FRI Project Record

No. 5028

PROJECTING INVENTORY DATA: PREDICTING INDIVIDUAL TREE HEIGHT GROWTH

A.D. GORDON

REPORT No 47. **APRIL 1996**

FRI / Industry Research Cooperatives

EXECUTIVE SUMMARY

Methods are needed to accurately project inventory data through time so that product yield can be predicted not just at the time of inventory but at any future time. Stand-level growth models can be used, but can predict individual tree height increment only approximately.

This report describes the development of a method for predicting individual tree height from site, stand and tree information.

Initially the possibility of directly predicting height increment was examined but results were not promising.

Existing and new methods, based on the relationship between tree diameter and height, were compared. The new equation relates tree height to a proportion of stand mean top height based on the relative diameter of the tree. It requires information on the site (region, altitude), the stand (stocking, mean top height, mean top diameter) and the tree (*Dbh* and position in the *Dbh* distribution), at the start and end of the prediction period. This equation is shown to predict tree height accurately.

Using this method at the start and end of a period gives an estimate of tree height increment which can be used in the projection of inventory data.

PROJECTING INVENTORY DATA: PREDICTING INDIVIDUAL TREE HEIGHT GROWTH

Introduction

Forest inventory enables the forest manager to assess the potential yield of a stand of trees. Young stands can be adequately assessed without detailed measurements, but as a stand reaches mid-rotation more detail is required. The size of the yield and its break-down into possible log products must be determined.

Commonly used inventory procedures can provide the requisite level of detail but the results still represent a snap-shot of the stand at one point in time. What is needed is a method to take the inventory description of the stand and use it as a basis for predicting the potential yield at any point in the stands future. Techniques are available for using stand-level growth models to project inventory data through time, but they involve aggregating the tree information before projecting the growth, then disaggregating the "grown" stand back to its individual trees.

Comparisons of methods used to project tree diameters have shown that simply scaling the initial diameter, by a factor based on stand growth model basal area predictions, under-estimates the variance of the resulting diameter distribution (Gordon and Lawrence 1992). To address this problem an individual tree diameter increment function was developed (Gordon and Lawrence 1994), which can be used as an adjunct to a stand level growth model. This exploits the stability of stand level models, which are usually well-tested and trusted over a range of sites, silvicultural practices and projection periods, while allowing the increment to be apportioned to trees on the basis of the individual tree model.

Although stand models predict the growth of some measure of stand height (mean top height, predominant mean height) individual tree height increment is needed when projecting inventory data. Simply using the increment in stand height growth, as a common estimate of the height growth of each tree, is a viable approximation but it does not use any of the information available about each tree.

This report describes the development of a method for predicting individual tree heights from site, stand and tree information. Using this method at the start and end of a period gives an estimate of individual tree height increment which can be used in the projection of inventory data.

Data

Plots were selected from the Permanent Sample Plot data base (McEwen, 1978; Pilaar and Dunlop, 1989) from seven growth modelling regions, to form a general data set. At least 15 trees measured for diameter were required in each plot to reduce the variability of estimates of stand parameters derived from each plot. This ruled out very small plots, particularly where the stocking was low. Plots were required to have had three or more consecutive measurements, with the first measurement somewhere between ages 15 and 25 inclusive. Only plots with "normal" levels of mortality (ie. excluding windthrow, poison thinnings etc.) were chosen, and all thinning operations were required to have been completed prior to the initial measurement. An estimate of mean top height was required.

A total of 291 plots were selected which met these criteria. They are listed by Gordon and Lawrence (1994). All the trees which had been measured for height were extracted from the Permanent Sample Plot system giving 3458 trees with between 2 and 13 height measurements per tree.

Screening Tree Height Increments

Procedures were developed to calculate the annual height increment of each height tree after each plot measurement excluding the last. This involved interpolation where re-measurement did not take place 12 months later. Trees which died were excluded from the data set from the last re-measurement where their diameter increment was less than or equal to zero. At each measurement point, from which increments were derived, corresponding stand statistics were calculated.

Tree height increments were found to be highly variable. Although this was not unexpected, a surprisingly large number of increments were found that were negative or exceeded 4m. Overlaid graphs of height and height increment against age were examined for all trees with unusual increment values. In many of the graphs there appeared to be several sequences in the height measurements; the breaks in these sequences possibly due to changes in the position in the sample plot from where the measurements were taken. Other graphs simply contained individual extreme measurements which were usually associated with two unusual increments.

Height increment tends to slowly decrease with age (in the age range of interest), so the following approach was taken to screen out incorrect measurements: For each height tree a regression of increment on age was fitted . The difference between the actual and the conditional mean increment at each age was then compared with the confidence interval around the mean. This interval was bounded at 2.0 m, as in some sample plots with only a few measurements the degrees of freedom were too low to produce a reasonable confidence interval. Those observations that deviated from the expected increment by more than this were removed from the data set. This method of screening the data allowed each tree to have its own increment pattern and included some allowance for between-year differences which might be caused by differences in annual rainfall for example.

Further screening was made on the basis of comparing individual height increments with the increment of tree diameter and stand mean top height for the same period. Graphical plots showed up anomalous observations for checking. The data set which resulted comprised 13440 observations. Table 1 shows the range of site index and age by growth modelling region. Table 2 shows the range of tree heights and height increments and the number of observations by growth modelling region.

Table 1. Range of Site Index and Age.

	Site (m)			Age (years)		
Region	Min	Mean	Max	Min	Mean	Max
CANTY	21.1	24.91	29.0	15.1	20.72	29.8
CLAYS	23.4	29.64	34.5	15.0	21.92	29.1
GDNS	20.0	26.93	31.1	14.8	19.91	30.1
HBAY	23.2	28.38	33.5	14.7	21.74	33.0
KANG	22.9	32.59	38.1	15.0	20.28	40.0
SANDS	19.5	25.23	32.7	14.6	23.30	45.7
SOUTH	20.6	23.55	27.9	14.6	23.65	38.0
ALL	19.5	28.32	38.1	14.6	21.41	45.7

Table 2. Range of Tree height and height increment.

	Number	Tree Height (m)			Height Increment (m)		
Region	of obs.	Min	Mean	Max	Min	Mean	Max
CANTY	301	15.8	24.57	37.9	0.0	0.97	3.9
CLAYS	831	15.5	30.43	43.7	0.0	1.00	3.6
GDNS	2538	10.3	25.28	41.0	0.0	1.16	3.9
HBAY	1365	15.2	29.58	45.3	0.0	1.23	3.6
KANG	4449	10.2	31.65	51.8	0.0	1.30	3.7
SANDS	1959	14.3	26.37	40.4	0.0	0.78	2.9
SOUTH	1997	11.3	26.23	47.6	0.0	1.05	3.5
ALL	13440	10.2	28.43	51.8	0.0	1.12	3.9

The variables available in the data set can be grouped loosely as:

- Site variables including growth modelling region, plot identifier, site index, latitude, longitude, altitude, spring, summer and annual rainfall, distance from the sea, and sufficiency scores for the elements N,K,P,B and Mg from the nutritional atlas (Hunter *et al*, 1991).
- Stand variables such as year of measurement, mean top diameter, basal area, mean top height and its increment and stocking.
- Tree variables including *Dbh*, *Dbh* increment, height, height increment, basal area in trees larger than the subject tree and age.

