Lidar for Biomass Estimation

Yashar Fallah Vazirabad and Mahmut Onur Karslioglu

Middle East Technical University
Turkey

1. Introduction

Great attention has been paid to biomass estimation in recent years because biomass can simply be converted to carbon storage which is very important to understand the carbon cycle in the environment. Biomass is typically defined as the oven-dry mass of the above ground portion of a group of trees in forestry (Brown, 1997, 2002; Bartolot and Wynne, 2005; Momba and Bux, 2010). However, there are a few studies about below ground biomass estimation. Conventionally, it is estimated using measurements which are recorded on the ground. On the other hand, the large number of studies have confirmed that Lidar as a kind of active remote sensing system is able to estimate biomass properly (Popescu, 2007). Hence, time-consuming field works can be avoided and unavailable regions become accessible using a relatively low cost and automated Lidar system. (Nelson et al., 2004; Drake et al., 2002, 2003; Popescu et al., 2003, 2004).

Traditional remote sensing systems detect vegetation cover using active and passive optical imaging sensors (Moorthy et al., 2011). Passive systems depend on the variability in vegetation spectral responses from the visible and near-infrared spectral regions. Widely accepted algorithms such as the Normalized Difference Vegetation Index (NDVI) have been empirically correlated to structural parameters (Jonckheere et al., 2006; Solberg et al., 2009; Morsdorf et al., 2004, 2006) such as Leaf Area Index (LAI) of canopy-level. On the contrary to passive optical imaging sensors, which are only capable of providing detailed measurements of horizontal distributions in vegetation canopies, Lidar systems can produce more accurate data in both the horizontal and vertical dimensions (Lim et al., 2003). Lidar-based instruments from space-borne, airborne, and terrestrial platforms provide a direct means of measuring forest characteristics which were unachievable previously by passive remote sensing imagery.

Developments in remote sensing technologies, in particular laser scanning techniques, have led to innovative methods and models in the estimation of forest inventories in terms of efficiency and scales (Hudak et al., 2008; Tomppo et al., 2002; Tomppo and Halme, 2004; Zhao et al., 2009; Koch, 2010; Yu et al., 2011). Lidar experiments and researches within the remote sensing community are now focusing to develop robust methodologies. These methods and models employ very precise 3D point cloud data (Omasa et al., 2007) to direct process and retrieve vegetation structural attributes which are validated by in situ measurements of vegetation biophysical parameters (Maas et al., 2008; Cote et al., 2011).

Lidar scanning systems have been used to extract various kinds of parameters, such as tree height, crown size, diameter at breast height (dbh), canopy density, crown volume, and tree...
species (Donoghue et al., 2007; Means et al., 1999, 2000; Magnussen et al. 1999). Most authors concentrate on the above-ground biomass while there are a few known studies focusing on the below-ground biomass (Kock, 2010; Nasset, 2004).

Bortlot and Wynne (2005) used Lidar data to generate canopy height models. Tree heights detected from image processing are entered as variables in a stepwise multiple linear regression to find an equation for biomass estimation. The method skips detecting small trees. They are not included in the process of estimation. A previous work by Lefsky et al. (1999) presented the prediction of two forest structure attributes, crown size and aboveground biomass from Lidar data. They analyzed the full waveform of the return pulses to define the beginning of canopy return. Linear regression was used to develop biomass estimation equation based on a defined canopy height index. Finally, they proposed stepwise multiple regression model to predict canopy volume and relatively biomass. They concluded that tree height is highly correlated with dbh in a square power function.

Van Aardt et al. (2008) evaluated the potential of an object oriented approach to forest classification as well as volume and biomass estimation using small footprint, multiple return Lidar data. A hierarchical segmentation method was applied to a canopy height model (CHM). An empirical model is employed to estimate the canopy volume and biomass. They performed stepwise discriminant analysis as a part of classification steps for variable reduction. Fallah Vazirabad and Karslioglu (2009) investigated the biomass estimation based on single tree detection method. This method is used to locate trees and detect the height of each tree top. Diameter at breast height is extracted from the close relation to the tree height which is defined by field measurements. A Log transformed model is applied for biomass estimation taking into account the dbh variable.

Airborne lidar is confirmed as the most ideal technology to obtain accurate CHM over large forested areas because of its high precision and its ability to receive ground returns over vegetated areas. Spaceborne geoscience laser altimeter system (GLAS) data on the other hand are intended to use mainly for scientific studies of sea ice elevation (Zwally et al., 2002; Kurtz et al., 2008; Xing et al., 2010), but it is also suitable for the estimation of the canopy height map (Lefsky et al., 2005; Simard et al., 2008; Chen, 2010; Duncanson et al., 2010).

The reason for the applications of GLAS data to canopy height mapping is to estimate the dynamic global carbon stock. Xing et al. (2010) analyzed the deforestation and forest degradation as a carbon source estimation model. They also investigated the forest growth model for afforestation and reforestation. Forest carbon stocks, fluxes, and biomass are directly related to each other (Garcia-Gonzalo et al., 2001; Widlowski et al., 2004). Therefore, accurate estimation of biomass of stocks and fluxes is essential for terrestrial carbon content and greenhouse gas inventories (Muukkonen and Heiskanan, 2007; Xing et al, 2010).

