8-2016

Linking in situ LAI and Fine Resolution Remote Sensing Data to Map Reference LAI over Cropland and Grassland Using Geostatistical Regression Method

Yaqian He
Beijing Normal University

Yanchen Bo
Beijing Normal University

Leilei Chai
Beijing Normal University

Xiaolong Liu
Yunnan Normal University

Aihua Li
Boise State University

Publication Information

He, Yaqian; Bo, Yanchen; Chai, Leilei; Liu, Xiaolong; and Li, Aihua. (2016). "Linking in situ LAI and Fine Resolution Remote Sensing Data to Map Reference LAI over Cropland and Grassland Using Geostatistical Regression Method". *International Journal of Applied Earth Observation and Geoinformation, 50*, 26-38. https://doi.org/10.1016/j.jag.2016.02.010

Linking in situ LAI and fine resolution remote sensing data to map reference LAI over cropland and grassland using geostatistical regression method

Yaqian He[a, b, c], Yanchen Bo[a, b, *], Leilei Chai[a, b], Xiaolong Liu[d], Aihua Li[e]

[a] State Key Laboratory of Remote Sensing Science, Research Center for Remote Sensing and GIS, and School of Geography, Beijing Normal University, Beijing 100875, China

[b] Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, Beijing 100875, China

[c] Department of Geology and Geography, West Virginia University, Morgantown, WV 26506, USA

[d] College of Tourism & Geography Science, Yunnan Normal University, Kunming, Yunnan Province 650500, China

[e] Department of Geoscience, Boise State University, Boise, ID 83725, USA

* Corresponding Author:

Dr. Yanchen Bo
State Key Laboratory of Remote Sensing Science, Research Center for Remote Sensing and GIS, and School of Geography, Beijing Normal University, Beijing 100875, China
Tel.: +86-10-58802062
Fax: +86-10-58805274
E-Mail: boyc@bnu.edu.cn

Abstract: Leaf Area Index (LAI) is an important parameter of vegetation structure. A number of moderate resolution LAI products have been produced in urgent need of large scale vegetation monitoring. High resolution LAI reference maps are necessary to validate these LAI products. This study used a geostatistical regression (GR) method to estimate LAI reference maps by linking in situ LAI and Landsat TM/ETM+ and SPOT-HRV data over two cropland and two grassland sites. To explore the discrepancies of employing different vegetation indices (VIs) on estimating LAI reference maps, this study established the GR models for different VIs, including difference vegetation index (DVI), normalized difference vegetation index (NDVI), and ratio vegetation index (RVI). To further assess the performance of the GR model, the results from the GR and Reduced Major Axis (RMA) models were compared. The results show that the performance of the GR model varies between the cropland and grassland sites. At the cropland sites, the GR model based on DVI provides the best estimation, while at the grassland sites, the GR model based on DVI performs poorly. Compared to the RMA model, the GR model improves the accuracy of reference LAI maps in terms of root mean square errors (RMSE) and bias.

Keywords: Leaf Area Index; Up-scaling; Geostatistical Regression; Reduced Major Axis; Vegetation Index

1. Introduction

Leaf Area Index (LAI), defined as half the total leaf area per unit ground surface areas (Chen and Black, 1992), is an important parameter of vegetation structure and function (Abuelgasim et al., 2006). LAI provides substantial information on the exchange of energy, mass, and momentum flux between the Earth's surface and its atmosphere (Morisette et al., 2006; Myneni et al., 1997). LAI has been widely used as an input in climate, hydrology, and biogeochemistry models (Berterretche et al., 2005; Knyazikhin et al., 1998; Morisette et al., 2006). To date, a number of global and regional moderate-resolution LAI products have been produced, including Moderate Resolution Imaging Spectroradiometer (MODIS), Carbon Cycle and Change in Land Observational Products from and Ensemble of Satellites (CYCLOPES), Canada Centre for Remote Sensing (CCRS), and Global Land Surface Satellite (GLASS) (Chen et al., 2002; Tian et al., 2000; Weiss et al., 2007; Xiao et al., 2014). Owing to the influence of model algorithms, vegetation heterogeneity, and observation conditions, these LAI products inevitably have inherent uncertainties (Chen et al., 2002), which subsequently may impact the accuracy of any resulting modeling activities. Specifying the uncertainties of these coarse spatial resolution LAI products is essential for users to determine the most appropriate dataset for their applications, and for producers to improve methodological algorithms. However, a direct comparison between in situ LAI measurements and these corresponding moderate resolution LAI products is not recommended because of scale-mismatch, geolocation errors, and land surface heterogeneity (Huang et al., 2006; Yang et al., 2006). The proposed way to validate coarse resolution remote sensing products is using fine reference maps derived from up-scaling in situ measurements (Fernandes et al., 2014;

Iiames et al., 2015; Kang et al., 2015; Morisette et al., 2006; Wang et al., 2014). Previous studies have generated fine resolution LAI reference maps through fusing in situ LAI measurements and fine resolution remote sensing images (e.g. TM, ETM+, ASTER, SPOT) (Baret et al., 2005; Chen et al., 2002; Cohen and Justice, 1999; Garrigues et al., 2008; Li et al., 2013a; Martinez et al., 2009; Morisette et al., 2006; Pisek and Chen, 2007).

There are three categories of methods for estimating reference LAI maps using in situ LAI observations and fine spatial resolution remote sensing data, including regression, vegetation radiation transfer equation inversion, and geostatistical methods (Cohen et al., 2003; Martinez et al., 2010; Yang et al., 2006). Of these, the radiation transfer equation inversion method is not used widely due to the difficulty in collecting certain model parameters (e.g. canopy structure) and the fact that the solution of the model is not unique (Yang et al., 2006). Geostatistical methods have become popular in linking field data to image data, and been applied to estimate forest parameters (basal area, height, health conditions, etc), detect land use and land cover change, and map vegetation index (e.g., normalized difference vegetation index: NDVI and LAI) (Van der Meer, 2012). Traditional geostatistical methods, such as Kriging, predict unknown points through spatially interpolating surrounding field observations (Berterretche et al., 2005; Li et al., 2013a; Li et al., 2013b). The limited number of field observations and the spatial non-stationarity of in situ observations distribution could lead to uncertainty of predicting results. Regression methods, such as ordinary least squares regression, attempt to improve the predicting accuracy through accounting for high resolution remote sensing data (e.g., reflectance or vegetation indices (VI) derived from Landsat ETM+). Cohen et al. (2003) compared three regression methods (i.e., traditional ordinary least squares regression, inverse ordinary least square regression, and reduced major axis: RMA) over the BigFoot AGRO and NOBS sites. They reported that the performance of RMA method was superior to the other two. However, none of the regression methods consider the spatial/temporal correlation of in situ observations and high resolution reflectance or VI data, which may lead to an underestimation of the uncertainty along with the regression coefficients (Chatfield, 2003).

Geostatistical regression (GR) method conserves merits from both traditional geostatistical methods and regression methods. It has been used in examining the relationships between terrestrial carbon dioxide flux and its primary environmental drivers (Mueller et al., 2010), and estimating snow cover and gross primary productivity (Erickson et al., 2005; Yadav et al., 2010). Compared to traditional regression methods, the GR method is improved in one distinct way, which is the ability to account for the spatial/temporal correlation of the residuals from in situ observations (such as field LAI measurements) and auxiliary data (such as NDVI) (Erickson et al., 2005; Mueller et al., 2010; Yadav et al., 2010). Unlike traditional geostatistcal methods (e.g., Kriging), the GR method attempts to provide better estimating of unknown points by exploring the correlation between high resolution remote sensing data and field observations. To our knowledge, no attempts have been made to use the GR method to estimate LAI reference maps. This study applied the GR method to estimate high resolution LAI reference maps over cropland and grassland sites through fusing in situ LAI measurements and high resolution remote sensing images (i.e., Landsat TM/ETM+ and SPOT). To investigate the discrepancy of employing different VIs on estimating LAI reference maps, this study established the GR models for the following VIs: difference vegetation index (DVI), NDVI, and ratio vegetation index (RVI). To robustly assess the performance of the GR model, the results from GR and RMA models were compared.

2. Methodology

2.1. Geostatistical regression method

The GR method not only models the relationships between auxiliary variables (DVI, NDVI, and RVI in this study) and field measurements (in situ LAI measurements in this study), but also accounts for the spatial/temporal correlation of the regression residuals (Erickson et al., 2005). As with the linear regression method, the GR method decomposes LAI into a deterministic and a stochastic component:

$$LAI = X\beta + \varepsilon \quad (1)$$

Where $X(n \times P)$ is the DVI, NDVI, and RVI, respectively, $\beta(P \times 1)$ is the corresponding regression coefficient, and $\varepsilon(n \times 1)$ is assumed to be second-order stationary and zero-mean residuals for DVI, NDVI, and RVI (Leung and Cooley, 2014; Mueller et al., 2010; Yadav et al., 2010). Unlike the traditional linear regression approach, which regards ε as white noise, the GR method uses spatial covariance to recognize the spatial autocorrelation structure of the regression residuals ε. The experimental covariance of residuals ε for DVI, NDVI and RVI, respectively, is:

$$Q(h) = E(\varepsilon(X)\varepsilon(X + h)) \quad (2)$$

Where h is the spatial and/ or temporal distance, $Q(h)$ is the covariance of residual at separation distance h (Erickson et al., 2005). Many theoretical covariance functions (such as nugget, exponential, spherical, and Gaussian functions) can be used to model the experimental covariance (Schabenberger and Pierce, 2001). In this study, a linear combination of nugget and exponential functions is used following the previous studies (Erickson et al., 2005; Li et al., 2013a; Mueller et al., 2010). This function is defined as:

$$Q(h) = \begin{cases} \sigma_N^2 + \sigma_S^2, h = 0 \\ \sigma_S^2 \exp\left(-\frac{h}{l}\right), h > 0 \end{cases} \tag{3}$$

σ_N^2 is the measurement error or the variability at small scale that is uncorrelated in space and/or time, σ_S^2 is the variance of the variability correlated in space and/or time, and l is the correlation range parameters (Leung and Cooley, 2014). The Restricted Maximum Likelihood (RML), which maximizes the marginal distribution of the covariance function parameters, is used to estimate the parameters (σ_N, σ_s, l) (Kitanidis and Shen, 1996).

