forest industry research cooperatives

EUCALYPT COOPERATIVE

Report No. 1

Stand-level Growth Models for *Eucalyptus nitens* Plantations in New Zealand

by
Steven G. Candy
Forestry Tasmania

Date: June 2003

SCION
Next generation biomaterials

Stand-level Growth Models for *Eucalyptus nitens* Plantations in New Zealand

S. G. Candy

Forestry Tasmania

Eucalypt Cooperative

Report No. 1 **June 2003**

Stand-level Growth Models for *Eucalyptus nitens* plantations in New Zealand

July – 2002

Steven G. Candy
Forestry Tasmania
Hobart, Tasmania
Email: Steve.Candy@forestry.tas.gov.au

Summary

A suite of empirical stand-level growth models has been constructed using data from permanent sample plots from trials at 42 sites in the north and south islands of New Zealand. Stand models of predominant height (PDH), basal area (BA), stocking (*N*) and volume (*V*) are given. Site index (*S*) can be predicted from the PDH model and is defined as PDH at age 15. The stand basal area model is a modification of Clutter's projection model that includes effects due to age and intensity of thinning. The site index term in the asymptote of the BA projection model was found to be negative and statistically significant. When altitude and latitude terms were included both these terms were non-significant while the *S* term was significant and negative. Since such a negative site index term is difficult to explain it was dropped from the BA model but further work, possibly employing a site-productivity measure predicted from physiologically-based models such as 3PG to replace or augment the site index term, is recommended.

A model which predicts stand basal area at age 10 for unthinned stands with stocking between 850 and 1100 sph is also evaluated in order to facilitate simulation of hypothetical stands given only site index. Stands can be grown forwards or backwards from age 10 and various thinning regimes applied if required. When this suite of models is combined with a stem taper model, yield by product type can be predicted by 'growing-on' inventory or experimental plots. Model fitting used mixed model methodology and included a nonlinear mixed model fitting procedure to fit the stand basal area projection model.

Introduction

The suite of models described here are empirical models developed from data obtained from permanent sample plots (PSPs) provided by the Management of Eucalypts Cooperative managed by the Forest Research Institute. The models have been described previously in Candy (1997) and with slight modifications the same stand-level models are used here. Measurements up to and including those carried out in 2000 were used to fit the models.

Data

Data Summaries

From the 42 trials distributed across both North and South islands, 266 PSPs were measured. Site Index (*S*) was estimated from the PDH model as described later using the measurement of the plot closest to the index age of 15 yrs and assigned that estimated value for all measurements of the plot. If the plot was measured at age 15 then the *S* value is estimated exactly as the PDH at that measurement. Due to missing values of PDH, *S* was calculated for 242 of the 266 PSPs. The term Mean Dominant Height (MDH) is used here synonymously with PDH, even though in Candy (1997) it represents the mean of the tallest 50 stems per hectare (sph) rather than the tallest 100 sph as is the definition for PDH. Summaries of the data are given in Tables 1 and 2 below.

Table 1. Stand data summary

Statistic	M'mnts[1] Per plot	Age (yr)	Site Index (SI) (m)	CAI[2] for Volume ($m^3ha^{-1}yr^{-1}$)	Altitude (masl)	Latitude (deg)
Mean	3.31	8.1	30.4	27.4	281	40.61
Minimum	0	1.8	15.9	1.0	80	35.20
Maximum	10	24.0	38.8	87.0	700	46.10
L-quartile	1	6.0	26.9	16.0	150	38.10
U-quartile	5	9.1	33.9	36.8	450	45.30

[1] Suitable measurements (excluding missing values of PDH).
[2] Current annual increment.

Table 2. Thinning measurements data summary

Statistic	Age at Thinning (yr)	Stems felled (sph)	Stems retained (sph)	Percentage of BA removed (m^2/ha)
Mean	6.26	471.6	532.7	35.4
Minimum	2.30	8.0	50.0	1.8
Maximum	13.60	1932.0	1477.0	90.3
L-quartile	4.10	100.0	150.0	22.1
U-quartile	7.00	700.0	800.0	44.0

Figure 1 shows a histogram of age of measurement excluding plots with missing site index and measurements with missing PDH while Figure 2 shows a histogram of site index values.

Thinning Data

There were 167 plot-measurements made at a thinning. Of the 116 plots that were thinned, 86 were thinned once, 9 were thinned twice, and 21 were thinned three times. Table 2 summarises attributes of the 167 plot-measurements.

Statistical Methods

The statistical methods described in Candy (1997) were used with a few minor changes that are described later.

MDH Increment/Site Index Model

The state space (Garcia, 1994) or projection form of the 3-parameter Richards model was used to model MDH (=*H*)

$$H_2 = A\left[1-\left\{1-\left(\frac{H_1}{A}\right)^{1/\alpha}\right\}^{\frac{T_2}{T_1}}\right]^{\alpha} \tag{1}$$

where T_1 is the age at the start and T_2 the age at the end of the projection period, H_1 is the MDH at the start and H_2 the MDH at the end of the projection period, A is the parameter representing the asymptotic MDH and α is a shape parameter.

The equation for obtaining site index given an MDH of H at age T is obtained by substituting H for H_1, T for T_1, 15 for T_2 in (1) with H_2 then corresponding to site index (base age 15). Figure 2 shows the distribution of plot site index where MDH at age 15 is used directly if an age 15 measurement of the plot was taken or an estimate was obtained using the measurement age closest to 15 and model (1) with parameter estimates obtained below.

Model (1) was fitted using nonlinear least squares using GENSTAT's (Genstat 5 Committee, 1997a,b) FITNONLINEAR directive. The data was first arranged into measurement pairs $\{H_2, H_1\}$ at ages $\{T_2, T_1\}$ with H_2 the response variable that represents MDHs at the second measurement through to the last measurement on each plot. H_1 represents the MDH at the start of each projection period at age T_1. For a more rigorous notation we could replace the index 2 by k+1 and the index 1 by k where k=1,...,m-1 and m is the number of measurements on the plot (Candy, 1989) but the simpler notation given by the measurement pairs above will be used throughout. This approach to fitting corresponds to the state-space (Garcia, 1994) or projection form of the model (Clutter *et al*., 1983). When model (1) was fitted using ordinary least squares (OLS) the asymptote $\ln(A)$ was estimated at 3.7222 (s.e.=0.0254) giving $\hat{A}$=41.4 and the shape parameter estimate was $\hat{\alpha}$=1.2668 (s.e.=0.0273). This model accounted for 98.0% of the variance in PDH with a residual mean square (RMS) obtained from the OLS fit of 0.9649. The estimate of the asymptote is clearly unrealistically low since the tallest measured PDH was 39.7 m. Figure 3 shows ordinary residuals (Res1) from the fit of model (1) to H_2 versus fitted value ($\hat{H}_2$).

