# Prediction of Timber Quality Parameters of Forest Stands by Means of Small Footprint Airborne Laser Scanner Data

**Ole Martin Bollandsås***
**Matti Maltamo**
**Terje Gobakken**
**Vegard Lien**
**Erik Næsset**

## Abstract

The aim of this study was to explore the capability of airborne laser scanner (ALS) data to explain the variation in field-measured variables representing timber quality within square 0.25 ha grid cells in a mature conifer forest in the southeast of Norway. These variables were the mean ratio between stem diameter at six m above ground and the diameter at breast height ($R_{D6}$), the volume of saw logs ($V_{SL}$), the proportion of saw logs relative to the total volume ($P_{SL}$), the ratio between tree height and diameter at breast height (HD), mean basal area diameter ($D_g$), and crown height (CH). Each of these variables was modeled using a mixed modeling approach. Model fit was expressed by the Pseudo-$R^2$, and were 0.85, 0.50, 0.78, 0.57, 0.74, and 0.58 for the respective quality variables. Furthermore, much of the residual error could be attributed to the different forest stands from which the grid cells originated even though we used field-observed tree species proportions as auxiliary information. It was concluded that more auxiliary information is needed to estimate models that are general across stands, but that the relationships between ALS-data and the quality variables considered here seem strong enough to be utilized for example to prioritize between stands in relation to harvest when specific quality distributions are sought.

*Keywords:* Airborne laser scanning, timber quality, forest stands, predictive models.

## Introduction

The quality of timber is dependent on several internal and external tree properties. Malinen et al. (2003, 2005) list the following key factors for internal quality: annual ring width, wood density, decay, amount of heartwood, location and proportion of compression wood, diameter and proportion of possible juvenile wood, pitch pockets, checks, and knot multitude. Also knot sizes can be added to the list together with the so-called green-knot-cylinder (e.g. Øvrum et al. 2008, Ikonen et al. 2009). The green-knot-cylinder is defined as the maximum diameter of a cylinder in the trunk where all knots are sound (living and completely fixed to the surrounding wood). The entire extent of these internal factors are not revealed before the logs are processed at the sawmill. However, visible external factors are correlated to these internal factors (e.g., Øvrum et al. 2008), enabling them to be assessed or predicted prior to harvest. Furthermore, some of the external factors directly determine timber quality and some of the internal factors are even visible on the outside of the trunk. External factors are here defined as knot sizes and multitude, stem bending, and taper. At the mills, the quality characterizations of logs are made based on the external factors and they are directly related to what will be the end product. For example, timber suited for load-bearing construction material is labeled differently compared to timber suited for panel boards. Thus, these quality assessments determine the range of utilization of the logs.

The authors' affiliations are: Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway; olebo@umb.no, telephone: +47 64 96 57 53; matti.maltamo@uef.fi; terje.gobakken@umb.no; vegard.lien@umb.no; and erik.naesset@umb.no.

*Corresponding author


International Journal of Forest Engineering
Volume 22, No. 1

The characterization of timber quality of a marked stand before harvest has been found to be very important, but laborious, costly and difficult. For example, in Finland there was an operational pre-harvest measurement system called PMP (PystyMittaus ja Palkanlaskenta) used in the 1970s and 1980s (Malinen 2003). It included measurements of diameter at breast height (DBH) from each tree to be cut and, furthermore, a lot of sample tree measurements of tree heights and upper diameters. Still, this system was found to be insufficient to describe tree quality accurately enough. In the period after the end of the 1980s, there have only been some visual-based pre-harvest assessment systems where the purpose has been to describe timber quantities only and not the quality. However, nowadays there is an increased need for such a system because of relatively small economical margins in the timber market, creating a need for precise information on both volume and quality to be able to deliver timber according to demand.

Forest stand density, most frequently expressed as basal area or stem number, is found to have a significant effect on timber quality. For example, Hein et al. (2008) found that increasing density was negatively correlated with maximum knot size in Douglas-fir stands in Germany. A relationship between density regulation by silviculture and diameter of branches has also been documented (e.g. Maguire et al. 1991, Schmidt 2004). The study of Hein et al. (2008) also found that branch angle in the middle part of the stem was decreasing with increasing density. Increasing branch angle negatively affects timber quality. Furthermore, the proportion of living branches was found to be lower for dense stands. The increasing proportion of dead branches means lower timber quality. Density regulation also has been found to influence timber quality in boreal conifer forest stands (e.g. Ikonen et al. 2008). Stand density will also affect the taper (Fulton 1999, Sharma and Zhang 2004) because diameter growth is more sensitive to increasing density than height growth (Assmann 1970). Taper, or the relationship between diameter and height, is a direct determinant of timber quality because it limits the dimensions of the final material for a given log length. At the tree level, inter-tree competition also affects the relationship between diameter and height (Loetsch et al. 1973). In this context it is not only density, defined as basal area, that is important, but also canopy structure.

Forest structure, defined as the vertical and horizontal distribution of the canopy and stems, is effectively measured by means of airborne laser scanner (ALS) data (Maltamo et al. 2005). ALS is a remote sensing technique where laser pulses are emitted in a scanning pattern from an airborne sensor. Hence, the laser pulses are distributed in a corridor on the ground along the flight line creating a 3-D representation of objects on the ground. The point density is usually in the range of one point per $m^2$ to 10 points per $m^2$ because data in this density range suit most of the current applications. In practice, however, it is possible to collect laser data of any density. The width of the corridor (the swath) depends on the maximum scan angle applied and the flying altitude, but is typically a few hundred meters. The laser pulses that are emitted from the sensor hit either the ground or whatever objects found within the swath, and then they are reflected back to the sensor. The sensor measures the elapsed time between emission and return, and thus enables the calculation of the distance between the sensor and the reflecting object. Georeferencing (xyz-coordinates) of each echo is possible by means of global navigation satellite systems (GNSS) for accurate positioning of the aircraft as the pulse is emitted, and inertial navigation systems (INS) that correct for yaw, pitch, and roll of the aircraft together with information describing the scan angle of a specific pulse. The planimetric position (xy) accuracies of the echoes are still dependent on the flying altitude, and the sensor providers report this to be 1/10,000 × flying altitude (Optech 2008). The elevation (z) accuracies of single echoes are typically 5 to 30 cm (Optech 2008). The ground echoes are distinguished from the vegetation echoes by mathematical algorithms (Axelsson 2000). Those echoes defined as ground reflections are then used to model the terrain. This model is usually called the digital terrain model (DTM) and typical vertical accuracies for such models are 20 to 30 cm (Hodgson and Bresnahan 2004, Kraus and Pfeifer 1998, Reutebuch et al. 2003). After the establishment of the DTM it is possible to calculate the height of all other echoes relative to the DTM. This yields a 3-D point cloud from which height and spatial distribution can be exploited for modeling purposes. For technical details on laser scanning, please refer to Wehr and Lohr (1999).