The best linear unbiased estimator of β on the basis of Aitken (1935) is the generalized-least-squares estimator, that is, the value of β that minimizes $(LAI - X\beta)^T Q^{-1}(LAI - X\beta)$. Thus,

$$\hat{\beta} = (X^T Q^{-1} X)^{-1} X^T Q^{-1} LAI \tag{4}$$

2.2. Reduced major axis method

To robustly assess the performance of the GR model, we compare the results from GR and RMA models. We choose RMA method because it is regarded as the 'standard' method for estimating LAI reference map in BigFoot project (Berterretche et al., 2005; Cohen et al., 2003), which is a well known project linking in situ measurements, remote sensing and models to validate MODIS products including LAI product. The form of RMA is identical to a simple linear regression method:

$$LAI = \beta_0 + \beta_1 X + \varepsilon \quad (5)$$

Where X is DIV, NDVI, and RVI, respectively. ε is white noise residual.

RMA method is superior to traditional ordinary least squares regression when both dependent (LAI in this study) and independent variables (DVI, NDVI, and RVI in this study) are measured with errors (Cohen et al., 2003; Smith, 2009). The estimating of β_0 and β_1 is different with the traditional ordinary least square regression. The traditional ordinary least square regression estimates the regression coefficients by minimizing the sum of squares of the residuals, while RMA minimizes the areas of triangles formed by the deviation of a point from the regression line in both horizontal and vertical directions (Smith, 2009). The equations for calculating β_0 and β_1 are $\beta_0 = \overline{LAI} - \frac{\sigma_Y}{\sigma_X}\bar{X}$ and $\beta_1 = \frac{\sigma_Y}{\sigma_X}$.

3. Data

3.1. Study Sites

Two cropland sites (AGRO and Plan-de-dieu sites) and two grassland sites (Hulun Buir and Zhangbei sites) were used in this study. The AGRO site is from the BigFoot project (http://www.fsl.orst.edu/larse/bigfoot/index.html), which is funded by NASA'S Terrestrial Ecology Program (Morisette et al., 2006; Pisek and Chen, 2007). Nine validation sites are in the BigFoot project with each of them covering a 5 km × 5 km extent (Morisette et al., 2006). The field LAI values in the AGRO site were measured by the allometric destructive method. The Hulun Buir site is one of the validation sites for the GLASS LAI product, which is a newly released LAI product generated by Beijing Normal University, China (Liang et al., 2014). The coverage of the Hulun Buir site is about 32 km × 28 km. The in situ LAI values in the Hulun Buir were measured by LAI-

2000. The Plan-de-dieu and Zhangbei sites are from the VALERI project (http://w3.avignon.inra.fr/valeri/), which has served to provide high spatial resolution maps of biophysical variables (e.g., LAI, fAPAR, fCover) to validate products derived from satellite observations (e.g., VEGETATION, MERIS, POLDER, AVHRR, and MODIS) (Baret et al., 2005). The VALERI project has 33 sites, each of them covering around 3 km × 3 km. The in situ LAI values in the VALERI project were measured by LAI-2000 or hemispherical images.

The AGRO site is located in Bondville, Illinois, USA. The main crop types of the AGRO site are corn and soybean (Pisek and Chen, 2007). The Plan-de-dieu site, with its main crop being vineyards, is located at Cotes du Rhone Village, France (Rossello, 2007). The Hulun Buir and Zhangbei grassland sites are located in Inner Mongolia and Hebei, China, respectively. The Landsat TM/ETM+ for AGRO and Hulun Buir sites were employed in this study as high resolution remote sensing images, because they are commonly used in up-scaling field measurements (Berterretche et al., 2005; Cohen et al., 2003) and could be easily obtained. We chose SPOT-HRV for Plan-de-dieu and Zhangbei sites because the Landsat TM/ETM+ corresponded to the date of in situ LAI in these two sites has gaps and does not have good quality, while SPOT-HRV images have been collected for many sites in VALERI project including Plan-de-dieu and Zhangbei sites (Baret et al., 2005) . The in situ LAI, TM/ETM+, and HRV data on the exact same date were not available. Therefore the data on the closest dates were chosen. The detailed information of the four sites is described in Table 1. The locations of the four study sites and the corresponding distribution of the in situ LAI locations in each site are shown in Figure 1.

Table 1. Information of the four study sites.

Sites	UTM X Coord	UTM Y Coord	UTM Zone	Location	Vegetation types	Datasets used	Datasets when obtained	In situ LAI measurement method
AGRO	389764	4429295	16N	Illinois USA	Corn and Soybean	Field LAI; ETM+; Land cover	7/24/2000 7/15/2000 2000	Allometric destructive means
Plan-de-dieu	655669	4895787	31N	Cotes du Rhone Village France	Vineyards	Field LAI; SPOT;	7/05-7/09/2004 6/29/2004	Hemispherical images
Hulun Buir	717675	5473425	50N	Inner Mongolia China	Grassland	Field LAI; TM; Land cover	6/26/2010 6/21/2010 2010	LAI-2000
Zhangbei	306354	4572278	50N	Hebei China	Grassland	Field LAI; SPOT;	8/08/-8/10/2002 8/23/2002	Hemispherical images

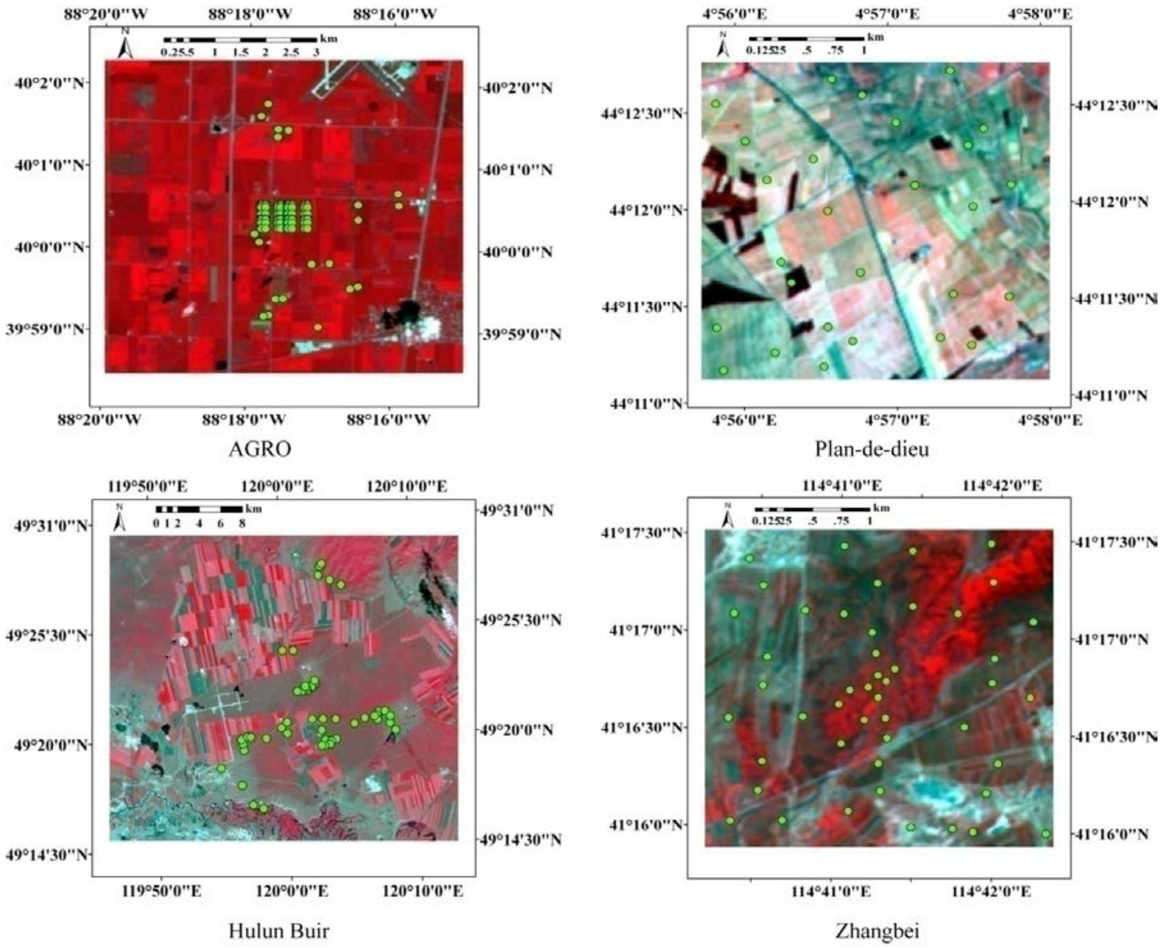


Figure 1. Study sites of the AGRO, Plan-de-dieu, Hulun Buir, and Zhangbei (the background is the standard false color composited image, and the green points are the in situ LAI locations).

3.2. Data pre-processing

Landsat TM/ETM+ data with 30 m spatial resolution used in this study were downloaded from the USGS website (http://glovis.usgs.gov/). The TM/ETM+ data are Level 1T data that have been systematically, radiometrically, and geometrically corrected. A large proportion of images are contaminated due to the influence of aerosols, clouds, and cloud shadows (Liang et al., 2001). The TM/ETM+ data were atmospherically corrected by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al., 2006). The two study areas, the AGRO and Hulun Buir sites, were extracted using ENVI software (Figure 1). The SPORT-HRV data with a spatial resolution of 20 m over the Plan-de-dieu and Zhangbei sites were obtained from the VALERI project (see the link in 3.1). Though they were geometrically corrected, no atmospheric corrections were applied to the images since no atmospheric data were available (Rossello, 2007, 2008). Rossello (2007) stated that atmospheric effects were assumed to be the same over the whole 3 km × 3 km extent, since the SPOT images were used to compute empirical relationships between reflectance and biophysical variables. The biophysical variables in the VALERI project over most of the 33 sites were based on SPOT-HRV top of atmosphere (TOA) reflectance (Baret et al., 2005). Following previous studies, this study also used the SPOT-HRV TOA reflectance to obtain the LAI values over the Plan-de-dieu and Zhangbei sites.