When the asymptote was fixed at 60m, as in Candy (1997), a shape parameter estimate of 1.0644 was obtained with standard error 0.0153. This model accounted for 97.8% of the variance in PDH with residual mean square (RMS) of 1.047. Figure 4 shows ordinary residuals (Res2) from the fit of model (1) to H_2 versus fitted value ($\hat{H}_2$).

In both Figs. 3 and 4 the regression line fitted to the residuals is shown. In both cases the regression is significant, implying some bias in the predictions, but with the regression explaining only 1.9% of the variance in the Res1 residuals the bias in this case is relatively small. For the Res2 residuals the bias is more concerning with the regression explaining 14.4% of the variance and the regression line in Figure 4 shows a more serious departure of the constant-zero line than in Figure 3. Further, to investigate the trends in both the Res1 and Res2 residuals these were plotted against the altitude and latitude for each of the PSPs. The regression of the residuals against the combined linear regression of altitude and latitude as a linear mixed model with random plot effects was significant in each case, as judged by Wald tests (Genstat 5 Committee, 1997a,b) but only explained a small proportion of the variance (5 and 7% for Res1 and Res2 respectively).

To compare model (1) with the anamorphic projection form of the Richard's model given by eqn (7) of Candy (1989) this last model was fitted again using OLS. The parameter estimates, for this model were $\hat{\alpha}$=0.04876 (s.e.=0.00577) and $\hat{\beta}$=1.0173 (s.e.=0.0242) with RMS of 1.210. There was little difference between predictions from the polymorphic model (1) with $\hat{A} = 41.4$ and those from the 2-parameter anamorphic projection model for the range of ages of interest. For the remainder the former model was used to estimate site index and given preference over the version of model (1) with A fixed at 60 m due to the bias in the latter as discussed above. However, model (1) with $\hat{A} = 41.4$ gives site index curves that differ quite dramatically from those in Candy (1997) particularly for ages above 15 years (Fig. 5) due to the very different values for the asymptote.

Site index is required as a predictor variable in some of the models described below. An estimate of site index can be obtained at each measurement of each plot in the dataset. However, site index would then vary with age, which contradicts the intent that site index be a measure of site productivity. So, for the remainder site index, S, was estimated using the measurement age closest to the index age of 15, and this estimate was used for all measurement periods for the particular plot to calibrate the models which rely on site index as input.

Stand basal area projection model

The basic form of the model used (i.e. excluding thinning effects) is given by eqns (4.42) and (4.43) in Clutter *et al*. (1983). The stand basal area, B, instantaneous growth rate function is

$$\frac{dB}{dT} = T^{-1}B\left[\alpha_0 + \alpha_1 S - \alpha_2 \ln(B)\right]$$

where α_0, α_1, α_2 are parameters to be estimated and *S* is site index defined earlier. The projection form of the model, which is at this stage a state-space model, is given by

$$\ln(B_2) = \left(\frac{T_1}{T_2}\right)^{\alpha_2} \ln(B_1) + \left(\frac{\alpha_0}{\alpha_2} + \frac{\alpha_1}{\alpha_2} S\right)\left\{1 - \left(\frac{T_1}{T_2}\right)^{\alpha_2}\right\} \tag{2}$$

where (B_2, B_1) are the stand basal areas giving response and conditioning variable defined in the same way as those given for MDH. Stand basal area B_2 is net basal area, that is, it excludes trees that have died in the projection period.

Model (2) is consistent (i.e. $B_2 = B_1$ if $T_2 = T_1$) and path invariant. The property of path invariance ensures that the same predicted value of *B* is obtained irrespective of the number of intermediate ages at which *B* is predicted (Clutter *et al*., 1983) and is a function of the state-space definition of (2). However, this property needs qualification when a thinning has been carried out or is simulated. If a thinning occurs at age T_t then for projections of B_2 after thinning, B_1 in (2) should be set to the residual basal area at the thinning age (or initialised at a later age if inventory is at age $T_1 \geq T_t$). Model (2) has two less desirable features. First, the growth rate (1) is not defined at age zero, as a result a starting basal area is required in (2) to obtain a yield curve for a given site index. The yield curve corresponding to (2) is given by

$$B = \exp\left(\frac{\alpha_0}{\alpha_2} + \frac{\alpha_1}{\alpha_2} S + \beta T^{-\alpha_2}\right)$$

with the projection model (2) derived from the yield equation by solving for β in terms of initial age, T_1, and basal area, B_1. So starting values of basal area and age as well as site index are required to determine the yield curve. Second, if $\ln(B_1) > \frac{\alpha_0}{\alpha_2} + \frac{\alpha_1}{\alpha_2} S$ then the relative growth, $\ln(B_2) - \ln(B_1)$, is negative. Since the right hand side of the above inequality is the logarithm of the asymptotic basal area, then theoretically this inequality should never be satisfied. However, the asymptote must be estimated and it is possible that a measured basal area is greater than the estimated asymptote but this is unlikely to occur in practice. Candy (1989, 1997) obtained a projection model from the same instantaneous growth rate function by solving for the asymptote rather than β and then modelling β and α_2 parameters as functions of site index and other stand variables. The above inequality can never occur with this 'anamorphic' projection model.

The above properties of model (2) do not in general limit its utility as discussed later.

Thinning effects

The simplest way to handle the effect of thinning is to assume that there is no effect of thinning on relative growth [i.e. 'grow on' the residual basal area after thinning using (2)]. Alternatively, thinning may affect the relative growth compared to an unthinned stand of the same initial basal area, initial age, and site index. Commonly, thinning effects are incorporated in basal area projection models using combinations of age, intensity, and nature of thinning (Bailey and Ware, 1983; Pienaar and Shiver, 1986; Candy, 1989). Bailey and Ware (1983) incorporate thinning effects in model (2) with the constraint that $\alpha_2 = 1$. The yield model in this case is a simple log/reciprocal model which is often not flexible enough to model yield or increment which is the reason Clutter *et al*. (1983) introduced the extra parameter, α_2. Here thinning was incorporated in (2), while maintaining path invariance, as follows.

Let 'linear age components' be

$$\eta_1 = T_1 + \alpha_3 P_t + \alpha_4 P_t T_t + \alpha_5 P_t T_t S$$
$$\eta_2 = T_2 + \alpha_3 P_t + \alpha_4 P_t T_t + \alpha_5 P_t T_t S$$

then

$$\ln(B_2) = \left(\frac{\eta_1}{\eta_2}\right)^{\alpha_2} \ln(B_1) + \left(\frac{\alpha_0}{\alpha_2} + \frac{\alpha_1}{\alpha_2} S\right)\left\{1 - \left(\frac{\eta_1}{\eta_2}\right)^{\alpha_2}\right\} \qquad (3)$$

where

P_t is the proportion of basal area removed in thinning at age T_t. For unthinned stands or ages prior to thinning in thinned stands P_t and T_t are each set to zero. After thinning, P_t and T_t remain fixed at their thinning-age values for all later projection periods unless another thinning is applied in which case they are changed to this later thinning's values (note that the data used here involved at most one thinning). As described earlier, the age at thinning should start a new projection period with $T_1 = T_t$ and B_1 set to the retained basal area. It can be seen that these thinning terms affect growth by modifying the age terms in (2). Depending on the age and intensity of thinning and the sign and magnitude of the parameters $\alpha_3, \alpha_4, \alpha_5$, the 'age shift' may be to an earlier or later age. Since growth rate depends on age, as in (2), then relative growth may be increased or decreased by thinning. Since model (3) is purely an empirical construct, expression of the effect of thinning in relative basal area growth depends largely on the thinning data available. Note that model (3) no longer has a state-space representation since given $T_1 > T_t$ future values of the state variable, stand basal area, do not depend solely on the state variables (i.e. at age T_1).

