

Toward new site index curves for Douglas-fir in the Netherlands *

Oscar García

December 16, 1996

Abstract

Some data analysis of Netherlands Douglas-fir growth data is reported. The reduction of variability and biases in measures of top height was studied, and site index models were fitted.

Noise can be reduced considerably by using heights estimated from height-diameter regressions instead of using the dominant height measurements directly. The problem of bias when estimating top height from sample plots of various sizes was examined, and a satisfactory estimator was developed.

Site index equations were obtained through maximum-likelihood estimation with stochastic differential equations. There are indications of heterogeneity in the data, suggesting a need to stratify by regions or other categories, or to separate some plots that show peculiar growth patterns. An analysis of this issue employing local knowledge and additional information would be necessary for producing a wholly satisfactory site index model.

Data

Using height-diameter regressions

The standard method of calculating top height in the Netherlands is based on the tallest (or thickest?) tree in 1/100 ha. The measured height in a sub-plot of that area is generally used. Unfortunately, height measurements are difficult and imprecise, and using the observed height directly introduces considerable noise into the data. An alternative would be to fit a regression of height over dbh with all the plot height measurements, and use the estimated regression heights for the largest sub-plot trees. This would smooth-out the height measurement errors.

To test this approach, either the Henriksen semi-logarithmic equation $H = a + b \log D$ or its quadratic extension $H = a + b \log D + c(\log D)^2$ was

*Working Paper. Royal Veterinary and Agricultural University, Department of Economics and Natural Resources, Unit of Forestry.

used. The best equation was selected for each plot according to the adjusted R^2 criterion [5]. The Henriksen is simple and convenient to use and, although occasionally can predict negative heights for small trees (not a problem in this application), has usually been found among the most accurate height-dbh equations [7, 4, 6, 1]. The quadratic semi-log can be justified by the relatively large height samples (often over 50) found in this data, and our interest in predicting height just for the top diameter.

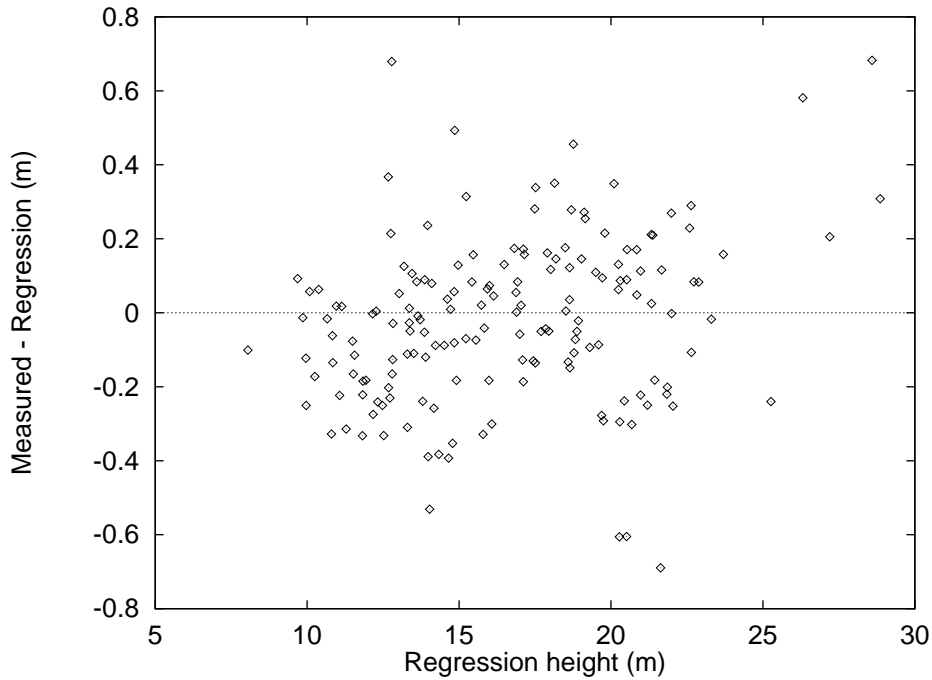


Figure 1: Differences between top heights from measured and from regression values. Mean of heights of the largest tree in 16 sub-plots of 1/100 ha for each 0.16 ha plot

Figure 1 shows that there is no appreciable bias when using the regression heights in place of the actual measurements. The variability, however, is smaller (Figs. 2, 3.) Regressed heights have therefore been used in what follows.

