

Article

# Seamless Upscaling of the Field-Measured Grassland Aboveground Biomass Based on Gaussian Process Regression and Gap-Filled Landsat 8 OLI Reflectance

Gaofei Yin <sup>1</sup>, Ainong Li <sup>1,\*</sup>, Chaoyang Wu <sup>2</sup>, Jiyan Wang <sup>3</sup>, Qiaoyun Xie <sup>4</sup>, Zhengjian Zhang <sup>1</sup>, Xi Nan <sup>1</sup>, Huaan Jin <sup>1</sup>, Jinhu Bian <sup>1</sup>  and Guangbin Lei <sup>1</sup> 

<sup>1</sup> Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610041, China; gaofeiyin@imde.ac.cn (G.Y.); zhangzj@imde.ac.cn (Z.Z.); nanxi@imde.ac.cn (X.N.); jinhuaan@imde.ac.cn (H.J.); bianjinhu@imde.ac.cn (J.B.); leiguangbin@imde.ac.cn (G.L.)

<sup>2</sup> Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; wucy@igsrr.ac.cn

<sup>3</sup> School of Civil Engineering & Architecture, Southwest Petroleum University, Chengdu 610500, China; wjiyan0909@163.com

<sup>4</sup> Climate Change Cluster (C3), University of Technology Sydney, Sydney, NSW 2007, Australia; Qiaoyun.Xie@uts.edu.au

\* Correspondence: ainongli@imde.ac.cn; Tel./Fax: +86-028-85224131

Received: 3 May 2018; Accepted: 18 June 2018; Published: 22 June 2018



**Abstract:** The spatially explicit aboveground biomass (AGB) generated through upscaling field measurements is critical for carbon cycle simulation and optimized management of grasslands. However, the spatial gaps that exist in the optical remote sensing data, underutilization of the multispectral data cube and unavailability of uncertainty information hinder the generation of seamless and accurate AGB maps. This study proposes a novel framework to address the above challenges. The proposed framework filled the spatial gaps in the remote sensing data via the consistent adjustment of the climatology to actual observations (CACAO) method. Gaussian process regression (GPR) was used to fully exploit the multispectral data cube and generated the pixelwise uncertainty concurrent with the AGB estimation. A case study in a 100 km × 100 km area located in the Zoige Plateau, China was used to evaluate this framework. The results show that the CACAO method can fill almost all of the gaps, accounting for 93.1% of the study area, with satisfactory accuracy. The generated AGB map from the GPR was characterized by a relatively high accuracy ( $R^2 = 0.64$ ,  $RMSE = 48.13 \text{ g/m}^2$ ) compared to vegetation index-derived ones, and was accompanied by a corresponding uncertainty map that provides a new source of information on the credibility of each pixel. This study demonstrates the potential of the joint use of gap-filling and machine-learning methods to generate spatially explicit AGB.

**Keywords:** aboveground biomass (AGB); uncertainty; consistent adjustment of the climatology to actual observations (CACAO); Gaussian process regression (GPR)

## 1. Introduction

Grasslands are one of the most widespread ecosystems worldwide and cover over 30% of the Earth's land area, and their primary productivity accounts for approximately 20% of the total terrestrial productivity [1]. Therefore, grasslands are an important component in global carbon cycling. Furthermore, grasslands also provide many ecological and economic services for human beings [2]. Aboveground biomass (AGB), defined as the total mass of plant material per unit area, is an important indicator of carbon cycling and is identified as an essential climate variable by the Global Climate

Observing System (GCOS) [3]. For grassland ecosystems, AGB determines forage availability and thus constrains the herbivore carrying capacity [4]. Spatially explicit AGB information is critical for spatial–temporal variation analysis of grasslands under the context of climate change, and facilitates the sustainable utilization and optimized management of livestock husbandry [5].

Traditional measurement of grassland AGB through field clippings, laboratory drying and weighing is spatially sparse. Upscaling is needed to extrapolate the sparse field measurements in a seamless manner [2]. Multisource information, for example, synthetic aperture radar [6], terrestrial light detection and ranging [7], and growth models [8], have been demonstrated to be useful for mapping grassland AGB, and multisource information-synergized methods have been recommended for improving the accuracy of the grassland AGB estimation [9,10]. However, a fully optical remote sensing-driven estimation method is still preferred in most cases because of its low cost and easy implementation [4,5,10–12].

Optical remote sensing-based estimation methods mostly rely on the functional relationship between the field-measured AGB and remote sensing observations. Although this type of method is relatively mature, three issues have not been fully addressed. First, the spectral information in optical remote sensing has not been fully exploited. Generally, vegetation indices (VIs) are employed to reduce the influences of nonvegetation factors [13,14]. The normalized difference vegetation index (NDVI) is the most widely used VI for monitoring grassland AGB [15–17]. The soil-adjusted vegetation index (SAVI) and enhanced vegetation index (EVI) have also been widely used because they can suppress the influences of the soil background and atmosphere, respectively [14,18]. Although VIs can enhance the vegetation information as well as reduce nonvegetation information, they fundamentally underexploit the full potential of the multispectral data cube in vegetation monitoring [19], which explains the increasing popularity of machine-learning algorithms for the estimation of AGB by directly inputting the spectral reflectance instead of the VI [12,20]. Second, the uncertainty imbedded in the estimated AGB map has not been fully quantified. In existing studies concerning AGB estimation, the coefficient of determination ( $R^2$ ) and root-mean-square error (RMSE) between the field-measured and remote sensing-estimated AGBs are often used to quantify the general accuracy of the resulting AGB map [18]. These criteria can only provide a general view of the uncertainty in the AGB map. Pixelwise uncertainty is urgently needed for precision management of livestock husbandry and carbon cycle simulation in a data assimilation framework [18]. The third problem is the gaps that exist in the optical remote sensing image, which are caused by the presence of cloudiness and instrument problems [21]. This is a commonly encountered, yet long-overlooked problem in AGB estimation that hampers the generation of seamless AGB maps.

To address the first two issues, a unified mathematical framework is needed that can both fully exploit the multispectral remote sensing data and concomitantly generate AGB and uncertainty maps. Gaussian process regression (GPR) is a good choice to fulfill the above two targets [22,23]. It has been reported that the GPR can fully exploit multispectral remote sensing data and provide a superior accuracy than function regression in biophysical parameter estimation [19]. In addition, GPR provides uncertainty estimates according to Gaussian probability [24]. Due to its excellent performance in estimation accuracy and its capacity for generating pixelwise uncertainty, GPR has been widely used for the estimation of the leaf area index [25], leaf chlorophyll content [24] and water chlorophyll content [26]. However, to the best of our knowledge, its potential for grassland AGB estimation and uncertainty quantification have not yet been tested.