Model fitting

The response variable in the fit of (3) was $\ln(B_2)$. Since the calibration data involved repeat measurements of sample growth/experimental plots, a mixed model was constructed from (3) by adding random plot effects to the linear age and site index components of the model. At the same time to facilitate the fitting procedure the parameters in the linear site index term, $\frac{\alpha_0}{\alpha_2} + \frac{\alpha_1}{\alpha_2} S$, were reparameterised as $\alpha_0' + \alpha_1' S$. The nonlinear mixed model (NLMM) is then

$$\eta_{1i} = T_{1i} + \alpha_3 P_{ti} + \alpha_4 P_{ti} T_{ti} + \alpha_5 P_{ti} T_{ti} S_i + b_{2i}$$
$$\eta_{2i} = T_{2i} + \alpha_3 P_{ti} + \alpha_4 P_{ti} T_{ti} + \alpha_5 P_{ti} T_{ti} S_i + b_{2i}$$

$$\ln(B_{2i}) = \left(\frac{\eta_{1i}}{\eta_{2i}}\right)^{\alpha_2} \ln(B_{1i}) + \gamma_i \left\{1 - \left(\frac{\eta_{1i}}{\eta_{2i}}\right)^{\alpha_2}\right\} + \varepsilon_{2i} \qquad (4)$$

$$\gamma_i = \alpha_0' + \alpha_1' S_i + b_{1i} \qquad (5)$$

where the $i = 1,\ldots,n$ subscript is introduced to represent the ith plot in a total of n plots and b_{1i} and b_{2i} are random plot effects with n elements each and ε_{2i} is a random error independent of the random effects.

A more conservative iterative weight function to that used by Candy (1997) was employed here. The weight function is

$$w_i = \left\{1 - \left(\frac{\eta_{1i}}{\eta_{2i}}\right)^{\alpha_2}\right\}^{-\kappa}$$

where $\kappa = 2$ was used in Candy (1997) while $\kappa = 1$ was used here since it was noted here and in Candy (1997) that using $\kappa = 2$ over-corrects for the trend of increasing variance of (conditional) residuals with increasing w_i^{-1}.

A further departure from the estimation method of Candy (1997) was to include a covariance parameter for the covariance between b_1 and b_2. Therefore $Cov(b_1, b_2) = \mathbf{D}$ where **D** is the 2 x 2 matrix

$$\mathbf{D} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Therefore

$$\hat{Var}\{\ln(B_{2i})\} = \mathbf{z}_i^T \mathbf{D}\, \mathbf{z}_i + w_i^{-2}\hat{\sigma}^2,$$

where $\mathbf{z}_i^T = (z_{i1}, z_{i2})$, $z_{i1} = \hat{w}_i^{-1}$, and

$$z_{i2} = \hat{\alpha}_2 \left(\frac{\hat{\eta}_{1i}}{\hat{\eta}_{2i}}\right)^{\hat{\alpha}_2 - 1} \frac{\hat{\eta}_{2i} - \hat{\eta}_{1i}}{\hat{\eta}_{2i}^2} \{\ln(B_{1i}) - (\hat{\alpha}_0' + \hat{\alpha}_1' S_i)\}.$$

A problem with the performance of the fitting algorithm of Candy (1997) (as modified above) was encountered here. In Candy (1997, Appendix 1), model (4,5) was fitted using the 'marginal' method or, more specifically, using marginal weighted generalized least squares (WGLS) which is the normal errors equivalent of marginal quasi-likelihood of Breslow and Clayton (1993). The marginal method in effect simply removes the random plot effects from the calculation of the η_1, η_2, γ terms and the working response variable but retains them as linear random effects in the linear mixed model of the working response variable. This is an approximate solution to producing 'population-average' estimates of the parameters $(\alpha_0,....,\alpha_6)$ in model (4,5). However, this approximate method may fail if the random effects are relatively large. Here problems were encountered in obtaining convergence in parameter estimates when using the marginal method. The damping of consecutive iteration's estimates was used to improve convergence properties of the marginal method in Candy (1997, Appendix 1) but this failed here. The subject-specific (i.e. plot-specific) fitting method on the other hand retains the estimated random plot effects in the calculation of the η_1, η_2, γ terms and the working response variable. The algorithm then proceeds as for the marginal method. Although this subject-specific (SS) fitting algorithm performed well in terms of the convergence of parameter estimates it was found that it gave an unrealistic estimate of α_2 (=1.29) when typically $\hat{\alpha}_2$ is less than 0.9 (e.g. Candy, 1997). Apart from this difficulty, which resulted in serious under-prediction of stand basal area at ages above 15, the random effects must be integrated out of SS predictions (Lee and Nelder, 1996) in order to obtain the required PA predictions (Candy, 1997).

For the above reasons ordinary least squares (OLS), as used by Clutter *et al*. (1983), was used to provide estimates of the parameters in model (4,5). However, after screening out eight outliers and fixing α_2 at the OLS estimate of 0.8385, the marginal method of fitting the NLMM converged successfully giving WGLS estimates which were then used as the final parameter estimates.

The data set used in the fit consisted of 872 measurement pairs after excluding missing values for site index, *S*, (due to missing values of PDH), measurement pairs for which $T_1 \le 3$, and the outliers mentioned above.

When model (4,5) was fitted, using the WGLS fitting algorithm the estimate of the parameter for the site index term in model (5) was negative and statistically significant. The estimate of $\hat{\alpha}_1'$ was –0.0536 (s.e.=0.0132) while the OLS fit gave an estimate of -0.0450 (s.e.=0.0141). The negative site index term in (5) indicates that this term modifies (i.e. by reducing) the positive relationship between BA increment and initial BA (i.e. B_1) as quantified by the first term on the right hand side of (4). This last term is always positive so increasing B_1 increases the BA increment in proportion to B_1 for fixed values of γ_i, η_{1i} and η_{2i}. This can be seen by expressing the predicted BA increment from (4,5) as

$$\hat{B}_{2i} - B_{1i} = B_{1i}\left[\exp\left\{\left(\frac{\hat{\eta}_{1i}}{\hat{\eta}_{2i}}\right)^{\hat{\alpha}_2}\right\} - 1\right]\exp\left[\hat{\gamma}_i\left\{1-\left(\frac{\hat{\eta}_{1i}}{\hat{\eta}_{2i}}\right)^{\hat{\alpha}_2}\right\}\right].$$

In contrast, for fixed B_{1i}, η_{1i} and η_{2i} then BA increment decreases with increasing site index via the term γ_i if α_1' is negative. Nevertheless, for a given stocking rate and initial age, T_1, then B_1 will be an increasing function of site index as quantified by the B10 model in Candy (1997). The overall effect is that BA increment will generally increase with site index since the positive effect of initial BA (and thus indirect effect of site index) on BA increment will 'overpower' the negative effect of site index within the γ_i term. However, model (4,5) with negative α_1' can give counter-intuitive predictions of decreasing BA increment with increasing site index if B_1 and *S* are 'unlinked' by fixing B_1 while varying *S*. This discussion does not include the effect of thinning on BA increment.