Estimating top height

As used in the Netherlands, top height may be defined as the expected height of the largest tree in a 1/100 ha plot (the largest dbh tree, for convenience.) In the WAU data the permanent sample plots are subdivided into 1/100 sub-plots, and it is known to which sub-plot each tree belongs. It is therefore straightforward to compute the plot top height as the mean of the largest trees from each sub-plot.

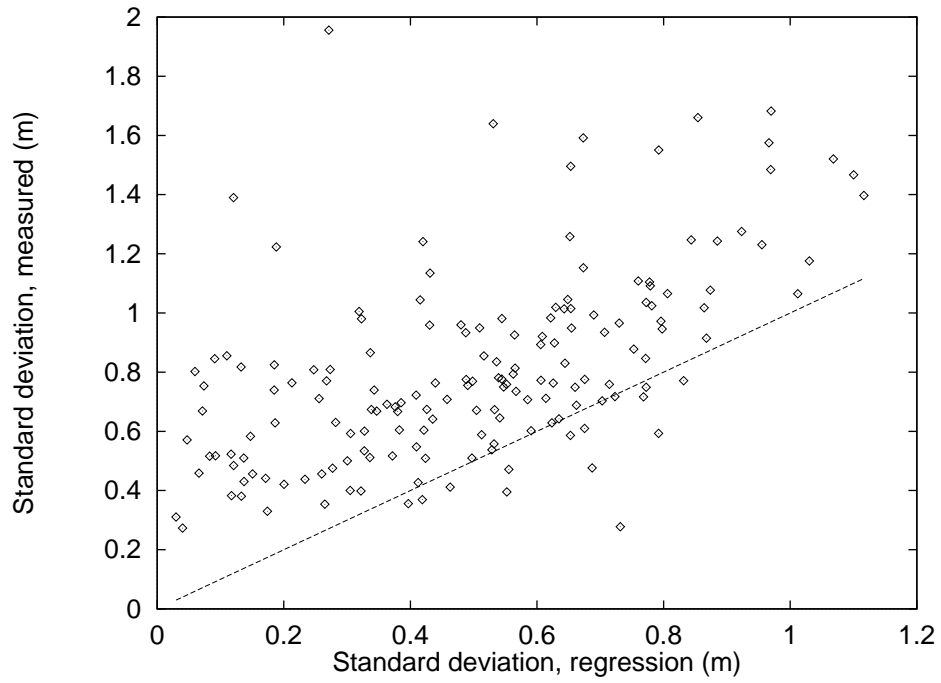


Figure 2: Standard deviations of measured and regression heights of the largest sub-plot trees. Each point based on 16 1/100 ha sub-plots in a 0.16 ha plot

