F.R.I. PROJECT RECORD

NO. 2611

INVESTIGATIONS INTO THE POSSIBLE CAUSES OF BIAS IN THE NELSON GROWTH MODEL

M.E. LAWRENCE

REPORT NO. 19 NOVEMBER 1990

FRI/INDUSTRY RESEARCH COOPERATIVES

EXECUTIVE SUMMARY

Concern over the high rates of mortality exhibited by the earlier Nelson growth model GDNS (Garcia, 1981) lead to the development of a new model for radiata pine in the Nelson/Marlborough region during 1989/90 (Law, 1990). The new model, though superior to the earlier effort, over-predicts basal area in stands carrying more than 50 m^2/ha. This problem has now been rectified. An investigation into the likely causes of the problem and results are described in this report.

INTRODUCTION

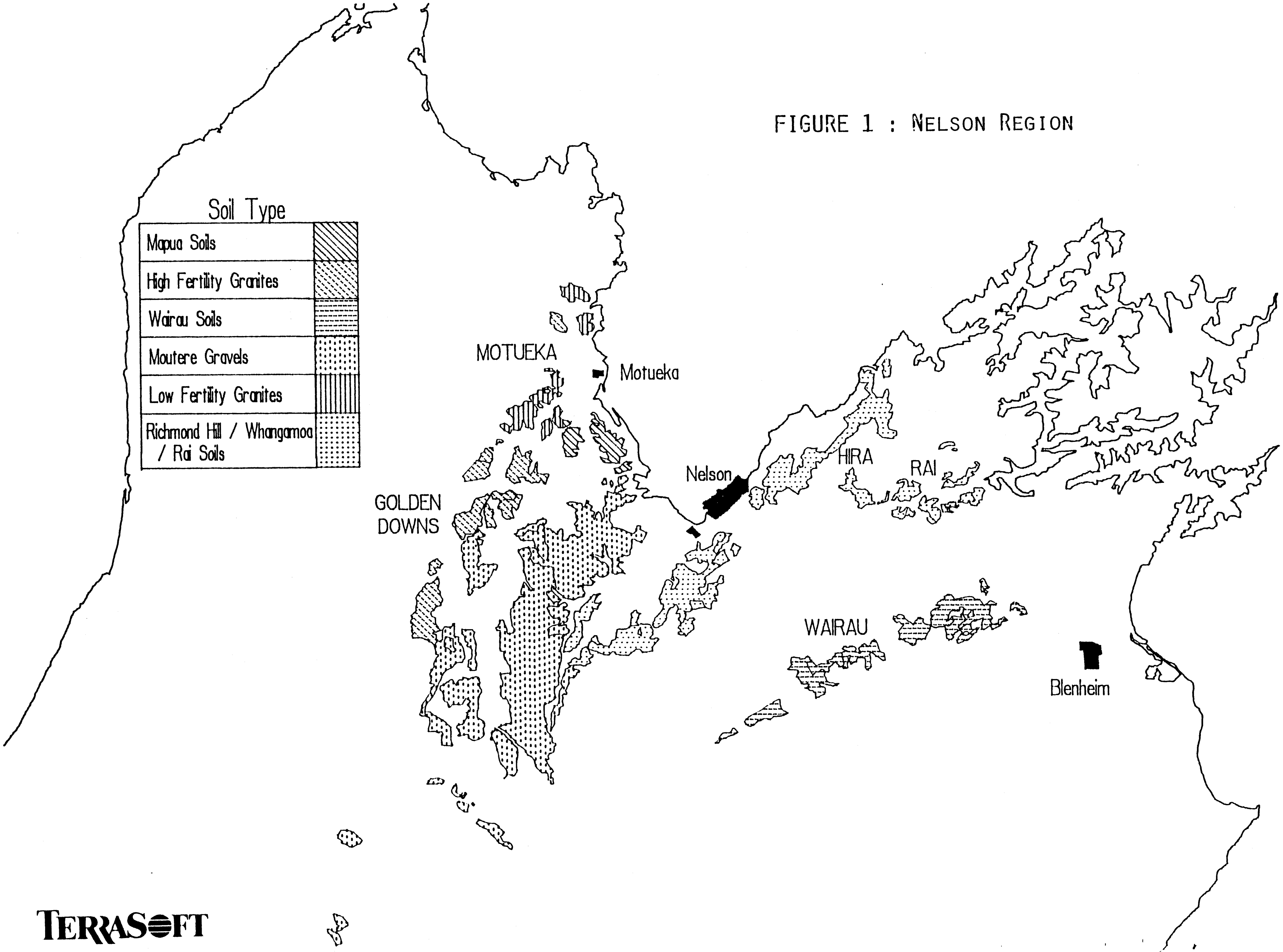
In 1987 the Growth Modelling Cooperative was presented with *Pinus radiata* data from Baigents' and Golden Downs forests showing the mortality to be significantly less than that predicted by the then current Golden Downs growth model (Garcia, 1984). Conflicting evidence from O. Garcia (Garcia, 1987) showed that the model fitted the original data well and it was concluded, therefore, that the original data no longer reflected the current situation. In July 1988 Timberlands and Baigents provided lists of plots from a number of forests in the Nelson/Marlborough region (Figure 1). These data were segregated by forest and/or soil type into the following categories:

Mapua
Wairau
Richmond
Moutere
High fertility granites
Low fertility granites

and provided the basis for a new regional growth model which was completed in February 1990 (Law, 1990).