To solve the third problem resulting from data gaps, many spatial and temporal reflectance fusion methods have been proposed [27,28]. Most spatiotemporal fusion methods assume that at least one pair of fine- and coarse-resolution images can be collected during the study period, which may not be possible in unstable meteorology conditions, for example, in alpine grasslands. In addition, none of the current algorithms account for the nonlinear dynamics of vegetation growth, which is critical to temporally match field measurements [29], and may induce uncertainty in AGB upscaling. Therefore, temporal filtering methods are preferred for biophysical parameter estimation [30]. Several

temporal filtering methods have been used to fill gaps in remote sensing data, including polynomial fitting [31] and the Savitzky–Golay (SG) filter [32]. These methods have been successfully applied to coarse-resolution remote sensing data with a high temporal frequency. As for fine-spatial-resolution remote sensing data with a low temporal frequency, the temporal filtering methods may not be directly applicable because of the lack of samples for function fitting. In fact, coarse-resolution satellite data can be used to constrain the temporal profiles and, consequently, to fill the temporal gaps in fine-resolution satellite data [30,33]. Consistent adjustment of the climatology to actual observations (CACAO) is one such promising temporal-constraint method [34]. For each pixel, the phenology model constructed from coarse-resolution satellite data is adjusted by shifting its time reference and scaling its magnitude to best fit the actual fine-resolution observations. Then, the value at an arbitrary day can be straightforwardly extracted from the adjusted phenology model [34]. It has been demonstrated that the CACAO method can effectively fill temporal gaps, even for long periods of missing data [34,35].

The overall objective of this study is to upscale the field measurements to generate seamless grassland AGB and concomitant uncertainty maps. The scientific questions that we addressed are as follows: (1) Can the gaps in the remote sensing data be successfully removed via CACAO method? (2) Does the GPR outperform function regression methods for AGB estimation? (3) Can a pixelwise uncertainty map be retrieved or not?

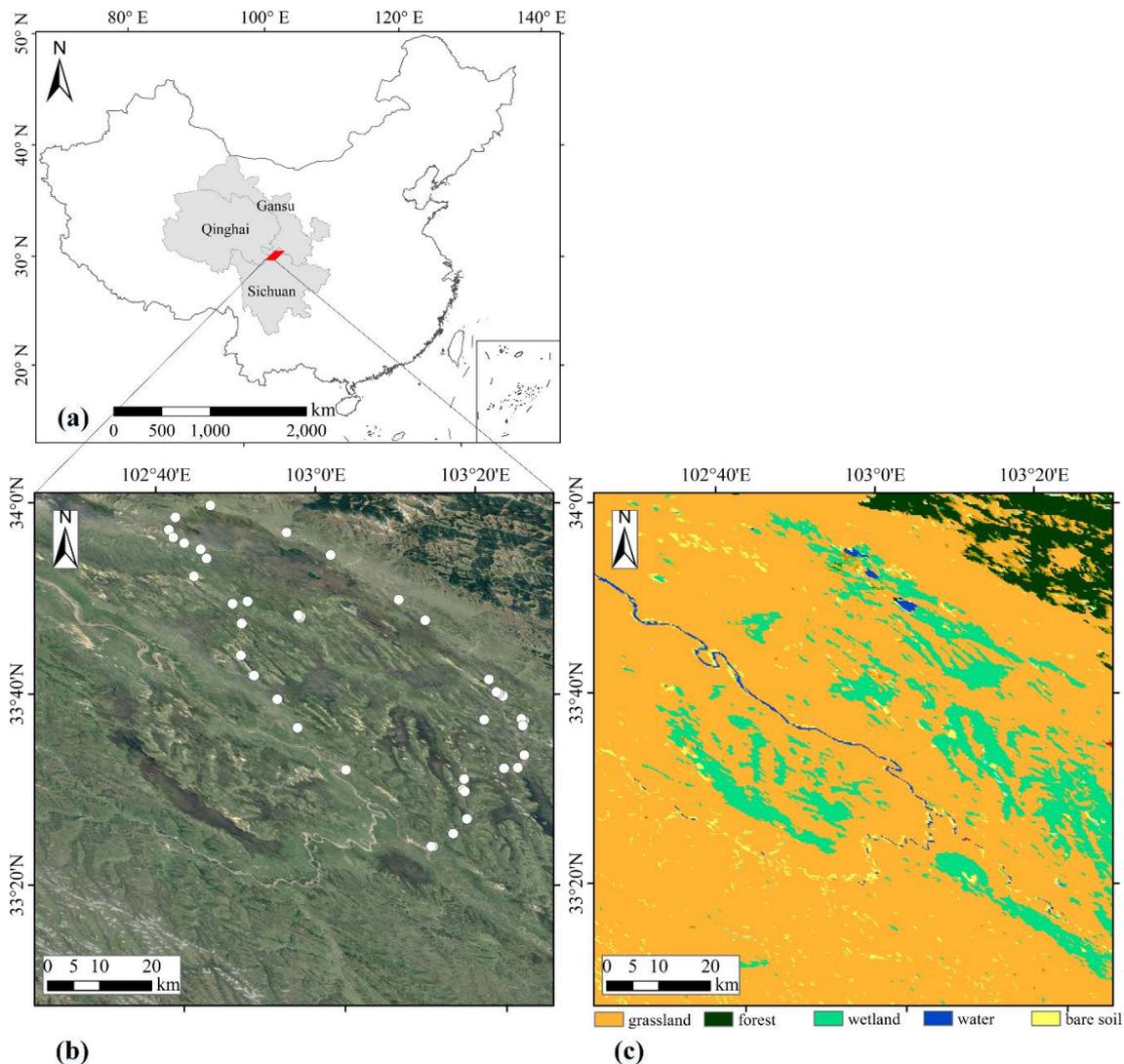
## 2. Data Collection

### 2.1. Study Area and Field Measurements

This study was conducted in a 100 km × 100 km region centered at ~33°35' N, 102°27' E (Figure 1), which is in the Zoige Plateau. This region is located at the junction of Sichuan, Gansu and Qinghai provinces, China. According to the land cover map generated by the European Space Agency Climate Change Initiative (CCI) project (available at <http://maps.elie.ucl.ac.be/CCI/viewer/download.php>), the dominant ecosystem in the study area is grassland (80.8%), followed by wetland (11.1%) and forest (4.2%) (Figure 1c). Only grassland was selected to estimate the AGB, and other land covers were masked out.

In recent years, the degradation of grasslands has become an increasingly serious problem, which is directly caused by the influence of human activities, such as overgrazing [36,37].

The field campaign was conducted from 14 to 19 August 2017. Forty-four homogeneous plots, each with an area of approximately 100 m × 100 m, were selected to perform the field measurement (Figure 1). A GPS was used to record the geographical positions of the plot centers. In each plot, three 0.5 m × 0.5 m quadrats were selected. All plants in each quadrat were clipped at the ground surface, transported to the laboratory, and oven-dried at 80 °C for 48 hours until a constant dry biomass was obtained. The average AGB value of the three quadrats within one plot was used to represent the AGB at the plot scale.



**Figure 1.** (a) Location of the study area; (b,c) a Google Earth image and the land cover of the study area, respectively. Sampling plots for the field aboveground biomass measurements are displayed as white dots in (b).