A number of variables, such as relative diameter, *Dbh* to height ratio and relative spacing, were derived from combinations or transformations of the variables available in the data set.

Possible Approaches

An approach currently in use (implemented in the GroMARVL software for example) uses the increment of stand height (mean top height) as an estimate of individual tree height increment. In this way tree height growth is tied to the stand height growth which allows increment levels to be set by the stand height model used and the site index. However this does not allow for any differences in height growth within the stand which may be due to differences in growth rates between trees of different crown classes, for example.

Other approaches which are connected with stand height growth use some form of tree height / *Dbh* relationship to predict the tree height directly from its *Dbh* and stand parameters. Increment can then be determined if the stand parameters and *Dbh* are known at the beginning and end of the increment period. This method is available through the "Weibull tables" (Lawrence 1990) used in the stand table (diameter distributions) generator in STANDPAK, but is not used in STANDPAK for this purpose.

Alternatively, individual tree height increment can be predicted directly using a height / age growth curve for each tree, allowing the shape of the curve to alter with changing tree and stand parameters. Although this promises a great deal of flexibility, there is no guarantee of consistency with stand height growth.

Initially individual increments were examined to determine which variables could be used to predict height increment and to test how successful this approach could be. Two methods based around stand height growth were then evaluated. The first used equations that are already in place that have been derived regionally for stand table generation in conjunction with Weibull *Dbh* distributions (Lawrence 1990). The Petterson curve is used to relate tree height (h) and *Dbh* (d) in the following way:

$$h = 1.4 + \left(\beta_0 + \frac{\beta_1}{d} \right)^{-2.5} \qquad \text{Equation 1.}$$

Lawrence (1990) provided a series of equations to predict the β_0 coefficient of the curve from the parameters of the stand (age, mean top height, basal area and stocking). If stand mean top height is known, and mean top diameter can be calculated, then the β_1 coefficient can be solved directly. So height increment can be calculated when the stand parameters and tree *Dbh* are known at the beginning and end of the period.

The second method was suggested by considerations of consistency between the tree heights and the stand mean top height which, by definition, is the average height of the tree of mean top diameter. If tree height is estimated by adjusting mean top height by the ratio of *Dbh* to mean top diameter, consistency is assured. The following equation was used:

$$h = \bar{h}_{100}\left(\frac{d}{\bar{d}_{100}}\right)^{f()} \qquad \text{Equation 2.}$$

where *f()* may be some function of site, stand and tree variables.

Methods and Results

Direct Prediction of Height Increment

The variables available in the data set were tested as linear predictors to tree height increment using "all-subsets" regression, with the three best (largest R-square) regressions calculated for 1 to 12 predictor variables. The R-square values rose rapidly with the number of predictors, tending toward an asymptote of around 0.2, implying that a maximum of only 20% of the variation in height increment could be accounted for. There was considerable interchange in the variables selected in the best regressions for each number of predictors.

To give a better idea of which were the important predictor variables, the same approach was tried using the logarithm of tree height increment as the dependent variable, under the assumption that any factors affecting the size of the increment would be likely to interact and hence a multiplicative model should be used. A similar proportion of the variation was accounted for but this analysis indicated that the important variables included age, site index, stand mean top height, the *Dbh* to height ratio and distance from the sea, or altitude.

Repeating the analysis by region added little information. Site index, stand mean top height and the *Dbh* to height ratio all appeared as predictors but apart from these there was no clear or consistent selection of predictor variables. Tree height, and height relative to mean top height, did not show up as useful predictors.

Principal component regression was used in an attempt to determine how many underlying "factors" were influencing height increment. 87% of the variation in the predictor variables was accounted for by the first 7 components. The regressions showed components 1,2 and 6 to be the most useful predictors. Unfortunately components 1 and 2 were formed from fairly equal contributions of a large number of the predictor variables. Component 6 however showed that relative diameter ($\frac{d}{\bar{d}_{100}}$) and basal area in trees larger than the subject tree ($G_{>d}$) were both important in the prediction of height increment. These variables are closely related to the *Dbh* to height ratio as measures of the position of the tree within the diameter distribution.

No clear results were forthcoming from this analysis and this approach was not pursued further.

Height Increment from *Dbh* and Stand Parameters

Equation 2 was used as a basis for prediction of tree height. By rearranging this equation a new variable was generated:

$$\frac{\ln\left(\frac{h}{\bar{h}_{100}}\right)}{\ln\left(\frac{d}{\bar{d}_{100}}\right)}$$

and added to the data set. This variable can be thought of as the curvature of the line relating relative diameter to relative height. Missing values were introduced for those trees of *Dbh* equal to the stand mean top diameter. Again, "all-regression" procedures were used to find potential predictor variables. Only weak correlations were found but the variables that were consistently selected were altitude, stand relative spacing and $G_{>d}$. When analysed by growth modelling region the results were less satisfactory. It appeared that only a minor gain was made by including these variables as well as a constant term so equation 2 was refitted by region with only a constant term, that is $f()$ was replaced by β_0. An analysis of the residual sums of squares was made to test the hypothesis that one model could be used for all regions (Table 3). The reduction in the residual by including the growth modelling region was significant which indicates that a single model for all regions cannot be used.

Table 3. Analysis of Variance of Residual Sums of Squares for Equation 2.

Source	d.f.	Sum of Squares	Mean Square
Residual about Hypothesis	13439	37419.46	2.784393
Residual about Maximum			
Canty	300	556.08	
Clays	830	3249.97	
Gdns	2537	8361.01	
Hbay	1364	4035.33	
Kang	4448	10591.68	
Sands	1958	4983.84	
South	1996	4743.37	
Combined	13433	36521.28	2.718773
Differenc e	6	898.18	149.6967

$$F = \frac{\text{Difference Mean Square}}{\text{Max. Model R.M.S.}} = 55.06037$$

Rather than exclude the effect of altitude, stand relative spacing and $G_{>d}$ by simply using regional models with one constant term each, a composite fit was then attempted by using boolean variables to represent the regions in the constant term.

At this stage the data set was divided at random into two sets. To avoid unbalancing the first set, it included only 25% of the Central North Island data (Kang) and 40% of the Nelson data (Gdns) while the other region's data sets where divided approximately in two. The first set was used to estimate parameters while the second set was used for independent tests.

The composite model for all regions is:

$$h = \bar{h}_{100}\left(\frac{d}{\bar{d}_{100}}\right)^{f()}$$

where

$$f() = \beta_R + \beta_1 \frac{1}{\bar{h}_{100}\sqrt{N}} + \beta_2 G_{>d} + \beta_3 altitude^2$$

Equation 3.

Table 4. Coefficients and their Standard Errors

	Coefficient	Standard Error
B_R		
CANTY	0.2555	0.0349
CLAYS	0.1901	0.0301
GDNS	0.3316	0.0275
HBAY	0.1992	0.0318
KANG	0.2308	0.0260
SANDS	0.2192	0.0307
SOUTH	0.2567	0.0309
B_1	21.47	10.05
B_2	8.842E-4	3.760E-4
B_3	-2.131E-7	0.549E-7

All coefficients (table 4) are significant. Residuals were examined, by region, for trends against predicted values, the predictor variables and other variables in the data set but none were found.

