

## Research Article

# Aboveground Forest Biomass Estimation with Landsat and LiDAR Data and Uncertainty Analysis of the Estimates

Dengsheng Lu,<sup>1</sup> Qi Chen,<sup>2</sup> Guangxing Wang,<sup>3</sup> Emilio Moran,<sup>1</sup> Mateus Batistella,<sup>4</sup> Maozhen Zhang,<sup>5</sup> Gaia Vaglio Laurin,<sup>6</sup> and David Saah<sup>7</sup>

<sup>1</sup>Anthropological Center for Training and Research on Global Environmental Change (ACT), Indiana University, Bloomington, IN 47405, USA

<sup>2</sup>Department of Geography, University of Hawaii at Manoa, 2424 Maile Way, Honolulu, HI 96822, USA

<sup>3</sup>Department of Geography and Environmental Resources, Southern Illinois University at Carbondale (SIUC), Carbondale, IL 62901, USA

<sup>4</sup>Embrapa Satellite Monitoring, Fazenda Chapadão, 13070-115 Campinas, SP, Brazil

<sup>5</sup>School of Environment and Resources, Zhejiang A&F University, Lin'An 311300, China

<sup>6</sup>Department of Computer, System and Production Engineering, University of Tor Vergata, 00133 Rome, Italy

<sup>7</sup>Spatial Informatics Group, LLC, 3248 Northampton Ct., Pleasanton, CA 94588, USA

Correspondence should be addressed to Dengsheng Lu, dlu@indiana.edu

Received 1 October 2011; Revised 6 January 2012; Accepted 20 January 2012

Academic Editor: Jingxin Wang

Copyright © 2012 Dengsheng Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Landsat Thematic mapper (TM) image has long been the dominate data source, and recently LiDAR has offered an important new structural data stream for forest biomass estimations. On the other hand, forest biomass uncertainty analysis research has only recently obtained sufficient attention due to the difficulty in collecting reference data. This paper provides a brief overview of current forest biomass estimation methods using both TM and LiDAR data. A case study is then presented that demonstrates the forest biomass estimation methods and uncertainty analysis. Results indicate that Landsat TM data can provide adequate biomass estimates for secondary succession but are not suitable for mature forest biomass estimates due to data saturation problems. LiDAR can overcome TM's shortcoming providing better biomass estimation performance but has not been extensively applied in practice due to data availability constraints. The uncertainty analysis indicates that various sources affect the performance of forest biomass/carbon estimation. With that said, the clear dominate sources of uncertainty are the variation of input sample plot data and data saturation problem related to optical sensors. A possible solution to increasing the confidence in forest biomass estimates is to integrate the strengths of multisensor data.

## 1. Introduction

It is well known that forest ecosystems provide an important carbon reservoir. Continued deforestation and forest degradation will result in the loss of forest biomass/carbon stocks magnifying the global negative effects of climate change. The policy and management decisions governing these resources will play a critical role in mitigating these effects, thus requiring a robust method of monitoring the spatial and temporal patterns of forest biomass. The rates of carbon emission are considered as the largest source of uncertainty in climate change scenarios due to the difficulty in spatial

explicitly estimating the carbon stocks and dynamic changes. One solution is to develop robust approaches for estimating biomass/carbon changes in land cover using remotely sensed data. The past three decades have produced significant advances in estimating forest biomass including the application of different sensor data (e.g., Landsat, radar, and LiDAR) and the development of advanced techniques such as regression analysis, neural network, and process-based ecosystem models [1–12].

There are four critical issues associated with remote sensing-based biomass estimation methods: (1) sufficient number of available sample plots, (2) selection of suitable

metrics for biomass modeling, (3) use of suitable algorithms to establish biomass estimation models, and (4) use of suitable methods to conduct uncertainty analysis of the estimates. The literature is rich with examples that highlight these issues. Lu [5] has summarized the biomass estimation methods and discussed the potential methods to improve biomass estimation performance. The application of LiDAR, radar, and hyperspectral data for biomass estimation was reviewed by Koch [13], while Gleason and Im [14] summarized biomass estimation methods based on literature in the past ten years.

Landsat Thematic Mapper (TM) data with suitable spectral and spatial resolution and relatively long history of data availability have made it a primary data source for biomass estimation. However, it does have a critical limitation in that data saturation generates large uncertainty for sites with high biomass density or sites having complex forest stand structure [15]. LiDAR data has recently received more attention because it can overcome the data saturation shortcoming of Landsat providing more robust biomass estimations. The confidence of these measures has long been recognized as an important part in forest biomass estimation; however, research on biomass uncertainty analysis has only recently obtained sufficient attention due to the difficulty in collecting reference data. The objective of this paper is to briefly overview aboveground forest biomass estimation methods based on Landsat and LiDAR data and the methods for uncertainty analysis and to present case studies. This paper is organized as follows: Section 2 provides an overview of biomass estimation methods with Landsat TM and LiDAR data, and uncertainty analysis methodology; Section 3 provides the case study corresponding to each review topic; Section 4 provides conclusions and discusses the challenges.

## 2. An Overview of Biomass Estimation and Uncertainty Analysis Methods

*2.1. Biomass Estimation with Optical Sensor Data.* The remote sensing-based biomass estimation methods assume that forest stand information captured by sensors is highly correlated with aboveground biomass. According to this assumption, the keys for biomass estimation are to use appropriate variables and to develop suitable estimation models if sufficient sample plots are available. Many new variables such as vegetation indices and textures can be calculated from the multispectral bands [16–18]. Lu and his colleagues examined the relationships of aboveground forest biomass with Landsat TM spectral signatures and vegetation indices [16], and textural images [17] in the moist tropical regions of the Brazilian Amazon, and then extensively investigated the potential variables for biomass estimation. They found that a combination of spectral response and textural images can effectively improve biomass estimation performance, especially in the areas with complex forest stand structures [15]. They also found that the use of subpixel information offered better estimation results than per-pixel-based spectral signatures [19]. However, the complex forest stand structures and biophysical environments often result in data

saturation in Landsat images that was a major source of uncertainty in biomass estimation [14, 15].

In addition to the selection of suitable spectral variables, the use of proper algorithms for establishing biomass estimation models is also critical. Common algorithms include multiple regression analysis, neural networks, and K-nearest neighbor (KNN) [5]. Multiple regression analysis may be the most frequently used approach for developing biomass estimation models [11, 12, 15, 20–25]. This approach will typically use the sample plot biomass measure as the dependent variable; spectral signatures, vegetation indices, and/or textures from Landsat TM image are used as independent variables. The regression models assume that the biomass variable is linearly correlated with spectral responses and that limited correlations exist between independent variables. This assumption is usually not met in practice because remotely sensed data are often highly correlated each other [16]. There is also an issue with the regression approach in that the selected variables may have a nonlinear relationship with forest biomass [26]. An alternative to these limitations is to use nonparametric approaches such as neural network and KNN.

Neural networks have universal approximation properties that are regarded as a more robust solution for complicated and nonlinear problems compared to conventional parametric approaches [27] and have been used for biomass or volume estimation [27–32]. In the KNN approach, it is assumed that the spectral responses of pixels are only dependent on the state of forests. The estimate of each location was computed as a weighted mean of K spectrally nearest neighbors by inverse distance weighting. This approach has been used to update national forest inventory databases in Nordic countries such as Finland and Sweden based on the combination of plot inventory data and Landsat images [33–38]. Both neural networks and KNN methods are not as extensively applied as multiple regression methods for biomass estimation. Both methods require a large number of observations and a good representation of the variation of forest variables limiting their applications.

Most forest biomass estimation based exclusively on TM images was conducted in the 1990s and 2000s. Newer sensors such as radar and LiDAR have recently offered an important new structural data stream for forest biomass estimations. Applications using these new datasets to estimate biomass have obtained increasing attention [8, 11, 12, 39–42]. The following section focuses on the overview of biomass estimation methods with LiDAR data.

*2.2. Biomass Estimation with LiDAR Data.* LiDAR remote sensing systems can be distinguished based on the way with which returned signals are recorded (discrete return or wave form), scanning pattern (profiling or scanning), platforms (airborne, spaceborne, or ground based), and footprint size (small footprint: ~1 m or smaller; medium footprint: ~10–30 m; or large footprint: ~50 m or larger). The most common configurations of LiDAR systems are (1) airborne discrete-return scanning LiDAR, (2) airborne discrete-return profiling LiDAR, (3) airborne small-footprint waveform

LiDAR, (4) airborne medium-footprint waveform LiDAR, and (5) satellite waveform systems.

Airborne discrete-return scanning LiDAR may be the most widely used LiDAR system for biomass estimation. Although airborne discrete-return LiDAR could be scanning or profiling, for simplicity we refer airborne discrete-return LiDAR to the scanning sensors specifically because the use of profiling systems has mainly been limited to researchers in NASA (e.g., [39, 43]). Because a LiDAR sensor can directly measure components of vegetation canopy structure, such as canopy height, previous research has indicated that the use of LiDAR data is a promising approach for biophysical parameter estimation [8, 44–47]. Lim et al. [45] had reviewed the application of LiDAR data to forest studies. More detailed review of using LiDAR data for biomass estimation was also provided by Koch [13] and Gleason and Im [14].

Previous studies often used regression models to estimate biomass based on the metrics derived from LiDAR data. Because most allometric equations for calculating biomass in the field are power models [48], biomass and LiDAR metrics are usually log transformed when fitting a regression model [45, 49, 50]. In this case, a simple linear regression corresponds to a simple power model, and a multiple regression model corresponds to a multiplicative power model.

For airborne discrete-return LiDAR systems, the metrics were usually extracted from the canopy height distribution of the point cloud (e.g., [45, 51]). LiDAR metrics can be calculated based on all returns or a specific type of returns (such as first returns). Although most studies have used all returns, there are a few studies that used LiDAR metrics derived from first returns for predicting biomass (e.g., [52, 53]). Kim et al. [53] found that the model using first returns improved the  $R^2$  by 0.1 for predicting the total aboveground biomass in the mixed coniferous forest in Arizona compared to the one using all returns.

Instead of using point cloud, some studies extracted LiDAR metrics by first interpolating the canopy height of LiDAR point clouds to generate a canopy height model (CHM) and then derived height metrics from the CHM cells [49, 50]. CHM usually records the maximum canopy height within each grid cell. Therefore, the height metrics derived from CHM is more related to upper canopy instead of the complete vertical profile of vegetation structure. However, few studies have reported the difference between metrics derived from different returns of LiDAR point cloud and those from CHM for biomass estimation.