To further investigate the effect of S within the γ_i term additional covariates of plot altitude (E) [i.e. elevation (masl)] and latitude (L) (^{0}S) were added to give

$$\gamma_i = \alpha_0' + \alpha_1' S_i + \alpha_6 E_i + \alpha_7 L_i + b_{1i} \qquad (5a).$$

The estimate of α_7 was not statistically different from zero with an estimate of 0.0062 (s.e. 0.0118). Dropping the latitude term gives

$$\gamma_i = \alpha_0' + \alpha_1' S_i + \alpha_6 E_i + b_{1i} \qquad (5b)$$

with estimates of α_1' and α_6 being –0.0207 (s.e. 0.0130) and $-0.358\text{x}10^{-3}$ (s.e. $0.205\text{x}10^{-3}$) respectively. The OLS estimates for model (5b) of α_1' and α_6 were –0.0435 (s.e. 0.0141) and $-0.120\text{x}10^{-3}$ (s.e. $0.213\text{x}10^{-3}$). Dropping the elevation term gives model (5).

Table 3 gives the OLS parameter estimates.

The final version of the model fitted was model (4,5c) where model (5c) is the simple constant log-asymptote model

$$\gamma_i = \alpha_0' + b_{1i} \qquad (5c).$$

The OLS parameter estimates and fit statistics are given for model (4,5c) in Table 4.

Table 3. Parameter estimates for the stand basal area projection model (4,5) using OLS

Parameter	α_0'	α_1'	α_2	α_3	α_4	α_5
Estimate	6.0864	-0.0450	0.7800	10.220	2.552	-0.1426
Standard error	0.4740	0.0141	0.0356	1.880	1.140	0.0454

fit statistics	*Value /estimate*
RSS{ $\ln(B_{2i})$ }	12.92 d.f.=867
$\hat{\sigma}^2$	0.01490

Dropping site index, to give model (4,5c), and using an OLS fit gave an estimate of α_2 of 0.8385 (s.e. = 0.0296). As discussed earlier model (4,5c) was then fitted using the marginal method for the NLMM. The parameter estimates are given in Table 4.

Conditional (i.e. plot-specific) predictions (or fitted values) from the WGLS fit of model (4,5c) are given by $\hat{\ln}\left(B_2^{(\hat{b})}\right) = \hat{\ln}\left(B_2 \mid b_1 = \hat{b}_1, b_2 = \hat{b}_2\right)$ where best linear unbiased predictions (BLUPs) (Robinson, 1991) of b_1 and b_2 were used to give $\hat{b}_1$ and $\hat{b}_2$, respectively. Figure 6 gives the conditional residuals given by $\left\{\ln(B_2) - \hat{\ln}\left(B_2^{(\hat{b})}\right)\right\}$ versus conditional predictions (or fitted values).

Table 4. Parameter estimates for the stand basal area projection model (4,5c) using WGLS

Parameter	α'_0	α'_1	α_2	α_3	α_4	α_5
Estimate	4.6078	-	0.8385	10.8218	1.5193	-0.1048
Standard error	0.0472		-	1.9394	0.9412	0.0383

fit statistics	*Value /estimate*	*Standard Error*
$RSS\left\{\ln\left(B_2^{(b=0)}\right)\right\}$	12.96	-
$RSS\left\{\ln\left(B_2^{(b=\hat{b})}\right)\right\}$	5.14 (d.f.=865)	-
$\hat{\sigma}_{11}^2$	0.3659	0.1160
$\hat{\sigma}_{22}^2$	2.5291	0.7369
$\hat{\sigma}_{12}$	0.5546	0.2628
$\hat{\sigma}$	0.0583	0.0037

Figure 7 shows marginal residuals, $\left\{\ln(B_2) - \hat{\ln}\left(B_2^{(b=0)}\right)\right\}$, where $\hat{\ln}\left(B_2^{(b=0)}\right) = \hat{\ln}\left(B_2 \,|\, b_1 = 0, b_2 = 0\right)$, versus marginal predicted values $\hat{\ln}\left(B_2^{(b=0)}\right)$. The $RSS\left\{\ln\left(B_2^{(b=0)}\right)\right\}$ in Table 4 is the marginal residual sum of squares. Figure 8 shows weighted conditional residuals given by $\left\{\ln(B_2) - \hat{\ln}\left(B_2^{(\hat{b})}\right)\right\} w^{\frac{1}{2}}$ versus conditional fitted values. Figure 9 shows values of current annual increment (CAI) in stand basal area, $(B_2 - B_1)/(T_2 - T_1)$ versus marginal predictions given by $\left\{\exp\left[\hat{\ln}\left(B_2^{(b=0)}\right)\right] - B_1\right\}(T_2 - T_1)^{-1}$. Figure 10 shows the equivalent figure to Figure 9 using conditional predictions of CAI. Note how improved the fit is with a value of the conditional sum of squares, $RSS\left\{\ln\left(B_2^{(b=\hat{b})}\right)\right\}$, of 5.14 (see Table 4) compared to 12.96 for the marginal sum of squares. Figures 11 and 12 show conditional and marginal residuals in terms of B_2 versus age, T_2. Figure 13 shows values of current annual increment (CAI) in stand basal versus predictions calculated using OLS estimates of parameters in model (4,5c).

It is clear from Figures 9 and 13 that there are a substantial number of projection periods for which the observed CAIs for stand basal area range between 8 and 10.7 $m^2ha^{-1}yr^{-1}$ (i.e. 100

projection periods) whereas predictions range from no more than 8 $m^2ha^{-1}yr^{-1}$, often substantially less, down to 0.5 $m^2ha^{-1}yr^{-1}$. For these projection periods most of the ages at period end are less than 8 yrs with the upper quartile being 6.4 yrs. The upper and lower quartiles for the period length, $T_2 - T_1$, are 0.7 and 0.9 yr respectively. These very large observed CAIs were not explained by high stockings with the lower and upper quartiles being 1067 and 1690 sph respectively. The 100 periods came from 75 plots most of which contributed a single period. Given the predictor variable values in η_1, η_2, and B_1 it is unclear why these projection periods have observed CAIs in stand basal area so much greater than the predicted values.

