Estimation of aboveground biomass using airborne LiDAR data

ANDREAS JOCHEM *†, MARKUS HOLLAUS ‡, MARTIN RUTZINGER §, BERNHARD HÖFLE ¶, KLEMENS SCHADAUER |, BERNHARD MAIER ¤

† alpS-Centre for Natural Hazard Management, Grabenweg 3, 6020 Innsbruck, Austria, (jochem@alps-gmbh.at)

‡ Vienna University of Technology, Institute of Photogrammetry and Remote Sensing, 1040 Vienna, Austria

§ Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7500 AE Enschede, The Netherlands

¶ University of Heidelberg, Department of Geography, 69120 Heidelberg, Germany

| Department of Forest Inventory at the Federal Research and Training Center for Forests, Natural Hazards and Landscape, Seckendorff-Gudent-Weg, 1130 Vienna, Austria

¤ Stand Montafon Forstfonds, Montafonerstraße 21, 6780 Schruns, Austria

Abstract

In this study a semi-empirical model that was originally developed for stem volume estimation is used for aboveground biomass (AGB) estimation. The semi-empirical model is based on the relative heights of first echo LiDAR point cloud data and assumes a linear relationship between AGB and canopy volume. However, the usage of point cloud data leads to a computationally demanding task when processing large point cloud datasets for the generation of area-wide AGB maps. In the presented study the effects of using rasterized LiDAR data as input for the AGB model are investigated in order to speed up processing and to make use of the model on large spatial datasets. The canopy volumes are calculated from a Canopy Height Model (CHM). The optimum resolution of the CHM is determined by analyzing the effects of varying cell sizes (1.0 m, 1.5 m, 2.0 m, 3.0 m) on the achievable accuracies. Calibrating the model with rasterized input data having a spatial resolution of 2.0 m instead of using first echo point cloud data leads to a slight increase of the coefficient of determination ($R^2 = 0.70$ to $R^2 = 0.72$) and a slight decrease of the standard deviation of the prediction errors. For calibrating the model reference AGB is calculated per sample plot from local forest inventory data by means of averaged weighted (according to tree species and age class composition) extension factors. The influence of using rasterized LiDAR input data on the achievable accuracy of the assessed AGB is investigated for a coniferous dominated study area in Vorarlberg, Austria.

1 Introduction

Aboveground biomass (AGB) is defined as the total amount of aboveground oven dry mass of a tree that is expressed in tons per unit area (Brown 1997). Accurate estimation of AGB, also referred to as dry total biomass, in forested areas provides an indication of the potential energy that is stored in cellulosic material. Gaining knowledge about the spatial distribution of the potential bioenergy is essential for developing sustainable low carbon climate friendly

* Corresponding author. Email address: jochem@alps-gmbh.at

strategies such as the optimization of timber harvesting chains. In contrast to time consuming and expensive field methods remote sensing is capable for mapping area-wide forest inventory (FI) data in a less expensive, fast and accurate way. This procedure is mainly based on the extrapolation of FI data measured at stand or plot level. In recent years Airborne Laser Scanning (ALS), also referred to as Light Detection And Ranging (LiDAR), has been established as a standard technology for the acquisition of high precision topographic data and has been widely used for mapping vegetation and forest inventory data, respectively (Lim et al. 2004, Næsset et al. 2004, Hyyppä et al. 2008). In contrast to multi-spectral satellite imagery or arial photographs ALS data represent the horizontal and vertical distribution of the forest canopies and does not suffer from saturation in spectral response to dense canopies with high biomass (Zhao et al. 2009).

