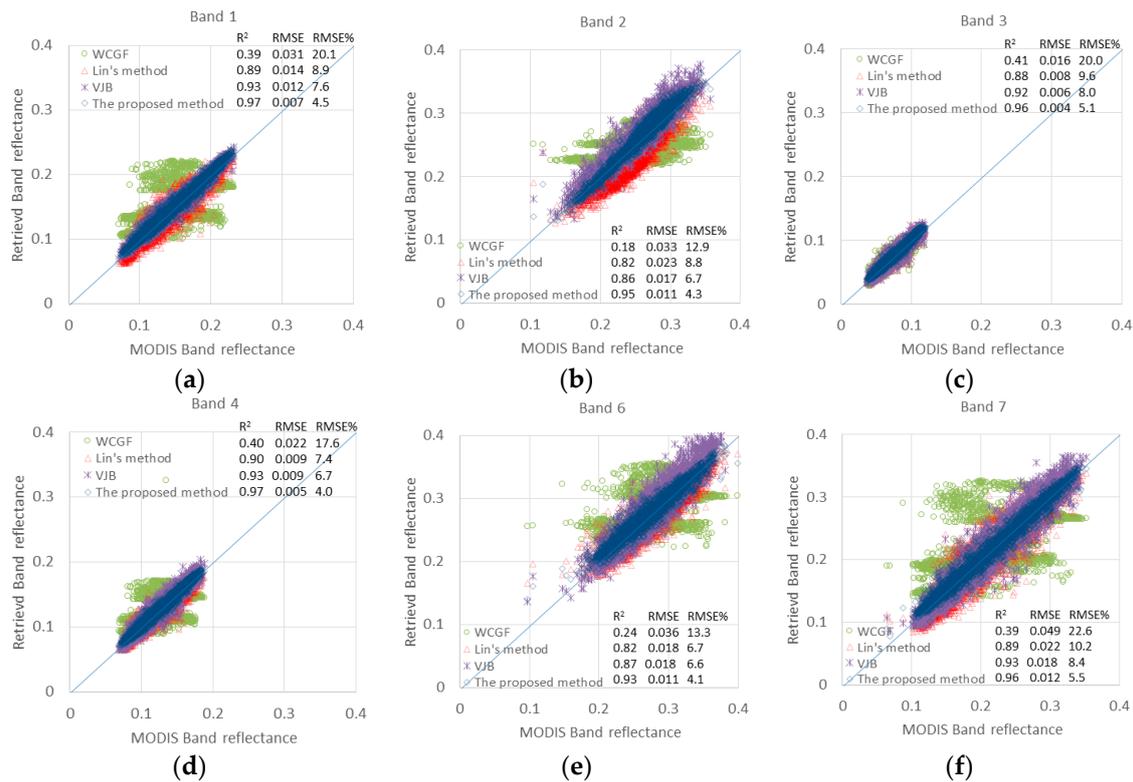


Figure 8. The 2D scattering plots show the precision statistics between the spectral reflectance for MODIS bands and the retrieved spectral reflectance for MODIS bands using the four methods in the simulated cloud area from Figure 6b: (a–f) MODIS bands 1, 2, 3, 4, 6, and 7 (Band 5 is not used because of bad stripe).

Based on the statistics from the 2D scattering plots in Figure 8, the R² and RMSE values show that the proposed method is the most accurate method among all four methods; the VJB method is second best after the proposed method; Lin's method shows intermediate performance; and WCGF performs the worst.