The new model was based on the combined data from all six forest/soil groupings after a graphical comparison of the growth of the primary 'state' variables top height (m), basal area (m^2/ha) and mortality (stems/ha) showed no significant differences between the categories. Most predictions from the model are notably precise when compared with the other regional growth models, particularly those for stocking/mortality. Unfortunately, however, the basal area predictions appear to be biassed (over-predict) at higher stand densities. Law found that all the model variations tested exhibited this same characteristic, and re-examined the possible effects of soil type, climate and establishment in an attempt to discover the possible causes - to no avail. Following the presentation of these results to the Stand Growth Modelling Cooperative, it was decided to further investigate possible causes of the bias and hopefully obviate the problem and produce an updated and unbiassed model.

FIGURE 1 : NELSON REGION



TERRASOFT

DATA

The total database for the Nelson/Marlborough model illustrated in Figure 2 covers a wide range of soil types and climate zones. Unfortunately the plot coverage is not uniform in number or diversity (see Appendices 1 & 2) and the major soil types are over-represented at the expense of the younger or smaller forest areas. To remove possible reservations about the amalgamation of the six groupings, investigations were concentrated on the dominant soil type in the region - the Moutere gravels upon which Timberlands' Golden Downs Forest and Baigent's forests are based.

SUMMARY OF INVESTIGATIONS

Investigations into the possible causes of the bias fell into four categories : checking of growth model formulation and related computer software, analysis of the residuals, detailed data screening and manipulation, and comparisons with other models and databases. These are described below.

1. Model formulation and software

The development of a growth model using the techniques propounded by Garcia (1979) and adopted by the Cooperative involves the use of two major computer programs: one to estimate the height model parameters and a second to estimate the coefficients of the 'growth' equations. Both are complex programs requiring the definition of a particular model via one or more subroutines prior to 'fitting', and subsequent analysis of the results. In constructing a model, this procedure is carried out a number of times to evaluate the effects of adding or subtracting parameters from the model, or to define relationships between different variables. In addition, there are programs to derive further coefficients based on those from the parameter estimation procedures, and to calculate predictions for testing the efficacy of any particular model construct.

There are three possible sources of error. The first is in the parameter estimation programs themselves, the second in the user-controlled formulation of any particular model and the third in one of the programs used for deriving predictions and model testing. In the first instance, the programs have evolved over a number of years and been transferred/converted through several mainframe versions to their current state on a personal computer. At each stage they have been thoroughly tested to ensure no errors have been accidentally introduced. In this case, data from earlier growth models (developed on a different program version) were re-analysed on the current program to check the estimation procedures and, conversely, the Nelson data were analysed on an much earlier version of the program. In both cases, there were no discrepancies.

The second possible source of error is readily checked. Any particular model is formulated in a user-defined subroutine which ultimately forms part of the analysis program. Law found that all the models she tried exhibited the same bias in the basal area predictions, so it was simply a matter of setting up a very simple routine with a minimum of parameters to be estimated, and checking the results.

The final check was made in a manner similar to the first. Earlier program versions (different languages and computers) were re-run and checked against the current versions. In addition, the program code was checked manually. No differences were discovered.