To evaluate the impacts of different vegetation indices on the GR and RMA models, this study employed DVI, NDVI, and RVI. The forms of these vegetation indices are: (Colombo et al., 2003; Huete et al., 2002).

$$DVI = NIR - R \tag{6}$$

$$NDVI = (NIR - R) / (NIR + R) \tag{7}$$

$$RVI = NIR / R \tag{8}$$

NIR is reflectance of near infrared band and *R* is reflectance of red band.

The scatter plots of DVI, NDVI, and RVI with the in situ LAI measurements at the four study sites are shown in Figure 2. At the AGRO site, DVI, NDVI, and RVI of the corn and soybean crop types have apparent boundaries. This study thus established the GR and RMA models for these two crop types, respectively. The land cover data from the BigFoot project was used to distinguish the corn and soybean over the AGRO site

(Table 1). As the Hulun Buir covered around 896 km^2, which may include other types of vegetation (e.g., forest), the land cover data used in this study to mask the non-grassland regions was provided by Tsinghua University (Table 1), China (Gong et al., 2013; Yu et al., 2013).

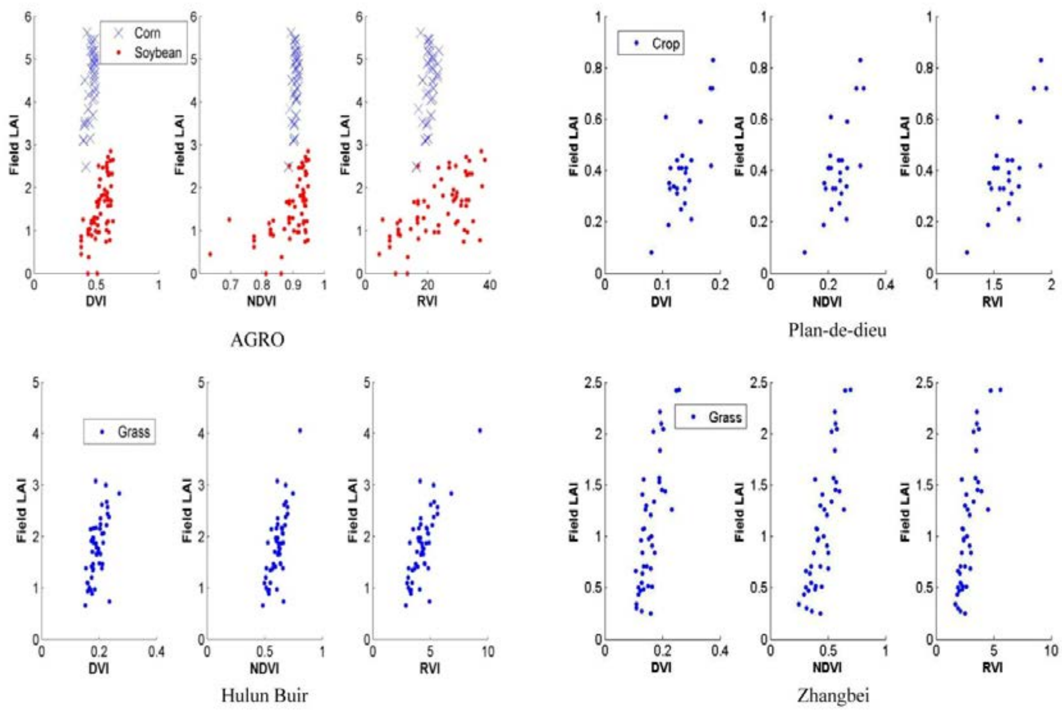


Figure 2. The scatter plots of DVI, NDVI, and RVI with the in situ LAI at the AGRO, Plan-de-dieu, Hulun Buir, and Zhangbei sites.

The total in situ LAI measurements for the AGRO, Plan-de-dieu, Hulun Buir, and Zhangbei sites are 98, 26, 51, and 42, respectively. This study randomly selected around 65% of the LAI points to establish and specify the GR and RMA models. The 35% of the LAI points were used to validate the results. This was repeated another five times for the GR models, in order to cross validate the robustness of performance of the models.

4. Results and Discussion

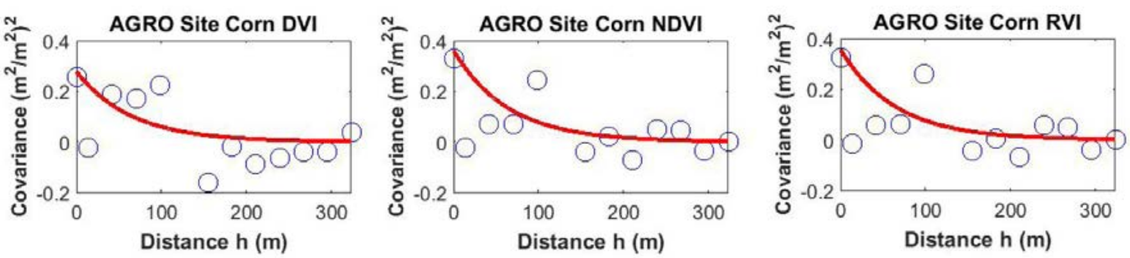
4.1. Spatial covariance models

As stated in 2.1, the residuals for Equation (1) were assumed to be second-order stationary with zero-mean, we calculated the experimental isotropic covariance of the residuals using least square method (Li et al., 2013b). The experimental covariances were modeled with exponential functions. The parameters of exponential functions were obtained through RML method. Table 2 shows the parameters of exponential functions for different VIs at four sites, respectively. The experimental and modeled covariances are shown in Figure 3. The parameters of covariance function in the same site have very similar values, which indicate similar spatial structure happens in the same site no matter what the VI is. At different sites the parameters are quit different (Table 2), depending on the locations of in situ LAI measurements and associations between LAI and VIs in that site. In addition to nugget variance for DVI at AGRO (Corn) site, all of the nugget values are larger than zero, which may be due to the heterogeneous of LAI of sub-samples within each sample, since the in situ LAI value for each sample is calculated from sub-samples in that sample (Baret et al., 2005; BigFoot, 1999). For example, each in situ LAI sample plot in Zhangbei site covers around 20 m x 20 m. In each sample plot, 12 sub-samples are used to calculate the corresponding LAI value for that sample plot (Baret et al., 2005).

Table 2. Parameters of the covariance function

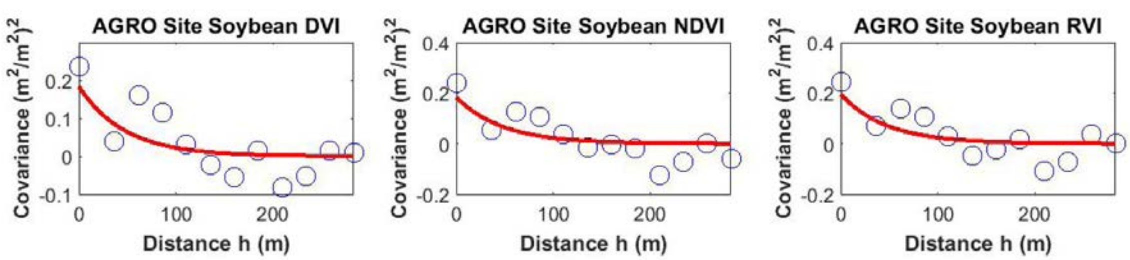
Site	**VIs**	σ_N	σ_s	l
AGRO (Corn)	DVI	0.000	0.278	193.862
	NDVI	0.099	0.361	193.862
	RVI	0.104	0.356	193.862

AGRO (Soybean)	DVI	0.069	0.184	142.228
	NDVI	0.067	0.184	142.228
	RVI	0.062	0.194	142.228
Plan-de-dieu	DVI	0.003	0.010	1505.988
	NDVI	0.003	0.011	1505.988
	RVI	0.003	0.011	1505.988
HulunBuir	DVI	0.053	0.212	2501.122
	NDVI	0.047	0.187	2501.122
	RVI	0.044	0.174	2501.122
Zhangbei	DVI	0.026	0.104	699.860
	NDVI	0.024	0.096	699.860
	RVI	0.022	0.088	699.860



(1) (2) (3)

AGRO (Corn)



(4) (5) (6)

AGRO (Soybean)