ALS data for forest inventory purposes have traditionally been applied using two different approaches. The most frequently used approach for predicting biophysical forest properties such as total volume, mean diameter, mean tree height, etc., has been to consider the height distribution of the laser echoes for a fixed area, for example, 250 $m^2$, which can be called the measurement unit or the resolution. First, a number of ground sample plots are measured in the forest for the properties of interest and accurately positioned. Then the ALS data collected for these plots are extracted and different variables describing the vertical and horizontal distribution of the laser echoes are calculated. Parametric or non-parametric models for the biophysical properties of interest are then estimated with the laser variables as independent variables. These developed models are then used to predict the dependent variables for each area unit where there only exist ALS data. In the literature, this approach is called the area-based method. For further details, please refer to Næsset (2002, 2004) and Packalén and Maltamo (2008). The second approach is the so-called single-tree method where single trees are considered instead of mean values over a certain area. With this approach single trees have to be recognized in the ALS data so that each echo is assigned to a specific tree. This process is called segmentation and is based on finding local maxima in the ALS data that are potential treetops and delineating the extent of the tree crown around these maxima. As a result, an estimate for tree height is obtained and volume and other properties on single-tree level are then predicted using regression models. Obviously, this approach gives data with higher resolution and level of details than the area-based method, because the measurement units are single trees. However, the segmentation process can be challenging. Many trees are shaded by larger trees and some trees are partly mixed together in clumps. This makes the delineation process difficult. Some trees can also be shaped in such a way that they appear with more than one local height maximum in the ALS data. Thus, a segmentation process can yield both commission errors (false trees) and omission errors (trees that are not detected). The segmentation process is especially difficult for complex canopy structures. Furthermore, the prediction of tree properties based on tree height at tree level is rather inaccurate. For more details about single-tree segmentation, please refer to e.g. Hyyppä et al. (2001), Persson et al. (2002), Morsdorf et al. (2004) and Solberg et al. (2006).

Numerous studies have taken advantage of the strong correlation between the forest structure and the three-dimensional point clouds resulting from ALS measurements. Several biophysical properties such as timber volume, mean diameter, basal area, dominant height, mean height, and stem number have successfully been modeled using ALS data (e.g., Næsset 2002, 2004). Even higher resolution forest information such as the distribution of diameters has been modeled by means of ALS (Gobakken and Næsset 2004, 2005, Bollandsås and Næsset 2007, Packalén and Maltamo 2008). The diameter distribution is a very interesting property to model because it enables volume calculations by tree size classes at the same time as it gives important information regarding the timber quality.

A few recent studies have tried to model quality parameters more directly related to those quality assessments that are made at the sawmill when classifying timber (Peuhkurinen et al. 2007, 2008, Fonweban et al. 2008, Korhonen et al. 2008, Moberg et al. 2008, Maltamo et al. 2009b). Timber quality assessments are highly relevant using the single-tree method. However, as mentioned above, omission and commission errors are inevitable. It is also still much more expensive to carry out a single-tree inventory compared to the area-based method. Nevertheless, some studies of the relationship between ALS and single-tree quality have been carried out. Peuhkurinen et al. (2007) segmented single trees from high pulse density ALS data (6.4 pulses $m^{-2}$). Diameters were estimated and Finnish taper functions were applied. Then these segmented trees were used in bucking simulations and the results were compared to real data collected by harvesters. The authors characterized their results as promising. Peuhkurinen et al. (2008) modeled proportions of saw log volume by means of the non-parametric k-Nearest Neighbor (k-NN) method and a "data bank" of actual bucked trees measured with a harvester. The results were not satisfactory, and they concluded that it might be expedient to include auxiliary information that explains the local height to diameter relationship to improve the predictions. Furthermore, Korhonen et al. (2008) predicted stand-level saw log volumes from laser data by means of a mixed effects regression approach. Compared to the results of Peuhkurinen et al. (2008), their results were more promising.

One of the most frequently considered quality parameters in ALS studies has been the crown height (CH). The definition of CH is the height from the stump to the beginning of the crown. The definition of the start of the crown may, however, vary between studies. Crown height has been characterized both at tree and stand levels (Næsset and Økland 2002, Maltamo et al. 2006, 2010, Dean et al. 2009). Crown height has either been predicted by using characteristics of the ALS height distribution (Næsset and Økland 2002, Maltamo et al. 2006, 2009b, 2010) or there has been direct analysis of the 3-D or height distribution of the laser-derived points (e.g., Morsdorf et al. 2004, Solberg et al. 2006, Popescu and Zhao 2008, Vauhkonen 2008, Maltamo et al. 2010). The accuracy of these approaches has varied from 1 m to about 4 m in terms of RMSE.

As indicated above, timber quality can also be expressed by variables representing taper. For example the ratio between diameters at 6 m and 1.3 m ($R_{D6}$), and the mean ratio between tree height and diameter at breast height (HD) are variables that represent this quality measure. The taper is also to some extent correlated to the mean diameter in a stand ($D_g$). However, $D_g$ is mainly interesting as a measure of quality because tree size per se is a determinant of the range of utilization of the trees in a stand. Both $R_{D6}$, HD, together with $D_g$ are correlated to vertical and horizontal forest structure. Since ALS accurately depicts forest structure, it should be possible to also model these measures of timber quality.