This conclusion also provides an analysis of the advantages of the proposed method compared with other methods. WCGF does not consider the spatial distribution in the cloudy area and only replaces the value of the cloudy pixel by averaging the value of the neighboring pixels without clouds but with the same land cover as that of the cloudy pixel; this leads to an unreasonable variation in the magnitude and spatial distribution. Compared to WCGF, Lin's method accounts for the spatial distribution in the cloudy area by constructing a gradient field based on the optimal multi-temporal data without clouds at the same location as the cloudy pixel and smooths the variation in the magnitude along the boundary of the simulated cloud area and the no-cloud area. This leads to a result that is more reasonable and accurate than WCGF, as shown in Figures 7 and 8. However, a critical disadvantage of Lin's method is that the BRDF of the land surface is ignored, and the BRDF leads to a variation in the spatial distribution of MODIS data observed at different times due to varying solar-viewing geometries. Therefore, we can see that the spatial distribution of the results from Lin's method shows less similarity with the distribution of MODIS data than the proposed method in the red square area in Figures 7 and 8. The disadvantages of the two previous methods are considered and addressed by the proposed method, thereby increasing the accuracy of the proposed method.

Meanwhile, the VJB method can also be used for directly retrieving the spectral reflectance of the cloudy pixels in MODIS bands 1, 2, 3, 4, 6, and 7 based on the fact that the empirical database of BRDF is driven by the same MODIS reflectance that is used in the proposed method. Furthermore, the results produced by the VJB method are less robust than those produced by the proposed method; however,

we still believe that VJB method's theoretical basis is reasonable because it is similar to that of the proposed method which assumes that the variation of reflectance observed by direct effect of BRDF and the BRDF's imperfection for describing the anisotropic reflectance of land surface. VJB assumes that the variation of reflectance observed by the sensor is caused by daily direct effects in a short time. The proposed method adds to the VJB method a local fitting scheme that uses GWR to take spatial heterogeneity into account. In addition, the proposed method's specialized design (e.g., the use of a local fitting scheme based on GWR) can ensure higher accuracy than the VJB method. The results produced by the VJB method show two conclusions: (1) the VJB method is better than both WCGF and Lin's method, which provides a strong theoretical basis for VJB having a wider range of application and better accuracy than methods based on pure mathematics; and (2) the VJB method is not as good as the proposed method, which means that VJB needs to be adjusted and redesigned for retrieving the spectral reflectance of cloudy pixels in the MODIS bands, an adjustment that the proposed method has already made (through, e.g., introducing spatiotemporal samples, designing the scheme of local fitting, adopting GWR).

However, this comparison does not ensure that the accuracy of the proposed method meets the requirements for retrieving high-level land surface parameters with high accuracy because the robustness of the method still needs to be checked. Therefore, the retrievals based on the simulated cloudy areas in Figure 6 (S-a: 15 km × 15 km and S-c: 45 km × 45 km) were compared with the results in Figure 7 in order to check the precision of the methods on a different cloudy area (Table 1).

Table 1. Statistics for the retrieved spectral reflectance for MODIS band from the four methods based on the simulating cloud areas in Figure 6a–c (% is the percent of RMSE).