Figure 14 gives a histogram of conditional residuals, $\hat{\varepsilon}_{2i}$, and Figure 15 histograms of the BLUPs of b_1, and b_2. Figure 16 shows a scatterplot of $\hat{b}_1$ versus $\hat{b}_2$. Figure 17 gives the partial residuals from the fit of the BLUPs of b_1 as a multiple regression on *S* and *E*. The fitted relationships in Fig. 17 are shown as lines and the partial residuals are positioned with respect to the lines to represent the observed data [GenStat Release 4.2 (Fifth Edition), ©2001, Lawes Agricultural Trust]. It can be seen that Figure 17 confirms the negative estimates of α_1' and α_6 from the fit of model (4,5b).

Stand basal area at age 10

The typical use of stand projection models is to 'grow-on' inventory plots established in existing stands that are old enough for inventory to reliably represent each site's growth potential. Another use of these models is the growth simulation of hypothetical stands. To carry out these simulations a starting age and corresponding MDH (or site index) and stand basal area are required. Stands can be grown forwards or backwards from this starting age. To make this task possible given only site index, a model has been constructed to predict stand basal area at age 10 (*B*10) for unthinned stands with stocking in the range 900 to 1100 sph from a single input of site index. The data used to fit this model was obtained by selecting measurements for which the stocking was in the required range and no prior thinning had occurred. Figure 18 shows values of *B*10 versus site index. The curve in Figure 18 is the model fitted in Candy (1997) given by

$$B10 = \exp\left(4.8432 - 38.52\, S^{-1}\right) \qquad (6)$$

since the data was too restricted in the range of site index to allow model (6) to be re-calibrated here. From Figure 18 it can be seen that apart from a few obvious outliers the model of Candy (1997) fits the data reasonably well.

Stand volume

Stand volume, *V*, was defined earlier as the total of individual-tree entire stem volumes (i.e. volume from ground to tip) for the plot divided by plot area. The Schumacher model was used to predict *V* given measured MDH (*H*) and stand basal area (*B*) at any stand age. The model is given by

$$V = \exp\left\{\beta_0 + \beta_1 \ln(H) + \beta_2 \ln(B)\right\} + \varepsilon \qquad (7).$$

Candy (1989) fitted model (7) as a GLM with gamma error and log link function. Candy (1997) fitted model (7) as a GLMM with a single random effect for the intercept. Alternatively model (7) can be fitted as a log-linear model assuming normally distributed errors on the log scale to give

$$\ln(V) = \beta_0 + \beta_1 \ln(H) + \beta_2 \ln(B) + \varepsilon' \qquad (8).$$

Excluding missing values of *H*, there were 1417 observed values of *V*, *H*, and *B* across 266 plots, available for model fitting and testing. Observed and fitted values of stand volume from the fit of model (7) as a gamma/log link GLM are shown in Figure 19 while Figure 20 shows the corresponding ordinary residuals, $\hat{\varepsilon}$. The corresponding standardized residuals (i.e. deviance residuals scaled by the leverages, GenStat Reference Manual, pg.364) are given in Figure 21. Alternatively, model (8) was fitted using OLS and Figure 22 shows ordinary residuals, $\hat{\varepsilon}'$, versus observed values. All these residual plots demonstrate a significant lack of fit of Schumacher's model. Due to the different methods of fitting the trend in the residuals with observed values is different but in both cases unacceptable. Figures 23 and 24 show observed versus fitted values and ordinary residuals on the log scale for the simple linear regression of $\ln(V)$ on the log of the product of MDH and live stand basal area given by

$\ln(V) = \beta_0 + \beta_1 \ln(HB) + \varepsilon''$.

Figure 23 indicates that there is a change in slope of the linear relationship between $\ln(V)$ and $\ln(HB)$ which occurs somewhere between 2 and 4 on the $\ln(HB)$ scale which is also demonstrated by the trend in residuals in Figure 24.

To remove this trend in residuals, a nonlinear model which smoothly joins two linear regression lines was fitted where this model given by equation (4.5.6) of Ratkowsky (1989, pg.119) is

$$\ln(V) = \beta_0 + \beta_1 \left\{\ln(HB) - \gamma\right\} + \beta_2 \left[\left\{\ln(HB) - \gamma\right\}^2 + \theta^2\right] + e \qquad (9).$$

Table 5 gives the parameter estimates and some fit statistics for model (9) whereby the nonlinear parameter estimates of (γ, θ) were obtained using an OLS fit of model (9) and the linear parameters $(\beta_0, \beta_1, \beta_2)$ were fitted using GLS by fitting (9) as a LMM conditional on the OLS estimates of (γ, θ). For the OLS fit, model (9) explained 99.8% of the variance in $\ln(V)$. A random plot-level intercept term, $(\beta_0 + b_1)$, and slope term, $(\beta_2 + b_2)$, were fitted as part of the LMM using the GLS/REML fitting algorithm. Fitting was unsuccessful when all three linear parameters in model (9) were allowed to incorporate random plot effects that co-vary across random effects. The log-likelihood for the above LMM was greater than that obtained when a random plot-effect was incorporated in β_1 rather than β_2. For comparison with the GLS estimates and their standard errors, Table 6 gives the full set of OLS parameter estimates. Figures 25 and 26 show observed values and ordinary residuals versus $\ln(HB)$ respectively from the OLS fit with the fitted regression line shown in Figure 25 based on the parameter estimates in Table 6. Figure 27 corresponds to Figure 25 with the fitted model in the former case based on the estimates in Table 5. Figure 28 shows the marginal residuals (i.e. excluding the estimated random effects from fitted predictions) and Figure 29 the conditional residuals (i.e. including the estimated random effects in predictions) from the fit of the LMM versus $\ln(HB)$. Figure 30 shows the random effects estimates $\hat{b}_1$ versus $\hat{b}_2$. Figure 31 shows a histogram of the conditional residuals, $\hat{e}$, while Figure 32 shows histograms of $\hat{b}_1$ and $\hat{b}_2$.

Table 5. Stand volume model parameter estimates and fit statistics for the GLS/OLS fit for linear/nonlinear parameters

Parameter	β_0	β_1	β_2	γ	θ
estimate	1.6750	0.7883	0.1340	2.4971	-0.8780
standard error	0.0084	0.0013	0.0027	0.0919	0.1780

fit statistics	*Value /estimate*	*standard error*
RSS{ln(*V* |***b***)	2.075 (df=1409)	-
$Var(b_1)$	0.01163	0.00159
$Var(b_2)$	0.00076	0.00013
$Cov(b_1,b_2)$	-0.00275	0.00043
σ^2	0.00190	0.00009

Table 6. Stand volume model parameter estimates and fit statistics for the OLS fit for linear parameters

Parameter	β_0	β_1	β_2
estimate	1.6309	0.7953	0.1418
standard error	0.0963	0.0055	0.0080

fit statistics	*value/estimate*
RSS{ln(*V*)}	6.769 (df=1412)
σ^2	0.004794

Stand mortality

The proportional hazards model of Candy (1989, 1997) was fitted here. The model is given by

$$M_2 = N_1\left\{1 - \exp\left[-\int_{T_1}^{T_2} \eta(t)dt\right]\right\} + \varepsilon \qquad (10)$$

where

M_2 is the mortality on the plot between ages T_1 and T_2, N_1 is the number of live trees on the plot at age T_1, and $\eta(t)$ is the hazard function given by

$$\eta(t) = \exp(\beta_0 + \beta_1 S + \beta_2 t + \beta_3 t^2)$$

and ε is a binomial error conditional on N_1.

