# ALS based estimation of plot volume and site index in a eucalyptus plantation with a nonlinear mixed effect model that accounts for the clone effect

PETTERI PACKALÉN *†, LAURI MEHTÄTALO† and MATTI MALTAMO†

† Faculty of Science and Forestry, University of Eastern Finland, P.O. Box 111, FI-80101 Joensuu, Finland.

Most ALS studies have been carried out in semi-natural forests but some research has also been carried out in plantations. Results indicate that methods similar to those which are used in semi-natural forest are also usable in plantation forestry. The aims of this study are to investigate (1) how accurately the plot volume (V) may be estimated by ALS data in eucalyptus plantations, and (2) how to estimate the site index (SI) directly by combining ALS data and stand age. The study was conducted in a pulpwood plantation growing *Eucalyptus urograndis* in Bahia state, Brazil. The data included field measurements from 55 stands and 28 different clones; all the trees in a stand belong to the same clone. The V and SI were estimated by means of nonlinear mixed effect modelling in order to take into account the stand within clone hierarchy of the data. The obtained accuracies are quite good if compared to those obtained in semi-natural forests. The RMSE was 8.2% for V and 2.7% for SI when clone effect was used in prediction. It is worth noting that in eucalyptus plantations the clone is always known and therefore its use does not restrict the applicability of this method. If none of the sample plots is located within a particular clone, only the fixed part of the model may be used. In the estimation of SI also the stand age was taken from the stand register data. This does not restrict the applicability of the method either, since the planting date is always known and trees are of the same age. Therefore the SI may be predicted in a wall-to-wall manner across the whole plantation. Precision forestry applied in plantations differs many ways from the forestry practiced in semi-natural environment. ALS based forest inventory has a lot of potential in pulpwood plantations when the unique features of plantation forestry are taken into account.

## 1. Introduction

Although most ALS (Airborne Laser Scanning) studies have been done in semi-natural forests, some research has also been carried out in plantations (e.g. McCombs et al. 2003, Wack et al. 2003, Roberts et al. 2005, Donoghue et al. 2007, Rombouts et al. 2008, Hopkinson et al. 2008, Tesfamichael 2009, Zonete et al. 2010). The focus in these studies has been the accuracy of ALS based stand attribute estimates including leaf area and growth. Results indicate that methods similar to those used in semi-natural forests also seem to be appropriate for plantation forestry. So far, most studies have dealt with conifer plantations, an exception being the articles by Wack et al. (2003), Zonete et al. (2010), and the PhD thesis by Tesfamichael (2009, cf. journal publications) in which eucalyptus was considered. Both area based method (Næsset 1997) and individual tree detection (Hyyppä and Inkinen 1999) have been used studies in plantations.

Plantation forestry differs in many ways from the forestry practiced in semi-natural environments. There are several kinds of forest plantations. The aim may be to

*Corresponding author. Email address: petteri.packalen@uef.fi

provide wildlife-habitat, biological diversity, or other services in addition to wood production: in such cases multi-species forest plantations are favoured (FAO 2002). The type of plantation considered here, however, is a mono-species eucalyptus plantation which is targeted to the pulpwood production only. In eucalyptus plantations, trees are planted in rows using a fixed stand density. There is no natural regeneration, thus trees within a stand are of the same age and single storied. Typically, all trees in a stand belong to the same clone; the aim is to select the most productive clone to a particular stand. Productivity is quantified by the site index (SI), which is usually based on the development of a dominant height over the rotation period.

Plantation forestry also occupies a special situation regarding remote sensing based forest assessment. Although the stand level information about age and tree species is often available also for semi-natural forests, this information is especially accurate in even age plantations where trees are of the same age and also the clone is known in certain cases. This makes it more feasible to combine existing stand register data and remote sensing data and to carry out estimation in a wall-to-wall manner. Accurate age at tree level also enables the estimation of attributes which are normally not feasibly estimated by remote sensing, such as the variation of site index (SI) within a stand.

The aims of this study are to investigate (1) how accurately the plot volume (V) may be estimated by ALS data in Eucalyptus plantation, and (2) how to estimate SI directly by combining ALS data and age taken from the stand register. In order to utilize efficiently all available information, the estimation was carried out with a nonlinear mixed effect model that accounts for the clone effect.

