Light detection and ranging and hyperspectral data for estimation of forest biomass: a review

Qixia Man
Pinliang Dong
Huadong Guo
Guang Liu
Runhe Shi
Light detection and ranging and hyperspectral data for estimation of forest biomass: a review

Qixia Man, Pinliang Dong, Huadong Guo, Guang Liu, Runhe Shi

Abstract. Forests are one of the most important sinks for carbon. Estimating the amount of carbon stored in forests is a major task for understanding the global carbon cycle. From local to global scales, remote sensing has been extensively used for forest biomass estimation. With the availability of multisensor image data, fusion has become a valuable method in remote sensing applications. Light detection and ranging (LiDAR) can provide information on the vertical structure of forests, whereas hyperspectral images can provide detailed spectral information of forests. Effective fusion of LiDAR and hyperspectral data is expected to help extract important biophysical parameters of forests. However, it is still unclear as to how forest biophysical and biochemical attributes derived from hyperspectral data relate to structural attributes derived from LiDAR data. A summary of previous research on LiDAR-hyperspectral fusion for forest biomass estimation is valuable for further improvement of biomass estimation methods. A review on the status of hyperspectral data, LiDAR data, and the fusion of these two data sources for forest biomass estimation in the last decade is provided. Some future research topics and major challenges are also discussed. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JRS.8.081598]

Keywords: light detection and ranging; hyperspectral; data fusion; aboveground biomass.

1 Introduction

The complex structure and diverse material resources of a forest make it a perfect resource pool and biological gene bank for nature. Forest ecosystems have an irreplaceable role in improving the ecological environment and maintaining ecological balance. In addition, forest ecosystems are also an important component of the carbon cycle, in which forest ecosystems account for 80% of the aboveground carbon stocks and 40% of the underground carbon stocks.1,2 In recent years, carbon sequestration has been a hot topic in climate change studies3 and carbon balance estimation.4 From local to global scales, it is of increasing importance to quantify forest carbon exchange and stocks because of international policies to reduce greenhouse gases, such as the United Nations Framework Convention on Climate Change (UNFCCC),5 the Kyoto protocol,4 and the program for reducing emissions from deforestation and forest degradation (REDD).5,6
Forest biomass is one of the main biophysical parameters that describe the carbon content of the forest. Therefore, accurate estimation and assessment of forest biomass are important to quantify terrestrial carbon, control greenhouse gases, keep forest sustainably managed, and make policies which can reduce CO₂ mission. Field measurements are the traditional methods for forest biomass estimation. However, it is destructive, labor intensive, costly, time consuming, and sometimes inapplicable due to poor accessibility. Remote sensing has been frequently used as a practical and economical means for forest biochemical and biophysical parameters estimation, and a primary source for forest biomass estimation. At local to sub-regional scales, intermediate and fine resolution remote sensors have been used for biomass estimation, such as Landsat ETM+, SPOT, and WorldView-2. At regional to national scales, coarse resolution sensors were used such as National Oceanic and Atmospheric Administration (NOAA), Advanced Very High Resolution Radiometer (AVHRR), and Moderate Resolution Imaging Spectroradiometer (MODIS).

Hyperspectral sensors acquire hundreds of narrow bands of the electromagnetic spectrum from visible to short-wave infrared wavelengths which could provide detailed and continuous spectral information of forests, whereas light detection and ranging (LiDAR) has been known as a vital method for characterizing forest vertical structures, including height, volume, and biomass. With the availability of multisensor image data, effective fusion of the two complementary data sources can improve the estimation of forest biomass and other forest structure parameters.

Many previous papers have reviewed forest biomass estimation using remote sensing. Lu reviewed the potential and challenge of remote sensing-based biomass estimation and pointed out that biomass estimation is still a challenging task, especially for areas with complex forest structure and environmental conditions. Koch reviewed the status and future of three remote sensing technologies (LiDAR, synthetic aperture radar, and hyperspectral remote sensing) for forest biomass estimation. Treitz and Howarth published a review on hyperspectral remote sensing for estimation of forest biophysical parameters. Govender et al. provided a review of hyperspectral remote sensing and its application in vegetation and water resource studies. Adam et al. reviewed multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation. Lim et al. reviewed the recent research progresses of LiDAR on forest structure extraction, including canopy height, volume, and biomass. Van Leeuwen and Nieuwenhuis (2010) reviewed the methods and challenges in forest inventory using LiDAR. Frolking et al. reviewed the impacts of forest disturbance and recovery on aboveground biomass (AGB) and canopy structure in the context of space-borne remote sensing. Since the review by Koch, many papers, which will be included in this review, have been published on the fusion of hyperspectral and LiDAR data for forest biomass estimation.

Although research on biomass estimation using remote sensing has been investigated in the past decades, a comprehensive review on the fusion of LiDAR and hyperspectral data for forest biomass estimation is still lacking. It is unclear as to how forest biophysical and biochemical parameters derived from hyperspectral data relate to the structural attributes from LiDAR data. Many studies have involved the fusion of hyperspectral and LiDAR data in forest biomass application, which makes it possible to provide an overview of the techniques that have been used and to identify the aspects that still need further investigation. This review consists of the following six major parts: (1) overview of LiDAR and hyperspectral remote sensing systems and concepts of image fusion; (2) AGB estimation using LiDAR data; (3) AGB estimation using hyperspectral data; (4) fusion of LiDAR and hyperspectral data for biomass estimation; (5) discussion of current challenges and research needs; and (6) general conclusions.