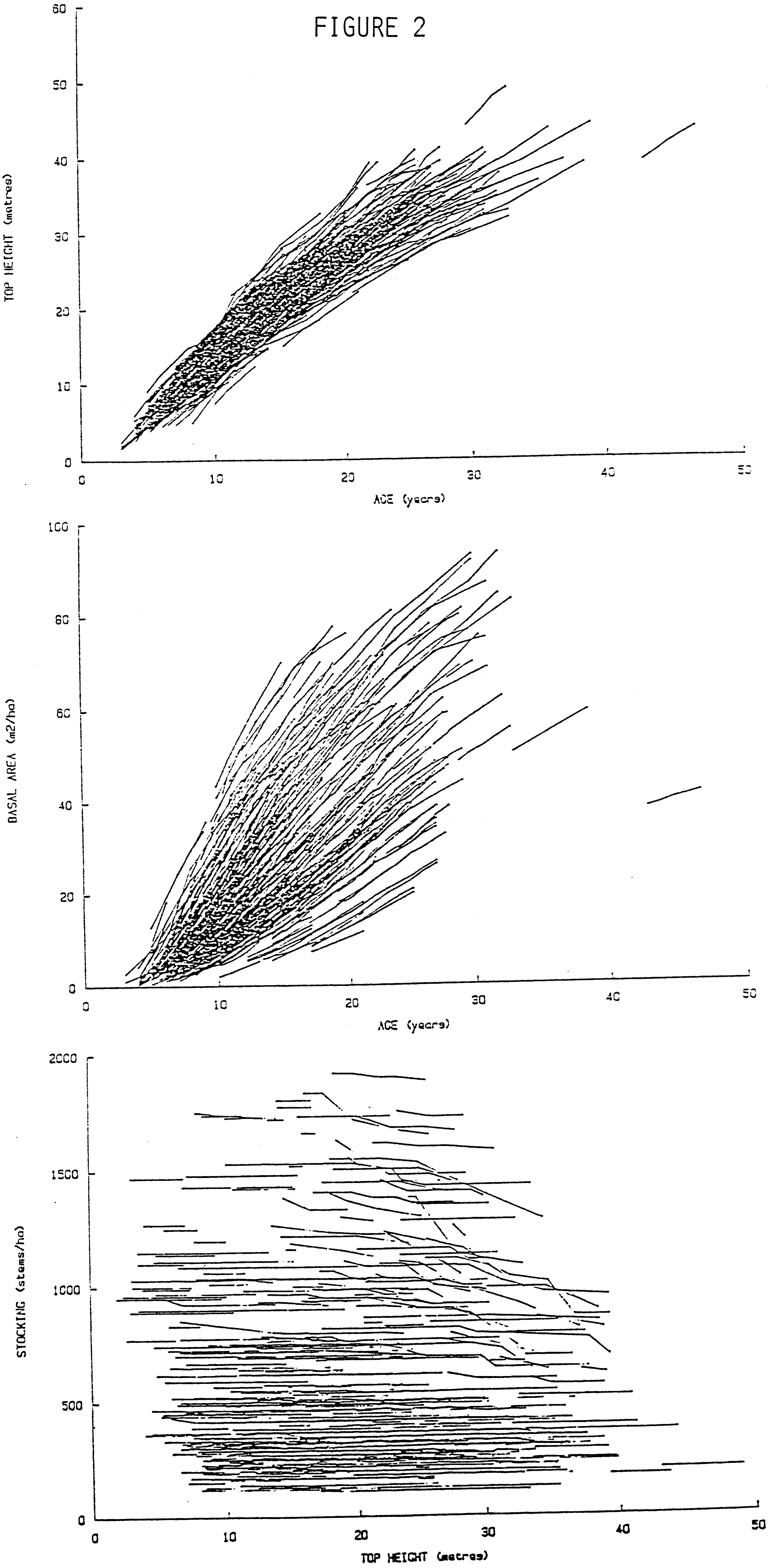
2. Residual analysis

An analysis of the basal area residuals (Figure 3a) was carried out in an attempt to discover possible trends in the data relating to particular plots or experiments, management histories or site/location effects. Initially, all the plots over $50m^2/ha$ in which the basal area was over-predicted were identified, their locations mapped and their plot histories studied. The information available for most of the plots was fairly limited and no common or unusual features could be identified.

The residuals were then graphed by plot and against stocking class, density and other variables in a further attempt to identify possible trends. Additional graphs of increments were also studied. A few unexplained 'outliers' were identified (including some fertilised plots missed previously) and removed from the database, but nothing of any significance was found.

The distribution of residuals is heavily weighted to measurements with less than $40m^2/ha$. This means that the tendency to over-predict at high basal areas is easily recognised, but trends at lower basal areas can be obscured by the mass of points. To overcome this problem, the residuals were filtered by dividing the residuals into $5m^2/ha$ basal area classes and randomly removing points to achieve an equal number in each class. It was observed that the over-predictions were in fact partially offset by an **under**-prediction between basal areas of $10\text{-}45m^2/ha$; an observation subsequently confirmed by a similar analysis of the full dataset.

FIGURE 2



TOP HEIGHT (metres)
AGE (years)
BASAL AREA (m2/ha)
AGE (years)
STOCKING (stems/ha)
TOP HEIGHT (metres)