Candy (1989) fitted this model as a GLM with binomial error for M_2 conditional on N_1, complementary log-log link, linear predictor

$$\eta = O + \beta_0 + \beta_1 S + \beta_2 T_m + \beta_3 T_m^2$$

where $T_m = (T_2 + T_1)/2$,

and where O is an 'offset' given by $\ln(T_2 - T_1)$. This model approximates the integral in model (10) using the mid-point rule and the observed measurement periods (Candy, 1986). Candy (1989) applied prior weights to scale the binomial error variance but this was not done here, instead model (10) was fitted initially simply as a standard GLM.

Some screening of the data was required to exclude periods in which the mortality was too excessive to be due to regular mortality. For all 975 projection periods that had a site index estimate available, the maximum annual mortality rate (i.e. $R = M_2 / [(T_2 - T_1)P_A]$ where P_A is plot area in hectares) was 875 stems/ha. Projection periods for which R was greater than 250 stems/ha were excluded giving 956 periods. Figure 33 shows signed deviance residuals (McCullagh and Nelder, 1989) versus age from the fit of model (10) as a GLM. If model (10) truly represents the systematic and random variability in M_2 then the absolute value of each point in Figure 33 should have close to a chi-square distribution with single degree of freedom. The residual mean deviance was 1.397 which would normally indicate a lack of serious overdispersion. However, the majority of the values of M_2 were zero (i.e. 717 out of 956 values) which makes the usual assumption that the residual deviance has an asymptotic chi-square distribution unlikely to be even approximately satisfied. Figure 34 shows the observed compared to predicted values of *R* and demonstrates the large influence of the substantial proportion of zero values of M_2 on the predictive model.

Figures 35 and 36 correspond to Figures 33 and 34 respectively with the difference being that the signed deviance residuals and predictions of *R* are based on the fit of model (10) with the zero values of M_2 excluded (i.e. 956-717=239 values). Figure 37 corresponds to Figure 36 with the difference that predictions of *R* were based on the fit of model (10) as in Candy (1997) as a Generalized Linear Mixed Model (GLMM) using the marginal method of fitting (Breslow and Clayton, 1993) and the GenStat procedure GLMM. Model (10) was also fitted as a Heirarchical Generalized Linear Model (HGLM) (Lee and Nelder, 1996,2001) and gave similar results to the GLMM. Although, the HGLM is theoretically more appealing, since it only gives 'subject-specific' parameter estimates, predictions must be obtained by integrating random effects out using their estimated distribution (Lee and Nelder, 1996; Candy, 2002). This makes implementing the HGLM unwieldy so it was not pursued here. The GLMM incorporated a conditional binomial error and a random plot effect, *b*, within the linear predictor so the hazard function is now

$$\eta_b(t) = \exp(\beta_0 + \beta_1 S + \beta_2 t + \beta_3 t^2 + b_1)$$

where b_1 has a normal distribution with variance σ_1^2. The conditional (i.e. given b_1) binomial variance is given by $\phi\ \hat{p}(1-\hat{p})/N_1$ where $\hat{p} = \hat{M}_{\hat{b}_1}/N_1$ are the conditional predictions. The value of the dispersion factor, ϕ, was set to unity. Table 7 gives the parameter estimates and some fit statistics.

Given the condition that the parameter estimates for model (10) given In Table 7 are based only on periods that had non-zero mortality less than 250 stems/ha/yr, predictions in general will over-estimate periodic mortality.

Table 7 Mortality model (10) parameter estimates and their standard errors

Parameter	β_0	β_1	β_2	β_3
estimate	-5.4650	0.0336	0.3571	-0.0186
standard error	0.6664	0.0220	0.0954	0.0072

fit statistic	*value/estimate*	*Standard error*
σ_1^2	0.3610	0.0790
ϕ	1.0	

To give unbiased predictions, model (10) was combined with a model that predicts the probability of non-zero mortality. To construct this last model a binary variable, *Y*, was constructed whereby *Y*=1 for a projection period if *R*>0 and *Y*=0 if *R*=0. After screening potential predictor variables using Wald tests obtained from the VDISPLAY directive of GenStat (Genstat 5 Committee, 1997a,b) the final GLMM model fitted, using the GenStat GLMM procedure (Payne *et al*., 1997) was

$$\Pr(Y = 1) = \exp(\eta_0)/\{1 + \exp(\eta_0)\} \tag{11}$$

where

$$\eta_0 = \alpha_0 + \alpha_1 O + \alpha_2 N_A + \alpha_3 S + \alpha_4 F + \alpha_5 F.S + b_0$$

$O = \ln(T_2 - T_1)$, $N_A = N_1 / P_A$ is the stocking (stems/ha) at the start of the period, *F*=1 if the plot has been thinned prior to age T_1 and *F*=0 otherwise, *Y* has a binomial distribution conditional on b_0 and b_0 has a normal distribution with variance σ_0^2. Table 8 gives the parameter estimates and some fit statistics obtained from the fit of the GLMM.

To combine models (10) and (11) in order to predict *R* for $R \geq 0$ the marginal predictions (i.e. setting $b_0 = b_1 = 0$) are given by

$$E(M_2) = E(M_2 \mid Y = 1)\Pr(Y = 1) + E(M_2 \mid Y = 0)\Pr(Y = 0).$$

However, since the second term on the right hand side of the above equation is zero the model simplifies to

$$E(M_2) = E(M_2 \mid Y = 1)\Pr(Y = 1) \quad (12)$$

where $E(M_2 \mid Y = 1)$ is given by model (10) with parameter estimates given in Table 7 since these estimates represent the fit with zero values of *M*, and thus *R*, excluded. The values of $\Pr(Y = 1)$ are obtained from model (11). Figure 38 shows observed and predicted values of *R* obtained from model (12) where the observed values of zero are included.

Table 8 Occurrence-of-Mortality model (11) parameter estimates and their standard errors

Parameter	α_0	α_1	α_2	α_3	α_4	α_5
estimate	-6.1660	0.7306	0.000402	0.1702	-9.0630	0.2423
standard error	0.9652	0.2404	0.000162	0.0350	3.2830	0.1039

fit statistic	*value/estimate*	*Standard error*
σ_0^2	0.441	0.180
ϕ	1.0	

Figure 39 shows histograms of the estimated random effects from the fits of model (10) and model (11).