## 2. Material

### *2.1 Study area and field data*

The study area is a pulpwood plantation growing eucalyptus in Bahia state, Brazil. The plantation is owned by Veracel and the pulp mill itself is sited within the test area. The cultivated *Eucalyptus* species is *E. urograndis*, which is a hybrid between *E. grandis* and *E. urophylla*. This eucalyptus hybrid is highly productive and is one of the main species currently used for pulp production in Brazil (Siverio et al. 2007).

A network of 195 circular sample plots with a radius of 13 meters was established and measured in August–September 2008. Sample plots were placed in 55 forest stands with three or four plots per stand. Satellite positioning was used to determine the position of each plot center using a real-time differential correction signal from the OmniSTAR satellite (http://www.omnistar.com). The trees were growing in rows and the tree spacing was fixed, giving a density of 833 stems per hectare. All trees were recorded in the field for diameter at breast height ($d$) and quality, and every seventh tree on each plot was measured for height ($h$). Näslund's (1937) $h$-$d$ curve was fitted by stands and used to predict heights for trees without height measurement. Stem volumes were calculated as a function of $d$ and $h$ using a clone and age class specific model constructed in-house in Veracel.

Plot volume (V) was calculated by aggregating from tree to plot level, and dominant height (HD) was calculated as the mean height of the 100 thickest trees at breast height per hectare. Stand age (T) was taken from the plantation database in which it was registered with the precision of one month. There were 28 different clones in the 55 stands from which the field data were collected. All trees in a stand belonged to the same clone.

The site index (SI) was predicted using the following form of the Chapman–Richards Equation (e.g., Clutter et al. 1983):

$$\mathrm{SI} = \mathrm{HD}\left(\frac{1-e^{-\beta_1 t_{reference}}}{1-e^{-\beta_1 t_{current}}}\right)^{\beta_2}, \quad (1)$$

where $t_{reference}$ is the reference age of 7 years, HD is the current dominant height, $t_{current}$ is the current age, and $\beta_1$ (0.3341) and $\beta_2$ (1.1442) are known model parameters in Veracel.

### *2.2 Airborne laser scanning data*

ALS data were collected on August 16, 2008 using an Optech ALTM 3100 laser scanning system. The test site was measured from an altitude of approximately 1200 m above ground level using a field of view of 30 degrees. Pulse repetition frequency was set to 50 000 pulses per second, which resulted a nominal sampling density of about 1.5 measurements per square meter. The footprint was about 35 cm at ground level.

A digital terrain model (DTM) was generated from the ALS data. First, laser points were classified as ground and non-ground points using the method reported by Axelsson (2000). Then a raster DTM with a 1-meter pixel size was interpolated using ground points and an inverse distance weighting algorithm (Lloyd and Atkinson 2002). Finally, the raster DTM was subtracted from the ellipsoidal heights of laser points in order to scale the ALS data to the above-ground level (AGL).

## 3. Methods

### *3.1 Explanatory variables*

The laser scanner used captures a maximum of four range measurements for each submitted pulse. These echo categories are ‘first of many’, ‘last of many’, ‘only’ and ‘intermediate’. After preliminary tests it was decided that only the echo categories ‘first of many’, and ‘only’ would be used in this study, since the exclusion of ‘last of many’ and ‘intermediate’ echoes did not significantly decrease the accuracy of the SI and V estimates. This set contains all of the first – or surface – echoes since an ‘only’ echo may also be considered as a first echo.

Numerous height and density metrics were calculated from the combined set of ‘first of many’ and ‘only’ echoes. The principle of the area based method was used here. The first step was to calculate height distributions for each sample plot using the heights of the AGL data. All the laser hits were considered, also those lying on the ground. Height quantiles for 5, 10, 20,…, 80, 90, 95% (*h5,…, h95*) were computed, and the corresponding densities (*p5,…, p95*) were calculated for the respective quantiles. Height quantiles were calculated by summing the heights at AGL. For instance, the metric h50 is the height at which 50% of the cumulative height has accumulated and p50 is the number of laser hits below h50 divided by all the laser hits on the plot. In addition, the mean (*hmean*) and standard deviation (*hstd*) of heights at AGL were calculated by plots. These metrics form a set of candidate explanatory variables used for modeling SI and V.