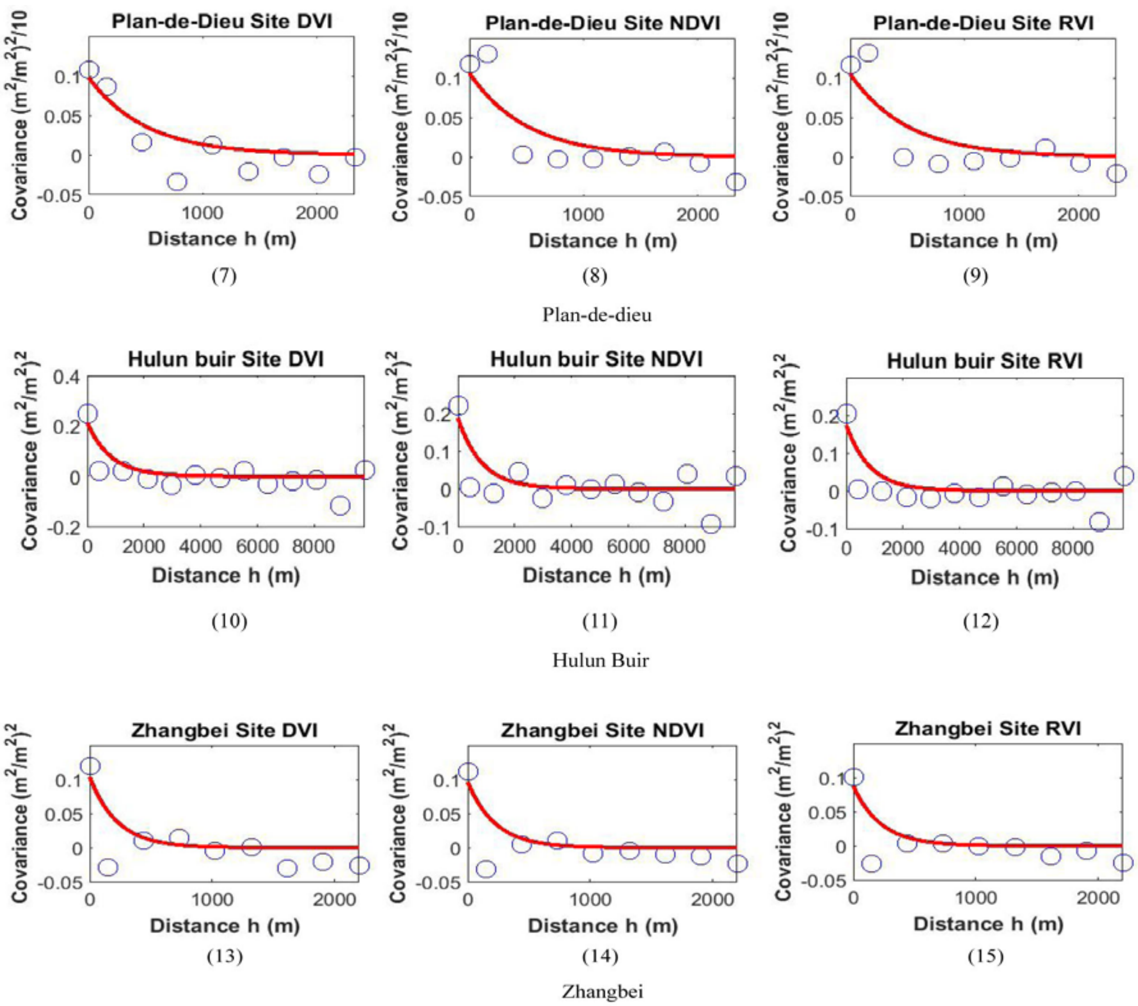


Figure 3. The experimental and modeled covariance (blue circle is experimental covariance, and red line is modeled covariance)

4.2. GR models for the four study sites

Table 3 shows the GR models for the AGRO, Plan-de-dieu, Hulun Buir, and Zhangbei sites. The values in parentheses are standard deviations for slope and intercept. The significance of slope and intercept are tested by Student's t test. Besides slopes for NDVI and RVI in the AGRO (corn) sites, all slopes are significant at 1% level, indicating the reliability of the models. The majority of intercepts are not significant at 1% level, excepting the intercepts in Zhangbei site. The insignificance may be due to small samples, such as the AGRO (corn) and Plan-de-dieu sites. The negative values of intercept may be attributed to the uncertainty of retrieving DIV, NDVI, and RVI from TM/ETM+ and HRV images, as there is no accurate atmosphere information for each sites, thereby the band reflectance from these images has errors. In addition, the in situ LAI values also have measurement errors. Therefore, the negative values of intercept are shown when conducting statistical analysis.

The coefficient of determination (R^2) varies among different models in different sites. At the AGRO site, the R^2 value for corn ranges from 0.28 to 0.44, and for soybean 0.38 to 0.40. The R^2 value of DVI model is the highest for the AGRO site compared to the R^2 values for NDVI and RVI models. As with the AGRO site, the R^2 value of DVI model in the Plan-de-dieu site is the highest. The R^2 value for the Hulun Buir and Zhangbei grassland sites ranges from 0.53 to 0.61, 0.63 to 0.69, respectively. In contrast to the cropland sites (i.e., the AGRO and Plan-de-dieu sites), the R^2 values of DVI models over the two grassland sites are the lowest. Excepting for Zhangbei site, the R^2 values are not high, which maybe because of the poor relationships between DVI, NDVI, and RVI and original in situ LAI values (Figure 2). However, the GR models with DVI perform best over the two cropland sites, while for the two grassland sites, the GR models with DVI have the poorest performance.

Table 3. GR models at the four study sites.

Site	**VIs**	**R^2**	**Slope**	**Intercept**
AGRO (Corn)	DVI	0.44	14.62** (4.02)	-2.29 (1.81)
	NDVI	0.28	23.56 (20.35)	-17.16 (18.40)
	RVI	0.29	0.10 (0.09)	2.20 (1.76)

AGRO (Soybean)	DVI	0.4	5.35** (1.09)	-1.23* (0.56)
	NDVI	0.4	8.55** (1.71)	-6.23** (1.55)
	RVI	0.38	0.05** (0.01)	0.41 (0.24)
Plan-de-dieu	DVI	0.57	4.47** (1.01)	-0.17 (0.14)
	NDVI	0.53	2.53** (0.61)	-0.16 (0.14)
	RVI	0.54	0.75** (0.18)	-0.78* (0.29)
HulunBuir	DVI	0.53	15.46** (2.62)	-1.25* (0.52)
	NDVI	0.58	8.41** (1.28)	-3.39** (0.79)
	RVI	0.61	0.50** (0.07)	-0.41 (0.32)
Zhangbei	DVI	0.63	13.82** (2.13)	-1.11** (0.34)
	NDVI	0.65	4.84** (0.70)	-1.11** (0.32)
	RVI	0.69	0.66** (0.09)	-0.75** (0.25)

* significant at 5% level, ** significant at 1% level.

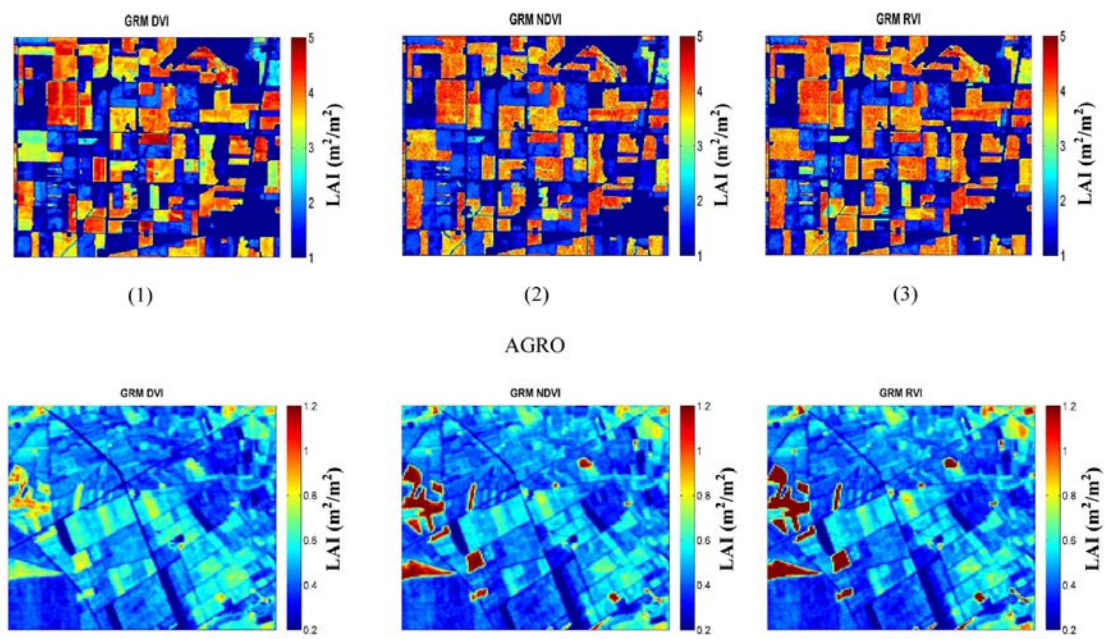
4.3. Estimating and validating the reference LAI maps based on GR models

Figure 4 presents the reference LAI maps estimated by the GR models based on Table 3. The validation results are shown in Figure 5 and Table 4. Most of the R^2 values in Table 4 are nearly equal to the R^2 values in Table 3, which indicates that the GR models are robust. However, some GR models may not be robust (e.g., GR model with DVI for corn at the AGRO site). We discuss the problem in detail at the end of this section. As mentioned in section 4.2, the low R^2 values for GR models at the AGRO, Plan-de-dieu, and Hulun Buir sites may be due to the poor relationships of DVI, NDVI, and RVI with in situ LAI observations. For example, there is one very low in situ LAI observation at the AGRO (corn) and Plan-die-dieu sites, and one very high in situ LAI observation at the Hulun Buir site. These abnormal in situ LAI observations may be owing to measurement errors. Regardless, the R^2 values show the same pattern as that in section 4.2. That is, in terms of R^2 values, the GR models with DVI have the best performance over the two cropland sites, while the GR models with DVI at the two grassland sites perform more poorly. The values of root mean square errors (RMSE) indicate that all the sites have the same trend within same vegetation types, excepting for the Plan-de-dieu site. The RMSE values are lowest for DVI at the AGRO site (0.88 for corn and 0.59 for soybean). This implies that the standard deviation of the differences between the estimated LAI based on DVI and the field LAI is lowest. However, at the Hunlun Buir and Zhangbei sites, the RMSE values are highest for DVI (0.40 and 0.46, respectively). In terms of bias, there are no clear common characteristics. For example, the value of absolute bias for the AGRO (corn) site is lowest based on DVI, while for the AGRO (soybean) site, the value of absolute bias is lowest based on RVI. In summary, the GR models based on DVI have the best estimations for the two cropland sites, while for the two grassland sites, the GR models based on DVI perform poorly.

