21 LiDAR Remote Sensing of Vegetation Biomass

Qi Chen

CONTENTS
21.1 Introduction........................................................................................................................399
21.2 Remote Sensing of Vegetation Biomass Using Different LiDAR Systems..............401
  21.2.1 Airborne Small-Footprint Discrete-Return Scanning LiDAR..........................401
    21.2.1.1 Area-Based Methods ..............................................................................402
    21.2.1.2 Individual Tree Analysis Methods .......................................................405
  21.2.2 Airborne Small-Footprint Discrete-Return Profiling LiDAR .........................405
    21.2.2.1 Useful Profiling LiDAR Metrics for Biomass Estimation ..................406
    21.2.2.2 Challenges of Profiling LiDAR for Biomass Mapping ....................406
  21.2.3 Airborne Medium- and Small-Footprint Waveform LiDARs ......................407
    21.2.3.1 Useful Waveform Metrics for Biomass Estimation .........................407
    21.2.3.2 Importance of Correcting Attenuated Waveforms ..........................407
  21.2.4 Satellite Large-Footprint Waveform LiDAR ..............................................408
    21.2.4.1 Challenges of Large-Footprint Satellite LiDAR for Biomass Mapping .................................................................................................................................408
  21.2.5 Ground-Based Discrete-Return Scanning LiDAR ........................................409
    21.2.5.1 Challenges of Ground-Based LiDAR for Biomass Estimation .............409
  21.3 General Challenges and Future Directions .........................................................409
    21.3.1 Fusion of LiDAR with Other Remotely Sensed Data ..........................409
    21.3.1.1 Fusion with Optical Imagery ...............................................................409
    21.3.1.2 Fusion with Radar .................................................................................410
  21.3.2 Generalizability of Biomass Models ..........................................................410
  21.3.3 Biological and Ecological Interpretation of Biomass Models ..................412
  21.4 Conclusions .................................................................................................................412
References .................................................................................................................................413

21.1 Introduction

Accurate estimates of vegetation biomass are critical for calibrating and validating biogeochemical models (Hurtt et al. 2010), quantifying carbon fluxes from land use and land cover change (Shukla et al. 1990; Houghton et al. 2001), and supporting the United Nations Framework Convention on Climate Change (UNFCCC) program to reduce deforestation and forest degradation (Reducing Emissions from Deforestation and Forest Degradation) (Asner 2009). For instance, it was argued that at least half of the uncertainty in the estimates of emissions of carbon from land use change results from uncertain estimates of biomass density (Houghton 2005; Houghton et al. 2009).
Vegetation biomass is commonly defined as the amount of organic matter in living and dead plant materials expressed as dry weight per unit area (Brown 1997). Estimates of biomass over large areas have been determined by a wide range of approaches, including interpolating measurements from a relatively small number of plots typically for ecological studies (Houghton et al. 2001; Malhi et al. 2006); averaging measurements from a large number of forest inventory plots determined by rigorous statistical sampling (Brown and Schroeder 1999; Jenkins et al. 2001); stratifying and multiplying field measurements with categorical land cover or biome maps (Fearnside 1997; Gibbs et al. 2007); geographic information system modeling with environmental variables such as soil, elevation, and precipitation (Olson et al. 1983; Brown et al. 1993; Iverson et al. 1994; Brown and Gaston 1995; Gaston et al. 1998); process-based biogeochemical modeling (Potter 1999; Potter et al. 2001); and direct mapping of the continuous biomass field based on remotely sensed imagery and relevant vegetation products (Houghton et al. 2001; Blackard et al. 2008). Among them, the remote sensing–based approach is appealing because it can provide spatially explicit estimates of the actual biomass at each pixel location, instead of only the average/total biomass within a given inventory unit (such as county, state, and country) (Jenkins et al. 2001) or the potential/equilibrium biomass (Iverson et al. 1994; Potter et al. 2001). The fine-grained biomass maps generated from remotely sensed data will enhance our understanding of contemporary environmental changes since natural and human disturbances can occur at scales as small as 10–1000 m (Houghton 2005; Houghton et al. 2009).

Remote sensing of biomass has been based on reflectance (e.g., Blackard et al. 2008), vegetation indices (e.g., Myneni et al. 2001; Tan et al. 2007), leaf area index (e.g., Feng et al. 2007), vegetation cover (e.g., Houghton et al. 2001), or their combinations (e.g., Blackard et al. 2008). Conventionally, the remotely sensed data used are optical imagery and radio detecting and ranging (radar), such as data from Landsat (e.g., Foody et al. 2003), Moderate Resolution Imaging Spectroradiometer (e.g., Blackard et al. 2008), advanced very high resolution radiometer (e.g., Myneni et al. 2001), hyperspectral imagery such as Hyperion (e.g., Thenkabail et al. 2004), high-spatial-resolution imagery such as IKONOS and QuickBird (e.g., Clark et al. 2004), and synthetic aperture radar (SAR) (e.g., Le Toan et al. 1992; Kasischke et al. 1997). The main challenge for optical and radar data is the insensitivity or saturation of remotely sensed signals at moderate to high biomass levels (Waring et al. 1995). For example, Le Toan et al. (2004) summarized that the saturation of AirSAR and E-SAR occurs at around 30, 50, and 150–200 Mg/ha at C, L, and P bands, respectively, over different types of temperate, boreal, and tropical forests. Relatively new radar technologies such as InSAR (Balzter et al. 2007; Treuhaft et al. 2009; Brown et al. 2010) and PolInSAR (Cloude and Papathanassiou 1998; Mette et al. 2003; Garestier et al. 2008; Neumann et al. 2010) can retrieve canopy height (CH), which can be used to improve biomass mapping. However, their performance over complex forests of high biomass is to be rigorously tested.