2 Definitions

2.1 Overview of LiDAR Systems

LiDAR is an active remote sensing technology that determines distances based on the speed of light and the time required for an emitted laser to reach a target object. It can simultaneously capture vertical and horizontal forest structure and terrain morphology with high accuracy. The components of LiDAR systems include laser and scanning subsystems, a global position system (GPS), and radar.
system, and an inertial measurement unit. The basic ranging principle can be expressed with the following equation:

\[ R = \frac{ct}{2}, \]  

where \( R \) represents the distance from the sensor to the object; \( c \) is the speed of light; and \( t \) is the round-trip transmission time from the sensor to the measured target.

Dubayah and Drake\(^{33}\) summarized three characteristics that can be used to classify LiDAR systems for forestry applications, including: (1) the manner in which the return signal is recorded as either discrete return LiDAR, which typically includes the first, last, and several intermediate returns, versus the full-waveform LiDAR, which characterizes the returned energy in a continuous manner;\(^{34}\) (2) footprint size—small (a few centimeters) or large (tens of meters); and (3) sampling rate and scanning pattern.

### 2.2 Overview of Hyperspectral Systems

Hyperspectral remote sensing is a technology that acquires hundreds of narrow continuous spectral bands between 400 and 2500 nm, throughout the visible (400 to 700 nm), near-infrared (700 to 1000 nm), and short-wave infrared (1000 to 2500 nm) sections of the electromagnetic spectrum.\(^{35}\) It is also known as imaging spectroscopy or imaging spectrometry.

### 2.3 Methods of Forest AGB Estimation

Biomass is the dry weight of living and dead organisms.\(^{36,37}\) In forests, aboveground living biomass mainly includes the wood of canopy trees, vine, epiphyte, canopy leaf, understory, and groundcover biomass and would exclude all aboveground dead material.\(^{38}\)

In remote sensing applications, biomass normally refers to the AGB. Measurements of biomass can be divided into direct and indirect methods. As for direct estimations, based on the relationship between remote sensing response and the biomass, AGB can be estimated with different methods, such as multiple regression analysis, \( K \) nearest neighbor classification, neural network, and statistical ensemble methods.\(^{3,39-43}\) For indirect methods, the mean tree height or diameter at breast height are derived from remote sensing images.\(^{44-47}\) Using these parameters, biomass is obtained through allometric equations.\(^{48}\) For individual trees, biomass is added to get the AGB for the whole plot.

### 2.4 Concepts of Image Fusion

Image fusion can be defined as the combination of two or more different images into a new image using algorithms.\(^{49}\) Based on the stage at which fusion occurs, image fusion can be divided into three levels: pixel level; feature level; and decision level.\(^{50,51}\) (1) Pixel level: After coregistration, a series of raster data layers are directly added to an image which has more abundant and reliable information. Pixel-level fusion is the lowest level of image fusion, in which information synthesis and analysis based on the original information of various images are conducted. Its merit is keeping most of the original information which could provide subtle information, while the shortcomings are mainly reflected in the following four aspects: (a) large amount of information, long processing time, and high cost; (b) because of the uncertainty, insecurity, and instability of the original information, it requires high error correction capability during the fusion process; (c) low anti-interference ability; and (d) the fusion process requires a pixel calibration accuracy. (2) Feature level: Using segmentation procedures, remote sensing images are processed individually using feature extraction to generate unidentified features. The feature extraction processes mainly depend on the image elements such as shape, extent, and neighborhood.\(^{52}\) The extracted objects from multiple data sources are then fused for further assessment using statistical approaches such as artificial neural networks (ANN).\(^{50}\) By combining the features of the two datasets, the identification of features is carried out. Feature-level fusion could keep sufficient important information, achieve objective data compression and guarantee a certain accuracy. (3) Decision level: datasets are processed completely separately, and only the final
results are combined in geographic information systems. Apart from the three fusion levels, image fusion can also be applied to various data types, e.g., single sensor and multisensor. Figure 1 shows the three levels of data fusion.

3 ABG Estimation with LiDAR Data

LiDAR technology can provide horizontal and vertical information about the forest canopy which makes it one of the most applicable technologies in forest monitoring. LiDAR data has been used to derive tree height, estimate stem volume, and classify tree species. An overview of LiDAR for forest applications can be found in the papers by Lim et al., Hyyppä et al., and Mallet and Bretar. These papers reviewed recent research progress on the extraction of canopy height, estimation of ABG, and canopy volume from LiDAR as well as the status of small footprint, multiple point or full-wave LiDAR data for forest applications. Table 1 shows the previous studies of forest biomass estimation using LiDAR data.

Compared with other sensors, Airborne LiDAR data is much more effective for forest biomass estimation. Many studies have described the approaches of biomass estimation from LiDAR data, including single regression between LiDAR-derived height metrics, tree crown delineation and biomass, stochastic simulation and machine learning approaches. There are two levels of forest biomass estimation: individual-tree level and plot-level.