### *3.2 Modeling of site index and plot volume*

The same form of the Chapman-Richards Equation which is used to predict the SI in this plantation is used as the starting point in the modeling of SI (see Eq. 1). However, instead of modeling the HD separately and inserting the resulting predictions into Eq. 1, the SI was modeled directly. This was done by replacing the HD of Eq. 1 with a linear dominant height model that uses ALS based explanatory variables, and fitting the equation as a nonlinear model into the modeling data. We also incorporated nested random effects into the model to account for the effects of the hierarchical data (sample plot within stand within clone) on the model residuals. To begin with, however, the form of the HD model was chosen by using linear regression. At this stage, both manual insertion and deletion of explanatory variables and a stepwise selection based on the Akaike information criterion were used.

Preliminary analyses for the model of V showed a linear relationship between ln(V) and the predictors. That is why a nonlinear model of the exponential form was used to model the total volume. This led to a model that has the observed linear relationship and is unbiased for V. This model was also fitted as a mixed effects model to account for data hierarchy.

The nonlinear mixed effect models were fitted by using the *nlme* routine (Lindstrom and Bates 1990, Pinheiro and Bates 2000) in the R environment (R Development Core Team 2009) by using the method of maximum likelihood.

### *3.3 Accuracy assessment*

Accuracy assessment was carried out at three different levels: for the fixed part of the mixed effect model, by using the predicted clone effect, and by using the predicted stand and clone effects. The accuracy of estimates was evaluated in terms of relative root mean squared error and bias at the plot level:

$$\text{RMSE-}\% = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}}}{y_{mean}} \times 100 \qquad (2)$$

$$\text{BIAS-}\% = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)}{n}}}{y_{mean}} \times 100 \qquad (3)$$

where $n$ is the number of plots, $y_i$ is the observed value for plot $i$, $\hat{y}_i$ is the predicted value for plot *I*, and $y_{mean}$ is the observed mean of the variable in question.

## 4. Results

### *4.1 Nonlinear mixed effect models*

The model for the SI is as follows:

$$SI_{cki} = \left(\frac{1}{\beta_3+\beta_4 h90_{cki}+\beta_5 h95_{cki}^{\;2}+\beta_6\sqrt{hmean_{cki}}}\right)\left(\frac{1-e^{-\beta_1 t_{reference}}}{1-e^{-\beta_1 t_{current,ck}}}\right)^{\beta_2} + b_c + b_{ck} + \varepsilon_{cki} \qquad (4)$$

where $SI_{cki}$ is the SI for clone $c$, stand $k$, and plot $i$; $b_c$ is a random clone effect; $b_{ck}$ is a random stand effect (stands are nested within clones); and $\varepsilon_{cki}$ is the residual for clone $c$, stand $k$, and plot $i$. $\beta_1$ and $\beta_2$ are the known parameters of the SI model (the same as in Eq. 1). The random effects and residuals are assumed to be normally distributed with mean zero and constant variance.

The model for the plot volume is as follows:

$$V_{cki} = e^{\left(\beta_7+\beta_8\log(\hbar 10)+\beta_9\log(\hbar 50)+\beta_{10}\sqrt{p90}+b_c+b_{ck}\right)} + \varepsilon_{cki} \tag{5}$$

where $V_{cki}$ is V for clone $c$, stand $k$, and plot $I$; $b_c$ is a random clone effect; $b_{ck}$ is a random stand effect (stands are nested within clones); and $\varepsilon_{cki}$ is the residual for clone $c$, stand $k$, and plot $i$. The random effects and residuals are assumed to be normally distributed with mean zero and constant variance. Preliminary analyses showed increasing residual variance with respect to the prediction, so the residual was assumed to be normally distributed with mean zero and variance var($\varepsilon_{cki}$)=$\sigma^2 h50^{2\delta}$ , where $\sigma$ and $\delta$ are the parameters of the variance function. The parameter estimates for both models are listed in Table 1.