A general overview of forest applications is provided by recent studies (Hyyppä et al., 2009; Dees and Koch, 2008; Mallet and Bretar, 2009; Koch, 2010). They show that the information related to the height or structure of forests can be extracted with high quality.

Apart from the land cover classification Lidar intensity data can be used to differentiate materials such as asphalt, grass, roof, and trees (Hasegawa, 2006; Donoghue et al., 2007; Kim, 2009; Song et al., 2002). To identify the position and diameter of tree stems within a forest the intensity of Lidar returns has been successfully used (Lovell et al., 2011). Hopkinson and Chasmer (2009) compared four lidar-based models of canopy fractional cover and found that those models which included the intensity of the returns were less
affected by differences in canopy structure and sensor configuration. This is because the intensity measurements provide some quantification of the surface areas interacting with the laser beam. Reitberger et al. (2008) used a waveform decomposition method to extract intensity and concluded that detection of small trees below the main canopy was improved. The ability to acquire laser pulse echoes from the bottom part of vegetation canopies is restricted in the spaceborne and airborne Lidar system. This is referred to the system properties such as laser footprint size, recording frequency, as well as the natural placement of the crown elements, for example dense or open canopies. But to provide detailed specification of canopy and individual tree crowns characterization it is logical to introduce a terrestrial platform which has a much higher resolution laser pulse records than others. However, terrestrial data for tree 3D models have some problems such as overlapping crowns and under-story vegetation which cause shadowing effects.

Deriving forest data from Lidar data to model the canopy height distribution and its statistical analysis was proposed by (Holmgren and Persson, 2004; Lim et al., 2003, 2004; Næsset, 2002). The single tree detection, its location and characteristics on the basis of statistical analysis have been studied by (Hyyppä and Inkinen, 1999; Fallah Vazirabad and Karslioglu, 2010; Yu et al. 2011).

2. Lidar for biomass estimation

This section comprises two parts: systems and data acquisition. In the first part space-borne, airborne, and terrestrial systems and their sensors in relation to the biomass estimation are presented. The appropriate and useful laser band for vegetation detection is also discussed in the same part. In the second part, types of laser data acquisition such as first return, last return and multi-return are described and the applications of each type are discussed. Additionally, the new technology of light detection, namely full waveform and its utilization will be emphasized as the state of the art. The results of recent researches and studies related to the waveform for the feature extraction are highlighted.

2.1 Systems

Lidar systems make use of the time of flight principle or phase-based differences to measure the distances of objects. For this, the time interval is detected between sent and return laser pulses which are backscattered from an abject. Lidar point cloud of returns generate a 3D digital representation of the vegetation structure in which each point is characterized by XYZ coordinates (Maas et al., 2008; Cote et al, 2011).

Lidar System consists of a laser ranging unit, a scanning instrument like an oscillating mirror or rotating prism and a direct geo-referencing navigation unit (using global positioning system – GPS and inertial navigation system - INS). The choice of the platform depends mainly on the application. Space-borne systems map the globe for researches and experimental purposes. Airborne systems are collecting the data for national or regional investigations. Terrestrial platforms are frequently used to produce 3D models of man-made structures or natural resources like trees. Thus, the basic principle and technical specification for a sensor installed on a platform such as Earth orbiting satellite, airplane, helicopter, tripod, or vehicles change due to the variety of the applications (Shan and Toth, 2009). Some engineering and environmental studies require information about the shallow water basin. The Bathymetric Lidar systems are capable to provide this information in the coastal zones.
or rivers deep to 50 meters in clear water (Bathymetric system is irrelevant to our discussions so, we will have no further dealings with it in this chapter).

Generally, commercial systems are designed to receive data from small-footprint (0.20-3.00m diameter, depending on flying height and beam divergence) with higher repetition frequency (Mallet and Bretar, 2009). These systems acquire a high point density and an accurate height determination. However, small-footprint systems often miss tree tops which cause under estimation in tree height. Therefore, it is hard to define whether the ground has been detected under dense vegetation or not. Consequently, ground and tree heights cannot be well estimated (Dubayah and Blair, 2000). Large-footprint systems (10-70 m diameter) increase the chance to both hit the ground and the tree top and eliminate the biases of small-footprint systems. Thus, the return waveform gives a record of vertical distribution of the captured surface within a wider area which provides important information for biomass estimation. First experimental full waveform topographic systems were large-footprint systems and mostly carried by satellite platforms. With a higher flying height, pulses must be fired at a lower frequency and with a higher energy to penetrate into the forest canopy as much as possible (Mallet and Bretar, 2009).

2.1.1 Space-borne systems

The geoscience laser altimeter system (GLAS) is the only Lidar operating space-borne system. GLAS is the important part of NASA earth science enterprise carried on the ice, cloud and land elevation satellite (ICESat) from 12 January 2003 (Afzal et al., 2007). This instrument has three lasers, each of which has a 1064 nm lidar channel for surface altimetry and dense cloud heights, and a 532 nm lidar channel for the vertical distribution of clouds and aerosols (NASA, 2007). The three lasers have been operated one at a time, sequentially throughout the mission. The mission mode involved 33 day to 56 day campaign, numerous times per year, to extend the operation life. The main objective of the GLAS instrument is to measure the ice sheet elevations and changes in elevation through time. Second objective is the cloud detections and measurements, atmospheric aerosol vertical profiles, terrain elevation, vegetation cover, and sea ice thickness. The figure 1 shows the world elevation maps for 2009 ICESat elevation data (national snow and ice data center, NSIDC, available online at: http://nsidc.org/data/icesat/world_track_laser2F.html).