3.1 Individual Tree Level

With individual tree level forest biomass estimation, a crown height map (CHM) was produced from raw LiDAR point cloud data; then, individual trees were identified by applying algorithms to locate the maximum heights in CHM, such as local maxima filtering; finally, biomass was calculated by using regression between the tree height and biomass. There were numerous methods for local maxima filtering based on different search window sizes. Popescu identified the crown diameter from a CHM using local maxima filtering and quantified forest biomass by using regression algorithms which included crown width as a parameter. His study managed to explain 93% of the biomass using individual tree metrics. Recently, some researchers have developed advanced methods to identify individual trees. Bortolot and Wynne proposed a new individual tree-based algorithm for forest biomass estimation using small footprint LiDAR data. Kwak et al. proposed a watershed segmentation algorithm to identify individual trees. While biomass
estimation at the individual tree level is promising, there are several limitations: (1) for small footprint LiDAR, the laser pulse may miss the tree top. Therefore, to accurately extract individual trees, point density is an important factor. Previous studies have shown that a point density lower than \(4 \text{m}^{-2}\) might be insufficient for the identification of individual trees;\(^{59,69,71}\) (2) in CHM applications, crown overlapping was the major issue for tree crown identification; (3) individual tree level methods were usually tedious, time consuming, and expensive for field data collection and validation; (4) subdominant trees could not be detected using LiDAR returns data, and the aggregation of individual trees within a plot underestimated the entire plot biomass; and (5) in mixed forest such as tropical forest, it would be difficult to acquire enough species-allometric. The individual tree level methods for biomass estimation could be greatly improved by fusion with spectral data to classify tree species. It is difficult to identify tree species using LiDAR data alone because there are no published standards for radiometric calibration of LiDAR intensity data. Few studies have showed classification of tree species using LiDAR intensity data alone. Donoghue et al.\(^{72}\) studied tree species classification using LiDAR intensity data, but only two species were included in the study area. Even with the disadvantages listed above, this method could be greatly improved by fusion with other spectral remote sense data to identify tree species. Many studies have already illustrated the combined performance of LiDAR-derived parameters and spectral data for forest biomass estimation.\(^{73-75}\) Furthermore, an adaptable model and LiDAR-derived parameters is needed to automatically identify trees and then calculate the forest biomass based on tree height.\(^{63,68-70,76,77}\)

<table>
<thead>
<tr>
<th>References</th>
<th>Parameters</th>
<th>Regression</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson et al.(^{58})</td>
<td>Average of the three greatest laser heights, mean plot height (all pulses and canopy pulses), distance between the top of canopy and a point 2, 5 or 10 m aboveground</td>
<td>Two logarithmic equations</td>
<td>(R^2 = 0.55)</td>
</tr>
<tr>
<td>Means et al.(^{59})</td>
<td>LiDAR canopy height, quadratic mean canopy height, canopy reflectance sum</td>
<td>Allometric equations on DBH</td>
<td>(R^2 = 0.96)</td>
</tr>
<tr>
<td>Lefsky et al.(^{60})</td>
<td>Max/min canopy height, canopy cover, variability in the upper canopy surface, total volumes of foliage and empty space in canopy</td>
<td>Stepwise multiple regression</td>
<td>(R^2 = 0.86)</td>
</tr>
<tr>
<td>Drake et al.(^{61})</td>
<td>LiDAR canopy height, height of median energy, height/median ratio, ground return ratio</td>
<td>The tropical wet allometric equation</td>
<td>(R^2 = 0.93)</td>
</tr>
<tr>
<td>Nelson et al.(^{62})</td>
<td>Quadratic mean height of pulses in the forest canopy</td>
<td>Parametric linear regression, nonparametric linear regression</td>
<td>(R^2 = 0.66)</td>
</tr>
<tr>
<td>Popescu et al.(^{63})</td>
<td>Average/maximum crown diameter; maximum height</td>
<td>Regression models</td>
<td>(R^2 = 0.82) (Pine), (R^2 = 0.33) (hardwoods)</td>
</tr>
<tr>
<td>Zhao et al. 2009(^{64})</td>
<td>LiDAR-derived canopy height distributions (CHD) canopy height quartile functions (CHQ)</td>
<td>A linear functional model and an equivalent nonlinear model</td>
<td>(R^2 = 0.95)</td>
</tr>
<tr>
<td>Garcia et al.(^{11})</td>
<td>LiDAR height, intensity or height combined with intensity data</td>
<td>A stepwise regression</td>
<td>(R^2 = 0.85) (Pine), (R^2 = 0.70) (Spanish juniper), (R^2 = 0.90) (Holmoak)</td>
</tr>
<tr>
<td>Zhao et al.(^{65})</td>
<td>LiDAR composite metrics</td>
<td>Support vector machine and Gaussian processes (GP).</td>
<td>RMSE = 21.4 (40.5) Mg/ha</td>
</tr>
</tbody>
</table>
3.2 Plot Level

When the scope of the study is the plot, regression models were always used for forest biomass estimation based on LiDAR-derived statistics, such as height and canopy density metrics. Regression models usually use a simple/multiple and stepwise linear regression. One of the pioneering studies is Nelson et al. using two logarithmic equations in conjunction with six LiDAR-derived canopy measurements to estimate canopy volume and biomass. The model that used the mean plot height metric derived from all LiDAR pluses as an input parameter was identified as the best model. Previous studies indicated that tree height and crown diameter were highly correlated with biomass. Some recent studies focused on canopy height metrics that take into account structural data at multiple heights throughout the whole forest canopy, such as quadratic mean canopy height, height of median energy (HOME), height/median ratio (HTRT), simple ground return ratio (GRND), CH0.5-1.5 (as the proportion of laser hits above 0.5 m that belong to the height interval of 0.5–1.5 m), CH1.5-2.5, CH2.5-3.5, and CH3.5-4.5. Lefsky et al. employed simple linear models involving the parameters of maximum canopy height, mean canopy height, mean canopy eight, and quadratic mean canopy height and found that these parameters account for 80%, 70%, 73% and 80% of the variation in the ABG estimation, respectively. Drake et al. derived four metrics from LiDAR data: LiDAR canopy height, HOME, HTRT and a GRND. The four metrics were then input into a stepwise regression procedure to predict field-estimated AGB with $r^2$ (correlation coefficient) up to 0.93. Means et al. and Lefsky et al. employed almost the same data processing and analysis techniques to demonstrate the capabilities of Scanning LiDAR Imager of Canopies by Echo Recovery (SLICER) for AGB. Means et al. concluded that models using (1) mean canopy height, (2) quadratic mean canopy height and mean canopy height, and (3) the sum of the portion of waveform return from the canopy as predictors account for 90%, 94%, and 96% of variation in AGB estimation. Riggins et al. calculated multiple height percentiles (5, 15, 25, 35, 45, 55, 65, 75, 85, 95 and 100th percentiles) from small-footprint aerial LiDAR data and used a regression-tree model to get the forest biomass. Lefsky et al. applied a single regression including the parameters of LiDAR measured canopy structure in three different biomes and explained that these parameters account for 84% of the variance in AGB estimation ($P < 0.0001$). Lim and Treitz introduced quartile estimators (at 0, 25, 50, 75 and 100th percentiles) derived from airborne discrete return laser scanner data to estimate forest AGB. The coefficient of determination ($r^2$) for each model was >0.8. Zhao et al. proposed a scale-invariant forest biomass estimation method and obtained promising results. Rowell et al. estimated conifer-mixed forest biomass using both generalized and species-specific allometric models.