Band	WCGF			Lin			VJB			The Proposed			
	R ²	RMSE	%										
S-a	B1	0.25	0.030	17.5	0.90	0.011	6.7	0.92	0.012	6.9	0.97	0.007	3.9
	B2	0.11	0.032	12.1	0.82	0.021	8.1	0.86	0.019	7.3	0.94	0.012	4.4
	B3	0.27	0.015	17.7	0.89	0.006	7.1	0.92	0.006	6.6	0.96	0.004	4.5
	B4	0.27	0.022	15.6	0.90	0.008	5.7	0.92	0.008	5.9	0.97	0.005	3.4
	B6	0.08	0.031	10.6	0.76	0.014	4.7	0.79	0.017	5.8	0.86	0.011	3.9
	B7	0.27	0.047	19.9	0.92	0.017	7.1	0.94	0.015	6.5	0.97	0.011	4.4
S-b	B1	0.39	0.031	20.1	0.89	0.014	8.9	0.93	0.012	7.6	0.97	0.007	4.5
	B2	0.18	0.033	12.9	0.82	0.023	8.8	0.88	0.017	6.7	0.95	0.011	4.3
	B3	0.41	0.016	20.0	0.88	0.008	9.6	0.92	0.006	8.0	0.97	0.004	5.1
	B4	0.40	0.022	17.6	0.90	0.009	7.4	0.93	0.009	6.7	0.97	0.005	4.0
	B6	0.24	0.036	13.3	0.82	0.018	6.7	0.87	0.018	6.6	0.93	0.011	4.1
	B7	0.39	0.049	22.6	0.89	0.022	10.2	0.93	0.018	8.4	0.96	0.012	5.5
S-c	B1	0.41	0.032	22.3	0.89	0.015	10.2	0.93	0.012	8.6	0.97	0.007	4.9
	B2	0.00	0.033	13.1	0.70	0.020	8.0	0.84	0.018	7.0	0.91	0.010	4.1
	B3	0.41	0.016	22.1	0.88	0.008	11.5	0.93	0.007	9.5	0.96	0.004	5.6
	B4	0.40	0.023	19.5	0.91	0.010	8.5	0.94	0.009	7.8	0.97	0.005	4.5
	B6	0.31	0.039	14.7	0.87	0.020	7.4	0.91	0.019	7.1	0.95	0.011	4.1
	B7	0.46	0.050	23.8	0.91	0.025	12.0	0.94	0.019	9.0	0.97	0.012	5.8

To analyze and compare the results more intuitively, the values in the columns of Table 1 were plotted in Figure 9. The proposed method shows the best accuracy in terms of all statistical indices for different simulated cloudy areas and bands. In addition, the retrieval precision at different MODIS bands from the proposed method shows a better stability than the other three methods, which appear to experience a severe drop in their retrieval precision. The blue curves in Figure 9c highlight the trend of RMSE with enlarging simulated cloudy area and shows the minimal variation in the precision of the proposed method.