Nevertheless, only a small number of studies have used airborne lidar data to evaluate the DTM which was derived from satellite laser altimetry GLAS data over forested areas. GLAS which is only operating on board ICESat, records the full waveform returns, and provides a high precision elevation data with nearly global spatial coverage at a low end user cost (Fricker et al., 2005; Martin et al., 2005; Schutz et al., 2005; Magruder et al., 2007; Neuenschwander et al., 2008). Space-borne data are mainly used to model the global canopy height for evaluating carbon budget (Xing et al., 2010).

Recently, Duong et al. (2007, 2009) compared terrain and feature heights derived from the satellite (GLAS) observations with a nationwide airborne lidar dataset (the Actual Height model of the Netherlands: AHN). They found that the average differences between GLAS- and AHN-derived terrain heights are below 25 cm over bare ground and urban areas. Over forests, the differences are even smaller but with a slightly larger standard deviation of about 60 cm (Chen, 2010). Harding et al. (2001) utilized GLAS full waveform data to generate the average forest CHM, and the results presented the variations of important canopy attributes, such as height, depth, and the over-story, mid-story, and under-story
forest layers. Sun et al. (2007, 2008) applied GLAS waveforms to estimate the forest canopy height in the flat area in Northern China mountains, and found that the ICESat-derived forest height indices was well correlated with the field-measured maximum forest height 
\[
R^2 = 0.75
\]
where \( R^2 \) is the coefficient of determination.

Fig. 1. Example of ICESat World Elevation Map

2.1.2 Airborne systems

An extensive test of laser profiler was performed at the Stuttgart University (1990) where Differential Global Positioning System (DGPS) and Inertial Measurement Unit (IMU) was integrated in the laser system for the first time to provide precise positioning and orientation (attitude) of the airborne platform. Soon after that, the scanning mechanism was designed by Optech company (Canada - ALTM system).

Laser profiler was developed in the forestry research by NASA’s Goddard space flight center (GSFC) on the basis of Riegl laser rangefinder with 20 ns wide laser pulse and repetition rate of 2 kHz. There are three main commercial suppliers of airborne laser scanning systems, Optech International Inc., Leica Geosystem, and Riegl which are producing the data for the forest inventory and biomass estimation researches. Generally, other companies completed their systems which utilize these three laser scanner instruments. Besides these commercial systems, a number of other systems built by US government research agencies are offered for scientific research purposes, like NASA, ATM, RASCAL, SLICER, Laser Vegetation Imaging Sensor (LVIS), and ScaLARS. LVIS has been developed by NASA for the topography mapping, elevation and the forest growing on it. A special design of scanning system such as the full waveform is required for the scanning of vegetation covered regions to capture the reflected pulse in different returns. This scanner has been used in USA (California, eastern states), Central America (Costa Rica and Panama). It was also applied in Amazonian forests of Brazil to generate direct measurements of canopy height and relatively aboveground biomass map. (Shan and Toth, 2009)
2.1.3 Terrestrial systems
The primary classification with respect to measuring principle is described by two techniques namely pulse ranging or time of flight (TOF) and phase measuring technique. Another classification is also available in accordance with the angular scanning technique and coverages of scanner which consist of Panorama, Hybrid, and Camera scanners (Shan and Toth, 2009). Panorama scanners carry out distance and angular measurements providing 360° angular coverage within the horizontal plane. Types of laser scanners, which perform unrestricted scanning around the rotation axis, fall in the category of Hybrid scanners. The third category of scanners carrying out distance and angular measurements over a limited angular range and in a specific field of view is called Camera scanners (Shan and Toth, 2009). For the range measurements, it is necessary to obtain information about the exterior orientation elements (positions and orientation or attitude angles) of platforms of the terrestrial laser scanner. Precise exterior orientation elements can be detected during the calibration procedure. Sensitivity of tree volume estimates which are related to different error sources in the spatial trajectory of the terrestrial Lidar has been analyzed by (Palleja et al., 2010). Their tests have demonstrated that the tree volume is very sensitive to the errors in the determinations of distance and the orientation angle. Cote et al. (2011) proposed to estimate the tree structure attributes by means of terrestrial Lidar. They concluded that the main limitation of the use of terrestrial system was the effect of object shading and wind. In context with the precise biomass estimation terrestrial laser scanning can be considered as a support system for airborne and space borne Lidar.