There are two methodological approaches for utilizing LiDAR data for AGB assessment. The (i) single-tree-based approaches and the (ii) area-based approaches. Both approaches mainly involve the use of empirical or semi-empirical models by using linear or non-linear regression analysis (e.g. Næsset 2004a, Popescu 2007, García et al. 2010). Single tree based approaches require LiDAR data with high point densities (>5 points/m²) and are mostly based on regression models focusing on a relationship between LiDAR derived individual tree parameters (e.g. tree height, crown dimensions) and field based estimates of AGB. Area-based methods can also be used for lower point densities but require an extensive set of reference data. Such methods analyze the vertical distribution of the laser echoes at stand or plot level by deriving various statistical quantities and estimate area-based forest inventory parameters (e.g. mean tree height, basal area, stem volume) and AGB, respectively.

LiDAR based estimations of AGB can be performed by means of both rasterized and point cloud data. Using rasterized data requires an aggregation of the 3D point cloud to 2.5D raster cells, meaning that the canopy surface is represented by a single-valued function. This procedure is accompanied with an irreversible loss of the 3rd dimension but makes processing less time consuming and reduces the amount of the storage size drastically. Hence, using rasterized data for the generation of area-wide digital high resolution maps is computationally more efficient.

In this paper a semi-empirical model (Hollaus et al. 2009) that was originally developed for stem volume estimation is used for AGB estimation. Furthermore, the model is investigated concerning the effects on the achievable accuracies of using rasterized instead of point cloud input data in order to make processing a less computationally intensive task and to apply the model on large spatial datasets. The optimum resolution of the rasterized input data is determined by analyzing the effects of different cell sizes on the accuracy of the assessed AGB. For this study an area of about 560 km² in the western part of Austria is analyzed.

This paper is structured as follows: In Section 2 the study area, the available ALS datasets as well as the estimation of the reference AGB are presented. The methodology including a short description of the semi-empirical model is subject of Section 3. In Section 4 the results are presented and discussed. A conclusion and outlook on future studies are given in Section 5.

2 Study area and datasets

The coniferous dominated study area is located in the southern part of the Federal State of Vorarlberg (Austria) in the so-called Montafon region, and covers an area of about 560 km². The main tree species in the area are Norway spruce (*Picea abies*) with 96% and fir (*Abies alba*) with 3%. The used ALS data are provided by the Land Survey Adminstration Feldkirch and were retrieved during several flight campaigns in the framework of a Vorarlberg-wide terrain mapping project. The ALS data were acquired under snow-free conditions in the years 2002 to 2004 using Optech Mapper systems (ALTM 1225, ALTM 2050) and Leica ALS-50 scanner. The Optech sensors have a beam divergence of 0.3 mrad and the ALS-50 scanner of

0.33 mrad, which resulted in a mean footprint diameter of 0.33 m and 0.36 m, respectively for the average flying height of 1100 m above ground. The mean point density within the study area varies between 0.9 point/m² and 2.7 points/m².
Besides the original ALS point cloud data a Digital Surface Model (DSM) and a Digital Terrain Model (DTM) with a spatial resolution of 1 m are available. The DTM was generated by using last echoes only and applying the hierarchic robust filter technique as described e.g. in Kraus and Pfeifer (1998). By subtracting the DTM from the DSM a Canopy Height Model (CHM) is produced. The relative height value of each laser point is derived by subtracting the underlying DTM elevation from each laser point.
For the investigated forests, FI data from 500 sample plots, which are regularly distributed in a 350 m grid are available. They were provided by the forest administration Stand Montafon Forstfonds and were collected in the year 2002 using the angle count sampling method (Bitterlich 1948), meaning that the plot areas and number of sampled trees may vary strongly from sample plot to sample plot. For each sample plot the stem volume per unit area [m³ ha^{-1}] was determined from tree specific parameters, such as tree species, tree height and diameter at breast height (DBH). The possible inaccuracies in the spatial positions between the LiDAR data and the FI data are corrected by performing co-registration as described in Dorigo et al. (2009).
The reference AGB that is used as target variable in this study and is estimated by means of averaged weighted (according to tree species and age class composition) extension factors taken from Weiss et al. 2000. The extension factor is determined for each sample plot separately. The AGB is estimated from the stem volume that is assessed from FI data for each sample plot as described in Hollaus et al. (2009). The first step contains the transformation of the estimated stem volume into dry stem biomass by using tree specific average raw density factors (Weiss et al. 2000, p.29). In the following step the dry stem biomass is converted in dry total tree biomass by means of extension factors given in Weiss et al. (2000, p.31).