Based on prior studies, it seems possible to retrieve information about timber quality by means of ALS data. In this study we wanted to test the capability of ALS data to explain the variation of timber quality as expressed by $R_{D6}$, volume of saw logs ($V_{SL}$), proportion of saw logs ($P_{SL}$), HD, $D_g$, and CH.

## Materials and Methods

**Study Area**

This study was conducted in the Aurskog-Høland municipality (59°50'N 11°30'E, 120-390 m a.s.l), southeastern Norway. The total area of the municipality is 960 $km^2$. The dominant tree species in the area were Norway spruce (*Picea abies (L.) Karst*.) and Scots pine (*Pinus sylvestris L*.).

**Field Data Collection**

The field data were collected during winter and fall of 2007 as part of a large clear-felling operation of 29 different forest stands over a total area of 40 ha. The data were collected using John Deere 1070D harvesters equipped with Timberjack/John Deere H754 harvester heads. The bucking software was the Timbermatic 300, version 2.4.9. The harvesters recorded length and diameter of each stem for every 10 cm section along the stem. Volume of each stem was calculated section-wise and summed. Because the tops of the trees did not pass through the harvester head they were not measured. A function for the length of the top based on the last diameter measured by the harvester was therefore calibrated from maximum diameter and top length data observed in field. The volumes of the tops were calculated under the assumption that their shapes were strictly conical.

No exact tree coordinates were registered by the harvester, but each tree could be attributed to the position of the harvester when each tree was cut. The harvester used a single-frequency GlobalSat® SiRF Star III GPS-receiver observing pseudo-range. The antenna was mounted on the roof of the harvester. The accuracy of such systems is based on prior experience with similar receiver types, expected to be approximately 10 m. Furthermore, the harvester targeted mostly merchantable trees and some small trees were therefore left standing. Some trees were also left to meet biodiversity standards for forest operations. Thus, a field effort was carried out to measure single isolated trees left by the harvester. These trees were added to the stand data collected by the harvesters. If there were several remaining trees in a clump, the clump was geo-referenced by differential Global Navigation Satellite Systems (dGNSS) using a 20-channel dual frequency Topcon rover receiver and an identical receiver as a base station to correct the rover observations in real time (real-time-kinetic mode). The accuracy of these positions was approximately 50 cm. The accuracy assessment was made on the basis of estimated accuracies observed on the field controller for each position. The accuracy estimate also includes random errors related to the determination of the stand border in the field. The uncut tree clumps were kept out of the subsequent analyses. Our intention was that this should yield a complete inventory of every tree in the study area, either measured by the harvester or by subsequent manual field inventory if left untouched by the harvester. However, it is likely that some of the small trees were registered neither by the harvester nor in the field inventory because harvesters sometimes run them over.

The harvester's diameter measurements were recorded in millimeters. Breast height was set to 1.1 m from the stump and diameter at breast height (DBH) was measured at this height. Furthermore, $R_{D6}$ and HD were calculated for each tree and mean values were calculated for each measurement unit (see "Data preparation" below). These two attributes were chosen to describe average stem form. Moreover, mean diameter ($D_g$) defined as diameter corresponding to mean tree basal area (mean quadratic diameter) was calculated as well. Also, more direct quality indicators were available in this data because each log was classified by the harvester operator during the harvesting operation. Thus, we were able to calculate mean proportions of saw logs ($P_{SL}$), and saw log volume ($V_{SL}$). Finally, the operators of the harvesters recorded the crown height from each cut tree visually as the tree passed through the harvester head. Mean basal area weighted crown height (CH) was calculated for each measurement unit. The definition of CH was the height from the stump to the lowest point on the stem where at least two living branches where found in the same whorl. The data are dominated by pine and spruce and a summary of the data appears in Table 1. The different proportions for spruce, pine, and deciduous species are labeled $P_s$, $P_p$, and $P_d$, respectively. These proportions were calculated as the ratio between the volume of the respective tree species and total volume. All volumes were calculated from the harvester measurements.

**Laser Scanner Data**

Laser data for this study were acquired under leaf-on conditions on June 8-10, 2005. Additional data were acquired on September 6, 2005, to fill in a minor gap in the data acquired in June. Laser data were collected with an Optech ALTM 3100 laser scanner mounted on a fixed-wing aircraft that flew 75 m $s^{-1}$ in an altitude of 1850 m a.g.l. The pulse repetition frequency was 50 kHz and the scan angle 15°. However, pulses emitted at angles >13° were discarded during subsequent processing. These setup parameters yielded a point density on the ground of approximately 0.7 $m^{-2}$. The ALTM 3100 sensor is capable of recording up to four echoes per pulse. In this study, we used the two echo categories classified as "first of many" and "single." Echoes of these two categories were pooled into one dataset, and this aggregated dataset was denoted as "first" echoes. The laser data is also described by Maltamo et al. (2009a).

**Table 1.** Harvester ground data summary

| Variable | Abbreviation | N | Mean | Min | Max | SD |
|---|---|---|---|---|---|---|
| Ratio between diameter at 6 meter and DBH (cm/cm) | $R_{D6}$ | 256[a] | 0.75 | 0.65 | 0.82 | 0.03 |
| Volume of saw logs ($m^3ha^{-1}$) | $V_{SL}$ | 256[a] | 170.8 | 0.00 | 380.3 | 77.3 |
| Volume of saw logs relative to total volume | $P_{SL}$ | 256[a] | 0.65 | 0.00 | 0.91 | 0.17 |
| Mean ratio between tree height and DBH (m/cm) | HD | 256[a] | 0.75 | 0.54 | 0.91 | 0.06 |
| Quadratic mean diameter (cm) | $D_g$ | 256[a] | 27.8 | 16.9 | 40.0 | 4.29 |
| Crown height (m) | CH | 256[a] | 7.48 | 3.83 | 13.3 | 1.51 |
| Proportion of spruce volume relative to total volume (%) | $P_s$ | 29[b] | 79.9 | 0 | 100 | 27.3 |
| Proportion of pine volume relative to total volume (%) | $P_p$ | 29[b] | 18.1 | 0 | 100 | 27.4 |
| Proportion of deciduous tree volume relative to total volume (%) | $P_d$ | $29_b$ | 2.0 | 0 | 12 | 2.9 |

[a] Number of grid cells
[b] In the table the variable is calculated on stand level as total species-specific volume for an entire stand relative to the total stand volume. When used in models, grid cell values were used.