2.2 Data acquisition
Measurement process of laser scanner can be represented by the frequency, intensity, phase and the travel time of the sent and returned signal. The transmitted and received energy are formulated similar to the Radar (radio detection and ranging) equation (Shan and Toth, 2009). This can be expressed as an integral (Mallet and Bretar, 2009) and the range is measured in pulsed systems as 
\[ R = c.t/2 , \]
where c is the speed of light, t is two way laser light travel time, \( R \) is the distance to be measured (Shan and Toth, 2009). The equation of the continuous waveform is 
\[ R = 0.5 (\phi/2\pi)\lambda \]
where \( \phi \) is the phase difference and \( \lambda \) is the wavelength which is operationally between 600 and 1000 nm (Electromagnetic infrared range). This interval is not eye-safe. Therefore, the optimum performance has to be balanced against safety considerations.

In addition to positional data, each Lidar observation must also contain the scan angle for each shot together with the measurement of reflectance from the target. Since the calculation of range for the detected pulse involves the elapsed time the precision of time measurement is of vital importance considering that 7 ns sensitivity is needed to distinguish 1 m object. This plays in turn a decisive role in the scanning of vegetated areas. In some methods they use a fraction which is a constant in the sent and return pulse. But, in others, they take the centroid of the pulses as a time reference.

The characteristics of forest inventory from both discrete return (first, last, multi returns) and full waveform recordings are extensively studied by different Lidar approaches such as tree crown detection and biomass estimation (Harding et al., 2001; Coopes et al., 2004; Jang et al., 2008; Brantberg et al., 2003).

2.2.1 First return, last return
Lidar systems can be categorized by the way they process the waveform reflections for each pulse and also by the size of the footprint they record. Systems that record footprints up to
100 cm are often called small footprint systems typically at frequencies around 15 kHz (Heritage and Large, 2009). Early small footprint systems recorded the range only up to the first reflecting object or the first pulse in discrete returns. In principle, the map of all first pulses results in such a model showing only the height of all surface objects. This requires to record the last reflecting object in each return signal if there is more than one reflectance, which is often referred to the last pulse. Although the last pulse data has clearly the potential to penetrate vegetation canopies, it can never be guaranteed that the last pulse can reach the ground and is not reflected from the higher point of canopy. Furthermore, where low vegetation is involved, the first and last pulse may be too close together to generate a reliable range and leads consequently to over estimation of the terrain height.

Coopes et al. (2004) used airborne discrete returns to indicate canopy crown and height. Lim and Treitz (2004) collected the airborne discrete first and last returns for above ground biomass estimation. In Jang et al. (2008) the apple tree inventory are extracted from discrete return without explaining their effect on the results. First and last returns are used by Thomas et al. (2006) but the effects of which are not explained on the results of canopy height models. Fallah Vazirabad and Karslioglu (2010) extracted the tree tops empirically from the first pulse data because it contains more canopy returns than the ground ones. In discrete return systems, the small diameter of footprints and the high repetition rates of these systems made possible to have high spatial resolution, which can yield dense distributions of sampled points. Thus, discrete return systems are preferred for detailed mapping of ground and canopy surface. Finally, these data are readily and widely available, with ongoing and rapid development in forestry.

### 2.2.2 Multi return

The capability of detecting different returns in the closely placed terrain surfaces depends on instrument parameters such as the laser pulse width (the shorter the better), detector sensitivity, response time, the system signal to noise performance, and others. In case of discrete returns more detectors are needed. With this technology the number of pulses between first pulse and last pulse can be recorded as many as the number of detectors. Thus, there are systems with second and third pulse beside first and last pulse record. In contrast to small footprint systems, large footprint systems (10-100 m) open up the possibility of recording the entire return pulse. Discrete return airborne laser systems (ALS) have the benefit of providing data over a large area, but are restricted by their laser pulse return density as points/m² ratio. Multiple return recording capabilities of system produce point cloud density between 1 and 20 points/m² optimistically. Often this level of point density is unsatisfactory to produce a comprehensive 3D model, especially in the vertical view (Moorthy et al. 2011).

### 2.2.2 Full waveform

The problems which are mentioned before in first and last pulse systems for vegetated regions can be solved with full waveform technology making an important contribution to biomass estimation (Shan and Toth, 2009). The waveform is usually digitized by recording the amplitude of the return signal at fixed time intervals (figure 2). To analyze the signal of emitted short duration laser pulse with only a few ns pulse-width, higher digitizer sampling
rate is required. These devices have been primarily designed for measuring vegetation properties. Extensive researches (Harding et al., 2001; Lefsky et al., 2001, 2002; Reitberger et al., 2009) have shown that waveform shape is directly related to canopy biophysical parameters including canopy height, crown size, vertical distribution of canopy, biomass, and leaf area index.

Harding et al. (2001) discussed about canopy height profile detection from full waveform raw data provided by SLICER. They studied the laser energy from the full waveform Gaussian distribution. The advantages of full waveform recording include an enhanced ability to characterize canopy structure, the ability to concisely describe canopy information over increasingly large areas, and the availability of global data sets. The examples of these data are airborne like SLICER and LVIS, and satellite data like GLAS. The other advantage of full waveform systems is that they record the entire time varying power of the return signal from all illuminated surfaces on canopy structure. It should also be stated that Lidar data, which is collected from space globally, provides only full waveform recordings (Lefsky et al., 2002).