3 Methods

3.1 Semi-empirical model

The semi-empirical model is based on the assumption that there is a linear relationship between AGB and the canopy volume (V_{can}) that is defined as the entire volume between the terrain surface and the topmost tree surface. V_{can} is determined for circular areas A around the center coordinates of the FI sample plots and is based on the relative heights of the first echoes. The relative heights are derived by subtracting the DTM height from the absolute heights of the first echoes. In order to take the height-dependent differences in canopy structure into account the relative height above terrain surface of each first echo point is used to classify the points into m different height classes, whereas all points having a relative height value of less than 2.0 m are classified as points being reflected from the ground, stones or bushes (Naesset, 2004b) and are not included into the canopy volume calculation. Based on the findings of the study from Hollaus et al. (2009) four canopy height classes having a canopy height interval of 10 m are most suitable for the calculation of the canopy volume. $V_{can,1}$ ranges between 2 m and 12 m, $V_{can,2}$ ranges between 12 m and 22 m, $V_{can,3}$ ranges between 22 m and 32 m and $V_{can,4}$ contains all first echoes having a relative height greater than 32 m. $V_{can,i}$ is calculated as:

$$V_{can,i} = p_{fe,i} * ch_{mean} \tag{1}$$

where ch_{mean} is the mean canopy height of all first echoes within the corresponding canopy height class. $p_{fe,i}$ is the relative portion of first echo points (between 0 and 1) within the canopy height range i. The linear regression model is formulated as:

$$AGB = \sum_{i=1}^{m} \beta_i * V_{can,i} \tag{2}$$

where β_i are the unknown model coefficients that can be interpreted as the fraction of the corresponding canopy volume that is occupied by AGB.

3.2 Determining of the optimum sample plot size

The estimation of the reference AGB is based on FI data that was collected using the angle count sampling method (Bitterlich 1948), meaning that the reference AGB is not related to a defined sample plot size. The challenge is to find the optimum sample plot size in order to allow a proper comparison of the ALS data with the FI data. Therefore, an approach introduced by Hollaus et al. (2007) is chosen that analyzes different sample plot sizes with radii ranging from 8.0 m to 16.0 m. This procedure is based on all co-registered sample plots containing at least 90% coniferous trees. The 90% coniferous trees threshold is introduced to avoid effects of different flight dates (winter/summer) on the calibration of the model for different tree species i.e. coniferous and deciduous trees. The sample plot size leading to the highest coefficient of determination (R^2) and the lowest standard deviation (SD) of the residuals derived from cross-validation (Section 3.4) is taken for further analyses.

3.3 Effects of rasterizing input data

Generating biomass maps for large spatial datasets based on first echo point cloud data is very time consuming and computationally intensive. Using rasterized data as input for the semi-empirical model speeds up processing and overcomes the disadvantages related to point cloud data (huge amount of storage size, computationally intensive task when performing spatial queries within the point cloud) when computing large area AGB maps. The effects of using rasterized input data on the achievable accuracies are analyzed by using a CHM for deriving the canopy volumes. The CHM is generated by aggregation of all first echo points into a regular grid, whereas the maximum relative elevation is chosen as cell value. Cells containing no laser point at all, obtain an elevation value of zero and consequently are not considered for the calculation of the canopy volumes. The optimum spatial resolution of the CHM is determined by investigating the effects of different cell sizes (1.0 m, 1.5 m, 2.0 m, 3.0 m) on R^2 and SD of the prediction errors, respectively.

3.4 Validation of the semi-empirical model

The predictive accuracy of the calibrated model is assessed by performing a leave one out cross validation procedure. This means that the model is fitted *n* times (*n* is the number of available sample plots), whereas for each step one sample plot is excluded and serves for the calculation of the prediction error. The remaining sample plots are used for the calibration of the model. Finally, one gets *n* prediction errors that are used for the calculation of statistical parameters such as minimum, maximum, mean and standard deviation.

4 Results and discussion

4.1 Determination of the optimum sample plot radius

The 90% coniferous trees threshold resulted in 450 out of 488 successfully co-registered sample plots that are taken as input for the determination of the optimum circular sample plot size. A sample plot radius of 12.0 m results in the highest R^2 (0.66) and the lowest SD of the prediction errors (109.0 tha^{-1}).

Sample plot radius [m]	8.0	10.0	12.0	14.0	16.0
R^2	0.60	0.64	0.66	0.64	0.61
SD [tha^{-1}]	120.2	111.4	109.0	111.2	115.7

Table 1: Determination of the optimum circular sample plot size. Various radii are analyzed according to their R^2 and SD of the prediction errors.

This procedure is followed by the selection of sample plots containing all trees measured in the field for reference AGB estimation within a radius of 12.0 m. This results in a selection of 196 out of 450 sample plots that are taken for the calibration of both the point cloud and the raster based semi-empirical model.

4.2 Calibration of the point cloud based model

The calibration of the point cloud based model is based on the 196 selected sample plots (Section 4.1) using the optimum sample plot radius of 12.0 m and four canopy height classes having a height interval of 10.0 m. The calibrated model achieved an R^2 of 0.70 and a SD of the prediction errors of 87.6 tha^{-1} (37.0 %). The scatter plot of the reference AGB versus the AGB estimated from LiDAR data is shown in figure 1.

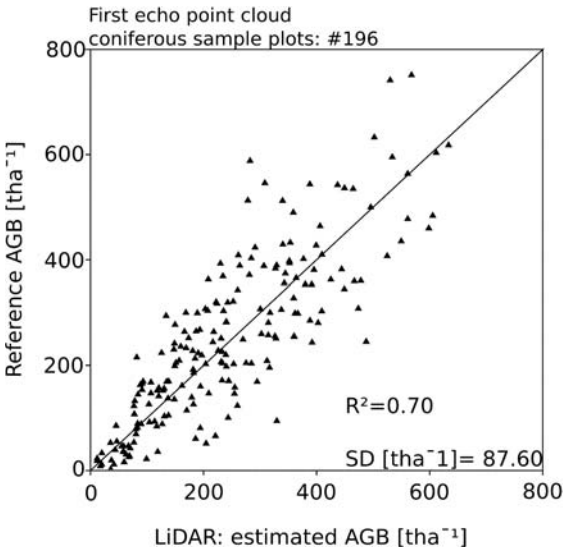

Figure 1: Scatter plot of reference AGB versus AGB estimated from first echo LiDAR data.

Hollaus et al. (2009) achieved R^2 values up to 0.86 for the estimation of stem volume. The deviations to the presented approach can be explained by several reasons: (i) they used stem volume instead of AGB, (ii) the conversion of stem volume per sample plot to the reference AGB is accompanied with several uncertainties and (iii) the different sizes of the study areas. The study area in Hollaus et al. (2009) is about three times less in size and results in the usage of 103 reference sample plots. Investigating a larger study area means both an increase of the heterogeneity of the LiDAR data (e.g. varying flying heights, acquisition dates, different point densities) and an increase of the spatial variability of the forest stand properties. The latter is due to the alpine topography of the Montafon region that influences forest growing conditions.