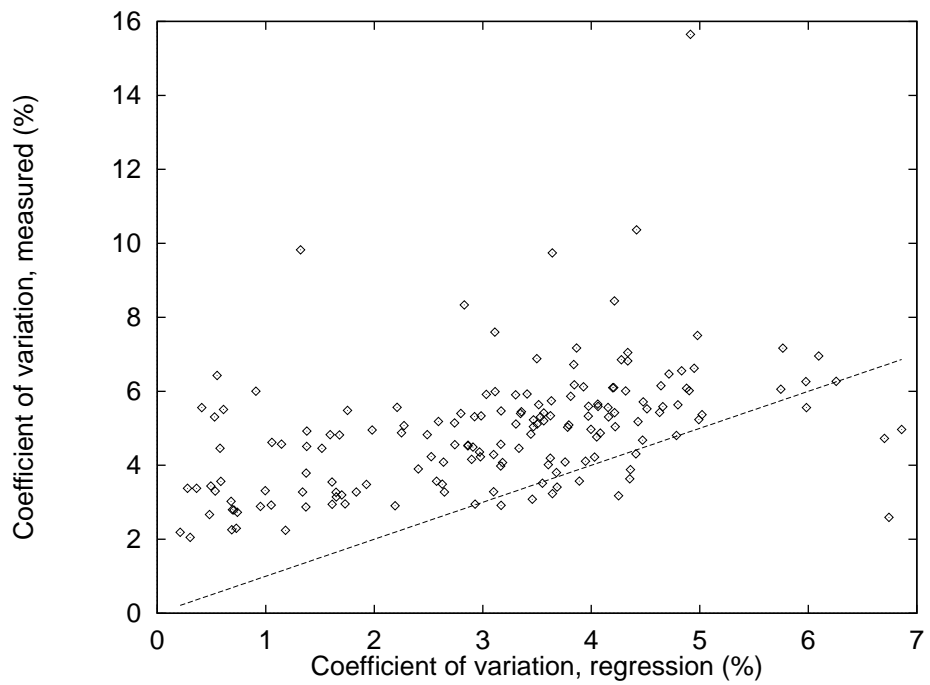


Figure 3: Coefficients of variation for measured and regression heights of the largest sub-plot trees. Each point based on 16 1/100 ha sub-plots in a 0.16 ha plot

In the IBN-DLO data, however, no such partition in subplots is available. This would be also the case in many model applications based on inventory data. It is necessary, then, to estimate top height for plots where the relative positions of the individual trees are not known.

The usual way of calculating top height for plots of varying sizes is as the mean height of a number of largest trees corresponding to the proportion of 1 in 1/100 ha. For example, with a plot of 0.1 ha the largest 10 trees in the plot would be used, or the largest 4 in a plot of 0.04 ha. This, however, produces a systematic overestimation of the top height as defined above [11, 13, 12]. This is easier to see in the 0.04 ha example: the mean of the 4 largest trees in the plot tends to be greater than that of 4 trees constrained to be each in a different 0.01 ha sub-plot.

Using the 0.16 ha plots in the WAU database, the extent of the potential biases and the performance of alternative estimators was studied. This research will be described in detail elsewhere, but the main results may be summarized as follows.

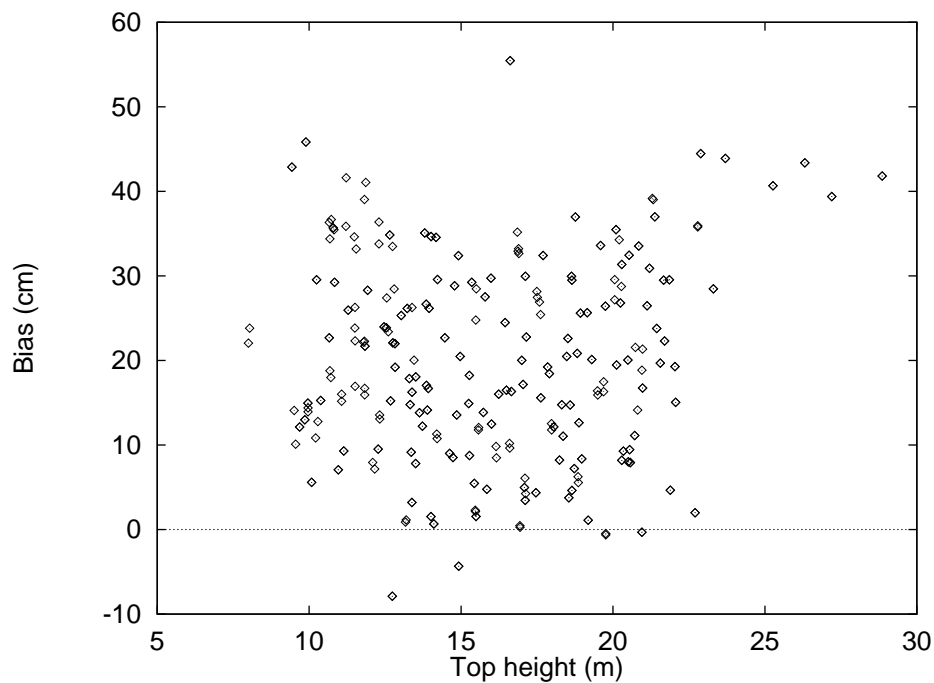


Figure 4: Bias of conventional top height estimates. Mean height of the 16 largest dbh trees in 0.16 ha plots minus the mean of the largest in each 0.01 ha sub-plot

Figure 4 Shows the differences between the conventional method of averaging the heights of the largest (by diameter) 16 trees in the plot and the mean for the largest in each 0.01 ha sub-plot. The bias might be considered as not too large, especially when expressed on a relative basis. But it is not negligible, and it would be desirable to reduce it as much as possible.