As far as data analysis is concerned, a wide variety of machine learning models have been effectively used in forest applications such as ANN, support vector regressions (SVRs), random forest (RF), cubist, bagging, and various algorithms based on decision trees (DTs) such as single and ensemble regression trees. Van Aardt et al. assessed a LiDAR-based, object-oriented approach to forest AGB models. The results showed that the new method was better than previous stand-based and plot-based approaches. Zhao et al. used two kernel machines, a support vector machine (SVM), and Gaussian processes to relate canopy characteristics to high-dimensional LiDAR metrics. Results illustrated that two machine learning models in conjunction with LiDAR metrics outperformed traditional approaches such as maximum likelihood classifier (MLC) and linear regression models. Gleason and Im used machine learning approaches to estimate forest biomass. In their study, four modeling techniques were used for moderately dense forest biomass estimation, including linear mixed effects regression, RF, SVR, and cubist. Results indicated that when estimating biomass at the plot level, the SVR modeling technique produced the most accurate biomass, whereas at the individual tree level, similar results were obtained by all models. The relationship between crown identification accuracy and biomass estimation accuracy is complex and needs to be further investigated. Figure 2 is a diagram summarizing some existing methods for biomass estimation using LiDAR data.

The space-borne full-wave laser system Ice, Cloud, and land Elevation satellite (ICESat)/geoscience laser altimeter system (GLAS) has been used to estimate vegetation height and forest biomass in large areas of the globe. These studies were mainly based on waveform decomposition. Duncanson et al. estimated forest canopy height from GLAS waveform.
metrics. Results indicated that GLAS waveforms could estimate forest height and AGB accurately, especially in flat areas with homogeneous forest conditions. Sun et al.86 also demonstrated that the vertical information derived from the GLAS waveform was very similar to that of the laser vegetation imaging sensor (LVIS) waveform which has been successfully used for forest structural parameters estimation. Although there are many studies on sub-boreal forest systems using ICESat/GLAS data, hardly any research exists on temperate, dry or tropical forests. This was mainly due to the sparse sampling density. Additionally, ICESat/GLAS data were often integrated with imaging sensors.34,88 Boudreau et al.88 combined multiple data sources to estimate biomass, including GLAS, SRTM, Landsat ETM+, airborne LiDAR, ground inventory plots, and vegetation zone maps. Their study showed that space-borne remote sensing measurements could be efficiently used for biomass estimation over large areas.

Increasingly, more studies focus on the detection of change in the AGB. By correlating LiDAR to forest inventory data, Jubanski et al.90 attempted to estimate AGB and its variability across large areas of tropical lowland forests. Huang et al.91 used NASA’s Laser Vegetation Imaging Sensor data for mapping biomass change. Meyer et al.92 also researched detecting tropical forest biomass dynamics from repeat LiDAR measurements. Næsset et al.93 detected change of forest biomass over an 11-year period using airborne LiDAR data. Englhart et al.94 quantified changes of tropical peat swamp forest with multitemporal LiDAR datasets. In the future, the fusion of LiDAR with other sensors is the tendency for biomass estimation. He95 fused LiDAR with SPOT-5 data to estimate coniferous forest biomass. Tsui et al.96 conducted a study to fuse multifrequency radar and discrete-return LiDAR measurements for AGB estimation in a coastal temperate forest. Tsui et al.97 focused on the fusion of LiDAR and radar for forest biomass estimation.

4 AGB Estimation with Hyperspectral Data

Hyperspectral imagery can provide numerous narrow bands. Compared with traditional multispectral imagery, hyperspectral can separate subtle changes of the biophysical parameters of forest.98 Because of this strength, hyperspectral imagery has been used for classifying vegetation species,23,99,100–108 extracting tree health information,109,110 deriving biophysical parameters,26,109,111,112 and estimating biomass.116 Pu and Gong117 used EO-1 Hyperion data to map forest crown closure (CC) and leaf area index (LAI). Three methods were used in their study, including: band selection, principal component analysis (PCA), and wavelet transform (WT). Results show that WT was the most effective method for mapping forest CC and LAI (mapping accuracy for CC = 84.9%, LAIMA = 75.39%). Zhang et al.109 explored a process-based method to retrieve leaf chlorophyll content from hyperspectral remote sensing imagery in complex forest canopies. Dalponte et al.107 evaluated two high spectral and spatial resolution hyperspectral sensors for tree species classification, where SVM, RF, and Gaussian maximum likelihood (ML) classification methods were used. Their results suggest that there was no
significant difference between SVM and RF classifiers and that the image spatial resolution had a strong effect on the classification accuracy.