Comparison of Accuracy of Equation 3 and Petterson Curves

Equation 3 was then applied to the independent data set and the errors in the height prediction were calculated. For comparison, tree heights were also calculated via the equations derived by Lawrence (1990) for predicting the Petterson curve coefficient for stand table generation. The errors in the height prediction are shown in table 5 which lists the standard deviation of the error and the probability of obtaining a mean error of that size under null hypothesis that the mean is zero (no bias). Probability values less than 0.05 indicate a significant bias in the predictions.

Table 5. Summary of errors from Stand table Petterson curves.

Region	number of observations	mean error (m)	standard deviation	P(t); mean=0.0
Canty	154	-0.305	1.314	0.00
Clays	408	-0.137	1.986	0.16
Gdns	1563	0.241	1.810	0.00
Hbay	668	0.054	1.780	0.44
Kang	3410	-0.003	1.724	0.93
Sands	996	0.090	1.613	0.08
South	977	0.112	1.577	0.03
Combined	8176	0.061	1.726	0.00

Significantly biased predictions were made in three regions. Table 6 shows the errors from equation 3.

Table 6. Summary of errors from Composite Equation (3) with Regional constants..

Region	number of observations	mean error (m)	standard deviation	P(t); mean=0.0
Canty	154	-0.009	1.355	0.93
Clays	408	-0.150	1.973	0.13
Gdns	1563	-0.087	1.778	0.05
Hbay	668	0.121	1.737	0.07
Kang	3410	-0.098	1.525	0.00
Sands	996	-0.013	1.606	0.80
South	977	-0.011	1.540	0.83
Combined	8176	-0.058	1.628	0.00

As table 6 shows, equation 3 gave slightly more accurate and precise predictions of the tree heights in the independent data set. Only the central North Island data (Kang) shows a significant bias (table 6) which is due in part to the much larger sample, as the precision is higher than all regions except Canterbury. Significant bias appears in 3 regions in table 5.

This composite model (equation 3) was then tested further by examining the prediction errors for correlations with other variables. No trends in the errors were found.

Discussion

Equation 3 was then examined to check its consistency. Figure 1 shows a series of tree height predictions over age and *Dbh* for an example plot with a long series of measurement on a wide range of height trees. Corresponding graphs for example plots in the other regions are shown in Appendix 1.

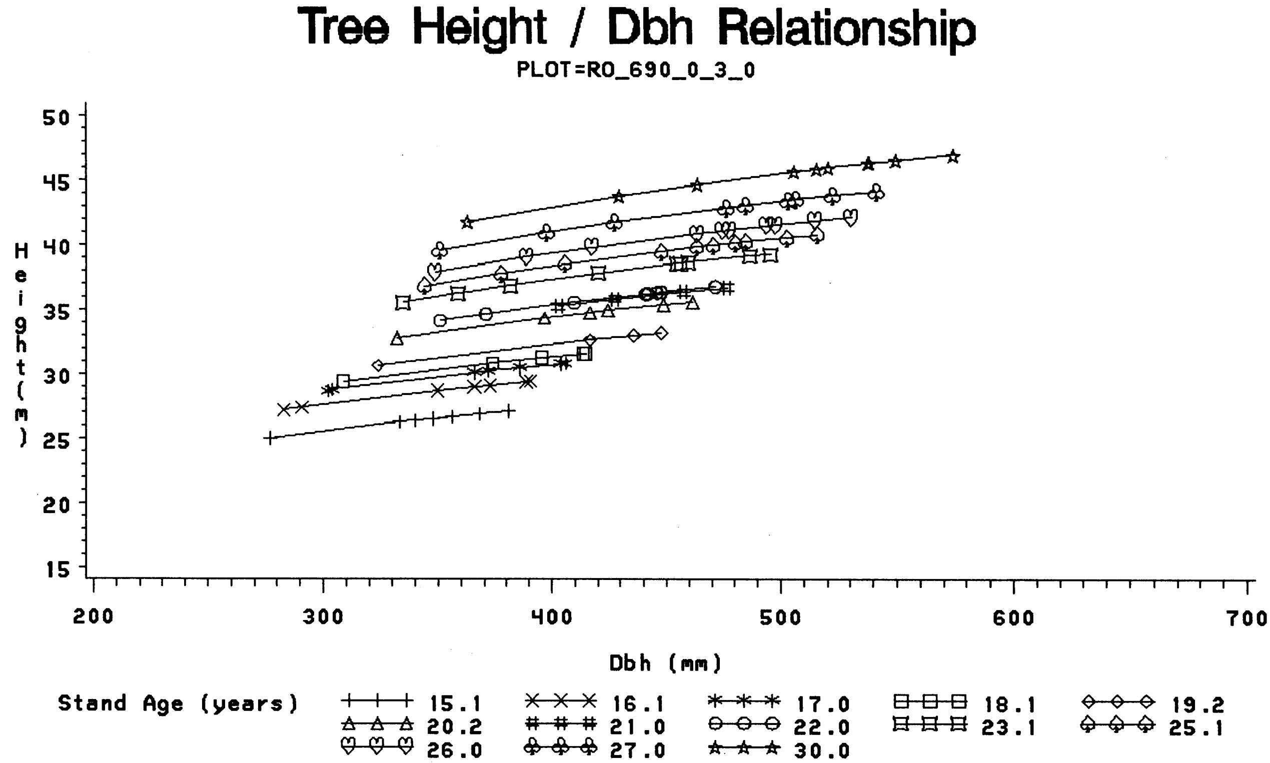

Figure 1. Height Predictions through Time.

Height increment can be calculated from the successive height predictions, which require information on the site (region, altitude), the stand (stocking, mean top height, mean top diameter) and the tree (*Dbh* and position in the *Dbh* distribution), at the start and end of the prediction period. The predicted tree height growth is strongly tied to the stand mean top height growth and will respond directly to changes in the site index and the stand height / age relationship. All trees with *Dbh* equal to mean top diameter will grow at the same rate as mean top height. In figure 1 the plot estimate of mean top height was used in equation 3 to calculate tree height, which is why the curves are not displaced in closer correspondence to the time interval between measurements.

In a given stand, height increment calculated in this way does not alter with initial tree height. This approach is supported by Tanaka (1988), who found no relationship between height increment and height and argued that the amount of apical meristem is relatively constant amongst trees regardless of height. However predictions of height increment will alter with changes in relative diameter.

The regional constants in equation 3 (Table 4) are similar over six of the regions, varying from 0.1901 for Clays to 0.2567 for Southland. The Nelson constant is the largest with a value of 0.3316, which indicates a stronger (steeper) relationship between tree height and diameter in this region.

Equation 3 can be expanded for use as an individual tree height / age curve by replacing mean top height with its estimate from a stand height / age curve, that is, a function of age and site index. If the other stand and tree predictor variables are also calculated as functions of age, a height / age curve for an individual tree is obtained.

Altitude has a clear effect on the height diameter relationship as shown in figure 2.