Fig. 2. Return pulse forms (Harding et al, 2001)
3. Methods and models for Biomass estimation

This section is organized in terms of three subsections containing data pre-processing, methods and models, and applications. Data pre-processing methods in turn are divided into four parts. For the filtering methods some efficient algorithms are explained. Apart from different interpolation methods the generation of the digital terrain model (DTM), digital surface model (DSM), and canopy height model (CHM) is treated. Quality assessment of laser data is carried out within another subsection. Additionally, the quality of filtering methods, interpolation methods, DTMs, DSMs, CHMs results and their performances are also evaluated. The subsection ‘methods and models’ consider the methods and models in biomass estimation, among others single tree and tree characteristics detection. The last subsection presents applications of Lidar using the models for biomass estimation to recognize the advantages of Lidar systems in the biomass estimation.

3.1 Data pre-processing

The critical step in using Lidar data is the data pre-processing. Choosing the proper filtering method plays an important role in the quality of results. Actually, it cannot be expected that the quality of the result should be better than the data accuracy itself. On the other side, all interpolation methods have no difficulties to generate precise 3D models since dense enough Lidar data is available.

3.1.1 Filtering

The purpose of filtering is to remove the vegetation points. Figure 3 shows all points before filtering (figure 3, left) and terrain points after filtering (figure 3, right).

![Fig. 3. Removing vegetation points](image)

The terrain points extracted from the point cloud of Lidar data set are used as an input to generate a DTM. The first pulse data sets contain vegetation points and terrain points in the forest area. Numerous kinds of filtering methods are developed to classify the terrain and vegetation points in the point cloud (Pfeifer et. al., 2004; Tovari and Pfeifer, 2005). Different concepts for filtering, with different complexity and performance characteristics have been proposed in mainly four categories such as morphological, progressive densification, surface based, segmentation based filter. There are also developments, extensions, and variants for these filter methods.

The morphological filter was derived by Vosselman (2000) from the mathematical morphology definition. It works in such a way that the smaller are the distances between a ground point and its neighboring points, the lesser is the height difference. Based on this criterion the method can properly eliminate the outliers. The progressive densification filter is developed by Axelsson (Axelsson, 2000). This filter works progressively by classifying...
points which belong to the ground. Surface based filters assume at the beginning that all
points lying on the ground form a surface. Then a fitting procedure is applied to extract the
points which do not belong to the ground. This method goes back to Pfeifer et al. (2001).
Segmentation filters are developed as the fourth category. Segment is a group of points
which are located within defined thresholds such as the distance and height difference
between neighbor points. Sithole (2005) introduced a segment classification method by
performing region growing techniques referring to Tovari and Pfeifer (2005). It works by
classifying segments into as many classes as possible (Filin and Pfeifer, 2006).
The experimental comparison of filtering algorithms with manual methods for DTM
extraction is introduced by Sithole and Vosselman (2004) to show the suitability of filters
with the terrain shape. In comparison with other filtering methods, segment base filter is
turned out to be a more reliable method in steep slope terrain extraction using a surface
growing method (Sithole and Vosselman 2005).

![Fig. 4. Segmentation method, point cloud from vertical view](image)

The most important part in this method is the accuracy assessment and parameter tuning.
These processes for the segmentation method are performed by Vazirabad and Karslioglu
(2009) as shown in figure 4. Segmented terrain points are coloured as brown and green
while white points are assumed to be the vegetation points in forest area.

### 3.1.2 Interpolation

Interpolation is necessary to produce digital models from Lidar point cloud. The simple idea
of the interpolation is referred to the nearest neighbor interpolation method to estimate the
elevation (Maune, 2007). It searches for the set of nearest points, thus the new elevation
value is selected as the same value of the nearest point instead of taking the average of all
points. An important problem here is the zigzag appearance of the surface. This is in fact
due to the selecting of the nearest point method by defining Voroni diagrams or Theissen
polygons. For this reason, some kinds of averaging methods should be applied to the set of
known nearest elevation points. Therefore, a weighted average like inverse distance
weighting (IDW) is introduced which is working with the distances between these points
(Monnet et al, 2010; Bater and Coops, 2009).
In Lidar data especially in vegetated areas distances are not related to the elevations. In contrast, kriging or geostatistical approaches provide better results (Heritage and Large, 2009). However, they require more mathematically complex and computationally intensive algorithms. Since dense data is always available, rapid interpolation methods such as the nearest neighbor are preferred to use for the rough surfaces in the forest areas (Fallah Vazirabad and Karslioglu, 2010).

Riano et al. (2003) investigated the performances of spline and nearest neighbor interpolation methods to generate DTM. Spline interpolation is a special form of piecewise polynomial. The interpolation error in the DTM can be small even applying the low degree polynomial. They concluded that there were no large differences between the spline and nearest neighbor results while the spline computation was three times slower. Hollaus et al. (2010) described the derivation of DSM employing the least square fitting method to compare it with kriging interpolation. They introduced a moving least square fitting technique which selects the highest points in the search window as surface points. This technique finds the best fitting surface to the set of points by minimizing the sum of squares of the residuals of the points from surface. The results of this study showed that the least square fitting technique produced high precision DSM on rough surfaces while it needs more computational time.