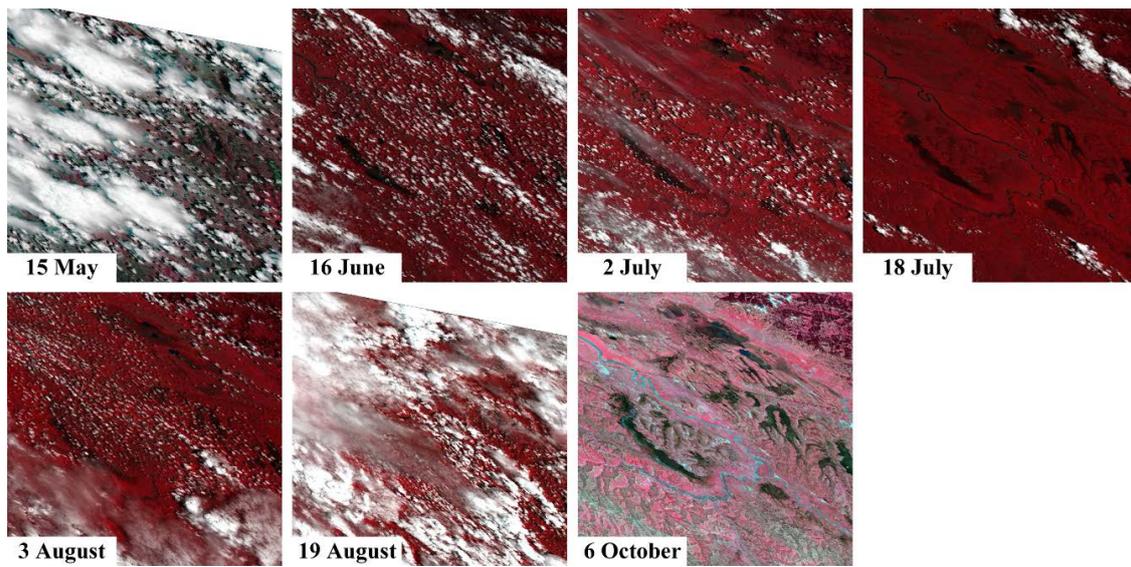
## 2.2. Landsat 8 OLI Reflectance Data

Landsat 8 Operational Land Imager (OLI) data, acquired through the Earth Resources Observation and Science Center Science Processing Architecture (ESPA) on-demand interface (<https://espa.cr.usgs.gov/>), were used in this study. The ESPA provides canopy reflectance through an atmospheric correction algorithm based on the Second Simulation of the Satellite Signal in the Solar Spectrum Vectorial (6SV) model [38]. The 6SV model was integrated into the Landsat 8 Surface Reflectance Code (LaSRC), which was used to generate the high-level Landsat data products.

All of the Landsat 8 OLI scenes were collected during the growing season in 2017, covering the study area (path 131 and row 36, and path 131 and row 37), with a cloud cover of less than 80%. When two scenes, with respective rows of 36 and 37, were obtained for a single date, they were composited to achieve a more widespread coverage of the study area. During the composition, a clear pixel without cloudy contamination was preferentially selected. When the pixels from the two scenes were both clear, the scene of path 131 and row 36 was used. The clear pixel was identified through the pixel quality layer (with pixel value of 322) accompanied by the reflectance data provided by the LaSRC [39]. This composition procedure resulted in 7 remote sensing images that were acquired on

15 May, 16 June, 2 July, 18 July, 3 August, 19 August and 6 October, 2017. The compositing results are shown in Figure 2. The image that was temporally synchronous with the field measurements (19 August) was severely contaminated by clouds (93.1%), and it was impossible to upscale the field-measured AGB to a seamless map without gap filling.

All potentially useful bands, including blue, green, red, near infrared (NIR), and the two shortwave infrared (SWIR1 and SWIR2) bands, were used in the following steps. The resolution of the OLI data was resampled to 100 m using an aggregation sampling method to be spatially consistent with the plot scale (with area of 100 m × 100 m).



**Figure 2.** Compositing Landsat OLI 8 images covering the study area in 2017.

### 2.3. MCD43A4 Reflectance Data

The MODIS Nadir Bidirectional Reflectance Distribution Function (BRDF) Adjusted Reflectance (NBAR) product (MCD43A4) provides 500 m reflectance data that are adjusted using a kernel-driven BRDF model to predict the values as if they were taken from nadir views [40]. This product is produced every 8 days using 16 days of combined data from the Aqua and Terra satellites. All scenes with the tile number of h26v05 covering the study area in 2017 were downloaded from the Land Processes Distributed Active Archive Center (LPDAAC) (<https://lpdaac.usgs.gov>). Similar to Landsat 8 OLI, the corresponding six bands in the MCD43A4 data were selected. The MCD43A4 data were also resampled to a 100 m spatial resolution using the nearest-neighbor sampling method.

## 3. Methodology

### 3.1. CACAO Method

CACAO is a gap-filling method with a temporal constraint [34]. The temporal constraint of CACAO makes it suitable to capture temporal variations of vegetation, which is important for grassland AGB estimation. The CACAO method is composed of two main steps [34]. A phenology model was first constructed using MCD43A4 data for each pixel at each band, and then, the phenology model was shifted and scaled to minimize its difference from the Landsat 8 OLI data in terms of the RMSE.

The MCD43A4 reflectance time series that was smoothed through SG filtering [32] was used to construct the phenology model. SG filtering smooths the time series by temporally fitting successive subsets of adjacent data points with a low-degree polynomial via the method of linear least squares. In this study, the degree of the smoothing polynomial and half-width of the smoothing window were both specified as 2.

Then, minimization of the RMSE was implemented, which can be formalized as:

$$\arg \min_{scale, shift} \sqrt{\frac{1}{n} \sum_{i=1}^n (R_{OLI}(t_i) - scale \cdot R_{pm}(t_i + shift))^2} \quad (1)$$

where  $R_{OLI}$  is the Landsat 8 OLI reflectance;  $R_{pm}$  is the reflectance extracted from the phenology model;  $n$  is the number of available Landsat 8 OLI observations (in this study,  $n = 7$ );  $t$  is the time in days; and  $scale$  and  $shift$  are the requested parameters, which can be determined through the method of least squares. Minimization of the RMSE was done pixel by pixel and band by band. After obtaining the values of  $scale$  and  $shift$  for each pixel in each band, the reconstructed reflectance at the targeted date (19 August) was easily extracted from the adjusted phenology model.

In the reconstructed reflectance from CACAO, the temporal and spatial details are from MCD43A4 and Landsat 8 OLI data, respectively. Note that the spectral normalization between MCD43A4 and Landsat 8 OLI data was not implemented because the difference in their spectral response functions can be implicitly accounted for during the shifting and scaling procedure, considering the significant linear correlation between the MODIS and OLI reflectance products [38].