**2.3. Uncertainty Analysis.** Uncertainty analysis of biomass estimates from remote sensing-based methods has only recently received sufficient attention [54, 55]. In forest biomass/carbon estimation, there are many sources of uncertainties, such as biased sampling, plot location errors, measurement errors of tree variables (diameter and height), improper use of allometric equations for biomass calculation, uncertainties from biomass and carbon conversion factors and from geometrical and radiometric correction of remotely sensed data, and the selection of input parameters

used in modeling algorithms. The keys to reduce uncertainties are to identify the sources causing the uncertainty, to model their accumulation and propagation and to quantify the amount of uncertainties contributed to the final output. Friedl et al. [56] summarized three primary sources: (1) errors introduced through the image acquisition process, (2) errors produced by the application of data processing techniques, and (3) errors associated with interactions between instrument resolution and the scale of an ecological process on the ground. The uncertainty of estimating forest carbon may be much higher than the change to be achieved in carbon sequestration through changes in forest management [57]. The sampling error, measurement error for tree height and diameter, and regression error for tree volume are often recognized as major error components [58].

The map accuracy is traditionally assessed by calculating correlation or root mean square error (RMSE) between estimates and observed values. However, the accuracy often varies spatially depending on the complexity of landscape, density of sampled data, and accuracy of the remotely sensed data used [59, 60]. Pixel-based uncertainty analysis is needed to account for spatial variation of uncertainties. Moreover, the assessment of map quality can be conducted by quantifying and assessing the impacts of various input variations on the variation of outputs or by estimating the amounts of various input and output uncertainties and then conducting an uncertainty budget, that is, portioning the output uncertainty into the input uncertainty components. The former is called sensitivity study and the latter called uncertainty analysis [57].

As a specialized form of sensitivity analysis, an uncertainty and error budget shows the impacts of individual input uncertainties and their groups on the quality of output estimates [61]. The goal for developing an uncertainty and error budget thus is to account for all major sources of uncertainties and further provide modelers with guidelines for reduction of uncertainties and provide policy makers with suggestions for risk assessment of decision making. For the purpose, the widely used methods include the Monte Carlo [62], Fourier Amplitude Sensitivity Test (FAST) [63, 64], Taylor series [65, 66], polynomial regression [67], and response surface modeling [68, 69]. Helton and Davis [70] summarized the advantages and disadvantages of these methods. Although these methods were originally developed for building uncertainty and error budgets for a spatial models, the methods can be improved and applied to conduct uncertainty analysis on a pixel-by-pixel basis [71–73]. One example is that Gertner et al. [74, 75] extended the polynomial regression method to carry out uncertainty and error budgets for spatially explicit estimates of soil erosion.

To date, there are only few reports on spatial uncertainty analysis and error budget of mapping forest biomass/carbon. For this purpose, Wang et al. [54] demonstrated a methodological framework by integrating an image-aided spatial co-simulation algorithm and a polynomial regression model. The co-simulation was used to combine forest inventory plot data with TM images and generate forest carbon maps with uncertainties of estimates, and the polynomial regression model was applied to establish the relationships

of input uncertainties with output uncertainties. They found that the variation of image data had more impact on the accuracy of forest carbon-predicted values than the sample plot data. Moreover, Wang et al. [55] investigated the impacts of plot location errors on accuracy of forest carbon estimates and concluded that the perturbations of plot locations statistically did not lead to biased population mean predictions of forest carbon but increased the RMSE of predicted values, and plot location errors smaller than 800 m did not lead to noticeably different spatial distributions of the predicted values.

### 3. Case Studies

This section provides three case studies to demonstrate the application of biomass estimation methods with Landsat TM and LiDAR data, respectively, and uncertainty analysis. The first case study estimates aboveground biomass using Landsat TM. The second case study estimates biomass using LiDAR data. The third case study illustrates the methods for conducting uncertainty analysis in order to understand the major sources influencing biomass estimation performance.

*3.1. Case Study I: Aboveground Biomass Estimation with Landsat TM Image.* Since the 1970s, the Brazilian Amazon has experienced high deforestation due to road-building, logging, agricultural, and cattle-raising expansion [76]. About 20–50% deforested areas in the overall Amazon basin are in certain stages of successional forests [77–80]. Accurate delineation of successional and mature forest biomass distribution becomes considerably significant in reducing the uncertainty of carbon emission and sequestration. It is also critical for understanding their roles in influencing soil fertility and land degradation or restoration and in environmental processes and sustainability [81]. This case study selected Machadinho d'Oeste in northeastern Rondônia, Brazil to examine the capability of using Landsat TM image for aboveground biomass estimation.

*3.1.1. Methods.* Fieldwork was conducted in July–August 1999. A detailed description of the field measurement and the calculation of aboveground forest biomass based on diameter at breast height (DBH) and tree height was provided in Lu et al. [19]. A total of 26 sample plots for secondary succession with data range of 24–160 T/ha and a mean value of 89 T/ha, and a total of 14 sample plots for mature forest with data range of 111–495 T/ha and a mean value of 248 T/ha were collected and used in this research. Based on the probability of tree height distribution in a sample plot, entropy was calculated and used to evaluate the complexity of forest stand structures [15].

The TM image, which was acquired on 18 June 1998, was geometrically rectified using control points taken from topographic maps at 1 : 100,000 scale (Universal Transverse Mercator, south 20 zone) with a RMSE of less than 0.5 pixels. An improved image-based dark object subtraction model was used to implement radiometric and atmospheric correction [82, 83]. The surface reflectance values after calibration

were rescaled to the range of 0–100 by multiplying 100 for each pixel. Eight grey level cooccurrence matrix- (GLCM-) based texture measures (i.e., mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation) associated with five TM bands (bands TM 2, 3, 4, 5, and 7), and seven different sizes of moving windows (i.e.,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ,  $11 \times 11$ ,  $15 \times 15$ ,  $19 \times 19$ , and  $25 \times 25$ ) were tested in order to identify the best variables for biomass estimation [17].

The multiple regression analysis was used for biomass estimation in this study. The sample plot data were linked to image variables to extract the mean value for each sample plot based on a window size of  $3 \times 3$  pixels. Because so many textural images were available but not all were needed in the biomass estimation, it was important to identify suitable textural images. Pearson correlation analysis was used to analyze relationships between biomass and TM-derived variables, including TM-spectral signatures and textures, and between the textural images each other. Only the textural images having significantly high correlation coefficients with biomass but having low correlation each other were selected and used for biomass estimation. The aboveground biomass from sample plots was used as a dependent variable and the selected TM-derived variables were used as independent variables. A stepwise regression analysis was used to develop the biomass estimation model. The coefficient of determination ( $R^2$ ) was calculated to evaluate a regression model performance.

*3.1.2. Resultant Analysis.* The regression analysis results (Table 1) indicated that spectral signature led to much better estimation performance than textural images for secondary forest, but the result was inverse for mature forest. Neither spectral signatures nor textural images could effectively estimate mature forest biomass. A combination of spectral signature and textural images slightly improved secondary forest biomass estimation performance, but the improvement was considerable for mature forest biomass estimation. This research also indicated that textural image was a critical variable for mature forest biomass estimation, but not for secondary forest. As Figure 1 illustrated, if entropy was used to express the complexity of a forest stand structure, the entropy value was linearly related to secondary forest biomass, implying that the stand structure became complex as biomass increases in successional stages. However, the entropy value was similar for mature forests even if there was a significantly large range of biomass density (from approximately 110 to 500 T/ha in this study). This implied the data saturation problem in Landsat TM-spectral data due to the impact of forest stand structure in mature forest. In this case, use of spatial information based on textural measures was valuable to reduce the impacts of the forest stand structure, especially the heterogeneity of forest stands; thus, incorporation of textural images was necessary for mature forest biomass estimation.

*3.1.3. Discussion and Summary.* The aboveground forest biomass estimation based on TM image has indicated that

TABLE 1: A summary of regression analysis results in the western Brazilian Amazon.

Variables	Models	$R^2$	Beta value
Spectral signature	$AGB_{SF} = 667.72 - 13.92 \times f_{b4}$	0.75	
	$AGB_{MF} = 1024.14 - 54.96 \times f_{b5}$	0.16	
Texture	$AGB_{SF} = 164.62 - 2.27 \times f_{var}$	0.23	
	$AGB_{MF} = 134.57 + 19.29 \times f_{con}$	0.39	
Combination	$AGB_{SF} = 480.82 - 8.06 \times f_{b4} - 0.98 \times f_{var}$	0.76	-0.71 (sp), -0.24 (txt)
	$AGB_{MF} = 753.31 - 43.21 \times f_{b5} + 17.89 \times f_{con}$	0.50	-0.31 (sp), 0.65 (txt)

(1)  $AGB_{SF}$  and  $AGB_{MF}$  represent biomass estimates for successional forest and mature forest.

(2)  $f_{b4}$  and  $f_{b5}$  represent reflectance values of TM bands 4 and 5.

(3)  $f_{var}$  and  $f_{con}$  represent the textural image developed with the variance texture measure with TM band 4 and a window size of  $15 \times 15$  pixels, and the textural image developed with the contrast texture measure with TM band 5 and a window size of  $19 \times 19$  pixels.

(4) sp and txt represent the variables of spectral signatures and textures.

(5)  $R^2$  is the coefficient of determination.

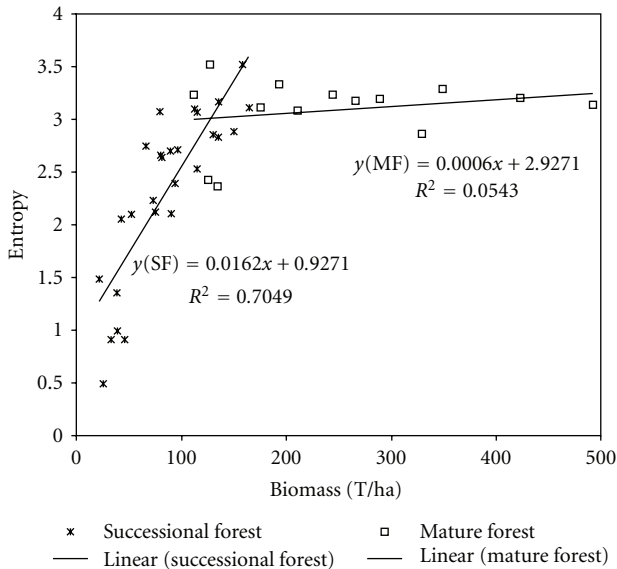


FIGURE 1: Comparison of aboveground biomass-entropy relationships in successional and mature forests in Machadinho d'Oeste in northeastern Rondônia, Brazil.

texture is an important variable for improving biomass estimation performance for the areas with complex forest stand structure. One critical step is to identify suitable textures but remains a challenge because textures vary with the characteristics of the landscape under investigation and images used. Identifying suitable textures involves the determination of appropriate texture measures, moving window sizes, image bands, and so on [17, 84, 85]. In practice, it is still difficult to identify which texture measures, window sizes, and image bands are suitable for a specific research topic, and there is also a lack of guidelines on how to select an appropriate texture [17].