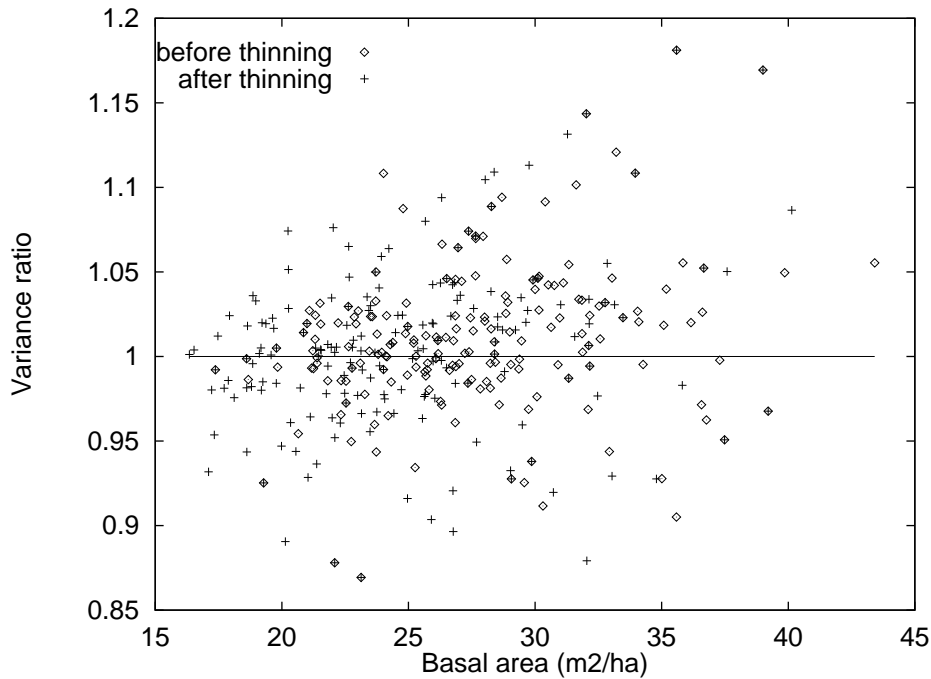


Figure 5: Ratio of dbh variances for 0.01 ha subplots to those for the whole 0.16 ha plots

Developing less biased estimators becomes complicated if tree sizes are spatially correlated. Then, in addition to the purely sampling effect described above, spatial structure may also contribute to biases in a complex manner. A simple way of testing for spatial structure is to compare the diameter variance computed for a plot with that obtained for sub-plots [9]. Figure 5 shows the ratios between the two variances for the 0.16 ha plots from the WAU database. Similar results were obtained subdividing into 0.04 ha plots. For comparison, data from young unthinned eucalypt plantations in Chile is shown in Figure 6. There is no evidence of any substantial spatial structure in the Douglas-fir data. It has been observed that thinning can destroy spatial correlations arising from competition and microsite effects [14, 9], and the frequent thinnings might play a role here. Note however that very young stands are not represented in the data, so it is not clear if the absence of spatial structure would apply to them.

Under the assumption of no spatial structure, and various other assumptions about the tree size distributions, several theoretically unbiased or nearly unbiased estimators of top height can be obtained. Estimators were tested with the WAU 0.16 ha plots. Both bias and variance need to be considered, as well as ease of use.