### 3.2. Gaussian Process Regression

GPR is a probabilistic approximation of nonparametric kernel-based regression that can provide both a predictive mean (represents pixelwise estimate of the AGB in this study) and predictive variance (error bar for the AGB prediction). The details of GPR can be found in the literature [19,22,24], and we only provide a brief description here.

The GPR model transforms the input to the output in the form of [23]:

$$\hat{y} = f(x) = \sum_{i=1}^N \alpha_i K(x_i, x) \quad (2)$$

where  $\{x_i\}_{i=1}^N$  are the reflectance values in the training dataset;  $\alpha_i$  is the weight assigned to each one of the values; and  $K$  is a kernel function that evaluates the similarity between the test reflectance  $x$  and all  $N$  ( $N = 44$  in this study) training reflectance values. In this study, a scaled Gaussian kernel function was used, and its specific expression can be found in [19,24].

For prediction purposes, GPR is obtained by computing the posterior distribution over the unknown output  $y_*$ ,  $p(y_* | x_*, D)$ , where  $D = \{x_n, y_n | n = 1, 2, \dots, N\}$  is the training dataset, that is, the 44 reflectance–AGB pairs in this study. This posterior is shown to be a Gaussian distribution,  $p(y_* | x_*, D) = N(\mu_*, \sigma_*)$ , for which one can estimate the predictive mean value (AGB predictions):

$$\mu_* = K_*^T (K + \sigma_n^2 I)^{-1} y \quad (3)$$

and the predictive variance:

$$\sigma_*^2 = K_{**} - K_*^T (K + \sigma_n^2 I)^{-1} K_* \quad (4)$$

Note that the variance is the difference between two terms: the first term,  $K_{**}$ , is the prior covariance, from which a positive term is subtracted. Therefore, the posterior variance is always smaller than the prior variance since the observations contribute additional information [23]. Theoretically, the GPR-estimated uncertainty depends on the training reflectance. If the test reflectance is similar to the training reflectance, the uncertainty of the test set is relatively small because we are dealing with similar input features [24]. To account for the dependence of the variance on the AGB value, the coefficient of variation (CV), rather than the variance per se, was used to represent the quality of the AGB. The CV is calculated as:

$$CV = \sigma_* / \mu_* \quad (5)$$

### 3.3. Function Regression between Field Measurements and Vegetation Indices

To assess the gains from the introduction of GPR into AGB upscaling, it was compared with function regression methods. Three commonly used vegetation indices were selected for the parameter regression methods, that is, NDVI, SAVI and EVI. They are formulated as:

$$\text{NDVI} = \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\rho_{\text{NIR}} + \rho_{\text{red}}} \quad (6)$$

$$\text{SAVI} = \frac{1.5(\rho_{\text{NIR}} - \rho_{\text{red}})}{\rho_{\text{NIR}} + \rho_{\text{red}} + 0.5} \quad (7)$$

$$\text{EVI} = \frac{2.5(\rho_{\text{NIR}} - \rho_{\text{red}})}{\rho_{\text{NIR}} + 6\rho_{\text{red}} - 7.5\rho_{\text{blue}} + 1} \quad (8)$$

where  $\rho_{\text{NIR}}$ ,  $\rho_{\text{red}}$ , and  $\rho_{\text{blue}}$  are the reflectance in near-infrared, red and blue bands, respectively. Three function types, that is, linear, exponential and logarithmic functions, were tested. Those with the lowest RMSE were selected for each VI.

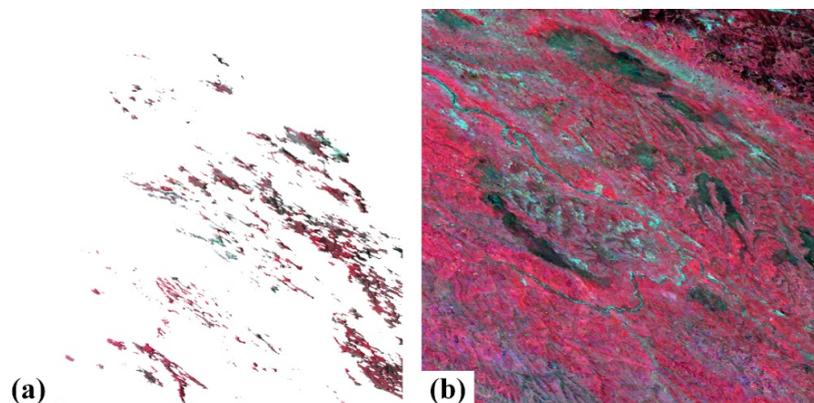
### 3.4. Assessment Method

The leave-one-out cross-validation (LOOCV) scheme was used to assess the performance of GPR and the function regression methods. For a given measured AGB to be validated, the remaining measured AGB values were used to construct the transfer function from the remote sensing observation to the AGB, and then, this constructed transfer function was used to predict the validated AGB. This process was iterated for each measured AGB. Finally, we compared all of the measured and corresponding predicted AGB values. The performance of each upscaling method was evaluated using the RMSE to assess the accuracy and  $R^2$  to account for the goodness of fit.

## 4. Results

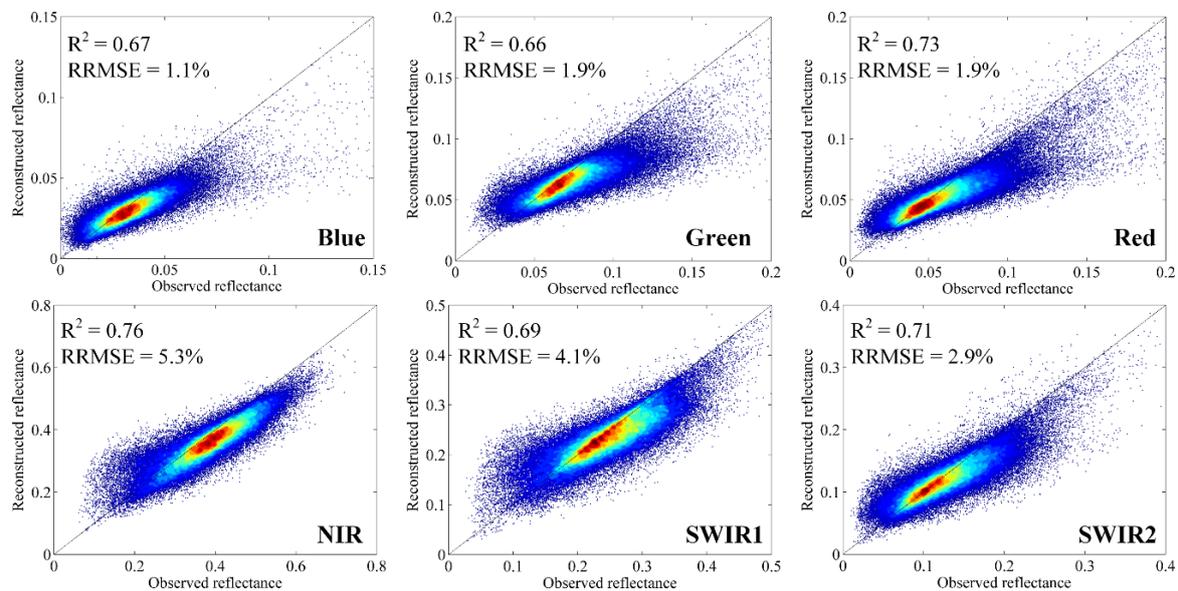
### 4.1. Gap-Filled Reflectance

Gap filling is the prerequisite for upscaling the field-measured grassland AGB using remote sensing images with gaps. The images temporally synchronous with the field campaign (19 August 2017) before and after gap filling are shown in Figure 3. Pixels contaminated by clouds were all filtered out in Figure 3. Before gap filling, the image was extremely cloudy (see also Figure 2) and the clear pixels without cloud contamination only accounted for 6.9% of the total study area. It is impossible to upscale the field-measured grassland AGB using the original image without gap filling. CACAO interpolated nearly all of the gaps and successfully reconstructed the spatial pattern of the study area.