3.1.3 DTM, DSM, CHM
The terrain model function \( z = f(x, y) \) is computed from 3D points, \( p_i = (x_i, y_i, z_i), i = 1, \ldots, n \), where \( n \) is the number of points (Shan and Toth, 2009). Heights are stored at discrete, regularly aligned points, and the interpolated height as the height of the grid has to be given within a grid mesh. These grid heights are obtained by interpolation methods explained before in the subsection 3.1.2. These methods consist of nearest neighbor, IDW, kriging, spline, and least square fitting.

An alternative method to the interpolations is so called triangular irregular network (TIN) data structure. The original points are used for reconstructing the surface in the form of TIN. For large point sets, triangular networks are more effective than the time consuming methods which are mentioned before. Digital surface model (DSM) is generated from noise removed Lidar data and represents the canopy top model. Digital terrain model (DTM) is basically produced by the laser pulse returns which are assumed to be on the terrain. (van Aardt et al., 2008). By subtracting DTM from DSM, CHM can be obtained which is presented in figure 5. Hence, CHM is a digital description of the difference between tree canopy points and the corresponding terrain points.

3.1.4 Quality assessment
The quality assessment is necessary for each step of the pre-processing. Pfeifer et al. (2004) reported an RMSE of 57 cm for DTM in wooded areas using data point spacing about 3 m. Hyyppa et al. (1999) reported a random error of 22 cm for fluctuating forest terrain using data point density 10 pts/m². They analyzed the effects of the date, flight attitude, pulse mode, terrain slope, and forest cover within plot variation on the DTM accuracy in the boreal forest zone. Hyyppa and Inkinen (1999) reported the CHM with an RMSE of 0.98 m and a negative bias of 0.41 m (nominal point density about 10 pts/m²). Yu et al. (2004) reported a systematic underestimation of CHM of 0.67 m for the data acquired in 2000 and 0.54 for another acquisition in 1998. The filtering methods mentioned before are likely to fail
Fig. 5. DSM (up) and CHM (down)

<table>
<thead>
<tr>
<th>Filter</th>
<th>Reduced incentives</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Terrain</td>
<td>Off-terrain</td>
</tr>
<tr>
<td>Original</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Off-terrain</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>Sum</td>
<td>A+C</td>
<td>B+D</td>
</tr>
<tr>
<td></td>
<td></td>
<td>T=(A+B+C+D)</td>
</tr>
</tbody>
</table>

Type I = \(\frac{B \times 100}{A+B}\) & Type II = \(\frac{C \times 100}{C+D}\)

Total Errors = \(\frac{(B+C) \times 100}{T}\)

Table 1. Type I and Type II errors

facing with (i) outliers in the data, (ii) complexity of the terrain, (iii) small vegetation which is completely attached to the terrain like bushes. Most of filter algorithms start with the minimum height in data. Thus the most effective error is the negative outliers which are originated from multi path errors and errors in range finder. The vegetation on the slope also produces difficulties in filter algorithms because of the reflected pulses returning from the neighbor points. Therefore, filtering methods need some initial threshold values, which
are usually defined by experience and a-priory information about the data and terrain characteristics.

Fallah Vazirabad and Karslioglu (2011) demonstrate that the quality of segmentation filter deteriorates with increasing point spacing of ALS point cloud looking at Type I and Type II errors (table 1). Large Type I error leads to a reduced DTM accuracy as a consequence, because many vegetation points will be included in DTM generation. The Type II error induces some effects resulting from the fact that measured elevation values in Lidar data are replaced by interpolated values for DTM, which cause a zig-zag pattern in the DTM modeling (figure 6).

Fig. 6. Poorly filtered (left), good filtered (right).

3.2 Methods and models

Extracting the forest characteristics from Lidar data for biomass estimation is classified into two categories, height distribution with its statistical analysis, and single tree detection containing its location and characteristics.

3.2.1 Methods and models used in biomass estimation

A conventional model of biomass estimation is introduced by Thomas et al. (2006), which is given as: $b \times dbh^2 \times height$, where $b$ is the coefficient. This equation was developed for the whole tree as well as the components of the stem wood, stem bark, branches, and foliage. As soon as the metrics (dbh and height) are measured for each plot, the equation can be established to estimate biomass and biomass components. The coefficient $b$ is a variable which is related to the species of trees. Measurements for the deriving forest biomass are destructive sampling which is the input of regression modeling. For this, sample trees are measured and then cut and weighted (Popescu et al, 2004). The mass of components of each tree is regressed to one or more dimensions of the standing tree. As discussed in the introduction section, biomass has also been estimated by means of previously developed models using Lidar which relies on tree characteristics extraction like height, dbh, and crown size. Crown size is not used directly in the estimation procedure but it is useful for extracting the tree species. All developed models and their parameters for biomass
estimation must be calibrated on the basis of tree characteristics. For this, four models were studied by Salmaca (2007). These are power function, Log transformed model, fractional power transformation, and explanatory function. The Power function is developed for North of USA, the Log transformed model is described by a linear function, the fractional power transformation is referred to linearized curvilinear model, and the explanatory function is constituted by a polynomial model. Under these models the Log transformed model is recommended which delivers the results with the unit of kilogram per every tree (Fallah Vazirabad, 2007). Consequently, tree characteristics extraction by Lidar data plays an important role in the biomass estimation model.