FIGURE 3A

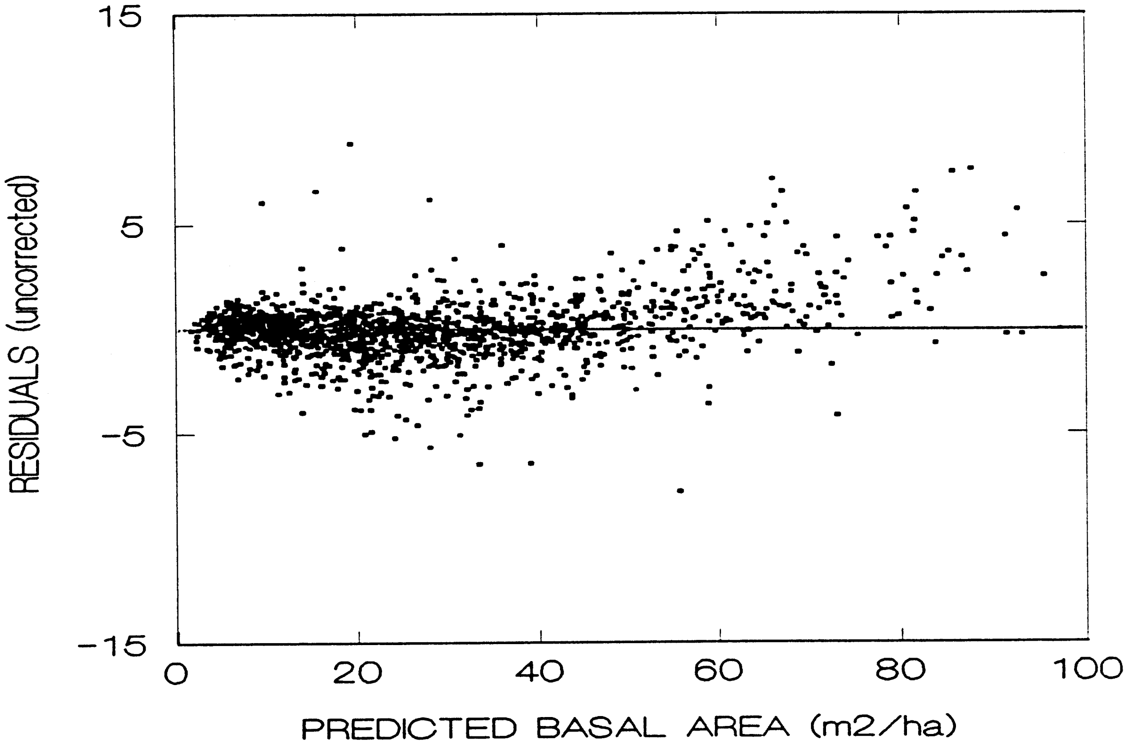
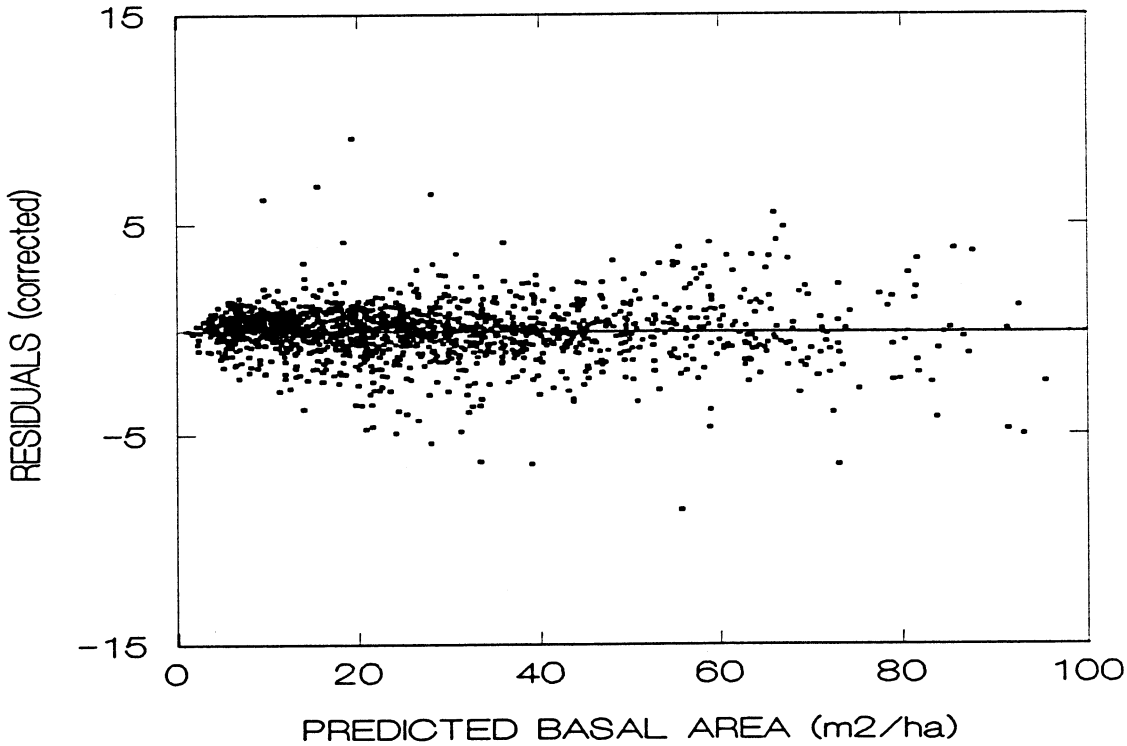


