韓國林學會誌 87(1) : 40 - 49. 1998
Jour. Korean For. Soc. 87(1) : 40 - 49. 1998

Selection of Growth Projection Intervals for Improving Parameter Estimation of Stand Growth Model[1]

Sang Hyun Lee[2]

林分 生長 모델의 母數 推定 能力 向上을 爲한 生長 測定間隔의 選擇[1]

李 尙 鉉[2]

ABSTRACT

This study aimed to provide a strategy for selecting an adequate combination of growth intervals(i. e. times between age T_1 and age T_2) to be used to improve the reality of the growth equation through obtaining better precision of parameter estimates. Variety of growth functions were fitted to the data and one equation which best fitted the data was chosen for the analysis. A modified Schumacher projection equation, selected as a best equation, that included dummy variables representing locality as a predictor variable was fitted for basal area and height equations with nonoverlapping growth interval and all possible growth interval data sets of Douglas-fir(*Pseudotsuga menziesii* Mirb.Franco). The data were measured in all parts of the South Island of New Zealand. It was found that the precision of parameter estimates was increased in both basal area and height equations by using data set which contained a range of measurement intervals from short to long term.

Key words: *Dummy variables, Basal area model, Mean top height model,* Pseudotsuga menziesii *Mirb.Franco, Schumacher projection equation.*

要 約

본 연구는 보다 정확한 母數 추정을 통한 生長모델의 현실성을 향상시키는데 이용되는 생장 측정간격(임목의 측정 초기 연령 T_1과 재측정 연령 T_2의 기간)의 적합한 조합을 선택하기 위한 계획을 제공하는데 목적이 있다.

다양한 생장식을 데이터에 적용한 후 가장 적합한 것으로 판정된 생장식을 분석에 이용하였다. 여러 생장식을 분석한 결과 최적의 생장식으로 판명된 더미 변수를 포함하는 변형 Schumacher 방정식을 임분 胸高斷面積 생장식과 平均樹高 생장식을 얻기 위하여 이용하였다. 그리고 사용된 자료는 뉴질랜드 남섬 전역에서 측정된 美松(*Pseudotsuga menziesii* Mirb.Franco)의 생장 측정기간이 변형되지 않은 데이터와 모든 가능한 생장 측정기간을 포함하는 변형된 2종류의 데이터이었다. 단기의 측정기간에서부터 장기의 측정기간의 범위를 포함하는 데이터(모든 가능한 생장 측정기간을 포함하는 데이터)를 사용할 때 흉고단면적 생장식과 임분 평균수고 생장식에서 모수 추정의 정확성이 증가되는 것이 발견되었다.

[1] 接受 1997年 12月 2日 Received on December 2, 1997.
[2] School of Forestry, University of Canterbury, Private Bag 4800, Christchurch, New Zealand

INTRODUCTION

Forest growth is defined as the increase in diameters of one or more individuals in a forest stand over a given period time(e. g. growth in m^3/ha/year). Yield refers to the total production over a given time period(e. g. volume in m^3/ha). Yield is the directly related to growth in that it is the accumulated growth to a specified point in time, and can be derived mathematically by integrating a growth function. The derivative of a yield function will therefore give a growth function(Vanclay, 1994).

The general purpose of growth and yield models, despite their complexity of structure, can be explained simply as ; given a set of stand or tree characteristics, such as basal area and stems per hectare at one point in time (T_1), to predict by how much these characteristics will have changed at a future time (T_2) given specified stand or tree treatments.

Growth interval data obtained from remeasured permanent plots or trees are referred to as repeated measurements or a real growth series. The term means that *N* experimental subjects are observed in each of *k* successive occasions that possibly correspond to different experimental conditions, the i_{th} subject yielding y_{ij} on the j_{th} occasion. Such repeated measurements data can be used for investigating tree dynamics as well as for modeling growth.