Bortlot et al. (2005) proposed to locate trees by image processing module assuming that the tree crown is circular, trees are taller than surroundings, and tree tops tend to be convex. They used the data of small footprint Lidar system. The algorithm starts by generating a CHM and works by shadow search method to find the crown boundaries which is related to tree tops. After defining a threshold and fitting the circles to the smoothed and generalized CHM, the circles should present the top of actual trees. The algorithm eliminates the small trees which are close to tall ones, because it searches for related high point neighboring. They conclude that tree heights are associated with canopy volume and therefore should be related to the biomass. They used the tree heights detected from image processing as variables for a stepwise multiple linear regression to find an equation for biomass prediction. They evaluated the results with highly significant (>95%) carrying out an efficient field measurement to calibrate the number of trees which are detected by an algorithm based on their height. Small trees are not included in this evaluation.

Lefsky et al. (1999) developed equations relating height indices to canopy area and biomass. They indicated that there are some differences in the predictive ability of the height indices; these differences are small, and statistically nonsignificant. However, the canopy structure information which is summarized in the median, mean, and quadratic mean canopy height indices, improved the stand canopy estimation related to the maximum canopy height. They defined the relation between tree height, H and dbh as: \[ \text{dbh} = (H/19.1)^{21} \]. They concluded that the result of the model using stepwise multiple regressions causes a higher variance value than those from the simple linear regression referring to the CHM. But, the predictions of the stand attributes were less applicable to the CHM than the height indices. Stepwise multiple regressions of basal area and biomass using the canopy height profile vector as independent variables increase the importance of the field measured regression equations.

Fallah Vazirabad and Karslioglu (2009) investigated the biomass estimation with the method of single tree detection. Lidar data segmentation filtering method is applied to point clouds to distinguish canopy points from the terrain points which are used for the generation of a DTM. The CHM is obtained by subtracting the DSM (from original data) from DTM. A single tree detection method is employed to locate trees and detect the height of each tree top. Diameter at breast height (at 1.37 m from ground) is extracted from the close relation with the tree height which is defined by field measurements for the evaluation. A Log transformed model is applied for biomass estimation on the basis of the dbh variable.

### 3.2.2 Single tree detection, tree characteristics detection

The objective of many previous studies was to validate the tree detection, tree height estimation, crown size estimation for volume and biomass estimation of different forest...
types. Nelson et al (1988) used discrete Lidar data to collect forest canopy height data. Two logarithmic equations were tested to find the best model. They used a height distribution method and analyzed a statistical approach. Falkowski et al (2006) described and evaluated spatial wavelet analysis techniques to estimate the location, height, and crown diameter of individual trees from Lidar data. Two dimensional hat wavelets were convolved with a CHM to identify local maxima within the wavelet transformation image. Maltamo et al. (2004) examined the CHM local maxima search method for high dense forest regions to detect individual trees. Because of the dense understory tree layer in most area, about 40% of all trees were detected. However, the detected tree heights were obtained with an accuracy of ±50 cm.

Anderson et al. (2006) developed a methodology for acquiring accurate individual tree height field measurements within 2 cm accuracy using a total station instrument. They utilized these measurements to establish the expected accuracy of tree height derived from small and large footprint Lidar data. It turned out that the accuracy of small footprint Lidar data changes according to the tree species. The comparison has shown that tree heights which are retrieved from small footprint Lidar are more accurate than the result of large footprint data. Hopkinson (2007) investigated the influence of flight altitude, beam divergence, and pulse repetition frequency on laser pulse return intensities and vertical frequency distributions within a vegetated environment. The investigation showed that the reduction in the pulse power concentration by widening the beam, increasing the flight altitude, or increasing the pulse repetition frequency results in (i) slightly reduced penetration into short canopy foliage and (ii) increased penetration into tall canopy foliage, while reducing the maximum canopy return heights.

Yu et al. (2004) demonstrated the applicability of small footprint, multi return Lidar data for forest change detection like forest growth or harvested trees. An object oriented algorithm was used for tree detections referred to the tree to tree matching method and statistical analysis. The small trees could not be detected by the algorithm. The forest growth is estimated about 5 cm in canopy crown and 10-15 cm in tree height.

Fallah Vazirabad and Karslioglu (2010) used a technique based on the searching for the local maximum canopy height to detect individual tree with variable window size and shape. The method detects tree location, number of trees, and the height of each single tree. The variable window size and shape solved the problems of small tree detection and not detectable CHM margin regions. The importance of field measurements and reference information (like orthophoto) are emphasized for evaluation. Popescu and Zhao (2008) developed a method for assessing crown base height for individual tree using Lidar data in forest to detect single tree crown. They also investigated the Fourier and wavelet filtering, polynomial fit, and percentile analysis for characterizing the vertical structure of individual tree crowns. Fourier filtering used for smoothing the vertical crown profile. The investigation resulted in the detection of 80% of tree crown correctly.