FIGURE 3B



3. Data screening and manipulation

The data were screened very carefully with the aid of specific programs and detailed graphs of individual plots and experiments. Because graphs are normally limited to only two dimensions, a plotting routine developed by O. Garcia (pers.comm.) for three dimensional plots of data was used to study the three state variables; basal area, stocking (actually average spacing), and top height. This program takes the *x,y,z* coordinates for each measurement and draws them within a frame which can then be rotated and inverted under 'mouse control'. Plot measurements are linked with solid lines (not just points) so that the development of all three variables over time can be studied for each plot, and in relation to all the other plots. By highlighting specific plots in different colours, points of interest can be located and checked.

As a result of the checks, a few plots with abnormally high mortality (at stockings above 800 stems/ha) were removed and some minor errors in stocking were corrected. (Some model coefficients are particularly sensitive to an increase in stocking of one stem/ha caused by rounding). Of some interest was a marked 'twist' in the data when viewed from certain perspectives of the 3-D graph, an effect which appeared to be related to average spacing. The ability of the growth model to generate the same pattern was checked by calculating the predictions for the identical data and plotting them also in 3-D. The resulting graph exhibited an almost identical surface, removing any doubts about the models flexibilty.

Other than in Hawkes Bay, no relationship has been found between site index and basal area. Although checked by Law during development of the Nelson model, there was some evidence of a relationship once the data had been restricted to the one major soil type and the data screened more closely. This relationship can be modelled (as in Hawkes Bay) and the magnitude of the effect assessed. The result was not at all significant, especially considering the degree of over-prediction shown by the model.

In an effort to comprehend more fully the interactions between the model coefficients, additional data were 'manufactured' to complement the existing high basal area plots. By adding more 'weight' to this end of the database and observing changes in the coefficients, some of the possible interactions could be identified. As a result, some of the coefficients were fixed at seemingly appropriate 'average' values and the others re-estimated. It proved feasible to obtain coefficients that gave reasonable (though not perfect) basal area predictions, but the resulting model was far from robust and only useful over a very limited range of conditions.

Finally, the data were split by ownership into Timberlands' Golden Downs group and those belonging to Baigents. Both groups are from the same soil type but their silvicutural philosophies are somewhat different; Baigents maintaining a comparatively high (600 stems/ha) unpruned final crop stocking compared to Timberlands' high pruned, low (200 stems/ha) final crop stockings. The data from the two groups therefore, are somewhat complementary rather than intermingled. As was the case for the 'manufactured' data, it was possible to obtain reasonable basal area predictions from the Baigents data, but the resulting model was very limited in scope. There was no improvement in the predictions for the Golden Downs data.

4. Regional comparisons

Regional comparisons of the Nelson data were made with Kaingaroa, Hawkes Bay and Canterbury. Firstly, estimated coefficients of a common model were compared and differences noted and secondly, three dimensional graphs of the regional data enabled visual comparisons of the differences in the development of the major state variables basal area, stocking and top height. Once again, no significant differences were found. From some perspectives, the 'twist' in the Nelson data noted previously is slightly more pronounced than in other regions, but it is not peculiar to the region, nor is it incorrectly modelled by the current Nelson growth model.

BIAS CORRECTION

Despite exhaustive investigations into the possible causes of the bias, no ready answers have been found. The absence of an obvious relationship between stand density and mortality in such a large database is baffling (though possibly a result of the manner in which plots were selected originally), but unlikely to be the root of the problem. It is conceivable that changes in management practice that have improved the growth of young and mid-rotation stands have not yet filtered through into the older age classes, which are currently represented by older and slower growing plots. If this is true, then predictions from the model at (say) age 30 may be correct for stands currently less than 20 years of age. **This remains unproven**.

The bias in basal area predictions above $50m^2/ha$ in the model presented by Law is not acceptable. Although it is disappointing not to have discovered the cause, it is possible to modify the predictions and thereby remove any error. During analysis of the residuals a definite trend was found in the predictions, from a slight **under**-prediction below $45m^2/ha$ to a more obvious **over**-prediction above $50m^2/ha$. Fitting a simple polynomial curve to the residuals enables a correction to be made to the predicted basal area which varies in accordance with the trend. The correction can be incorporated in the growth model program via a simple routine which is called immediately prior to be basal area prediction being displayed. It is important to remember that the correction is applied to the result only, and that the 'internal' value used within the model remains unaltered.

RESULTS

The function to correct the bias works well. The new residuals are shown in Figure 3b and the effect of the residual function is demonstrated in Table 1, which gives model predictions for the three regimes used by Law.

The regimes used for the comparison are:

1. Initial stocking: 1250 stems/ha. Unthinned

2. Initial stocking: 1100 stems/ha. Thin to 600 stems/ha at top height 6m.

3. Initial stocking: 1250 stems/ha. Thin to 600 stems/ha at top height 6m, thin to 200 stems/ha at top height 12m.

All values are at age 30 years, mean site index 28.5m.