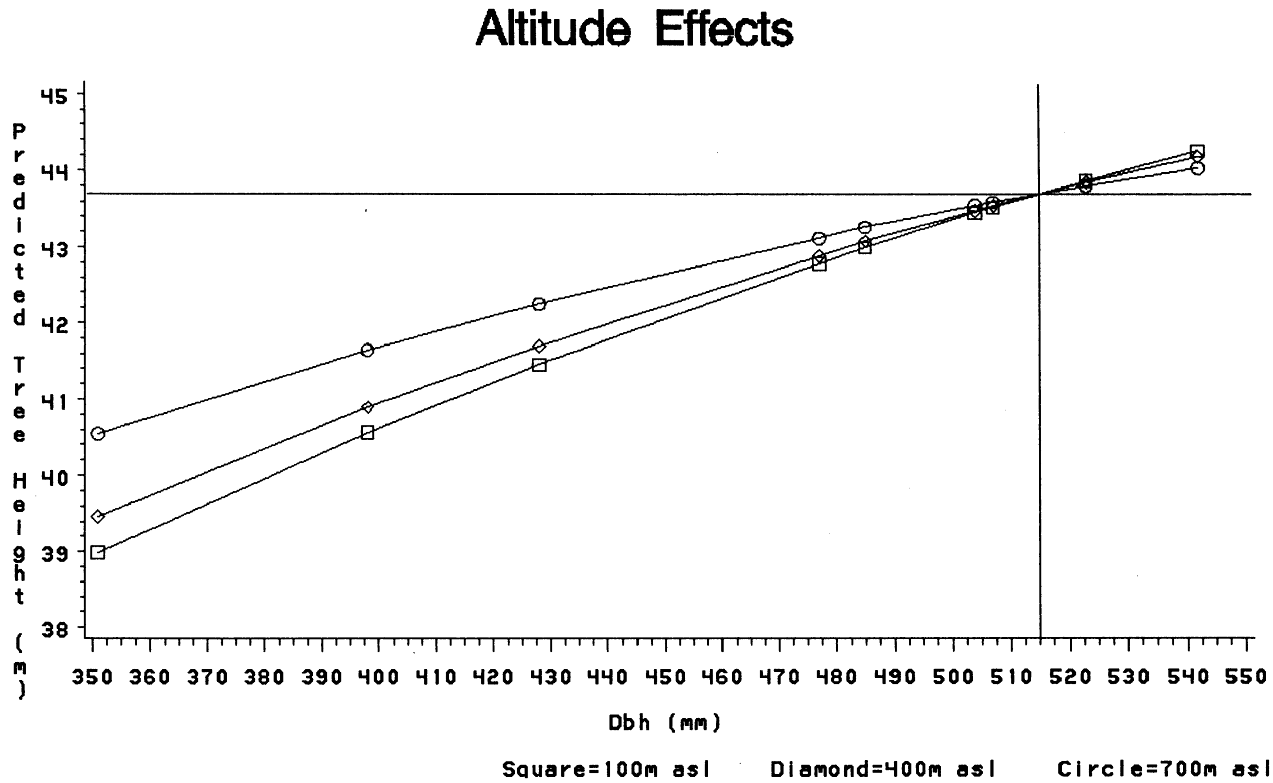

Figure 2. Changes in Height Diameter relationship with Altitude

Increasing altitude tends to weaken the relationship and reduce the variation in tree height within a stand. Figure 2 shows the predicted height / diameter relationship based on plot RO 690 0 3 0 at age 27. The curves were calculated by altering only the altitude while keeping all other variables constant. As the mean top height and diameter have not changed all the curves pass through this point. The effect is not linear as can be seen in the smaller trees, where the change in the predicted heights gets larger even though the change in altitude is the same between the three curves.

Spacing (stand density) exerts a small effect on the height / diameter relationship. Figure 3 shows the same plot but at different stockings: 150, 300 and 600 stems per hectare.

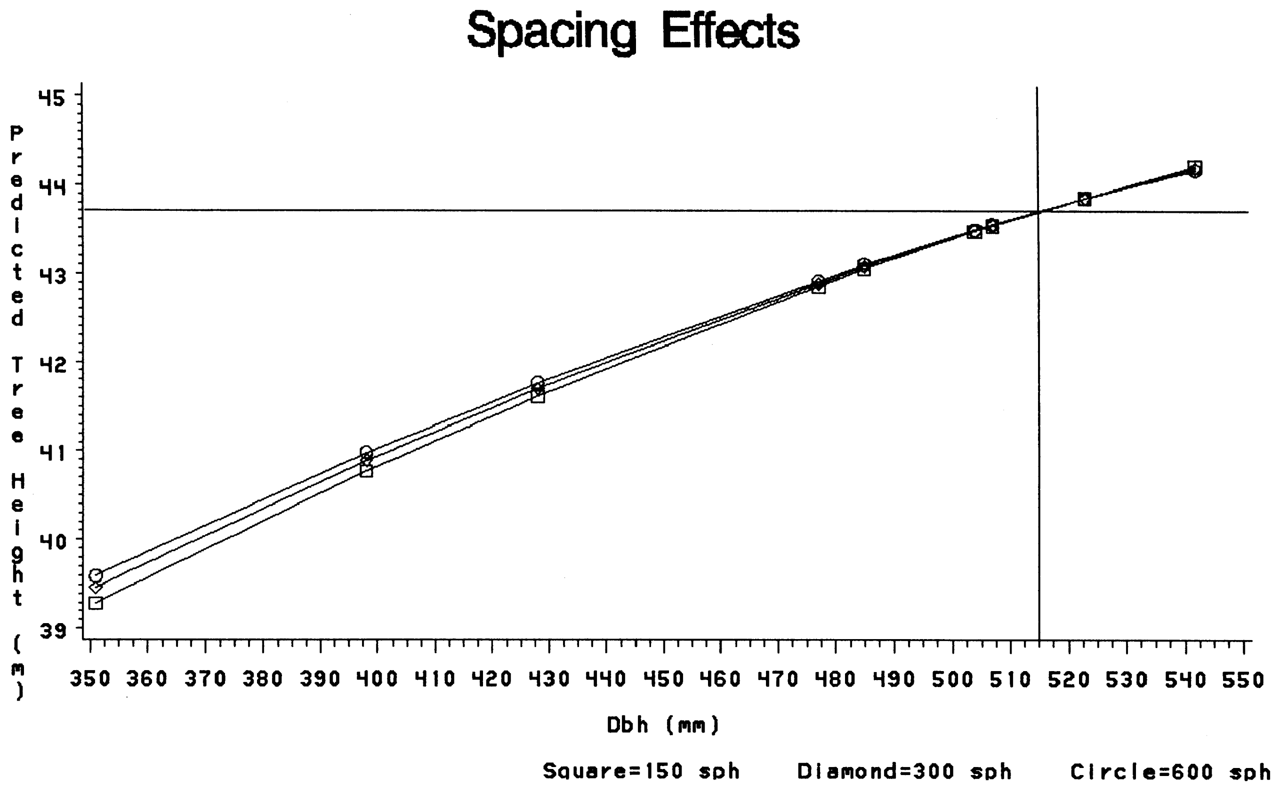

Figure 3. Spacing effects on the Height Diameter relationship

This is a small effect which tends to weaken the height diameter relationship as stocking increases.

The shape of the *Dbh* distribution has an effect through the $G_{>d}$ variable, which represents the cumulative tree basal area distribution. As the variation in the *Dbh* distribution increases so this variable changes more slowly from the smallest to the largest tree. Plot RO 690 0 3 0 was used to examine the effect of this variable by artificially changing the shape of the cumulative distribution of tree basal area. Two extreme and one intermediate distributions were constructed. The trees in the plot were then ranked on *Dbh* and these ranks were used to determine the proportion of the total plot basal area from the distributions that was assigned to the $G_{>d}$ variable. The intermediate distribution was similar to the original plot distribution.