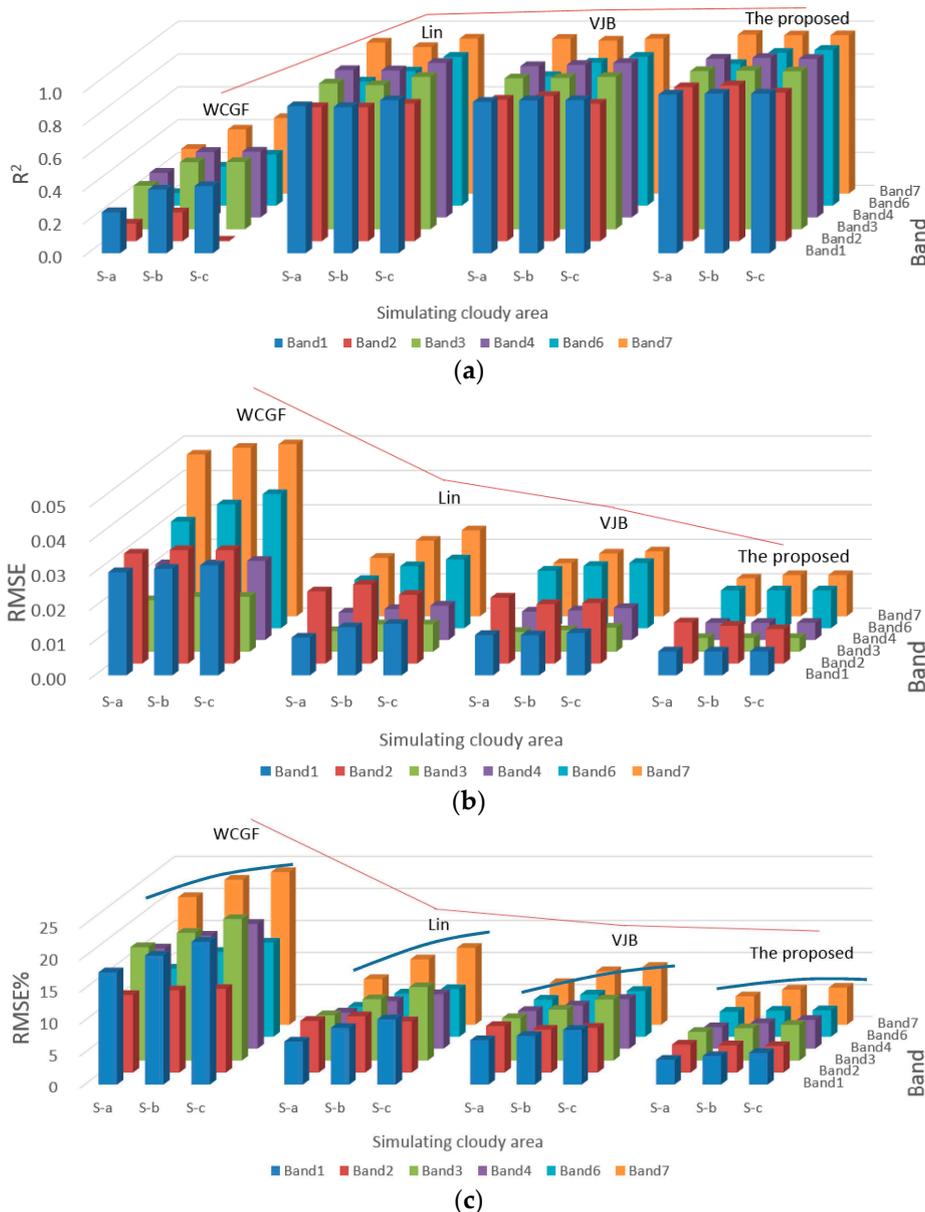


Figure 9. Column plots of the precision statistics based on Table 1: (a) R^2 versus modeling method for bands 1, 2, 3, 4, 6 and 7; (b) RMSE versus modeling method for bands 1, 2, 3, 4, 6 and 7; and (c) RMSE% versus modeling method for bands 1, 2, 3, 4, 6 and 7.

Table 1 and Figure 9 demonstrate that the results from the proposed method has the best and most stable precision of the four methods and that the statistical accuracy (e.g., R^2 , RMSE, and RMSE%) do not significantly diminish with larger simulated cloudy areas, which means that the proposed method is very robust owing to its physical mechanism and strict retrieving control. Furthermore, other methods based on mathematical statistics may not be stable in enlarged cloudy area.

5.2. Retrieving White-Sky Albedo of Cloudy Pixels

Based on the previous section, the proposed method is expected to have a very good accuracy; however, the target of this study is to retrieve spectral reflectance for MODIS band that can be used to derive high-level land surface parameters with high accuracy. Therefore, a test was conducted to retrieve the band white-sky albedo based on the spectral reflectance for MODIS band with the original reflectance and the reflectance retrieved by the proposed method for 12–27 September 2012.

The algorithm to retrieve the band white-sky albedo is given in the Algorithm for MODIS Bidirectional Reflectance Anisotropies of the Land Surface (AMBRALS).

Based on the statistics of the accumulation of cloudy pixels shown in Figure 10, 2D scattering plots of the spectral albedo for the MODIS band retrieved by no-cloud data and the mixed data that include no-cloud and restored cloudy pixels are shown according to the different numbers (e.g., 3, 4, and 5) of cloudy pixels during the retrieval period (Figure 11). Pixels with different numbers, 3–5, account for approximately 80% of the total pixels. From Figure 11, the worst RMSE% is approximately 6% and the best RMSE% is less than 2%.

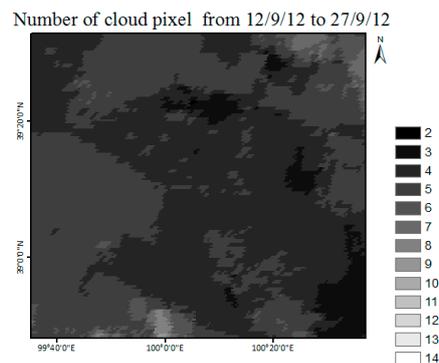


Figure 10. The accumulation of cloudy pixels during the retrieval period (12–27 September 2012).