**Data Preparation**

Performing the analyses of this study on a stand level is problematic since each stand is quite large. Averaging our response variables over such large areas will to a great extent dampen the correlation to the explanatory laser variables because the range of variation is narrowed. This was a problem in this study, since we had no exact tree positions, which meant that our analyses could not be performed with a high spatial resolution. However, we did have the position of the harvester for every tree, which enabled us to split the data into sub-stand measurement units. We therefore divided the study area into square grid cells of 50x50 m (0.25 ha), and we used the harvester position to determine if a tree was inside or outside a certain grid cell. Some of the grid cells partly overlapped the stand boundaries, and for this reason, the grid cells have different sizes. If the grid cell was less than 50 $m^2$, it was excluded because the positioning error would be overwhelming.

The height distributions created from the first echoes were used to calculate percentiles for 5, 10, 20, …, 80, 90, 95, and 100% of the heights ($H_5$, $H_{10}$, $H_{20,\ldots,}$ $H_{80}$, $H_{90}$, $H_{95}$, $H_{100}$) and cumulative proportional canopy densities ($D_5$, $D_{10}$, $D_{20,\ldots,}$ $D_{80}$, $D_{90}$, $D_{95}$) for each grid cell (see Næsset 2002). The height distributions contained only those laser points which were classified as above-ground echoes, using a threshold value of 2 m (Nilsson 1996). This threshold was used to exclude echoes from stones, bushes, and other objects close to the ground that would introduce noise from a modeling point of view. The $H_5$ variable, for example, denotes the height above ground at which the accumulation of laser echoes in the vegetation is 5%. Thus, if the value of $H_5$ is 3 m, then 5% of the laser pulses that hit the grid cell hit below this specific height. These variables are referred to as height variables. Furthermore, $D_5$ denotes the proportion of laser echoes accumulating at 5% of maximum height relative to the total number of echoes. More specifically, for each grid cell, the height range of the laser echoes was divided into percentiles (fractions of equal length). Then the number of echoes below selected percentiles were counted and divided by the total number of echoes. Thus, the $D_5$ variable is the number of echoes below 5% of the maximum laser height relative to the total number of echoes. These variables are also called density variables or density metrics. Other variables calculated were the proportion of ground echoes versus canopy echoes (VEG) using a threshold value of 2 m, and the average height ($H_{mean}$), the standard deviation ($H_{std}$), and coefficient of the variation ($H_{coeffvar}$) of echoes >2 m above ground level.

**Data Analyses**

Using all of the laser variables as potential explanatory variables, we modeled all of our proxies for timber quality: $R_{D6}$, $P_{SL}$, $V_{SL}$, HD, $D_g$, and CH. An initial selection of the most powerful explanatory variables was first carried out using a stepwise regression procedure evaluating each included variable by its statistical significance. The stepwise analysis was carried out using PROC REG of SAS (SAS 2010). With the selected variables from the stepwise procedure as a basis, each response variable was then modeled using a mixed modeling procedure since the data have a hierarchical structure where inter-correlated grid cells come from 29 stands. For this we used the nlme (nonlinear mixed-effects models) package of the R software (Pinhero et al. 2007). The model had the following form:

$$Y_{ip} = (\beta_0 + \alpha_{0p}) + \sum_{r=1}^{k} \beta_r X_{rip} + e_i + e_p$$

where $Y_{ip}$ is the observed response value of grid cell $i$ in stand $p$. $X_{rip}$ represents observed value of explanatory variable $r$ at grid cell $i$ in stand $p$. $\beta_0$ (intercept) and $\beta_r$ are parameters for fixed effects to be estimated. Correspondingly, the $\alpha_0$ represents random effects for the intercept in stand $p$. $e_i$ is the unexplained random error and $e_p$ is the error explained by stand. The final models were estimated by manually entering and removing explanatory variables while evaluating model fit and variable significance. Since the laser variables are correlated, not only the selected variables after the stepwise procedure were used as candidate variables, but also the adjacent variables. For example, if the 80th percentile ($H_{90}$) was selected by the stepwise procedure, also the 90th and the 70th percentile were candidates in the mixed modeling. In addition to the explanatory laser variables, the proportions of pine-, spruce-, and deciduous volumes ($P_p$, $P_s$, and $P_d$) in each grid cell were used. This was because the relationships between laser variables and properties of the forest are dependent on tree species. In the results section we report the selected variables, model fit, RMSE, and error proportions due to stand effect ($error_{stand}$) and unexplained residual error ($error_{random}$). For some of the dependent variables we also report more than one model to illustrate the model sensitivity to certain explanatory variables. Model fit was assessed by the Pseudo-$R^2$. The Pseudo-$R^2$ was computed as 1 minus the ratio between the sum of residual sum of squares (SSR) and the corrected total sum of squares (CSST), i.e.

$$Pseudo-R^2 = 1 - \left(\frac{SSR}{CSST}\right)$$

Furthermore, RMSE was calculated on the basis of the differences between average predicted values and average observed values over grid cells within stands:

$$RMSE_j = \sqrt{\frac{1}{n}\sum_{p=1}^{n}(\bar{\hat{y}}_{jp} - \bar{y}_{jp})^2}$$

where $\bar{\hat{y}}_{jp}$ is the mean of the predicted quality variable $j$ in stand $p$, and $\bar{y}_{jp}$ is the corresponding mean observed value.

The different models are reported in Table 2. The sign of the relationship between the explanatory variables and the different dependent variables are indicated with "p" for positive and "n" for negative.

We also modeled each response using an ordinary least square model. This analysis was carried out to be able to isolate the stand effect of the model fit estimate. If the ordinary least square $R^2$ is much less compared to the Pseudo-$R^2$ from the mixed modeling, it indicates that the unexplained variance attributed to the stand is large. These models are reported in Table 3, and corresponding to Table 2 the relationships are indicated with "p" and "n" for positive and negative relationship, respectively.

**Table 2.** Explanatory variables, RMSE, model fit (Pseudo-$R^2$), and percentages of residual errors caused by stand effects ($error_{stand}$) and randomness ($error_{random}$) for mixed models of different measures of timber quality (Dependent variable). See Table 1 for definitions of Dependent Variables, $P_d$ and $P_p$. $H_{60}$, $D_5$, etc., are height distribution percentiles and cumulative proportional canopy density percentiles.

| Dependent variable | Explanatory variables (relationship)[a,b] | RMSE | Pseudo-$R^2$ | $error_{stand}$ (%) | $error_{random}$ (%) |
|---|---|---|---|---|---|
| $R_{D6}$(cm/cm) | $H_{60}$, $D_5$, $P_d$ (p, p, n) | 0.01 | 0.85 | 63 | 37 |
| $V_{SL}$($m^3ha^{-1}$) | $H_{90}$, $D_5$, $P_p$ (p, n, p) | 41.3 | 0.50 | 61 | 39 |
| $P_{SL}$ | $H_{50}$, $D_{40}$, $P_d$, $P_p$ (p, p, n, p) | 0.13 | 0.78 | 77 | 23 |
| HD (m/cm) | $D_{30}$, $P_p$ (n, n) | 0.03 | 0.57 | 60 | 40 |
| | $H_{90}$, $D_{10}$ (p, n) | 0.03 | 0.54 | 61 | 39 |
| $D_g$ (cm) | $H_{95}$, $D_{70}$, $D_{20}$, $P_p$ (p, n, p, p) | 1.65 | 0.74 | 56 | 44 |
| | $H_{95}$, $D_{70}$, $D_{20}$ (p, n, p) | 1.93 | 0.70 | 59 | 41 |
| | $H_{95}$, $D_{20}$ (p, p) | 2.23 | 0.65 | 59 | 41 |
| CH (m) | $H_{95}$, $D_{10}$, $P_p$ (p, n, p) | 0.54 | 0.58 | 55 | 45 |
| | $H_{95}$, $D_{10}$ (p, p) | 1.14 | 0.48 | 63 | 37 |

[a] Letters in parentheses, "p" or "n," indicate if the relationships between the explanatory variables and the dependent variables were positive or negative, respectively.
[b] Selection criterion for models where Pseudo-$R^2$ and RMSE.

## Results

Table 2 shows the results from modeling of the six variables describing timber quality by means of laser derived variables and tree species proportions. A general observation for all of our responses was that a substantial part of the modeling errors was an effect of stand as the $error_{stand}$ ranged between 55% and 77%.

For $R_{D6}$ the model fit was quite good. The Pseudo-$R^2$ was 0.85 and the RMSE was 0.01 which is just over 1% of the observed mean value of 0.75 (Table 1). The selected explanatory variables were the 60th height percentile ($H_{60}$), the proportion of laser echoes accumulating at 5% of maximum height relative to the total number of echoes ($D_5$), and the proportion of deciduous species in the grid cell.

The model fit expressed by the Pseudo-$R^2$ for $V_{SL}$ was 0.50 and the RMSE was 41.3 $m^3 ha^{-1}$. This translates into 24% of the observed mean value. The selected explanatory variables were a height percentile near maximum height ($H_{90}$), $D_5$, and the proportion of pine.

In addition to the actual saw log volume, we also modeled the proportion of saw timber relative to total volume ($P_{SL}$). Compared to the actual saw log volume, the model fit for $P_{SL}$ was better with a Pseudo-$R^2$ of 0.78 and a RMSE of 0.13, corresponding to 20% of the observed mean value. As opposed to the $V_{SL}$, the explanatory variables did not come from the extremes of the height or density metrics.

The mean ratio between tree height and diameter at breast height (HD) for each grid cell was also modeled. The results showed that a model including $D_{30}$ and the proportion of pine maximized the Pseudo-$R^2$ and minimized RMSE. However, an almost equally good fit was achieved with $H_{90}$ and $D_{10}$. When these two variables were selected the proportion of pine was not statistically significant (p>0.05).

**Table 3.** Explanatory variables, RMSE, and model fit ($R^2$) for Ordinary Least Squares models of different measures of timber quality (Dependent Variable). See Table 1 for definitions of Dependent Variables. $H_{60}$, $D_5$, etc., are height distribution percentiles and cumulative proportional canopy density percentiles.

| Dependent variable | Explanatory variables (relationship)[a] | RMSE | $R^2$ |
|---|---|---|---|
| $R_{D6}$(cm/cm) | $H_{60}$, $D_5$, $P_d$ (p, p, n) | 0.01 | 0.77 |
| $V_{SL}$($m^3ha^{-1}$) | $H_{90}$, $D_5$, $P_p$ (p, n, p) | 41.8 | 0.28 |
| $P_{SL}$ | $H_{50}$, $D_{40}$, $P_d$, $P_p$ (p, p, n, p) | 0.12 | 0.30 |
| HD (m/cm) | $D_{30}$, $P_p$ (n, n) | 0.03 | 0.37 |
| $D_g$ (cm) | $H_{95}$, $D_{70}$, $D_{20}$, $P_p$ (p, n, p, p) | 1.63 | 0.66 |
| CH (m) | $H_{95}$, $D_{10}$, $P_p$ (p, n, p) | 0.54 | 0.49 |

[a] Letters in parentheses, "p" or "n," indicate if the relationships between the explanatory variables and the dependent variables were positive or negative, respectively.

The model of $D_g$, the quadratic mean diameter, showed a pseudo-$R^2$ of 0.74 and a corresponding RMSE of 1.65 cm, which is 6% of the observed value. Also, this response variable was best explained by variables both from the height percentiles and the density variables. However, it was necessary to include both $D_{70}$ and $D_{20}$ in order to get a statistically significant intercept when the proportion of pine was included in the model. Without the proportion pine the value of Pseudo-$R^2$ declined, and all parameters were significant.

The last response variable that we modeled in this study was mean basal area weighted crown height (CH) of each grid cell. Similar to the models for most of the other response variables, one height percentile variable and one density variable was included in the model. The proportion of pine improved the model fit from Pseudo-$R^2$ of 0.48 to 0.58 and lowered the RMSE from 1.14 m to 0.54 m.

Table 3 shows the results from modeling the same response values with ordinary least squares (OLS) models without mixed effects. The table shows that the RMSE values are close to identical as for the mixed models. The $R^2$ values that do not account for the stand effect are generally lower compared to those of Table 2.

## Discussion

The geo-referencing of trees were based on the assumption that tree and harvester had the same position when the tree was cut. Since the harvesters usually work in a certain direction and the maximum reach of the type of harvester used here is 10 m, we assumed that the positioning accuracy was at an acceptable level. A harvester will seldom operate at full reach, and we assumed that an average distance from the harvester to a tree was 5 m. Following this reasoning, it is likely that the positioning errors cause that 10% of the trees are assigned to the wrong grid cell. Adding also the error from the positioning of the harvester, we end up with a total error between 20 and 25%. However, the effect of this positioning error will be dampened because of the spatial correlations that exist in a forest stand. For example, Gobakken and Næsset (2009) found that estimates of mean height, basal area, and volume from circular field plots could be quite accurately retrieved from low-density laser data even if the positioning error of the field plot was equal to the plot radius. Thus, in this study we were confident that analyses at the grid cell level would be relevant, also bearing in mind the advantage of better spatial resolution compared to stand level analyses.

Table 2 shows that the stand effect of the modeling errors was large for most of the variables expressing timber quality, and if Table 2 and Table 3 are compared, it is also evident that the model fits were quite different if the stand effect was accounted for or not. This means that for some of the quality variables it was difficult to find models based on laser variables that were general across stands. Nevertheless, Table 2 shows that the model fits were quite good at the stand level. The proportion of saw logs ($P_{SL}$) and the saw log volume ($V_{SL}$) for example, which can be regarded as the most direct measures of quality in this study, seemed to have a quite variable relationship to the laser variables between stands. Also, the mean ratio between tree height and diameter at breast height (HD) seemed to have the same variable relationship to the laser variables. To some extent it can be difficult to relate relative quality variables such as $P_{SL}$ and HD to absolute laser measurements, and for such variables some stand effects must be expected, even though we assumed that this would be a minor issue in the present study since we were dealing with mature forest only. However, the model for $V_{SL}$ which is an absolute measure also had a relatively low fit and large stand effect, so the fact that some of the variables are relative is not the explanation for the low model fit. This was also commented on by Peuhkurinen et al. (2008). They suggested that more auxiliary information that explains local (stand) variability of timber quality is needed in addition to laser metrics. In the present study we allowed the inclusion of proportions of tree species as such information, but there still seems to be a need for other types of auxiliary information in order to develop general relationships that could work well across stands.

The saw log volume estimates of this study can be compared to those of Korhonen et al. (2008). In their study the corresponding absolute and relative RMSE values for spruce dominated stands were 32.0 $m^3 ha^{-1}$ and 20.2%, respectively, indicating a slightly better fit. However, Korhonen et al. (2008) did not model harvester-based values, but used the harvester data for validation purposes. Values used in the modeling were based on measurements of accurately located sample plots where saw log volumes were calculated using taper curves and model based saw log reduction factors which take into account decay and other defects (Mehtätalo 2002).

The grid cells used in this study were quite large (0.25 ha) compared to what is common practice [0.02-0.04 ha, e.g. Næsset (2004)] using the area-based method. The large size of the grid cells was necessary because we only had rough tree coordinates, which meant that we had to increase grid cell size in order to decrease the impact of the geo-referencing error between field and laser data. At the same time we needed to keep the grid cell sizes small enough to ensure that the relationship between the field and laser variables wasn't too much dampened. Since the laser variables express the three-dimensional distribution of the laser pulses over the entire area of each grid cell, the correlation will be dampened when the area increases because the relative effect of change in the field data on the laser variables decreases with increasing area. However, if we had not been able to balance these two concerns of geo-referencing errors and grid cell size, at least some of our quality measures would suffer from low correlation with the laser variables. However, since the study showed that the relationship between $D_g$ and laser data was similar to other studies (e.g., Næsset et al. 2005, Næsset 2007) using normally sized ground plots and good geo-referencing, it seems that there is a good balance between these concerns. However, different variables could have different sensitivity both to geo-referencing errors and plot size (Gobakken and Næsset 2009).

The proportional stand effect for mean crown height is the lowest of all considered quality measures, but still rather high. The $R^2$ value of the model is not very high, but it should be mentioned though that the absolute RMSE value of 0.54 m is lower than in earlier studies where corresponding errors for mean crown height have usually been over 1 m (Næsset and Økland 2002, Dean et al. 2009, Maltamo et al. 2010).

While the variation of some of the quality variables seems to be hard to explain, there are other variables where the relation to the laser variables seems to be more general. Even if it is a relative measure, the $R_{D6}$ seemed to have a strong general relationship to the laser variables, although the low RMSE value must be seen in relation to the relatively low standard deviation (SD) in the observed values of $R_{D6}$ (Table 1). A comparison of the model fits of Tables 2 and 3 shows that the difference is less than for the other variables. The $R_{D6}$ represents the mean taper or the mean decrease in diameter from breast height to six meters height of each grid cell. On the other hand, the other stem-form variable HD did not have as strong a relationship to the laser variables. However, the model may be affected by the fact that tree height was not exactly known in the harvester data as the length of the top was predicted. It is likely that the HD model would be improved by having more accurate observed tree heights.

