Diameter distribution modelling 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

The Weibull distribution is one of the most frequently used functions in forestry to fit diameter or height distributions. However, estimating the location parameter of the Weibull distribution frequently causes numerical problems because it is highly correlated with the scale and shape parameters. The location parameter is therefore usually fixed to a certain value. We propose to use the reversed generalized extreme value distribution (RGE) to overcome this limitation. The RGE is a reparametrization of the Weibull distribution that allows estimation of the location parameter. We apply the RGE in the context of a generalized linear model (GLM). In the GLM, the tree diameter is assumed to be the RGE distributed response. It is estimated using area-based methods (vegetation height metrics). While visual comparison reveals a good conformity of the RGE with observed diameter distributions, the smallest diameter (location parameter) is in tendency underestimated by the RGE. For the distribution means, the RMSE is 2.12 cm with a bias of 0.29 cm.

Keywords: reversed generalized extreme value distribution, Weibull distribution, GAMLSS, lidar

1. Introduction

Several studies proved that small footprint airborne laser scanner data (ALS) can be used for estimating forest parameters either using single tree detection algorithms (e.g., Persson et al. 2002, Peuhkurinen et al. 2007) or area-based (also referred to as *plot-wise*) approaches (e.g., Nilsson 1996, Magnussen & Boudewyn 1998, Næsset 2002). Due to their robustness, the latter are used in Scandinavia on operational scales since several years (Næsset 2004). Area based approaches usually provide plot level estimates such as total volume, basal area or mean height. However, for predicting timber assortments, the diameter distribution of a forest stand is needed as an important parameter.

Generally, non-parametric (e.g., Maltamo & Kangas 1998) and parametric (e.g., Hafley & Schreuder 1977) methods can be used to model diameter distributions. Aim of parametric methods is to estimate the parameters of a distribution function. Due to the possibility of using biological interpretable parameters, parametric methods enjoy a high popularity. The Weibull distribution is an often-used function to model diameter distributions (e.g., Bailey & Dell 1973, Nagel & Biging 1995, Cao 2004). Several authors also used ALS to estimate the parameters of Weibull distributions (e.g., Gobakken & Næsset 2004, Mehtätalo et al. 2007).

In managed forests, it frequently occurs that the smallest diameter on sample plots with large trees is larger than zero or the smallest measured tree (calliper limit). However, while the

Weibull distribution has a location parameter, its estimation causes numerical instability since the scale and the shape parameter are highly correlated with it. The location parameter is therefore usually fixed to a certain value (e.g., Breidenbach et al. 2008, Gobakken & Næsset 2004, Cao 2004). The probability of the occurrence (density) of small trees will then be overestimated.

In this paper we describe the use of the generalized extreme value distribution (Johnson et al. 1995), which is provided by (Rigby & Stasinopoulos 2005) as the reversed generalized extreme value distribution (RGE). It is a reparametrization of the three-parameter Weibull distribution and can be used to estimate also the location parameter. We applied a generalized linear model (GLM, Nelder & Wedderburn 1972) with the diameter as the RGE distributed response. The parameters of the RGE distribution are estimated using plot-wise vegetation height metrics derived from small footprint, low density ALS data. Due to the plot design, a combination of several truncated RGE distributions was used.

2. Material and Methods

2.1 Study area

The tree species composition of the 50 km² study site is a managed forest, 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.

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

The reversed generalized extreme value distribution (RGE) is obtained from the generalized extreme value distribution (Johnson et al. 1995, p.76) by replacing y with -y and ξ by -ξ (Rigby & Stasinopoulos 2005). It has the density

$$f(y \mid \xi,\theta,\gamma) = \frac{1}{\theta}\left\{1+\gamma\left(\frac{y-\xi}{\theta}\right)\right\}^{\frac{1}{\gamma-1}} \cdot S(y \mid \xi,\theta,\gamma) \qquad (1)$$

defined for $\xi - \frac{\theta}{\gamma} < y < \infty$

where the equation $S(y \mid \xi,\theta,\gamma)$ is given by

$$\exp\left(-\left\{1+\gamma\left(\frac{y-\xi}{\theta}\right)\right\}^{\frac{1}{\gamma}}\right)$$

which is defined for $-\infty < \xi < y + \frac{\theta}{\gamma}$, $\theta,\gamma > 0$.