Hyperspectral remote sensing data such as MODIS, Hyperion, AVHRR, AISA, HyMap, AVIRIS, and DAIS have also been frequently used for quantifying biomass from local to global scales. Among them, the spatial resolutions of MODIS and AVHRR are 250 m to 1 km and 1.1 km, respectively, and are usually used for forest biomass estimation at a regional or global scale. Dong et al. utilized normalized difference vegetation index (NDVI) derived from AVHRR images for forest biomass estimation. Because MODIS is a hyperspectral sensor, numerous band and index combinations can be used for regression modeling. At regional and global scales, MODIS and AVHRR are promising for forest biomass estimation. However, there are two major limitations of these sensors for forest biomass study: (1) because of low spatial resolution, it was not appropriate for small-area forest research; (2) because of long lapses, it was difficult to avoid the influence of cloud cover, especially the missing information of interesting areas. To make up for these limitations, fusion with other sensors might provide more accurate results in forest biomass estimation, especially the combination of spectral information and structural information.

For airborne hyperspectral sensors, most of the existing studies utilized different ways for forest biomass estimation, including raw spectral bands, regression analysis, and machine learning methods such as SVMs, end-member methods, ML classification, spectral angle mapper (SAM), RF, genetic algorithms, regression trees, ANN, and DT classifiers. Hansen and Schjoerring used NDVI in a linear regression analysis for estimating green biomass. The results showed that partial least square regression analysis may provide a useful tool when applied to hyperspectral data. Dong et al. used a multiple linear regression analysis to investigate the relationship between field estimates of AGB and various vegetation indices (VIs) acquired by hyperspectral data. Cho et al. used spectral indices and partial least squares regression to estimate green grass/herb biomass in a seminature landscape from airborne hyperspectral imagery. Their results showed that partial least squares regression combined with airborne hyperspectral imagery provides a better result than unvaried regression involving hyperspectral indices for grass/herb biomass estimation in Majella National Park, Italy. Gong et al. utilized Airborne Imaging Spectrometer Applications (AISA) hyperspectral imagery for forest biomass estimations for three forest crops. In their study, VIs and red edge positions (REPs) were derived from hyperspectral imagery and then regression models were used for the estimation of forest biomass. Results indicated that both VIs and REPs were effective for forest biomass estimation.

Some studies argued that a direct estimation of AGB cannot be achieved from hyperspectral imagery alone due to the weak relationship between vegetation biomass and spectral indices, especially in dense forests. However, it seems that fusion of hyperspectral and other types of remotely sensed data for biomass estimation is a promising area that needs further investigation.

5 AGB Estimation with Fusion of Hyperspectral and LiDAR Data

As no single data type could fulfill all requirements in AGB estimation, the complementary information of the fused data has obtained increasing interest. For tree-level biomass estimation, the species type is needed for application of species-specific allometric equations. Compared with multispectral data, hyperspectral data have shown to be promising in species classification and spectral attributes. Therefore, species classification maps derived from hyperspectral data could be used as an additional parameter to refine models based on LiDAR-derived metrics and intensity information. To date, many studies have been focused on the fusion of hyperspectral and LiDAR data for a variety of applications, including crown identification, AGB estimation, sagebrush distribution mapping, and fuel type mapping. Furthermore, numerous studies have particularly investigated forest species classification, vegetation type classification, and species-level discrimination. Additionally, the LiDAR/hyperspectral fusion has also been shown to increase the capacity of image segmentation and object-based classification.
Although numerous previous studies used the fusion of hyperspectral and LiDAR data in forest applications, most of the studies focused on forest species classification. For the fusion of hyperspectral and LiDAR data, there were three general levels as introduced in Sec. 2.4: pixel-level fusion, feature-level fusion, and decision-level fusion. Most of the existing studies used pixel-level fusion, which includes the following major steps: (1) preprocessing of hyperspectral and LiDAR data; (2) creating a combined dataset; (3) applying machine learning methods for forest species classification, such as SVM,129 RF,106,125 object-based classification,127 and Gaussian ML.103 Hill and Thomson100 fused Hymap images and airborne LiDAR data on the pixel level to classify woodland species. In their study, a digital canopy height model and the first two principal components from HyMap data were processed by a parcel-based unsupervised classification approach. Naidoo et al.106 also conducted an experiment to accurately classify and map individual trees at the species level in a savanna ecosystem. Hyperspectral- and LiDAR-derived parameters were grouped into seven predictor datasets and then an automated RF modeling approach was applied to classify eight common savanna tree species in the Greater Kruger National Park region, South Africa. The results showed that the most significant predictors were the NDVI, the chlorophyll b wavelength (chlorophyll includes chlorophyll a, chlorophyll b, chlorophyll c, chlorophyll d, and chlorophyll f. The range of chlorophyll b absorption wavelength is from 460 to 645 nm) and a selection of raw, continuum removed and SAM bands. Naidoo et al.106 also concluded that RF modeling with hybrid datasets yielded the highest accuracy for the eight tree species with an overall accuracy of 87.68%. There are other classification methods for information extraction from LiDAR–hyperspectral datasets, for example, object-oriented classification methods,127 RF classifier,125 SVM, and Gaussian ML.103 Feature-level fusion is a little different from pixel-level fusion. The major differences between them are in step (2) and step (3). In feature-level fusion, step (2) is used to derive LiDAR and hyperspectral metrics, such as canopy height and NDVI, and step (3) is used to form the combined dataset (combination of LiDAR and hyperspectral-derived metrics). Puttonen et al.129 used datasets consisting of two reflectance and two shape parameters to classify coniferous and deciduous trees and individual tree species through a SVM method, and the best classification result was 95.8% for the separation of coniferous and deciduous trees. Dalponte et al.125 extracted LiDAR (i.e., $H_{\text{min}}$, $H_{\text{max}}$) and hyperspectral metrics, respectively, and then a feature-selection technique was used to extract variables that have the most information. Several fused bands were formed including all hyperspectral bands, spectral bands + max height (LiDAR low density), spectral bands + max height (LiDAR high density), and spectral bands + height features (LiDAR height density). A nonlinear SVM and an RF classifier were used to classify tree species. When combined with either hyperspectral or multispectral data, high-density LiDAR data could provide more information for tree species than low density LiDAR data. As few studies used decision-level fusion of multiple data, decision-level fusion will not be discussed here. Figure 3 shows the preprocessing of LiDAR and hyperspectral data before fusion. Figure 4 illustrates the pixel-level fusion of hyperspectral and LiDAR for forest species distribution mapping. Table 2 shows the previous studies of forest biomass estimation using the fusion of hyperspectral and LiDAR data.

