Estimation of bivariate diameter and height distributions using ALS

Johannes Breidenbach[1], Christian Gläser[2] & Matthias Schmidt[3]

[1]Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg, Abteilung Biometrie und Informatik, Johannes.Breidenbach@forst.bwl.de
[2]Georg-August-Universität Göttingen, Institut für Statistik und Ökonometrie, Christian_Glaeser@gmx.de
[3] Nordwestdeutsche Forstliche Versuchsanstalt, Abteilung Waldwachstum, Matthias.Schmidt@nw-fva.de

Abstract

In this paper, we present a method for estimating a bivariate height and diameter distribution based on airborne laser scanner data (ALS). ALS are analyzed with the area-based method. To construct a bivariate likelihood function, we assume that the diameters are Weibull distributed and the heights are normal distributed where the expectation value is modelled with the Näslund function. Comparison of predicted and observed mean diameter distributions result in an RMSE of 2.78 cm and a bias of 0.48 cm. Comparisons of observed and predicted height distributions are still to come.

Keywords Weibull distribution, Näslund function, lidar, GLM

1. Introduction

High density (i.e., >1 return m^{-1}, Magnusson et al. 2007) Airborne laser scanning data (ALS) can be used to estimate attributes of single trees (e.g., Persson et al. 2002, Popescu et al. 2003). Low density ALS data are usually analyzed using area-based methods (Næsset 2002) that result in plot-level estimates of, for example, timber volume or basal area. The lower flying height required to produce higher density ALS data causes longer flight times per area. Therefore, costs obviously increase with the density of the raw data. This may be one reason for why area-based methods are already used in operational forest inventories (Næsset 2004). Additionally, some studies show, that the density of the laser data may be reduced in a wide range without a significant lost of information (Magnusson et al. 2007, Gobakken and Næsset 2007). Consequently, further cost reductions may be possible with future generation of laser scanners that allow larger flying heights.

While plot level estimates of, for example volume, is an important piece of information, for the actually wanted prediction of assortments, at least diameter, but also height distributions are needed. Magnussen & Boudewyn (1998) estimated height distributions using ALS. Methods for estimating diameter distributions were shown, for example by Gobakken & Næsset (2004). However, height and diameter distributions cannot be combined if they are estimated independently. While Schreuder & Hafley (1977) and Zucchini et al. (2001) proposed functions to fit bivariate distributions, only Mehtätalo et al. (2007) used a regression model to fit a bivariate height and diameter distribution conditional on ALS data. They first used the method of moments to fit regression models for the height distribution as a function of the diameter and the diameter distribution as a function of ALS data. They then applied the method of parameter recovery to obtain the parameters of the bivariate distribution.

In this study, we use the log-likelihood method to estimate the bivariate distribution of height and diameter. The diameters are assumed to follow the Weibull distribution given metrics from ALS data. Opposed to Mehtätalo et al. (2007), we need to deal with several tree species. Therefore, tree height is modelled using the Näslund function given diameter and metrics from ALS data. However, different tree species are not addressed separately.

2. Material and Methods

2.1 Study area

The tree species composition of the 50 km² managed forest that served as study site is dominated by Norway spruce (*Picea abies* L. Karst.) with a 70% proportion by area, beech (*Fagus sylvatica* L.) with 11% and silver fir (*Abies alba* Mill.) with 10%. More details on the forest structure are given in Table 1.