Table 4. Statistics of estimated LAI of the GR and RMA models compared to the in situ LAI.

Site	VIs	R^2		RMSE		bias	
		GR	RMA	GR	RMA	GR	RMA
AGRO	DVI	0.23	0.23	0.88	0.89	0.05	0.10
	NDVI	0.18	0.18	0.94	1.10	-0.17	0.01

(Corn)	RVI	0.17	0.17	0.94	1.10	-0.16	0.01
AGRO (Soybean)	DVI	0.43	0.43	0.59	0.68	-0.15	-0.22
	NDVI	0.29	0.29	0.73	0.99	-0.22	-0.33
	RVI	0.38	0.38	0.60	0.68	-0.12	-0.16
Plan-de-dieu	DVI	0.52	0.52	0.16	0.17	0.10	-0.12
	NDVI	0.43	0.43	0.16	0.17	0.08	-0.10
	RVI	0.45	0.45	0.16	0.17	0.08	-0.10
HulunBuir	DVI	0.45	0.45	0.40	0.43	-0.11	-0.12
	NDVI	0.55	0.55	0.39	0.48	-0.14	-0.15
	RVI	0.56	0.56	0.38	0.43	-0.15	-0.16
Zhangbei	DVI	0.53	0.53	0.46	0.52	-0.02	0.02
	NDVI	0.67	0.67	0.38	0.42	-0.01	0.02
	RVI	0.63	0.63	0.43	0.50	0.05	0.10



(1) (2) (3)

AGRO

(4) (5) (6)

Plan-de-dieu

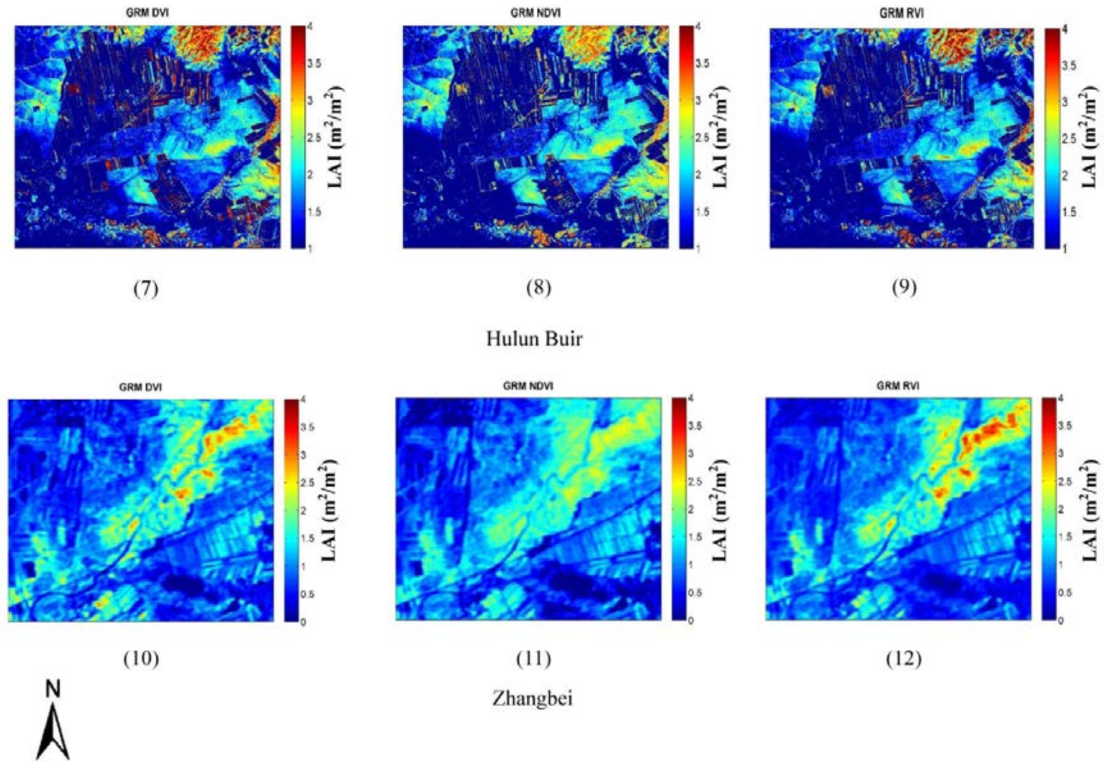
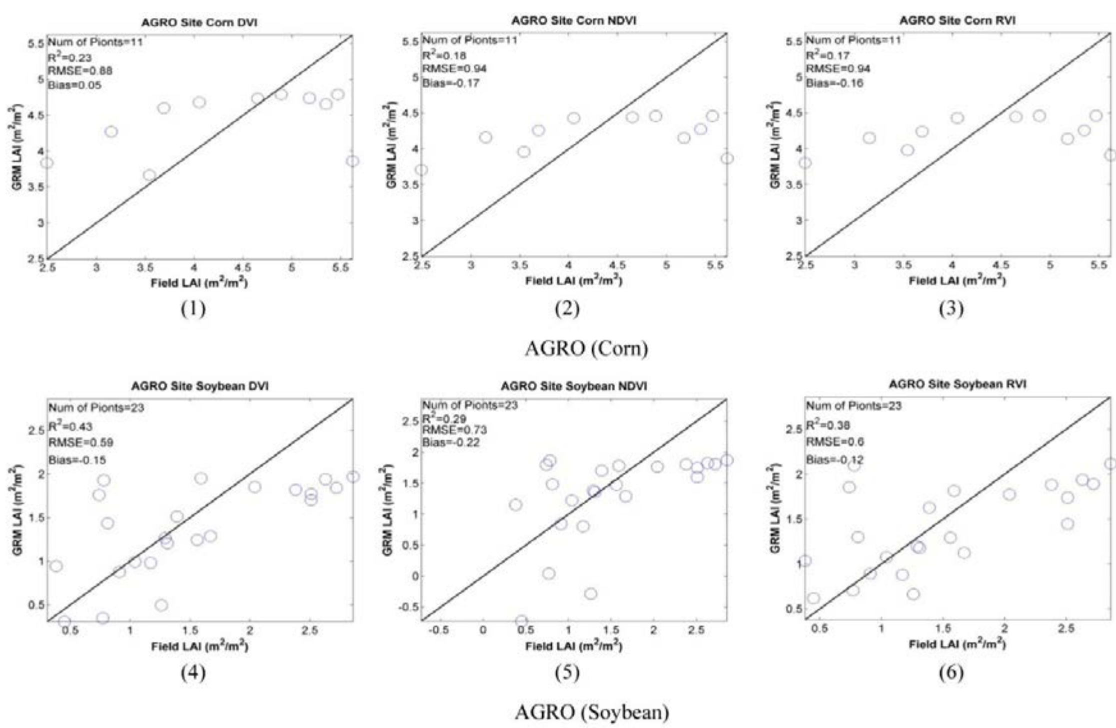


Figure 4. Reference LAI maps estimated by the GR models at the AGRO, Plan-de-dieu, Hulun Buir, and Zhangbei sites.



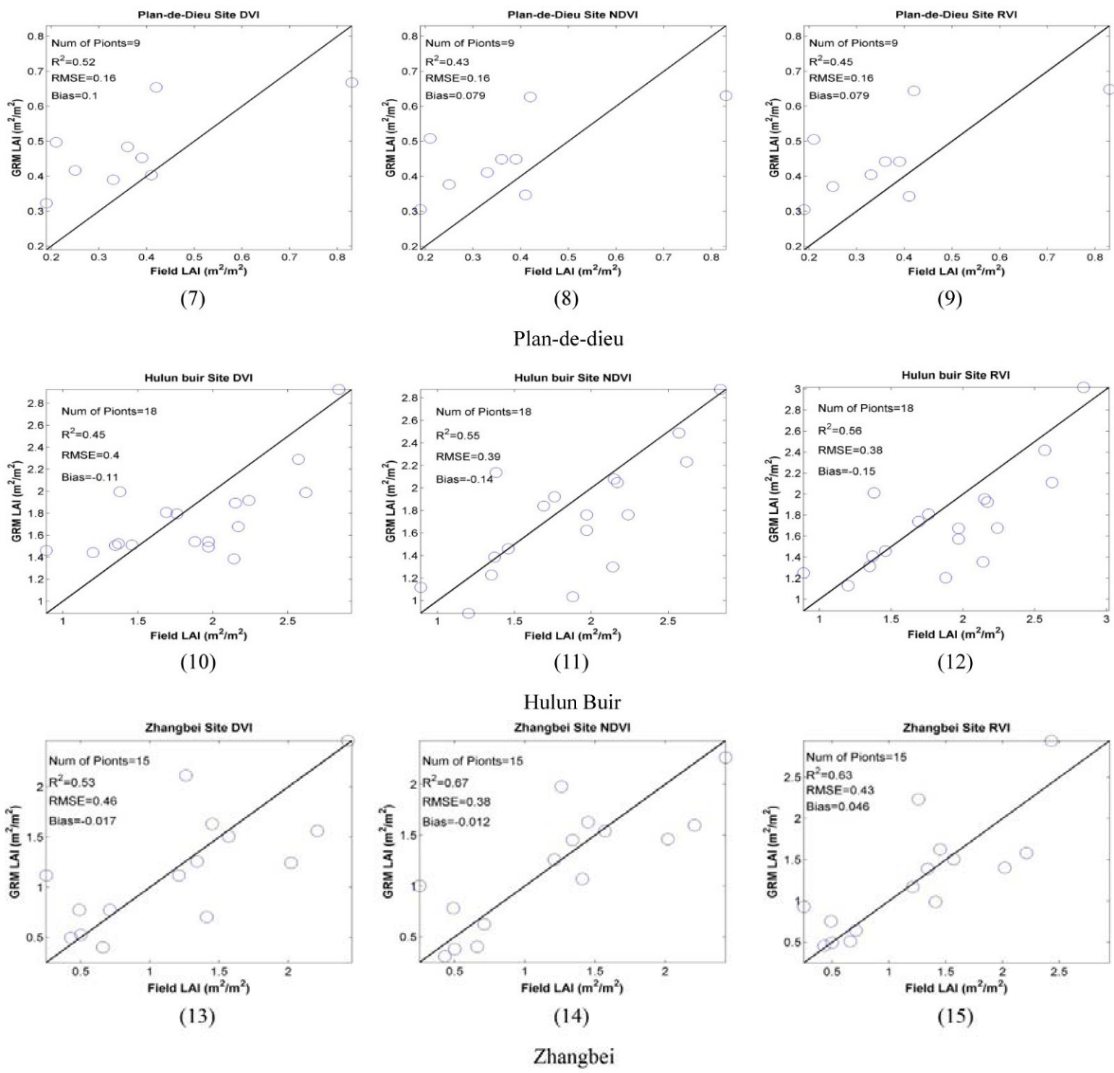


Figure 5. Validation results of the GR models at the AGRO, Plan-de-dieu, Hulun Buir, and Zhangbei sites.