In addition to species classification, hyperspectral and LiDAR fusion data were also used for forest biomass estimation. Hyperspectral sensors were used for species classification, and then LiDAR data were used to perform biomass estimation of each classified species. The fusion of LiDAR and hyperspectral data included the following steps: (1) preprocessing of each
data (i.e., removing clear outlier points from LiDAR data, atmospheric correction of hyperspectral data, co-register LiDAR and hyperspectral data); (2) deriving LiDAR and hyperspectral metrics (e.g., canopy height model from LiDAR data, height metrics from CHM, varieties of vegetation index); (3) classifying forest species using hyperspectral data and creating the combined dataset; and (4) estimating forest biomass using regression methods or a species-specific allometric. In a complex forest, Lucas\textsuperscript{74} used airborne LiDAR data and CASI hyperspectral image data to automatically identify trees and estimate their biomass. In their study, a Jackknife linear regression was also used to estimate plot-level forest biomass using six LiDAR strata heights and crown cover. Results showed that the Jackknife linear regression method was more robust for forest biomass estimation and showed a closer relationship with plot-sale ground data ($r^2 = 0.90$, RSE = 11.8 Mg/ha, $n = 31$). The fusion study also required methods which could deal with high-density LiDAR data and complex forests. Clark et al.\textsuperscript{8} estimated tropical forest biomass using the fusion of hyperspectral and small-footprint, discrete-return LiDAR data. LiDAR metrics (i.e., mean height, maximum height) and hyperspectral metrics (i.e., NDVI) were retrieved, respectively. Then single and two-variable linear regression analyses were used to relate plot-scale LiDAR and hyperspectral metrics to field-derived biomass from all plots. The results showed that the best model was created using all 83 biomass plots including two LiDAR height metrics, plot-level mean height, and maximum height with an $r^2$ of 0.90 and RMSE of 38.3 Mg/ha. However, analysis combined for plantation plots had the most accuracy field data with $r^2$ increased to 0.96 and RMSE of 10.8 Mg/ha ($n = 32$). Swatantran et al.\textsuperscript{25} used LVIS metrics, AVIRIS spectral indices, multiple endmember spectral mixture analysis fraction, and linear and stepwise regressions to map species-specific biomass and stress at landscape scale. The results showed that the accuracy of prediction by LVIS after species stratification of the field data was up to $r^2 = 0.77$, and RMSE = 70.12 Mg/ha. The results also suggest that LiDAR data were better for biomass estimation, whereas the hyperspectral data were used to adjust the LiDAR predictive models for species. Latifi et al.\textsuperscript{132} fused LiDAR-hyperspectral data for forest structure modeling, including models of stem density and aboveground total biomass. Their results indicated that LiDAR

**Fig. 4** Pixel-level fusion of hyperspectral and LiDAR for forest species distribution mapping, CVPs: canopy volume profiles; PRI: photochemical reflectance index; CRI: carotenoid reflectance index; PCA: principal component analysis; MNF: minimum noise fraction; NDVI: normalized difference vegetation index.
provided most of the information for the combination, whereas hyperspectral data only made a modest contribution. Figure 5 shows two methods of AGB estimation by fusion of LiDAR-hyperspectral dataset.

The above studies showed good results of data fusion for forest biomass estimation, yet there are still many aspects which need refinement for improving biomass estimation through fusion of hyperspectral and LiDAR data. The major limitations are: (1) a very high LiDAR sampling rate is required to correctly identify treetops;133,134 (2) forest biomass estimation was affected by the plot size for some sensors; (3) at the individual tree level, if the spatial resolution is coarser than that of individual crown areas, it would be difficult to identify individual trees, especially in complex forests; and (4) in complex forests, subcanopy trees and stems are usually omitted from the traditional canopy height models, which would underestimate the biomass.

Table 2 List of references on the fusion of hyperspectral/LiDAR for biomass estimation and forest species classification.