Table 1: Forest characteristics of the study site

	Minimum	Median	Mean	Maximum
Stem number [ha^{-1}]	22.1	397.8	497.3	2829
Stem volume [m^3 ha^{-1}]	7.2	412.7	413.2	1193
Basal area [m^2 ha^{-1}]	1.8	36.8	36.8	81.9
Basal area mean diameter [cm]	7.5	35	35.8	68.8
Mean height [m]	5.1	25	24.6	40.7

2.1.1 Plot establishment

In 2002, a permanent sample-plot inventory was carried out on a 100 m (easting) by 200 m (northing) grid. Trees with a diameter at breast height (dbh) of at least 7 cm were measured on concentric sample plots with a maximum diameter of 12 m. To increase the efficiency of the inventory, trees with a dbh <30 cm were sampled on plots with smaller radii. This results in four possible plot sizes of 2, 3, 6 and 12 m, where trees with a minimum dbh of 7, 10, 15 and 30 cm are measured. On each sample plot the height of the two largest trees per species were measured using angle measurement instruments. The height of the other trees was estimated based on local diameter-height curves calibrated with the measured trees.

2.1.2 Laser data

The laser scan data were collected with an Optech ALTM 1225 laser scanner in winter 2003/2004, i.e. about one year after the inventory took place. A flight altitude of approx. 900 m above ground yielded an average distance of 1 m between scan points on the ground. The first as well as the last pulse data were automatically classified by the data provider into vegetation- and ground points (reflection from terrain surface).

A digital terrain model (DTM) with one meter pixel spacing was computed from the ground returns using the average height of returns if several reflections were located within one pixel and bilinear interpolation if no return was within the pixel. The value of the respective DTM pixel was subtracted from the first pulse vegetation raw data to obtain vegetation heights. Vegetation height metrics (e.g., percentiles and mean) were derived for every sample plot (Næsset 2002).

2.2 Parameter estimation

If *a* is the location, *b* the scale and *c* the shape parameter, the density of the Weibull distribution is denoted by

$$f(y \mid a,b,c) = \frac{c}{b}\left(\frac{y-a}{b}\right)^{c-1} \exp\left[-\left(\frac{y-a}{b}\right)^{c}\right] \tag{1}$$

for $b,c > 0$.

To estimate the diameter distribution, equation 1 was extended to $f\left(d_i \mid a,b_i,c_i\right)$ where *d* is a diameter, to conditional the scale and shape parameters on predictor variables. The location parameter (*a*) was set to the calliper limit (7 cm) because estimation of this parameter frequently causes numerical problems (Gobakken and Næsset 2004).

Due to the concentric sample plot design, we constructed four censored Weibull distributions for every possible plot radii by

$$g_R\left(d_i \mid a,b_i,c_i\right) = \frac{f(d_i \mid a,b_i,c_i)}{\int_L^U f(x \mid a,b_i,c_i)\,\mathrm{d}x} \tag{2}$$

where *U* and *L* are the upper and lower bounds of the diameters for the concentric sample plot with radius *R*, respectively. This resulted in the functions g_2, g_3, g_6, g_{12}.

The parameters are bound to the predictor variables (laser metrics) with link functions $h_{(P)}^{-1}(x'_{(P),i}\,\beta_{(P)})$,
where $x_{(P)}$ are the predictor variables, β are the coefficients and *P* is either *b* or *c*. The link function for both parameters is the natural logarithm.

The height distribution given DBH and ALS data is assumed to follow a normal distribution:

$$f(h_i \mid \mu_i(P_\mu),\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\left(\frac{h_i - \mu_i(P_\mu)}{\sigma}\right)^2\right) \tag{3}$$

where the expectation value μ is an extended Näslund function that includes metrics from ALS and P_μ are its parameters:

$$\mu_i(d_i,x_i,\beta_{0(Näs)},\beta_{1(Näs)},\beta_{2(Näs)}) = 1.3 + \left(\frac{d_i}{\beta_{0(Näs)} + \beta_{1(Näs)}x_i + \beta_{2(Näs)}d_i}\right)^3 \tag{4}$$

where $\beta_{0(Näs)}$, $\beta_{1(Näs)}$ and $\beta_{2(Näs)}$ are coefficients of the Näslund function and x_i is a predictor variable derived from ALS data.