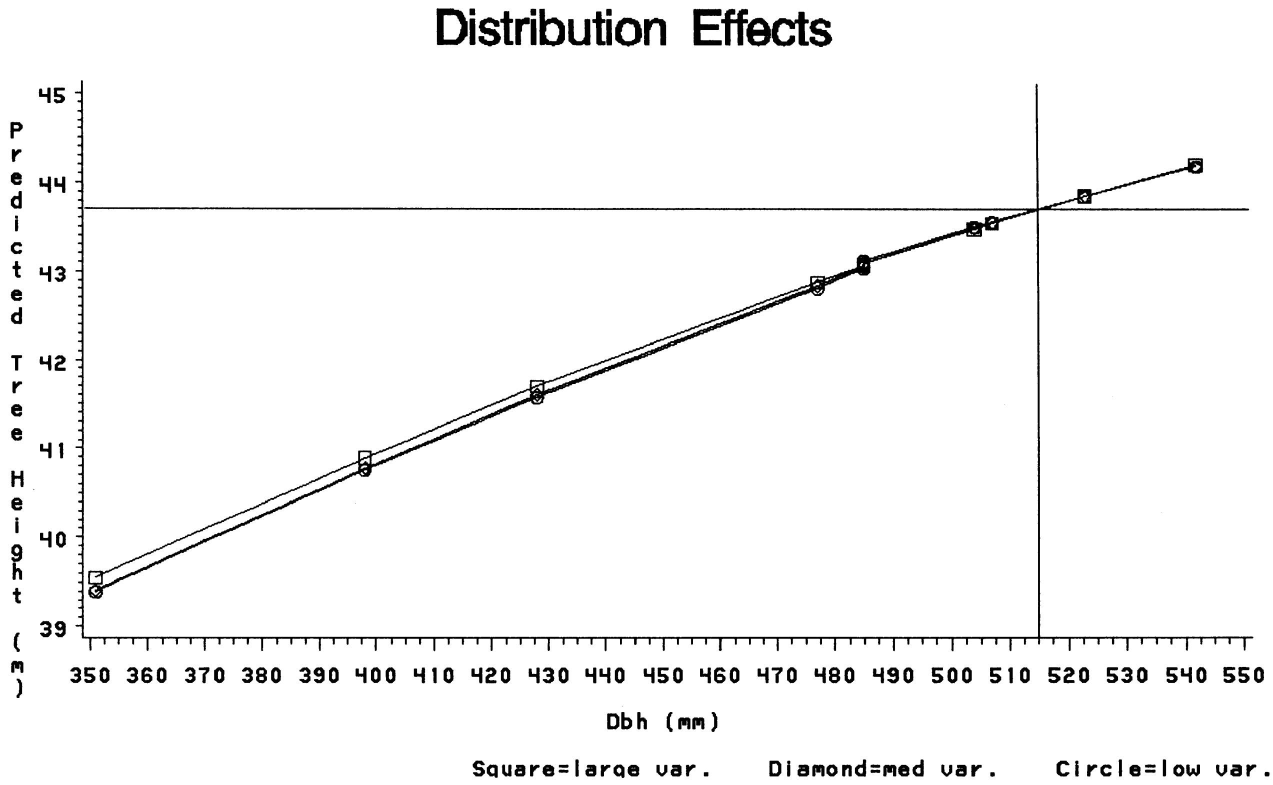

Figure 4. Effect of the shape of the *Dbh* distribution on the Height Diameter relationship

Figure 4 shows the height diameter relationship for this plot under the three different distributions. The effect is quite minor and shows that increasing the variation in the *Dbh* distribution leads to a slight decrease in the slope of the height diameter relationship. That is, a small tree with less stand basal area above it will be slightly taller than one of the same size but in a more suppressed condition.

Summary

Equation 3 can be used to predict the height growth of individual trees as is required when projecting inventory data. This equation requires information on the site (region, altitude), the stand (stocking, mean top height, mean top diameter) and the tree (*Dbh* and position in the *Dbh* distribution), at the start and end of the prediction period, so it must used in growth prediction algorithms after the diameter increment has been calculated.

Because equation 3 is tightly connected with stand mean top height it should be possible to implement it in conjunction with stand-level growth models, so that the individual tree height increments will respond directly to changes in the site index and the stand height / age relationship.

References

CAO, Q.V. 1993: Estimating coefficients of base-age-invariant site index equations. Canadian Journal of Forest Research, 23: 2343-2347.

GORDON, A.D. and LAWRENCE, M.E. 1992: A Comparison of Methods to Predict Individual Tree Diameter Growth. FRI/Industry Stand Growth Modelling Research Cooperative. Report No. 30.

GORDON, A.D. and LAWRENCE, M.E. 1994: Projecting Inventory Data: Predicting Individual Tree Diameter Growth. FRI/Industry Stand Growth Modelling Research Cooperative. Report No. 34.

HUNTER, I.R., RODGERS, B.E., DUNNINGHAM, A., PRINCE, J.M., THORN, A.J. 1991: An Atlas of Radiata Pine Nutrition in New Zealand. Forest Research Institute, Ministry of Forestry, FRI Bulletin No 165.

LAWRENCE, M. E. 1990: Diameter distributions for the regional stand growth models. . FRI/Industry Stand Growth Modelling Research Cooperative. Report No. 13. February 1990.

McEWEN, A.D. 1978: New Zealand Forest Service Computer System for Permanent Sample Plots. Pp 235-52 **in** Elliot, D.A. (Ed.) Mensuration for Management Planning of Exotic Forest Plantations. New Zealand Forest Service, Forest Research Institute Symposium No. 20.

PILAAR, C.H. and DUNLOP, J.D. 1989: The Permanent Sample Plot system of the New Zealand Ministry of Forestry. Proceedings of the IUFRO Conference: Forest Growth Data: Capture, Retrieval and Dissemination. April 3-5, 1989, Faculty of Agriculture, Gembloux, Belgium.

TANAKA, K. 1988: A stochastic model of height growth in an even-aged pure forest stand. Journal of the Japanese Forestry society 70(1):20-9.