If laser data are going to be used for prediction of biophysical properties of forest stands, there has to be some calibration of the relationship between laser data and ground values. For some of the variables representing timber quality it can be difficult to obtain these ground truth values. Taper, for example, is difficult to measure while a tree is still standing. However, recent technical development of terrestrial laser scanners may make it possible to collect accurate detailed information about each stem on a field plot (e.g. Strahler et al. 2008). A terrestrial laser scanner is based on the same technology as the airborne, but collects data fixed to a tripod at a specific point on the ground. These scanners collect data in a near spherical space enabling detection of stems and branches in the scanner's circumference. Since trees near the scanner can shade other trees from being detected by the scanner, the instrument has to collect data from, say, three different points to fully cover a normal field plot area.

## Conclusions

This study has examined the relationship between laser measurements and some quality variables. We found that many variables can be hard to predict accurately using airborne laser scanner data. The reason seems to be that the relationships are not general, but dependent on local variation that is not captured by the laser data. Such local factors could, for example, reflect previous silvicultural treatments of the stands. However, some of the variables seem to be quite easy to model. It must be remembered though, that interior quality variables are out of the scope of this kind of approach. The accuracy levels obtained here tell about the general applicability of our laser-based approach. While detailed quality descriptions do not seem to be possible, at least the relationships found in this study seem to be strong enough to be able to prioritize between stands before harvest according to where it is most likely to find a certain log quality distribution.

Future research dealing with predictions of timber quality from airborne laser data should focus on finding auxiliary information or using the laser data in new ways that can make quality models more general across stands. Since most of these variables are somehow related to stand density and canopy structure, there should still be great potential to exploit laser scanner data for these purposes.

## References

Assmann, E. 1970. The principle of forest yield study. Oxford, England: Pergamon.

Axelsson, P. 2000. DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*. 33:110-117.

Bollandsås, O.M. and Næsset, E. 2007. Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. *Scandinavian Journal of Forest Research*. 22(1):33-47.

Dean, T.J., Cao, Q.V., Roberts S.D. and Evans, D.L. 2009. Measuring heights to crown base and crown median with LiDAR in a mature, even-aged loblolly pine stand. *Forest Ecology and Management*. 257(1):126-133.

Fonweban, J., Gardiner, B., Suarez, J., Lundqvist, S.O. and Keane, E. 2008. Linking growth and quality models with terrestrial and airborne scanning to predict the British forest resource. In: Connection between Forest Resources and Wood Quality: Modelling Approaches and Simulation Software. Peltola, H. Ed. International Union of Forest Research Organizations WP 5.01.04, June 8-14, 2008, Koli, Finland.

Fulton, M.R. 1999. Patterns in height-diameter relationships for selected tree species and sites in eastern Texas. *Canadian Journal of Forest Research*. 29(9):1445-1449.

Gobakken, T. and Næsset, E. 2004. Estimation of diameter and basal area distributions in coniferous forest by means of airborne laser scanner data. *Scandinavian Journal of Forest Research*. 19(6):529-542.

Gobakken, T. and Næsset, E. 2005. Weibull and percentile models for LIDAR-based estimation of basal area distribution. *Scandinavian Journal of Forest Research*. 20 (6):490-502.

Gobakken, T. and Næsset, E. 2009. Assessing effects of positioning errors and sample plot size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*. 39(5):1036-1052.

Hein, S., Weiskittel, A.R. and Kohnle, U. 2008. Effect of wide spacing on tree growth, branch and sapwood properties of young Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco] in southwestern Germany. *European Journal of Forest Research*. 127(6):481-493.

Hodgson, M.E. and Bresnahan, P. 2004. Accuracy of airborne LIDAR-derived elevation: Empirical assessment and error budget. *Photogrammetric Engineering and Remote Sensing*. 70(3):331-339.

Hyyppä, J., Kelle, O., Lehikoinen, M. and Inkinen, M. 2001. A segmentation-based method to retrieve stem volume estimates from 3-D tree height models produced by laser scanners. *IEEE Transactions on Geoscience and Remote Sensing*. 39(5):969-975.

Ikonen, V.P., Kellomäki, S. and Peltola, H. 2009. Sawn timber properties of Scots pine as affected by initial stand density, thinning and pruning: a simulation based approach. *Silva Fennica*. 43(3):411-431.

Ikonen, V.P., Peltola, H., Wilhelmsson, L., Kilpeläinen, A., Väisänen, H., Nuutinen, T. and Kellomäki, S. 2008. Modelling the distribution of wood properties along the stems of Scots pine (*Pinus sylvestris L.*) and Norway spruce (*Picea abies (L.) Karst.*) as affected by silvicultural management. *Forest Ecology and Management.* 256(6):1356-1371.

Korhonen, L., Peuhkurinen, J., Malinen, J., Suvanto, A., Maltamo, M., Packalén, P. and Kangas, J. 2008. The use of airborne laser scanning to estimate saw log volumes. *Forestry.* 81(4):499-510.

Kraus, K. and Pfeifer, N. 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*. 53(4): 193-203.

Loetsch, F., Zöhrer, F. and Haller, K.E. 1973. Forest inventory, Volume II. BLV, München, Bern, Wien.

Maguire, D.A., Kershaw, J.A. and Hann, D.W. 1991. Predicting the effect of silvicultural regime on branch size and crown wood core in Douglas-fir. *Forest Science.* 37(5):1409-1428.

Malinen, J. 2003. Prediction of characteristics of marked stand and metrics for similarity of log distribution for wood-procurement management. D.Sc. (Agr. and For.) thesis. University of Joensuu. Faculty of Forestry. Research Notes 151.

Malinen, J., Maltamo, M. and Verkasalo, E. 2003. Predicting the internal quality and value of Norway spruce trees by using two non-parametric nearest neighbor methods. *Forest Products Journal.* 53(4):85-94.

Malinen, J., Maltamo, M. and Verkasalo, E. 2005. Wood Properties of Mature Norway spruce on Drained Peatlands Compared to Mineral Forest Lands in Southern Finland. *Baltic Forestry*. 11(1):21-37.

Maltamo, M., Bollandsås, O.M., Vauhkonen, J., Breidenbach, J., Gobakken, T. and Næsset, E. 2010. Comparing different methods for prediction of mean crown height in Norway spruce stands using airborne laser scanner data. *Forestry.* 83(3):257-268.