In order to check the robustness of the predictive ability of the GR models, this study used cross validation. Considering the intensive computation of the GR models that involve spatial covariance modeling and geostatistical estimation, this study was repeated five times by randomly selecting 65% of the LAI points for establishing the GR models, with the remainder of the LAI points used for model validation. The mean RMSE values (μ_{RMSE}) of the five repetitions were calculated following previous studies Lee et al. (2008a, b). Figure 6 shows the results of cross validation. The blue bar is the μ_{RMSE} of the five repetitions for each GR model, the black error bar is $\mu_{RMSE} \pm \sigma_{RMSE}$ (σ_{RMSE} is the standard deviation) of the five repetitions for each GR model, and the brown square is the RMSE value from Table 4. In comparison to the μ_{RMSE} in Figure 6, most of the RMSE values in Table 4 are nearly within [$\mu_{RMSE} - \sigma_{RMSE}$, $\mu_{RMSE} + \sigma_{RMSE}$], which indicates that the GR models are robust. The RMSE value of the GR model for DVI at the AGRO (corn) site slightly exceeds the upper limits of the error bar ($\mu_{RMSE} + \sigma_{RMSE}$), which confirms that the GR model with DVI for corn at the AGRO site is not robust. This is presumably due to the poor association of DVI and the in situ LAI values (Figure 2). The RMSE values of the GR model for DVI and RVI at the Zhangbei site also exceed upper limits of the error bar, which may be due to the limited repetitions. More repetitions are needed for robust validation.

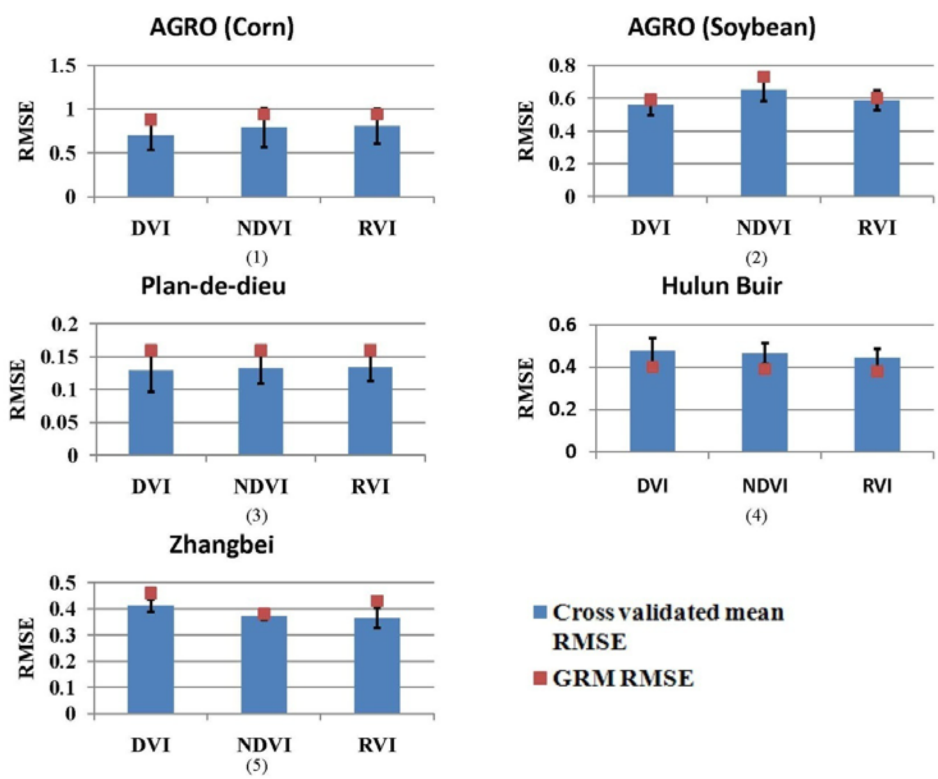


Figure 6. Cross validation for the GR models

4.4. Comparing the results of GR and RMA models

For robust assessment of the performance of the GR models, the results from the GR and RMA models were compared. Based on equation (5), the high resolution reference LAI maps estimated by the RMA model are depicted in Figure 7. The validation results are displayed in Figure 8 and Table 4. In terms of R^2, the GR models have identical values with the RMA models at the four study sites. The RMSE values for the GR models are lower than the RMA models for all of the sites, which may due to the consideration of spatial correlations of regression residuals. The GR models have lower biases than the RMA models, excluding the GR models with NDVI and RVI at the AGRO (corn) site. In summation, the GR models improve the accuracy of reference LAI maps compared to the RMA models.

In addition, the GR and RMA models had consistent performance at cropland and grassland sites. Both GR and RMA models have the best estimating ability based on DVI at the cropland sites (AGRO and Plan-de-dieu sites), while the GR and RMA models perform poorly based on DVI at the grassland sites (Hulun Buir and Zhangbei sites).

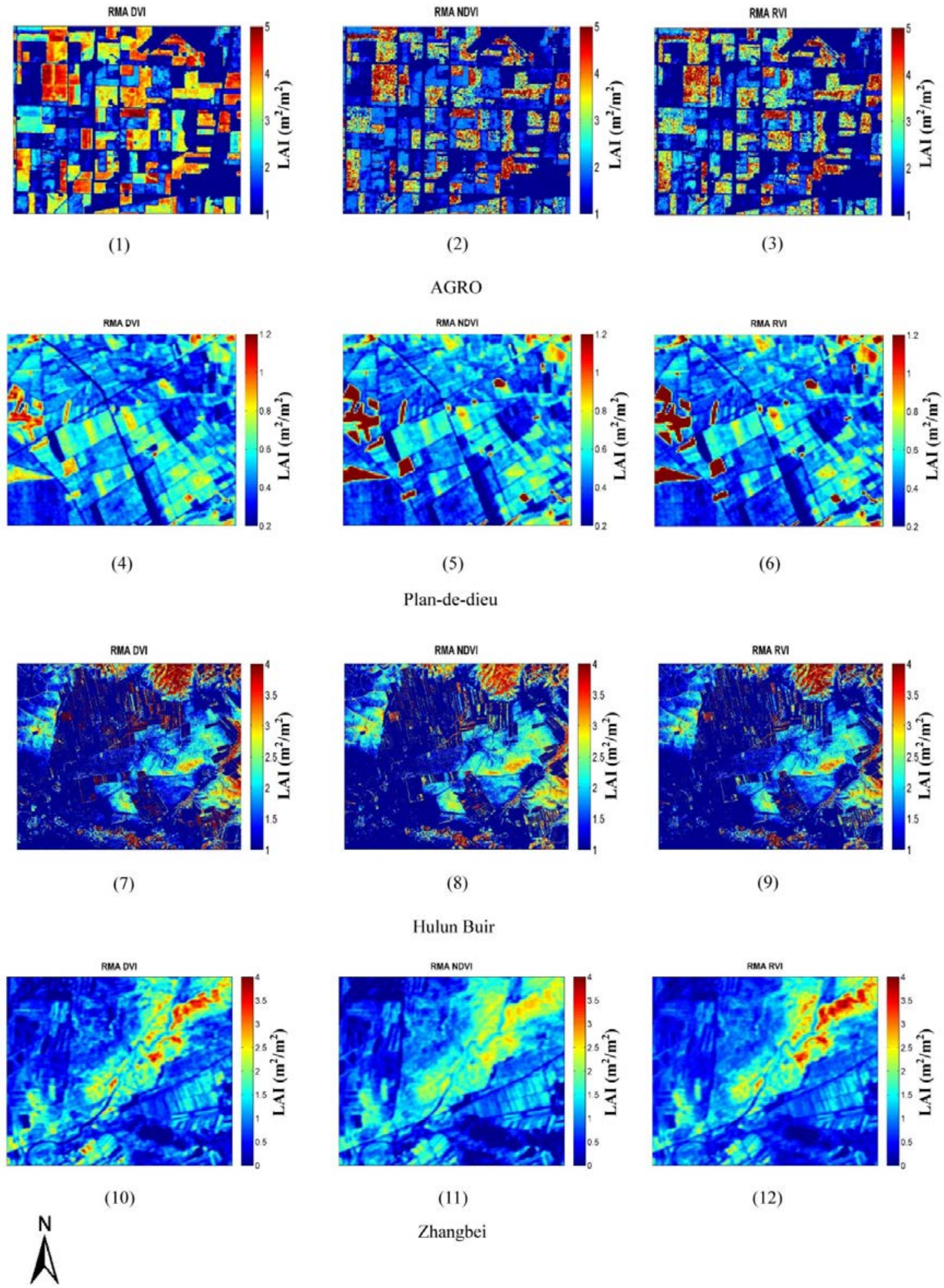
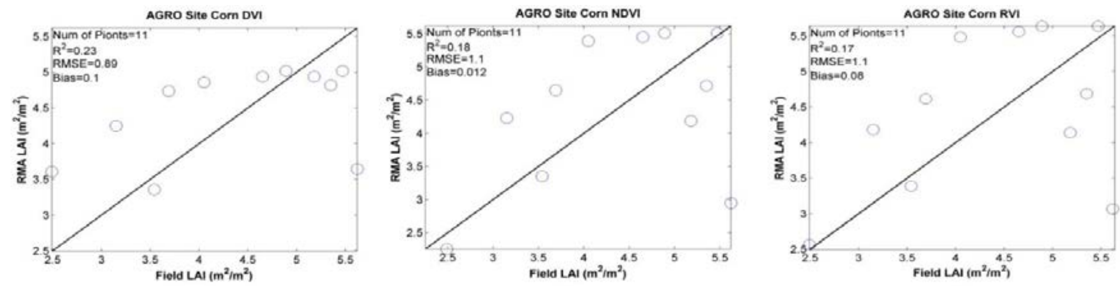
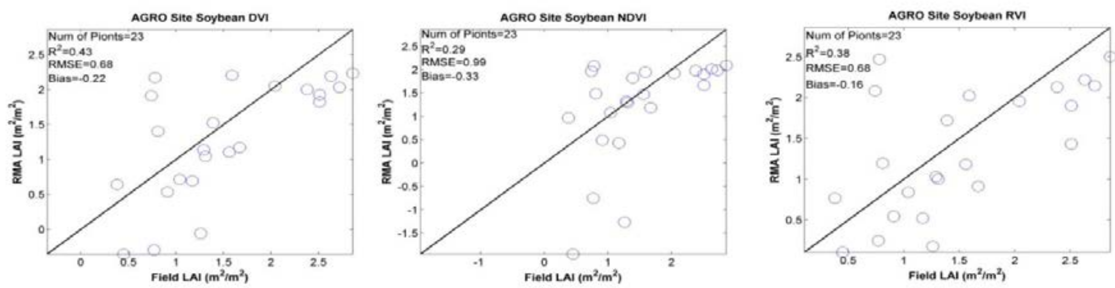