Appendix 1

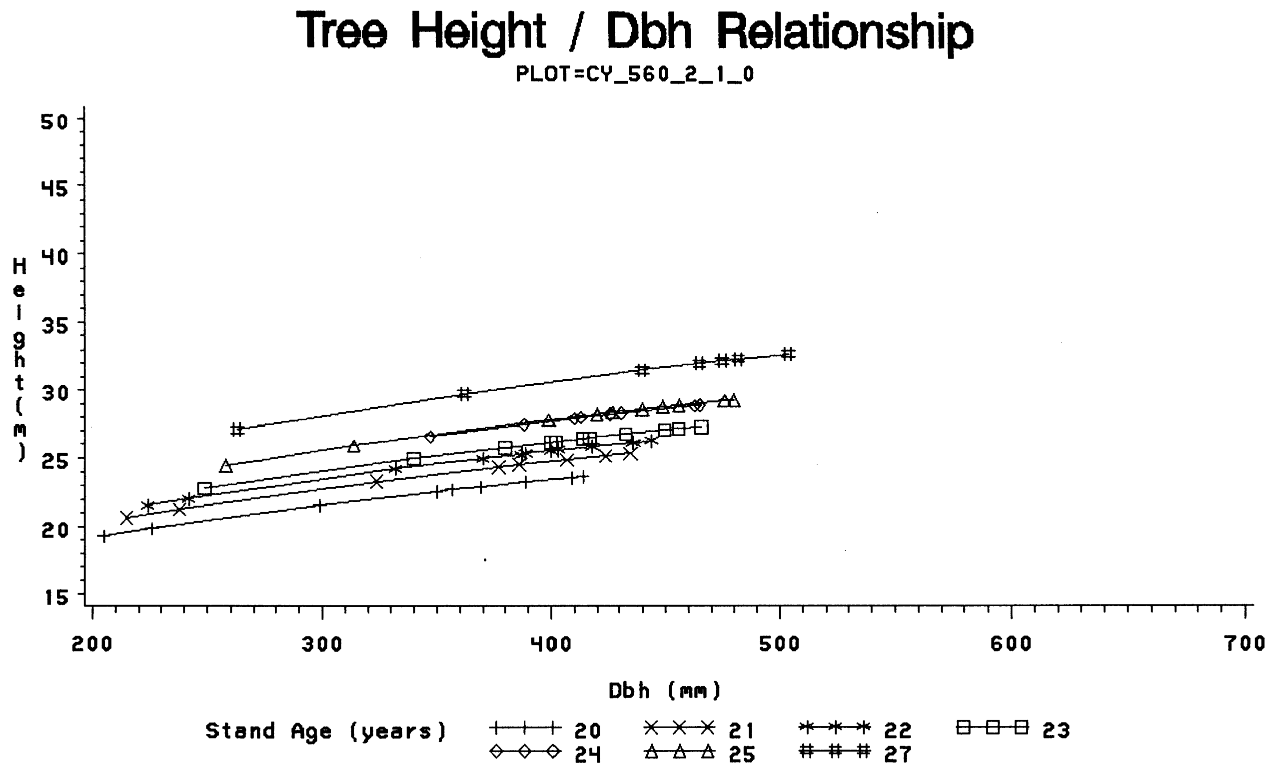

Figure 1. Example from Canterbury region

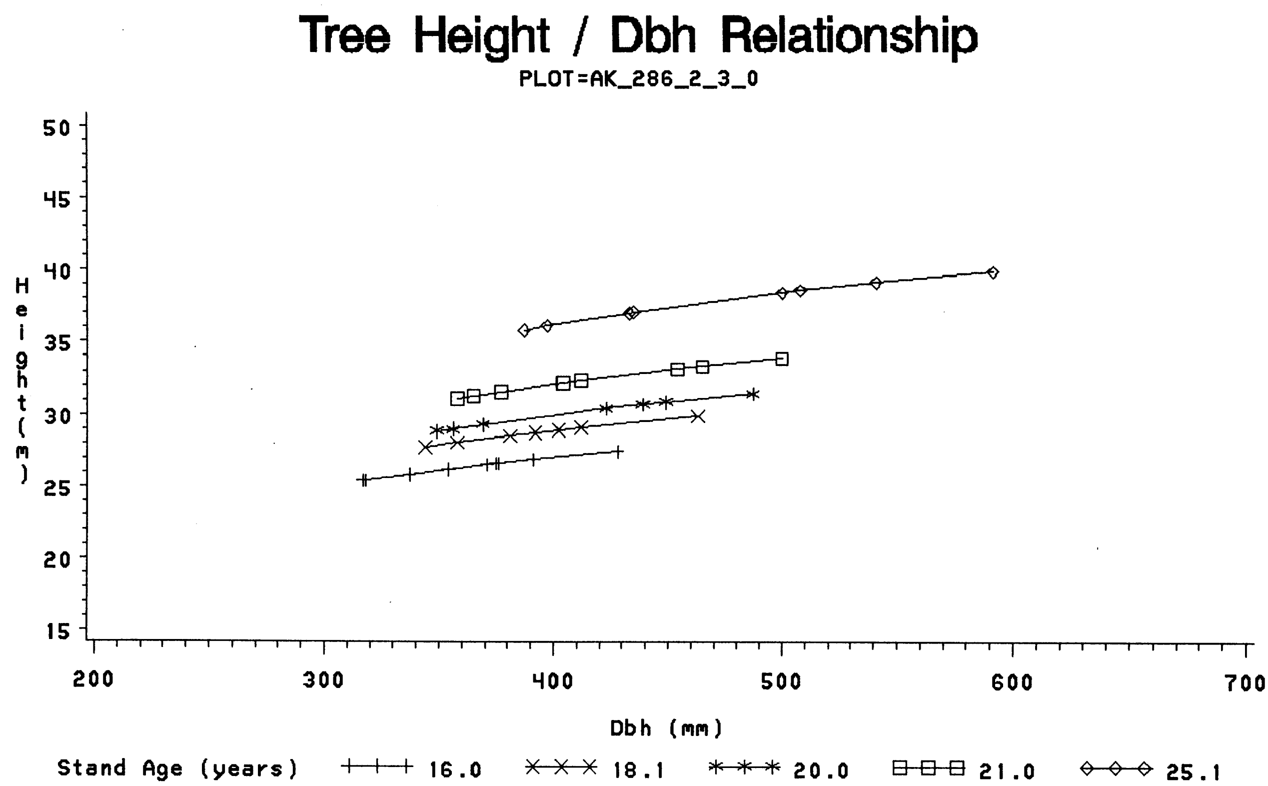

Figure 2. Example from Clays region

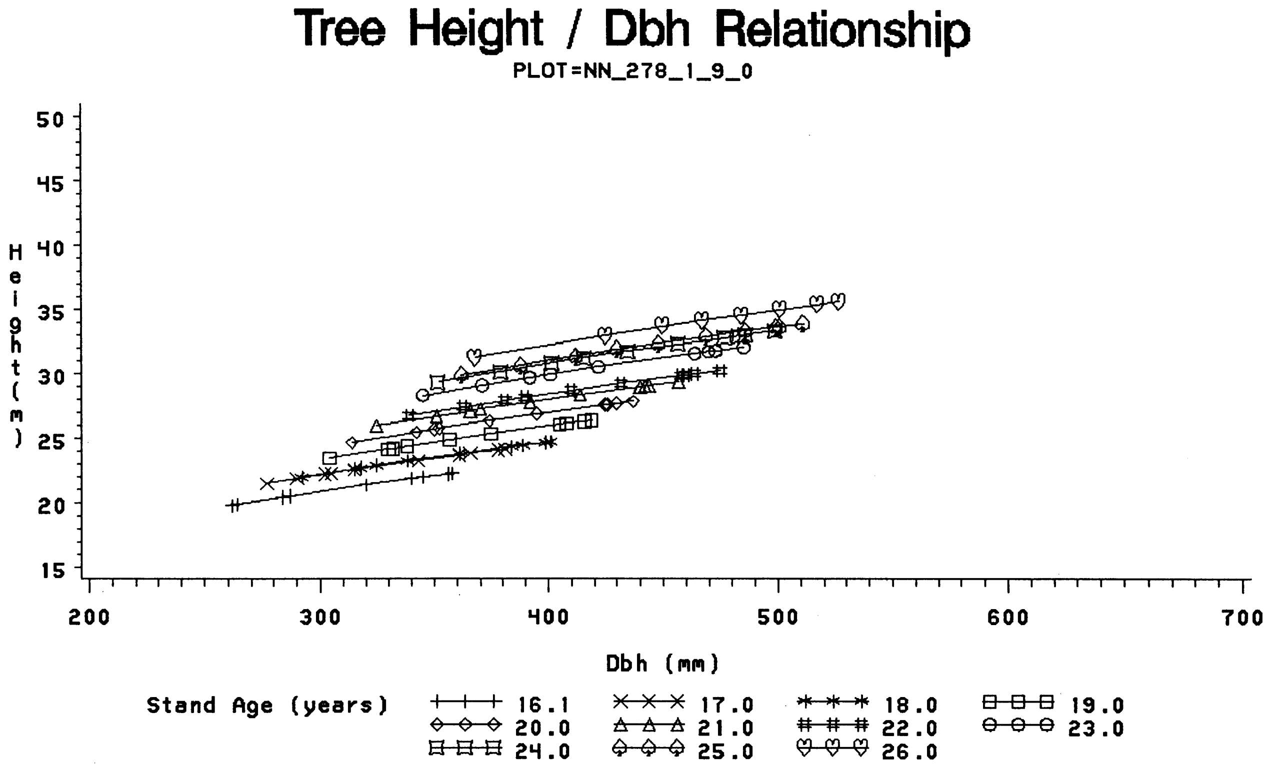