The effective expression of the existing relationship between growth and yield was first presented by Clutter(1963) and has been referred to algebraic difference equation(Borders et al., 1984). That is

$$Y_2 = f(Y_1,\ T_1,\ T_2,\ \theta,\ MR) \qquad (1)$$

where

Y_2 = value of a continuous variable derived for a tree or stand at age T_2
Y_1 = value of the same variable at initial measurement
T_1 = tree or stand age at initial measurement
T_2 = tree or stand age at next re-measurement
θ = set of parameters of equation
and MR = management regime.

In this approach, at a given time, the future state of the variables and the transition functions or changes in the state variable are a function of the initial state of the variable, time elapsed, management inputs and prevailing environment. In using projection equation of this function form, growth series data can be used in fitting routines to obtain sample estimates of parameters of equation that best described the growth and yield of the selected stand variables.

The most common procedure for establishing parameters in algebraic differential equations is to use only the nonoverlapping growth interval (i.e. A1 to A2, A2 to A3, --- An - 1 to An) (Borders et al., 1984 ; Cluter, 1963 ; Sullivian and Clutter, 1972). Another possibility is to use all possible intervals for each unit. If permanent sample plots have been measured n times there are ${}_nC_2$ combinations of different intervals between time T_1 and T_2 that can be derived and used to build equations.

This paper was intended to compare and contrast the two data sets formulations namely, Nonoverlapping growth interval and all possible growth interval, to provide a strategy for selecting an appropriate combination of projection intervals.

MATERIALS AND METHODS

Data for this study came from a large database of Douglas-fir(*Pseudotsuga menziesii* Mirb. Franco) permanent sample plots maintained by New Zealand Forest Research Institute. The data were measured in all parts of the South Island of New Zealand. All of 355 permanent sample plots ranging from 0.02 to 0.04 ha and 1844 sets of measurements were used in this analysis. Table 1 is the summary of the mean and extreme values of age, mean top height and basal area per hectare of the raw data.

Table 1. Summary of mean and extreme values extracted from permanent sample plots data

Region	No. of observations	Variable*	Mean	Minimum	Maximum
Canterbury	241	Age(years)	32.8	9.0	61.0
		H_{100}(m)	22.9	2.9	39.3
		Altitude(m)	326.1	150.0	790.0
		G(m^2/ha)	46.3	0.43	116.2
Nelson	929	Age(years)	27.5	7.0	58.0
		H_{100}(m)	22.9	5.6	47.8
		Altitude(m)	438.1	183.0	625.0
		G(m^2/ha)	42.2	1.2	109.4
Southland	449	Age(years)	33.6	7.0	78.0
		H_{100}(m)	24.1	4.1	47.4
		Altitude(m)	251.1	50.0	625.0
		G(m^2/ha)	51.3	1.1	141.7
Westland	225	Age(years)	26.9	5.0	59.1
		H_{100}(m)	18.8	1.9	37.5
		Altitude(m)	229.0	0.0	330.0
		G(m^2/ha)	29.9	0.01	123.8
Sum	1844	–	–	–	–

* H_{100} = mean top height of the stand
G = net basal area of the stand

The two data sets were created from raw data base namely, nonoverlapping interval data set that consisted of about 1600 sets of measurements and all possible interval data set which had 6173 sets of measurements in through lag and put statements in Statistical Analyses System(SAS).

Prior to any model estimation the data were verified and screened to ascertain the reliability of the data. Examples of data validation included that ensuring number of stems/ha at age_2 were not greater than that of at age_1, basal area/ha at age_2 were greater than that of at age1 and age_2 were greater than age_1. Residuals were also used to detect outliers. Outliers were observations that had residuals greater than ±3.5 standard deviation from zero(Xu, 1990). Standard residuals Si defined as

$$S_i = \frac{\text{Residual}}{\text{MSE}} \qquad (2)$$

where in (2)

Residual=(Observed value) - (Fitted value) and
MSE = Residual mean square

As a rule, in all the model fitting routines, observations which have value of S_i greater than ±3.5 were considered to be outliers.

The main standard analytical procedures used was non-linear ordinary least-squares regression. Regression equation can be fitted variables sets of any sort, but in this study it was ensured that the dependent and independent variables conform to biologically and mathematically sound relationship, the functions used were of appropriate form to represent the independent relationship and parameter estimates were free of apparent bias.