Table 1 : Model and Regime Comparison

	VARIABLE	*LAW*	*NEL*	*GDN*
Regime 1:	Top height (m)	39.1	39.1	38.9
Plant 1250	Stocking (stems/ha)	1055	1055	797
Unthinned	Basal area (m^2/ha)	102.6	96.5	84.3
	Mean DBH (cm)	35.2	34.1	36.7
	Volume (m^3/ha)	1244	1170	1032
	VARIABLE	*LAW*	*NEL*	*GDN*
Regime 2:	Top height (m)	39.1	39.1	38.9
Plant 1100	Stocking (stems/ha)	553	553	443
Thin_600	Basal area (m^2/ha)	76.8	74.1	66.3
	Mean DBH (cm)	42.0	41.3	43.7
	Volume (m^3/ha)	944	911	816

Table 1 : Continued

	VARIABLE	*LAW*	*NEL*	*GDN*
Regime 3:	Top height (m)	39.1	39.1	38.9
Plant 1250	Stocking (stems/ha)	197	197	171
Thin_600	Basal area (m^2/ha)	48.1	48.0	43.9
Thin_200	Mean DBH (cm)	55.8	55.7	57.1
	Volume (m^3/ha)	612	610	548

There is no difference between the new Nelson model and the Law model in the 200 stems/ha regime. This is to be expected since the basal areas are below the point where over-prediction in the original model occurred. For Regime 2, (600 stems/ha), there is a decrease of 3.5% in basal area, 5.5% in mean DBH, and 3.5% in total stand volume at age 30. The unthinned regime shows a slightly greater decrease of 5.9% in basal area, 7.1% in DBH and 5.9% in volume.

Table 2 gives the new residual mean square values for Nelson as well as those for the Kaingaroa, Canterbury and Hawkes Bay models. It can be seen that the Nelson values are particularly low for stocking and (not surprisingly) mean DBH, and are good for top height, basal area and volume in comparison with other regions.

Table 2 : Residual Mean Squares *

	NELSON	KGM3	CANT	HBAY
Height	0.65	0.54	0.69	0.85
Barea	1.61	1.00	2.09	1.98
Stock	14.30	36.10	33.80	32.14
Volume	44.72	34.59	62.80	61.04
Mean DBH	0.16	0.55	0.72	0.89

* The mean residual (or deviation) is $\pm$ the square root of the RMS value.

WARNING

It is important to remember that unlike the other growth models developed using this methodology, **no windthrow** is included in the database. If, as appears to be the case, wind is a more significant factor in Nelson than elsewhere, forest managers and users of the model should make suitable allowances by reducing the predicted stockings and volumes as appropriate for a given age.

ACKNOWLEDGEMENTS

I would like to thank Ms Kirsty Law for her assistance in a review of much of her original work and Dr Oscar Garcia for his guidance in reviewing the methodology and for providing the graphics program to enable a detailed study of the data. I am also grateful to the Stand Growth Model Cooperative Technical Committee for their patience over the duration of this project.

REFERENCES

DUNNINGHAM, A.G.; LAWRENCE, M.E. 1987: An "interim" stand growth model for radiata pine grown on the Central North Island pumice plateau. **Ministry of Forestry, Forest Research Institute, Project Record No.1497 (Unpublished).**

GARCIA, O.P. 1979: Modelling stand development with stochastic differential equations. Pp 315-7 in Elliott, D.A. (Ed.) "Mensuration for Management Planning of Exotic Forest Plantations". **New Zealand Forest Service, Forest Research Institute Symposium No.20.**

GARCIA, O.P. 1983: A stochastic differential equation model for the height growth of forest stands. **Biometrics 39:1059-72.**

GARCIA, O.P. 1984a: New class of growth models for even-aged stands - *Pinus radiata* in Golden Downs Forest. **New Zealand Journal of Forestry Science 14(1):65-8.**

GARCIA, O.P. 1984b: Performance of the Golden Downs growth model on low-stocking stands. **Ministry of Forestry, Forest Research Institute, Project Record No.575 (Unpublished).**

GARCIA, O.P. 1987: Mortality in the Golden Downs model. **Ministry of Forestry, Forest Research Institute, Project Record No.1457 (Unpublished).**

LAW, K.R.N. 1990: A growth model for radiata pine grown in the Nelson/Marlborough region. **Ministry of Forestry, Forest Research Institute, Project Record No.2381 (Unpublished).**

LAWRENCE, M.E. 1988: A growth model for radiata pine grown in Canterbury. **Ministry of Forestry, Forest Research Institute, Project Record No.1862 (Unpublished).**