Figure 7. Reference LAI maps estimated by the RMA models at the AGRO, Plan-de-dieu, Hulun Buir, and Zhangbei sites.



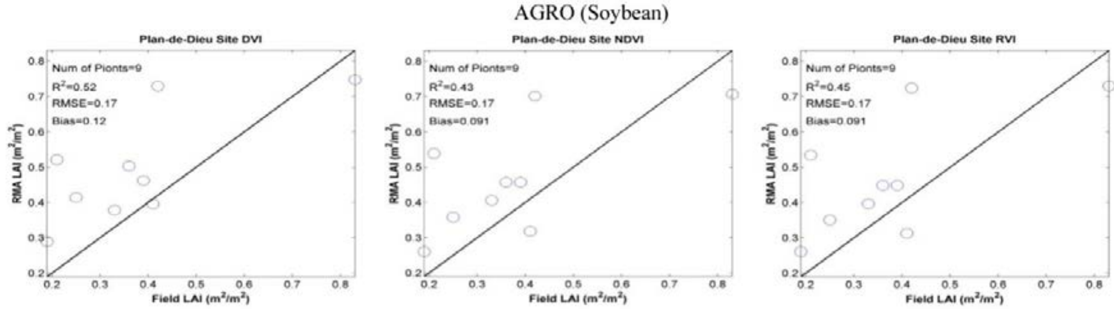
(1) (2) (3)

AGRO (Corn)



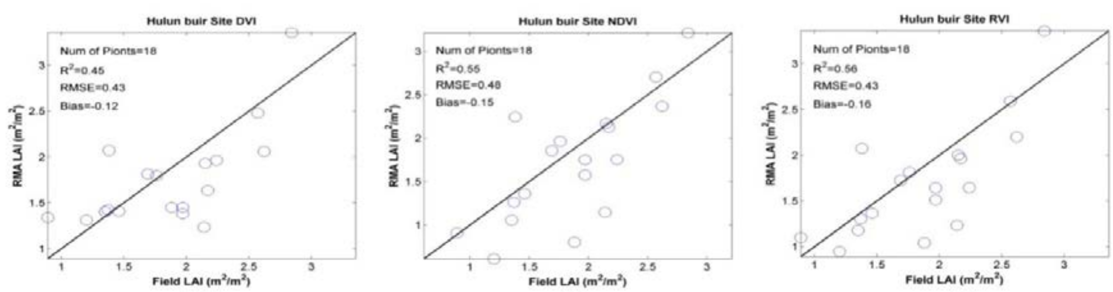
(4) (5) (6)

AGRO (Soybean)



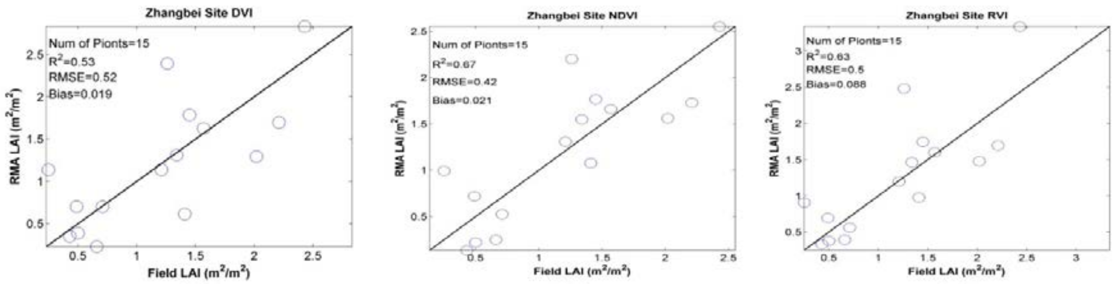
(7) (8) (9)

Plan-de-dieu



(10) (11) (12)

Hulun Buir



(13) (14) (15)

Zhangbei

Figure 8. Validation results of the RMA models at the AGRO, Plan-de-dieu, Hulun Buir, and Zhangbei sites.

5. Conclusions

Spatial scale issue commonly exits in remote sensing studies. Van der Meer et al. (2001) explored spatial scale effects on vegetation indices estimation through calculating vegetation indices, including NDVI, perpendicular vegetation index, weighted difference vegetation index, etc., from the Medium Resolution Imaging Spectrometer (MERIS) at the spatial resolutions ranging from 6 to 300 m. The proposed way to validate coarse resolution remote sensing products is using fine reference maps derived from up-scaling in situ measurements. This study up-scaled the field LAI measurements to high resolution LAI reference map through linking the in situ LAI measurements and Landsat TM/ETM+ and SPOT-HRV data using the geostatistical regression method. To analyze the discrepancy of employing different vegetation indices on estimating LAI reference maps, this study established the GR models for DVI, NDVI and RVI. To further assess the performance of the GR model, this study compared the results from GR and RMA models. The results show that the performances of GR models over the cropland and grassland sites are different. The GR models based on DVI provide the best estimation at the cropland sites (AGRO and Plan-de-dieu sites), while the GR models perform poorly based on DVI at the grassland sites (Hulun Buir and Zhangbei sites). By considering the spatial/temporal correlations of in situ LAI observations and high resolution DVI, NDVI, and RVI data, this study reveals that the performance of the GR models is better than the RMA models in terms of RMSE and bias.

In summary, the GR method inherits the merits from both traditional geostatistical methods and regression methods. Compared to regression methods (e.g., RMA), the GR method is improved in accounting for the spatial/temporal correlation of residuals from the regressions of LAI observations and high resolution remote sensing data (e.g., DVI, NDVI and RVI data in this study). In contrast to traditional geostatistcal methods (e.g., Kriging), the GR method attempts to provide better estimating of unknown points by exploring the association between high resolution remote sensing data and field observations. Our study confirmed the performance of the GR models is better than the RMA models in terms of RMSE and bias, which indicates the potential of GR method to up-scale other in situ biophysical and geophysical measurements (e.g., fAPAR and soil moisture) to high resolution reference data to validate other coarse resolution products.

Acknowledgments

This work was jointly supported by the National Key Basic Research Program of China (973 Program) Project (Grant No. 2013CB733403), the National Natural Science Foundation of China (Grant No. 41271347 and 91125004), and the High-Tech Research and Development Program of China (863 Program) Project (Grant No. 2012AA12A305). We thank Dr. Gaolong Zhu from the Department of Geography at Minjiang University, BigFoot project, VALERI project, and USGS for providing in situ LAI measurements, SPOT-HRV data, and Landsat TM/ETM+ data. We are grateful to Dr. Anna Michalak's team in the Department of Global Ecology, the Carnegie Institution for Science for providing original GR Matlab code. We appreciate Matthew Purtill and Jothiganesh Shanmugasundaram in the Department of Geology and Geography, West Virginia University for their helpful comments in improving the English language. We also appreciate two anonymous referees for their valuable suggestions to make the manuscript better.

References

Abuelgasim, A.A., Fernandes, R.A., Leblanc, S.G., 2006. Evaluation of national and global LAI products derived from optical remote sensing instruments over Canada. IEEE Transactions on Geoscience and Remote Sensing 44, 1872-1884.

Aitken, A.C., 1935. On least squares and linear combinations of observations. Proceedings of the Royal Society of Edinburgh 55, 42-48.

Baret, F., Weiss, M., Allard, D., Garrigues, S., Leroy, M., Jeanjean, H., Fernandes, R., Myneni, R., Privette, J., Morisette, J., 2005. VALERI: a network of sites and a methodology for the validation of medium spatial resolution land satellite products. Remote Sensing of Environment, in press.

Berterretche, M., Hudak, A.T., Cohen, W.B., Maiersperger, T.K., Gower, S.T., Dungan, J., 2005. Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest. Remote Sensing of Environment 96, 49-61.