In order to apply any regression to data sets of variables, it is assumed that the residual errors were independent, had a mean of zero and constant variance, and plot of residuals followed a normal distribution(Sokal and Rohlf, 1981 ; Draper and Smith, 1981). Variety of sigmoid shaped functions were applied to the data sets using the

PROC NLIN procedure of the SAS package(SAS institute Inc., 1990) and the derivative free algorithm method for non-liner least squares(Raltson and Jennrich, 1978) to find which function could be used to compare the two data sets.

Anamorphic and polymorphic forms of Schumacher(Clutter et al., 1983), Chapman-Richard(Pienaar and Turnbull, 1973), Hossfeld(Xu, 1990) and Gompertz equations were fitted to two data sets. The equation formulations used are listed in Table 2.

The residuals resulting from fitting each of these equations were analyzed using following methods :

1. comparison of mean square errors(MSE) ;
2. examination of plots of residuals against predictor variables and predicted values to provide ocular estimates of their normality of errors ; and
3. comparison of extreme deviation and moments of the residual on the assumption that they should be normally and independently distributed with mean zero and constant variance δ^2.

PROC UNIVARIATE(SAS institute Inc., 1990) procedure was also used to ascertain the goodness of the equations and normality because this procedure provides a wide range of statistics to supplement the residual patterns inferences. After fitting equation chosen as a best model with two data sets, coefficients of parameters were compared.

RESULTS AND DISCUSSION

Two compatible projection equations for basal area/ha and mean top height were derived and compared.

1. Basal Area Model

Most of the anamorphic equations were found large bias in residuals pattern. The Chapman-Richard functions displayed difficulty in convergence and once parameters were estimated bias was noted in the graphical representation of residuals. The coefficients of the general equations fitted to the data are presented in Table 3 and 4 with respective mean square error(MSE) value.

Since the equation with the least biased residual plots was found the lowest MSE value, the values of this in Table 3 and 4 were used to indicate the best fitting equation. The Schumacher polymorphic function with MSE of 4.15 and 19.10 for nonoverlapping and all possible data sets respectively, were found to give a better fit than the rest of the equations. This equation, therefore, was considered for further examination.

After trying numerous modifications to the Schumacher equation, with the addition and sub-

Table 2. Equation forms applied to data sets

Equation name	Equation Forms*
Hossfeld Polymorphic	$Y_2 = 1/((1/Y_1)\ (T_1/T_2)^{\gamma} + (1/\alpha)\ (1 - (T_1/T_2)^{\gamma}))$
Hossfeld Anamorphic	$Y_2 = 1/((1/Y_1) + \theta\,(1/T_2^{\beta} - T1^{\beta}))$
Schumacher Polymorphic	$Y_2 = \exp(\mathrm{In}(Y_1)\ (T_1/T_2\ )^{\beta} + \alpha\,(1 - (T_1/T_2)^{\beta}))$
Schumacher Anamorphic	$Y_2 = Y_1\ \exp(\beta\,(1/T_1^{\alpha} - 1/T_2^{\alpha}))$
Chapman-Richards Polymorphic	$Y_2 = (\alpha / \gamma)^{[1/(1-\beta)]}(1 - (1 - (\gamma / \alpha)Y_1^{(1-\beta)})\ \exp(-\gamma\,(1-\beta)(T_2 - T_1))^{[1/(1-\beta)]}$
Chapman-Richards Anamorphic	$Y_2 = Y_1((1 - \exp(-\beta T_1)) / (1 - \exp(-\beta T_2)))^{1/(1-\gamma)}$
Gompertz polymorphic	$Y_2 = \exp(\mathrm{In}\ (Y_1)\ \exp(-\beta\,(T_2 - T_1) + \gamma\,(T_2^2 - T_1^2)) + \alpha\,(1 - \exp(-\beta\,(T_2 - T_1) + \gamma\,(T_2^2 - T_1^2))))$

* Y_1 = net basal area of the stand or mean top height at age T_1
Y_2 = net basal area of the stand or mean top height at age T_2
α, β, γ and θ are parameters to be estimated

Table 3. Coefficients for general equation fitted to nonoverlapping basal area data

Model Name	Parameters				MSE
	α	β	γ	θ	
Hossfeld Polymorphic	112.65	-	2.53	-	4.77
Hossfeld Anamorphic	-	1.48	-	1.36	21.79
Schumacher Polymorphic	5.11	1.08	-	-	4.15
Schumacher Anamorphic	0.81	19.94	-	-	8.60
Chapman-Richards Polymorphic	1.05	0.36	0.04	-	8.87
Chapman-Richards Anamorphic	-	0.002	1.79	-	24.20
Gompertz Polymorphic	4.93	0.09	0.006	-	5.83

Table 4. Coefficients for general equation fitted to all possible basal area data

Model Name	Parameters				MSE
	α	β	γ	θ	
Hossfeld Polymorphic	100.45	-	2.84	-	21.11
Hossfeld Anamorphic	-	1.66	-	1.72	189.35
Schumacher Polymorphic	4.98	1.17	-	-	19.10
Schumacher Anamorphic	0.82	18.62	-	-	81.88
Chapman-Richards Polymorphic	0.91	0.48	0.07	-	29.12
Chapman-Richards Anamorphic	-	0.004	1.59	-	191.97
Gompertz Polymorphic	4.79	0.09	0.006	-	20.02

traction of various predictor variables and alteration to the equation form, the modified Schumacher polymorphic equation (3) that included dummy variables representing locality was found to give the best fit for the both two data sets.

$$G_2 = G_1(T_1/T_2)^{\beta}((\alpha + \beta_1 k1 + \beta_2 k2 + \beta_3 k3)(1-(T_1/T_2)^{\beta}) \quad (3)$$

where in (3)

G_2 = net basal area of the stand(m^2/ha) at T_2
G_1 = net basal area of the stand(m^2/ha) at T_1
T_1 = age in years at the beginning of a growth periods
T_2 = age in years at the end of a growth period
K1, K2 and K3 = dummy variables for region
α, β, β_1, β_2 and β_3 = parameters to be estimated.

The goodness of the fit was evaluated through plots of residuals against predicted values as shown Fig. 1 and 2. And both are acceptable with no apparent bias.

Parameter estimates were slightly different between the two cases as shown in Table 5. The total regression and residual sums of square are seen to be much less for nonoverlapping intervals. This arises because there was only an annual interval in that case, while the all possible intervals had much wider scatter.

The t statistic, estimated by dividing the parameter estimate by asymptotic standard error, was much greater for all parameters included in the all possible interval model. That t statistic test provides indicative guidelines on the relative precision. In this case the all possible interval model estimates provide a better fit to the data. In addition, the confidence intervals for both models show that all estimates are significant at $p<0.05$ as none includes zero, the 95% confidence intervals for the all possible model are much tighter.

2. Mean Top Height Model

A range of growth equations in different form were fitted to the data sets and analyzed. Anamorphic Schumacher, polymorphic Schumacher and polymorphic Hossfeld equations were found

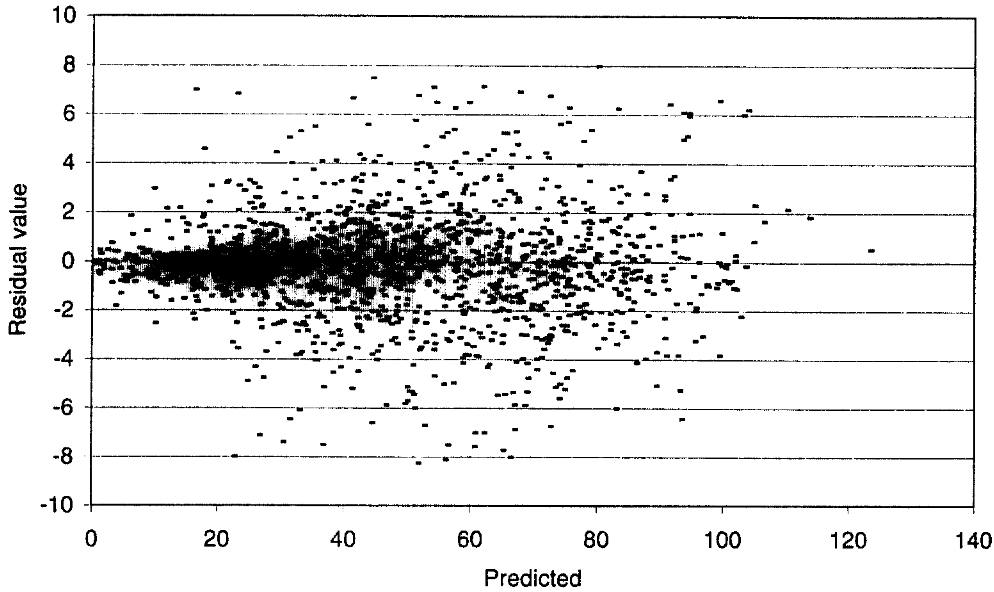

Fig. 1. Plot of residual Vs predicted for nonoverlapping intervals basal area equation

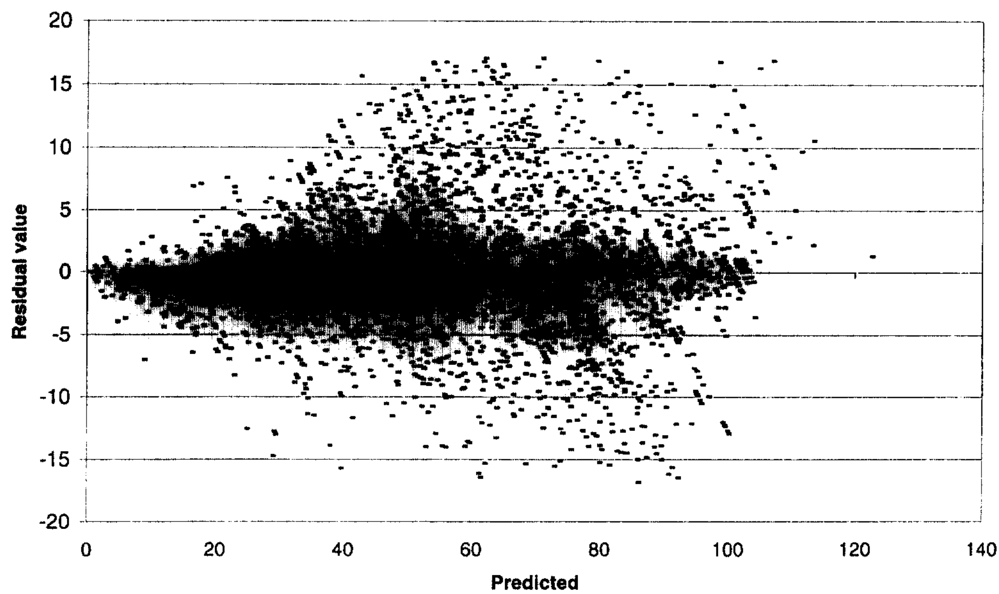

Fig. 2. Plot of residual Vs predicted for all possible intervals basal area equation

to give a better fit than rest of the equations based on analysis of residuals. The coefficients and the residual mean square error for seven candidate equations are presented in the Table 6 and 7.

The Schumacher polymorphic equation with mean square error(MSE) of 0.59 and 0.96 for nonoverlapping and all possible data sets respectively, were found to give the best fit after comparing residual patterns, mean residual error and PROC UNIVARIATE statistics. Therefore, this equation was chosen for further improvement by incorporating several explanatory variables in a logical manner. After modification of this equa-

Table 5. Non linear least squares summary statistics for basal area models

SOURCE	Degree of Freedom		Sums of Squares		Mean Squares	
	Nonover-lapping	All possible	Nonover-lapping	All possible	Nonover-lapping	All possible
Regression	5	5	4,508,417	1,7528,522	901,683	3,505,704
Residual	1,794	6,105	6,414	104,319	3.5755	17.088
Uncorrected total	1,799	6,110	4,514,831	1,7632,842		
Corrected total	1,798	6,109	1,017,708	3,650,294		

Parameter	Estimates		Std.Error / t statistics		95% Confidence Interval			
	Nonover-lapping	All possible	Nonover-lapping	All possible	Nonover-lapping		All possible	
					Lower	Upper	Lower	Upper
α	4.9149	4.7274	0.0268 / 183.12	0.0166 / 284.05	4.86	4.97	4.69	4.76
β	1.0801	1.1904	0.0183 / 59.08	0.0083 / 143.95	1.04	1.12	1.17	1.20
β_1	0.1293	0.2143	0.0225 / 5.75	0.0158 / 13.55	0.09	1.17	0.18	0.24
β_2	0.3201	0.3506	0.0224 / 14.26	0.0167 / 21.02	0.28	0.36	0.32	0.38
β_3	0.3120	0.4063	0.0296 / 10.53	0.0210 / 19.39	0.25	0.37	0.37	0.45

Table 6. Coefficients for general equation fitted to nonoverlapping mean top height data

Model Name	Parameters				MSE
	α	β	γ	θ	
Hossfeld Polymorphic	74.80	-	1.39	-	0.61
Hossfeld Anamorphic	-	1.27	-	2.25	0.96
Schumacher Polymorphic	5.63	0.38	-	-	0.59
Schumacher Anamorphic	0.32	8.67	-	-	0.64
Chapman-Richards Polymorphic	0.56	0.32	0.03	-	0.69
Chapman-Richards Anamorphic	-	0.001	1.83	-	3.08
Gompertz Polymorphic	4.16	0.05	0.0003	-	0.67

Table 7. Coefficients for general equation fitted to all possible mean top height data

Model Name	Parameters				MSE
	α	β	γ	θ	
Hossfeld Polymorphic	77.04	-	1.44	-	0.98
Hossfeld Anamorphic	-	1.27	-	2.12	2.05
Schumacher Polymorphic	5.69	0.39	-	-	0.96
Schumacher Anamorphic	0.31	8.90	-	-	1.10
Chapman-Richards Polymorphic	0.71	0.28	0.04	-	1.43
Chapman-Richards Anamorphic	-	0.003	1.64	-	21.37
Gompertz Polymorphic	4.19	0.05	0.003	-	1.22

tion through adding and subtracting of explanatory variables, a modified Schumacher polymorphic equation (4) that include dummy variables representing locality gave the best fit for two data sets.

$$H_{100,2} = H_{100,1}(T_1/T_2)^{\beta}((\alpha + \beta_1 k1 + \beta_2 k2)(1-(T_1/T_2)^{\beta}) \quad (4)$$

where in (4)

$H_{100,2}$ = mean top height in meters at age T_2
$H_{100,1}$ = mean top height in meters at age T_1
T_1 = age in years at the beginning of a growth periods
T_2 = age in years at the end of a growth period
K1, K2 and K3 = dummy variables for region
α, β, β_1 and β_2 = parameters to be estimated.

Altitude and dummy variables were found to improve the model when they were introduced independently to the basic Schumacher equation. Though altitude has been found to be an important variable for explaining variations in mean top height growth(Woollons and Hayward, 1985 ; Mason, 1992), it was not included in the finial formulations. The reason is that the modification of the model through adding and subtracting of these two variables was not superior to equation including only dummy variables. Parameter estimates for both equations are summarized in Table 8.

All the parameter estimates were significant at least 5% level. Fig. 3 and 4 show the plot of residuals against predicted values. Both models were tested in terms of actual observation minus predictions for the data sets. The precision achieved in those overall equations was better than any other models.

The residuals about the predicted values never exceed $\mp$4.0m. The mean residual values for the nonoverlapping intervals model was 0.0335 about one third greater than that for all possible intervals model at -0.011. The ideal would be a mean residual value of 0. Skewness for the all possible data form had a value of 0.1492 compared with 0.2497 for the nonoverlapping form. The ideal would, of course, be a skewness value of 0. The t statistic of all possible intervals equation was also much greater than that of nonoverlapping intervals equation.