4.3 Effects of rasterizing input data

Calibrating the model with canopy volumes derived from rasterized instead of point cloud data does not change the achievable accuracy significantly. A slight increase of R^2 (0.72) as well as a slight decrease of SD of the prediction errors (84.11 tha^{-1}) can be observed when using a CHM having a cell size of 2.0 m. This is considered to be the optimum spatial resolution when generating AGB maps for the investigated study area. The effects of varying cell sizes, ranging from 1.0 m to 3.0 m on the accuracy statistics and the β coefficients are shown in table 2. Concerning R^2 and SD of the prediction errors the AGB model seems to be very robust against the varying cell sizes of the CHM. Analyzing the β coefficients confirmed the findings of Hollaus et al. (2009) where canopy heights between 22 m and 32 m are the highest contributors to growing stock and AGB, respectively. This applies to both the first echo LiDAR based (Section 4.2) and the CHM based AGB model. However, the coarser the spatial resolution of the CHM gets, the lower the values of the β coefficients $\beta 1, \beta 2, \beta 3$. This can be explained by the overestimation of the canopy volume with increasing cell size as small gaps in the forest canopy are neglected. The variation of the corresponding $\beta 4$ values is less distinct. The CHM is generated by using the maximum relative elevation within one cell as the corresponding cell value, meaning that points of lower canopy height classes are not considered anymore if one or more points within a cell fall into a upper canopy height class. Hence, the bigger the cell size the lower the fraction of the lower canopy height classes to the reference AGB. The minor change of the $\beta 4$ values might be due to the small amount of first echo points falling into the canopy height class greater than 32 m.

	Point Cloud			**Resolution of CHM**	
Parameters	**(0.9 – 2.7 p/m²)**	**1.0 m**	**1.5 m**	**2.0 m**	**3.0 m**
R^2	0.70	0.70	0.70	0.72	0.71
SD [tha^{-1}]	87.60 (37.0%)	88.84 (37.5%)	88.60 (37.2%)	84.11 (34.4%)	84.6 (34.6%)
$\beta 1$	$7.71 \cdot 10^{-4}$	$6.97 \cdot 10^{-4}$	$4.56 \cdot 10^{-4}$	$2.70 \cdot 10^{-4}$	$1.38 \cdot 10^{-4}$
$\beta 2$	$19.91 \cdot 10^{-4}$	$19.19 \cdot 10^{-4}$	$17.34 \cdot 10^{-4}$	$14.62 \cdot 10^{-4}$	$10.98 \cdot 10^{-4}$
$\beta 3$	$29.75 \cdot 10^{-4}$	$28.36 \cdot 10^{-4}$	$25.40 \cdot 10^{-4}$	$22.80 \cdot 10^{-4}$	$18.87 \cdot 10^{-4}$
$\beta 4$	$15.87 \cdot 10^{-4}$	$15.85 \cdot 10^{-4}$	$15.84 \cdot 10^{-4}$	$15.46 \cdot 10^{-4}$	$15.48 \cdot 10^{-4}$

Table 2: Accuracy statistic of the AGB model when using rasterized and point-based input data. R^2, SD of the prediction errors and the estimated β coefficients are shown.

The scatter plots of the reference AGB versus the AGB estimated from the canopy volumes derived from a CHM are shown in figure 2.