A promising new approach for estimating biomass is light detection and ranging (LiDAR), which does not saturate even at very high biomass levels (>1000 Mg/ha) (Means et al. 1999). LiDAR provides unprecedented accuracy in measuring vegetation structures such as height (Nelson et al. 2000; Lefsky et al. 2005a; Chen et al. 2006; Dolan et al. 2009; Chen 2010b), crown size (Riano et al. 2003; Chen et al. 2006), basal area (Nelson et al. 1997; Gobakken and Naesset 2005; Chen et al. 2007a), stem volume (Holmgren et al. 2003; Hollaus et al. 2007; Chen et al. 2007a), and vertical profile (Harding et al. 2001; Zhao et al. 2009), all of which can be related to biomass. The potential of LiDAR for vegetation biomass retrieval was demonstrated more than two decades ago using airborne profiling LiDAR systems (Nelson et al. 1984, 1988a). During the last decade, the body of literature on LiDAR remote sensing of biomass grew rapidly, mainly due to the advances made in sensors,
LiDAR Remote Sensing of Vegetation Biomass

Recent reviews on LiDAR remote sensing of biomass can be found in the works of Koch (2010), Gleason and Im (2011), Lu et al. (2012), and Rosette et al. (2012). This chapter summarizes past research advances on biomass estimation using airborne, spaceborne, and ground-based LiDAR systems and discusses key research questions and challenges and future directions.

21.2 Remote Sensing of Vegetation Biomass Using Different LiDAR Systems

LiDAR measures distances between a sensor and objects based on the time lags between transmitting light amplification by stimulated emission of radiation (laser) beams from the sensor and receiving signals reflected from the illuminated objects. The distances derived from LiDAR, combined with the position of the sensor and the direction of the laser beam, uniquely determine the three-dimensional (3D) coordinates of the objects illuminated. The errors of 3D coordinates vary with a myriad of factors such as laser range sampling interval, global positioning system (GPS) positioning, inertial measurement unit, flying altitude, and surface reflectivity. But, in general, the vertical precision of position is on the order of decimeters for airborne and satellite LiDAR (Zwally et al. 2002; Chen 2010a).

LiDAR remote sensing systems can be distinguished based on the way in which returned signals are recorded (discrete return or waveform), scanning patterns (profiling or scanning), platforms (airborne, spaceborne, or ground based), and footprint sizes (small footprint: ~1 m or smaller, medium footprint: ~10–30 m, or large footprint: ~50 m or larger) (Lu et al. 2012). The most common configurations of LiDAR systems are airborne small-footprint discrete-return scanning LiDAR, airborne small-footprint discrete-return profiling LiDAR, airborne medium-footprint waveform LiDAR, and satellite large-footprint waveform systems. Ground-based systems and airborne small-footprint waveform systems are also emerging. The key findings and issues on the use of these LiDAR systems for biomass retrieval are summarized and discussed in Sections 21.2.1 through 21.2.5.

21.2.1 Airborne Small-Footprint Discrete-Return Scanning LiDAR

An airborne small-footprint discrete-return scanning LiDAR system typically includes three key components: (1) a laser range finder to measure distance, (2) a differential GPS system to measure the 3D position of the airplane/sensor, and (3) an inertial navigation system to measure the attitude of the airplane. Such systems can record the information of one or multiple returns of a transmitted laser pulse. The information of a laser return includes, at a minimum, 3D coordinates and usually intensity, laser pulse angle, the number of the return (first, last, or other), and so on, among which the 3D coordinates are the key information. So, the data acquired by a discrete-return LiDAR system is also called a point cloud. Over the last decade, airborne small-footprint discrete-return scanning LiDAR has experienced thrilling developments in hardware and software so that today data can be acquired with a much lower cost and be processed more effectively (Chen 2007, 2009; Chen et al. 2007b). For example, the laser pulse repetition rate has reached 500 kHz, whereas it was only 10–20 kHz a decade ago. Some earlier systems only record either the first or the last return (Omasa et al. 2003), whereas most current systems can record both
first and last returns and even up to four returns (Table 21.1). As a mainstream LiDAR system for terrain mapping and various environmental applications, airborne discrete-return scanning LiDAR has been used in a large number of studies for mapping biomass, mainly using two kinds of approaches: (i) area-based and individual tree–based methods.

### 21.2.1.1 Area-Based Methods

Area-based methods develop statistical models to relate biomass with metrics derived from a LiDAR point cloud at the plot or stand level and apply the models over the whole study area. Such methods have been used to estimate the biomass of boreal forests (Lim et al. 2003; Naesset and Gobakken 2008; St-Onge et al. 2008; Thomas et al. 2006), temperate forests (Hall et al. 2005; Bortolot 2006; van Aardt et al. 2006; Zhao et al. 2009; Chen et al. 2012; Gleason and Im 2012; Lu et al. 2012), savanna woodlands (Lucas et al. 2006), and tropical forests (Asner 2009; Asner et al. 2009; Kronseuder et al. 2012). The development of statistical models requires field data for calibration and validation. Almost all of these studies were conducted at the landscape level; so the numbers of field plots or stands used were relatively small (~20–60) except in a few studies such as the one by Naesset and Gobakken (2008), which used more than 1000 forest inventory plots to estimate the biomass for 10 different forest areas in south Norway. Most of the field plots had a fixed radius of ~10–50 m, whereas some used plots of variable radii (e.g., van Aardt et al. 2006). Besides estimating total aboveground biomass (AGBM), some also modeled the biomass