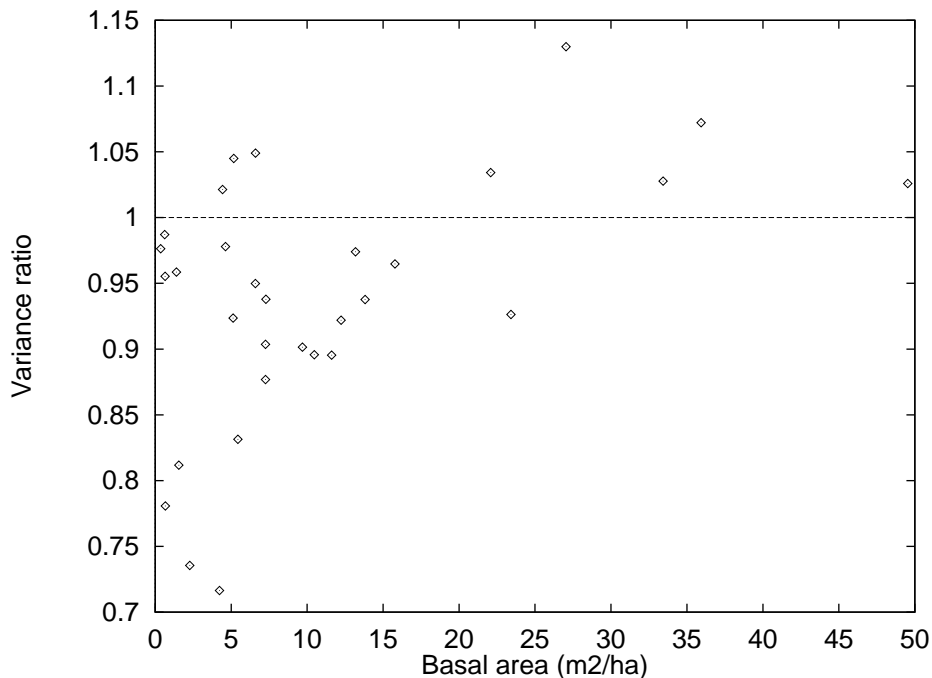


Figure 6: Ratio of dbh variances for 0.0111 ha subplots to those for 0.10 ha plots, eucalypt plantations in Chile

Best results were obtained estimating the logarithm of top dbh by

$$\log D_{\text{top}} = \sum_{i \leq n-m+1} \frac{\binom{n-i}{m-1}}{\binom{n}{m}} \log D_i, \quad (1)$$

where D_i are the dbh of the n trees in the plot, ordered from largest to smallest, and m is the corresponding number of trees in 1/100 ha (interpolation can be used if m is not an integer.) It is shown that (1) is a non-parametric uniformly minimum variance unbiased (UMVU) estimator of the expected largest $\log D$ in samples of size m (expected largest order statistic), given a random sample of size n . Using Henriksen's height-dbh equation, this gives us the UMVU estimator of top height, if there is no spatial structure and we neglect any effects of the variability in m . The empirical tests showed no appreciable bias. For consistency, all top heights were computed with this estimator.

Data screening

The data was plotted and examined in a number of ways, to check for errors and irregularities. Only a small number of obviously anomalous measure-

ments were rejected on this basis. The height-dbh relationships were also assessed, and measurements with less than 10 height-dbh pairs were not used.

As usual, ages were computed changing close to the middle of the growing season (1st July). However, because of the variable proportions of growth achieved near the culmination of the growing season and the consequent uncertain “physiological age”, measurements carried out during the months of May, June and July were excluded from the site index computations. A total of 838 measurements from 153 plots were used.

Site index models

Model and parameter estimation

The methods described in [8] were used. The site index curves are based on the commonly-used von Bertalanffy-Richards model. In differential equation form this can be written as

$$\frac{dH^c}{dt} = b(a^c - H^c) , \quad (2)$$

with the general solution

$$H = a[1 - (1 - H_0^c)e^{-b(t-t_0)}]^{1/c} . \quad (3)$$

Here H is top height, t is age, and (H_0, t_0) is an initial condition, i. e., some height-age point on the curve. To obtain site index curves one of the parameters a , b , or c , or some function of these, is assumed to be site-dependent, varying across plots (“local” parameter), while the rest are common to all plots (“global”.) The initial t_0 and H_0 can be fixed at 0, or one of them can be included as an additional parameter to be estimated.

For estimation, environmental perturbations are represented by a Wiener stochastic process with variance σ^2 added to the right-hand-side of (2), and the observed H^c are assumed to deviate from their actual values by independent normal measurement errors of variance η^2 . A different perturbation variance of average magnitude σ_0^2 for young stands (before the first measurement) can also be included.

The resulting stochastic differential equation can be integrated to obtain the likelihood function. Then the computed likelihood is maximized with a specialized numerical procedure over all the global and local parameters ($a, b, c, \sigma^2, \eta^2$, and possibly t_0, H_0 , and σ_0) to obtain their maximum-likelihood estimates. See [8] for details.

Several versions of the basic model were tried:

1. a local. This gives anamorphic site index curves, with different asymptotes a , a popular choice in the literature.