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

$$f(y \mid a,b,c) = \frac{c}{b}\left(\frac{y-a}{b}\right)^{c-1} \exp\left[-\left(\frac{y-a}{b}\right)^{c}\right] \qquad (2)$$

for $b,c > 0$.

The RGE is a reparametrization of the Weibull distribution in that

$a = \xi - \frac{\theta}{\gamma}$, $b = \frac{\theta}{\gamma}$ and $c = \frac{1}{\gamma}$.

The parameters of the RGE distribution were estimated using plot-wise derived vegetation height metrics from ALS raw data. The equation $f(y \mid \xi,\theta,\gamma)$ was therefore extended to $f(y_i \mid \xi_i,\theta_i,\gamma_i)$.

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

$$g_R(y_i \mid \xi_i,\theta_i,\gamma_i) = \frac{f(y_i \mid \xi_i,\theta_i,\gamma_i)}{\int_L^U f(x \mid \xi_i,\theta_i,\gamma_i)\,\mathrm{d}x} \qquad (3)$$

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 likelihood function for the parameter estimation is therefore

$$L = \sum_{i=1}^{n} \ln \begin{matrix} (g_2(y_i \mid \xi_i, \theta_i, \gamma_i) 1_2(y_i) + g_3(y_i \mid \xi_i, \theta_i, \gamma_i) 1_3(y_i) + \\ g_6(y_i \mid \xi_i, \theta_i, \gamma_i) 1_6(y_i) + g_{12}(y_i \mid \xi_i, \theta_i, \gamma_i)\ 1_{12}(y_i)) \end{matrix} \quad (4)$$

with $1_U(y_i)$ as size-class dependent indicator functions If $U \in \Re$ then

$$1_U(y_i) = \begin{cases} 1 & y_i \in U \\ 0 & y_i \notin U \end{cases}. \quad (5)$$

The parameters are bound to the predictor variables with link functions *h*:

$$\xi_i = h_1^{-1}(x'_{(\xi),i}\, \beta_{(\xi)}) \qquad \theta_i = h_1^{-1}(x'_{(\theta),i}\, \beta_{(\theta)}) \qquad \gamma_i = h_1^{-1}(x'_{(\gamma),i}\, \beta_{(\gamma)}).$$

where *x* are the predictor variables, β are the coefficients. The identity is the link function for ξ and the natural logarithm is the link function for θ as well as γ.

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 RGE distribution can be compared with the histogram of the observations.

3. Results

The first and third quartile (Qu1 and Qu3) of the vegetation height were selected as predictor variables for all parameters. Their interaction term (Qu1 * Qu3) was considered as predictor variable for the ξ and θ parameters.

The parameters of the RGE distribution can be predicted by

$$\begin{aligned} \xi_i &= 4.15 + -1.20\ Qu1_i + 1.93\ Qu3_i + 0.02\ Qu1_i\ Qu3_i \\ \theta_i &= 0.97 + -0.03\ Qu1_i + 0.11\ Qu3_i + -0.001\ Qu1_i\ Qu3_i \\ \gamma_i &= -0.31 + 0.03\ Qu1_i + -0.05\ Qu3_i \end{aligned} \quad (6)$$

Compared with a Weibull distribution (location parameter fixed at the calliper limit) directly fitted to the observations, the RGE distribution matches well to the observed diameter distributions (Figure). The smallest estimated diameter of the RGE distribution is usually above the calliper limit and especially for plots with large trees, larger than for the Weibull distribution. However, compared with the actual observations, the size of the smallest diameter is still underestimated (Figure).