### TABLE 21.1

Summary of Different LiDAR Systems

<table>
<thead>
<tr>
<th>Type</th>
<th>Characteristics</th>
<th>Examples of Developers (Systems)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airborne small-footprint</td>
<td>High horizontal and vertical position accuracy (on the order of 1 cm); footprint</td>
<td>Optech (ALTM);</td>
</tr>
<tr>
<td>discrete-return scanning</td>
<td>size less than 1 m (at the flying altitude of 1000 m); PRF up to a few hundred</td>
<td>Leica (ALS);</td>
</tr>
<tr>
<td>LiDAR</td>
<td>kHz; usually records both first and last returns and up to four returns</td>
<td>Riegl (LMS)</td>
</tr>
<tr>
<td>Airborne small-footprint</td>
<td>The same as above, except that waveform can be recorded with the addition of a</td>
<td>Optech (ALTM);</td>
</tr>
<tr>
<td>waveform scanning</td>
<td>waveform digitizer</td>
<td>Leica (ALS);</td>
</tr>
<tr>
<td>LiDAR</td>
<td></td>
<td>Riegl (LMS, VQ)</td>
</tr>
<tr>
<td>Airborne small-footprint</td>
<td>Horizontal positional accuracy is low (on the order of 10 m); footprint size</td>
<td>NASA (PALS)</td>
</tr>
<tr>
<td>discrete-return profiling</td>
<td>less than 1 m (at the flying altitude of 1000 m); PRF of a few kHz; usually</td>
<td></td>
</tr>
<tr>
<td>LiDAR</td>
<td>records only first return and/or last return</td>
<td></td>
</tr>
<tr>
<td>Airborne medium-footprint</td>
<td>High horizontal and vertical position accuracy (on the order of decimeters);</td>
<td>NASA (SLICER and LVIS)</td>
</tr>
<tr>
<td>footprint waveform LiDAR</td>
<td>footprint size 10–30 m with a flying altitude of ~10 km; PRF of tens to hundreds</td>
<td></td>
</tr>
<tr>
<td></td>
<td>of Hz; records the received waveform</td>
<td></td>
</tr>
<tr>
<td>Satellite large-footprint</td>
<td>High vertical position accuracy (on the order of 10 cm); nominal footprint size</td>
<td>NASA (GLAS)</td>
</tr>
<tr>
<td>waveform LiDAR</td>
<td>60 m with a flying altitude of ~600 km; PRF of 40 Hz; records the received waveform</td>
<td></td>
</tr>
<tr>
<td>Ground-based small-footprint</td>
<td>Very high horizontal and vertical position accuracy (on the order of millimeters</td>
<td>Leica (ScanStation);</td>
</tr>
<tr>
<td>discrete-return scanning</td>
<td>depending on the distance); PRF of a few to tens of kHz; records the first or</td>
<td>Optech (ILRIS);</td>
</tr>
<tr>
<td>LiDAR</td>
<td>last return; range varies from a few hundred to a few thousand meters</td>
<td>Riegl (LMS, VZ)</td>
</tr>
</tbody>
</table>

**Note:** PRF, pulse repetition frequency, kHz, thousand Hertz.
of individual components such as foliage (Lim and Treitz 2004; Hall et al. 2005; Thomas et al. 2006), branches (Lim and Treitz 2004; Thomas et al. 2006), stems (Lim and Treitz 2004; Thomas et al. 2006), and roots (Thomas et al. 2006). All studies modeled live vegetation biomass. Kim et al. (2009) reported the results for modeling the biomass of dead trees as well.

21.2.1.1.1 Different Ways of Deriving Area-Based LiDAR Metrics

The most widely used area-based LiDAR metrics for biomass prediction are various height metrics (Lim et al. 2003; Lim and Treitz 2004; Patenaude et al. 2004; Hall et al. 2005). Height metrics are usually calculated from vegetation returns, which are typically defined as returns with a certain height (such as 0.5 m [Lucas et al. 2006], 2 m [Naesset and Gobakken 2008], or 3 m [Hall et al. 2005; Kim et al. 2009]) above the ground surface. More complicated methods than using thresholds were also developed. For example, in a mature hardwood forest in Ontario, Canada, Lim and Trietz (2004) used the expectation–maximization algorithm to identify overstory canopy returns and obtained an $R^2$ of 0.90 using the 25th percentile height. However, this algorithm assumes that overstory and understory returns are relatively well separated. Therefore, its application over dense and continuous canopy could be unsuccessful, especially when LiDAR point density is low (Thomas et al. 2006).

LiDAR metrics can be calculated based on first, last, or all returns. There is no one-to-one correspondence between first returns and vegetation or last returns and the ground. However, first returns are usually more related to canopy surface structure than last returns (Thomas et al. 2006). A few studies have used LiDAR metrics derived from first returns for predicting biomass. For example, Hall et al. (2005) found that a canopy cover metric derived from first returns was able to predict foliage biomass and total AGBM in a Ponderosa pine forest in Colorado with $R^2$ values of 0.79 and 0.74, respectively. Kim et al. (2009) found that the model using first returns improved $R^2$ by 0.1 for predicting the total AGBM in the mixed coniferous forest in Arizona compared with the one using all returns.

Instead of directly using the point cloud, height metrics can also be calculated from grids of the CH model (CHM) (Asner et al. 2009; Patenaude et al. 2004; St-Onge et al. 2008; Zhao et al. 2009). CHM usually records the maximum CH within each grid cell. Therefore, the height metrics derived from CHM are more related to upper canopy instead of the complete vertical profile of vegetation structure. Lu et al. (2012) found that the LiDAR metrics generated from CHM, that is, first returns, last returns, and all returns, have a similar performance (with an $R^2$ difference of up to 0.02) in predicting biomass over a mixed coniferous forest in Sierra Nevada of California.

Different cell sizes have been used in generating CHM, ranging from 0.5 m (Zhao et al. 2009) and 1 m (Patenaude et al. 2004; St-Onge et al. 2008) to 5 m (Asner et al. 2009). There was no discussion in these studies on how cell sizes were determined. However, it is expected that a larger cell size will produce a smoother CHM that can weigh taller canopy elements more since a CHM usually records the height of the highest canopy returns within a grid cell.