2. b local. The curves have a common asymptote a , differing in the time scaling factor b . This was found to be the best model in all but one of the regional growth models for radiata pine and Douglas-fir developed in New Zealand [10].
3. Burkhart and Tennent’s model [3], obtained by substituting $b = \beta S$ and $a = S/(1 - e^{-\beta S k})^{1/c}$, where β is a new global parameter, S is the site index (local), and k is the chosen site index age [10]. Values of 20 and 50 years were tried for k . Note that, in more direct analogy with the models that follow, this could also be written keeping b as the local variable and making $a = pb/(1 - e^{-bk})^{1/c}$.
4. A new composite model that can take forms intermediate between 1 and 2. This is very general, in that covers situations in which a and b are replaced by any power transformations of the form $\mu^\alpha \theta^\beta$, where θ is a local parameter and the others are global. Actually, it is seen that it is sufficient to keep b as a local parameter, and to substitute pb^q for a , where p and q are globals.
5. Like 4, but with linear transformations: b local and $a = p + qb$.

In addition, versions with and without the additional parameters t_0 , H_0 or σ_0^2 were tried. These and the variances σ^2 and η^2 were treated as global.

Results

Little or no improvement in the likelihood was obtained by allowing the origin to differ from zero, or by introducing the different variance for young stands. The estimated t_0 was always small, below one year. Therefore, only the results for $t_0 = H_0 = 0$ and $\sigma_0 = \sigma$ are shown in Table 1. Model 3 with index ages 20 and 50 are indicated by 3/20 and 3/50, respectively. Units are meters for height and years for age.

Table 1: Results

Model	log-likelihood	a	b	c	σ	η
1	1200.0	local	0.03928	0.5763	0.02973	0.03627
2	1236.5	34.30	local	0.6029	0.03036	0.05371
3/20	1235.0	see text	0.003144 <i>S</i>	0.6155	0.03170	0.05553
3/50	1247.9	see text	0.001610 <i>S</i>	0.5618	0.02521	0.04955
4	1247.4	81.29 <i>b</i> ^{0.2827}	local	0.5673	0.02577	0.05019
5	1248.2	23.71 + 222.6 <i>b</i>	local	0.5665	0.02564	0.05030

When comparing the log-likelihoods for models with the same number of parameters, differences of more than about two units may be seen as “significant”, and between 1/2 and 3 units should be discounted for each additional parameter [8, 2]. By this criterion there is little to chose between

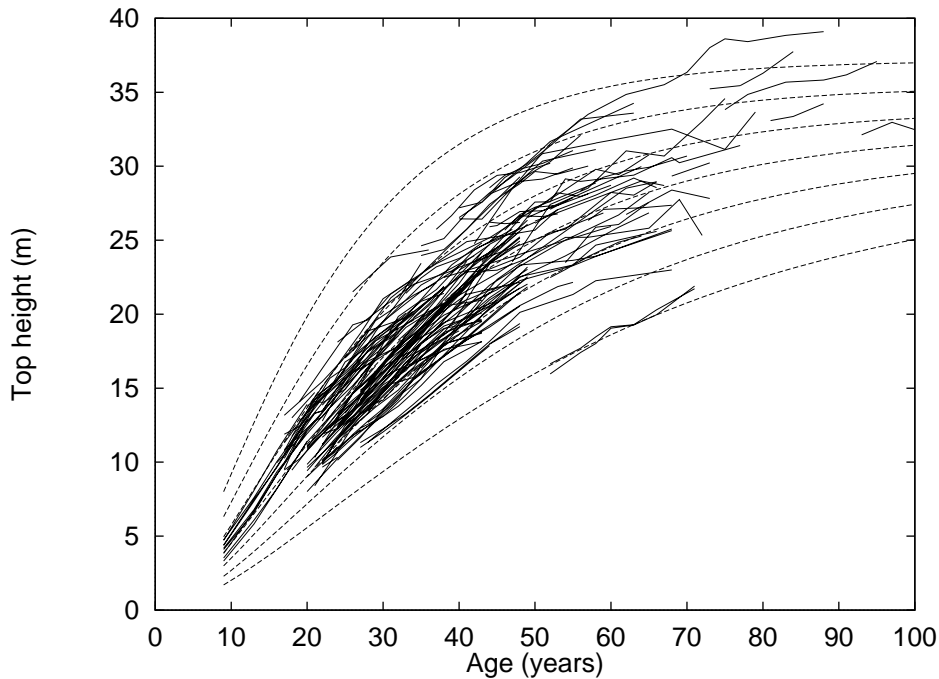


Figure 7: Data, and site index curves for model 5. Site indices 16, 19, . . . , 34 for index age 50