**Figure 3.** (a) Original and (b) gap-filled Landsat OLI image on 19 August 2017. The cloudy pixels in the original image were masked out.

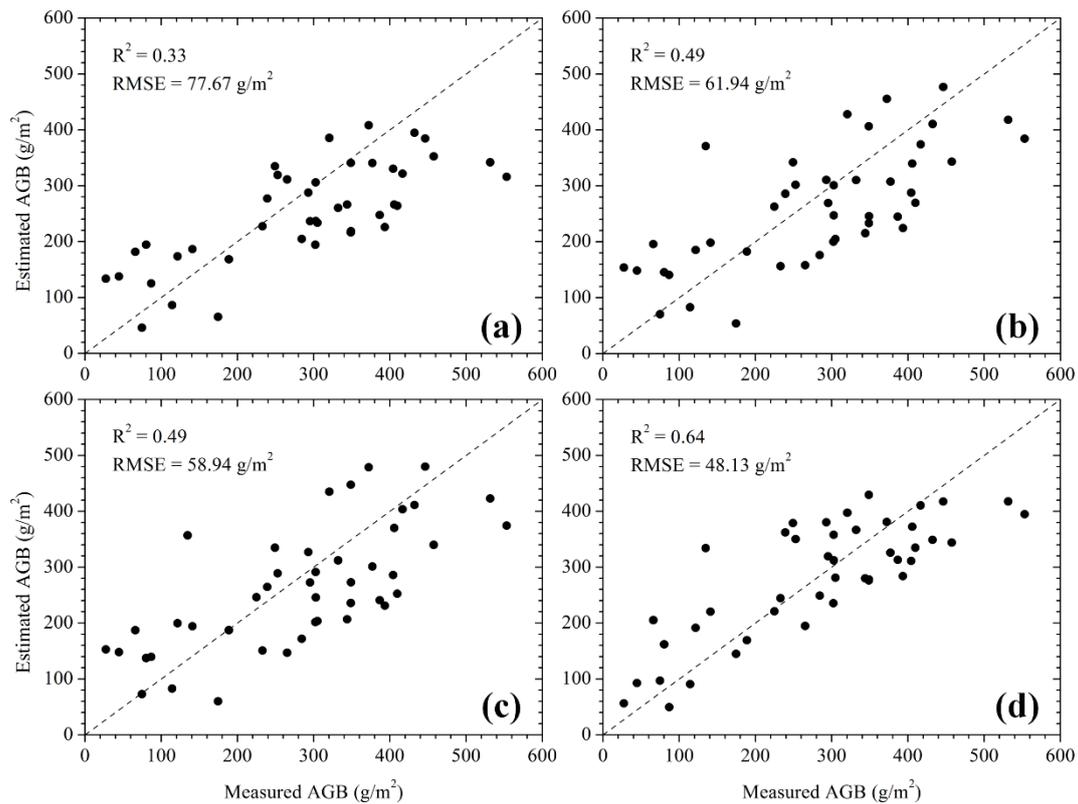
To quantitatively evaluate the gap-filled image, we compared the reconstructed and original reflectance of the clear pixels on 19 August 2017. The comparison results for the selected six bands are shown in Figure 4. The points cluster around the 1:1 lines, with  $R^2$  ranging between 0.66 and 0.76 and the relative RMSE (RRMSE) ranging between 1.1% and 5.3%. Generally, CACAO reconstructed the original reflectance with high accuracy.



**Figure 4.** Density scatterplots between CACAO (consistent adjustment of the climatology to actual observations) reconstructed and OLI-observed reflectance for each band on 19 August 2017.

#### 4.2. Upscaling Assessment

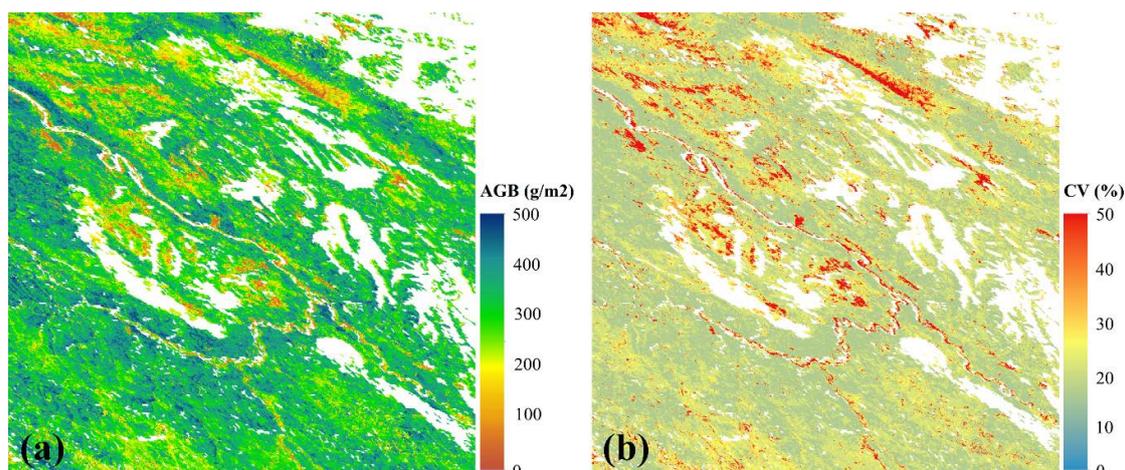
The LOOCV results for the NDVI, SAVI and EVI using function regression and the reflectance using GPR are shown in Figure 5. The function regression method using NDVI performed the poorest ( $R^2 = 0.33$ ,  $RMSE = 77.67 \text{ g/m}^2$ ) compared with the other methods. The estimated AGB gradually became saturated when the AGB was greater than  $400 \text{ g/m}^2$  due to the saturation problem of the NDVI in detecting dense vegetation [41]. Function regressions using SAVI and EVI have similar performances. SAVI and EVI both obtained better results than the NDVI because of the suppression of nonvegetation information, for example, the soil background and atmosphere, in the formulation of these two VIs. GPR outperformed the three function regression methods significantly, with  $R^2 = 0.64$  and  $RMSE = 48.13 \text{ g/m}^2$ . Therefore, the GPR method was selected to upscale the field-measured AGB.