This case study indicates that Landsat TM image can provide reasonable good estimation results for secondary forest biomass estimation, but the estimates have relatively poor for mature forests due to the data saturation problem

in Landsat TM image. As shown in this case study, when biomass density is greater than approximately 150 T/ha in the moist tropical forest of the Brazilian Amazon basin, Landsat TM spectral signatures become insensitive due to the complex forest stand structure. The impacts of different biophysical environment on optical sensor data and different composition of vegetation species make it difficult in the model transferability, as previous research shown [15, 81]. It is important to develop reliable and stable variables from multispectral image. Spectral mixture analysis may provide an alternative because it can decompose the multispectral image into fractions having biophysical meanings [19].

The first case study shows the importance of Landsat TM image for successional forest biomass estimation, but incapability for mature forest because of data saturation problem caused by complex forest stand structure. Since LiDAR data can capture canopy height information, the use of LiDAR data may significantly improve biomass estimation performance. The second case study shows the application of LiDAR data for biomass estimation.

### 3.2. Case Study II: Biomass Estimation with LiDAR Data.

Sierra Nevada in California is a mountain range as a home to many national resorts such as Lake Tahoe and Yosemite National Park. It is also a major source of water to the Central Valley, the most productive agricultural area in the United States. The vegetation in Sierra Nevada, including montane forests in higher elevation and woodland in lower elevation, plays a key role in regulating water and carbon cycling and providing critical ecological services. In this case study, LiDAR data were used to estimate vegetation biomass in the Sagehen Experimental Forest, which is approximately 32 km north of Lake Tahoe on the eastern slope of the Sierra Nevada. The major tree species in this area include white fir (*Abies concolor*), red fir (*Abies magnifica*), mountain hemlock (*Tsuga mertensiana*), lodgepole pine (*Pinus contorta*), Jeffrey pine (*Pinus jeffreyi*), sugar pine (*Pinus lambertiana*), and western white pine (*Pinus monticola*).

3.2.1. Methods. A total of 81 circular field plots with each of 0.05 ha were collected in 2006. A Trimble GeoXH handheld

GPS with Zephyr Geodetic antenna was used to remeasure the center of 81 individual plots. The average horizontal accuracy of the new GPS measurements was 0.1 m with the majority of less than 0.2 m, and, at the worse case, it was 1.5 m. All trees greater than 5 cm in DBH within each plot were measured with a nested sampling design. Canopy trees (DBH  $\geq$  19.5 cm) were tagged and measured within the whole plot; understory trees (DBH between 5 and 19.5 cm) were measured for one third of the plot. Tree measurements include species, DBH, tree height, canopy base height, vigor, and crown position. Allometric equations of individual tree species were selected from the literature [86, 87] to predict biomass of individual trees based on their DBH and, in some cases, height. The individual tree biomass was summed at the plot level to derive aboveground live tree biomass density in T/ha.

LiDAR data were collected on September 14–17, 2005 for the study area using an Optech ALTM 2050 system, which acquired up to two returns per pulse at a pulse frequency of 50 kHz, scan frequency of 38 Hz, and a maximum scan angle of 15°, creating a swath width of  $\sim$ 580 m. The point density was about 2–4 returns per square meters. The LiDAR point clouds were first filtered to generate a Digital Elevation Model (DEM) of 1 m cell size. The canopy height of individual points was calculated by subtracting the terrain elevation from the original Z values. A number of LiDAR metrics were generated based on LiDAR point heights within any given plot: mean ( $h_u$ ), standard deviation ( $h_{std}$ ), skewness ( $h_{skn}$ ), and kurtosis ( $h_{kurt}$ ); proportion of LiDAR points within different height bins (0 to 5 m, 5 to 10 m, ..., 45 to 50 m, and  $>$ 50 m, denoted as  $p_{0\text{ to }5}$ ,  $p_{5\text{ to }10}$ , ...,  $p_{45\text{ to }50}$ , and  $p_{>50}$ , resp.); percentile heights (5, 10, ..., 100 percentile, denoted as  $h_5$ ,  $h_{10}$ , ...,  $h_{100}$ , resp.); quadratic mean height ( $h_{qm}$ ). To examine the impacts of laser returns on biomass estimation, we generated LiDAR metrics from point clouds based on first returns and all returns, respectively. Besides using point clouds, we also generated LiDAR metrics based on the 1-m Canopy Height Model (CHM) cells within individual plots. CHM was obtained by subtracting DEM from a DSM (Digital Surface Model), which was generated by interpolating the highest point within each cell. All of the above LiDAR data processing was conducted using the Tiffs (Toolbox for LiDAR Data Filtering and Forest Studies) software [88].

We first did log transformation for both biomass and individual LiDAR metrics and then fit linear models using stepwise regression to select the most statistically significant LiDAR metrics. The developed linear models were multiplicative in the original scale. A total of four models were developed depending on the data sources used to extract LiDAR metrics: (1) all LiDAR returns, (2) first returns only, (3) last returns only, and (4) CHM cells. Among the 81 plots used for model development, one plot had questionable GPS accuracy. Three plots were found to have large residuals ( $>$ 3 standard residuals) in the regression models and had obvious mismatch between field measurements and LiDAR point clouds after carefully visual inspection. Therefore, these four plots were removed, resulting in a total of 77 plots in statistical analysis.

TABLE 2: Regression models and their statistics.

Data source	Independent variables	$R^2$	RMSE (T/ha)
All returns	$h_{qm}, h_{35\text{ to }40}$	0.76	82.1
First returns	$h_{qm}, h_{35\text{ to }40}$	0.75	83.5
Last returns	$h_{qm}, h_{40}, h_{35\text{ to }40}$	0.76	81.1
CHM cells	$h_{qm}, h_{35\text{ to }40}$	0.77	81.0

Note:  $h_{qm}$  is the quadratic mean height,  $h_{35\text{ to }40}$  means the proportion of LiDAR points or CHM cells within the height bin of 35–40 m, and  $h_{40}$  means the 40th percentile of LiDAR points.  $R^2$  and RMSE are the coefficient of determination and root mean square error, respectively.

**3.2.2. Resultant Analysis.** The LiDAR metrics chosen for the four regression models and their model fitting statistics are summarized in Table 2. Among the four models, the model based on CHM cells has the highest  $R^2$ , followed by the one based on last returns and the one based on all returns. The model based on first returns has the highest RMSE. However, the differences among these four models are too small ( $R^2$ : 0.75–0.77; RMSE: 81.0–83.5 T/ha) to be considered as significantly different. If all field plots are used to fit the regression model, the biomass model based on LiDAR metrics from all returns is

$$\ln(\text{AGB}) = 2.10 + 1.56 \ln(h_{qm}) + 0.05 \ln(h_{35\text{ to }40}), \quad (1)$$

where AGB is the aboveground live tree biomass in T/ha,  $h_{qm}$  is the quadratic mean height in meter, and  $h_{35\text{ to }40}$  is the proportion of LiDAR points between 35 m and 40 m. The maximum canopy height of the field plots ranges from 15.7 m to 44.7 m with a mean value of 27.9 m. The above model indicates that if there are more points in the upper vertical stratum (higher  $h_{35\text{ to }40}$  values), a plot will have higher aboveground biomass. The quadratic mean height  $h_{qm}$  is also a metric giving more weight to the higher canopy points [89]. The selection of these two metrics in the model agrees with the fact that the biomass of a field plot is highly influenced by larger and taller trees. Figure 2 shows the predicted biomass in comparison to the field biomass when the LiDAR metrics from all returns are used to develop the biomass model. For the 77 field plots, the average field biomass is  $239.1 \pm 166.4$  T/ha and the maximum biomass is 808.9 T/ha. This figure clearly shows that LiDAR is able to predict biomass even for the plots of very high biomass.

**3.2.3. Discussion and Summary.** In previous studies, the coefficients of determination ( $R^2$ ) of the statistical models for LiDAR-based biomass estimation typically vary from 0.6 to 0.9 depending on the specific vegetation conditions, the amount of field observations, and the specific approach used for statistical modeling. Our results are in the range of the coefficients from these studies and are close to those from similar forest conditions (e.g., [90]). Although previous research indicated that the use of first returns instead of all returns can improve  $R^2$  by 0.1 [53], this study found a negligible difference ( $\sim$ 0.01) to use first, last, or all returns for biomass estimation. This research also indicated no significant difference by using LiDAR metrics derived from point clouds or CHM (Table 2).

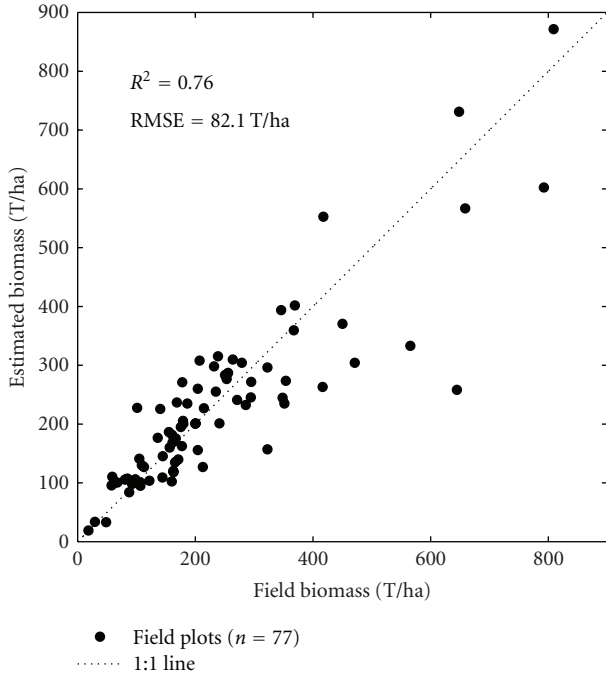


FIGURE 2: LiDAR-predicted versus field biomass over the 77 field plots in the Sagehen Experimental Forest in Sierra Nevada, California.