BigFoot, 1999. Linking In Situ Measurements, Remote Sensing, and Models to Validate MODIS Products Related to the Terrestrial Carbon Cycle. http://www.fsl.orst.edu/larse/bigfoot/ovr_dsgn.html (accessed 01.02.2016).

Chatfield, C., 2003. The analysis of time series: an introduction. CRC press, Boca Raton.

Chen, J.M., Black, T., 1992. Defining leaf area index for non‐ flat leaves. Plant, Cell & Environment 15, 421-429.

Chen, J.M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S., White, H., Hall, R., Peddle, D., King, D., Trofymow, J., 2002. Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. Remote Sensing of environment 80, 165-184.

Cohen, W.B., Justice, C.O., 1999. Validating MODIS terrestrial ecology products: linking in situ and satellite measurements. Remote Sensing of Environment 70, 1-3.

Cohen, W.B., Maiersperger, T.K., Gower, S.T., Turner, D.P., 2003. An improved strategy for regression of biophysical variables and Landsat ETM+ data. Remote Sensing of Environment 84, 561-571.

Colombo, R., Bellingeri, D., Fasolini, D., Marino, C.M., 2003. Retrieval of leaf area index in different vegetation types using high resolution satellite data. Remote Sensing of environment 86, 120-131.

Erickson, T.A., Williams, M.W., Winstral, A., 2005. Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States. Water Resources Research 41, W04014.

Fernandes, R., Plummer, S., Nightingale, J., Baret, F., Camacho, F., Fang, H., Garrigues, S., Gobron, N., Lang, M., Lacaze, R., LeBlanc, S., Meroni, M., Martinez, B., Nilson, T., Pinty, B., Pisek, J., Sonnentag, O., Verger, A., Welles, J., Weiss, M., Widlowski, J.L., 2014. Global Leaf Area Index Product Validation Good Practices, in: Schaepman-Strub, G., Román, M., Nickeson, J. (Eds.), Best Practice for Satellite-Derived Land Product Validation. Land Product Validation Subgroup (WGCV/CEOS).

Garrigues, S., Lacaze, R., Baret, F., Morisette, J., Weiss, M., Nickeson, J., Fernandes, R., Plummer, S., Shabanov, N., Myneni, R., 2008. Validation and intercomparison of global Leaf Area Index products derived from remote sensing data. Journal of Geophysical Research 113, G02028.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. International Journal of Remote Sensing 34, 2607-2654.

Huang, D., Yang, W., Tan, B., Rautiainen, M., Zhang, P., Hu, J., Shabanov, N.V., Linder, S., Knyazikhin, Y., Myneni, R.B., 2006. The importance of measurement errors for deriving accurate reference leaf area index maps for validation of moderate-resolution satellite LAI products. IEEE Transactions on Geoscience and Remote Sensing 44, 1866-1871.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote sensing of environment 83, 195-213.

Iiames, J.S., Congalton, R.G., Lewis, T.E., Pilant, A.N., 2015. Uncertainty Analysis in the Creation of a Fine-Resolution Leaf Area Index (LAI) Reference Map for Validation of Moderate Resolution LAI Products. Remote Sensing 7, 1397-1421.

Kang, J., Jin, R., Li, X., 2015. Regression Kriging-Based Upscaling of Soil Moisture Measurements From a Wireless Sensor Network and Multiresource Remote Sensing Information Over Heterogeneous Cropland. IEEE Geoscience and Remote Sensing Letters 12, 92-96.

Kitanidis, P.K., Shen, K.-F., 1996. Geostatistical interpolation of chemical concentration. Advances in Water Resources 19, 369-378.

Knyazikhin, Y., Martonchik, J., Myneni, R., Diner, D., Running, S., 1998. Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. Journal of Geophysical Research 103, 32257-32276.

Lee, E., Chase, T.N., Rajagopalan, B., 2008a. Highly improved predictive skill in the forecasting of the East Asian summer monsoon. Water resources research 44.

Lee, E., Chase, T.N., Rajagopalan, B., 2008b. Seasonal forecasting of East Asian summer monsoon based on oceanic heat sources. International Journal of Climatology 28, 667-678.

Leung, S., Cooley, D., 2014. A comparison of a traditional geostatistical regression approach and a general Gaussian process approach for spatial prediction. Stat 3, 228-239.

Li, A., Bo, Y., Chen, L., 2013a. Bayesian maximum entropy data fusion of field-observed leaf area index (LAI) and Landsat Enhanced Thematic Mapper Plus-derived LAI. International Journal of Remote Sensing 34, 227-246.

Li, A., Bo, Y., Zhu, Y., Guo, P., Bi, J., He, Y., 2013b. Blending multi-resolution satellite sea surface temperature (SST) products using Bayesian maximum entropy method. Remote Sensing of Environment 135, 52-63.

Liang, S., Fang, H., Chen, M., 2001. Atmospheric correction of Landsat ETM+ land surface imagery. I. Methods. IEEE Transactions on Geoscience and Remote Sensing 39, 2490-2498.

Liang, S., Zhang, X., Xiao, Z., Cheng, J., Liu, Q., Zhao, X., 2014. Challenges and Prospects, Global LAnd Surface Satellite (GLASS) Products. Springer, Berlin.

Martinez, B., Cassiraga, E., Camacho, F., Garcia-Haro, J., 2010. Geostatistics for mapping leaf area index over a cropland landscape: Efficiency sampling assessment. Remote Sensing 2, 2584-2606.

Martinez, B., García-Haro, F., Camacho-de Coca, F., 2009. Derivation of high-resolution leaf area index maps in support of validation activities: Application to the cropland Barrax site. Agricultural and Forest Meteorology 149, 130-145.

Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., Lim, T.-K., 2006. A Landsat surface reflectance dataset for North America, 1990-2000. IEEE Geoscience and Remote Sensing Letters 3, 68-72.

Morisette, J.T., Baret, F., Privette, J.L., Myneni, R.B., Nickeson, J.E., Garrigues, S., Shabanov, N.V., Weiss, M., Fernandes, R.A., Leblanc, S.G., 2006. Validation of global moderate-resolution LAI products: A framework proposed within the CEOS land product validation subgroup. IEEE Transactions on Geoscience and Remote Sensing 44, 1804-1817.

Mueller, K.L., Yadav, V., Curtis, P.S., Vogel, C., Michalak, A.M., 2010. Attributing the variability of eddy‐ covariance CO2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest. Global Biogeochemical Cycles 24.

Myneni, R.B., Ramakrishna, R., Nemani, R., Running, S., 1997. Estimation of global leaf area index and absorbed PAR using radiative transfer models. IEEE Transactions on Geoscience and Remote Sensing 35, 1380-1393.

Pisek, J., Chen, J.M., 2007. Comparison and validation of MODIS and VEGETATION global LAI products over four BigFoot sites in North America. Remote Sensing of Environment 109, 81-94.

Rossello, P., 2007. Ground data processing & production of the level 1 high resolution maps VALERI 2004 plan-de-dieu site http://w3.avignon.inra.fr/valeri/Europe/France/PACA/Plan-de-dieu/2004/biomap/PlandeDieu2004FTReport.pdf (accessed 17.07.2015).

Rossello, P., 2008. Ground data processing & production of the level 1 high resolution maps VALERI 2002 zhangbei site http://w3.avignon.inra.fr/valeri/asie/chine/zhangbei/2002/biomap/ZhangBei2002FTReport.pdf (accessed 17.07.2015).

Schabenberger, O., Pierce, F.J., 2001. Statistical models for spatial data, in: Schabenberger, O., Pierce, F.J. (Eds), Contemporary statistical models for the plant and soil sciences. CRC press, Boca Raton, pp. 561-697.

Smith, R.J., 2009. Use and misuse of the reduced major axis for line‐ fitting. American Journal of Physical Anthropology 140, 476-486.

Tian, Y., Zhang, Y., Knyazikhin, Y., Myneni, R.B., Glassy, J.M., Dedieu, G., Running, S.W., 2000. Prototyping of MODIS LAI and FPAR algorithm with LASUR and LANDSAT data. IEEE Transactions on Geoscience and Remote Sensing 38, 2387-2401.

Van der Meer, F., 2012. Remote-sensing image analysis and geostatistics. International Journal of Remote Sensing 33, 5644-5676.

Van der Meer, F., Bakker, W., Scholte, K., Skidmore, A., De Jong, S., Clevers, J., Addink, E., Epema, G., 2001. Spatial scale variations in vegetation indices and

above-ground biomass estimates: implications for MERIS. International Journal of Remote Sensing 22, 3381-3396.
Wang, J., Ge, Y., Song, Y., Li, X., 2014. A Geostatistical Approach to Upscale Soil Moisture With Unequal Precision Observations. IEEE Geoscience and Remote Sensing Letters 11, 2125-2129.
Weiss, M., Baret, F., Garrigues, S., Lacaze, R., 2007. LAI and fAPAR CYCLOPES global products derived from VEGETATION. Part 2: validation and comparison with MODIS collection 4 products. Remote sensing of Environment 110, 317-331.
Xiao, Z., Liang, S., Wang, J., Chen, P., Yin, X., Zhang, L., Song, J., 2014. Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance. IEEE Transactions on Geoscience and Remote Sensing 52, 209-223.
Yadav, V., Mueller, K., Dragoni, D., Michalak, A., 2010. A geostatistical synthesis study of factors affecting gross primary productivity in various ecosystems of North America. Biogeosciences 7, 2655-2671.
Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N.V., Wang, Y., Privette, J.L., Huemmrich, K.F., Fensholt, R., Sandholt, I., 2006. MODIS leaf area index products: From validation to algorithm improvement. IEEE Transactions on Geoscience and Remote Sensing 44, 1885-1898.
Yu, L., Wang, J., Clinton, N., Xin, Q., Zhong, L., Chen, Y., Gong, P., 2013. FROM-GC: 30 m global cropland extent derived through multisource data integration. International Journal of Digital Earth 6, 521-533.