**Figure 5.** Results of the leave-one-out cross-validation. (a–c) Cases from the function regression between the field-measured aboveground biomass (AGB) measurements and the NDVI, SAVI and EVI, respectively. (d) Gaussian process regression.

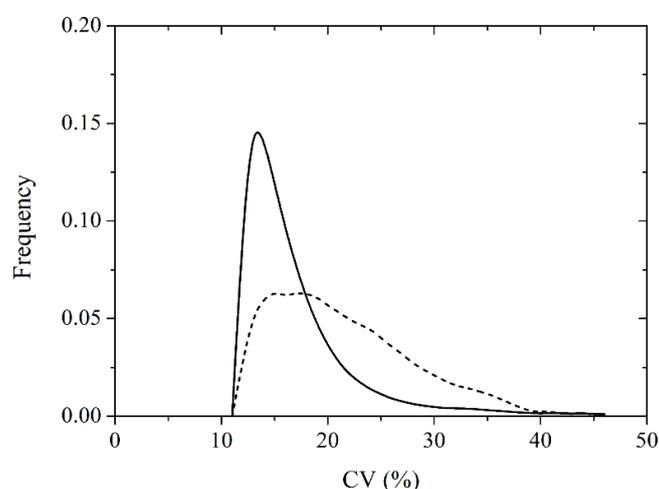
#### 4.3. Aboveground Biomass and Uncertainty Maps

GPR trained with all of the field-measured AGB and the corresponding reconstructed reflectance was used to map the concomitant AGB and CV (Figure 6). The AGB map provides detailed information regarding the spatial dynamics of vegetation. For example, areas near rivers have higher AGB values than areas far from rivers because of the availability of moisture. A closer inspection of the CV map shows that areas with low AGB values (less than  $100 \text{ g/m}^2$ ) correspond to high CV values (implying a high uncertainty). Three main factors can explain this phenomenon. First, the uncertainty derived from GPR quantifies the similarity between the test and training datasets [24], and values less than  $100 \text{ g/m}^2$  were slightly undersampled (see Figure 5). Therefore, GPR can provide valuable feedback regarding the improvement of the sampling design. Second, the CV was formulated as the ratio between the predictive mean and standard deviation, so a low mean tends to result in a high CV. Finally, the CCI land-cover product assigns some nongrassland covers as grassland. For example, a careful comparison between Figures 1c and 3b revealed that many bare soil and river pixels were identified as grassland in the CCI product, and these misclassified pixels almost correspond to the high CV values.



**Figure 6.** (a) Aboveground biomass (AGB) and (b) coefficient of variation (CV) maps generated from Gaussian process regression. The white parts represent nongrassland land-cover types.

Verrelst et al. [24] argued that an uncertainty map can be used to exclude nongrassland pixels through thresholding. To test this, we analyzed the CV frequency distribution histograms for grassland and nongrassland pixels (see Figure 7). The histogram for grassland showed a sharp shape, with most of the pixels meeting the 20% uncertainty threshold established by GCOS [42]. By contrast, the nongrassland histogram had a much flatter distribution. Nevertheless, no obvious threshold was able to be readily established to separate grassland and nongrassland pixels according to the CV values. This phenomenon can be attributed to the large amount of wetland pixels (accounting for 11.1%) in our study area. The wetland was covered by grass, making its spectral signal similar to that of grassland.



**Figure 7.** CV frequency distribution histograms for grassland (solid line) and nongrassland (dashed line).

## 5. Discussion

### 5.1. Gap Filling for Grassland AGB Estimation

Gap filling is the prerequisite for upscaling field-measured AGB to a seamless AGB map when the used remote sensing image is cloud contaminated [21,30,33]. This has been long neglected in AGB upscaling. In this study, we use CACAO to fill the spatial gaps in Landsat 8 OLI data by borrowing temporal information from MCD43A4 data. This process can fill nearly all of the gaps with satisfactory accuracy (see Figures 3 and 4).

However, CACAO assumes that the temporal curves of the reflectance from fine- and coarse-resolution data have similar shapes, with only differences in the magnitude and shift [29]. This assumption may be not always the case when the land surface is heterogeneous [43]. The uncertainty in the gap-filled reflectance caused by spatial heterogeneity and its propagation into the final AGB map requires a deep analysis. In addition, a more advanced method to account for the scale difference between fine- and coarse-resolution data is needed to further improve the CACAO method. Prognostic phenology models, which allow the temporal dynamics of vegetation under variable weather conditions to be simulated [44], may be used to constrain the temporal curves of fine-resolution data.

### 5.2. Benefits from GPR

GPR can lead to an improved performance compared to the traditional function regression methods (Figure 5). Two main reasons can explain this improvement. First, the reformulation of reflectance to VIs loses some degree of vegetation information, as has been shown in [12] and [20]. Second, GPR can cope with the strong nonlinearity of the functional dependence between the biophysical parameter and observed remote sensing signal. A comparison study performed by Verrelst et al. [45] shows that GPR outperforms most of the existing machine-learning methods, including kernel ridge regression, regression trees, the relevance vector machine, neural network, and the partial-least-squares regression, in biophysical parameter retrieval.

One of the most obvious advantages of GPR over other methods is that it can generate a pixelwise uncertainty map concurrent with the AGB map [24]. Concomitant uncertainty is a prerequisite for the proper use of the upscaled AGB. For example, the uncertainty map can be evaluated against the uncertainty threshold of 20% for the AGB proposed by the GCOS [42]. The highly reliable retrieval meeting this uncertainty threshold can be easily extracted for subsequent applications. In addition, the concomitant uncertainty map facilitates the integration of remote sensing and process models for carbon cycle simulations [46]. For example, the spatially explicit AGB can be easily assimilated by process models because their contribution to an assimilation framework can be automatically weighted according to the concomitant uncertainty maps.

One of the main concerns regarding the use of a multispectral cube through GPR, rather than VIs, is that it may face the so-called curse of dimensionality problem. For example, Verrelst et al. [19] found that the collinearity of adjacent bands can lead to information redundancy in biophysical parameter retrieval using GPR with hyperspectral data. However, the spectral resolution of Landsat 8 OLI data is relatively low, so it should not lead to a significant collinearity of different bands [38]. In addition, GPR has the potential to best exploit useful information from remote sensing data to estimate biophysical parameters, even against high-dimensional data [22].

### 5.3. Selection of Vegetation Indices

In the function regression methods used for comparison with GPR, only three VIs were selected, that is, the NDVI, EVI and SAVI, because they are the most commonly used VIs in biophysical parameter estimation [13,16,18]. Many studies revealed that narrow-band VIs, especially those based on a red edge band, can provide an improved estimate of the AGB [41,47]. However, sensors equipped with narrow bands are still limited. This study demonstrates that wide-band remote sensing can also obtain high accuracy in AGB estimation, provided that the proper method is selected.