The bivariate height- and diameter distribution is given by

$$g_R\left(d_i \mid a,b_i,c_i\right) * f(h_i \mid \mu_i(d_i,x_i,\beta_{0(Näs)},\beta_{1(Näs)},\beta_{2(Näs)}),\sigma) = f(d_i,h_i \mid x_i,\beta). \tag{5}$$

The bivariate likelihood function can then be denoted as

$$LLK = \sum_{i=1}^{n} \ln\begin{pmatrix} g_2(d_i \mid a, b_i, c_i) 1_2(y_i) + g_3(d_i \mid a, b_i, c_i) 1_3(y_i) + \\ g_6(d_i \mid a, b_i, c_i) 1_6(y_i) + g_{12}(d_i \mid a, b_i, c_i)\ 1_{12}(y_i) \end{pmatrix} + \sum_{i=1}^{n} \ln\big(f\big(h_i \mid \mu_i(P_\mu), \sigma\big)\big). \tag{6}$$

where $1_U(y_i)$ are size-class dependent indicator functions. If $U \in \Re$ then $1_U(y_i) = \begin{cases} 1 & y_i \in U \\ 0 & y_i \notin U \end{cases}$.

The likelihood function was maximized using the Nelder-Mead algorithm implemented in the function *optim* (Venables & Ripley 2002), within an *R* environment (R Development Core Team 2007)

On average, 12 trees were measured on a sample plot. The predicted distribution can therefore not be compared with observations from one sample plot. Therefore, the observations from plots similar with respect to the explanatory variables are aggregated to what we will call *vegetation height quartile classes* for the remainder of the text. Then, the predicted distributions can be compared with the histogram of the observations.

3. Results

The first and third quartile (Qu1 and Qu3) of the vegetation height and their interaction term (Qu1 * Qu3) were considered as predictor variables for the Weibull scale and shape parameters. The Qu3 was used as additional predictor variable for the height distributions besides the DBH.

The parameters of the Weibull distribution can be predicted by

$$b_i = 1.05 + 0.04\ \mathrm{Qu1}_i + 0.11\ \mathrm{Qu3}_i - 0.002\ \mathrm{Qu1}_i\ \mathrm{Qu3}_i$$
$$c_i = -0.19 + 0.09\ \mathrm{Qu1}_i + 0.02\ \mathrm{Qu3}_i - 0.002\ \mathrm{Qu1}_i\ \mathrm{Qu3}_i \tag{7}$$

Tree height can be predicted by

$\mu_i = d_i / (4.03 - 0.10\ \mathrm{Qu3}_i + 0.30\ d_i)$ with standard deviation 2.23.