CONCLUSION

This research showed that use of the nonoverlapping data set form to build growth and yield models was efficient, appropriate and resulted in estimates of parameters that were generally precise. However, The more precise estimates of parameters were achieved when more efficient mix of projection intervals was used. This research provides some positive evidences that by using data sets which contains a range of $Age_{(T2)} - Age_{(T1)}$ time intervals, as opposed to the more common method of using annual or bi-

Table 8. Non linear least squares summary statistics for mean top height model

SOURCE	Degree of Freedom		Sums of Squares		Mean Squares	
	Nonoverlapping	All possible	Nonoverlapping	All possible	Nonoverlapping	All possible
Regression	4	4	1,018,458	3,896,357	254,614	974,089
Residual	1,681	5,790	990	5,325	0.5894	0.9197
Uncorrected total	1,685	5,794	1,019,449	3,901,682		
Corrected total	1,684	5,793	132,944	412,275		

Parameter	Estimates		Std.Error / t statistics		95% Confidence Interval			
	Nonoverlapping	All possible	Nonoverlapping	All possible	Nonoverlapping		All possible	
					Lower	Upper	Lower	Upper
α	5.2913	5.5352	0.1250 / 42.34	0.0415 / 133.39	5.05	5.54	5.45	5.62
β	0.4101	0.3867	0.0210 / 19.95	0.0058 / 67.02	0.37	0.45	0.38	0.40
β_1	0.2178	0.1956	0.0392 / 5.56	0.0161 / 12.13	0.14	0.29	0.15	0.23
β_2	0.1693	0.0408	0.0443 / 3.82	1.0187 / 2.17	0.08	0.26	0.01	0.08

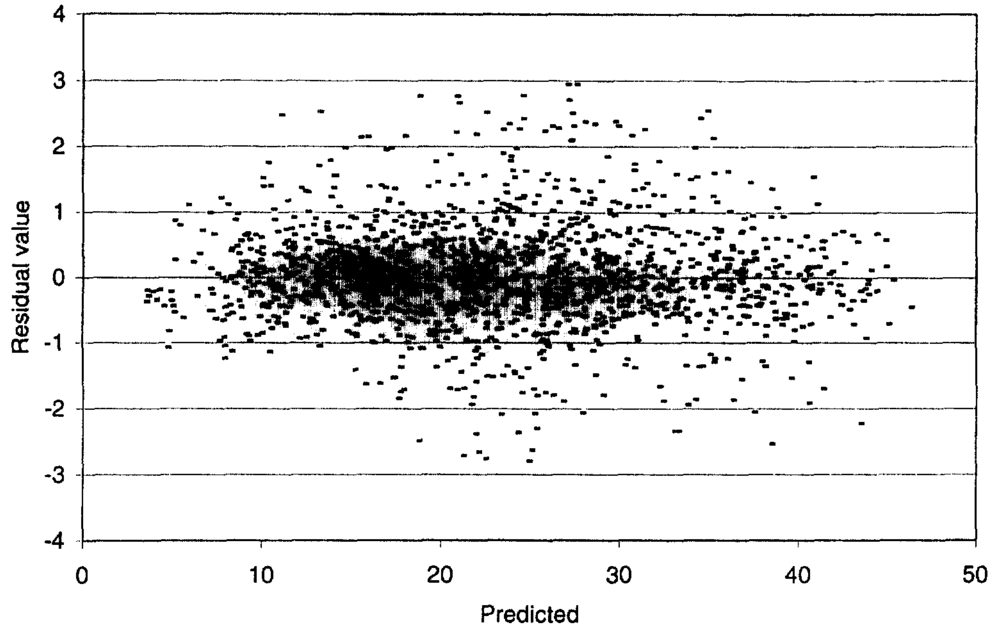

Fig. 3. Plot of residual Vs predicted for nonoverlapping intervals mean top height equation