In addition, several studies have shown that besides VIs, many other factors can explain the variations of grassland AGB, for example, the leaf area index [15,20], gross primary production [48] and topography [20]. Therefore, the multifactor modeling of grassland AGB recommended by Liang et al. [9] should be tested in future work to further improve the accuracy of the resulting AGB map.

## 6. Conclusions

In this study, we proposed a framework to upscale field-measured AGB to a seamless map based on the integrated use of the CACAO method to fill the gaps in remote sensing data, and GPR to fully exploit the multispectral data cube and generate an uncertainty map concurrent with the AGB map. The performance of the proposed method was evaluated in a 100 km × 100 km study area located in the Zoige Plateau, China. The results show that CACAO can fill nearly all of the gaps, accounting for 93.1% of the study area, with a satisfactory accuracy. The AGB map generated from GPR has a relatively high accuracy ( $R^2 = 0.64$ , RMSE = 48.13 g/m<sup>2</sup>) compared to those generated from vegetation indices, and was accompanied by a corresponding uncertainty map. This study contributes to carbon cycle simulations and optimized management of grassland.

**Author Contributions:** G.Y. and A.L. conceived and designed the experiments; G.Y. performed the experiments; H.J., J.B. and G.L. analyzed the data; J.W., Z.Z. and X.N. contributed to the field measurements. G.Y. wrote the paper; C.W. and Q.X. revised the paper.

**Funding:** This work was funded by the National Natural Science Foundation of China (41571373, 41601403, 41631180 and 41531174), the GF6 Project (30-Y20A03-90030-17/18), the National Key Research and Development Program of China (2016YFA0600103) and the Youth Talent Team Program of the Institute of Mountain Hazards and Environment, CAS (SDSQB-2015-02).

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

## References

1. Scurlock, J.M.O.; Hall, D.O. The global carbon sink: A grassland perspective. *Glob. Chang. Biol.* **1998**, *4*, 229–233. [[CrossRef](#)]
2. Shoko, C.; Mutanga, O.; Dube, T. Progress in the remote sensing of C3 and C4 grass species aboveground biomass over time and space. *ISPRS J. Photogramm. Remote Sens.* **2016**, *120*, 13–24. [[CrossRef](#)]
3. Bojinski, S.; Verstraete, M.; Peterson, T.C.; Richter, C.; Simmons, A.; Zemp, M. The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 1431–1443. [[CrossRef](#)]
4. Yang, Y.H.; Fang, J.Y.; Pan, Y.D.; Ji, C.J. Aboveground biomass in Tibetan grasslands. *J. Arid Environ.* **2009**, *73*, 91–95. [[CrossRef](#)]
5. Gao, T.; Yang, X.C.; Jin, Y.X.; Ma, H.L.; Li, J.Y.; Yu, H.D.; Yu, Q.Y.; Zheng, X.; Xu, B. Spatio-Temporal Variation in Vegetation Biomass and Its Relationships with Climate Factors in the Xilingol Grasslands, Northern China. *PLoS ONE* **2013**, *8*, e83824. [[CrossRef](#)] [[PubMed](#)]
6. Sinha, S.; Jeganathan, C.; Sharma, L.K.; Nathawat, M.S. A review of radar remote sensing for biomass estimation. *Int. J. Environ. Sci. Technol.* **2015**, *12*, 1779–1792. [[CrossRef](#)]
7. Greaves, H.E.; Vierling, L.A.; Eitel, J.U.H.; Boelman, N.T.; Magney, T.S.; Prager, C.M.; Griffin, K.L. Estimating aboveground biomass and leaf area of low-stature Arctic shrubs with terrestrial LiDAR. *Remote Sens. Environ.* **2015**, *164*, 26–35. [[CrossRef](#)]
8. He, B.B.; Li, X.; Quan, X.W.; Qiu, S. Estimating the aboveground dry biomass of grass by assimilation of retrieved LAI into a crop growth model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 550–561. [[CrossRef](#)]
9. Liang, T.G.; Yang, S.X.; Feng, Q.S.; Liu, B.K.; Zhang, R.P.; Huang, X.D.; Xie, H.J. Multi-factor modeling of above-ground biomass in alpine grassland: A case study in the Three-River Headwaters Region, China. *Remote Sens. Environ.* **2016**, *186*, 164–172. [[CrossRef](#)]
10. Reddersen, B.; Fricke, T.; Wachendorf, M. A multi-sensor approach for predicting biomass of extensively managed grassland. *Comput. Electron. Agric.* **2014**, *109*, 247–260. [[CrossRef](#)]
11. Quan, X.W.; He, B.B.; Yebra, M.; Yin, C.M.; Liao, Z.M.; Zhang, X.T.; Li, X. A radiative transfer model-based method for the estimation of grassland aboveground biomass. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *54*, 159–168. [[CrossRef](#)]
12. Sibanda, M.; Mutanga, O.; Rouget, M. Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS J. Photogramm. Remote Sens.* **2015**, *110*, 55–65. [[CrossRef](#)]