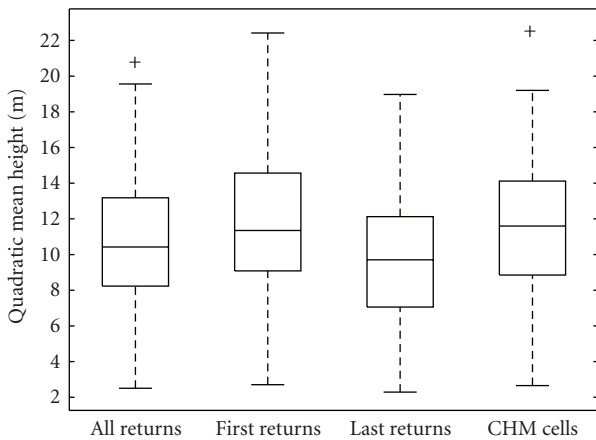


FIGURE 3: Boxplots of quadratic mean height derived from different LiDAR data sources.

Despite the models’ performance is similar, it was found that there are significant differences among the LiDAR metrics from some data sources. Figure 3 shows the box plots of quadratic mean height calculated from all returns, first returns, last returns, and CHM cells, respectively. Table 3 summarizes the relevant  $P$ -values of the pair-wise two-sample  $t$ -tests. It was observed that the quadratic mean height from first returns was higher than the one from last returns on average at the 5% significant level ( $P$ -value: 0.001). This makes sense since a laser usually penetrates deeper into the canopy to generate another return after the first one if it is not completely blocked. Figure 4(a) shows

TABLE 3:  $P$ -values of pair-wise two sample  $t$ -tests for quadratic mean heights derived from different data sources.

	First returns	Last returns	CHM cells
All returns	0.133	0.081	0.141
First returns	—	0.001	0.962
Last returns	—	—	0.001

an example of the LiDAR point cloud in one plot, with the first and last return linked by arrow-headed lines. It can be seen that there are large elevation differences between first and last returns for many laser pulses. Figure 4(b) shows the histogram of the differences between first and last returns for the same plot. Note that there are points with differences of 0.5 m–4 m, which is the so-called “blind-zone” of airborne LiDAR data.

The quadratic mean height from last returns is lower than the one from CHM cells on average at the 5% significant level ( $P$ -value = 0.001). This is also reasonable since CHM cells characterize the canopy surface height distribution while last returns correspond more to trunk, branches, and ground. The relevance of last returns to lower canopy elements such as branches and trunks, the major biomass components, might explain why the LiDAR metrics based on last returns have slightly better model performance ( $R^2$ : 0.76, RMSE: 81.1 T/ha) compared to the ones based on first returns ( $R^2$ : 0.75, RMSE: 83.5 T/ha). Again, this is a very small difference, and it is unclear whether this pattern will persist over other vegetation types.

Based on above observations, we conclude that different groups of LiDAR metrics (from all returns, first returns, last returns, or CHM cells) do not lead to significant difference in biomass estimation although they could be significant from each other by themselves. In other words, different group metrics are all sensitive to biomass variation in our study. Because our findings are different from the results from previous studies [53], more research is needed in the future along this line.

Although uncertainty analysis of the biomass estimation results is regarded as an important part in biomass modeling based on remote sensing methods, insufficient number of sample plots often make it difficult to conduct such analysis, as shown in much previous research and above two case studies. Therefore, the third case study in the following section shows the methods for conducting uncertainty analysis in a study area of Zhejiang province, China, in order to understand the major sources influencing biomass estimation performance.

3.3. Case Study III: Estimation and Uncertainty Analysis of Aboveground Forest Carbon. The third case study was conducted at Lin’An County, Zhejiang Province of East China for examining uncertainty of forest carbon estimates. Lin’An County has an area of 312,680 ha and is located in a subtropical zone with average annual temperature of 16.4°C and average annual precipitation of 1,628 mm. This is a mountainous area with an elevation range of 1770 m

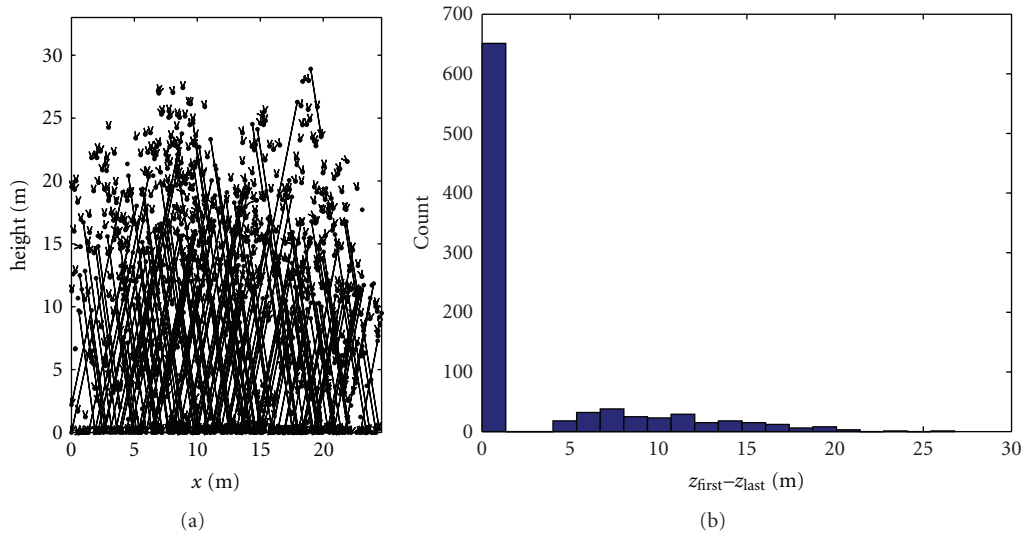


FIGURE 4: (a) LiDAR point cloud in one field plot with the first and last returns linked by arrow-headed lines, and (b) the histogram of the elevation differences between first and last returns for the same plot.

and characterized by plantation ecosystems consisting of coniferous forests, evergreen broad-leaf forests, deciduous and evergreen broadleaf-mixed forests [91].

**3.3.1. Methods.** A total of 766 plots covering 80% of the county which were measured in 2004 were used to generate an aboveground forest carbon map. Within each plot, tree DBHs and heights were measured. The plot tree volumes, including stump but excluding branches, were first computed by using the empirical models by species. The values of aboveground forest biomass and carbon were then obtained by using biomass expansion factors and carbon conversion factor [55].

A Landsat Enhanced Thematic Mapper Plus (ETM+) image, dated on August 3, 2004, was used for this study. This ETM+ image was geometrically rectified using a digital topographic map with the Universal Transverse Mercator (UTM) as reference coordinate system, and the resulting RMSE was less than one pixel. In order to improve relationships of the images with the aboveground forest carbon, six inversions of bands:  $1/b_1$ ,  $1/b_2$ ,  $1/b_3$ ,  $1/b_4$ ,  $1/b_5$ ,  $1/b_7$ , and ten band ratios:  $b_1/b_4$ ,  $b_2/b_4$ ,  $b_3/b_4$ ,  $b_5/b_4$ ,  $b_7/b_4$ ,  $b_8/b_4$ ,  $(b_4 - b_3)/(b_4 + b_3)$ ,  $(b_3 + b_5)/b_7$ ,  $(b_3 + b_5 + b_7)/b_4$ , and  $(b_2 + b_3 + b_5)/b_7$ , were calculated. Moreover, correlation matrix-based principal component analysis of six original ETM+ bands was made. The images that had the highest correlation with forest carbon were used to create estimates of forest carbon with an image-based cosimulation algorithm.

The 766 plots were divided into two groups by a random sampling. The first group of 600 plots (Figure 5(a)) was used to generate an aboveground forest carbon map by an image-based spatial co-simulation algorithm in which the plot data were combined with the ETM+ image [54]. The second group of 166 plots (Figure 5(b)) was applied to conduct the

accuracy assessment and uncertainty analysis of the obtained estimates. In the algorithm, it is assumed that aboveground forest carbon is spatially autocorrelated [55, 92]. That is, the closer the locations, the more similar the values of aboveground forest carbon. The similarity becomes weaker and weaker as the distance between two locations increases. Once the distance reaches a threshold value, the similarity will disappear. This feature of aboveground forest carbon can be characterized by using variogram. The threshold value for the distance between two locations is called the range of spatial autocorrelation. The values of aboveground forest carbon are spatially dependent on each other within the range and independent outside of it. In addition, it is assumed that aboveground forest carbon is spatially cross-correlated with the image-spectral variables. The spatial autocorrelation of the forest carbon and spectral variables and cross-correlation between them provide the basis on which the predicted values of the forest carbon at unobserved locations can be generated from the sample plot data using the co-simulation algorithm.

The value of aboveground forest carbon at each location is regarded as a realization of this random process, and the realization can be created using the following co-simulation algorithm [54, 55]. The study area is first divided into square cells or pixels, and a random path to visit each of the pixels is set up. When each pixel is predicted, a neighborhood in which plot data are used is determined based on the range of spatial autocorrelation. Within this neighborhood, the sample plot data, previously simulated values if any, and collocated image data are employed to obtain a conditional cumulative distribution function. This function is determined by calculating a conditional mean and variance by an unbiased collocated simple cokriging estimator that weights the data values and spatial variability. The weights vary depending on spatial configuration of the