The intensity of a laser return has also been used in a few studies to generate metrics for predicting biomass. For example, Lim et al. (2003) used the mean height of points with intensity $>200$ to predict the biomass of a mature hardwood forest in Ontario using a simple power model. It was found that using such a subset of laser points increases the coefficient of determination, $R^2$, from 0.72 to 0.85. Kim et al. (2009) used a stepwise regression model to predict the biomass in a mixed coniferous forest in Grand Canyon National Park, Arizona, with 17 LiDAR metrics extracted from the height and intensity information of LiDAR returns. It was found that the canopy volume of the LiDAR returns of high intensity and the low intensity peak count of the intensity distribution were among the best for predicting live and dead tree biomass, respectively.
21.2.1.1.2 Useful Height Metrics for Biomass Estimation

It would be ideal if a LiDAR height metric that can predict biomass well over different sites is identified. Unfortunately, the best height metric reported in the literature differs a lot in previous studies, including the 80th (Patenaude et al. 2004), 75th (St-Onge et al. 2008), 50th (Thomas et al. 2006), 30th (Stephens et al. 2012), and 25th (Lim and Treitz 2004) percentile heights. Such variability is likely the result of differences in vegetation structure, model form, and data processing procedure. But the dominant cause is unclear. Data processing procedure and/or model form might be as important as, if not more important than, vegetation structure. For example, the three studies conducted by Lim and Treitz (2004), St-Onge et al. (2008), and Thomas et al. (2006) all focused on boreal forests, but their best LiDAR metrics for biomass estimation differ significantly. These three studies differ in model form and data processing procedure: Lim and Trietz (2004) and St-Onge et al. (2008) used a simple power model, whereas Thomas et al. (2006) used a simple linear model with square root—transformed biomass; Lim and Treitz (2004) and Thomas et al. (2006) derived height metrics from a raw point cloud, whereas St-Onge et al. (2008) calculated height metrics based on CHM. In particular, the percentile height in the study by Lim and Treitz (2004) was calculated based on overstory canopy returns, which might explain why a much lower percentile was chosen. On the other hand, the best metrics in the studies of Patenaude et al. (2004) and St-Onge et al. (2008) are similar (80th and 75th percentile heights, respectively) even though they focused on temperate deciduous forest and boreal forest, respectively. This is likely because the data processing procedure and the model form in the two studies are the same in that percentile heights were extracted from a $1 \times 1$ m$^2$ CHM and biomass was modeled with a simple power model.

21.2.1.1.3 Other Useful Metrics for Biomass Estimation

Besides height metrics, other metrics such as the ones related to canopy cover (Hall et al. 2005), canopy density (Naesset and Gobakken 2008), and canopy volume (Kim et al. 2009) have been proved to be useful for predicting biomass as well. Canopy cover is useful especially over areas with low biomass density (Hall et al. 2005). Canopy cover and density can complement height metrics for a better characterization of 3D canopy structure because they are related to horizontal and vertical vegetation structures, respectively (Hall et al. 2005; Bortolot 2006; Naesset and Gobakken 2008). Among these metrics, canopy volume is unique given that it is a composite index for characterizing both horizontal and vertical canopy structures. This metric has been successfully used for predicting biomass (Hall et al. 2005), stem volume, and basal area (Chen et al. 2007a).

21.2.1.1.4 Common Area-Based Statistical Models

Power models have been used in a large number of studies for biomass estimation, probably because most allometric equations for calculating biomass in the field are power models. At the log–log scale, simple power models (Lim et al. 2003; Lim and Treitz 2004; Patenaude et al. 2004; St-Onge et al. 2008) correspond to simple linear regression models, whereas multiplicative power models (Naesset and Gobakken 2008) correspond to multiple linear regression models. Many studies have performed nonlinear transformation on LiDAR metrics or biomass. For example, the quadratic term of LiDAR metrics might be used (Asner 2009). Biomass might be log transformed (Hall et al. 2005) or square root transformed (Thomas et al. 2006) for reducing the heterogeneity of regression residual variance. Although nonlinear models have been widely used, there are a few studies that have used linear models for predicting biomass (Bortolot 2006; Lucas et al. 2006; van Aardt et al. 2006; Asner et al. 2009; Zhao et al. 2009). The area-based models are mostly statistical...
and thus have the risk of overfitting. If stepwise regression or similar statistical (e.g., CART [classification and regression tree]) methods are used to select the best subset of LiDAR metrics, it is not so intuitive to interpret the physical meaning of the model. More research will be needed to develop models with good generalizability in the estimation of biomass and canopy structure (Chen et al. 2007a; Chen 2010b).

21.2.1.2 Individual Tree Analysis Methods

Individual tree analysis (ITA) methods are to automatically identify individual tree crowns and extract individual tree information such as tree height and crown size, which can be related to biomass and other canopy structure variables through allometric equations (Chen et al. 2006; Chen et al. 2007a). If the individual tree boundaries can be delineated, LiDAR statistical metrics such as those used in area-based approaches can also be extracted by analyzing the LiDAR point cloud within individual tree polygons. Compared to area-based approaches, theoretically no fieldwork is needed for model development if tree crowns can be delineated with no errors and allometric equations exist to estimate biomass based on LiDAR metrics such as tree height and/or crown size. If there are no such allometric equations, the amount of fieldwork required is still much smaller than that for area-based approaches because field data are needed only for a sample of trees instead of a sample of plots or stands. In terms of biomass mapping, this is probably the most appealing factor of ITA methods compared with area-based approaches.

21.2.1.2.1 Challenges of ITA Methods for Biomass Estimation

Despite the benefits discussed in Section 21.2.1.2, ITA methods face several challenges that limit their broader applications. First of all, LiDAR data of high point density are usually needed to measure canopy surface morphology in sufficient detail so that trees can be automatically identified with intelligent algorithms. Although a minimal point density is not established in the literature, it is recommended to have LiDAR data with a spot spacing of 0.5 m or smaller for effective tree isolation. The LiDAR data density might not be a big issue nowadays due to the significant increase in laser repetition rates of recent airborne LiDAR sensors. The major limitation of ITA might be that tree recognition algorithms are prone to errors given the complexity of canopy surface, especially for broadleaf trees that do not have regular geometry and sharp treetops as coniferous trees do. Tree recognition is challenging if trees of different heights are clustered and there are no distinct elevation “valleys” among trees. It is also more difficult to detect understory and smaller trees than dominant/codominant trees in a forest. These problems can be alleviated if we can develop more accurate and computationally efficient algorithms for tree isolation (Chen et al. 2006) and develop LiDAR metrics and allometric equations that are less sensitive to errors in tree recognition (Chen et al. 2007a), which obviously are not trivial tasks. The advent of commercial airborne small-footprint waveform LiDAR provides better opportunities for detecting small trees under complex canopy, but the data analysis methods usually are complicated (Reitberger et al. 2009). Given the aforementioned challenges, only a few studies have used ITA methods to estimate biomass (e.g., Omasa et al. 2003; Bortolot and Wynne 2005; Popescu 2007).