Table 1. Parameter estimates for the fixed independent variables and estimated variances for the random effects at the clone, forest stand, and plot level.

| Equation 4 | | | Equation 5 | | |
|---|---|---|---|---|---|
| Coefficient | Estimate | SE | Coefficient | Estimate | SE |
| $\beta_3$ | 0.0977596 | 0.0028945 | $\beta_7$ | 11.415382 | 2.298940 |
| $\beta_4$ | -0.0034367 | 0.0001932 | $\beta_8$ | 0.142945 | 0.077417 |
| $\beta_5$ | 0.0000432 | 0.0000031 | $\beta_9$ | 1.460686 | 0.095138 |
| $\beta_6$ | -0.0008237 | 0.0003894 | $\beta_{10}$ | -1.133237 | 0.236214 |
| Random parameters | | | Random parameters | | |
| Var($b_c$) | $0.478231^2$ | | Var($b_c$) | $0.064235^2$ | |
| Var($b_{ck}$) | $0.726614^2$ | | Var($b_{ck}$) | $0.065542^2$ | |
| Var($\varepsilon_{cki}$) | $0.659612^2$ | | Var($\varepsilon_{cki}$) | $0.015319^2 h50^{4.236}$ | |

### *4.2 Accuracies*

The accuracy and bias of the estimates of SI and V are presented in Table 2. Level denotes the level of grouping used for obtaining the predictions. Level 'Fixed' means that only the fixed part of the model (Eq. 4 or 5) is used, level 'Clone' means that the predicted clone effect is used in prediction, and 'Stand' means that also the predicted random effect for the stand within clone is used in the prediction.

The estimates of V were already rather accurate without random effects (Table 2), however, the inclusion of random clone effects improved the accuracy from 11.82 to 8.22, and the further inclusion of random stand effects decreased the RMSE-% to 4.90. There was some bias in the estimates of V when only the fixed part of the model was used, without random components.

The accuracy of estimation of SI was 3.18% when only the fixed part of the model was used in the prediction. The inclusion of random effects improved the accuracy notably but not quite as much as in the case of V. The SI estimates were virtually unbiased at all levels.

Table 2. The RMSE and bias of the estimates of site index (SI) and volume (V) at different levels of grouping.

| | SI | | V | |
|---|---|---|---|---|
| Level | RMSE-% | BIAS-% | RMSE-% | BIAS-% |
| Fixed | 3.18 | 0.41 | 11.82 | 2.20 |
| Clone | 2.66 | 0.10 | 8.22 | 0.41 |
| Stand | 1.68 | 0.02 | 4.90 | -0.07 |

## 5. Discussion

Precision forestry applied to plantations differs in many ways from the forestry practiced in semi-natural environments. ALS based forest inventory and assessment has a lot of potential in pulpwood plantations when the unique features of plantation forestry are taken into account.

In this study, SI was modeled directly in one stage. Another option is a two-stage approach, where the HD is first modeled and then the resulting prediction is used in the known SI curve. Direct modeling approach is justified since the only use of the predicted HD is the prediction of SI. In addition, the model was fitted in a nonlinear form, where the SI was treated as the independent variable as such, without transformations. These two methodological choices led to unbiased prediction of SI, which could not have been guaranteed by using the two-stage method and/or a linearized form of the applied HD model. Also, the applied nonlinear model for plot volume led to unbiased prediction of volume. With the selected model form, the non-homogeneous variance was satisfactorily homogenized by a variance function of the power form.

Using the clone and stand effects in prediction improved the accuracy of the prediction considerably. This improvement resulted from using all the plot-specific observations of the response (V or SI) for the stand or clone under consideration in prediction. For clones that do not have measurements available, this level of accuracy is not possible, because the prediction can be done only at the ‘Fixed’ level. For clones with measurements from different stands, the prediction can be done at the ‘Clone’ level. Prediction at the ‘Stand’ level can be done only if sample plot measurements from that particular stand are available. The practical use of the model arises from the possibility to use the clone effect in prediction for stands with no measured sample plots. In those cases, the prediction could be close to the reported accuracy at the clone level.

The drawback of using predictors based on ALS data is that the models are local. The properties of point cloud produced by ALS is dependent on the scanning configuration, e.g., what is the pulse repetition frequency or flying altitude; also, the responses of different sensors differ (Naesset 2009). Therefore, such models as presented here must be made by plantations but in practical terms this is not a major restriction. If several sensors or different scanning configurations are used to collect ALS data from a plantation, the sensor effect may be taken into account as a random effect in a similar manner as were the clone and stand considered here. However, this definitely decreases the accuracy.