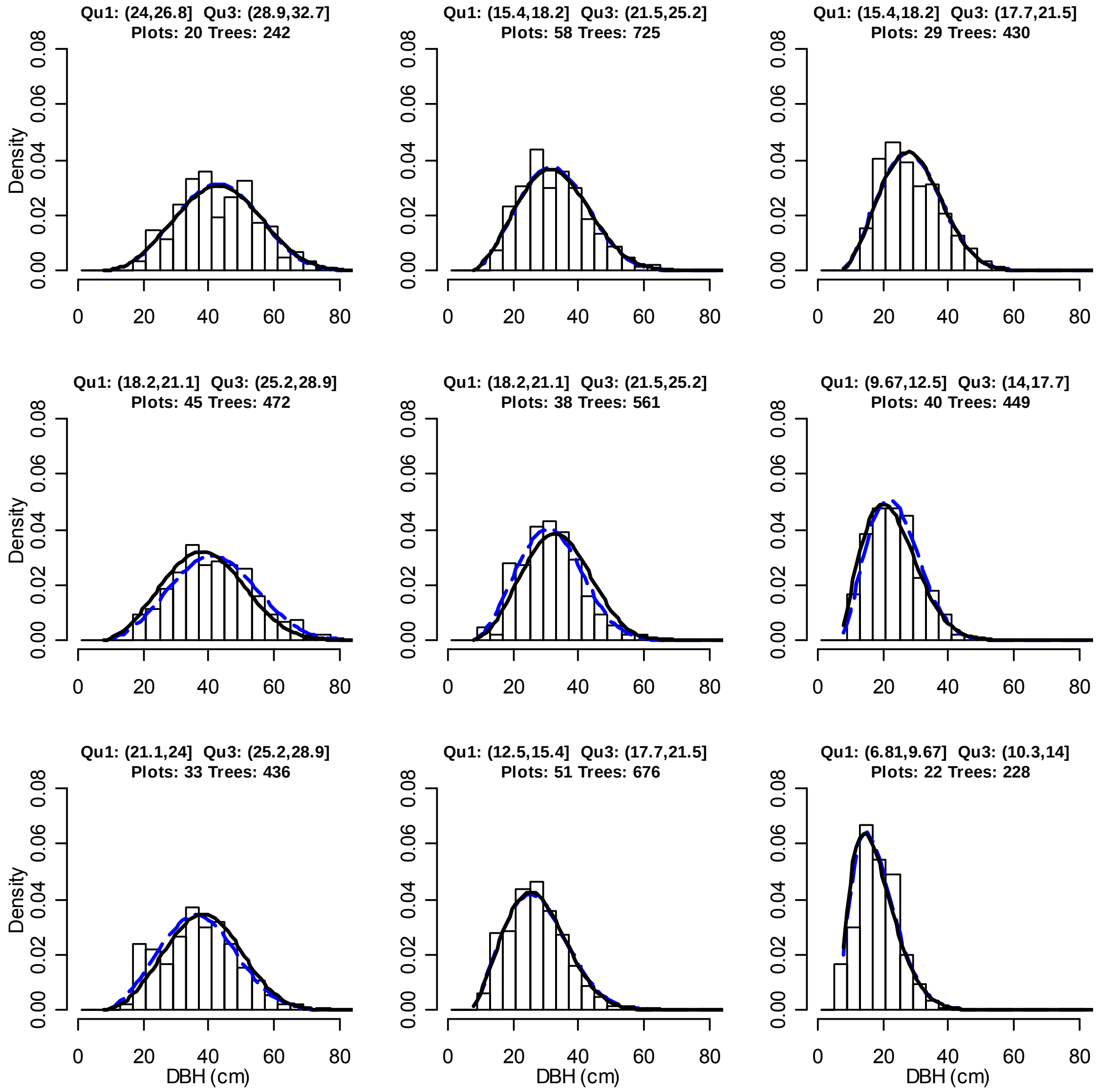

Figure 1: Probability density distribution of observed DBH (histogram) and predicted Weibull distributions (solid graph) for the 9 most densely populated laser-derived vegetation height quartile classes. The dashed curve marks the Weibull distribution which has been directly fitted to the observations. Qu1 denotes the class width of the first quartile (m) and Qu3 the class width of the third quartile (m). Plots and trees represent the number of sample plots and trees in the corresponding plot strata.

The predicted diameter distribution fits well to the observed diameter distributions (Figure). Results for height distributions are not yet available, however the comparison of the expectation value with observed heights is promising (Figure 2). For this graphic, Qu3 was estimated given DBH. (However, the regression model is weak with R^2=0.3.)

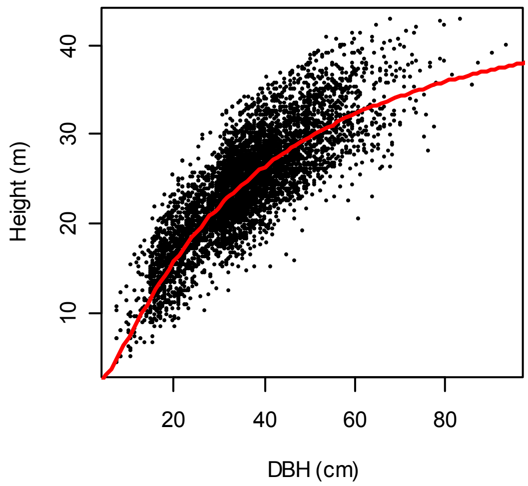

Figure 2: Predicted and observed height given DBH.