21.2.2 Airborne Small-Footprint Discrete-Return Profiling LiDAR

Unlike scanning LiDAR, airborne discrete-return profiling LiDAR does not include an inertial navigation system. So the positional accuracy of laser returns is much lower (on the order of 10 m) (Nelson et al. 2004). However, profiling LiDAR systems are
cost-effective for large-area vegetation inventory. For example, the cost of the portable airborne LiDAR system (PALS), a profiling LiDAR developed by the National Aeronautics and Space Agency (NASA), was around $30,000 (Nelson et al. 2003), whereas commercial airborne discrete-return scanning systems typically cost 1–2 million dollars. Nelson and colleagues have been pioneering the use of airborne profiling LiDAR for vegetation studies and have estimated the biomass of pine plantations in Georgia (Nelson et al. 1988a, b); tropical wet forests in Costa Rica (Nelson et al. 1997, 2000); statewide vegetation in Delaware (Nelson et al. 2004); the pinelands of New Jersey (Skowronski et al. 2007); and the forest south of the tree line in Québec, Canada (Boudreau et al. 2008; Nelson et al. 2009).

21.2.2.1 Useful Profiling LiDAR Metrics for Biomass Estimation

Similar to the studies that use scanning LiDAR systems, various metrics can be derived from LiDAR returns to estimate biomass. A variety of metrics have been demonstrated to be useful for predicting biomass, including canopy profile area (Nelson et al. 1988a, b), average CH of canopy returns (Nelson et al. 1997, 2000), coefficients of variation of the average height of canopy returns and all returns (Nelson et al. 1997, 2000), quadratic mean height (QMH) of canopy returns (Nelson et al. 2004), QMH of all returns (Boudreau et al. 2008; Nelson et al. 2009), and 80th percentile height (Skowronski et al. 2007). Note that the two-dimensional (2D) canopy profile area proposed by Nelson et al. (1988a, b) is analogous to the 3D canopy geometric volume generated from an airborne scanning LiDAR (Hall et al. 2005; Chen et al. 2007a).

21.2.2.2 Challenges of Profiling LiDAR for Biomass Mapping

Airborne profiling LiDARs have a low pulse repetition rate and sparse laser pulse density. Therefore, one of the main challenges of profiling LiDAR data processing is to automatically derive the ground elevation so that CH can be accurately estimated for each laser pulse. Ground returns were manually determined and connected in early studies (Nelson et al. 1988a). Later, Nelson et al. (2004) and Boudreau et al. (2008) proposed a semiautomatic approach to fit the lowest returns within a moving window with splines, followed by inspection and, if necessary, adjustment of the splines by an operator.

Another challenge of using profiling LiDAR is related to the relatively coarse positional accuracy of its LiDAR pulses, which makes it difficult to accurately register laser returns with field measurements of biomass, especially over areas such as the tropics where accessibility in the field is limited. To handle this problem, Nelson et al. (1997) proposed a simulator of laser returns based on the field data of tree characteristics when estimating biomass of three sites in the tropical wet forests of Costa Rica. Statistical models were developed to relate the biomass measured in the field with the simulated laser returns. Then, the developed statistical models were used to estimate biomass for a larger area. Such an approach was also used to estimate biomass for the state of Delaware (Nelson et al. 2004).

With its low cost and portability, PALS has been used to estimate biomass for relatively large areas such as the state of Delaware of 5205 km² (Nelson et al. 2004). Note that profiling LiDAR systems essentially can only sample vegetation along transects. Unlike scanning LiDAR systems, it is difficult to use them to directly generate wall-to-wall maps of biomass. To calculate the total biomass, a study area might need to be stratified with land cover maps and biomass is then estimated for each stratum (Nelson et al. 2004).
21.2.3 Airborne Medium- and Small-Footprint Waveform LiDARs

Compared to discrete-return LiDAR, waveform LiDAR can better characterize vegetation structure by recording the whole vertical profile of returned energy reflected from vegetation. However, its advantages might be marginal in terms of digital elevation model mapping, which is the major conventional application of airborne LiDAR. This might explain why there were no commercial airborne waveform LiDAR systems until 2004 (Mallet and Bretar 2009), almost 10 years behind the advent of the first commercial discrete-return LiDAR systems.

A number of studies have used experimental airborne medium-footprint waveform systems, specifically Scanning LiDAR Imager of Canopies by Echo Recovery (SLICER) (Blair et al. 1994) and Laser Vegetation Imaging Sensor (LVIS) (Blair et al. 1999), both developed by NASA. These two airborne waveform systems act as simulators for satellite LiDAR missions, so they were flown at relatively high altitudes (~10 km) and have a medium-sized (~10–30 m, equivalent to one or two typical crown width) footprint. SLICER has been used to estimate biomass for temperate deciduous forests in Maryland (Lefsky et al. 1999b); boreal coniferous forests in Manitoba, Canada (Lefsky et al. 2002); and temperate coniferous forests in the Pacific Northwest (Means et al. 1999; Lefsky et al. 1999a, 2001, 2005b). Means et al. (1999) found that a SLICER of footprint size 10 m can predict total biomass and leaf biomass with $R^2$ values of 0.96 and 0.84, respectively, for coniferous forests in the Pacific coast with maximum biomass density up to 1300 Mg/ha, which provides strong evidence that LiDAR can map forests with very high biomass where conventional optical and radar remote sensing systems have saturation problems.

There are only a few studies that have used small-footprint waveform LiDAR for biomass mapping (e.g., Sarrazin et al. 2011; Kronseder et al. 2012). The data volume of small-footprint waveform LiDAR is at least one magnitude higher than discrete-return LiDAR data of comparable pulse density, and there is a lack of methods specifically designed for processing small-footprint waveform LiDAR data. For example, Kronseder et al. (2012) simply used the point clouds generated from waveform LiDAR for biomass estimation, making the methods not essentially different from the ones for discrete-return LiDAR. There is an urgent need for more sophisticated methods for fully exploiting the rich information in waveforms.