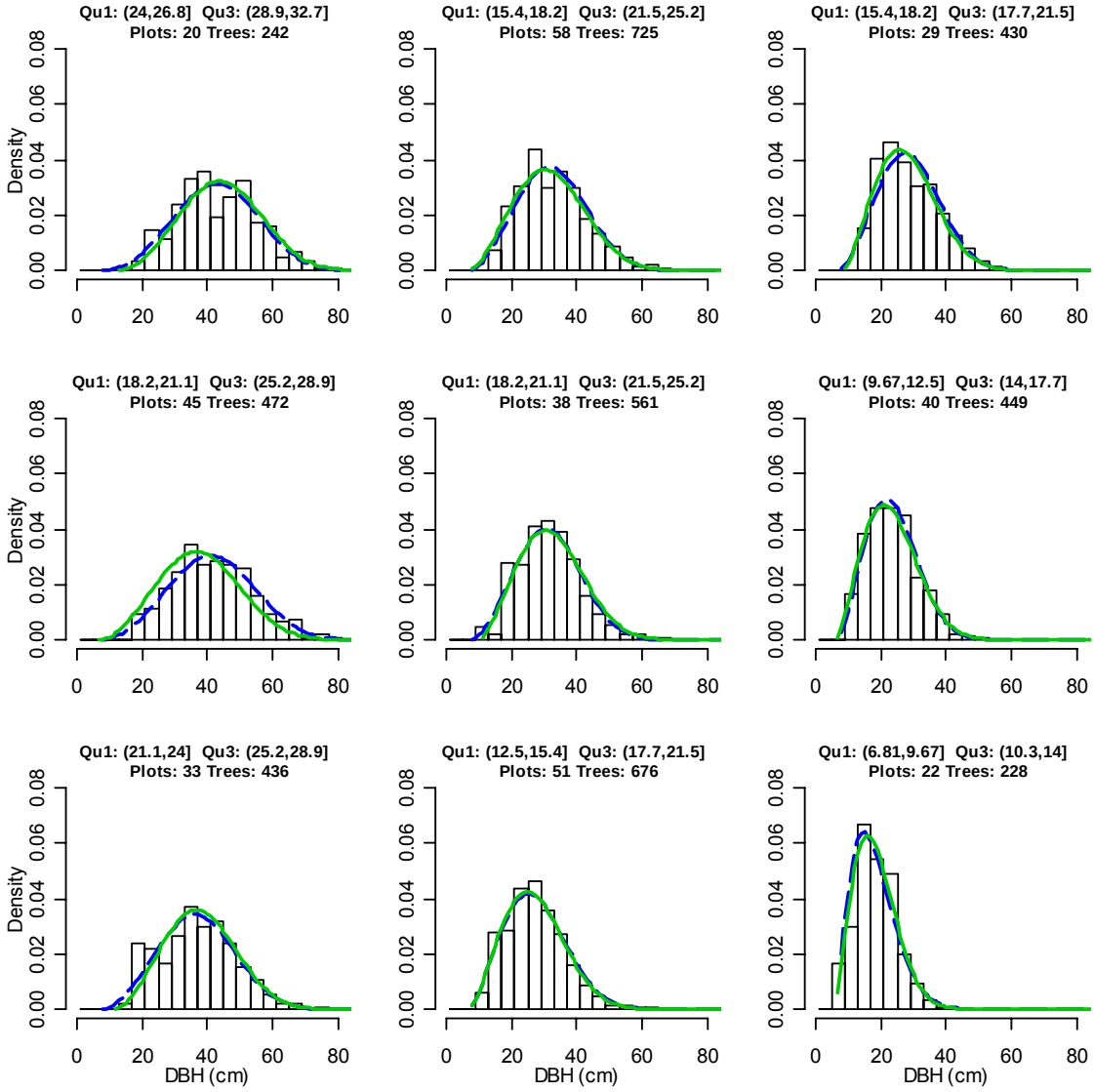

Figure 1: Probability density distribution of observed DBH (histogram) and predicted RGE distributions (solid graph) for the 9 most densely populated laser-derived vegetation height quartile classes. The dashed curve marks the Weibull distribution which has been directly fitted to the observations. Qu1 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.

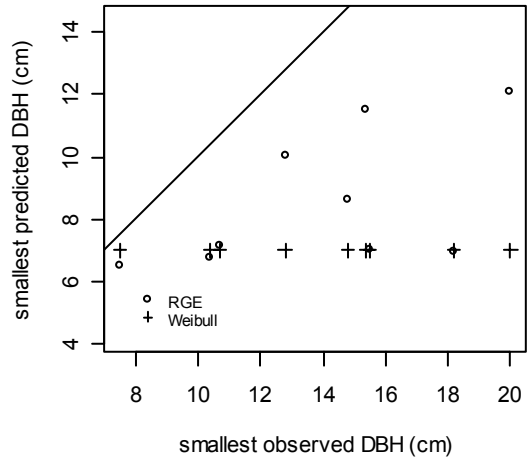

Figure 2: Smallest predicted DBH using the RGE and Weibull distribution and smallest observed DBH (solid line = 1:1 line).