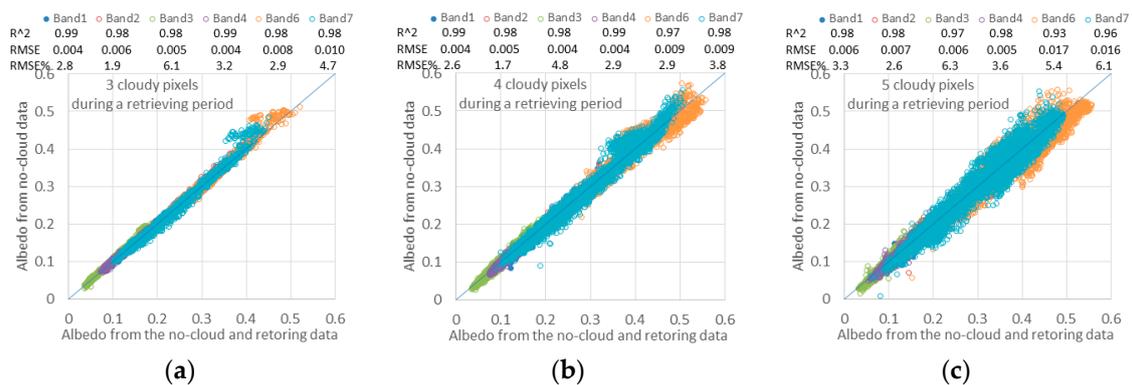


Figure 11. 2D scattering plots of the band white-sky albedo retrieved from no-cloud data and mixed data including no-cloud and restored cloudy pixels according to the different number of the cloudy pixels: (a) 3; (b) 4; and (c) 5.

The result is shown in Figure 12, assuming that the albedo retrieved using the no-cloud data is the true value and converting the spectral albedo for the MODIS band to shortwave albedo. With the restored cloudy data, which were used when the number of cloudy pixels increased from 2 to 5 (accounting for approximately 80% of the total area), the precision of the retrieved albedo shows fluctuations in the RMSE% from 0.9% to 3.9% and R² from 97% to 98%. The total precision of the method is high, with a total RMSE% of 3.1% and a total bias% of 2.0%. Considering the condition of the six restored cloudy data used to retrieve the albedo, the worst precision of the retrieved albedo is with an RMSE% of 7.7% and an R² of 94%; however, the total RMSE% of 3.6% and total bias% of 2.2% are still very accurate and stable, because the conditions of the six restored cloudy data used to retrieve the albedo only account for approximately 6% of the total.

Comparing earlier studies on albedo retrieval, Miguel et al. have shown that most of the retrieved NIR broadband albedos meet the 10% accuracy (bias) requirement; therefore, 10% accuracy is considered the threshold for judging the level of accuracy of a retrieval method [30]. Franch et al. validated the improved VJB method and acquired an overall RMSE of 7% [15], which is better than

an RMSE of 0.024 or 12% acquired by Shuai and coworkers' method [13]. Therefore, the retrieved spectral albedo for MODIS band based on the spectral reflectance for MODIS band retrieved using the proposed method is relatively more accurate. The white-sky albedo retrieved using the restored cloudy MODIS data shows very good precision and completely satisfies all statistical indices; therefore, the proposed method can retrieve the spectral reflectance for MODIS band of cloudy pixels with better precision than other methods (e.g., WCGF, Lin's method, and VJB method) and has sufficiently good accuracy to be used to retrieve high-level land surface parameters (e.g., the albedo).