APPENDIX 1 : PLOT DISTRIBUTION IN THE NELSON GROWTH MODEL

NZ Timberlands, Nelson

Forest	*No.*	*Forest*	*No.*	*Forest*	*No.*
Golden Dns	161	Motueka	27	Rai Valley	17
Hira	24	Wairau	46	Tutaki	2

Baigent Forests

Harakeke	6	Sth Pigeon	11	Waiwhero	8
Greenhill	15	Dovedale	20	Orinoco	2
Trass	12	Mahana	17	Hoults	8
Waimea West	2	Nth Pigeon	3	Sunrise	4
Tasman	1	Sunset	3	Redwoods	1
Riwa	1	Pece	3	Mariri	3
Lee	10	Richmond	3		

Total number of forests: 26

Total number of plots : 410

APPENDIX 2 : LIST OF PLOTS USED IN THE DEVELOPMENT OF THE NELSON GROWTH MODEL

Mapua

1500000100.	4960007108.	4980000100.
4960007111.	4960007242.	4980000200.
4960007113.	4960007447.	4980000300.
4960007114.	4960008058.	4980000400.
4960007124.	4960008059.	4980000500.
4960007226.	4960008069.	4980000600.
4960007228.	4960008083.	4980000700.
4960007229.	4960008085.	4980000800.
4960007231.	4960008181.	4980000900.
4960007232.	4960008291.	4980001000.
4960007233.	4960008402.	4980001100.
4960007241.	4960008403.	4980001200.

Wairau

4440000400.	4690007704.	4690007905.
4440001100.	4690007707.	4690007906.
4500008402.	4690007708.	4690008301.
4500008403.	4690007709.	4840000100.
4520215001.	4690007710.	4840000200.
4520235001.	4690007711.	4840000300.
4520325002.	4690007713.	4840000400.
4520325003.	4690007714.	4840000500.
4520325004.	4690007716.	4840000600.
4520339001.	4690007717.	4840000700.
4520415002.	4690007718.	4840000800.
4520435002.	4690007719.	4840000900.
4520525007.	4690007720.	4840001000.
4690007702.	4690007721.	4840001100.
4690007702.	4690007901.	4840001200.
4690007703.	4690007902.	

Lowfert

870000400.	1950000300.	1950001900.
870000800.	1950000500.	1950002100.
1410400700.	1950000700.	1950002300.
1930000100.	1950000900.	4210000400.
1930000500.	1950001000.	4210001000.
1930000700.	1950001200.	4210001400.
1930000900.	1950001400.	4210002600.
1930001000.	1950001500.	5140200300.
1930001500.	1950001700.	5140200400.

Moutere

790000200.	3100000600.	4930001200.
790001300.	3100000700.	4930001300.
790001700.	3100000800.	4930001400.
790001800.	3100000900.	4960007116.
790002000.	3100001000.	4960007117.
790005500.	3100001100.	4960007118.
790005600.	3100001200.	4960007119.
790005900.	3100001300.	4960007120.
790006000.	3100001400.	4960007122.
790006100.	3100001500.	4960007123.
790006200.	3240000100.	4960007203.
790006300.	3240000200.	4960007204.
790006400.	3240000300.	4960007205.
1510000100.	3240000400.	4960007207.
1510000200.	3240000600.	4960007209.
1510000300.	3240000800.	4960007215.
1510000400.	3240000900.	4960007225.
1620000100.	3240001100.	4960007230.
1620000200.	3240001200.	4960007235.
1620000300.	3710000100.	4960007236.
1620000400.	3710000200.	4960007237.
1620000500.	3710000300.	4960007239.
1620000600.	3710000400.	4960007243.
1620000700.	3710000500.	4960007345.
1620000800.	3710000600.	4960007346.
2390000300.	3710000700.	4960008061.
2390000500.	3710000800.	4960008062.
2390000600.	3790000200.	4960008063.
2620000100.	3790000500.	4960008064.

2620000300.	3790001200.	4960008065.
2620000500.	3920000100.	4960008067.
2670000100.	3920000800.	4960008068.
2670000200.	3920001100.	4960008070.
2670000300.	4140001400.	4960008071.
2670000400.	4460006802.	4960008072.
2670000500.	4460006803.	4960008074.
2670000600.	4460006804.	4960008075.
2670000700.	4460006806.	4960008076.
2670000800.	4460006807.	4960008078.
2670000900.	4460006816.	4960008079.
2780000100.	4460006828.	4960008080.
2780000200.	4460007502.	4960008082.
2780000300.	4460007503.	4960008293.
2780000400.	4460007504.	4960008294.
2780000500.	4460007505.	4960008396.
2780000600.	4460007506.	4960008397.
2780000700.	4460007507.	4960008404.
2780000800.	4460007609.	5140100100.
2780000900.	4460007610.	5140100200.
2780001000.	4460007611.	5140100500.
2780001100.	4460007612.	5140100600.
2780001200.	4460007613.	5140100700.
2780001300.	4460007614.	5140100800.
2990005001.	4460007615.	5140100900.
2990005002.	4460007701.	5140101000.
2990010001.	4460007702.	5140101100.
2990010002.	4460007708.	5140101300.
2990020001.	4460007710.	5140101500.
2990020002.	4460007711.	5140101600.
2990030001.	4460007814.	5140101700.
2990030002.	4460007844.	5140101900.
3000000100.	4460007866.	5140102000.
3000000200.	4460007878.	5140102100.
3000000300.	4930000100.	5460000400.
3000000400.	4930000200.	5460001000.
3000000500.	4930000300.	5460001500.
3000000600.	4930000400.	5460001600.
3000000800.	4930000600.	5460001700.
3000000900.	4930000700.	5460001900.
3100000100.	4930000800.	5740000100.
3100000200.	4930000900.	5740000200.