The values of *Y* were accumulated in six classes where these classes were allocated by dividing the estimated values of the linear predictor, $\hat{\eta}_0$, from model (11) into classes with cut-points of -2,-1,-0.5,0.0, and 0.5. The proportion of the *Y*'s that were one versus the class median value of the linear predictor is shown in Figure 40 along with the predicted curve obtained from model (11). The fitted curve in Figure 40 is not expected to be a least squares fit to the class proportions since model (11) was fitted to the binary data values and included random plot-effect estimates. However, examining Figure 40 is useful in confirming the general trend in Pr(*Y*=1) with predictor variables as quantified by model (11).

Figure 41 shows predicted stocking curves for unthinned stands with 3000, 2000, and 1000 initial stems per hectare at age 3 for each of site indices 20 and 30. The curves in Figure 41 were constructed by combining models (10) and (11), as given by (12). Using annual intervals, model (10) was evaluated using the mid-point rule to evaluate the integral $\int_{T_1}^{T_2} \eta(t)dt$ in the same way that the GLM and GLMM models were fitted (i.e. via the offset and the mid-points given by T_m). Stocking was updated at the end of each annual projection period as a required input to model (11). Since stocking and projection period length are included as a predictor variables in model (11), then the combined model (12) is path-variant. Path-variance means that projection of stocking between ages T_1 and T_3 in a single step rather than projecting to an intermediate age, say T_2 (where $T_1 < T_2 < T_3$) and then to age T_3 in a second step will result in different estimates of stocking. If model (11) was replaced by the following generalized logistic model (Monserud 1976) with the stocking and *O* terms dropped from the linear predictor

$$\Pr(Y=0) = \{1+\exp(-\eta_0)\}^{-(T_2-T_1)}$$

and with (12) redefined

$$E(M_2) = E(M_2 \mid Y=1)[1-\Pr(Y=0)]$$

then the combined model is path-invariant. However, there was a significant reduction in goodness-of-fit when stocking was removed from model (11) (deviance increase of 7.2; *P*<0.01 cf: chi square distribution with single degree of freedom). When the stocking term was included in the above generalized logistic model the residual deviance was greater by a value 4.2 (*P*<0.05) compared to that of model (11).

It was found that the two-level 'thinning status' factor (measurement periods that were not preceded by a thinning versus those that were) and its interaction with the age and site index terms in model (10) were not significant for the restricted data set defined by 0<*R*<250. However, the thinning factor and its interaction with site index were significant in model (11) (Table 8). The overall improvement in combining models (10) and (11) to give predictions via (12) compared to simply fitting model (10) (with separate parameters for each level of the thinning factor) for the data set including the zeros was measured by the reduction in the residual (marginal) deviance. The residual deviance was reduced from 1343 to 1331 for model (12) predictions. The larger deviance was calculated using model (10) predictions for the 956 values of *M* as obtained by fitting the model as a marginal GLMM to allow a fair comparison (i.e the GLM fit is designed to minimize the marginal binomial deviance). The improvement in fit is not very large, but as noted by Woollons (1998), modelling zero and non-zero mortality separately facilitates the exploration of candidate models. The parameter estimates for the version of model (10) with the thinning-status factor included, redefined as model (13), and calibrated with the zero values included in the data, are given in Table 9. These estimates were obtained from the GLMM fit mentioned above where model (13) is given by.

$$\eta = O + \beta_0 + \beta_1 S + \beta_2 T_m + \beta_3 T_m^2 + \beta_4 F + \beta_5 F.S + \beta_6 F.T_m + \beta_7 F.T_m^2 + b \qquad (13)$$

where F was defined for model (11) and b is a random plot effect. Figure 42 shows the observed and predicted values of R obtained from the fit of model (13) while Figure 43 shows the histogram of estimated random effects. Figure 44 shows predicted stocking curves calculated from model (13).

Table 9 Mortality model (13) parameter estimates and their standard errors

Parameter	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
estimate	-11.0106	0.1630	0.5520	-0.0309	-26.9716	0.3764	3.2941	-0.1833
S.e.	0.8629	0.0278	0.0943	0.0070	4.8488	0.0913	1.0859	0.0736

fit statistic.	*value/estimate*	*Standard error*
σ_1^2	0.964	0.161
ϕ	1.0	

It can be seen from Table 7 that the standard errors for the parameter estimate of β_1 indicate that the site index term in model (10) may not be significant. However, this term should be retained since self-thinning is *a priori* more intense at earlier ages in highly stocked stands on high quality sites relative to lower quality sites (Candy 1989). In addition, the site index term and its interaction with thinning status are both highly significant in model (11) and the site index term is highly significant in model (13) (Table 9). Overall the site index effect appears to be stronger in model (12) compared to model (13) (Figure 45).

The recommendation of which model should be implemented is not obvious. The path-variant model (12) (based on combining models (10) and (11) with parameter estimates given in Tables 7 and 8) can be implemented with annual steps used to update stocking as used to construct Figure 41. Alternatively, the path-invariant model (13) [note that model (10) is in theory path-invariant by way of the integral $\int_{T_1}^{T_2} \eta(t)dt$] using parameter estimates from Table 9 could be implemented. Woollons (1998) notes that models such as (11) can be implemented stochastically by using a random uniform number in the [0,1] interval to determine if a projection interval, such as the annual steps used to construct Figure 41, is assigned zero mortality. If for that interval non-zero mortality is assigned then model (10), using Table 7 parameter estimates, can be directly applied.

Discussion

The single asymptote polymorphic site index curves given in Figure 5 depend as expected very strongly on the estimate of the single asymptote. The estimated asymptote of 41.4 m seems unrealistically low given that the largest value of PDH of 39.7 was obtained at age 22. There were a number of PDHs above 36 m for stands ranging in age from 19 to 22. Further work could be carried out to determine the effect of site-dependent variables such as altitude, latitude, and other variables that determine physiological determinants of growth (such as average temperature, number of frost days) on the shape and asymptote of the PDH versus age trajectory.

The site index term in the asymptote of the BA projection model was found to be negative and statistically significant. Such a negative term in the asymptote can be considered counter-intuitive although it was described earlier how such a negative effect of site index can occur in a stand basal area projection model. A similar negative trend was also estimated as significant for *P. radiata* PSPs in Tasmania (Candy, 1989, 1997b) with that data being much more substantial in terms of number of plots and age range than the data used here. Despite this, applications of the projection model that ‘unlink’ site index and stand basal area can give nonsensical BA projections. It would be more reassuring to have either a positive or non-significant site index term. The deviations from the final BA projection, model (4,5c), when expressed in terms of current annual increment are noisier than expected as discussed earlier. In comparison, the model calibrated using only Tasmania data had variance component estimates that were substantially smaller than those given in Table 4. In addition the trends in random effects obtained from the fit of model (4,5c) were reverse of those seen in Figure 17. The large latitudinal range of the New Zealand PSPs of almost

11 degrees compared to a 3.5 degree range for mainland Tasmania could explain the large variability in model deviations. Further work, possibly employing a site-productivity measure predicted from physiologically-based models such as 3PG to replace or augment the site index term is recommended in order to improve the precision of predictions.