21.2.3.1 Useful Waveform Metrics for Biomass Estimation

A large number of metrics can be generated from waveforms. In the studies that use SLICER for biomass estimation, the best LiDAR metrics include mean CH (MCH) (Lefsky et al. 2001, 2002; Means et al. 1999), quadratic MCH (QMCH) (Means et al. 1999, Lefsky et al. 1999b), canopy volume (Lefsky et al. 1999a), and the number of waveforms taller than 55 m (Lefsky et al. 1999a). LVIS has been used to estimate biomass for tropical wet forests in Costa Rica (Drake et al. 2002a, b; Dubayah et al. 2010), coniferous forests in the Sierra Nevada mountain of California (Hyde et al. 2005), and temperate mixed forests in New Hampshire (Anderson et al. 2006). The best metrics for biomass estimation in LVIS-based studies include height of medium waveform elevation (HOME) relative to the ground peak elevation (Anderson et al. 2006; Drake et al. 2002a; Hyde et al. 2005), quantile heights (Drake et al. 2002b), mean, and minimum height (Hyde et al. 2005).

21.2.3.2 Importance of Correcting Attenuated Waveforms

The vertical profile of waveform energy does not necessarily resemble the vertical profile of vegetation structure because laser energy attenuates as it penetrates through a canopy.
Lefsky et al. (1999a) applied Beer’s law to estimate the canopy vertical profile (CHP) from raw waveform signals before calculating LiDAR metrics. However, when Drake et al. (2002b) used the quantile metrics from raw LiDAR waveforms as well as from the CHP estimated by Beer’s law to estimate total AGBM in the tropical wet forest in Costa Rica, it was found that the stepwise regression does not include the quantile metrics from CHP. In other words, this study indicated that the waveform transformation did not lead to better results for estimating total AGBM. Since the numbers of studies addressing this issue are few, more research is needed along this direction to reach a more general conclusion.

21.2.4 Satellite Large-Footprint Waveform LiDAR

The Geoscience Laser Altimeter System (GLAS) on board the Ice, Cloud, and Land Elevation Satellite (ICESat) was the only operating satellite LiDAR system from 2003 to 2009; today, a new-generation photon-counting satellite LiDAR is being developed (Abdalati et al. 2010). We focus on GLAS in this section. The lasers on GLAS transmit pulses from an altitude of ~600 km, producing 60 m nominal footprint size and ~170 m shot spacing along track on the ground. The receiver on GLAS records the vertical canopy structure up to 81.6 m for early laser campaigns (laser 1a and 2a periods) or 150 m for later campaigns to avoid the truncation of signals by tall objects or steep slopes (Chen 2010a).

Since its launch in 2003, GLAS has produced unprecedented datasets at the global scale. However, only a few studies have explored the use of GLAS in biomass mapping. This might be because of the high cost of collecting field measurements of biomass at the scale of GLAS footprints due to their large sizes. Also, a relatively large area has to be visited to get a representative sample for statistical modeling due to the sampling nature of GLAS waveforms, especially when the accessibility to specific shots is limited. One common solution to this problem is to use estimates from airborne LiDAR data as a proxy of ground truth (Chen 2010a, b) given the high accuracy and precision of airborne LiDAR–derived biomass estimates. Such a strategy has been used for GLAS-based biomass modeling in the boreal forest in Canada (Boudreau et al. 2008; Duncanson et al. 2010) and the peat swamp forest in Kalimantan, Indonesia (Ballhorn et al. 2011). Since GLAS is essentially sampling the vegetation along the orbit, it is usually combined with optical imagery to produce continuous maps of biomass (e.g., Duncanson et al. 2010; Saatchi et al. 2011; Mitchard et al. 2012).

21.2.4.1 Challenges of Large-Footprint Satellite LiDAR for Biomass Mapping

One of the main challenges for GLAS data processing is that signals from terrains and vegetation are convoluted and waveforms are broadened over mountainous areas due to its large footprint size. Usually, Gaussian decomposition is used to disentangle the terrain and vegetation signals to estimate terrain elevation (Chen 2010a; Liu and Chen 2012) and vegetation height (Chen 2010b), based on which biomass can be estimated. The large footprint size of GLAS was optimally set for detecting ice sheets and glaciers, not necessarily for retrieving vegetation information. Chen (2010b) simulated the different footprint sizes from 10 to 40 m for retrieving maximum vegetation height. It was found that a smaller footprint size can increase the performance in vegetation information retrieval and the footprint size of 10 m has a bias and a root-mean-square error of 1 m and 1–2 m, respectively, for estimating maximum CH even when models are applied across coniferous and woodland sites. Since the next satellite LiDAR mission, ICESat-2, carries a sensor of 10 m footprint (Abdalati et al. 2010), we expect that the terrain convolution problem will be greatly alleviated for vegetation information retrieval in the future.
21.2.5 Ground-Based Discrete-Return Scanning LiDAR

Compared with other LiDAR systems, ground-based LiDAR is unique in many aspects. Ground-based LiDAR systems can measure vegetation with very high precision without destructive measurement/sampling. It is also possible to distinguish different vegetation components (stem, branches, and foliage) from a point cloud. Besides, the upward scanning geometry of most ground-based LiDAR systems can help to better characterize understory and crown base height, which are important parameters for fire behavior prediction modeling. A few studies have used ground-based LiDAR for biomass estimation (e.g., Loudermilk et al. 2009; Seidel et al. 2011; Yao et al. 2011; Ku et al. 2012). Some studies (e.g., Loudermilk et al. 2009) have found that terrestrial LiDAR can measure canopy attributes with higher accuracy than the conventional field-based approach.

21.2.5.1 Challenges of Ground-Based LiDAR for Biomass Estimation

When ground-based LiDAR systems are used to measure trees in a plot or stand, the trees usually have to be scanned from multiple positions due to the shadowing effects caused by objects closer to the scanners. The 3D coordinates of laser returns collected at individual scanning positions are local and relative to the scanners. The individual datasets have to be georeferenced to a common coordinate system based on features visible to multiple positions, which is not a trivial task. To make it more difficult, ground-based LiDAR systems can acquire data with point densities 100–1000 times higher than the average point density of airborne small-footprint LiDAR systems. Such massive volumes of data pose a significant challenge in developing fast, automatic, and memory-efficient software for data processing and information extraction.