3100000300.	4930001000.
3100000400.	4930001100.
3100000500.	

<u>Richmond</u>

1830000100.	4620007201.	4960008184.
1840000100.	4620007202.	4960008186.
2400000300.	4620007801.	4960008289.
3760000100.	4620007804.	4960008290.
3760000200.	4620007805.	4960008295.
3760000300.	4620007806.	4960008398.
3760000400.	4620007807.	4960008401.
3760000500.	4620007810.	4960008406.
3760000600.	4620008303.	4960008407.
3760000700.	4860008101.	4960008499.
3760000800.	4860008102.	4960008509.
3760000900.	4860008103.	4960008610.
3760001000.	4860008104.	4960008611.
4620006901.	4860008105.	5730000200.
4620006903.	4860008106.	5750000300.
4620006904.	4860008107.	5750000500.
4620006905.	4860008108.	5750001200.
4620006908.	4860008309.	5750001300.

<u>Highfert</u>

3110000100.	4910000600.	4960008054.
3110000200.	4910000700.	4960008055.
3110000300.	4910000800.	4960008056.
3110000400.	4960007101.	4960008057.
3110000500.	4960007102.	4960008187.
4910000100.	4960007234.	4960008188.
4910000200.	4960008051.	4960008292.
4910000300.	4960008052.	5140101200.
4910000400.	4960008053.	5140101400.
		5140101700.

NELSON GROWTH MODEL
"NM90"
(1990)

K.R.N. Law
M.E. Lawrence

1. DATA

All the data are taken from the Ministry of Forestry's permanent sample plot system. This data set includes measurements from forests owned by NZ Timberlands Ltd and Baigent Forest Industries Ltd in the Nelson/Marlborough region.
1655 measurement pairs from 410 plots are used in the development of this model. Table 1. describes the range of data used in the model.

Table 1. Range of data

	Minimum	Mean	Maximum
Age (years)	3.0	27.4	46.7
Site index (metres)	18.7	28.5	35.7
Basal area (m2/ha)	0.3	24.3	86.8
Stocking (stems/ha)	110	597	4075
Top height (m)	2.0	38.2	49.0

2. SCOPE

The model may be used to derive yield tables and predict the effects of different management regimes on stands in the Nelson/Marlborough region. All ages, as defined by the database, and thinning treatments are covered by the model.

3. VALIDATION AND ACCCURACY

As illustrated in the full report describing this model (Law, 1990), there has been extensive plotting and analysis of residuals. A tendency to over-predict at high basal areas (>50m^2/ha) has now been rectified (Lawrence, 1990) and the program altered accordingly. The new residual mean squares (RMS) are shown in Table 2.

Table 2. Residual Mean Squares

	RMS
Basal area (m2/ha)	1.61
Stocking (stems/ha)	14.30
Top height (m)	0.65
Basal area * height (m3/ha)	44.72
Mean DBH (cm)	0.163
Average spacing (m)	0.024

4. INPUT DATA REQUIRED

Input data required includes at least 2 of the following : age, top height or site index. The user is also prompted for initial stocking and basal area, one of which must be known. If the effects of thinning are to be modelled then stocking before and after thinning must be known as well as the age or height when thinned.

5. PRUNING

The model does not distinguish between pruned and unpruned trees in the data set.

6. METHODOLOGY

The methodology used is the same as that in the Golden Downs Model. Refer to Garcia, 1984 in Section 7.

7. REFERENCES

GARCIA, 0. 1984: New class of growth models for even-aged stands - *Pinus radiata* in Golden Downs forest. *New Zealand Journal of Forestry 24(1): 108-124.*

LAW, K.R.N. 1990: A growth model for radiata pine grown in the Nelson/Marlborough region. *Ministry of Forestry, Forest Research Institute, Project Record No. 2381 (unpublished).*

LAWRENCE, M.E. 1990: Investigations into the possible causes of bias in the Nelson growth model. *Ministry of Forestry, Forest Research Institute, Project Record No. 2611 (unpublished).*