Maltamo, M., Næsset, E., Bollandsås, O.M., Gobakken, T. and Packalén, P. 2009a. Non-parametric prediction of diameter distributions using airborne laser scanner data. *Scandinavian Journal of Forest Research*. 24(6):541-553.

Maltamo, M., Hyyppä, J. and Malinen, J. 2006. A comparative study of the use of laser scanner data and field measurements in the prediction of crown height in boreal forests. *Scandinavian Journal of Forest Research*. 21(3):231-238.

Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J. and Pitkänen, J. 2005. Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management*. 216(1/3):41-50.

Maltamo, M., Peuhkurinen, J., Malinen, J., Vauhkonen, J., Packalén, P. and Tokola, T. 2009b. Predicting tree attributes and quality characteristics of Scots pine using airborne laser scanning data. *Silva Fennica*. 43(3):507-521.

Mehtätalo, L. 2002. Valtakunnalliset puukohtaiset tukkivähennysmallit männylle, kuuselle, koivulle ja haavalle. (National tree level saw log reduction models for pine, spruce, birch and aspen, in Finnish) *Metsätieteen aikakauskirja.* 4/2002:575-591.

Moberg, L., Wallerman, J., Holmgren, J. and Barth, A. 2008. High-precision inventory methods to predict wood properties for operational, pre-harvest planning. Connection between Forest Resources and Wood Quality: Modelling Approaches and Simulation Software. Peltola, H. Ed. International Union of Forest Research Organizations WP 5.01.04, June 8-14, 2008, Koli, Finland.

Morsdorf, F., Meier, E., Kotz, B., Itten, K.I., Dobbertin, M. and Allgower, B. 2004. LIDAR-based geometric reconstruction of boreal type forest stands at single-tree level for forest and wildland fire management. *Remote Sensing of Environment*. 92(3):353-362.

Næsset, E. 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*. 80(1):88-99.

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

Næsset, E. 2007. Airborne laser scanning as a method in operational forest inventory: Status of accuracy assessments accomplished in Scandinavia. *Scandinavian Journal of Forest Research*. 22(5):433-442.

Næsset, E., Bollandsås, O.M. and Gobakken, T. 2005. Comparing regression methods in estimation of biophysical properties of forest stands from two different inventories using laser scanner data. *Remote Sensing of Environment*. 94(4):541-553.

Næsset, E. and Økland, T. 2002. Estimating tree height and tree crown properties using airborne scanning laser in a boreal nature reserve. *Remote Sensing of Environment*. 79(1):105-115.

Nilsson, M. 1996. Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sensing of Environment*. 56(1):1-7.

Optech. 2008. Gemini ALTM 167. Available from http://www.optech.ca/pdf/Gemini167.pdf [accessed June 24, 2009].

Øvrum, A., Vestøl, G.I. and Høibø, O.A. 2008. Modeling the longitudinal variation of sawn timber grades in Norway spruce (*Picea abies (L.) Karst.). Holz als Roh- und Werkstoff*. 66(3):219-227.

Packalén, P. and Maltamo, M. 2008. The estimation of species-specific diameter distributions using airborne laser scanning and aerial photographs. *Canadian Journal of Forest Research*. 38(7):1750-1760.

Persson, Å., Holmgren J. and Söderman, U. 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*. 68(9):925-932.
Peuhkurinen, J., Maltamo, M. and Malinen, J. 2008. Estimating species-specific height-diameter distributions and saw log recoveries from ALS data and aerial photographs: A distribution-based approach. *Silva Fennica*. 42(4):625-641.
Peuhkurinen, J., Maltamo, M., Malinen, J., Pitkänen, J. and Packalén, P. 2007. Preharvest measurement of marked stands using airborne laser scanning. *Forest Science.* 53(6):653-661.
Pinhero, J., Bates, D., DebRoy, S. and Sarkar, D. 2007. nlme. Linear and non-linear mixed effects models. R package version 3, 1-85.
Popescu, S. and Zhao, K. 2008. A voxel-based lidar method for estimating crown base height for deciduous and pine trees. *Remote Sensing of Environment*. 112(3):767 -781.
Reutebuch, S.E., McGaughey, R.J., Andersen, H.E. and Carson, W.W. 2003. Accuracy of a high-resolution LIDAR terrain model under a conifer forest canopy. *Canadian Journal of Remote Sensing*. 29(5):527-535.
SAS 2010. SAS product documentation. Available from http://support.sas.com/documentation [accessed January 1, 2010].
Schmidt, M. 2004. Simulation on branchiness and quality for roundwood of Douglas-fir. *Allgemeine Forst und Jagdzeitung*. 175(3):49-60.
Sharma, M. and Zhang, S.Y. 2004. Height-diameter models using stand characteristics for *Pinus banksiana* and *Picea mariana. Scandinavian Journal of Forest Research*. 19(5):442-452.
Solberg, S., Næsset, E. and Bollandsås, O.M. 2006. Single tree segmentation using airborne laser scanner data in a structurally heterogeneous spruce forest. *Photogrammetric Engineering and Remote Sensing*. 72(12):1369-1378.
Strahler, A.H., Jupp, D.L.B., Woodcock, C.E., Schaaf, C.B., Yao, T., Zhao, F., Yang, X., Lovell, J., Culvenor, J., Newnham, G., Ni-Miester, W. and Boykin-Morris, W. 2008. Retrieval of forest structural parameters using a ground-based lidar instrument (Echidna®). *Canadian Journal of Remote Sensing*. 34(Suppl. 2):426-449.
Vauhkonen, J. 2008. Estimating crown base height for Scots pine by means of the 3D geometry of airborne laser scanning data. In: Proceedings of SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory. Hill, R., J. Rosette and J. Suárez. Eds. September 17–19, 2008. Heriot-Watt University, Edinburgh, UK, p. 616-624.
Wehr, A. and Lohr, U. 1999. Airborne laser scanning — An introduction and overview. ISPRS *Journal of Photogrammetry & Remote Sensing.* 54(2/3):8-82.