---

21.3 General Challenges and Future Directions

21.3.1 Fusion of LiDAR with Other Remotely Sensed Data

21.3.1.1 Fusion with Optical Imagery

LiDAR has limited spectral information usually with only one wavelength in the near-infrared region. It would be interesting to investigate whether the integration of multispectral or hyperspectral imagery with LiDAR can improve biomass estimation. Some previous studies have reported that the addition of multispectral or hyperspectral features to LiDAR resulted in only slight or no improvements in biomass estimation over mixed coniferous forests in California (Hyde et al. 2006), tropical forests in Costa Rica (Clark et al. 2011), and temperate coniferous forest in Germany (Latifi et al. 2012). The study by Anderson et al. (2008) is among the few studies that have reported significant improvements when integrating LiDAR with spectral features from optical data. They integrated LVIS and Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data to estimate AGBM in a northern temperate mixed coniferous and deciduous forest in Bartlett Experimental Forest (BEF) in central New Hampshire. It was found that 8%–9% more of the variation in AGBM was explained by the use of integrated sensor data in comparison with either AVIRIS or LVIS metrics applied singly.
21.3.1.2 Fusion with Radar

Radar is another technology that can potentially be fused with LiDAR data for biomass estimation. Interestingly enough, mixed results also have been reported in the literature. Hyde et al. (2006, 2007) have found that simply adding radar backscatters InSAR height at the locations of LiDAR data does not necessarily improve the performance of their biomass estimation for coniferous forests in California and Arizona. Nelson et al. (2007) made a similar conclusion when they used a low frequency (80–120 MHz) VHF radar (BioSAR) and a profiling LiDAR (PALS) to estimate biomass over loblolly pine plantations in North Carolina. In contrast, Banskota et al. (2011) obtained very difficult results when they used BioSAR and airborne profiling and scanning LiDAR for estimating forest biomass in Virginia: the combination of BioSAR with LiDAR can significantly increase model performance. They attributed the discrepancy between their findings and the earlier ones to the fact that the previous studies focused on more or less even-aged pine forests, whereas their studies involve not only conifers but also hardwood and mixed forests. Tsui et al. (2012) also found that the integration of C-band HH backscatter to a LiDAR-only biomass model can explain an additional 8.9% and 6.5% of the variability in total aboveground and stem biomass, respectively, whereas C-band polarimetric entropy explained an additional 17.9% of the variability in crown biomass.

In general, the dominant finding from previous studies is that the addition of radar and optical spectral features into LiDAR do not, or only slightly, improve biomass estimation. This is reasonable given that biomass is highly related to tree height, which can be accurately measured by LiDAR. The predictors derived from radar and image spectral features usually become insignificant once the LiDAR-derived height metrics are included in the model. One promising method for integrating LiDAR with other remote sensing data is to use the other data for vegetation classification and develop vegetation type–dependent biomass modeling approaches (e.g., Swatantran et al. 2011), especially with novel statistical skills such as mixed-effects modeling (e.g., Chen et al. 2012). Note that the discussion here only involves the integration of airborne LiDAR data and other remotely sensed imagery. Satellite LiDAR could have a very different strategy for data fusion because of its nature of sampling instead of wall-to-wall mapping. The data fusion issue related to satellite LiDAR is not included here due to space limitation.

21.3.2 Generalizability of Biomass Models

A key issue in biomass estimation with LiDAR data is model generalizability, the extent to which a model developed at one study area can be applied to other areas (Chen 2010b). This is critical for biomass estimation at the regional and global scales or for areas where field measurements of biomass are difficult or impossible to collect. There are two aspects from which model generalizability can be improved: first, to choose LiDAR metrics that are mostly related to biomass over a wide range of vegetation conditions and, second, to choose a model form that is applicable over different areas based on these metrics.

Several studies have attempted to identify LiDAR metrics with which a model can be applied across different sites. Lim et al. (2003) found that when maximum height was used to estimate biomass, different models should be fitted for different groups of plots in their mature hardwood forest in Ontario. However, when the mean height of the pulses with an intensity >200 is considered, there is no need to develop different models. Lefsky et al. (1999b) tested four height indices for estimating biomass. Of the four indices predicting AGBM, they found that QMCH had the best performance when the model based on it was applied over two different datasets.
A few studies have found that MCH is the best LiDAR metric that can be applied over different study areas (e.g., Lefsky et al. 2002; Asner et al. 2012). Lefsky et al. (2002) found that a simple model based on MCH (AGBM = 0.378MCH^2, where AGBM is in megagrams per hectare) can explain 84% of the biomass variability for 149 plots across three types of forests (temperate conifer, temperate deciduous, and boreal conifer) in North America. Asner et al. (2009, 2012) confirmed the usefulness of MCH for biomass mapping, but they (Asner et al. 2012) found that biomass is affected not only by MCH but also by basal area and wood density when they developed carbon (biomass) models across four different tropical study areas (Hawaii, Panama, Peru, and Madagascar). They found that the regional variations of MCH-based biomass models are dominated by the basal area–MCH relationship. This might partially explain why Anderson et al. (2006) found that the model by Lefsky et al. (2002) performed worse than HOME for estimating the AGBM of 20 circular plots in BEF using LVIS.