<table>
<thead>
<tr>
<th>References</th>
<th>Parameters</th>
<th>Methods</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hill et al.100</td>
<td>PCs 1 and 2 of HyMap and Digital Canopy Height Model (DCHM)</td>
<td>Segmentation and ISODATA</td>
<td>Mapping of woodland species composition and structure</td>
</tr>
<tr>
<td>Geerling et al.51</td>
<td>Pixel-level fusion of 10 CASI bands and 6 LiDAR texture bands (min, max, mean, median, SD, and range)</td>
<td>MLC</td>
<td>Classification of floodplain vegetation</td>
</tr>
<tr>
<td>Sugumaran et al.127</td>
<td>4 band Quickbird image with and without LiDAR; 24-band AISA hyperspectral image with and without LiDAR</td>
<td>Object-oriented classification</td>
<td>Tree species identification in an urban environment</td>
</tr>
<tr>
<td>Dalponte et al.103</td>
<td>25 hyperspectral bands, elevation and intensity of the first LiDAR return; 40 hyperspectral bands, elevation and intensity of the first LiDAR return; 126 hyperspectral bands, elevation and intensity of the first LiDAR return.</td>
<td>SVM classifier; GML-LOOC; K-NN</td>
<td>Classification of complex forest</td>
</tr>
<tr>
<td>Jones et al.104</td>
<td>AISA data (40 bands); 40 AISA bands and CHM; 40 AISA bands and 4 CVPs; 40 AISA bands, and 2 CVPs</td>
<td>A multiclass SVM classifier</td>
<td>Mapping species distribution</td>
</tr>
<tr>
<td>Onojeghuo et al.130</td>
<td>Spectral data only; spectrally compressed data: PCA transformed data and MNF transformed data; SSPCA transformed data; texture combined data: MNF 1-15-GLCM45; (Optical MNF + texture) and LiDAR derived measures; Optical SSPCA image and LiDAR derived measures.</td>
<td>Maximum likelihood classifier (MLC)</td>
<td>Optimize the use of LiDAR and hyperspectral data for reedbed habitats mapping</td>
</tr>
<tr>
<td>Naidoo et al.106</td>
<td>Height only dataset; Hyperspectral indices only dataset (CRI, NDVI, PRI, red edge NDVI); Height and indices dataset (Height, CRI, NDVI, PRI, red edge NDVI); Raw bands datasets (all 72 CAO raw bands); SAM selected bands dataset (Band Add-On); Nutrient and Leaf Mass Bands Dataset</td>
<td>Random forest module</td>
<td>Classification of savanna tree species</td>
</tr>
<tr>
<td>Anderson et al.131</td>
<td>24 AVIRIS MNF bands and 4 LVIS metrics (RH25, RH75, RH50 and RH 100)</td>
<td>Stepwise mixed linear regression techniques</td>
<td>LiDAR-hyperspectral fusion for inventory of a northern temperate forest</td>
</tr>
</tbody>
</table>

Note: CVPs, canopy volume profiles; PRI, photochemical reflectance index; CRI, carotenoid reflectance index.
6 Challenges and Future Research

LiDAR provides precise height information which could extract vertical and horizontal information of the forest and is less sensitive to saturation when compared with other sensors. Hyperspectral data can provide detailed spectral information which can be used to effectively classify forest species. The findings on forest biomass estimation in the last decades are encouraging, and there is a growing interest for hyperspectral, LiDAR, and fusion of LiDAR and hyperspectral data for forest biomass estimation. However, there are still many limitations that should be taken into account: (1) because a forest is a complex ecosystem, many factors may impact the estimation of forest biomass when using LiDAR and hyperspectral imagery, such as spectral and spatial resolution, co-registration of data from different sources (data from different platforms, different dates or with different resolutions), LiDAR point density, fusion framework, classification algorithm, allometric equations, plot-size, study area, stem density, canopy volume, and height; (2) although the Global Climate Observing System has suggested some accuracy levels, there is still no clear standard for forest biomass estimation. In addition, it is still unclear as to how forest biophysical and biochemical attributes derived from hyperspectral data relate to structural attributes from LiDAR data. The current methods normally add LiDAR elevation information as an extra band of hyperspectral data on the pixel level, or extract metrics, respectively, and classify fused images using different algorithms. In some other studies, hyperspectral and LiDAR data were processed separately. For example, hyperspectral data were used to classify forest species, and LiDAR data were used to identify individual trees, then specific allometric equations were used for forest biomass; (3) it is difficult to estimate forest biomass in forest with complex structures, especially tropical forests. In addition, high-quality ground truth data for validation is still lacking. For example, it is unrealistic to classify all tree species with hyperspectral data and classify all species-specific allometric equations; (4) for tree-level biomass estimation, although hyperspectral and LiDAR data have been successfully used, it is still problematic, because it is difficult to allocate individual trees to the proper species. Using a spatial resolution coarser than that of individual tree crown size, it is difficult to classify and attribute individual trees. In this case, tone can assume that there are several species within a pixel area, especially in heterogeneous forests. Using a spatial resolution finer than that of individual tree crown size, it is also difficult to classify and attribute individual trees. As in this case, an individual crown may contain several pixels with different illumination.

Although there are limitations as described above, the fusion of hyperspectral and LiDAR is still promising for forest biomass estimation, especially with the emergence of new sensors and new fusion methods.
6.1 New Sensors

Many dedicated missions will be launched.\textsuperscript{136,137} For example, forthcoming hyperspectral missions will also be oriented toward image fusion, such as Environmental Mapping and Analysis Program (EnMap), PRerorsore IperSpectrale of the application mission (PRISMA), Medium Resolution Imaging Spectrometer (MERIS), and Hyperspectral Infrared Imager (HyspIRI). A new processing and fusion framework will also need to deal with these new sensors.

6.2 New Fusion Methods

More advanced modeling methods are needed to quantify the biophysical characteristics of forest.\textsuperscript{120} Compared with pixel-level classification, an object-based classification method is more accurate when segmenting tree crown.\textsuperscript{74,138–140} Therefore, it is a reasonable step to implement simple crown segmentation algorithms to generate crown objects for future analysis, especially in a low density forest. However, the selection of optimal parameters remains a challenge in the case of a high-density forest area. The new fusion methods include the fusion of different sensors and the fusion of different modeling methods: (1) fusion of different sensors, including electro-optical, radar and LiDAR sensors; (2) fusion of direct and indirect methods. For example, biomass extracted from high resolution remotely sensed images over some study areas can be used in the less computationally demanding indirect regression methods; and (3) fusion of different modeling methods, such as SVM, RF, and object-based classification. The combination of these methods might improve the classification accuracy.