models 3/50, 4 and 5, at least if the index age in 3 is counted as a parameter. In fact, graphing the curves these are seen to be very close. Model 5 may be preferred for its higher likelihood value, and for not depending of the somewhat arbitrary selection of an index age. Figure 7 shows the data, and site index curves from model 5.

Unfortunately, these models are somewhat cumbersome to use. Unlike with 1 and 2, given an (H, t) pair it is not possible to solve analytically (3) for the local parameter (or site index), and iterative numerical methods have to be used. Four or five Newton iterations are sufficient, but it is not as convenient as having an explicit formula. Therefore, model 2 might still be worth considering. This model is compared to model 5 in Figure 8.

Most of the plots show a strong tendency for the height growth to slow down considerably at older ages. There are however a few that have grown to much greater heights than usual, and are not well represented by neither model. The low final heights suggest some soil limitation, or apical damage caused by wind, salt spray, or other factors (New Zealand Douglas-fir data, unpublished, shows strong height growth beyond 40 meters, well fitted by model 2 with an asymptote around 67 meters.) The plots that do not exhibit this slow down, notably plots 11, 12, 47, 977 and 978, might represent different or atypical growing conditions.

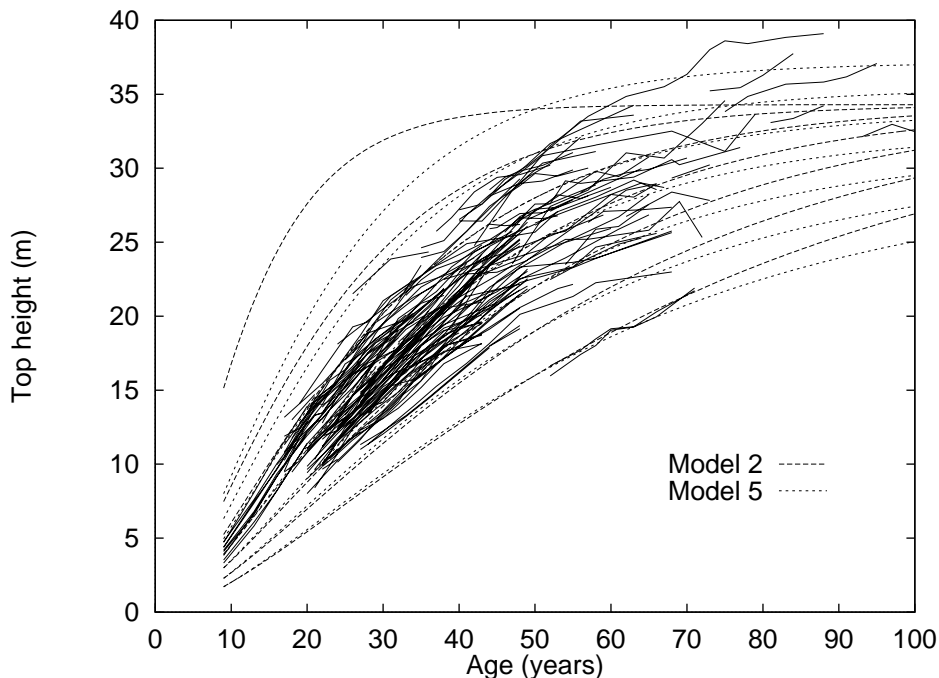


Figure 8: Data, and site index curves for models 2 and 5

Height prediction and site index estimation

As just mentioned, as long as the observed growth pattern differences are not clarified it is not possible to have much confidence in the models that we have obtained. It may be useful, however, to describe the necessary computations with the models for experimentation or other uses.

Given one or more height-age measurement pairs for a plot or stand, it is typically desired to estimate a site index and/or perform height projections. This is a two-stage process: first the local parameter (b) is estimated, and then this is used to predict the height(s) at other age(s) — at the index age for the site index.