The means of the Weibull and the observed distribution was computed for the 20 most densely populated quartile classes (containing at least 3 Plots). As the good conformity of the predicted distribution with the observed distribution supposes, the difference between the mean of the Weibull distribution and the mean of the observations is rather small (Figure 3). The RMSE is 2.78 cm with a bias of 0.48 cm.

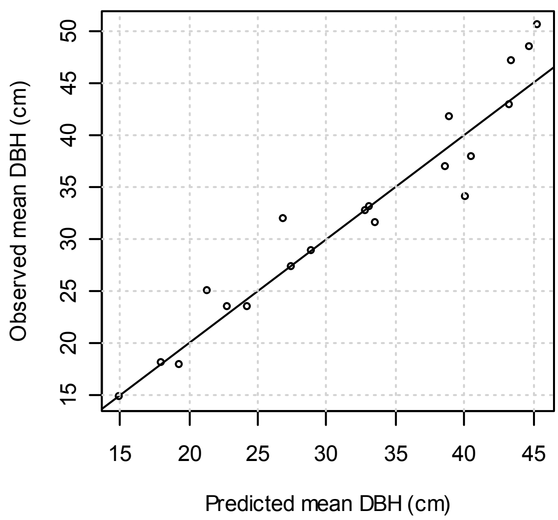

Figure 3: Observed versus predicted mean DBH for the 20 most densely populated quartile classes (circles) and 1:1 line (solid line).

4. Discussion

The proposed bivariate distribution can be used to estimate diameter and height distributions. Since we did not use the parameter recovery method (Mehtätalo et al. 2007), we obtain density distributions. They need to be combined with an estimation of stem number or basal area to get stem numbers per diameter class. For the prediction of assortments, information about tree species are also required. In this study, we assumed the observations to be independent of one another. One major drawback of the used data material is that not all tree heights were measured. This reduces the variance to be modelled.

Acknowledgements

We acknowledge the useful comments of two anonymous reviewers.

References

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, 529-542.

Gobakken, T. and Næsset, E., 2007. Assessing effects of laser point density on biophysical stand properties derived from airborne laser scanner data in mature forest In E. Rönnholm, P.; Hyyppä, H. & Hyyppä, J. (ed.), *ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*, 200.

Hafley, W. and Schreuder, H., 1977. Statistical distributions for fitting diameter and height data in even-aged stands. *Canadian Journal of Forest Research, NRC Research Press*, 7, 481-487.

Magnussen, S. and Boudewyn, P., 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. *Canadian Journal of Forest Research*, 28, 1016-1031.

Magnusson, M.; Fransson, J. and Holmgren, J., 2007. Effects on Estimation Accuracy of Forest Variables Using Different Pulse Density of Laser Data. *Forest Science*, 53, 619-626.

Mehtätalo, L., Maltamo, M. and Packalén, P., 2007. Recovering plot-specific diameter distribution and height-diameter curve using als based stand characteristics. In: P. Rönnholm, P.; Hyyppä, H. & Hyyppä, J. (Eds.). *ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007*.

Næsset, E., 2002. Data acquisition for forest planning using airborne scanning laser. *Forestsat Symposium 2002*.

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

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

Persson, A., Holmgren, J. and Söderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering and Remote Sensing*, 68, 925-932.

Popescu, S. C.; Wynne, R. H. and Nelson, R. F. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian Journal of Remote Sensing*, 2003, 29, 564–577

R Development Core Team. R., 2007. A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*.

Venables, W. & Ripley, B., 2002. Modern Applied Statistics with S. Springer, New York.

Zucchini, W.; Schmidt, M. and Gadow, K. v., 2001. A model for the diameter-height distribution in an uneven-aged beech forest and a method to assess the fit of such models *Silva Fennica*, 35, 169 -183