The means of the RGE 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 RGE distribution and the mean of the observations is rather small (Figure 3). The RMSE is 2.12 cm with a bias of 0.29 cm.

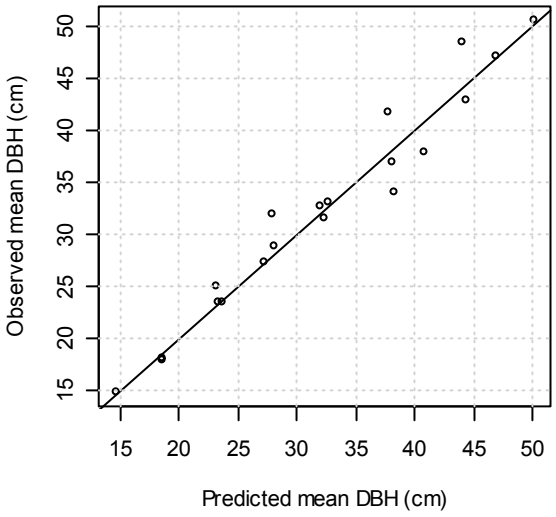

Figure 3: Observed versus predicted mean DBH for the 20 most densely populated quartile classes (circles) and 1:1 line (solid line).

4. Discussion

Numerical problems may occur while estimating the location parameter of the Weibull distribution (e.g., Gobakken & Næsset 2004) because the parameters are highly correlated. Therefore, the location parameter is usually fixed to zero or some other value (e.g., Cao 2004). The reversed generalized extreme value distribution (RGE), as described by Rigby &

Stasinopoulos (2005), is a reparametrization of the Weibull distribution. Since the location parameter of the Weibull distribution is obtained from a combination of the three parameters of the RGE distribution that are not strongly correlated, it can be estimated.

The proposed RGE distribution can be used to estimate diameter distributions. For the prediction of assortments, information about tree heights (for a solution see for example Mehtätalo et al. 2007) and tree species are also required. In this study, we assumed the observations to be independent of one another. Another topic of future research will be how spatial autocorrelation affects the statistical models. Standard errors of the coefficients will also need to be computed. To do so, derivations of the log-likelihood function can be used to compute the Fisher information matrix. The inversion of the Fisher information is the covariance matrix of the parameters.

The GLM used here is the state of the art method to fit conditional distributions. It allows the prediction of parameters, also if reference data stem from small sample plots. Consequently, potential multimodal distributions are not likely to occur since small patches of forest are relatively homogenous.

Acknowledgements

We like to thank Mikis Stasinopoulos and Bob Rigby, the developers of the generalized linear models for location, scale and shape (GAMLSS), for their help and the provision of the RGE distribution. We acknowledge the comments of two anonymous reviewers that helped to improve the quality of this paper.

References

Bailey, R. and Dell, T., 1973. Quantifying diameter distributions with the Weibull function. *Forest Science*, 19, 97-104.

Breidenbach, J., Gläser, C., Schmidt, M., 2008. Estimation of diameter distributions by means of airborne laser scanner data. *Canadian Journal of Forest Research*, 38(6), 1611-1620.

Cao, Q., 2004. Predicting parameters of a Weibull function for modeling diameter distribution. *Forest Science*, 50, 682-685.

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.

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.

Johnson, N., Kotz, S. and Balakrishnan, N., 1995. Distributions in Statistics: Continous Univariate Distributions, Vol. 2. Wiley, New York.

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.

Maltamo, M. and Kangas, A., 1998. Methods based on k-nearest neighbor regression in the prediction of basal area diameter distribution.*Canadian Journal of Forest Research*, 28, 1107-1115.

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.

Nagel, J. and Biging, G., 1995. Schätzung der Parameter der Weibullfunktion zur Genierung von Durchmesserverteilungen. *Allgemeine Forst- und Jagdzeitung*, 166, 185-189.

Nelder, J. and Wedderburn, R., 1972. Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General), JSTOR*, 135, 370-384.

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.

Peuhkurinen, J., Maltamo, M., Malinen, J., Pitkänen, J. and Packalén, P., 2007. Preharvest Measurement of Marked Stands Using Airborne Laser Scanning. *Forest Science, Society of American Foresters*, 53, 653-661.

Rigby, R. and Stasinopoulos, D., 2005. Generalized Additive Models for Location Scale and Shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54, 507-554.

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.