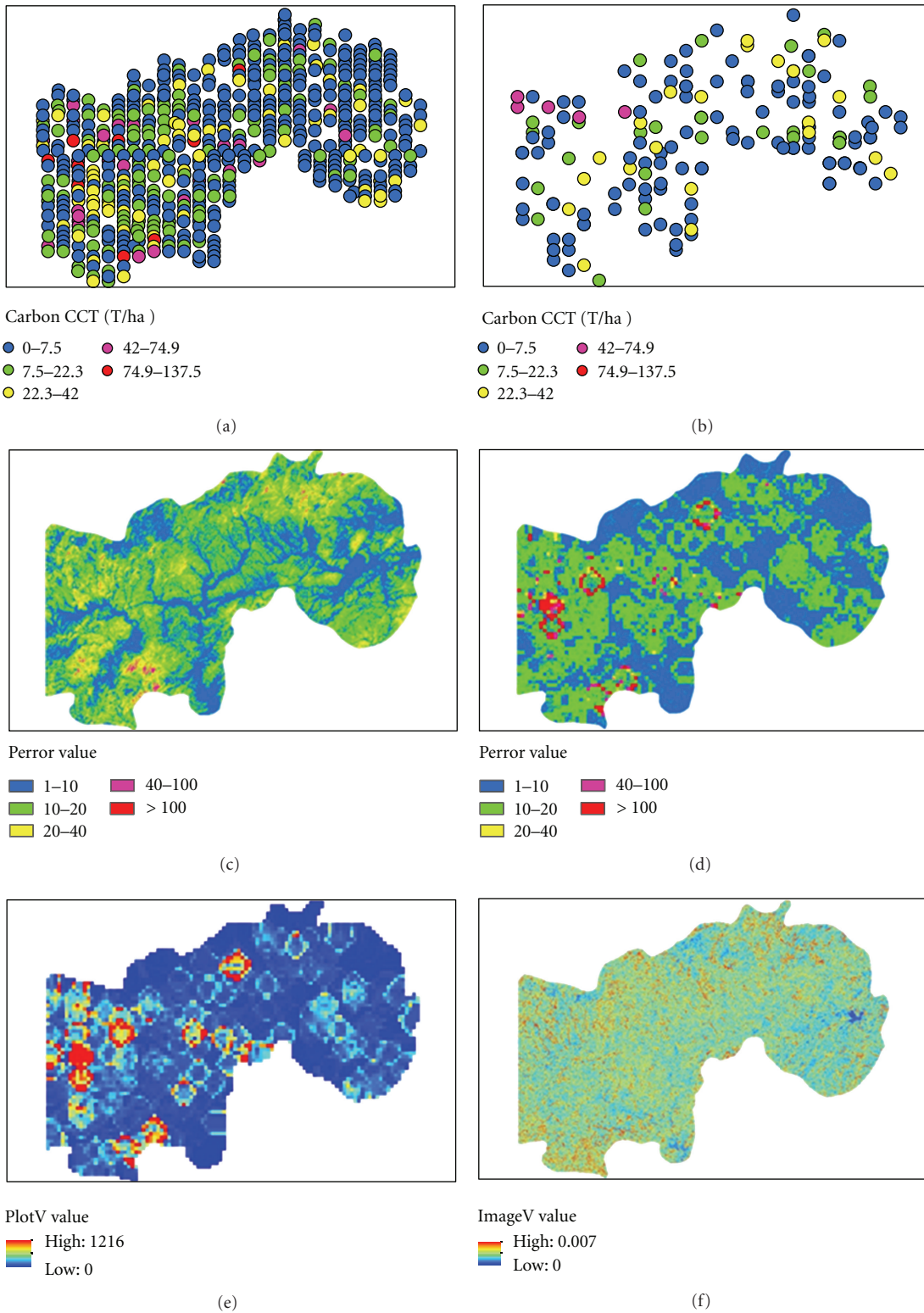


FIGURE 5: Continued.

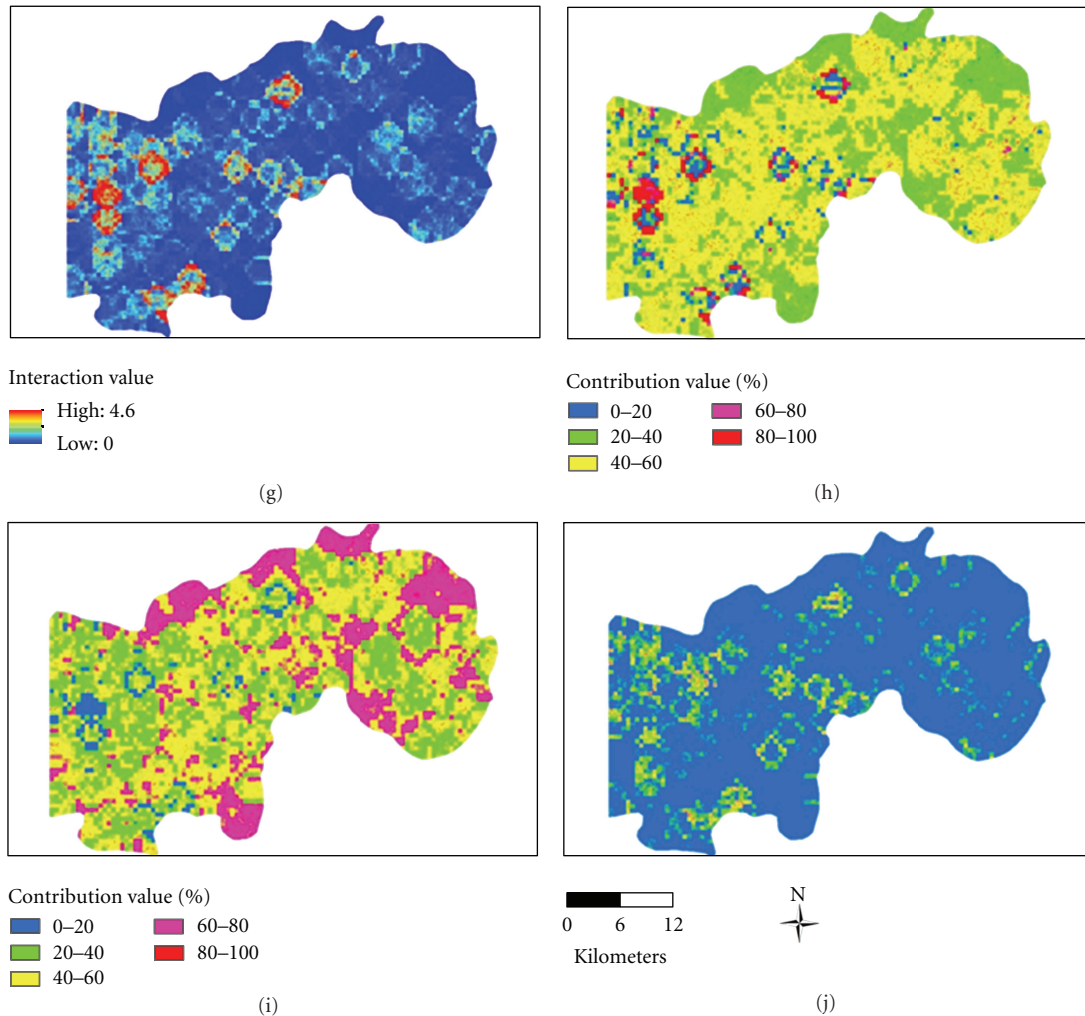


FIGURE 5: Aboveground forest carbon map obtained by combining forest inventory sample plot data and ETM+ images using image-based spatial co-simulation and its polynomial regression-based uncertainty analysis in Lin-An county, Zhejiang province of East China. [(a) forest carbon (T/ha) of sample plots used for development of algorithm; (b) forest carbon (T/ha) of sample plots used for test of results; (c) predicted forest carbon values (T/ha); (d) errors (T/ha) of predicted forest carbon values; (e) variance of input plot data; (f) variance of input image; (g) interaction of forest carbon and image variances; relative uncertainty contribution (%) from (h) the input plot data, (i) ETM+ image, and (j) their interaction].

data and predicted locations and their spatial autocorrelation and cross-correlation. Generally, the shorter the distance is, the greater the weight of the data is. From the distribution function, a value is then randomly drawn, regarded as a realization of the forest carbon at this location and also used as a conditional data for simulation of neighboring locations. A computer program is used to follow this random path to create a simulated value for each location. A map of the forest carbon is obtained after all pixels are visited. By setting up different random paths to visit each location and repeating the above process, more than one simulated values can be created. In this study, 400 simulated values for each pixel were obtained, and an E-type sample mean of the forest carbon was finally calculated and used as a predicted value.

An uncertainty and error budget method was developed based on the polynomial regression to explicitly model the

propagation of uncertainties from inputs to output for the above mapping [54]. The advantages of this method are that it can model spatial and temporal uncertainty propagation and contributions from sampling, plot location errors, measurement errors, improper allometric equations, biomass and carbon conversion factors, and so forth, and it can also handle the effect of interactions among input variables and the effect of neighboring information. First, the various sources of uncertainties are identified and quantified. A polynomial regression is then developed to define the relationships of uncertainties between the inputs and output. Finally, the regression model is used to estimate the relative uncertainty contributions of the inputs to the uncertainty of the output. In order to clearly explain this method, in this study only the variances of the used field plot data and ETM+ images were considered as the input uncertainties

for estimation of each location. The output uncertainty was the absolute difference between the estimated and observed values of aboveground forest carbon for each of the 166 test plots.

**3.3.2. Resultant Analysis.** The sample mean, standard deviation, and coefficient of variation of aboveground forest carbon for the 600 plots that were used for development of the co-simulation algorithm was 11.23 T/ha, 18.08 T/ha, and 161.08%, respectively. That is, the forest carbon at the plot level varied greatly due to a great range of elevation, and diverse soil and topographic features in the mountainous area. The spatial distribution of the predicted forest carbon values in Figure 5(c) was consistent with that of the sample plot data in Figure 5(a). That is, where the sample plots have higher forest carbon values, the predicted values are also higher and vice versa. Based on the test sample of 166 plots, the Pearson correlation coefficient and RMSE between the estimated and observed values of the forest carbon were 0.7146 and 9.98 T/ha, respectively. This RMSE value was relatively large, which was caused by large variation of the forest carbon in this area.

In order to model propagation of uncertainty, a total of sixteen 1st to 4th order polynomial regression models with and without constant regression coefficients, and with and without interactions of uncertainties between the forest carbon and spectral variable were obtained by using the 166 test sample plots. It was found that the 1st order polynomial regression model without the interaction and the 3rd order polynomial regressions models with and without the interaction, but without the constant regression coefficient, provided more accurate prediction. All other models led to negatively predicted values of the output uncertainty for the extremely small and large input errors and thus had poor ability of predicting uncertainty. Let  $F_u$ ,  $P_v$ , and  $I_v$  be the uncertainties of the output carbon, input plot variance and image variance, the models were expressed as follows:

$$F_u = 4.5349 + 0.07936P_v + 1007.191I_v, \quad (2)$$

$$F_u = 0.2768P_v + 9231.889I_v - 0.003657P_v^2 - 5840468I_v^2 + 0.00001315P_v^3 + 1280283000I_v^3, \quad (3)$$

$$F_u = 0.2452P_v + 10495.45I_v + 33.046P_v \times I_v - 0.003762P_v^2 - 7476639I_v^2 + 0.0000137P_v^3 + 1726692000I_v^3. \quad (4)$$

The values of regression coefficient  $R$  for models (2), (3), and (4) were 0.4147, 0.5033, and 0.5064, respectively. The regression coefficients were relatively low, mainly due to the great variations of the aboveground forest carbon, the input plot data and image variance, and the output uncertainties. Based on the model (4), the error map of the predicted values was calculated and illustrated in Figure 5(d). Its spatial distribution was mainly determined by the spatial patterns of the input variances from the field plot data in Figure 5(e) and

less affected by the input variances of the used image (Figure 5(f)). In the areas where the variation of the forest carbon was large (Figures 5(a) and 5(c)), the error of the predicted values was also large (Figure 5(d)). The relative uncertainty contributions from the input variances of the plot data were larger than 40% at the most part of the study area (Figure 5(h)), especially in the areas where the variation of the forest carbon was large. Thus, the plot data variance was the main source of uncertainty. In the areas where the uncertainty contributions from the plot data variances were smaller, the uncertainties of the predicted forest carbon values were mainly determined by the variances of the used image (Figure 5(i)). The variance contributions from the interaction of the aboveground forest carbon with the image were smaller but positive, implying that the variances from both plot data and ETM+ image increased the uncertainty of the predicted values (Figures 5(i) and 5(j)).