There is converging evidence showing that the relationships between biomass and LiDAR height metrics are vegetation type dependent (Maclean and Krabill 1986; Drake et al. 2003; Naesset and Gobakken 2008; Ni-Meister et al. 2010; Asner et al. 2012; Chen et al. 2012). This is reasonable given that vegetation biomass is determined not only by height but also by factors such as stem tapering and wood density (Niklas 1995; Chave et al. 2006), which could be vegetation specific and difficult to directly measure using LiDAR. There are at least two key research issues about incorporating vegetation type information into biomass modeling: First, how do we develop vegetation type–dependent biomass models? The most straightforward method is to fit a model per vegetation type, but the problem is that it will reduce the sample size per model if the total number of field plots is small and the number of vegetation types is large. Second, what level of detail (e.g., species, alliance, or association) should vegetation be stratified at? A very detailed stratification requires vegetation maps at the fine taxonomic level and a significant amount of effort for collecting field measurements for model calibration. On the other hand, a very coarse stratification can lead to a poor performance in biomass estimation. Chen et al. (2012) addressed the first issue using a mixed-effects model, a novel statistical technique, and the second issue by comparing two alliance-level vegetation classification systems (Society of American Foresters [SAF] and National Vegetation Classification [NVC]). They found that the mixed-effects model can deal with the issue of small sample size per vegetation type and improve biomass estimation, and the coarser SAF vegetation type has almost the same performance as the finer NVC types, indicating that the height–biomass relationships do not vary much at the fine taxonomic level. This is consistent with the finding of Nelson et al. (1988a), who showed that stratifying a pine plantation in Georgia into four southern pine species (loblolly pine, shortleaf pine, slash pine, and longleaf pine) did little to increase the accuracy of biomass estimation using airborne profiling LiDAR.

A more thorough study of the aforementioned issue depends on a good understanding of plant allometry between vegetation structure variables such as biomass, height, diameter-at-breast height, and wood density. Unfortunately, allometric equations are available only for a limited number of species, most of which are from developed countries (TerMikaelian and Korzukhin 1997; Jenkins et al. 2004; Lambert et al. 2005; Zhao et al. 2012). Over areas such as the tropics where species diversity is high, generalized biomass equations (Brown 1997; Chave et al. 2005) have been used, which may lead to a bias in estimating biomass for a particular species (Litton and Kauffman 2008). Moreover, developing or validating allometric equations is strenuous because it usually involves harvesting different tree components if a direct assessment is required. It is expected that accumulation of field measurements (Clark and Kellner 2012) and progress in ecological theory in plant allometry (Brown et al. 2005; Kozlowski and Konarzewski 2004) can shed more light on this research direction (Chen et al. 2007a).
It is very common for stepwise regression to be used to select predictors from a large number of LiDAR metrics for biomass estimation. However, stepwise regression usually suffers from the overfitting problem. Instead, models based on a priori knowledge and ecological theories might have better model generalizability when LiDAR metrics and model form are carefully chosen (Chen et al. 2007a). Magnussen et al. (2012) proposed a generic model to predict tree-size-related canopy attributes from the CH of LiDAR first returns. The model they proposed includes two predictors: mean and variance of CH. Using three separate datasets from Norway, they obtained compelling results when they estimated a number of canopy attributes (such as volume and basal area), not including biomass. It would be interesting to test such models for biomass estimation in the future.

### 21.3.3 Biological and Ecological Interpretation of Biomass Models

An issue relevant to the generalizability of LiDAR-based biomass models is the biological and ecological interpretation of the models. Most previous studies ran stepwise regression models with a large number of LiDAR metrics and reported the best model without much interpretation. As LiDAR-based biomass studies are rapidly accumulating, we should put more emphasis than before on interpreting the LiDAR metrics selected by a model so that other researchers have better ideas on what LiDAR metrics they should try for their specific studies and are able to make better sense of their own models.

For example, QMH has been found to be the best predictor for biomass in a couple of studies (e.g., Lefsky et al. 1999b; Chen et al. 2012; Lu et al. 2012). QMH can be calculated as

\[
\sqrt{\frac{1}{n} \times \sum_{i=1}^{n} h_i^2},
\]

where \( n \) is the number of laser points and \( h_i \) is the height above the ground for point \( i \). As indicated in the formula, QMH gives a higher weight to higher points. This is analogous to the fact that the biomass of a plot or stand is dominated by tall trees. QMH is also an index integrating the height of all points, similar to the fact that the biomass at a given area is an integration of the biomass of all trees. This makes it theoretically appealing for biomass estimation. As introduced earlier, MCH was also the best predictor of biomass in many studies (e.g., Lefsky et al. 2002; Asner et al. 2012). In most studies, the best models include either QMH or MCH, but not both. And there usually is a positive relationship between biomass and either metric. However, in coniferous forests on the Pacific coast, Means et al. (1999) found that their best model included both QMH and MCH. Moreover, they found that QMH is negatively related to biomass, whereas MCH has a positive effect on biomass. This was explained by the fact that mature stands have relatively lower biomass compared with old-growth stands at a given height, whereas the foliage of mature stands mostly is in the upper half of the canopy (note that QMH weighs higher canopy elements more than lower ones). This indicates that in the future both QMH and MCH should be attempted for predicting biomass if a study area includes old-growth forests.

### 21.4 Conclusions

Information on biomass is useful for numerous environmental applications and is critical for understanding and predicting global environmental changes. LiDAR is a state-of-the-art remote sensing technology of tremendous potential in biomass mapping without the
LiDAR Remote Sensing of Vegetation Biomass

limitation of saturation at high biomass levels. This chapter gives a thorough and in-depth review of the current status, future directions, and main challenges of biomass estimation using various satellite, airborne, and ground-based LiDAR systems.

Each type of LiDAR systems has its strengths and weaknesses, and it is possible to combine them for better biomass mapping (Nelson et al. 2009; Chen 2010a, b). Airborne and satellite LiDAR systems measure vegetation from above the canopy, missing much information on stem and understory. Instead, ground-based LiDAR systems can measure stem, branches, and understory in much more detail and thus can be used to calibrate and validate airborne and satellite LiDAR data and products. It is unlikely to generate a wall-to-wall map of biomass directly from satellite LiDAR due to its sparse spatial distribution of measurements. However, airborne LiDAR can be used to assess the sampling errors of satellite LiDAR and help optimize the design of next-generation satellite systems such as ICESat-2. It is expected that this comprehensive review will facilitate the use of different LiDAR systems in tandem for biomass mapping. Overall, research on using many LiDAR systems for biomass mapping is still in its infancy; much remains to be explored in the future to address issues such as data fusion and model generalizability.

References


LiDAR Remote Sensing of Vegetation Biomass


LiDAR Remote Sensing of Vegetation Biomass