Figure 3. Example from Nelson region

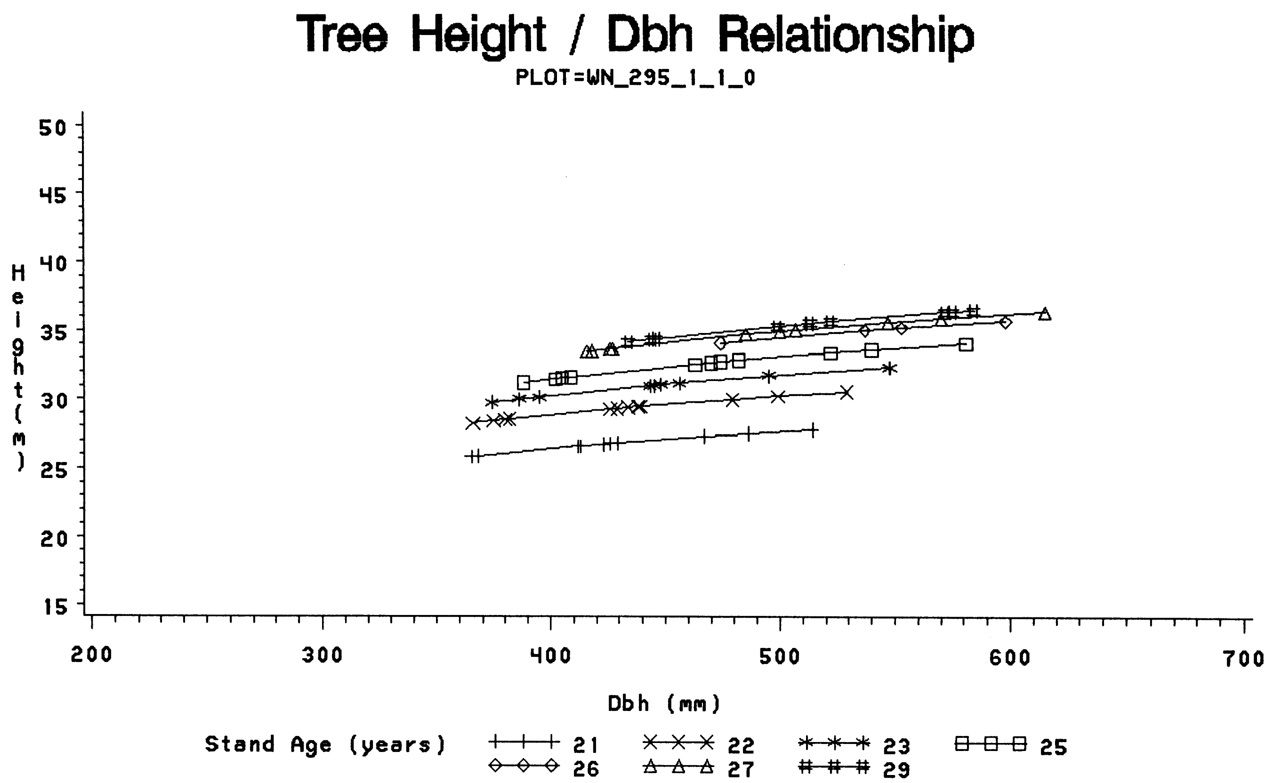

Figure 4. Example from Hawkes Bay region

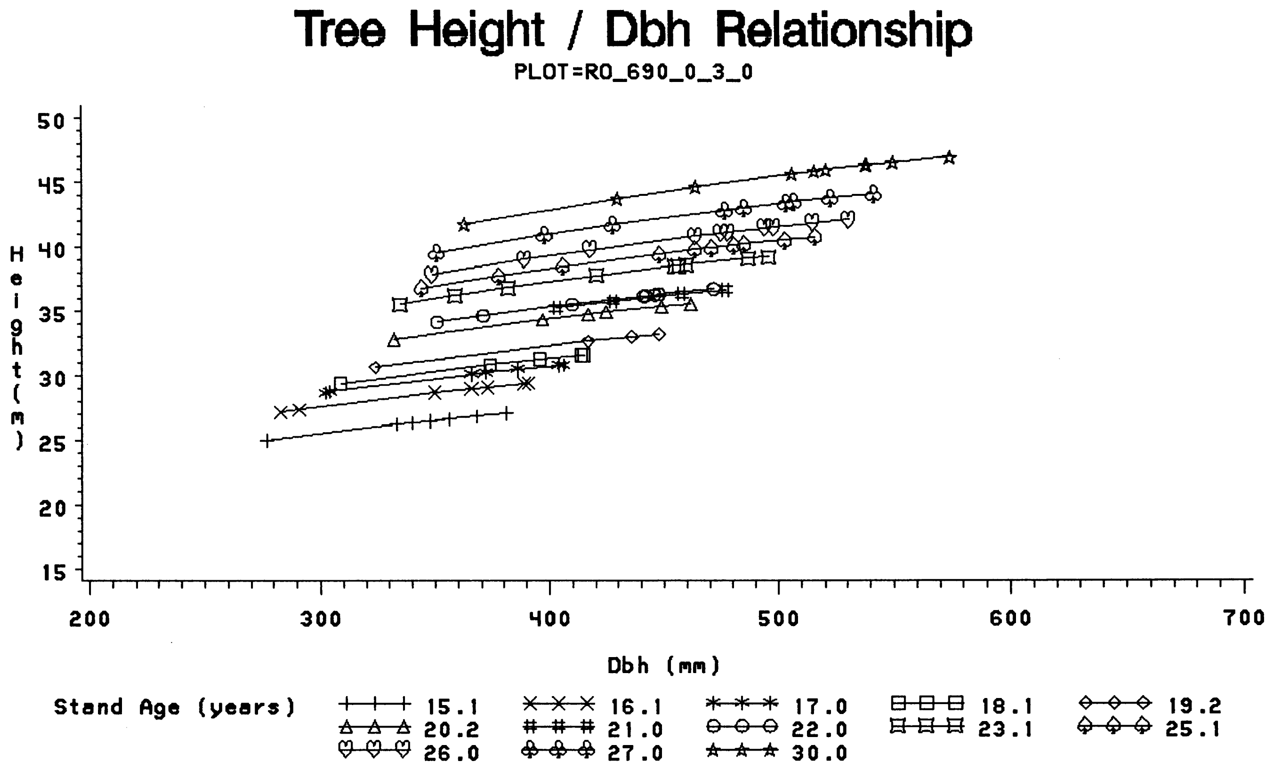