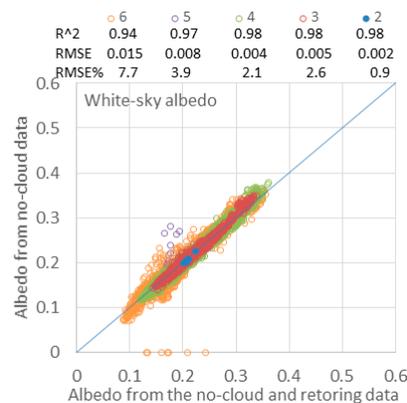


Figure 12. A 2D scattering plot of the land surface albedo (white-sky albedo) retrieved from no-cloud data and mixed data including no-cloud and restored cloudy pixels according to different cloudy pixel numbers (2, 3, 4, 5, and 6).

6. Conclusions

In this study, we not only proposed an algorithm to retrieve the spectral reflectance for MODIS band of cloudy pixels with sufficient precision to be used for quantitative remote sensing retrieval but also applied the improved method (e.g., Shuai [13] and Franch [15]) to solve the new problems related to clouds. The proposed method considers the following key factors when retrieving the reflectance of cloudy pixels: (1) determining the anisotropic reflectance of the land surface; (2) eliminating errors from spatial heterogeneities between the data samples; and (3) evaluating the precision of the proposed method based on a more comprehensive evaluation scheme. To address these three issues, the proposed method considers the anisotropic reflectance of the land surface and introduces the BRDF to calculate the spectral reflectance of each sample based on the solar-viewing geometry, which eliminates the errors from the anisotropic reflectance. The theoretical basis of this method is similar to that of the existing algorithms (e.g., Shuai) in which the sample pixels are matched with the target pixel in space. The proposed method uses a no-cloud pixel as a sample to model the cloudy pixel, which leads to a spatial mismatch between the sample pixels and the target pixel; therefore, GWR was used to optimize the samples and eliminate errors due to spatial heterogeneity. Finally, a comprehensive scheme was conducted to evaluate the results of the proposed method in comparison with other methods, and the retrieved spectral reflectance for MODIS band was used to retrieve the land surface albedo. The results were validated and good accuracy was demonstrated. The proposed method introduces the idea of quantitative remote sensing and geographical theory to address cloudy pixels, and the validation shows that the proposed method has a better accuracy than other methods and can meet the accuracy requirements for quantitative retrievals.

Even though the proposed method improved the accuracy effectively, some deficiencies exist. One deficiency is that the spectral reflectance within the shortwave band is subject to atmospheric conditions (e.g., the aerosol optical depth and water vapor), and the different spectral reflectances are affected by these conditions in varying degree, which will in turn reduce the accuracy. Another deficiency is that the proposed method cannot retrieve spectral reflectance accurately while an event

is happening (e.g., rain and/or snow) because the BRDF data does not have sufficient temporal resolution. Finally, validation should be conducted for other sensors, such as Enhanced Thematic Mapper (ETM) data, to further prove the superior applicability of the proposed method. These issues will be addressed in our future work.

Acknowledgments: This work was partly supported by the National Natural Science Foundation of China (No. 41130744/D0107, No. 41501380/D0106), the China Postdoctoral Science Foundation (No. 2016M591209), National Natural Science Foundation of China (No. 41571364), the Beijing Postdoctoral Research Foundation and the National Key Basic Research Program of China (No. 2015CB953702).

Author Contributions: Bo Gao designed the research, processed the remote sensing data and drafted the manuscript. The other authors provided supervision during manuscript construction and revision and analysis of results throughout the process.