There are several questions that need further investigation: (1) can alteration in the fusion framework (pixel-level, feature-level and decision-level) of hyperspectral and LiDAR data lead to different findings? (2) Can feature selection methods be further improved to increase the stability and accuracy of tree species classification from hyperspectral data? (3) Can additional LiDAR measures be connected to forest structure and canopy architecture?

7 Conclusions

In this paper, the application of LiDAR data, hyperspectral data, and LiDAR/hyperspectral data fusion for forest biomass estimation and the current difficulties and prospects were reviewed. Generally, using remote sensing data, forest biomass could be directly estimated with different methods, including multiple regression analysis, $K$ nearest-neighbor and neural network. Forest biomass could also be indirectly estimated. For example, canopy parameters such as crown diameter are derived from remotely sensed data and finally, allometric equations were used for forest biomass estimation.

Data from LiDAR systems such as ICESat/GLAS, SLICER, LVIS and discrete return LiDAR have been used for forest biomass estimation. Biomass estimation from airborne LiDAR metrics is more accurate than that from a satellite-borne GLAS instrument. The sparse sampling density of GLAS and the limited spatial extent of airborne LiDAR can be integrated with data from other imaging sensors. The methods of biomass estimation using LiDAR data include single regression between LiDAR-derived height metrics, tree crown delineation and biomass, stochastic simulation, and machine learning approaches. It is expected that more studies will focus on the variability of AGB estimated from LiDAR data and will multisensor measurements to estimate biomass, including LiDAR-radar fusion and LiDAR-hyperspectral fusion.

Hyperspectral data have been used for classifying vegetation species, monitoring health status, and deriving biophysical parameters. The main methods for information extraction from hyperspectral data include SVMs, ML, ANN, RF, DT, and end-member methods. Due to the loss of correlation, a few studies have used hyperspectral data for biomass estimation, especially in forests with high biomass and complex environmental conditions. Hyperspectral data fused with other types of remotely sensed data will be a promising research area for forest biomass estimation.

Since LiDAR data can characterize the vertical structure of forests with high accuracy and hyperspectral data can provide detailed spectral information on biophysical parameters, the fusion of LiDAR and hyperspectral data has received increased attention. There are three levels
of data fusion using LiDAR and hyperspectral data, including pixel level, feature level, and decision level. LiDAR/hyperspectral data fusion has been applied to AGB estimation and tree species classification. However, most of the studies were focused on forest species classification. As forest is a complex ecosystem, many factors may impact the estimation of forest biomass when using LiDAR and hyperspectral imagery, such as spectral and spatial resolution, co-registration of data from different sources (data from different platforms, different dates or with different resolutions), LiDAR point density, fusion framework, classification algorithm, allometric equations, plot-size, study area, stem density, canopy volume, and height. In addition, new systems which occupy complementary sensors on the same platform are emerging, such as Goddard’s LiDAR, Hyperspectral and Thermal (GLiHT) airborne imager which shows promising prospects for the fusion of hyperspectral and LiDAR data. In the near future, new satellite-based sensors will be launched, including the Medium Resolution Imaging Spectrometer (MERIS). These developments will provide more opportunities for multisensor fusion. Besides the fusion of hyperspectral-LiDAR data for forest biomass, there will be more fusions between multiple sensors and methods, including the fusion of different sensors (fusion of space-borne LiDAR, airborne LiDAR data, and hyperspectral data for upscaling forest biomass estimation), the fusion of direct and indirect methods for forest biomass (biomass extracted from high resolution remote sensing can be used in the less computationally demanding indirect regression methods), and the fusion of different machine learning methods.

Acknowledgments

This research was funded by the ABCC Program of the National Natural Science Foundation of China (Project 41120114001), and the National Key Technology R&D Program (Projects 2012BAC16B01 and 2012BAH27B05). The first author would like to thank the China Scholarship Council for the support of two years’ study in the Department of Geography, University of North Texas. The authors would like to thank anonymous reviewers for their helpful comments and suggestions.

References

Man et al.: Light detection and ranging and hyperspectral data for estimation of forest biomass...


Qixia Man is a PhD student at the East China Normal University. She studied at the University of North Texas as a visiting student from 2012 to 2014. Her research interests include digital image processing, LiDAR, and hyperspectral remote sensing.

Pinliang Dong is an associate professor in the Department of Geography, University of North Texas (UNT). He received his BSc degree from Peking University, China, MSc degree from the Institute of Remote Sensing Applications (IRSA), Chinese Academy of Sciences, and PhD degree from the University of New Brunswick, Fredericton, New Brunswick, Canada. His research interests are mainly in GIScience and remote sensing.

Huadong Guo is an academician and research professor at the Chinese Academy of Sciences. He graduated from the Geology Department at Nanjing University in 1977 and received his MSc degree from the Graduate University of the Chinese Academy of Science (CAS) in 1981. He is a guest professor of eight universities in China. His research interests include radar remote sensing, applications of Earth-observing technologies to global change, and Digital Earth.
Guang Liu received his BS/MS in physics from TsingHua University China in 1999/2002 and received his PhD degree from the Institute of Remote Sensing Applications of the CAS in 2008. He has been working with the Mathematic Geodesy and Positioning, Delft University of Technology of the Netherlands from 2006 to 2007 as a visiting researcher. He is an associate professor of Institute of Remote Sensing and Digital Earth, CAS. His work is focused on the study of the feasibility and potential applications of SAR image time series analysis.

Runhe Shi received his BS in East China Normal University in 2001, and received his PhD degree from Institute of Geographic Sciences and Natural Resources Research of the CAS. His research mainly focused on research and application of quantitative remote sensing algorithm.