Figure 5. Example from Central North Island region

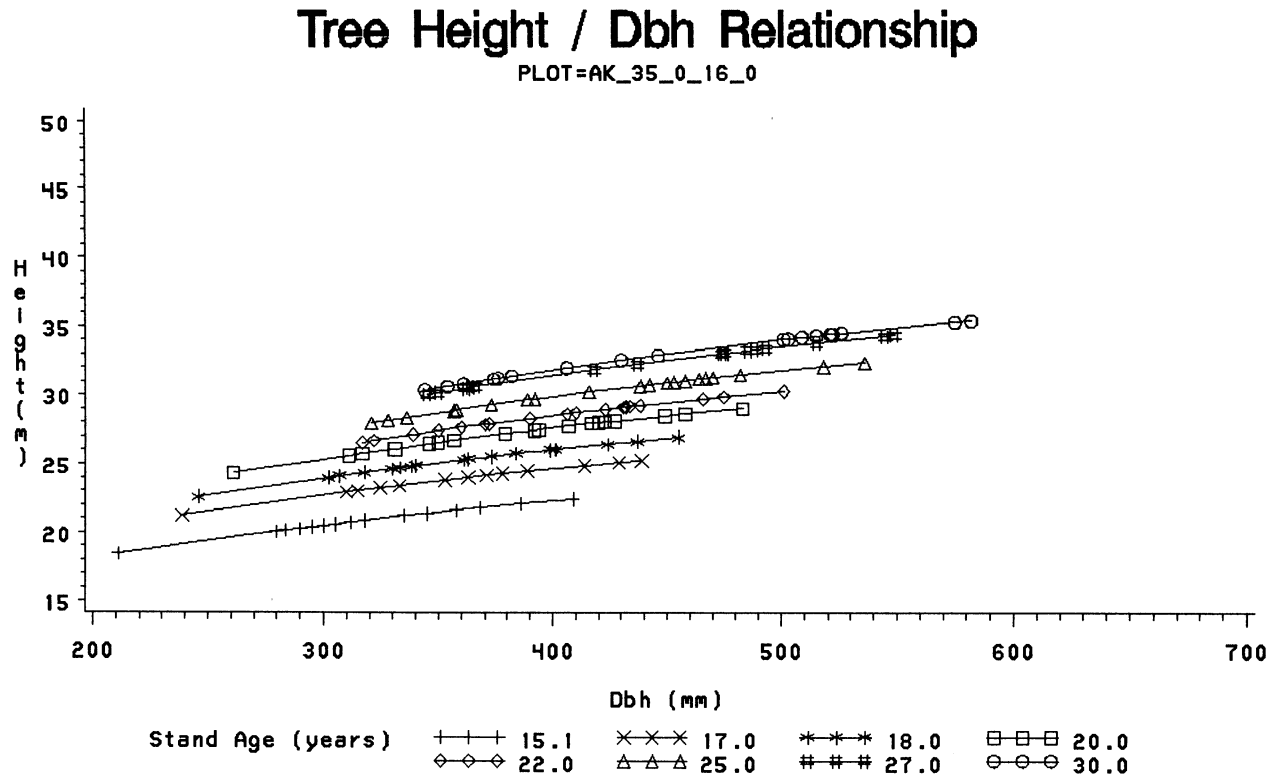

Figure 6. Example from Sands region

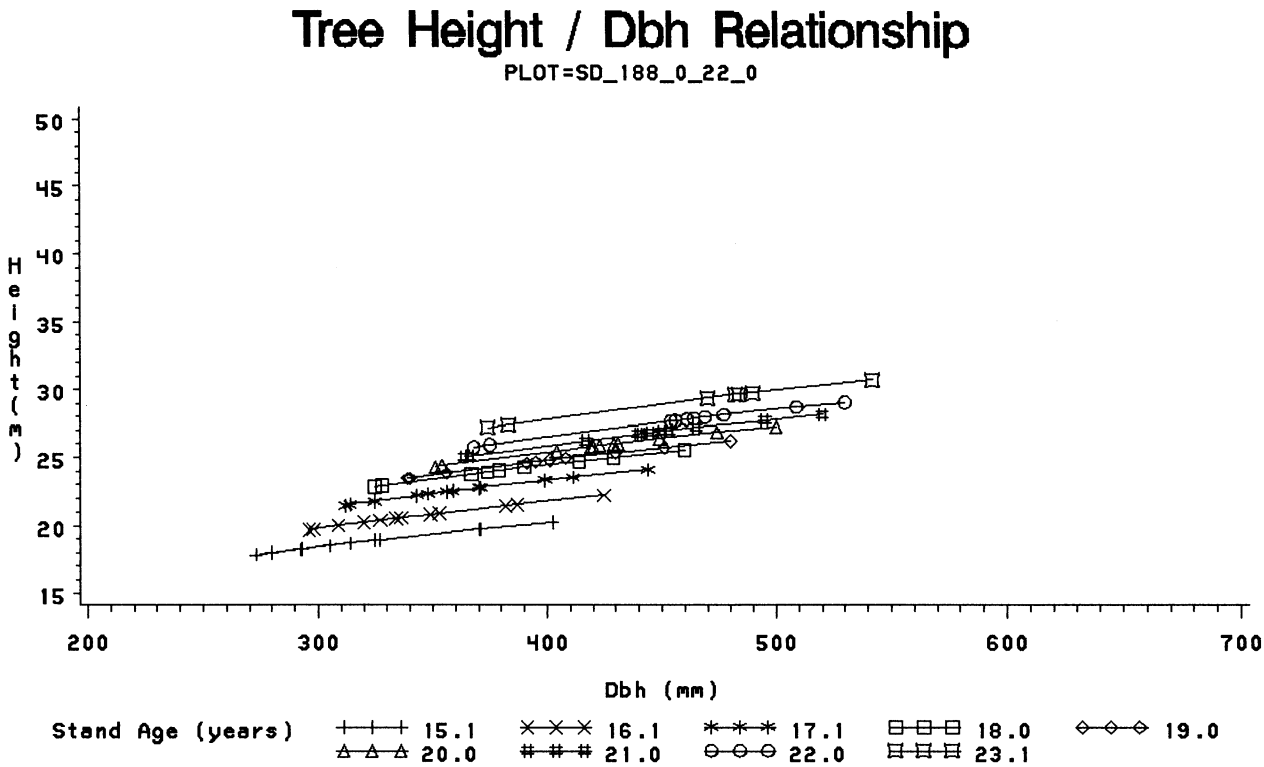

Figure 7. Example from Southland region