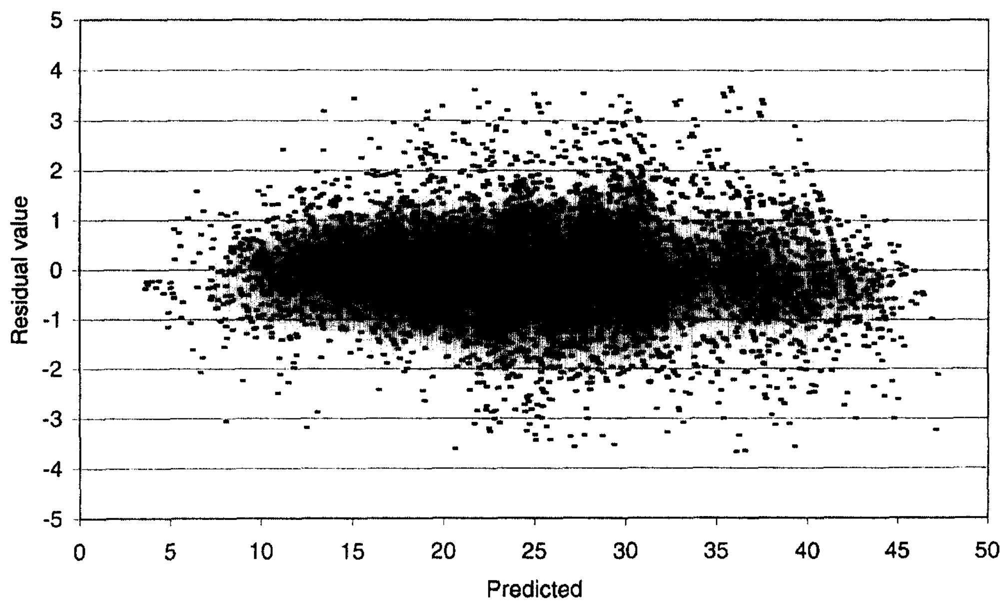

Fig. 4. Plot of residual Vs predicted for all possible intervals mean top height equation

annual intervals, the precision of parameter estimates could be increased.

ACKNOWLEDGEMENTS

Acknowledgements are expressed to the NZ Forest Research Institute for providing data and information.

LITERATURE CITED

1. Borders, B.E., R.L. Bailey and K.D.

Ware. 1984. Slash pine site index from a polymorphic model by joining(splining) non-polynomial segments with an algebraic difference method. Forest Science 30 : 411-423.
2. Clutter, J.L., C. Forsion, L.V. Pienaar, G.H. Brister and R.L. Bailey. 1983. Timber management ; a quantitative approach. John Wiley and Sons, New York. 333pp.
3. Clutter, J.L. 1963. Compatible growth and yield models for loblolly pine. Forest Science 9 : 354-371.
4. Draper, N.R. and H. Smith. 1981. Applied regression analysis. Wiley, New York. 709pp.
5. Mason, E.G. 1992. Decision support systems for establishing radiata pine in Central Northland of New Zealand. Ph.D Thesis. University of Canterbury. 303pp.
6. Pienaar, L.V. and K.J. Turnbull. 1973. The Chapman-Richards generalization of Von Bertalanffy's model for basal area growth and yield in even -aged stands. Forest Science 19 : 2-22.
7. Ralston, M.L. and R.I. Jennrich. 1978. Dud, a derivative free algorithm for nonlinear least squares. Technometrics 20(1) : 7-13.
8. SAS Institute Inc. 1990. SAS/STAT User's guide, Version 6. fourth edition. Cary, NC.
9. Sokal, R.R. and F.J. Rohlf, 1981. Biometry. The principles and practice of statistics in Biological research. Second edition, W.H. Freeman and Company. New York. 859pp.
10. Sullivan, A.D. and J.L. Clutter. 1972. A simultaneous growth and yield model for loblolly pine. Forest Science 18 : 76-86.
11. Vanclay, J. K.1994. Modeling forest growth and yield application to mixed tropical forests CAB International Wallingford, UK. 312pp.
12. Woollons, R.C. and W.J. Hayward. 1985. Revision of growth and yield model for radiata pine in New Zealand. Forest Ecology and Management 11 : 191-202.
13. Xu, L. 1990. Growth and yield of Douglas-fir plantation in the central North Island of New Zealand. Ph.D. thesis. University of Canterbury. 246pp.