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

References

1. Lin, C.H.; Lai, K.H.; Chen, Z.B.; Chen, J.Y. Patch-based information reconstruction of cloud-contaminated multitemporal images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 163–174. [[CrossRef](#)]
2. Wang, L.L.; Qu, J.J.; Xiong, X.X.; Hao, X.J.; Xie, Y.; Che, N.Z. A new method for retrieving band 6 of Aqua MODIS. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 267–270. [[CrossRef](#)]
3. Rakwatin, P.; Takeuchi, W.; Yasuoka, Y. Restoration of Aqua MODIS band 6 using histogram matching and local least squares fitting. *IEEE Trans. Geosci Remote Sens.* **2009**, *47*, 613–627. [[CrossRef](#)]
4. Shen, H.F.; Zeng, C.; Zhang, L.P. Recovering reflectance of Aqua MODIS band 6 based on within-class local fitting. *IEEE J.-Stars* **2011**, *4*, 185–192. [[CrossRef](#)]
5. Gladkova, I.; Grossberg, M.D.; Shahriar, F.; Bonev, G.; Romanov, P. Quantitative restoration for MODIS band 6 on Aqua. *IEEE Trans. Geosci Remote Sens.* **2012**, *50*, 2409–2416. [[CrossRef](#)]
6. Li, X.H.; Shen, H.F.; Zhang, L.P.; Zhang, H.Y.; Yuan, Q.Q. Dead pixel completion of Aqua MODIS band 6 using a robust m-estimator multiregression. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 768–772.
7. Shen, H.F.; Li, X.H.; Zhang, L.P.; Tao, D.C.; Zeng, C. Compressed sensing-based inpainting of Aqua moderate resolution imaging spectroradiometer band 6 using adaptive spectrum-weighted sparse bayesian dictionary learning. *IEEE Trans. Geosci Remote Sens.* **2014**, *52*, 894–906. [[CrossRef](#)]
8. United States Geological Survey (USGS). *Gap-Fill Algorithm: Slc-off Gap-Filled Products Gap-Fill Algorithm Methodology*; USGS: Reston, VA, USA, 2004.
9. Li, X.H.; Shen, H.F.; Zhang, L.P.; Zhang, H.Y.; Yuan, Q.Q.; Yang, G. Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning. *IEEE Trans. Geosci Remote Sens.* **2014**, *52*, 7086–7098.
10. Wanner, W.; Li, X.; Strahler, A.H. On the derivation of kernels for kernel-driven models of bidirectional reflectance. *J. Geophys. Res. Atmos.* **1995**, *100*, 21077–21089. [[CrossRef](#)]
11. Schaaf, C.B.; Gao, F.; Strahler, A.H.; Lucht, W.; Li, X.W.; Tsang, T.; Strugnell, N.C.; Zhang, X.Y.; Jin, Y.F.; Muller, J.P.; et al. First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sens. Environ.* **2002**, *83*, 135–148. [[CrossRef](#)]
12. Bacour, C.; Breon, F.M. Variability of biome reflectance directional signatures as seen by polder. *Remote Sens. Environ.* **2005**, *98*, 80–95. [[CrossRef](#)]
13. Shuai, Y.M.; Masek, J.G.; Gao, F.; Schaaf, C.B. An algorithm for the retrieval of 30-m snow-free albedo from landsat surface reflectance and MODIS BRDF. *Remote Sens. Environ.* **2011**, *115*, 2204–2216. [[CrossRef](#)]
14. Franch, B.; Vermote, E.F.; Claverie, M. Intercomparison of landsat albedo retrieval techniques and evaluation against in situ measurements across the us surfrad network. *Remote Sens. Environ.* **2014**, *152*, 627–637. [[CrossRef](#)]
15. Franch, B.; Vermote, E.F.; Sobrino, J.A.; Julien, Y. Retrieval of surface albedo on a daily basis: Application to MODIS data. *IEEE Trans. Geosci Remote Sens.* **2014**, *52*, 7549–7558. [[CrossRef](#)]
16. Gao, B.; Jia, L.; Wang, T.X. Derivation of land surface albedo at high resolution by combining HJ-1A/B reflectance observations with MODIS BRDF products. *Remote Sens.* **2014**, *6*, 8966–8985. [[CrossRef](#)]
17. Gao, B.; Gong, H.L.; Wang, T.X. A method for retrieving daily land surface albedo from space at 30-m resolution. *Remote Sens.* **2015**, *7*, 10951–10972. [[CrossRef](#)]