The clone level explained a considerable amount of the between-stand variation for both models, and including the clone effect in the prediction improves the accuracy considerably. This observation leads to suggesting the placing of sample plots over the whole inventory area so that some measurements are available for all the clones. Then the prediction could be carried on a grid to the whole plantation at the clone level. Final inventory results may be obtained by aggregation to stand level, or cell level results may be used as such. For instance, SI estimated in a wall-to-wall manner provides new possibities to carry out precision forestry.

**References**


CLUTTER, J.L., FORTSON, J.C., PIENAAR, L.V., BRISTER, G.H. and BAILEY, R.L., 1983, *Timber management: a quantitative approach*. (New York: John Wiley & Sons).

DONOGHUE, D.N.M., WATT, P.J., COX, N.J. and WILSON, J., 2007, Remote sensing of species mixtures in conifer plantations using Lidar height and intensity data. *Remote Sensing of Environment*, **110**, pp. 509–522.

FAO, 2002, *Forest plantation productivity. Report based on the work of W.J. Libby and C. Palmberg-Lerche.* In *Forest Plantation Thematic Papers, Working Paper 3.* Forest Resources Development Service, Forest Resources Division. FAO, Rome.

HOPKINSON, C., CHASMER, L. and HALL, R.J., 2008, The uncertainty in conifer plantation growth prediction from multi-temporal lidar datasets. *Remote Sensing of Environment*, **112**, pp. 1168–1180.

HYYPPÄ, J. and INKINEN, M., 1999, Detecting and estimating attributes for single trees using laser scanner. *The Photogrammetric Journal of Finland*, **16**, pp. 27–42.

LINDSTROM, M.J. and BATES, D.M., 1990, Nonlinear mixed effects models for repeated measures data. *Biometrics*, **46**, pp. 673–687.

MCCOMBS, J.W., ROBERTS, S.D. and EVANS, D.L., 2003, Influence of fusing lidar and multispectral imagery on remotely sensed estimates of stand density and mean tree height in a managed loblolly pine plantation. *Forest Science*, **49**, pp. 457–466.

NÆSSET, E., 1997, Estimating timber volume of forest stands using airborne laser scanner data. *Remote Sensing of Environment*, **51**, pp. 246–253.

NÆSSET, E. 2009, Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, **113**, pp. 148–159.

PINHEIRO, J.C. and BATES, D.M. 2000, *Mixed-effects models in S and S-PLUS*, (New York: Springer).

R Development Core Team, 2009, *R: A language and environment for statistical computing*. (Vienna: R Foundation for Statistical Computing). URL http://www.R-project.org.

ROBERTS, S.D., DEAN, T.J., EVANS, D.L., MCCOMBS, J.W., HARRINGTON, R.L. and GLASS, P.A., 2005, Estimating individual tree leaf area in loblolly pine plantations using LIDAR-derived measurements of height and crown dimensions. *Forest Ecology and Management*, **213**, pp. 54–70.

ROMBOUTS, J., FERGUSON, I.S., and LEECH, J.W., 2008, Variability of LiDAR volume prediction models for productivity assessment of radiata pine plantations in South Australia. In *Proc. of SilviLaser 2008, 8th international conference on LiDAR applications in forest assessment and inventory*, R. Hill, J. Rosette, and J. Suárez (eds). Edinburgh, pp. 39–49.

SIVERIO, F.O., BARBOSA, L.C.A., MALTHA, C.R.A., SILVESTRE, A.J.D., PILO-VELOSO, D. and GOMIDE, J.L., 2007, Characterization of lipophilic wood extractives from clones of *Eucalyptus urograndis* cultivated in Brazil. *BioResources*, **2**, pp. 157–168.

TESFAMICHAEL, S.G., 2009, Assessment of structural attributes of even-aged Eucalyptus grandis forest plantations using small-footprint discrete return lidar data. PhD thesis, University of KwaZulu-Natal, South Africa.

WACK, R., SCHARDT, M., LOHR, U., BARRUCHO, L. and OLIVEIRA, T., 2003, Forest inventory for Eucalyptus plantations based on airborne laser scanner data. In *Proc. of the International Society for Photogrammetry and Remote Sensing Symposium, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **34**, 3/W13. The Netherlands, pp. 40–46.