The failure of the Schumacher model to accurately predict stand volume as a function of PDH and stand basal area is unusual. Also the change of slope in the relationship between stand volume and *ln*(MDHxBAlive) seen in Figures 25 and 27 is difficult to explain. For example the change of slope was not due to thinning since when measurements at and after a thinning were excluded (i.e. 293 measurements) the graph of observed and fitted values versus *ln*(MDHxBAlive) did not change noticeable from Figures 25 or 27.

Mortality is one of the most difficult stand characteristics to model satisfactorily. The lack of older age measurements (i.e. > 20 years) available for model calibration results in a lack of confidence in predictions from either model (12) or (13) at ages greater than 20. If such data were available it might show that the stocking curves converge to a single lower asymptote independent of initial stocking (i.e. at age zero) (Xue and Hagihara, 2002) rather than separate asymptotes as seen in Figures 41, 44, and 45. The equivalent figure to Figure 44 given in Candy (1989, Fig.3) suggests that this hypothesis of a common asymptotic stocking has some support with the qualification that this asymptote, although independent of initial stocking, may depend on site index.

References

Bailey and Ware (1983) Compatible basal-area growth and yield model for thinned and unthinned stands. *Canadian Journal of Forest Research* **13**, 563-571.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9-25.

Candy, S.G. (1989). Growth and yield models for *Pinus radiata* in Tasmania. *New Zealand Journal of Forestry Science* **19**, 112-133.

Candy, S.G. (1997). Growth and yield models for *Eucalyptus nitens* plantations in Tasmania and New Zealand. *Tasforests* **9**, 167-198.

Candy, S.G. (1997b) Estimation in forest yield models using composite link functions with random effects. *Biometrics* **53**, 146-160.

Candy, S.G. (2002) Empirical Binomial Sampling Plans: Model Calibration and Testing Using Williams' Method III for Generalized Linear Models with Overdispersion. *Journal of Agricultural, Biological, and Environmental Statistics*, **7**, 373-388.

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.

Garcia, O. (1994). The state-space approach in growth modelling. *Canadian Journal of Forest Research* **24**, 1894-1903.

Genstat 5 Committee (1997a). **Genstat 5 Reference Manual**. Oxford: Clarendon Press.

Genstat 5 Committee (1997b). **Genstat 5 (Release 4.1) Reference Manual Supplement**. Oxford: Clarendon Press.

Lee, Y. and Nelder, J.A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*, **58**, 619-78.

Lee, Y. and Nelder, J.A. (2001). Hierarchical generalized linear models: a synthesis of generalized linear models, random-effect models and structured dispersions. *Biometrika*, **88**, 987-1006.

Monserud, R.A. (1976). Simulation of forest tree mortality. *Forest Science* **22**,438-444.

Payne, R.W., Arnold, G.M., and Morgan, G.W. (Eds.) (1997). Genstat 5 Procedure Library Manual. Release PL10. Lawes Agricultural Trust (Rothamsted Experimental Station)

Pienaar, L.V. and Shiver, B.D. (1986). Basal area prediction and projection equations for pine plantations. *Forest Science* **32**, 626-633.

Ratkowsky, D.A. (1989). *Handbook of Nonlinear Regression Models*. Marcel Dekker Inc., New York.

Robinson, G.K. (1991) That BLUP is a good thing: the estimation of random effects. *Statistical Science*, **6**, 15-51.

Woollons, R.C. (1998). Even-aged stand mortality estimation through a two-step regression process. *Forest Ecology and Management*, **105**, 189-195.

Xue, L. and Hagihara, A. (2002). Growth analysis on C-D effect in self-thinning Masson pine (*Pinus massoniana*) stands. *Forest Ecology and Management*, 165**, 249-256.**

Appendix 1: Projection stand basal area from an inventory

Two types of predictions of B_2 given model (4,5c) were given above, plot-specific predictions given by $\exp\left[\hat{\ln}\left(B_2^{(\hat{b})}\right)\right]$ and average-plot predictions $\hat{B}_2 = \exp\left[\hat{\ln}\left(B_2\right)\right]$. In application of model (4,5c) to projecting a set of inventory plots within a forest patch or compartment from the measurement age to a later harvest age, then plot-specific predictions are not available since estimation of b_1 and b_2 is not possible for the inventory plots. The average-plot predictions are available since these are obtained by setting b_1 and b_2 to zero as described above. However, population-average (PA) predictions can also be obtained using a bivariate normal distribution for b_1 and b_2, with variance-covariance matrix $\mathbf{D}$ for which estimates of its elements are given in Table 5, and the following integration:

$$\tilde{B}_2 = \exp\left[\tilde{\ln}\left(B_2\right)\right]$$

where

$$\tilde{\ln}\left(B_2\right) = \frac{1}{2\pi}\left|\mathbf{D}\right|^{-\frac{1}{2}} \iint_{b_2, b_1} \hat{\ln}\left(B_2 \mid b_1, b_2\right) \exp\left\{-\tfrac{1}{2}\left(b_1, b_2\right)\mathbf{D}^{-1}\left(b_1, b_2\right)^T\right\} db_1 db_2 .$$

Since the joint distribution of the random effects can be expressed from Bayes Theorem as $f\left(b_1, b_2\right) = f\left(b_1 \mid b_2\right).f\left(b_2\right)$, model (4,5c) is linear in b_1, and the expected value of b_1 is zero then the above integral can be simplified to

$$\tilde{\ln}\left(B_2\right) = \frac{1}{\sqrt{2\pi\sigma_{22}^2}} \int_{b_2} \hat{\ln}\left(B_2 \mid b_1 = 0, b_2\right) \exp\left\{-\tfrac{1}{2} b_2^2 / \sigma_{22}^2\right\} db_2 \qquad \text{(A1).}$$

The question arises then as to which estimate of the two possible predictions $\hat{B}_2$ (average-plot estimate) or $\tilde{B}_2$ (population-average estimate) should be used in 'growing on' a sample of inventory plots for a given area of forest. There are two methods of growing on a sample of inventory plots with each method suggesting the most appropriate of the two types of predictions. If each inventory plot is grown-on to age T_2, then it is recommended that PA predictions, $\widetilde{B}_2$, are calculated, requiring the above integral (A1) to be evaluated for each plot, and the average of the resultant sample of $\widetilde{B}_2$'s then used to represent the average stand basal area per hectare at age T_2. If however, an average of the plot attributes at age T_1 is used to provide an 'average plot' which is then grown-on to age T_2, then the average-plot prediction $\hat{B}_2$ should be used to represent the average stand basal area per hectare at age T_2.