**3.3.3. Discussion and Summary.** Although the co-simulation algorithm led to an aboveground forest carbon map that had a similar spatial distribution of estimates to that of the sample plot data, the overall relative RMSE of 88.9% was very large. This was mainly due to the large variation of the aboveground forest carbon in this case study. Developing robust polynomial regression models that can account for the relationships of input uncertainties with output uncertainties is a great challenge for uncertainty analysis of forest biomass/carbon estimates. It is often difficult to obtain a high-quality polynomial regression model in terms of regression coefficient or coefficient of determination due to large variations of input and output uncertainties, especially when outliers exist. Wang et al. [54] used a 4th-order polynomial regression with a regression coefficient of 0.4818. Although the regression coefficients of the obtained models were statistically significant in this case study, they were relatively small. The main reason was due to the large variations of the input plot data variances and image variances. In order to improve the quality of the models, in practice, the outliers can be removed using 2 standard deviations.

This case study led to a finding that the main source of the uncertainty for mapping aboveground forest carbon was the variances of the input field plot data and relatively the uncertainty contribution from the used image was small. Moreover, the interaction of the uncertainties from both the field plot data and image increased the uncertainties of the predicted values. Because of space limitation, this case study did not deal with other uncertainty sources including biased sampling, measurement errors of tree variables, uncertainties of allometric equation parameters for tree volume calculation, uncertainties from biomass and carbon conversion factors, uncertainties from geometrical and radiometric correction of remotely sensed data, and so forth. However, this case study did not lose its generalization as the demonstration of a methodology for mapping and uncertainty analysis of forest carbon by combining sample plot data and remotely sensed images. In future studies, more sources of uncertainties will be investigated.

#### 4. Conclusions and Challenges

Landsat TM imagery having proper spectral and spatial resolutions and relatively long historical datasets, as well as world-wide data availability with free has been extensively applied for forest biomass/carbon estimation. The TM imagery is suitable for forest biomass estimation for the area with relatively simple stand structure such as secondary forest but does not work well for the area with complex forest stand structure such as mature forest in the moist tropical region of the Brazilian Amazon. Due to the complexity of atmosphere, vegetation phenology, and different vegetation types and structures, application of Landsat TM images for biomass estimation is often site dependent, which means that the algorithm developed in one study area cannot be directly transferred to other study areas [15].

Many factors, such as collection of sample plots, biomass calculation with allometric equations, selection of spectral variables, and use of modeling algorithms, affect biomass estimation results. The uncertainty analysis based on co-simulation algorithm indicates that the variation of input field plots is the major source inducing uncertainties of biomass estimates. For the areas with complex forest stand structure and vegetation species composition, data saturation on the optical sensor data is another important source resulting in large uncertainty in the moist tropical region. Since LiDAR data can capture forest canopy height, its use has proven to offer an alternative for improvement of biomass estimation. However, LiDAR is mainly used to estimate canopy height. Although canopy height is a very important variable for biomass estimation, biomass is also related to other forest stand parameters such as tree species composition and tree density. LiDAR or optical sensor data alone cannot provide sufficiently accurate biomass estimation results. It necessitates making full use of different sensor features inherited in LiDAR and TM images. Therefore, identifying suitable methods to integrate both data sources for biomass estimation is valuable. This involves the selection of suitable variables for use in modeling and of suitable algorithms to establish estimation models. Through the uncertainty analysis, it is possible to identify the major sources affecting biomass estimation performance, and then we can adopt suitable measures to reduce the uncertainties.

When multiscale remotely sensed images are used for mapping forest biomass/carbon, a great challenge we are facing is how to combine the images that have inconsistent spatial resolutions. Traditional data fusion algorithms such as RGB (red-green-blue) to HIS (intensity-hue-saturation) and Brovey transformation [93] do not work well because the spatial resolutions of the results are limited to the original images. Multi-scale image fusion approaches such as multi-scale Kalman Filtering [94] are needed. Another great challenge is how to match the multi-scale images with sample plot data when the images and data have inconsistent spatial resolutions and further scale up the spatial data from finer spatial resolutions to coarser ones when sample plots are smaller than pixels of maps that are required for decision making. The upscaling of spatial data is often necessary because generally the sizes of sample plots are smaller than

30 m × 30 m because of cost limitation, while forest biomass/carbon maps at national scale are usually generated at spatial resolution of 1 km × 1 km. The challenge is mainly because the relationships of forest attributes with spectral variables vary depending on spatial resolution of data and a model that is developed based on the data at one spatial resolution cannot be used across resolutions. Wang et al. [54, 95] overcame this gap by developing an image-based spatial co-simulation algorithm in which the same distributions of spatial data are assumed when images and sample plot data are combined and scaled up from finer spatial resolutions to coarser ones.

Another challenge in forest biomass/carbon estimation is accuracy assessment and uncertainty analysis. Decision and policy makers need not only forest biomass estimates but also their measures of uncertainty. However, the latter is usually difficult to obtain. The challenge is mainly attributed to (1) high cost to collect independent sample plot data; (2) complicated landscapes and environments in which various soil and topographic factors lead to high variation of forest biomass/carbon; (3) multiple sources of uncertainties; (4) lack of algorithms that can be used to conduct uncertainty analysis. Wang et al. [54, 55] suggested four groups of uncertainty sources for forest biomass/carbon estimation. First of all, there are uncertainties of the used data, mainly including measurements errors of tree variables, plot location errors, image data errors due to remote sensing systems and atmospheric conditions, geometric errors due to mismatch of image and map coordinate systems, and so forth. The second group of uncertainty deals with variation of variables and parameters. For example, high variation of forest structures and complicated tree species composition will no doubt result in large uncertainties of forest biomass estimates. High variation of tree volume to biomass and biomass to carbon conversion factors will also lead to great uncertainties. Moreover, the uncertainties may also be due to uses of biased sampling methods, improper allometric equations for tree volume calculation, inaccurate interpolation algorithms, and so forth. In addition, knowledge gaps are another source of uncertainty.

Some of uncertainties may accumulate and propagate to the outputs through the estimation and mapping system of forest biomass/carbon and others may be cancelled out. Uncertainty analysis-modeling and quantifying the propagation of the input uncertainties to the outputs is thus a complicated process. Currently, robust approaches are still lacking. Larocquea et al. [96] discussed a theoretical framework to estimate error propagation for process-based models of forest ecosystem carbon cycle. Nabuurs et al. [57] compared the uncertainties of forest carbon estimates for a tropical and a temperate forest for model-based methods and found that carbon content, wood density, and current annual increment of stems were the parameters that exhibited the highest influence on carbon sequestration. Wang et al. [54] developed a general framework for uncertainty analysis of forest biomass/carbon estimates, and this study further demonstrated its application.

In this study, the first case study revealed data saturation of Landsat TM images for biomass estimation of mature

forest as one of uncertainty sources. The uncertainties can be attributed to both complex forest stand structures and inherent disadvantages of Landsat TM sensors. Because of its capability to capture canopy height information, the use of LiDAR data in the second case study implied the potential to improve biomass estimation performance by combining LiDAR and Landsat TM images. However, due to limitations of space and time, this study did not demonstrate the integration. Because of the insufficient number of sample plots, in addition, uncertainty analysis of this study was only conducted in the third case study. Thus, the important contribution of this study lies at providing the potential and suggestions for the future research direction.

In this study, an image-based spatial conditional co-simulation algorithm was used to map aboveground forest carbon in the third case study. This algorithm is based on spatial autocorrelation of variables and accounts for not only the local spatial variation of variables but also the uncertainties of the local estimates [54, 55]. In practice, the spatial autocorrelation of forest variables widely exist. The variables and the relationships between them also vary from place to place. Thus, this algorithm is very promising. Another similar choice is geographically weighted regression (GWR) that is a local form of linear regression [97]. GWR can be used to model spatially varying relationships and allows regression coefficients to vary spatially. In the first and second case studies of this research, however, a global regression modeling method was employed to map forest biomass, and the local variation and spatial autocorrelation of forest variables were not tested mainly because of limited sample plots. In the future study, it may be necessary to demonstrate the applications of GWR to these study areas.

## Acknowledgments

This research was partially supported by the Anthropological Center for Training and Research on Global Environmental Change (ACT), Indiana University and partially supported by the National Natural Science Foundation of China (no. 30972360).