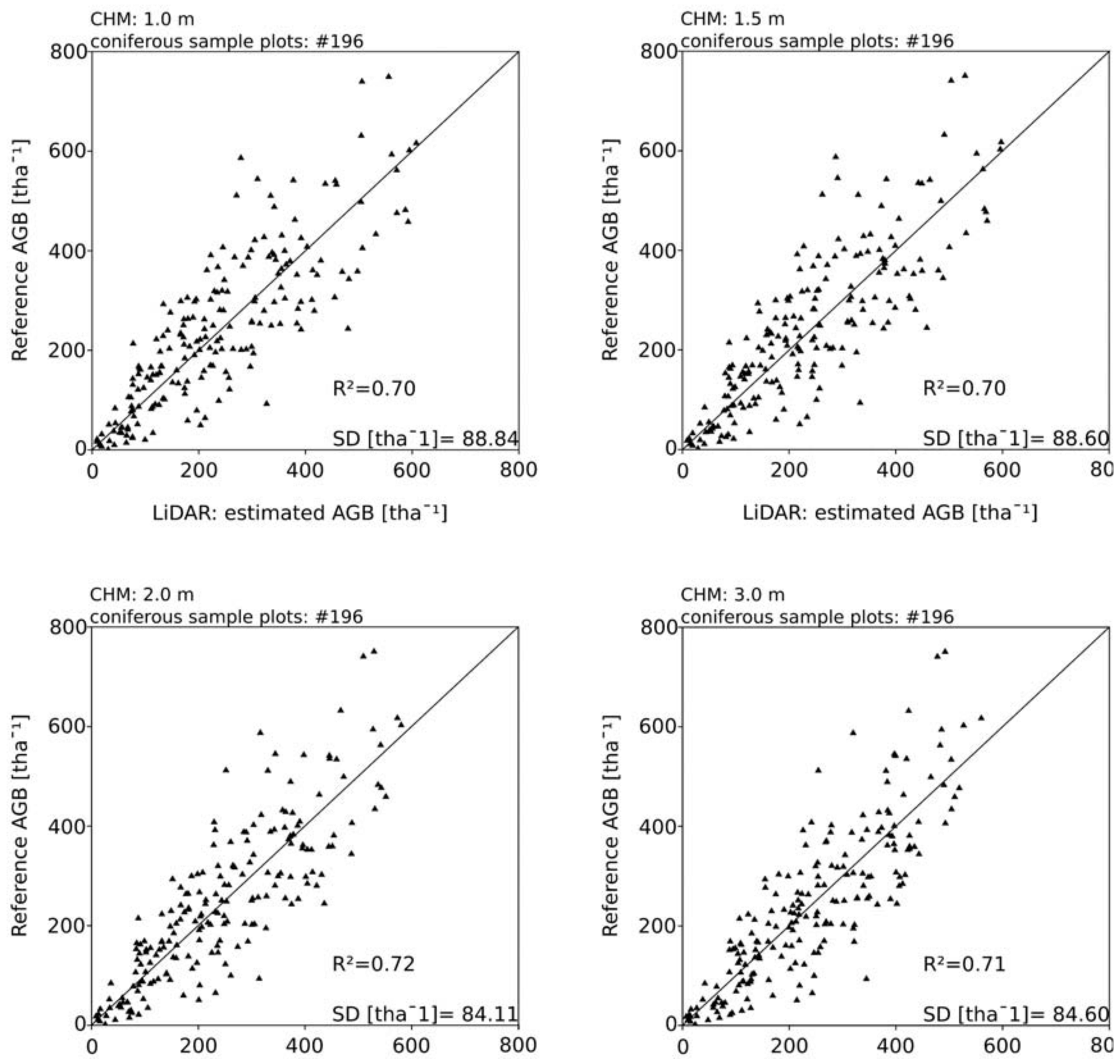

Figure 2: Scatter plots showing AGB derived from in-situ measurements versus AGB estimated from canopy volumes derived from rasterized input data. The x-axis represents the AGB of the sample plots estimated by the AGB model. The y-axis shows the reference AGB calculated from in-situ measurements.

5 Conclusion and outlook

In the presented study a semi-empirical model that was originally developed for stem volume estimation is investigated concerning its reliability for AGB estimation. The semi-empirical model assumes a linear relationship between AGB and canopy volume that is derived from first echo LiDAR point cloud data. Furthermore, the effects of using rasterized LiDAR data as input for the AGB model are analyzed in order to make processing a less computationally demanding task when applying the model on large spatial datasets. Therefore, the canopy volume is calculated from the 3D first echoes and the CHM, respectively. The effects of varying cell sizes on the achievable accuracy are investigated to find the optimum spatial resolution of the CHM-based canopy volume calculation. The results show that the semi-empirical model can also be used for AGB estimation of a spruce-dominated alpine forest and that AGB maps can be generated by means of rasterized input data. The usage of first echo point cloud data with leads to a R^2 of 0.70 and a SD of the prediction errors of 87.6 tha^{-1}. Calibrating the model with canopy volumes derived from a CHM does not change R^2 and SD