13. Li, F.; Zeng, Y.; Luo, J.H.; Ma, R.H.; Wu, B.F. Modeling grassland aboveground biomass using a pure vegetation index. *Ecol. Indic.* **2016**, *62*, 279–288. [[CrossRef](#)]
14. Ullah, S.; Si, Y.; Schlerf, M.; Skidmore, A.K.; Shafique, M.; Iqbal, I.A. Estimation of grassland biomass and nitrogen using MERIS data. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *19*, 196–204. [[CrossRef](#)]
15. Dusseux, P.; Hubert-Moy, L.; Corpetti, T.; Vertès, F. Evaluation of SPOT imagery for the estimation of grassland biomass. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *38*, 72–77. [[CrossRef](#)]
16. Liu, S.L.; Cheng, F.Y.; Dong, S.K.; Zhao, H.D.; Hou, X.Y.; Wu, X. Spatiotemporal dynamics of grassland aboveground biomass on the Qinghai-Tibet Plateau based on validated MODIS NDVI. *Sci. Rep.* **2017**, *7*, 4182. [[CrossRef](#)] [[PubMed](#)]
17. Piao, S.L.; Fang, J.Y.; Zhou, L.M.; Tan, K.; Tao, S. Changes in biomass carbon stocks in China's grasslands between 1982 and 1999. *Glob. Biogeochem. Cycles* **2007**, *21*, GB2002. [[CrossRef](#)]
18. Jia, W.X.; Liu, M.; Yang, Y.H.; He, H.L.; Zhu, X.D.; Yang, F.; Yin, C.; Xiang, W.N. Estimation and uncertainty analyses of grassland biomass in Northern China: Comparison of multiple remote sensing data sources and modeling approaches. *Ecol. Indic.* **2016**, *60*, 1031–1040. [[CrossRef](#)]
19. Verrelst, J.; Alonso, L.; Camps-Valls, G.; Delegido, J.; Moreno, J. Retrieval of vegetation biophysical parameters using Gaussian process techniques. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1832–1843. [[CrossRef](#)]
20. Xie, Y.C.; Sha, Z.Y.; Yu, M.; Bai, Y.F.; Zhang, L. A comparison of two models with Landsat data for estimating above ground grassland biomass in Inner Mongolia, China. *Ecol. Model.* **2009**, *220*, 1810–1818. [[CrossRef](#)]
21. Fang, H.L.; Liang, S.L.; Townshend, J.R.; Dickinson, R.E. Spatially and temporally continuous LAI data sets based on an integrated filtering method: Examples from North America. *Remote Sens. Environ.* **2008**, *112*, 75–93. [[CrossRef](#)]
22. Camps-Valls, G.; Gómez-Chova, L.; Muñoz-Marí, J.; Vila-Francés, J.; Amorós, J.; Valle-Tascon, S.D.; Calpe-Maravilla, J. Biophysical parameter estimation with adaptive Gaussian processes. In Proceedings of the 2009 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2009), Cape Town, South Africa, 12–17 July 2009; pp. 69–72.
23. Rasmussen, C.E.; Williams, C.K. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2006; Volume 2, p. 4.
24. Verrelst, J.; Rivera, J.P.; Moreno, J.; Camps-Valls, G. Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. *ISPRS J. Photogramm. Remote Sens.* **2013**, *86*, 157–167. [[CrossRef](#)]
25. Mousivand, A.; Menenti, M.; Gorte, B.; Verhoef, W. Multi-temporal, multi-sensor retrieval of terrestrial vegetation properties from spectral-directional radiometric data. *Remote Sens. Environ.* **2015**, *158*, 311–330. [[CrossRef](#)]
26. Bazi, Y.; Alajlan, N.; Melgani, F.; AlHichri, H.; Yager, R.R. Robust estimation of water chlorophyll concentrations with Gaussian process regression and IOWA aggregation operators. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3019–3028. [[CrossRef](#)]
27. 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.
28. 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](#)]
29. Yin, G.F.; Li, A.N.; Jin, H.A.; Zhao, W.; Bian, J.H.; Qu, Y.H.; Zeng, Y.L.; Xu, B.D. Derivation of temporally continuous LAI reference maps through combining the LAINet observation system with CACAO. *Agric. For. Meteorol.* **2017**, *233*, 209–221. [[CrossRef](#)]
30. Zhang, H.K.; Chen, J.M.; Huang, B.; Song, H.H.; Li, Y.R. Reconstructing Seasonal Variation of Landsat Vegetation Index Related to Leaf Area Index by Fusing with MODIS Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 950–960. [[CrossRef](#)]
31. Karnieli, A.; Gabai, A.; Ichoku, C.; Zaady, E.; Shachak, M. Temporal dynamics of soil and vegetation spectral responses in a semi-arid environment. *Int. J. Remote Sens.* **2002**, *23*, 4073–4087. [[CrossRef](#)]
32. Chen, J.; Jonsson, P.; Tamura, M.; Gu, Z.H.; Matsushita, B.; Eklundh, L. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sens. Environ.* **2004**, *91*, 332–344. [[CrossRef](#)]

33. Yin, G.; Li, A.; Jin, H.; Bian, J. Spatiotemporal fusion through the best linear unbiased estimator to generate fine spatial resolution NDVI time series. *Int. J. Remote Sens.* **2018**, *39*, 3287–3305. [[CrossRef](#)]
34. Verger, A.; Baret, F.; Weiss, M.; Kandasamy, S.; Vermote, E. The CACAO method for smoothing, gap filling, and characterizing seasonal anomalies in satellite time series. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 1963–1972. [[CrossRef](#)]
35. Verger, A.; Baret, F.; Weiss, M. Near real-time vegetation monitoring at global scale. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 3473–3481. [[CrossRef](#)]
36. Li, A.N.; Bian, J.H.; Lei, G.B.; Huang, C.Q. Estimating the maximal light use efficiency for different vegetation through the CASA model combined with time-series remote sensing data and ground measurements. *Remote Sens.* **2012**, *4*, 3857–3876. [[CrossRef](#)]
37. Wang, J.Y.; Li, A.N.; Bian, J.H. Simulation of the Grazing Effects on Grassland Aboveground Net Primary Production Using DNDC Model Combined with Time-Series Remote Sensing Data—A Case Study in Zoige Plateau, China. *Remote Sens.* **2016**, *8*, 168. [[CrossRef](#)]
38. Vermote, E.; Justice, C.; Claverie, M.; Franch, B. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* **2016**, *185*, 46–56. [[CrossRef](#)]
39. United States Geological Survey (USGS). *Landsat 8 Surface Reflectance Code (LaSRC) Product Guide*, USGS: Reston, VA, USA, 2017.
40. 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](#)]
41. Mutanga, O.; Skidmore, A.K. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *Int. J. Remote Sens.* **2004**, *25*, 3999–4014. [[CrossRef](#)]
42. GCOS. *Systematic Observation Requirements for Satellite-Based Products for Climate: Supplemental Details to the Satellite-Based Component of the “Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC”*; World Meteorological Organization: Geneva, Switzerland, 2011; p. 89.
43. Yin, G.F.; Li, J.; Liu, Q.H.; Li, L.H.; Zeng, Y.L.; Xu, B.D.; Yang, L.; Zhao, J. Improving leaf area index retrieval over heterogeneous surface by integrating textural and contextual information: A case study in the Heihe River Basin. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 359–363.
44. Savoy, P.; Mackay, D.S. Modeling the seasonal dynamics of leaf area index based on environmental constraints to canopy development. *Agric. For. Meteorol.* **2015**, *200*, 46–56. [[CrossRef](#)]
45. Verrelst, J.; Pablo Rivera, J.; Veroustraete, F.; Munoz-Mari, J.; Clevers, J.G.P.W.; Camps-Valls, G.; Moreno, J. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods—A comparison. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 260–272. [[CrossRef](#)]
46. Scholze, M.; Buchwitz, M.; Dorigo, W.; Guanter, L.; Shaun, Q.G. Reviews and syntheses: Systematic Earth observations for use in terrestrial carbon cycle data assimilation systems. *Biogeosciences* **2017**, *14*, 3401–3429. [[CrossRef](#)]
47. Ramoelo, A.; Cho, M.A.; Mathieu, R.; Madonsela, S.; van de Kerchove, R.; Kaszta, Z.; Wolff, E. Monitoring grass nutrients and biomass as indicators of rangeland quality and quantity using random forest modelling and World View-2 data. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *43*, 43–54. [[CrossRef](#)]
48. Fu, X.Y.; Tang, C.J.; Zhang, X.X.; Fu, J.Y.; Jiang, D. An improved indicator of simulated grassland production based on MODIS NDVI and GPP data: A case study in the Sichuan province, China. *Ecol. Indic.* **2014**, *40*, 102–108. [[CrossRef](#)]