## References

- [1] R. A. Houghton, J. L. Hackler, and K. T. Lawrence, "The U.S. carbon budget: contributions from land-use change," *Science*, vol. 285, no. 5427, pp. 574–578, 1999.
- [2] F. Achard, H. D. Eva, P. Mayaux, H. J. Stibig, and A. Belward, "Improved estimates of net carbon emissions from land cover change in the tropics for the 1990s," *Global Biogeochemical Cycles*, vol. 18, no. 2, article GB2008, 11 pages, 2004.
- [3] S. Hese, W. Lucht, C. Schmullius et al., "Global biomass mapping for an improved understanding of the CO<sub>2</sub> balance—the Earth observation mission Carbon-3D," *Remote Sensing of Environment*, vol. 94, no. 1, pp. 94–104, 2005.
- [4] R. A. Houghton, "Aboveground forest biomass and the global carbon balance," *Global Change Biology*, vol. 11, no. 6, pp. 945–958, 2005.
- [5] D. Lu, "The potential and challenge of remote sensing-based biomass estimation," *International Journal of Remote Sensing*, vol. 27, no. 7, pp. 1297–1328, 2006.
- [6] P. Muukkonen and J. Heiskanen, "Estimating biomass for boreal forests using ASTER satellite data combined with stand-wise forest inventory data," *Remote Sensing of Environment*, vol. 99, no. 4, pp. 434–447, 2005.
- [7] J. A. Blackard, M. V. Finco, E. H. Helmer et al., "Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information," *Remote Sensing of Environment*, vol. 112, no. 4, pp. 1658–1677, 2008.
- [8] M. García, D. Riaño, E. Chuvieco, and F. M. Danson, "Estimating biomass carbon stocks for a Mediterranean forest in central Spain using LiDAR height and intensity data," *Remote Sensing of Environment*, vol. 114, no. 4, pp. 816–830, 2010.
- [9] S. L. Powell, W. B. Cohen, S. P. Healey et al., "Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: a comparison of empirical modeling approaches," *Remote Sensing of Environment*, vol. 114, no. 5, pp. 1053–1068, 2010.
- [10] M. Chopping, C. B. Schaaf, F. Zhao et al., "Forest structure and aboveground biomass in the southwestern United States from MODIS and MISR," *Remote Sensing of Environment*, vol. 115, no. 11, pp. 2943–2953, 2011.
- [11] E. T.A. Mitchard, S. S. Saatchi, S. L. Lewis et al., "Measuring biomass changes due to woody encroachment and deforestation/degradation in a forest-savanna boundary region of central Africa using multi-temporal L-band radar backscatter," *Remote Sensing of Environment*, vol. 115, no. 11, pp. 2861–2873, 2011.
- [12] G. Sun, K. J. Ranson, Z. Guo, Z. Zhang, P. Montesano, and D. Kimes, "Forest biomass mapping from lidar and radar synergies," *Remote Sensing of Environment*, vol. 115, no. 11, pp. 2906–2916, 2011.
- [13] B. Koch, "Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 6, pp. 581–590, 2010.
- [14] C. Gleason and J. Im, "A review of remote sensing of forest biomass and biofuel: options for small-area applications," *GIScience and Remote Sensing*, vol. 48, no. 2, pp. 141–170, 2011.
- [15] D. Lu, "Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon," *International Journal of Remote Sensing*, vol. 26, no. 12, pp. 2509–2525, 2005.
- [16] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin," *Forest Ecology and Management*, vol. 198, no. 1–3, pp. 149–167, 2004.
- [17] D. Lu and M. Batistella, "Exploring TM image texture and its relationships with biomass estimation in Rondônia, Brazilian Amazon," *Acta Amazônica*, vol. 35, no. 2, pp. 249–257, 2005.
- [18] H. Du, R. Cui, G. Zhou et al., "The responses of Moso Bamboo (*Phyllostachys heterocycla* var. *pubescens*) forest aboveground biomass to Landsat TM spectral reflectance and NDVI," *Acta Ecologica Sinica*, vol. 30, no. 5, pp. 257–263, 2010.
- [19] D. Lu, M. Batistella, and E. Moran, "Satellite estimation of aboveground biomass and impacts of forest stand structure," *Photogrammetric Engineering and Remote Sensing*, vol. 71, no. 8, pp. 967–974, 2005.
- [20] P. S. Roy and S. A. Ravan, "Biomass estimation using satellite remote sensing data—an investigation on possible approaches for natural forest," *Journal of Biosciences*, vol. 21, no. 4, pp. 535–561, 1996.
- [21] R. F. Nelson, D. S. Kimes, W. A. Salas, and M. Routhier, "Secondary forest age and tropical forest biomass estimation using

- thematic mapper imagery," *BioScience*, vol. 50, no. 5, pp. 419–431, 2000.
- [22] R. Nelson, J. Jimenez, C. E. Schnell, G. S. Hartshorn, T. G. Gregoire, and R. Oderwald, "Canopy height models and airborne lasers to estimate forest biomass: two problems," *International Journal of Remote Sensing*, vol. 21, no. 11, pp. 2153–2162, 2000.
- [23] M. K. Steininger, "Satellite estimation of tropical secondary forest above-ground biomass: data from Brazil and Bolivia," *International Journal of Remote Sensing*, vol. 21, no. 6-7, pp. 1139–1157, 2000.
- [24] D. Zheng, J. Rademacher, J. Chen et al., "Estimating above-ground biomass using Landsat 7 ETM+ data across a managed landscape in northern Wisconsin, USA," *Remote Sensing of Environment*, vol. 93, no. 3, pp. 402–411, 2004.
- [25] J. Heiskanen, "Estimating aboveground tree biomass and leaf area index in a mountain birch forest using ASTER satellite data," *International Journal of Remote Sensing*, vol. 27, no. 6, pp. 1135–1158, 2006.
- [26] H. Li, P. Mausel, E. Brondizio, and D. Deardorff, "A framework for creating and validating a non-linear spectrum-biomass model to estimate the secondary succession biomass in moist tropical forests," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 2, pp. 241–254, 2010.
- [27] G. M. Foody, M. E. Cutler, J. McMorro et al., "Mapping the biomass of Bornean tropical rain forest from remotely sensed data," *Global Ecology and Biogeography*, vol. 10, no. 4, pp. 379–387, 2001.
- [28] Y. Q. Jin and C. Liu, "Biomass retrieval from high-dimensional active/passive remote sensing data by using artificial neural networks," *International Journal of Remote Sensing*, vol. 18, no. 4, pp. 971–979, 1997.
- [29] J. Hyypä, H. Hyypä, M. Inkinen, M. Engdahl, S. Linko, and Y. H. Zhu, "Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes," *Forest Ecology and Management*, vol. 128, no. 1-2, pp. 109–120, 2000.
- [30] G. M. Foody, D. S. Boyd, and M. E. J. Cutler, "Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions," *Remote Sensing of Environment*, vol. 85, no. 4, pp. 463–474, 2003.
- [31] F. Del Frate and D. Solimini, "On neural network algorithms for retrieving forest biomass from SAR data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 1, pp. 24–34, 2004.
- [32] P. Muukkonen and J. Heiskanen, "Biomass estimation over a large area based on standwise forest inventory data and ASTER and MODIS satellite data: a possibility to verify carbon inventories," *Remote Sensing of Environment*, vol. 107, no. 4, pp. 617–624, 2007.
- [33] Z. Fazakas, M. Nilsson, and H. Olsson, "Regional forest biomass and wood volume estimation using satellite data and ancillary data," *Agricultural and Forest Meteorology*, vol. 98-99, pp. 417–425, 1999.
- [34] M. Katila and E. Tomppo, "Selecting estimation parameters for the Finnish multisource national forest inventory," *Remote Sensing of Environment*, vol. 76, no. 1, pp. 16–32, 2001.
- [35] M. Halme and E. Tomppo, "Improving the accuracy of multi-source forest inventory estimates by reducing plot location error—a multicriteria approach," *Remote Sensing of Environment*, vol. 78, no. 3, pp. 321–327, 2001.
- [36] E. Tomppo, M. Nilsson, M. Rosengren, P. Aalto, and P. Kennedy, "Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and above-ground biomass," *Remote Sensing of Environment*, vol. 82, no. 1, pp. 156–171, 2002.
- [37] H. Mäkelä and A. Pekkarinen, "Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data," *Forest Ecology and Management*, vol. 196, no. 2-3, pp. 245–255, 2004.
- [38] F. Maselli, G. Chirici, L. Bottai, P. Corona, and M. Marchetti, "Estimation of Mediterranean forest attributes by the application of k-NN procedures to multitemporal Landsat ETM+ images," *International Journal of Remote Sensing*, vol. 26, no. 17, pp. 3781–3796, 2005.
- [39] R. Nelson, J. Boudreau, T. G. Gregoire et al., "Estimating Quebec provincial forest resources using ICESat/GLAS," *Canadian Journal of Forest Research*, vol. 39, no. 4, pp. 862–881, 2009.
- [40] Q. Chen, "Retrieving vegetation height of forests and woodlands over mountainous areas in the Pacific Coast region using satellite laser altimetry," *Remote Sensing of Environment*, vol. 114, no. 7, pp. 1610–1627, 2010.
- [41] Q. Chen, "Assessment of terrain elevation derived from satellite laser altimetry over mountainous forest areas using airborne lidar data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 1, pp. 111–122, 2010.
- [42] G. Sandberg, L. M. H. Ulander, J. E. S. Fransson, J. Holmgren, and T. Le Toan, "L- and P-band backscatter intensity for biomass retrieval in hemiboreal forest," *Remote Sensing of Environment*, vol. 115, no. 11, pp. 2874–2886, 2011.
- [43] R. Nelson, W. Krabill, and J. Tonelli, "Estimating forest biomass and volume using airborne laser data," *Remote Sensing of Environment*, vol. 24, no. 2, pp. 247–267, 1988.
- [44] J. B. Drake, R. O. Dubayah, R. G. Knox, D. B. Clark, and J. B. Blair, "Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest," *Remote Sensing of Environment*, vol. 81, no. 2-3, pp. 378–392, 2002.
- [45] K. Lim, P. Treitz, M. Wulder, B. St-Onge, and M. Flood, "LiDAR remote sensing of forest structure," *Progress in Physical Geography*, vol. 27, no. 1, pp. 88–106, 2003.
- [46] P. Hyde, R. Dubayah, B. Peterson et al., "Mapping forest structure for wildlife habitat analysis using waveform lidar: validation of montane ecosystems," *Remote Sensing of Environment*, vol. 96, no. 3-4, pp. 427–437, 2005.
- [47] K. Zhao, S. Popescu, and R. Nelson, "Lidar remote sensing of forest biomass: a scale-invariant estimation approach using airborne lasers," *Remote Sensing of Environment*, vol. 113, no. 1, pp. 182–196, 2009.
- [48] D. Zianis and M. Mencuccini, "On simplifying allometric analyses of forest biomass," *Forest Ecology and Management*, vol. 187, no. 2-3, pp. 311–332, 2004.
- [49] G. Patenaude, R. A. Hill, R. Milne, D. L. A. Gaveau, B. B. J. Briggs, and T. P. Dawson, "Quantifying forest above ground carbon content using LiDAR remote sensing," *Remote Sensing of Environment*, vol. 93, no. 3, pp. 368–380, 2004.
- [50] B. St-Onge, Y. Hu, and C. Vega, "Mapping the height and above-ground biomass of a mixed forest using lidar and stereo Ikonos images," *International Journal of Remote Sensing*, vol. 29, no. 5, pp. 1277–1294, 2008.
- [51] E. Næsset, O. M. Bollandsås, and T. Gobakken, "Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data," *Remote Sensing of Environment*, vol. 94, no. 4, pp. 541–553, 2005.
- [52] S. A. Hall, I. C. Burke, D. O. Box, M. R. Kaufmann, and J. M. Stoker, "Estimating stand structure using discrete-return lidar: an example from low density, fire prone ponderosa pine