18. Goodchild, M.F. Twenty years of progress: Giscience in 2010. *J. Spat. Inf. Sci.* **2010**, *1*, 3–20. [[CrossRef](#)]
19. Fotheringham, A.S.; Charlton, M.E.; Brunson, C. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environ. Plan. A* **1998**, *30*, 1905–1927. [[CrossRef](#)]
20. Chen, F.R.; Liu, Y.; Liu, Q.; Li, X. Spatial downscaling of TRMM 3B43 precipitation considering spatial heterogeneity. *Int. J. Remote Sens.* **2014**, *35*, 3074–3093. [[CrossRef](#)]
21. Ceccherini, G.; Ameztoy, I.; Hernandez, C.P.R.; Moreno, C.C. High-resolution precipitation datasets in South America and West Africa based on satellite-derived rainfall, enhanced vegetation index and digital elevation model. *Remote Sens.* **2015**, *7*, 6454–6488. [[CrossRef](#)]
22. Li, X.W.; Strahler, A.H. Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy—Effect of crown shape and mutual shadowing. *IEEE Trans. Geosci Remote Sens.* **1992**, *30*, 276–292. [[CrossRef](#)]
23. Roujean, J.L. A bidirectional reflectance model of the earth’s surface for the correction of remote sensing data. *J. Geophys. Res.* **1992**, *97*, 20455–20468. [[CrossRef](#)]
24. Li, X.; Cheng, G.D.; Liu, S.M.; Xiao, Q.; Ma, M.G.; Jin, R.; Che, T.; Liu, Q.H.; Wang, W.Z.; Qi, Y.; et al. Heihe watershed allied telemetry experimental research (Hiwater): Scientific objectives and experimental design. *Bull. Am. Meteorol. Soc.* **2013**, *94*, 1145–1160. [[CrossRef](#)]
25. Li, X.; Li, X.W.; Li, Z.Y.; Ma, M.G.; Wang, J.; Xiao, Q.; Liu, Q.; Che, T.; Chen, E.X.; Yan, G.J.; et al. Watershed allied telemetry experimental research. *J. Geophys. Res. Atmos.* **2009**, *114*, 2191–2196. [[CrossRef](#)]
26. Zhong, B.; Ma, P.; Nie, A.; Yang, A.; Yao, Y.; Lü, W.; Zhang, H.; Liu, Q. Land cover mapping using time series HJ-1/CCD data. *Sci. China Earth Sci.* **2014**, *57*, 1790–1799. [[CrossRef](#)]
27. Loveland, T.; Reed, B.; Brown, J.; Ohlen, D.; Zhu, Z.; Yang, L.; Merchant, J. Development of a global land cover characteristics database and IGBP discover from 1 km AVHRR data. *Int. J. Remote Sens.* **2000**, *21*, 1303–1330. [[CrossRef](#)]
28. Wang, Q.; Wu, C.; Li, Q.; Li, J. Chinese HJ-1A/B satellites and data characteristics. *Sci. China Earth Sci.* **2010**, *53*, 51–57. [[CrossRef](#)]
29. Vermote, E.; Justice, C.O.; Breon, F.M. Towards a generalized approach for correction of the BRDF effect in modis directional reflectances. *IEEE Trans. Geosci Remote Sens.* **2009**, *47*, 898–908. [[CrossRef](#)]
30. Roman, M.O.; Gatebe, C.K.; Shuai, Y.M.; Wang, Z.S.; Gao, F.; Masek, J.G.; He, T.; Liang, S.L.; Schaaf, C.B. Use of in situ and airborne multiangle data to assess MODIS- and Landsat-based estimates of directional reflectance and albedo. *IEEE Trans. Geosci Remote Sens.* **2013**, *51*, 1393–1404. [[CrossRef](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).