of the prediction errors significantly. A spatial resolution of 2.0 m leads to the highest R^2 (0.72) and the lowest SD of the prediction errors (84.11 tha^{-1}).
Future studies will concentrate on the application of the AGB model on regions that are characterized by both a wider range of tree species and higher point densities. Due to the different crown properties of deciduous and coniferous trees it is expected that the consideration of tree species (deciduous versus coniferous trees) in the AGB model will increase the accuracy of the assessed AGB. Furthermore, the improvement of the calculation of the reference AGB will be in the focus of future research including the consideration of the biomass compartments (stem, branches, needles, foliage).

6 Acknowledgements

The authors would like to thank the Landesvermessungsamt Feldkirch (Austria) for granting the use of the ALS data, and the Stand Montafon Forstfonds for supplying the forest inventory data. This study was carried out within the project LASER-WOOD (822030) which is funded by the Klima- und Energiefonds in the framework of the program NEUE ENERGIEN 2020.

References

Bitterlich, W., 1948. Die Winkelzählprobe. *Allgemeine Forst und Holzwirtschaftliche Zeitung* 59, pp. 4–5.

Brown, S., 1997. Estimating biomass and biomass change in tropical forests: a primer. In *FAO Forestry Paper,* Vol. 134.

Dorigo, W., Hollaus, M., Schadauer, K., Wagner, W., 2009. An application-oriented automated approach for co-registration of forest inventory and airborne laser scanning data. *International Journal of Remote Sensing*, 31(5), pp. 1133 – 1153.

García M., Riaño, D., Chuvieco, E., Mark Danson, F., 2010. Estimating biomass carbon stocks for a mediterranean forest in central spain using LiDAR height and intensity data. *Remote Sensing of Environment* 114(4), pp. 816–830.

Hollaus, M., Wagner, W., Maier, B., Schadauer, K., 2007. Airborne laser scanning of forest stem volume in a mountainous environment. *Sensors* 7(8), pp. 1559–1577.

Hollaus, M., Wagner, W., Schadauer, K., Maier, B., Gabler, K., 2009. Growing stock estimation for alpine forests in Austria: a robust lidar-based approach. *Canadian Journal of Forest Research* 39(7), 1387-1400.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing* 29(5), pp. 1339–1366.

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing* 54(4), pp. 193–203.

Lim, K., Treitz, P., 2004. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. *Scandinavian Journal of Forest Research* 19, pp. 558– 570.

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, A., Söderman, U., 2004. Laser scanning of forest resources: the nordic experience. *Scandinavian Journal of Forest Research* 19(6), pp. 482–499.

Næsset, E., 2004a. Estimation of above- and below-ground biomass in boreal forest ecosystems. In *International Archives of Photgrammetry, Remote Sensing and Spatial Information Sciences*. Vol. XXXVI (8/W2). Freiburg, Germany, pp. 145–148.

Næsset, E., 2004b. Practical large-scale forest stand inventory using a small foot print airborne scanning laser. *Scandinavian Journal of Forest Research* 19(2), pp. 164–179.

Popescu, S., 2007. Estimating biomass of individual pine trees using airborne LiDAR. *Biomass and Bioenergy* 31, pp. 646–655.

Weiss, P., Schieler, K., Schadauer, K., Radunsky, K., Englisch, M., 2000. Die Kohlenstoffbilanz des Österreichischen Waldes und Betrachtungen zum Kyoto Protokoll. *Tech. Rep.* 106, Umweltbundesamt, Wien.

Zhao, K., Popescu, S., Nelson, R., 2009. LiDAR remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment* 113(1), pp. 182–196.