- forests," *Forest Ecology and Management*, vol. 208, no. 1–3, pp. 189–209, 2005.
- [53] Y. Kim, Z. Yang, W. B. Cohen, D. Pflugmacher, C. L. Lauver, and J. L. Vankat, "Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint lidar data," *Remote Sensing of Environment*, vol. 113, no. 11, pp. 2499–2510, 2009.
- [54] G. Wang, T. Oyana, M. Zhang et al., "Mapping and spatial uncertainty analysis of forest vegetation carbon by combining national forest inventory data and satellite images," *Forest Ecology and Management*, vol. 258, no. 7, pp. 1275–1283, 2009.
- [55] G. Wang, M. Zhang, G. Z. Gertner, T. Oyana, R. E. McRoberts, and H. Ge, "Uncertainties of mapping aboveground forest carbon due to plot locations using national forest inventory plot and remotely sensed data," *Scandinavian Journal of Forest Research*, vol. 26, no. 4, pp. 360–373, 2011.
- [56] M. A. Friedl, K. C. McGwire, and D. K. McIver, "An overview of uncertainty in optical remotely sensed data for ecological applications," in *Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications*, C. T. Hunsaker, M. F. Goodchild, M. A. Friedl, and T. J. Case, Eds., pp. 258–283, Springer, New York, NY, USA, 2001.
- [57] G. J. Nabuurs, B. van Putten, T. S. Knippers, and G. M. J. Mohren, "Comparison of uncertainties in carbon sequestration estimates for a tropical and a temperate forest," *Forest Ecology and Management*, vol. 256, no. 3, pp. 237–245, 2008.
- [58] D. L. Phillips, S. L. Brown, P. E. Schroeder, and R. A. Birdsey, "Toward error analysis of large-scale forest carbon budgets," *Global Ecology and Biogeography*, vol. 9, no. 4, pp. 305–313, 2000.
- [59] R. G. Congalton, "Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data," *Photogrammetric Engineering & Remote Sensing*, vol. 54, no. 5, pp. 587–592, 1988.
- [60] B. M. Steele, J. Chris Winne, and R. L. Redmond, "Estimation and mapping of misclassification probabilities for thematic land cover maps," *Remote Sensing of Environment*, vol. 66, no. 2, pp. 192–202, 1998.
- [61] G. Z. Gertner and B. T. Guan, "Using an error budget to evaluate the importance of component models within a large scale simulation model," in *Proceedings of the Mathematical Modeling of Forest Ecosystems Conference*, pp. 62–74, Frankfurt am Main, Germany, 1991.
- [62] M. J. W. Jansen, "Analysis of variance designs for model output," *Computer Physics Communications*, vol. 117, no. 1, pp. 35–43, 1999.
- [63] D. C. Collins and R. Avissar, "An evaluation with the Fourier amplitude sensitivity test (FAST) of which land-surface parameters are of greatest importance in atmospheric modeling," *Journal of Climate*, vol. 7, no. 5, pp. 681–703, 1994.
- [64] S. Fang, G. Z. Gertner, S. Shinkareva, G. Wang, and A. Anderson, "Improved generalized Fourier amplitude sensitivity test (FAST) for model assessment," *Statistics and Computing*, vol. 13, no. 3, pp. 221–226, 2003.
- [65] G. Gertner, X. Cao, and H. Zhu, "A quality assessment of a Weibull based growth projection system," *Forest Ecology and Management*, vol. 71, no. 3, pp. 235–250, 1995.
- [66] S. Fang, G. Z. Gertner, and A. A. Anderson, "Estimation of sensitivity coefficients of nonlinear model input parameters which have a multinormal distribution," *Computer Physics Communications*, vol. 157, no. 1, pp. 9–16, 2004.
- [67] G. Gertner, P. Parysow, and B. Guan, "Projection variance partitioning of a conceptual forest growth model with orthogonal polynomials," *Forest Science*, vol. 42, no. 4, pp. 474–486, 1996.
- [68] D. J. Downing, R. H. Gardner, and F. O. Hoffman, "An examination of response-surface methodologies for uncertainty analysis in assessment models," *Technometrics*, vol. 27, no. 2, pp. 151–163, 1985.
- [69] R. L. Iman and J. C. Helton, "An investigation of uncertainty and sensitivity analysis techniques for computer models," *Risk Analysis*, vol. 8, no. 1, pp. 71–90, 1988.
- [70] J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering and System Safety*, vol. 81, no. 1, pp. 23–69, 2003.
- [71] G. B. M. Heuvelink, *Error Propagation in Environmental Modeling*, Taylor & Francis, 1998.
- [72] M. Crosetto and S. Tarantola, "Uncertainty and sensitivity analysis: tools for GIS-based model implementation," *International Journal of Geographical Information Science*, vol. 15, no. 5, pp. 415–437, 2001.
- [73] G. Wang, S. Fang, S. Shinkareva, G. Gertner, and A. Anderson, "Spatial uncertainty in prediction of the topographical factor for the Revised Universal Soil Loss Equation (RUSLE)," *Transactions of the American Society of Agricultural Engineers*, vol. 45, no. 1, pp. 109–118, 2002.
- [74] G. Gertner, G. Wang, S. Fang, and A. B. Anderson, "Mapping and uncertainty of predictions based on multiple primary variables from joint co-simulation with Landsat TM image and polynomial regression," *Remote Sensing of Environment*, vol. 83, no. 3, pp. 498–510, 2002.
- [75] G. Z. Gertner, G. Wang, S. Fang, and A. Anderson, "Partitioning spatial model uncertainty when inputs are from joint simulations of correlated multiple attributes," *Transactions in GIS*, vol. 8, no. 4, pp. 441–458, 2004.
- [76] W. F. Laurance, A. K. M. Albernaz, P. M. Fearnside, H. L. Vasconcelos, and L. V. Ferreira, "Deforestation in Amazonia," *Science*, vol. 304, no. 5674, p. 1109, 2004.
- [77] E. F. Moran, E. Brondizio, and P. Mausel, "Monitoring secondary succession and land-use change in Amazonia," *Research and Exploration*, vol. 10, no. 4, pp. 458–476, 1994.
- [78] D. L. Skole, W. H. Chomentowski, W. A. Salas, and A. D. Nobre, "Physical and human dimensions of deforestation in Amazonia," *BioScience*, vol. 44, no. 5, pp. 314–322, 1994.
- [79] R. M. Lucas, M. Honzak, P. J. Curran et al., "Mapping the regional extent of tropical forest regeneration stages in the Brazilian Legal Amazon using NOAA AVHRR data," *International Journal of Remote Sensing*, vol. 21, no. 15, pp. 2855–2881, 2000.
- [80] D. A. Roberts, I. Numata, K. Holmes et al., "Large area mapping of land-cover change in Rondônia using multitemporal spectral mixture analysis and decision tree classifiers," *Journal of Geophysical Research D*, vol. 107, no. 20, pp. LBA 40-1–LBA 40-18, 2002.
- [81] G. M. Foody, "Remote sensing of tropical forest environments: towards the monitoring of environmental resources for sustainable development," *International Journal of Remote Sensing*, vol. 24, no. 20, pp. 4035–4046, 2003.
- [82] P. S. Chavez Jr., "Image-based atmospheric corrections—revisited and improved," *Photogrammetric Engineering and Remote Sensing*, vol. 62, no. 9, pp. 1025–1036, 1996.
- [83] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Assessment of atmospheric correction methods for Landsat TM data applicable to Amazon basin LBA research," *International Journal of Remote Sensing*, vol. 23, no. 13, pp. 1671–2651, 2002.
- [84] S. E. Franklin, M. A. Wulder, and M. B. Lavigne, "Automated derivation of geographic window sizes for use in remote

- sensing digital image texture analysis,” *Computers and Geosciences*, vol. 22, no. 6, pp. 665–673, 1996.
- [85] D. Chen, D. A. Stow, and P. Gong, “Examining the effect of spatial resolution and texture window size on classification accuracy: an urban environment case,” *International Journal of Remote Sensing*, vol. 25, no. 11, pp. 2177–2192, 2004.
- [86] J. C. Jenkins, D. C. Chojnacky, L. S. Heath, and R. A. Birdsey, “Comprehensive database of diameter-based biomass regressions for North American tree species,” Tech. Rep. NE-319, US Department of Agriculture, Forest Service, 2004.
- [87] J. E. Means, H. A. Hansen, G. J. Koerper, P. B. Alaback, and M. W. Klopsch, “Software for computing plant biomass—BIOPAK users guide,” Tech. Rep. PNW-GTR-340, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, 1994.
- [88] Q. Chen, “Airborne lidar data processing and information extraction,” *Photogrammetric Engineering and Remote Sensing*, vol. 73, no. 2, pp. 109–112, 2007.
- [89] M. A. Lefsky, D. Harding, W. B. Cohen, G. Parker, and H. H. Shugart, “Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA,” *Remote Sensing of Environment*, vol. 67, no. 1, pp. 83–98, 1999.
- [90] P. Gonzalez, G. P. Asner, J. J. Battles, M. A. Lefsky, K. M. Waring, and M. Palace, “Forest carbon densities and uncertainties from Lidar, QuickBird, and field measurements in California,” *Remote Sensing of Environment*, vol. 114, no. 7, pp. 1561–1575, 2010.
- [91] M. Zhang, G. Wang, G. Zhou, H. Ge, L. Xu, and Y. Zhou, “Mapping of forest carbon by combining forest inventory data and satellite images with co-simulation based up-scaling method,” *Acta Ecologica Sinica*, vol. 29, no. 6, pp. 2919–2928, 2009 (Chinese).
- [92] P. Goovaerts, *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, NY, USA, 1997.
- [93] J. R. Jensen, *Introductory Digital Image Processing*, Pearson Education, Cranbury, NJ, USA, 2005.
- [94] T. Zhao, Y. Wang, and H. Wang, “Image fusion based on multi-scale Kalman filtering,” in *Proceedings of the 2nd International Workshop on Education Technology and Computer Science (ETCS '10)*, vol. 3, pp. 207–215, Wuhan, China, 2010.
- [95] G. Wang, G. Z. Gertner, and A. B. Anderson, “Spatial-variability-based algorithms for scaling-up spatial data and uncertainties,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 9, pp. 2004–2015, 2004.
- [96] G. R. Larocquea, J. S. Bhattib, R. B. Boutina, and O. Cherstovc, “Uncertainty analysis in carbon cycle models of forest ecosystems: research needs and development of a theoretical framework to estimate error propagation,” *Ecological Modelling*, vol. 129, no. 3-4, pp. 400–412, 2008.
- [97] A. S. Fotheringham, C. Brunsdon, and M. E. Charlton, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons, Chichester, UK, 2002.