For model 2 the basic equation is

$$H = a(1 - e^{-bt})^{1/c} , \quad (4)$$

where a and c are given in Table 1, and b is the site-dependent parameter varying from plot to plot. Given a height-age pair, (4) can be solved for b :

$$b = -\ln[1 - (H/a)^c]/t . \quad (5)$$

(Of course, this fails if the height is above the 34.3 m asymptote.) Then, the height at any age is estimated by substituting the age in equation (4). In particular, the site index is obtained by inserting the index age k . Or,

having calculated the site index S , the height at other ages is given by

$$H = a\{1 - [1 - (S/a)^c]^{t/k}\}^{1/c} .$$

If there are several $H - t$ pairs available, various procedures could be devised. For most applications it would be probably sufficient to use the average of the site indices calculated for each observation.

For model 5, $p + qb$ is used in place of a in (4), with the constants p and q given in Table 1. It is not possible to get an explicit expression like (5) for b . The following Newton iteration may be used instead:

$$b \leftarrow b + \frac{H - (p + qb)(1 - e^{-bt})^{1/c}}{(p + qb)t/c + [q - (p + qb)t/c](1 - e^{-bt})^{1/c}} .$$

This can be started with $b = 0.04$ (near the average value) and repeated until there is no significant change. Or, to simplify programming, just iterated 5 times, which would guarantee sufficient accuracy.

Acknowledgments

The data was collected by IBN-DLO in The Netherlands and provided by Dr. Frits Mohren. His assistance and that from his staff at the DLO Institute for Forestry and Nature Research, Wageningen, is gratefully acknowledged. Most of the work was carried out while the author was a visiting professor at the ENGREF, Nancy, France.

References

- [1] Arabatzis, A. A. and Burkhart, H. E. "An evaluation of sampling methods and model forms for estimating height-diameter relationships in loblolly pine plantations". *Forest Science* 38, 192–198. 1992.
- [2] Atkinson, A. C. "A Note on the Generalized Information Criterion for Choice of a Model". *Biometrika* 67, 413–418. 1980.
- [3] Burkhart, H. E. and Tennent, R. B. "Site Index Equations for Douglas fir in Kaingaroa Forest". *New Zealand Journal of Forestry Science* 7, 417–419. 1977.
- [4] Curtis, R. O. "Height-diameter and height-diameter-age equations for second-growth Douglas-fir". *Forest Science* 13, 365–375. 1967.
- [5] —

- [6] Ek, A. R. “Performance of regression models for tree height estimation with small sample sizes”. In: *Statistics in Forestry Research, Proceedings of the meeting of the IUFRO subject group S6.02, Vancouver, B. C., Canada*. 1973.
- [7] García, O. “Ecuaciones altura-diámetro para pino insigne” [height-diameter equations for radiata pine]. Instituto Forestal, Chile. Nota Técnica 19. 1974.
- [8] García, O. “A stochastic differential equation model for the height growth of forest stands”. *Biometrics* 39, 1059–1072. 1983.
- [9] García, O. “What is a diameter distribution?” In: Minowa, M. and Tsuyuki, S. (eds.) *Proceedings of the Symposium on Integrated Forest Management Information Systems*. Japan Society of Forest Planning Press. 1992.
- [10] García, O. and Lawrence, M. E. “Height growth of radiata pine in New Zealand”. Submitted for publication.
- [11] Matérn, B. “Om skattning av övre köjden” [On estimating the dominant height]. *Sveriges Skogsvårdsförbunds Tidskrift* 74, 51–53. 1976.
- [12] Pierrat, J. C., Houllier, F., Hervé, J. C., and Salas González, R. “Estimation de la moyenne des valeurs les plus élevées d’une population finie: Application aux inventaires forestiers”. *Biometrics* 51, 679–696. 1995.
- [13] Rennolls, K. ““Top height”; its definition and estimation”. *Commonwealth Forestry Review* 57, 215–219. 1978.
- [14] Schoonderwoerd, H. and Mohren, G. M. J. “Autocorrelation and competition in even-aged stands of Douglas-fir in the Netherlands”. In: Ek, A. R., Shifley, S. R., and Burke, T. E. (eds.) *Forest Growth Modelling and Prediction*. USDA Forest Service, General Technical Report NC-120. 1988.