Moorthy et al. (2011) utilized terrestrial laser scanning to investigate the individual tree crown. From the observed 3D laser pulse returns, quantitative retrievals of tree crown structure and foliage were obtained. Robust methodologies were developed to characterize diagnostic architectural parameters, such as tree height ($R^2 = 0.97, \text{rmse} = 0.21 \text{m}$), crown width ($R^2 = 0.97, \text{rmse} = 0.13 \text{m}$), crown height ($R^2 = 0.86, \text{rmse} = 0.14 \text{m}$), crown volume ($R^2 = 0.99, \text{rmse} = 2.6 \text{m}^3$). It seems that the first pulse return from the upside view of an individual tree in terrestrial laser scanning brought about the low performance in crown height while the other characteristics are detected well.
Riano et al. (2004) estimated leaf area index (LAI) and crown size using Lidar data. They concluded that LAI was better estimated using larger search windows while the crown size was better estimated using small window size. They generated the vegetation height above the ground for each laser pulse using interpolated values extracted from DTM. DTM was produced using the bisection principle. They also applied spline function interpolation in order to obtain the height above the ground. But in this work it is not obvious whether first or last return has been used to extract the canopy height, effecting the result significantly.

3.3 Applications
To provide reliable results on tree location, height, and number of detected trees the local maximum detection method is introduced by Vazirabad and Karslioglu (2009). This method determines the canopy height by applying a variable window size. The window size selection is related to the height and density of trees. High trees were easier to detect with large windows while short trees were easier to detect with small windows. The derivation of the appropriate window size to search for tree tops relies on the assumption that there is a relation between the height of trees and their crown size. In the 100*100 m test area, the correctness of single tree detection was calculated approximately 91%. The main reason for 9% error is referred to the not detected trees which are located in the corners and edges of the searched patch. To deal with this problem, the standard rectangle windows, variable size and variable shape are recommended (figure 6).

![Fig. 6. Search windows (left); Single tree detection, CHM horizontal view (right-back), test patch 5 (right-top corner), respected orthophoto (center), and result (right-bottom)](image)

Four window sizes such as standard 3*3 m, standard 5*5 m, rotated 3*3 m (5*5 m), and rotated 5*5 m (9*9 m) are employed (each pixel represents one meter). Tree heights from CHM show that they vary between 2 m to 25 m (figure 6, right). The single tree detection method works in several steps. First generation of a tree height model is required to obtain the tree height. In this model the algorithm looks for all nonzero values and then creates a sorted list depending on the point height above ground (reducing data makes searching procedure faster). In the second step a tree height specific filtering is accomplished, by moving the window pixel by pixel over the tree height model. By changing the window size and shape repeatedly the procedure is continuing up to the end. Six reference patches are
provided for counting manually the number of trees by using orthophotos. Density and height of trees are variable inside the patches. The total 7479 trees are detected in whole 1*1 km². Tree height, dbh, and crown diameters are estimated in the whole area. All this information is adapted to the Log Transformed model for biomass estimation. Hence the total biomass which is given in kilograms for every tree in vegetation cover area is calculated as 1,966,123.3 kg.

Fig. 8. Biomass model and dbh

4. Conclusion

A comprehensive review has been done within this chapter concerning the use of Lidar for biomass estimation. As a consequence it can be said that the reasons for the underestimation of biomass in relation to the tree height need further studies. The development of large footprint Lidar systems on the spaceborne platform GLAS will allow the biomass estimations on a global scale. Spaceborne systems are restricted to record regional and detailed forest data mainly due to the ground track resolution of the system. However, since they receive data continuously, biomass estimation and carbon storage studies are possible every time which can be regarded as a great benefit. Airborne Lidar has the advantages of variable height flying systems and hence collects more precise data with respect to the shape of the terrain. Taking advantages of intensity information from Lidar data provides more information about the interpretation of the ground surface. There are several full waveform airborne Lidar operational systems. But some substantial challenges still exist such as the huge data processing and the interpretation of waveform for complex objects like trees. The fast progresses in computer technologies will help overcome such problems. On the other hand, the high point density in terrestrial systems can help to evaluate the results of other systems. Besides, it allows to model vegetation canopy characteristics particularly concerning tree species estimations in detail. From the data acquisition point of view, it is obvious that models and methods need to exploit the whole potential of the full waveform data for biomass estimation in future. The investigation on the point density in Lidar data represents that having a sufficient number of points has a large impact on the filtering results. The result of the segmentation filtering shows a high capability of adaptation in different landscapes. But it requires choosing correct segmentation parameters by
considering the point density. Point spacing plays also an important role for the selection of the interpolation method with respect to the DTM, DSM, and CHM resolution. The methods for individual tree detection which are described and evaluated in the application part are performing well, but they are still under development. Hence more empirical studies are required for improving the quality of the approaches.

5. References


Cote et al, 2011; An architectural model of trees to estimate forest structural attributes using terrestrial LiDAR, Environmental Modelling & Software XXX, pp. 1-17.


Biomass has been an intimate companion of humans from the dawn of civilization to the present. Its use as food, energy source, body cover and as construction material established the key areas of biomass usage that extend to this day. Given the complexities of biomass as a source of multiple end products, this volume sheds new light to the whole spectrum of biomass related topics by highlighting the new and reviewing the existing methods of its detection, production and usage. We hope that the readers will find valuable information and exciting new material in its chapters.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following: