FORECASTING YIELD OF DOUGLAS FIR IN THE SOUTH ISLAND OF NEW ZEALAND

A THESIS

SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY IN FORESTRY

IN THE

UNIVERSITY OF CANTERBURY

by

Makitauwo Joseph Temu

University of Canterbury

1992

Abstract

Preparing forecasts of assorted yields for forest crops is fundamental to managing forests. Studies of growth and yield in the form of systems of prediction equations provide managers with information on instantaneous, periodic, and whole stand growth and yield information, which provide a first means of regulating yield from forests. The availability of appropriate compatible stem volume and taper equations provides further quantitative information about resources in terms of merchantable lengths, end diameters and volumes of particular sections of the stem of the tree necessary to determine the mix of products. This thesis describes in particular how a system for forecasting assorted yields of Douglas fir in the South Island of New Zealand was developed.

The system comprises two models, namely DfirTree and DfirStand.

> DfirTree is a compatible tree volume and taper prediction system, developed to cater for Douglas fir trees grown throughout the South Island, in Canterbury, Nelson and Southland. The volume - taper prediction system is based on the principle of splines (segmented polynomials) and provides two approaches with which to determine volumes of any part of the stem: (i) volume based and (ii) taper based.

> DfirStand is a simultaneous growth and yield model for simulating growth of Douglas fir in all four regions of the South Island, namely Canterbury, Nelson, Southland and Westland. DfirStand is developed through a *state - space approach*, the variables used to describe the state of the system at any time being mean top height, net stand basal area/ha, stocking, thinning history and *local environment*.

Both components of the overall yield prediction system show how regional attributes can be aggregated and modelled in a more realistic manner through the use of *dummy variables* to explain *locality adaptation(s)* where applicable. Rather than having a proliferation of models or an unwieldy quantity of adjustment factors, this system envisages a return to the traditional general volume, taper and yield prediction systems that can be developed with modern technology, ones which utilize the power of user - friendly computer hardware and software to provide the requisite sensitivity for forecasting assorted yields.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Douglas fir (Pseudotsuga menziesii [*Mirbel, Franco*]) ranks second to radiata pine (Pinus radiata [*D. Don*]) in New Zealand as a plantation tree species from which several domestically consumed wood products are currently derived. There is also a considerable potential to secure a position in the external market with Douglas fir manufactured products. The species was introduced in New Zealand as a trial plantation crop in 1897 (Kirkland, 1969), but its widespread establishment did not begin until the 1920's. The current total establishment is 63 130 hectares, with 28 784 hectares in the South Island and 34 346 hectares in the North Island (NEFD, 1989).

Douglas fir has good timber qualities, which are at least as good as, or even superior to those of radiata pine (James and Bunn, 1978). This is especially the case for engineering purposes (Hellawell, 1978). Nevertheless, the importance of radiata pine as a timber crop in New Zealand outweighs that of Douglas fir for two major reasons:

(1) radiata pine has a shorter technical rotation, about 30 years, while Douglas fir has financial rotations of between 50 and 80 years;

(2) Douglas fir plantations were heavily infected with Phaeocryptopus gaeumannii in the 1960's, a needle fungal parasite, which led to a significant decline in growth of the species (Hood and Kershaw, 1973, 1975; James and Bunn, 1978; Beekhuis, 1978).

However, it remains an important commercial timber species in New Zealand.

There are several volume and taper tables for the species currently in use in New Zealand. The volume tables are T15 covering the whole of New Zealand, which is now superceded by T136, T120 for Ashley forest and T228 for Longwood forest. The corresponding compatible taper tables are F136 for all of New Zealand and F228 for Longwood forest. All these equations are maintained by the New Zealand Ministry of Forestry. The existing computerized growth and yield models in use for the species are DFCNIGM (Liu Xiu, 1990) and SIDFIR (Law, 1990), which cater for the Central North Island and the South Island respectively.

There has been a tendency in many countries (including New Zealand) of prescribing growth and yield models even for very small populations (Whyte et al., 1992). This approach to modelling is frequently not justified. This study aims at reversing this trend by looking at the potential for creating general models that allow for regional or sub - population (local) adaptations. The study embraces the traditional yield model approach by prescribing such models to as large populations as possible. The reasons behind such a seemingly backward move are:

(1) there are all sorts of dangers through growth and yield model proliferation;

(2) users are misapplying the existing ones or at least demanding unrealistic sensitivity;

(3) the ability to incorporate regional and local adaptations provides a means of forecasting production yield with increased sensitivity.

1.1.1 Objectives of the Study

This study examines the extent to which a precise general growth model can be formulated for Douglas fir grown in the South Island of New Zealand, that caters for regional and local adaptations. The detailed objectives of this study, therefore, are to:

(1) study the different patterns of growth of Douglas fir in the South Island of New Zealand and where necessary to stratify the crop into different growth classes for modelling them individually;

(2) develop a whole stand growth and yield model with respect to objective (1) above that caters for variable thinning regimes;

(3) develop associated stem volume and compatible taper equations, volume - based and taper - based taper equations for Douglas fir in the South Island to provide the means for catering for log assortments.

1.1.2 Scope of the Study

The population for the whole stand growth and yield modelling conducted here refers to Douglas fir grown throughout the South Island of New Zealand. Sectional measurements for developing the new stem volume and taper equations come from 25 forests in Canterbury, Nelson and Southland (see appendix A.1), but no data were available for Westland. Permanent sample plot data (PSP) available for this study come from 21 forests in the four regions of the South Island, namely Canterbury, Nelson, Southland and Westland (see appendix A.2). The data set is such that the various models can be safely applied only between ages 5 and 78 years, and only for the South Island of New Zealand.

1.2 Notation

Throughout this thesis standard IUFRO notation is adopted. Unless otherwise stated, the following symbols and definitions apply.

α_i, β_i, γ_i: regression coefficients;

AL: altitude above mean sea level (m);

d': top diameter inside bark (cm) at height h' (m) from ground level;

d: diameter at breast height outside bark (cm);

SEE: standard error of parameter estimate;

ESS: Error sum of squares (residual sum of squares);

f: form factor;

G_i: net basal area per hectare at crop age T_i (m^2/ha);

h: total tree height from the ground level (m);

h': height above ground somewhere along its length in m ;

$h_{(100,i)}$: mean top height at crop age T_i (m);

K: a constant to convert d^2 in cm^2 to basal area in m^2;

ℓ: distance in m from the top of the tree to top diameter d' (cm) or $h - h'$;

MSE: mean squared error;

N: total number of observations in a population or stocking/ha depending on context;

N_i: number of stems per hectare (stocking) at crop age T_i;

NID(0,σ^2): normally and independently distributed with mean 0 and constant variance σ^2;

RND: random normal deviate;

RMS: root mean square;

S: site index - mean top height at age 40 years;

T_i: age of crop in years for period i;

T_t: age of thinning in years;

v: total stem volume of a tree (m^3);

v_m: merchantable volume (m^3) of a tree stem to specified height;

V_i: stand volume per hectare at crop age T_i (m^3/ha);

z: relative distance from the top of the tree, determined by the following relationships;

$$z = \frac{h - h'}{h} = \frac{\ell}{h}$$

σ^2: variance of a population;

s^2: variance of a sample;

Chapter 2

Review of Literature

2.1 Modelling in general

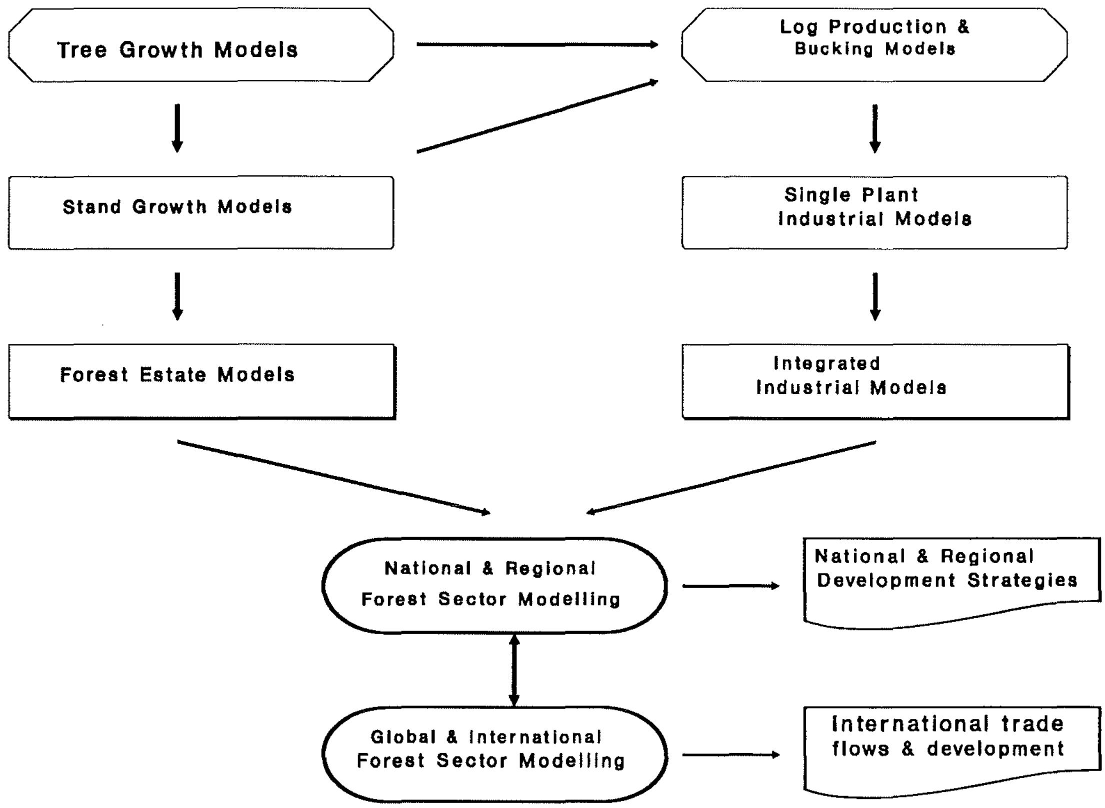
A model may be defined as a mathematical or physical system obeying specified conditions, the behaviour of which is used to understand a physical, biological or social system and to which it is analogous in some way (Ralston and Meek, 1976). The main purpose of having models is usually to aid in planning and decision making. The term *planning model* refers to any decision - making aid, ranging in complexity from the toss of a coin to sophisticated computer based models (Johnson, 1989).

Planning in forestry has evolved with the changing needs, priorities and objectives of both public and private forest owners. For example, the classical European concept of yield regulation evolved from the optimal rotation model (Faustmann, 1849), and arose out of fears of inadequate future supplies of wood. Today, the issues extend beyond just timber famines and are equally concerned with other benefits that forests can confer. Biomass production is a key issue in all of these.

At present the most effective research and analysis technique to represent a system, concept or operation is perhaps through use of a logical mathematical model. A mathematical model can be regarded as one or more mathematically expressed relationships among variables which may logically be expected to be obeyed. Generally computers are now used to analyze and study models because of large computational needs. Many basic phenomena are associated with certain mathematical planning models, examples in forestry being optimal allocation, replacement of equipment, routing of forest roads, queueing of logging trucks on a weigh-bridge, sequencing of logging trucks at some landing, predicting occurrence of forest fires, multiple objective decision analysis and many others. The use of computer technology has made it possible in forestry to create planning models and successfully apply them for harvest scheduling, fire fighting, tree breeding, log bucking and resource allocation problems, to mention just a few examples.
Forest planning models can be ranked according to the hierarchical order of Figure 2.1: tree growth models, stand growth models, forest estate models, national and regional forest models, global and international forest sector models (Johnson, 1989).

Forest sector planning models are those models developed specifically to incorporate the multitude of variables involved in examining an entire forest sector, and to indicate strategic alternatives for that sector. These are classified by their scope (national and regional, global and international) as suggested in Figure 2.1, as well as by the methodology they employ to generate alternative strategies (i.e. dynamic simulation, mathematical programming and econometric spatial equilibrium models).

Figure 2.1: The Hierarchy of Forest Planning Models [Johnson, 1989]



This thesis focusses specifically on computer based models that belong to tree and stand growth model categories, addressing the linkages with other kinds of models (in Figure 2.1) wherever applicable.

The following sub - sections describe each of these briefly, while sections 2.2 to 2.4 discuss tree volume, stem taper and stand growth and yield, the aspects most central to the research reported here.

2.1.1 Tree and Stand Growth models

One of the earliest forest planning needs, as discussed in the introduction to this chapter, was for the regulation of forest yields. Predictions of future growth of tree crops are essential for analyzing and interpreting forest supply capabilities and their sustainable supply capacities. Models for this purpose can be grouped into those dealing with natural forests and those dealing with plantations (Clutter et al., 1983), both of which groups are discussed in detail in section 2.5.

2.1.2 Forest Estate Models

Forest estate models are a modern form of forest working plans, taking advantage of the capacities of computers (Allison, 1987). Estate models address production of whole forests, while growth and yield models are restricted to modelling tree, or at most, stand growth. A forest, however, consists of many different stands, each belonging to its characteristic crop type (similar accessibility, silviculture, ownership, growth model representation, location, *et cetera*). Extending the optimal management strategies for single stands derived from a stand level growth model to an entire forest is therefore rarely optimal for the forest as a whole. The need to plan for aggregates of non - uniform forest stands gave rise to the concept of forest estate modelling. Such models incorporate inputs from relevant growth models in some form, as well as recognizing other distinguishing stand features.

Examples of forest estate models developed in New Zealand are IFS (Interactive Forest Simulator, Garcia, 1981), FOLPI (Forestry Oriented Linear Programming Interpreter, Garcia, 1984a), RMS - 2020 (Resource Maturity Simulator, Allison, 1987), and REGRAM

(McGuigan, 1992). RMS - 2020 and IFS are both simulation models, which allow the evaluation of different management alternatives for a given forest. In both, the forest is described in terms of crop types, while the state of the forest in a given period is defined by the area and potential yield in each age class within each crop type. For each period of simulation, areas may be harvested, planted or otherwise treated, using inputs supplied from growth and yield models. The user specifies the management actions to be followed each period and can produce a number of reports at the end of the simulation describing the results of those actions. While simulation models such as these are capable of finding acceptable management options, there is no guarantee that better alternatives for increasing the utility of the forest to its owner do not exist.

Linear programming has also been successfully employed to determine forest level management strategies through an optimization process, which maximizes utility for the forest owner under a range of constrains. Two basic formulations of this problem have been recognized (Johnson and Scheurman, 1977), depending on whether the identity of initial harvesting units is preserved throughout the planning horizon (Model I), or new harvesting units are created from the area regenerated in each period (Model II). The disadvantage of Model II is the loss of identity of initial harvesting units, rendering less accuracy than can be obtained from Model I, although its formulation significantly reduces the number of decision variables. Both formulations are subject to inadequacy in modelling the range of detailed silviculture options and management regimes that characterize modern plantation forestry.

In New Zealand, the major forest level optimization model is FOLPI. FOLPI and IFS are fully compatible, and same input data files can be used for FOLPI and IFS to generate identical reports. FOLPI, IFS and REGRAM use other than Model I or MODEL II formulations.

2.1.3 Log Production and Bucking Models

Forecasting growth and yield by itself does not provide information about the likely availability of the assortment of products, that is needed to plan their industrial use. The output of growth models can be used as inputs to models that predict the stand outturn, providing long term predictions of the quality, and volume by log sizes. PROD (Goulding and Shirley, 1979) is one such model. PROD predicts a (diameter at breast height) dbh distribution and stock table volumes of a mix of log products given the site index, age, top height, basal area/ha and stocking for each stand according to a simple set of user defined cutting patterns. The output, estimates of volume, may subsequently form part of the input to forest estate modelling indicated in previous section. PROD is an integral part of the silvicultural stand model SILMOD (Whiteside and Sutton, 1983), now replaced by STANDMOD (Whiteside et al., 1987), used widely in New Zealand for evaluating different silvicultural regimes for radiata pine. LOGRAD (see Whiteside et al., 1987) is a further refinement developed by the Conversion Planning Team at FRI - Rotorua, which transforms the log sizes from PROD into predictions of log grade outturn based on expected defects, taper, and sweep.

Optimal bucking models have also been developed explicitly for optimizing the cross - cutting or bucking decision process modelled in PROD. MARVL (A Method for Assessing Recoverable Volumes by Log Types) by Deadman and Goulding (1979) is one such model. MARVL is mainly used for analyzing pre - harvest inventory information. AVIS

(Assessment of Value by Individual Stems) by Threadgill and Twaddle (1985) is an audit/training system developed for comparing actual with optimal cutting patterns and determining the resulting loss in value for individual stems. Eng et al., (1986) used a combination of linear and dynamic programming to extend bucking strategies that were optimal for a year's harvest throughout the entire forest of Caribbean pine in Fiji so as to meet market demands.

2.1.4 Single Plant Industrial Models

Numerous single plant models involve simulations of the production process in question. Models used in New Zealand include GEMS (A General Energy and Material Simulator for Pulp and Paper Industry) by Edwards et al. (1987), PLYMILL (A Pulp Wood Mill Simulation) by Ward (1987), and a suite of simulation programs incorporating a linear program for saw pattern selection by van Wyk and Eng (1987).

2.1.5 Integrated Industrial Models

The vertical integration of several production facilities in a multi - product forest industry complex creates many economies of scale and eliminates several transport and processing costs. Such integration, however, requires enhanced planning, with a company having many possible sources of raw material, for example. The primary class of integrated industrial models in use today are log allocation models. The technique used in determining optimal management strategies in this class of models is mainly linear programming, sometimes in combination with dynamic programming. The use of log allocation in

New Zealand was first reported by van Wyk (1983), who developed the programme ROBUST, (a single period process selection model) LP model at FRI. Developments made by McGuigan (1984) produced a model LOGRAM essentially a modification of ROBUST capable of allowing for multiple manufacturing sites and forest allocations. It has remained a single period process selection model, however, until the development of REGRAM (McGuigan, 1992). Further research into development of optimal log bucking and allocation that includes multiple period consideration is currently under investigation (Ogweno, 1992).

2.1.6 National and Regional Forest Models

In their basic form these models are simply aggregations of the integrated industrial models within a nation or large region. TAMM (Timber Assessment Market Model) by Adams and Hynes (1980) is the best known model in this category: it is used by US Forest Service to assist in the analysis of long term trends in resource use and status.

2.1.7 Global and International Models

These models focus on the prediction of long term development of production, consumption and trade in forest products at international and global level. One such model is GTM (Global Trade Model) by Dykstra and Kallio (1986). GTM is a spatial market equilibrium economic model employing non-linear programming. GTM was developed by the forestry section of IIASA (International Institute of Applied Systems Analysis). GTM is an economic multi-period, multi-region model that considers each forest industry separately. But, due to its broad global emphasis, GTM does not cater for alternative forest management options. It incorporates four classes of growing stock, namely large and small, coniferous and non-coniferous, which are upgraded by simple growth functions between periods.

Nevertheless, all these models in 2.1.1 to 2.1.7 depend, as shown in their hierarchical arrangement in Figure 2.1, on having reliable tree volume, stem taper and stand growth models, which provide the focus for the research reported here in this thesis.

2.2 Tree Volume Equations

2.2.1 Past Work in Tree Volume Modelling

Models to predict whole stem and merchantable volume up to (i) a given height or (ii) diameter limit or (iii) diameter to a given height, are required for forest related raw material analysis and are vital components of growth and yield simulators. Compatible stem volume and taper equations form a useful class of models to predict these quantities (Demaerschalk, 1972; Goulding and Murray, 1976; Van Deusen et al., 1982; Byrne and Reed, 1986; McClure and Czaplewski 1986). Tree volume equations can be classified as follows.

2.2.1.1 One Dimensional Tree Volume Equations

These utilize only diameter at breast height to predict total stem volume. Local volume tables and tariff tables fall into this category, the term local being used because such tables are very restricted in their applicability because of height and stem taper interactions which are embedded in the table. Tariff tables provide sets of local volume tables for particular stands from a large data set, (Jolly, 1951; Hummel, 1955; Hummel et al., 1962; Turnbull and Hoyer, 1965); Gray, 1966; Carron, 1971. They have been used in Central Europe in the interests of consistency of estimation and were calibrated regularly against actual outturn. They were later used in Australia in the early 1950's, and subsequently in North America. They are well suited to growth modelling that incorporates a stand table projection capability.

2.2.1.2 Two Dimensional Tree Volume Equations

These utilize diameter at breast height and total or merchantable height in predicting total stem or merchantable volume. Examples falling into this category are the combined variable tree volume equation (Spurr, 1952), and various other formulations by Schumacher and Hall (1933), Gevorkiantz and Olsen (1955), Bennett et al. (1959), Romancier (1961), Honer (1965), Newham (1967), Brackett (1973), *et cetera.* Provided that there is reliable characterization of the average height of each dbhob class, these too are amenable to the stand table projection approach.

2.2.1.3 Multi - Dimensional Tree Volume Equations

Multi - dimensional tree volume equations are an extension of two dimensional ones: in addition to diameter at breast height and total or merchantable height, they use form and sometimes site index as independent variables (Bruce, 1926; Mesavage and Girard, 1946; Anon, 1948). Volume equations that utilize form as one of the independent variables are termed form - class volume tables, but they have become obsolete when it was realized that establishing diameter form - class relationships was rather elusive and that the variation of form between species and sites was considerable (Honer,1965). Multi - dimensional tree volume equations are not amenable to stand table projection unless their dimensions other than dbh and height are averaged, which then largely negates their adoption.

Perhaps the most important tree volume equation that appears in the literature is the combined variable equation (Spurr, 1952; Burkhart, 1977),

$$v = \beta_1 + \beta_2 d^2 h \tag{2.1}$$

where v is the tree volume either inside or outside bark, d is diameter at breast height (usually outside bark) and h is tree height. Comprehensive comparisons of volume equations have been made by Spurr (1952), Golding and Hall (1961) and Burkhart (1977). They all concluded that the combined variable equation was the best for their data and could not be improved by the addition of another variable. Other researchers have, however, found alternative tree volume equations which are superior for a particular species and locality (*e.g.* Newham, 1967; Brackett, 1973; Candy, 1989). Equation forms that can serve as mathematical models for construction of volume tables or as the basis to develop other models have been discussed in Husch et al. (1972), Loetsch and Haller (1973), Clutter et al. (1983), *et cetera.*

Until the early 1960's or so, volume equations were constructed independently of taper equations, and so it was not possible to have a unified system of predicting both total tree and merchantable volume. Until quite recently, the prediction of merchantable volumes to varying merchantability limits was usually accomplished by fitting a separate regression equation for each merchantability limit involved. Thus, for a single tree population, different formulae would be involved for predicting merchantable volumes to 10, and 15 cm top diameters inside bark. These have now been replaced by volume ratio equations which utilize the merchantability limit as an independent variable (Burkhart, 1977; Matney and Sullivan, 1982a; and many others).

2.2.2 Principles Underlying Construction of Tree Volume equations

Volume tables were first constructed mostly by graphical methods or, at best, with the harmonized curve method (e.g Chapman and Meyer, 1949). Alignment charts were later used as a means of graphically portraying and solving an equation or formula. In addition they were also used in constructing numerical volume tables (Bruce and Schumacher, 1950; Husch, 1963).

The contemporary principles underlying the construction of volume tables have utilized the same tree variables since the early years of the 19th century, namely dbh, total or merchantable height and tree form (Husch et al., 1972). Sometimes additional factors such as mean annual diameter increment d/T , (Candy, 1989), and site index have been included. The use of graphical methods and alignments charts have now been replaced by regression equations, now made easier with computer technology. The elegance and objectivity of modern statistical computations for fitting equations give them a large advantage over the other two methods, although many times the results are not necessarily significantly better (Cunia, 1964).

The development of science and statistics and the advent of electronic computers have now made such graphical applications obsolete, however, and nowadays regressions are fitted to sample tree data to produce formulae that explicitly define the relationship between the predicted tree volume and predictor variables used. Solutions of such formulae could still, of course, be presented in tabular or graphical forms.

2.2.3 Problems Associated With Modelling Tree Volume Equations

Even with the use of modern computers and regression analysis, construction of tree volume equations needs to be done with due consideration of the following major problems which prevail:

(i) normality of distribution of tree volume (Meyer, 1953),

(ii) homogeneity of variance of tree volume (Cunia, 1964),

(iii) tree sampling process (Cunia, 1964),

(iv) selection of variables and number to be used (Spurr, 1952).

These factors are discussed more fully in the subsections below.

2.2.3.1 Normality of Distribution of Tree Volume

Tree volume for a given dbh is not normally distributed, but is highly skewed. The dbh distribution of a population or sample of trees could theoretically be normal, but tree volume for a given dbh class will vary because of the differences in height. Consequently, this affects the variance of tree volume for a given dbh. In fitting regressions equations it is common to estimate the parameters for the overall data and not by dbh or height classes. While this has no great effect on estimation of regression coefficients, it will affect the standard error of the equation and hence the probability level of the significance tests and confidence limits (Meyer, 1953; Cunia, 1964). This problem is usually avoided by effectively screening the data, according to d^2h classes: a tree volume that deviates much from the rest constitutes an outlier, which needs to be carefully scrutinized and may have to be removed from the data, as discussed in section 3.3.3.

2.2.3.2 Homogeneity of Variance

Variance of tree volume is usually a function of the quantity d^2h. Deviations from the true "regression function", induced through the volume of large trees in the sample having disproportionate effects on the estimation of the least - squares coefficients, need to be compensated for. One of the earliest methods used to correct for this lack of homogeneity was the use of logarithmic volume equations (Bruce and Schumacher, 1950; Spurr, 1952; Meyer, 1953). An important drawback of this approach is that, by taking logarithms, the estimation about the arithmetic mean is automatically replaced by the geometric mean. Because the first is always larger than the second, the estimated coefficients are biased, and so this approach is not to be recommended. An unbiased and better way of correcting for non - homogeneity of variance is to estimate the regression coefficients by use of *weighted least - squares*. Schumacher and Chapman (1954) were the first foresters known to use this regression technique, (others who followed later include Gedney and Johnson, 1959; Buckman, 1961; Furnival, 1961, Cunia, 1962). One disadvantage of the method of weighted least - squares regression is that there is no formal general treatment on what are the best sets of weight to be used in any one particular case. Numerous researchers have discussed guidelines on appropriate weights (Cunia, 1962, 1964; Moser and Beers 1969; Draper and Smith, 1981; Clutter et al., 1983). Generalisations are probably illusory. The use of weighted least - squares regression does not solve all problems of valid and efficient calculations of tree volume equations. It too has problems associated with it. Thus, the weight finally chosen is a subjective decision, while the weight itself is simply an approximation.

2.2.3.3 Tree Sampling Process

An efficient tree volume equation should provide unbiased estimates for each diameter - height class. It is important that the sample data consist of adequate replication of all possible diameter - height classes throughout the range in that locality. Some authors like Gordon (1985) suggest that the selection of sample trees in each diameter - height class should be proportional to the frequency of occurrence of that class, the aim should be rather to have a reliable average volume for each diameter - height class that could be represented. Usually this means an approximately equal number of trees should be selected from each such class.

2.2.3.4 Selection of Variables and Their Combinations

Selection of variables to enter the equation and how they may be combined in a tree volume function to accurately predict volume have provided a major problem for researchers for a long time (Spurr, 1952). It is not always a straight - forward analysis to derive appropriate direct relationships between tree volume and predictor variables such as d and h. Empirical evidence has shown that frequently these variables have to be combined in a variety of different ways to provide reliable equations in different circumstances. There is no standard procedure: one has simply to examine all the likely combinations of variables and their interactions. Today, statistical knowledge, regression techniques and computer technology have allowed researchers to alleviate the computational tedium that this computational process involves. Stepwise regression, for instance, can be used to select variables or combinations of variables that are to enter an equation, but how the variables are to interact with each other remains a skill of the modeller. In this study, for example, locality effect was implicitly defined in the tree volume equation through the use of dummy and other predicting variables while an all - paths sequential routine for including variables was adopted.

2.3 Taper Equations

2.3.1 Past Work in Taper Modelling

This sub - section reviews the general structure of taper equations and also classifies the most important taper equations that have appeared in the literature. Taper equations express the expected stem diameter, either inside or outside bark, as a function of height above ground level, total tree height, diameter at breast height and sometimes total tree volume (Demaerschalk, 1972). Taper equations reported in the literature can be divided into three major groups:

(i) single equations that describe taper

(ii) segmented taper equations

(iii) variable exponent taper equations

These are discussed serially in more detail below.

2.3.1.1 Single Equations

The most common approach used to describe diameter changes from ground to top involves a single function of many forms (Hojer, 1903; Behre, 1923; 1927, 1935; Matte, 1949; Osumi, 1959; Kozak and Smith, 1966; Bruce et al., 1968; Kozak et al., 1969; Demaerschalk, 1971, 1972, 1973b; Bennett and Swindel, 1972; Munro and Demaerschalk, 1974; Clutter, 1980; Goulding and Murray, 1976; Gordon, 1983; Amidon, 1984; Bigin, 1984; Newberry and Burk, 1985; and others). The major weakness of all these models is the significant bias in estimating diameters close to ground as well as at some other parts of the tree. A trade - off between accuracy and precision has to be made for any one equation. The advantages of this approach are that they are easy to fit and usually easy to integrate for calculation of merchantable volume.

2.3.1.2 Segmented Taper Equations

In this approach more than one curve is used to represent all the various parts of the stem, and neighbouring ones are joined in such a way that their first derivatives are equal at the point of intersection (Heijbel, 1928; Ormerod, 1973; Max and Burkhart, 1976; Demaerschalk and Kozak, 1977; Cao et al., 1980; Martin, 1981; Byrne and Reed, 1986; McClure and Czaplewski, 1986; Candy, 1989; Whyte et al., 1992). Its advantage is that the diameters are predicted with less bias at most parts of the stem than by single functions. The disadvantages are that the parameter estimates are usually very difficult to derive and the formulae for calculating volume and merchantable height are cumbersome, sometimes non - existent.

2.3.1.3 Variable Exponent Taper Equations

This approach uses a single continuous function as described in (i) above, but with a changing exponent from the ground to compensate for the changing form of the stem (Hayward, 1987; Kozak, 1988). This kind of equations remain the least popular in forestry literature because of their complexity. Their level of precision is comparable to the segmented equation approach.

2.3.2 Integrating Volume and Taper Estimating Systems

Early taper equations developed before Demaerschalk (1971), had one common attribute: they were developed independently of the corresponding tree volume function. A theory was later developed early in the 1970's which greatly improved the understanding of relationships that exists between tree volume and taper functions (Demaerschalk, 1971, 1972, 1973b; Munro and Demaerschalk, 1974). In this theory, when taper and volume equations are treated as one and the same, they are deemed to be *compatible* (Demaerschalk, 1971, 1972). The accepted definition of *compatible* equations is : those taper functions which, when integrated over total tree height, give the same total volume as that given by a volume equation. There are two basically different techniques which can be used to obtain compatible systems of taper and volume: one involves fitting a taper equation on taper data and deriving from it a volume prediction system through integration; the other is more or less the opposite, in that a volume equation is fitted to the volume data from which a compatible taper equation is derived. The first is a taper - based system and the second is volume - based system (Demaerschalk, 1973a). The theory further proposed that a compatible taper equation should be a polynomial of general form shown in equation 2.2. (Munro and Demaerschalk, 1974).

$$d'^2 = \frac{(p+1)}{K}\frac{v}{h}z^p \tag{2.2}$$

where;

d' =top diameter to be predicted;

v =total tree volume;

z =relative length from the top of the tree to top diameter d;

$$z = \frac{\ell}{h} = \frac{h-h'}{h}$$

K =constant to convert square centimetres to square metres;

p =free parameter.

In this formulation, the polynomial is limited to the order p , and in most cases was not sufficient to describe the stem taper without bias. It was noted by Goulding and Murray (1976), that the polynomials of this kind described in equation 2.2 were not flexible enough to account for the butt swell, which is quite noticeable in large older trees. Goulding and Murray (1976) extended the theory of Munro and Demaerschalk (1974), therefore, to allow the polynomial taper function to acquire variable orders as shown in equation 2.3.

$$d'^2 = \frac{v}{kh}f(z) \tag{2.3}$$

where $f(z)$ is a polynomial in z . The coefficients of the polynomial are algebraically restricted so that consistent volume estimates are derived, just as if a volume equation was being employed. Such polynomials are usually fitted by conditioned linear or non linear least - squares regression. The methods used and restrictions on the parameters are discussed in chapter 4 of this thesis.

2.3.3 Summary

Tree volume equations are used in forest inventory to calculate total volume of the tree inside bark, summation of which adds up to total tree volume per hectare. Merchantable volume/ha to required utilization standards can be obtained by incorporating a compatible taper equation (Clutter et al., 1983). Taper equations are useful adjuncts to inventory because they provide (i) predictions of inside bark diameters at any point on the stem; (ii) estimates of total stem volume; (iii) estimates of merchantable volume and merchantable height to any top diameter and from any stump height; and (iv) estimates of individual log volumes. Tree volume and taper equations are an integral part of any detailed yield forecasting undertaking.

Because of the properties mentioned above, taper functions are essential components of forest harvesting and bucking models. For example,

(i) they provide estimates of the mix of products, such as saw logs and pulp logs, as described in Eng et al. (1986), without the need to collect additional data, even when utilization standards change;

(ii) they are used in many optimal bucking models such as in Pnevmaticos and Mann (1972); (iii) they are an integral part of many inventory models, such as MARVL (Deadman and Goulding, 1979), and planning models such as PROD (Goulding and Shirley, 1979);

(iv) taper functions can also be used in analytical estimates of tree weight and biomass (Clutter et al., 1983) and for determining the centre of gravity and mass of inertia of trees, an important aspect of full tree harvesting systems (Fridley and Tufts, 1989).

2.4 Forest Growth and Yield Modelling

Modelling the growth and yield of forests over time is a means of adequately characterizing or describing some or all of the many processes that make up the system collectively known as a forest. Avery and Burkhart (1983) described it concisely as a means of forecasting stand dynamics. The purpose of growth and yield models, despite their complexity of structure, can be explained simply as;

> given a set of stand or tree conditions, such as basal area and stems per hectare, which refer to one point in time (T_1) and to certain locality, by how much will these have changed at a future time (T_2) given specified stand or tree treatments (thinning, fertilization, *et cetera*).

Such a quantitative capability has been the forest manager's crucial need so that sound decisions can be taken on how wood supply from particular forests should be managed.

2.4.1 Need for Mathematical Models in Growth and Yield studies

Fitting mathematical models to forest growth and yield data is a widely accepted method of summarizing resource production information about individual trees, stands and forests. A mathematical growth model is a mathematical function, or system of functions, used to relate actual growth rates to measured tree, stand, and site variables. The advantages of fitting mathematical models to growth are set out below.

(i) An appropriate curve (equation) may conveniently summarize the information provided by the observations of a given data set. Thus, a large number of observations, collected over time on individual trees, stands and forests can be represented by only a few parameters.

(ii) Comparisons performed between growth data sets or at various ages of a single data set is more efficient if performed on the summary of parameter estimates of the fitted models (Hoel, 1964). This is particularly efficient when the growth functions can be integrated with respect to time as set out in Clutter (1963).

(iii) Problems of missing or unequally spaced data are alleviated by statistical readjustment or interpolation for missing observations.

(iv) Growth velocities and accelerations are easily estimated from fitted models by differentiation (Clutter, 1963; Berkey, 1982a).

2.4.2 Past Work in Forest Growth and Yield Modelling

The following sub - sections review the major concepts and historical development of growth and yield modelling from the early 19th century until today.

2.4.2.1 Normal Yield Tables — Use of Graphs and Tables

Forest growth and yield models formed the earliest forest planning models. They were derived empirically and represented in the form of graphs and tables. German foresters established the 'normal yield table' approach, as explained in Bickford et al. (1957); Fries (1967); Clutter et al. (1983); Avery and Burkhart (1983); to predict yield of even aged stands. The methods then adopted had the following limitations as pointed out by Ware et al. (1988).

(i) The modelling assumed fully stocked stands, and thus the variables used were constrained and not allowed to enter the model independently.

(ii) The graphical procedures required adjustments to be made when the tables were used for non - normal stands, which often produced major errors because of the inadequacy of the assumptions in the adjustment process.

(iii) A strong argument still persists today over the applicability of the concept of normality: some forest researchers have argued that this ideal condition is subjective and does not represent a rational management goal, because a non - fully stocked stand with proper silvicultural treatment could produce higher returns than a fully - stocked untended stand (Curtis, 1972). This substantiates point number (i) above that maintaining one variable (stocking) at its maximum will not always give the maximum yield and may well not be a relevant management objective anyway.

2.4.2.2 Variable Density Yield Tables

MacKinney et al., (1937) were among the first to propose a variable density growth and yield model for non - normal loblolly pine stands grown in North and South Carolina, USA. More important is that the modelling was performed objectively through the use of a statistical approach. One drawback of the early statistical analyses of growth was that they were performed using polynomial models (for example, MacKinney et al., 1937; Wishart, 1938), a consequence of polynomial growth curves being easy to fit and interpret

(Merrel, 1931). However, they fall short of ideal objectives because more often than not, high order polynomials are needed to provide empirical fits to many natural phenomena and no proper biological meaning can be attached to the parameters associated with them.

Empirical variable density yield tables were developed in North America as a means of refining the definition of normal forests to cope with existing natural forests there. The concept of average rather than normal values applied (e.g. Schumacher, 1939; Bennett et al., 1959; Avery and Burkhart, 1983 *et cetera*). However, adjustments still had to be made when tables were applied to stands that were not at the average level.

The proposed yield model by MacKinney et al., (1937) was a logarithmic polynomial model of the form shown in equation 2.4,

$$\ln[\frac{(\alpha - Y)}{Y}] = \beta_1 T + \beta_2 T^2 + \beta_3 T^3 + \beta_4 (SDI) + \beta_5 S + \beta_6 C \tag{2.4}$$

Where α is the maximum theoretical yield (asymptote), Y is yield, T is stand age, S is site index, SDI is Reineke's (1933) stand density and C is composition index. Schumacher (1939) proposed a log - reciprocal of time yield equation, which was later used by MacKinney and Chaiken (1939) as a refinement of their previous equation, this equation was

$$\ln(Y) = \beta_0 + \frac{\beta_1}{T} + \beta_2 S + \beta_3 log(SDI) + \beta_4 C \tag{2.5}$$

In both equations, 2.4, and 2.5 the stand density is part of the equation and a variable that is allowed to vary freely. This was indeed a major development in forest growth and yield research, because, apart from the use of regression techniques, the equations have the following desirable properties: first, the dependent variable yield (Y) is predicted from a specific combination of independent variables over a wide range; secondly, the logarithm of the yield is proportional to the reciprocal of age; and thirdly, the functions exhibit asymptotic growth. All these properties are essential components of modern growth and yield models (Clutter et al., 1983). Many researchers have used similar methods of constructing growth and yield equations since that first one in 1939 (for example, Schumacher and Coile, 1960; Brender, 1960; Avery and Burkhart, 1983; Bailey and Ware, 1983; Murphy and Farrar, 1988). Subsequent research and technological development have made growth and yield models more sophisticated through the incorporation of often complex mathematical equations and their implementation on fast computers.

2.4.2.3 Compatible Growth and Yield Models

Buckman (1962) and Clutter (1963) laid the foundation for the need for compatibility between growth and yield. Both demonstrated that, when cumulative growth is plotted over time, the yield curve which results can be derived mathematically by integration of the growth function: that is, the first derivative of the yield function results in a growth function (Clutter, 1963; Clutter et al., 1983). Compatibility between growth and yield is a major premise for the building of modern growth and yield models, in that the total area under the growth curve must equate yield. Biological theory, supported by much empirical evidence indicates that yield curves have a sigmoidal shape, as demonstrated in the Schumacher yield equation (using IUFRO conventional notation as outlined in section 1.2).

$$G = e^{(\alpha+\frac{\beta}{T})} \tag{2.6}$$

The first derivative of this equation will provide a corresponding growth function.

$$\frac{dG}{dt} = -\frac{\beta}{T^2}e^{(\alpha+\frac{\beta}{T})} \tag{2.7}$$

2.4.2.4 Simultaneous Growth and Yield Model

Sullivan and Clutter (1972) further extended the concept of compatibility and developed a simultaneous growth and yield model, by simultaneously estimating yield and cumulative growth as a function of initial age. When future age (T_2) equals current age (T_1), the equation is reduced to the conventional yield model. Thus, it is simultaneously a yield equation for the current condition and a projection model for the future. They used Clutter's (1963) volume yield equation 2.8 and basal area yield equation 2.9. Using IUFRO conventional symbols, their equations can be represented as

$$\ln(V) = \alpha_0 + \alpha_1 S + \frac{\alpha_2}{T} + \alpha_3 \ln(G) \tag{2.8}$$

$$\ln(G) = \beta_0 + \beta_1 S + \frac{\beta_2}{T} \tag{2.9}$$

and thus, from 2.9

$$\beta_2 = T(\ln(G) - (\beta_0 + \beta_1 S)) \tag{2.10}$$

Differentiating 2.9 with respect to T, and substituting β_2 as in 2.10 gives 2.11

$$\frac{d\ln(G)}{dT} = -\frac{(\ln(G) - (\beta_0 + \beta_1 S))}{T} \tag{2.11}$$

The above equation is in differential form and can be integrated, rearranged and presented in a projection form shown in 2.12

$$\ln(G_2) = (\frac{T_1}{T_2})\ln(G_1) + (\beta_0 + \beta_1 S)(1 - (\frac{T_1}{T_2})) \tag{2.12}$$

This equation represents a sigmoid curve, has an inflection point and an upper asymptote. It is mathematically compatible, numerically consistent and path invariant (Clutter, 1963). Future volumes can be predicted by substituting equation 2.12 in equation 2.8 for G_2, that is

$$\ln(V_2) = \alpha_0 + \alpha_1 S + \frac{\alpha_2}{T_2} + \alpha_3(\frac{T_1}{T_2})\ln(G_1) + (\alpha_4 + \alpha_5 S)(1 - (\frac{T_1}{T_2})) \tag{2.13}$$

Where,

$$\alpha_4 = \alpha_3 \beta_0 \tag{2.14}$$

$$\alpha_5 = \alpha_3 \beta_1 \tag{2.15}$$

There are several other functional forms of growth models that can be formulated as compatible/simultaneous, and have been used successfully in yield studies: they include Chapman - Richards, Gompertz, Levakovich, Hossfeld, Weibull, monomolecular and others. There is no one functional form that is superior to another, but the way they behave when fitted depends on the nature of the data (Woollons et al., 1990).

Numerous researchers have then adopted the compatible/simultaneous growth and yield methodology with different functions (Clutter, 1963; Brender and Clutter, 1970; Pienaar and Turnbull, 1973; Pienaar, 1979; Farrar, 1979; Clutter and Jones, 1980; Bailey and Ware, 1983; Pienaar and Shiver, 1984; Pienaar et al., 1985; Murphy and Farrar, 1988 *et cetera*).

2.4.2.5 Computer Technology in Growth and Yield Modelling

Without the use of computers growth and yield prediction can be difficult, or even impossible sometimes to effect. Nevertheless, evaluation of the role of computers as a necessary tool for growth and yield prediction has seldom been comprehensively carried out, perhaps because it is assumed to be well known by researchers. The paragraphs below attempt to categorize the most common uses of computers as they pertain to growth and yield modelling.

(1) **Development of Models**

With a large data set and without computers, regression equations can be difficult to fit and interpret. Computers revolutionized growth and yield modelling and made possible the fitting of multiple linear and non linear regression equations easily. One of the disadvantages pointed out by MacKinney et al. (1937), was that the process involves a rather laborious procedure as well as a knowledge of correlation analysis. Today, special statistical packages (software) and high speed computers (hardware) have been developed to perform regression analyses (linear and non linear), which have made possible an array of modelling alternatives, some very complicated and others which were infeasible with the past technology: for example it is possible to use individual trees as a basic unit of growth and yield modelling and to adopt alternative modelling approaches.

Other uses include evaluation of different models, verification, calibration of models to suit local conditions, validation and forest growth monitoring.

(2) **Routine use of the Models**

Routine use of growth and yield models is generally effected through use of computers. Forest managers use growth and yield models interfaced with harvest models, bucking models, economic models, *et cetera*, to simulate forest productivity and economic outturn, and so that sensitivity analysis of a set of management alternatives can be conducted reliably. This vital capability cannot easily be achieved without the use of computers.

2.4.3 Site Index — Overview and Historical Perspective

This section explores some of the major concepts of site index. Problems of applying site index as a measure of site quality are also highlighted.

Site quality can be evaluated indirectly or directly, as explained in the following paragraphs.

2.4.3.1 Indirect Estimation of Site Quality

Evaluation of site quality can be done indirectly through use of vegetative site types (Zon, 1913; Cajander, 1926; Ure, 1950; Hodkings, 1961; Daubenmire, 1961; Daubenmire and Daubenmire, 1968). The application of this approach has the following disadvantages.

(i) The deep soil horizons, for example, may have little influence on the under - storey vegetation but still much influence on the tree growing medium. Under - storey is also influenced by such factors as wildlife, fire and site preparation.

(ii) The evaluation cannot be quantified without reference to another species growing in the area, and hence it is difficult to quantify.

Inter - species relationships have also been used in evaluation of site index, For example Coile (1948) used this method to calculate site index of loblolly pine and short leaf pine. Olson and Della - Bianca (1959) used such an approach for mixed stands.
Another approach is the use of topographic features (such as elevation), and soil characteristics (physical and nutrient properties). Theoretically tree growth is controlled by

environmental factors such as soil nutrients, soil moisture, aspect, elevation and temperature. Data pertaining to these values can be identified and regression or correlation analysis conducted with respect to tree growth (Coile, 1952; Myers and Van Deusen, 1960; McGee, 1961; Lewis and Harding, 1963; Carmean, 1970; Steinbrenner, 1975; Alban, 1976; Clutter et al., 1983). The application of this method, although it more often provides a good inference is limited because it involves massive data collection, is laborious, very costly and sometimes impossible.

2.4.3.2 Direct Evaluation of Site Quality

The direct estimation of site quality falls under two major categories, namely (1) historical yield records, and (2) site Index. Historical yield records can be used as a method for directly evaluating site index in terms of a measure of production in physical quantity, like volume/ha (Bates, 1918). The disadvantage of this method is that physical quantities such as stand volume are influenced by other factors like rainfall, rotation, insects, disease, genetics, management and the fact that most forests lack such historical data. For stands in which factors that influence volume production can be strictly controlled, stand volume is the best indicator of quality (Lewis et al., 1976; Clutter et al., 1983). This method is not very practical, however, because the cost of controlling those factors would be enormously high.

Site index (the mean top height at an index age) is by far the most common measure of productivity (Spurr and Barnes, 1980). Site index is popular because it is relatively easy to measure and dominant height is fairly independent of stand density, except at extremes such as thinning from above (Spurr, 1952). Also there is a strong historical precedent for its use, in that volume production potential is positively related to height growth (Roth, 1916). Thus mean top height could be a good indicator of site quality. The approach involves fitting a family of curves of mean top height development at a specified age (base age). (See examples: Zahner, 1962; Coile and Schumacher, 1964; King, 1966; Brickell, 1968; Lundgren and Dolid, 1970; Beck, 1971; Carmean, 1972; Graney and Burkhart, 1973; Bailey and Clutter, 1974; Trousdell et al., 1974; Burkhart and Tennent, 1977; Newberry and Pienaar, 1978; Clutter and Jones, 1980; Boardes et al., 1984; Harrison et al., 1986; Bailey et al., 1989).

Site index as a measure of productivity has been subject to numerous problems (Monserud, 1984a, 1985a, 1987; Wykoff and Monserud, 1987), especially in irregular stands and with mixed species composition or uneven distribution of ages. Although careful site/species tree selection can overcome some of these problems (Monserud, 1984b, 1985b), the solution to others has remained a mystery. A direct and similar example is provided in this study, an examination of site productivity showed that elevation and locality were jointly the best predictors of the productivity of a site, rather than site index. Other researchers too have been able to predict site productivity by using equations that do not include site index (for example: Stage, 1973; Wykoff et al., 1982).

Data for estimation of site index estimation can be derived in two ways,

(i) by analysis of repeated measurements from permanent sample plots (PSP);

(ii) through stem analysis (Bruce, 1926).

The most widely applied technique in plantations is the use of psp data, because of its simplicity and in most cases such data are readily available. In natural forests, however, stem analysis may be the only practical technique.

There are three methods by which site index curves (equations) can be generated (Clutter et al., 1983). These are:

(a) guide curve

(b) difference equation

(c) parameter predictions

In most situations one method is usually clearly superior to the others, or else a combination of methods could be applied to good effect. The Douglas fir data analyzed here were from permanent sample plots (PSP's) measured and remeasured several times. Analysis showed that the difference equation and Schumacher form gave the best overall predictions. Site index equations are classified into three types: anamorphic, polymorphic - disjoint and polymorphic - non disjoint (Clutter et al., 1983; Boardes et al., 1984).

This study has explored environmental factors associated with site quality variation in Douglas fir grown in the South Island of New Zealand. It was noted that site index, the mean top height at age 40 years, was not the best predictor of the site productivity. Site index equations developed in this study and their relationship to site quality are described in detail in chapter 4.

2.5 Modelling Philosophy

There are at least three ways one can review modelling philosophy: one is to examine modelling objectives or what the modelling is characterizing (Munro, 1974); the second is to focus on the condition of the forest that is being modelled (Bruce and Wensel, 1987); and the third is to look at the mode of action of the model (Garcia, 1988).

Past attempts by researchers to distinguish particular modelling objectives rather than a particular modelling philosophy contributed to a confused nomenclature of forest growth and yield models (Munro, 1974). Models of this sort have appeared in the literature in many forms. Although empirical growth models differ widely, common basic elements appear in most of them. Estimates are made of the changes over time of tree diameter, height, form, volume, or all of these variables, and also changes in stocking (Bruce and Wensel 1987). The following sub - sections discuss and elaborate the nomenclature which frequently have appeared in the modelling philosophy.

2.5.1 Classification According to Objectives of the Model

Munro (1974) developed a classification in which growth and yield models are classified according to their general objectives or according to what they model.

(1) individual tree growth models

  (i) individual tree - distance dependent models,

  (ii) individual tree - distance independent models.

(2) Whole stand growth models.

(3) Diameter distribution models.

The paragraphs below look at this classification and the advantages and disadvantages of each approach.

2.5.1.1 Individual Tree Growth Models

(i) Individual Tree - Distance Dependent Models

The main concept behind this kind of model is based on the postulate that the amount of competition to which a tree is subjected is proportional to the amount the competition circle of a subject tree is overlapped by competition circles of neighbouring trees. Competition circles are defined as some function of the size of a tree. The actual amount of competition has been expressed by different researchers in units of area, circumference, or angles. The first such model to appear in the literature was that of Newham (1964), other examples include Lee (1967), Mitchell,(1969), Lin (1970), Arney (1972) *et cetera*. Individual tree - distance dependent models use individual tree values as inputs, which are then aggregated to provide estimates of stand growth and yield. They are capable of producing very detailed information about the structure of a stand. For example, potentially powerful uses result from those which incorporate crown dimensional increments (*e.g.*Mitchell, 19691; Arney, 1972), as they include studies of tree to tree competition, pruning impact, insect defoliation, top die back, bole form change and mistletoe infections. Clearly, they offer a potential to examine the effects of various cultural programmes and their interactions, such as thinning, spacing and fertilization in a very detailed way.

The major disadvantage of all distance dependent modelling is the difficulty of calculating a meaningful biological measure of competition and the high use of computer resources. Spatial information (coordinates), elevation, aspect, stem charts of the tree, for example, must be obtained in the field and supplied as model inputs. Such information is expensive to acquire and is not available for any but the most intensively monitored permanent sample plot (PSP) system. Most models of this type provide artificial generation of tree spatial distribution. When this is done, strong arguments can be made to suggest that the model is functioning as a distance - independent one. Nevertheless, with the development of modern aerial photographs and mapping equipment the cost of acquiring the needed information can be significantly reduced. Today, these models, are not of great utility because advances in individual tree distance - independent models show that much of the information can now be obtained without inter - tree distance data, specifically through knowledge about spacing, thinning and fertilization. The effects of these cultural operations can be mostly evaluated more efficiently for operational purposes with distance - independent models.

(ii) **Individual Tree - Distance Independent Models**

The basic difference between this and the previous kind of model is the absence of a measure of distance between trees. In individual tree - distance independent modelling, trees are growing with respect to several tree characteristics, individually or in groupings of similar diameters, according to some mathematical functions. These models have ranged from simple regression as in Lemmon and Schumacher (1962) to extremely complicated stochastic models such as the one proposed by Dress (1970). The most referenced and used model in this category is PROGNOSIS developed by Stage (1973): this model offers great potential, particularly through its designed ability to function as a feedback mechanism for localization of more general models. The major advantage of individual tree, distance - independent models is the elimination of the necessity for stem charts, which results in fast computing and permits testing of many alternative hypotheses of management. Models with this property are essential in the development of management decision making tools such as RAM (Navon, 1971). The disadvantage of individual tree, distance - independent models is their inability to predict the growth of a specific single tree with any reliability; consequently they cannot be used to forecast the crown shape, crown development, and bole shape changes or defoliation in individual trees.

2.5.1.2 Whole Stand Growth Models

Whole stand models are also referred to as whole stand, distance - independent models. Stand models have a common objective, namely to produce at some point or points in simulated time, summary tables which indicate the state of forest stands on a per unit area basis. That is, they use stand variables as inputs and produce stand outputs, such as age, basal area per hectare and stems per hectare.

Whole stand models are the most widely represented kind of growth model in forestry modelling. For many years, forest scientists have investigated stand growth by regression functions to express stand growth under prescribed management. Computer capability has enabled the development of complex models utilizing complicated mathematical functions which permit the solution of yield functions based on virtually unlimited parameters. Examples of such Douglas fir models are Myers (1971) model which is used operationally by the United States Forest Service, Hoyer's (1972) model which simulates forest management practices in Washington (USA), and SIDFIR (Law, 1990) which simulates the growth of the species in the South Island of New Zealand,

All are designed with the specific objective of producing managed stand yield tables. Most forest enterprises depend on whole stand models to provide necessary stand information for economic analyses. The main advantages of whole stand modelling are its ability to utilize conventional inventory information, fast computation time and simplicity of operation. The disadvantage is that specific individual tree or tree class information is totally lacking.

2.5.1.3 Diameter Distribution Growth Models

Diameter distribution models occupy an intermediate position between the whole stand and individual tree models in terms of state description detail, computational cost, and information requirements (Garcia, 1988). Diameter distribution models should be constrained to operate as whole stand models, but with the additional ability of inputting stand level information (variables) to produce not only stand level statistics but also diameter distributions of trees to aid in forecasting size class information. The use of mathematical equations by foresters to predict diameter distributions goes back as early as 1898, when de Liocourt (*in Meyer and Stevenson, 1943*) constructed a diameter distribution model for all - aged stands using geometric progression (Meyer and Stevenson, 1943). In 1943, Meyer and Stevenson successfully constructed a diameter distribution model following de Liocourt's theory but through use of the exponential distribution. Since that time much attention has been given to diameter distribution modelling because it provides a detailed structure of the stand in terms of size classes, an important requirement for financial analysis.

Probability density functions have been the key to the generation of diameter distributions. Examples include Gram - Charlier series (Meyer 1930), Pearl - Reed growth curve (Osborne and Schumacher 1935), Johnson's S_b distribution (Hafley and Schreuder, 1977), Gamma distribution (Nelson, 1964), Beta distribution (Clutter and Bennett, 1965), Weibull distribution (Bailey, 1972), and others. The most used distribution today is the Weibull function, because it offers the following desirable properties with respect to forest stand categorization:

(i) relative ease of mathematical manipulation (Bailey and Dell, 1973),

(ii) it has a closed form (Bailey and Dell, 1973; Clutter and Belcher, 1978),

(iii) flexibility of the model (Johnson and Kotz, 1970).

The early diameter distributions were derived by regressing the probability density functions directly to the stand variables such as site index, stems per hectare and age, in the so called parameter prediction technique. This produced inconsistent estimates of stand values between the diameter distribution model and the whole stand model (Clutter and Belcher 1978). Diameter distribution models are useful only when they are compatible with the whole stand model, because consistent estimates of various stand yield variables need to be derived. Compatibility has been achieved by employing a parameter recovery method, with which method the parameters of the probability function are estimated implicitly from stand estimates (Cao and Burkhart, 1984; Knoebel et al., 1986; Boardes et al.,1987). In growth and yield studies today, whole stand models are in the majority because their level of aggregation has been reasonably easy to work with. By pooling all stems together, a lot of variation due to genetics, site, climate, and the like are absorbed, making the fitting of functions easier.

A potential difficulty with tree size distributions arises from the spatial correlation of tree sizes: over very short distances there is usually a negative correlation due to competition: over longer distances, microsite similarity causes a positive correlation, decreasing with time (Garcia, 1988). This implies that tree size distribution must vary with area of land considered. In particular, the variance must vary with plot size, and distributions derived from sample plots are unlikely to apply to the whole stand. This aspect has been ignored by growth modellers although its importance has long been recognized in forest sampling (Garcia, 1988). Experience has shown that for estimating stand variables only minimal gains in precision are attained by these models and that they may not be justified considering the higher costs involved. Diameter distribution modelling is thus a compromise, one that has proved very effective (e.g. Clutter and Allison, 1974; Alder, 1979; Bailey et al., 1981).

2.5.2 Condition of the Stand Being Modelled

Bruce and Wensel (1987) recognized the relevance of Munro's classification just described, but they also put emphasis on the applicability of different models according to forest condition being modelled and on the purpose of the model. Stands with a uniform progression of frequencies in size classes throughout their range can be characterized in more ways than stands without them. For example, a single tree - distance dependent model with some modifications can be used in uniform and non-uniform stands, but not particular models for both situations. In even-aged stands, density and hence competition can be evaluated on a stand basis in terms of the state variables, namely basal area per hectare, volume per hectare and stocking. Uneven aged stands require a detailed description of individual trees and are seldom successfully simulated by simple models developed for regular stands. To model most stands that are irregular, the growth of each tree must be estimated individually, because no single measure of stand density can be used to represent the competition affecting individual trees. The best solution is to use an individual tree distance - dependent growth model where the size, vigour, and proximity of neighbouring trees are evaluated. This technique can also be used for regular stands, but it is not always necessary because they are expensive, while whole stand models are a cheaper alternative with an acceptable degree of accuracy in certain circumstances.

2.5.3 Mode of action of Growth Models

Garcia (1988) put emphasis on the function of the models themselves. Essentially, the evolution over time of any system can be modelled by specifying: (i) an adequate description of the system at any point in time; (ii) the rate of change of state as a function of current state and of current value of any external control variables (Garcia, 1988). The state is the stand/tree *(N/ha, G/ha,* h_{100},*)* and the external factors are time, climate, altitude, aspect, soil, cultural treatment *et cetera.* Growth models can be quantified as static or dynamic, depending on how they function.

(1) **Static Growth Models (Alder, 1980)**

Static growth models attempt to model the development over time of quantities of interest (volumes, mean diameter). This approach to modelling falls short of a capability to predict the rates of change: nevertheless, they can produce good results for unthinned stands or stands subject to a limited range of standardized treatments for which long - term experimental data are available. Examples of these are the Forestry Commission Management Tables (Johnson and Bradley, 1963) and South Australian Yield Tables (Lewis et al., 1976).

(2) **Dynamic Growth Models**

The modern approach to modelling relies on the capabilities of dynamic models. Unlike static models, dynamic models forecast growth over a wider range of external factors (such as initial spacing, various thinning and pruning intensities, and fertilization). Instead of modelling directly the course of values over time, these models predict rates of change under various conditions. The trajectories over time are then obtained by adding or integrating these rates. Thus, the Munro (1974) classification refers to dynamic growth models.

2.5.3.1 Summary

Prospective uses influence the choice of growth models. If for example, only an estimate of total volume is required at a given time, then little attention need to be paid to the irregularity of the stand. If a prediction of change of inventory by size class material is needed, then a diameter distribution model of some form will be adequate. Nevertheless, if some estimate of change of quality is required, each tree may have to be treated individually in the computer model, irrespective of the stand regularity. Comparisons of alternative treatments, especially those not previously applied and observed, may be inaccurately predicted, no matter how it was developed (Bruce and Wensel, 1987).

2.6 Modelling Approaches

The major drawback of least - squares methods is that the nature of forestry data in terms of repeated measurements have correlated errors. The application of generalized least - squares (Ferguson and Leech, 1978; Davis and West, 1981) on permanent sample plot data was an effort to remove the bias of estimating standard error of parameters, which occurs when ordinary least - squares techniques are applied. Generalized least - squares changed the parameters very little, however, and the technique has not been of much significance in forest growth and yield modelling.

Analysis of growth and yield can proceed using various modelling alternatives: stand - level only, diameter distribution, distance - dependent tree - level or distance - independent tree - level (Munro, 1974; Bruce and Wensel, 1987). The methodology used by many researchers is to develop regression equations by single variables (Clutter, 1963; Sullivan and Clutter, 1972; Smalley and Bailey, 1974;). Garcia (1979, 1984b, 1987) has improved growth and yield modelling through introducing *stochastic differential equations*. The stochastic differential equation models have different mathematical properties and have been attracting considerable interest throughout the world (Ware et al., 1988). In this approach, the state variables such as basal area per hectare, stems per hectare, and others are simultaneously projected over time. Models of this kind have proved to be satisfactory to use in practice (Garcia, 1984b; Dunningham and Lawrence, 1987). Nevertheless, stochastic differential equations like any other equations need to be subject to biological and statistical tests when fitted to data. Moreover, simultaneous estimation of parameters of different equations can be very restrictive. This approach has not, therefore, been adopted here. The approach used in this thesis is to model each stand variable singly, because this allows a wide range of equation forms to be evaluated for each individual variable that needs to be predicted.

2.7 Growth and Yield Models for Douglas fir

Growth and yield models for uneven - aged stands of Douglas fir have been developed in North America (Curtis et al., 1981; Newham and Smith, 1964; Bruce et al., 1977; and others). The Weyerhauser company in USA grows Douglas fir in plantations and has re-measured data that can be adequately modelled. The methodologies adopted in North America are also applicable in New Zealand, but judging from graphs of crop production these models would not be applicable to New Zealand conditions because of disparity of climatic and other growth factors.

The first Douglas fir growth and yield model in New Zealand was DFIR, developed by Mountfort (1978), specifically for Kaingaroa forest. NFIR was devised in 1979 to cater for production of that species in Nelson. Calibration of DFIR gave rise to DFPP, RODF, and SDFIR (Law, 1990). The first complete Douglas fir models in New Zealand were DFCNIGM 1 and DFCNIGM 2, of which are both whole stand models created in 1989 (Liu Xiu, 1990). DFCNIGM 3 is a diameter distribution model compatible with DFCNIGM 2, which was created in 1990 (Liu Xiu, 1990). DFCNIGM1 and DFCNIGM2 endeavour to identify the presence of intra - regional or temporal differences among subsets of data and to make due allowance for such effects, such as disease infection.

SIDFIR (Law, 1990), was an attempt to develop a whole stand growth and yield model for all Douglas fir grown in the South Island of New Zealand. SIDFIR does not account, however, for existing regional variability. DfirStand described in chapter 4 of this thesis is a whole stand model for Douglas fir grown throughout the South Island of New Zealand, the model incorporates both the existing local and regional adaptations.

2.8 Localising Growth and Yield Models

There has been a tendency in some countries to develop and use growth and yield models that are specific to increasingly restricted sub - populations. In New Zealand, for example, there are currently about 16 models in use for one species, radiata pine alone, when 20 years ago there were 2 (Whyte et al., 1992). Statistically one may argue that stratification of a large population into smaller components will result in more homogeneous sub - populations, but this is not always justified unless analysis of covariance or some other appropriate technique can confirm that all the parameter estimates of a growth forecasting equations are unequal across sub-populations. General forest growth projection systems are often developed for large geographic regions (e.g. PROGNOSIS, Stage, 1973; STEMS, Smith, 1981; SIDFIR, Law, 1990, *et cetera*). Developed in this overall way, however, these models will not necessarily provide adequate sensitivity of estimation for sub-regions (e.g. counties, forest districts and wood supply centres) for the following reasons.

(1) Existing Variations

Regional growth models usually do not fully account for sub-regional site quality, stocking variability, genotypic variability, local climatic fluctuations, growth variations over time, to cite just some examples. These unexplained factors may well average out for whole region estimates, but for sub-regional estimates they may not, resulting in estimates that are biased.

(2) Operational fall-down (Bruce, 1977).

These occur when a projection system has been derived from data acquired from permanent or temporary sample plots that are located in uniform stands that are undamaged, and are of very high quality. When this system is used to project stands that are not maintained under the same optimal conditions, predicted growth is commonly found to be higher than observed growth.

(3) Silvicultural Practice

For many regions and species, models have not been developed for different cultural regimes (fertilization, genetic improvement, site preparation, thinning, *et cetera*). Under such circumstances, a means for at least partially accounting for treatment response must be developed. Parameter estimates of a regional model estimated in accordance with a given range of silvicultural practice, may not necessarily give unbiased predictions if used in a sub-region that can be characterized with different parameters, arising from different regimes. Various techniques have been used for localising regional models or regional estimates to suit local conditions. These are described in the next sub-sections.

2.8.1 Stratification

Stratification involves modelling each different stratum individually. To justify this an hypothesis that all parameters of each stratum are different from those of other strata must be conducted.

2.8.2 Simple Means Ratio

This technique is probably the simplest: if P_r and $\bar{P}_r$ are taken to represent a regional prediction and the mean of regional prediction respectively, while P_s and $\bar{P}_s$ are corresponding sub-regional predictions and means, then the adjusted regional prediction to suit sub region can be given by

$$P_r' = P_s \times \frac{\bar{P}_s}{\bar{P}_r}$$

where, P_r' is the adjusted regional mean. Smith (1981) successfully used this technique to localise estimates of individual tree annual diameter growth provided by the regional growth projection system STEMS. Provided that the ratio estimator $\frac{\bar{P}_s}{\bar{P}_r}$ is determined accurately, this method can achieve good results in the short run.

2.8.3 Regression Revision to Adjust the Parameters

Regression revision has been employed to adjust some of the parameters of the regional model. For example, PROGNOSIS, Stage (1981) used this technique in localising the intercept of the model, while other parameters were kept constant. The disadvantage of this method is that, there has been no formal procedure for the technique and the parameters to be localised depend on the assumptions of the modellers themselves.

2.8.4 Bayesian Methods

Bayesian methods are based on the probability principle that, a *posterior* probability distribution function (pdf) can be derived from a *prior* pdf. Bayes' theorem is of the following form,

$$f(\beta_0|Y_1) = \frac{f(Y_1|\beta_0)f(\beta_0)}{f(Y_1)} \tag{2.16}$$

where

> $f(\beta_0)$ is the prior probability function of a random parameter β_0 that has been obtained from fitting the growth function over an entire region;
>
> $f(Y_1|\beta_0)$ is a conditional probability density function of observations taken from the sub region given the parameter β_0;
>
> $f(Y_1)$ is the probability density function of Y_1, which need not to be explicitly considered;
>
> $f(\beta_0|Y_1)$ is the posterior distribution of parameter β_0 given Y_1 which contains information from the entire region $f(\beta_0)$, as well as a sub region $f(Y_1|\beta_0)$.

If the error component of the posterior information has mean 0, constant variance σ, and normally and independently distributed, then a posterior parameter estimate can be obtained that maximizes $f(\beta_0|Y_1)$.

Bayesian techniques have been used successfully in localising growth and yield models. Berkey (1982b) and Green et al. (1992) have shown that if parameter estimates of the global model are similar to those of the sub region model in some respect, then considerable gains in, say, root mean square (RMS) may be realized by using empirical Bayesian regression over the use of the ordinary least-squares regression. Berkey (1982b) has reported a reduction in (RMS) of about 50% in parameter estimation in fitting the Jenss (Jenss and Bayley, 1937) growth model to a sample of children. Green et al. (1992) have reported a reduction of more than 50% in RMS in simultaneous parameter estimation for Honduran pine yield models for sub-populations with different soil types. The use of Bayesian technique to adjust the parameter estimates through time is known as a sequential Bayesian procedure; the technique remains the same, except it is done through time. Gertner (1984) used this method to localise a diameter increment model taken from STEMS (Shifley and Fairweather, 1983). He demonstrated that the parameter estimates of a growth model change with crop development, and that they reach an asymptote as the crop matures.

2.8.5 Use of Dummy Variables

Dummy variables have also been employed to localise growth and yield models. This mostly involves formulating an ANCOVA problem in which dummy variables are incorporated in regression equations. The general approach has been demonstrated by, for example, Gujarat (1970). Monserud (1984b) used dummy variables to estimate specific parameters in site index equations of inland Douglas fir according to habitat types, while Ferguson (1979) used dummy variables to localise a basal area increment equation for five forests of radiata pine in the Australian Capital Territory (ACT). Because of their usefulness, the next paragraphs explain how dummy variables can be employed in localising growth and yield.

Given a set of 3 populations (regions) and one covariate, one can formulate an ANCOVA problem as set out in the following equations.

(1) Minimal model

$$Y_{ij} = \beta_0 \tag{2.17}$$

(2) Simple linear regression

$$Y_{ij} = \beta_0 + \beta_3 X_{ij} + \epsilon_{ij} \tag{2.18}$$

(3) Regression slope varied by region

$$Y_{ij} = \beta_0 + \beta_1 J_1 + \beta_2 J_2 + \beta_3 X_{ij} + \epsilon_{ij} \tag{2.19}$$

(4) Maximal model

In maximal model all parameters are varied by region.

$$Y_{ij} = \beta_0 + \beta_1 J_1 + \beta_2 J_2 + \beta_3 X_{ij} + \beta_4 J_1 X_{ij} + \beta_5 J_2 X_{ij} + \epsilon_{ij} \tag{2.20}$$

The coefficients are interpreted as follows:

β_0, intercept for population 1;

β_1, differential intercept for population 2;

β_2, differential intercept for population 3;

β_3, slope of Y with respect to X for population 1;

β_4, differential slope of Y with respect to X for population 2;

β_5, differential slope of Y with respect to X for population 3;

J_1, J_2, are dummy variables;

ϵ_{ij}, NID(0,σ^2).

Equations 2.17, 2.18, 2.19 and 2.20 are nested, so that, an hypothesis test can be performed on slopes β_4 and β_5. If β_4 and β_5 are statistically different from one another, (Ho: β_4=β_5), then 2.20 is the appropriate model. If Ho is accepted, however, then equation 2.19 is preferred. The usual assumptions about the error term ϵ_{ij} hold here, namely that NID(0, σ^2). Initially, the two dummy variables are assigned to any two populations, while one remains as the default. Provided that all possible combinations are tested there is no chance of missing out the best combination. If for instance, the dummy variables J_1 and J_2 are allocated to populations 2 and 3 respectively, while population one enters the model freely, and 2.19 is the best equation then 2.19 will have a common slope β_3, but with varying intercepts, β_0 for population 1, $\beta_0 + \beta_1$ for population 2, and $\beta_0 + \beta_2$ for population 3. Since it is assumed that the error term is normally and independently distributed with mean 0 and constant variance σ^2, this formulation allows straight-forward tests of hypothesis associated with the confidence limits of parameter estimates through use of statistical packages like SAS, which have the capability of sorting data to their respective sub-populations. Although the above example applies to linear regression, the same principles are applicable to non-linear models. The use of dummy variables thus provides potential capabilities for testing the justification of having different models for different sub-populations. Ferguson (1979) used generalized linear least-squares regression with 4 dummy variables to localise basal area increment model for five forests, mainly to represent different rainfall patterns.

2.8.6 Summary

If adequate information is available and if the accuracy needed by the user demands it, any of the above methods for localising growth and yield can lead to satisfactory results, subject to evaluation and validation of the adjusted models. However, the applicability of Bayesian techniques relies on obtaining prior information, requires high statistical knowledge, and sometimes special algorithms for solving the parameters. These methods are therefore costly, although capable of providing very accurate estimates (Berkey, 1982b; Green et al., 1992). The gain in precision may accrue, however, only when the assumptions of Bayes' theory are met; that is, there needs to be similarity between the prior and posterior information, otherwise they may lead to costly unjustified results.

This study aims specifically to incorporate *locality adaptation* by including *dummy variables* among the predictor variables. The approach adopted here, to pool all the data, then assign dummy or other predictor variables to respective regions to account for locality variation, is preferred over the other two approaches because it should provide a better basic understanding of variation that is necessary for testing the stratification and Bayesian methods. The latter two could well be evaluated in studies following this.

Chapter 3

Data and Data Analysis

3.1 Tree Volume and Taper Data

3.1.1 Sources of Data

The data used in tree volume and taper modelling were sampled from 16 forests in the Canterbury region, 10 in Southland and one in Nelson (these are summarized in appendix A.1). The data were retrieved mainly from Forest Research Institute archives. Procedures for taking sectional measurements and making data entries are explained in detail by Gordon, (1985). Initial examination of the data showed that the Nelson region was represented by 32 trees only. This number was insufficient to represent Nelson, where 38% of the area of Douglas fir planted in the South Island is located. Thus, an additional 50 trees were measured in Golden Downs to strengthen the data base.

3.1.1.1 Selection of Additional 50 Trees

The following procedure was followed in the sample selection of the extra 50 trees, just mentioned. Sample trees were obtained systematically throughout the dbhob range, the allocation across the dbhob range being in roughly equal numbers by size class. Sample trees included all merchantable sizes, emanating from crops aged 40, 35, and 13 years. Samples from the tails of the observed diameter distribution were deliberately included because of their importance in estimating the coefficients effectively. Individual trees were selected from several stands throughout the forest according to the criteria set out below:

(i) reasonably straight stem with less than a 10 degree lean;

(ii) no leader die back , nor broken top, nor stem forking;

(iii) unblemished dbhob, unaffected by forking, fluting, abnormal taper, concavity, callous growth or scar tissue;

(iv) crown class normal for the tree size, very suppressed or grossly emergent trees being excluded from consideration.

This procedure aimed to supply a sample from which a volume function representative of the main tree population growing in the locality could be obtained.

3.1.2 Nature of Tree Volume and Taper Data

Sectional measurements were taken along tree stems in the manner prescribed by the Forest Research Institute (Ellis, 1979). The average number of sectional measurements per tree was 10, which constitutes an adequate set of repeated measurements for taper definition purposes.

3.1.3 Quantity and Quality of Sectional Measurement Data

Tables 3.1, 3.2 and 3.3 below show the frequency of diameter - height classes of trees separately for each region, Canterbury, Nelson and Southland respectively: Table 3.4 summarizes the whole data set used for tree volume and taper modelling. Table 3.4 shows that the initial data set consisted of 641 trees and about 7000 measurements, while Tables 3.1, 3.2 and 3.3 are distributions of diameter and height - classes represented in the samples from the three regions. The range of dbh - height classes of these data appears to be adequate, and their quality is good for the purpose of modelling volume and taper. To construct a new volume table the minimum recommended sample size is 100 trees (Gordon, 1985), though the sample may have to be increased if the variation in tree form is high, until the resulting equation has an acceptable reliability in terms of both the accuracy and precision required by the user. For a volume function that has validity for a greater area, several hundred or even thousand may be required according to Loetsch and Haller (1973), but this claim needs to be challenged on the basis of research reported later here.

Table 3.1: Dbh—Height Classes for Trees From Canterbury Region

Dbh class	Height Classes (5 m)									
(5 cm)	5	10	15	20	25	30	35	40	45	
	Number of trees									Total
5	10									10
10	5	13	17	1						36
15		10	51	13						74
20		1	34	72						107
25			8	60	1					69
30			4	12	1					17
35			3	4	1					8
40				1	1	4				6
45					1	8				9
50					1	5	1			8
55					1	7	2	2		11
60						4	1	2	1	8
65						3		2	1	6
70							2	1	1	4
75						1		1		2
80							3			3
85							1	3	1	5
90							1			1
Total	15	24	117	163	6	32	11	12	4	384

Table 3.2: Dbh—Height Classes for Trees From Nelson Region

Dbh class	Height Classes (5 m)								
(5 cm)	5	10	15	20	25	30	35	40	
	Number of trees								Total
5									
10									
15		1	3						4
20			7	2	4	1			14
25			5	1	5	6			17
30			1	1	7	8			17
35				3	1	4	1		9
40				1	3	3	6		13
45						1	1	1	3
50							2		2
55							1	2	3
60									
65									
70									
75									
80									
85									
90									
Total	-	1	16	8	20	23	11	3	82

Table 3.3: Dbh—Height Classes for Trees From Southland Region

Dbh class (5 cm)	Height Classes (5 m)								
	5	10	15	20	25	30	35	40	
	Number of trees								Total
5	20								20
10	12	1	1						14
15		5	1	5	1				12
20		6	5	13	3				27
25			4	5	14				23
30			3	7	11	4			25
35				5	8	5			18
40				2	8	1			11
45			1	1	4	1			7
50					1	4			5
55						1			1
60							2		2
65								1	1
70						2			2
75						1	1	1	3
80								1	1
85							1		1
90							1		1
95						1			1
Total	32	12	15	38	50	20	5	3	175

Table 3.4: Summary of Number of Trees and Sectional Measurements

Region	Number of Trees	Number of Sectional measurements
Canterbury	384	4241
Nelson	82	984
Southland	175	1771
Total	641	6996

An efficient procedure used to evaluate the efficacy of a volume equation is to measure the bias by diameter classes (Honer, 1965) and this technique is used later to justify the challenge to that European research assertion. The data cover all sites which grow substantial amounts of Douglas fir in the South Island of New Zealand (except for Westland), and represent a considerable range of dbh, height and age classes as well as different silvicultural histories. The spread of tree volume and stem data could have been even better, however, if they had also had disease information on Phaeocryptopus gaeumannii (Gilmour, 1966; Hood and Kershaw, 1973, 1975), fewer measurement errors, freedom from correlated errors and a full description of silvicultural practices applied to individual sample trees (e.g pruned vs unpruned stems). Measurement errors were considered as a possible source of variation while other factors were catered for as local adaptations, as will be explained later.

3.1.4 Sources of Variation

The main sources of variation in modelling at the tree level appeared to arise from:

(i) measurement errors

(ii) sampling errors

(iii) correlated errors

(iv) locality variations

Each of the above four factors was carefully considered and methods to eliminate or reduce their effects are reported in the next subsections.

3.1.4.1 Measurement Errors

Possible measurement errors were first identified by graphically plotting the data and observing outliers from the raw graphs. Any anomalous data were identified and then either corrected whenever the true values were available, or else they were removed from subsequent analysis, obviously in error but without objective evidence to correct them. Likely errors were also determined through fitting preliminary equations to the data and isolating those observations that had residual values of more than 3.5 RND (units of standard deviation) for detailed scrutiny. These tree measurements were again checked against the original data and corrections made, wherever feasible and if absolutely clear - cut. If no changes could be made, the measurements were retained despite their large deviations from average trends.

3.1.4.2 Sampling Errors

The ideal sampling approach for volume and taper function construction should provide representative estimates of each of the coefficients in any adopted equation form, and have as low standard errors of prediction as the sample size and population variability would allow. The aim should be, therefore, to obtain adequate representation in the sample of each dbhob and average volume within each class. The sampling process adopted in collecting the data fell a little short of such an ideal, as it relied largely on the one used by FRI. Each distinct stand in which sample trees were recorded was treated initially as a separate stratum. At the data processing and analysis stage, one or more strata were amalgamated to form groups. This approach allowed data to be grouped on the basis of observed differences rather than pre - allocating stands into strata. The construction of the volume equation was then done using these groupings, leading to volume equations T15, T120, and T228. One large aggregation which consisted of data from all over New Zealand was also formed, namely T136. Inadvertently, the Nelson region was not adequately represented and thus, an additional 50 trees were sampled as previously mentioned, with the specific aim of plugging the gaps in the Nelson region data.

In this study, the data were grouped according to their region of origin, but the formula used for the computation of the standard deviation of the tree volume equation was that of a simple random design. Tree locality effects were taken into account through use of two regional dummy variables, which acted as filters for the data according to their region of origin in the parameter estimation process. The formula used in computing the root mean square (RMS), which is the standard deviation of the residuals was

$$RMS = \sqrt{(\frac{\sum_1^N (y_i{}^2 - \bar{y}^2)}{(N - k - 1)})} \tag{3.1}$$

where N is the number of observations and k is the number of parameters in the equation.

3.1.4.3 Correlated Errors

Tree sectional measurement data consist of a series of repeated measurements on the same stem. When used to derive functions representing stem taper, they are obviously not altogether independent but correlated to some extent. The usual method of ensuring that such correlated errors are minimized is the method of randomization (Fisher, 1947), but this is virtually impossible to adopt with data used for analysis of stem taper data. Although such data are usually processed as if they were independent, this must be done with caution because the standard error estimates of the parameters may be biased and conventional statistical hypothesis testing becomes invalid. The errors arising from estimating least - squares regression coefficients from such data have a component of error ascribable to the degree of their correlation. Standard hypotheses tests on the parameter estimates are inadequate for testing the goodness of fit for equations developed from such data. Thus, in this study plots of residuals and the univariate procedure, as described in SAS (SAS Institute Inc., 1988), were used to ascertain those parameter estimates that appeared to be statistically worthwhile.

3.1.4.4 Locality Variations

Data for this study come from a wide range of localities as explained in section 3.1.1 Previous modelling of tree volume and taper equations for Douglas fir in New Zealand considered locality as a source of variation to be accounted for, but generally grouped the data according to only the year of sampling, a traditional approach that leads to national volume tables (FRI-Ministry of Forestry, New Zealand, 1992). The approach previously adopted was either to put an average through all available data irrespective of locality, silvicultural history or provenance, or to separate the data to form individual equations for combinations of locality and silvicultural history, but there is room for improvement in both these approaches. Stratification in all sorts of forestry applications is an effective tool widely used to derive various estimates for populations with distinctly grouped different characteristics (see Freese, 1962). Modelling each stratum individually has often been regarded as appropriate in obtaining unbiased estimates of parameters for regression relationships (Loetsch and Haller, 1973; Steel and Torrie, 1980). The approach taken here followed their recommendations, but in contrast to the previous approaches, it used all available data, and recognized possible sources of variation by means of explanatory or dummy variables.

3.2 Data for Stand Level Modelling

3.2.1 Sources of data

Data for stand level modelling were obtained from permanent sample plots maintained by F.R.I, Rotorua. These data originated from 4 forests in Canterbury, 3 forests in Nelson, 9 forests in Southland, and relatively fewer data from 5 forests in Westland. The data for the permanent sample plots were totally separate from and, hence, independent of the sectional measurement. Differences in climate among the four regions were not directly considered, but the impact of environmental variables, reflected in indices such as site class, altitude above sea level, distance from the sea, were evident from a comparison of the mean and extreme values of age, mean top height, basal area per hectare and volume production per hectare of the raw data and so revealed trends worth investigating. Table 3.5 summarizes this information. Figures in parenthesis in column 1 of Table 3.5 refer to the number of observations. Appendix A.2 shows in full the list of the forests by region from which the data were sampled.

3.2.2 Quantity and Quality of Stand Level Modelling data

Basic and derived variables were derived from 355 permanent sample plots. Table 3.6 shows the distribution of plots by regions. Routine validation of the basic variables h_{100}, G, N and V, was done for each plot through plotting the data over time and identifying abnormal growth patterns. Preliminary statistical analyses were also conducted to ascertain the reliability of the data, including use of the procedures PROC MEANS, PROC COMPARE, PROC PLOT, PROC FREQ and PROC UNIVARIATE in SAS package (SAS Institute INC., 1988). Where possible, errors were corrected and the corresponding measurement included, but in some cases suspicious measurements were deleted from the data base, including the following.

(i) Negative C.A.I. for any one variable . Such data were isolated from analysis, but other data in the plot without abnormalities were accepted.

(ii) Unnaturally high mortality which resulted in a decrease of basal area with time, (probably caused by wind, drought or disease).

(iii) Coding fault where the true values could not be ascertained.

(iv) Fertilized plots were discarded from this analysis because there were insufficient of them to quantify the effect of different fertilization treatments.

(v) Plots with fewer than three measurement were also discarded, because they will have only one set of repeated measurements. One or two measurements alone are not adequate for studying growth and yield.

Table 3.5: Summary of Mean and Extreme Values Extracted From Psp Data

Region	Variable	Mean	Minimum	Maximum
	T (years)	32.8	9.0	61.0
	h_{100} (m)	22.9	2.9	39.3
Canterbury	Altitude (m)	326.1	150.0	790.0
(241)	G (m^2/ha)	46.3	0.43	116.2
	V (m^3/ha)	414.8	1.4	1505.4
	T (years)	27.5	7.0	58.0
	h_{100} (m)	22.9	5.6	47.8
Nelson	Altitude (m)	438.1	183.0	625.0
(929)	G (m^2/ha)	42.2	1.18	109.4
	V (m^3/ha)	413.7	18.4	1723.6
	T (years)	33.6	7.0	78.0
	h_{100} (m)	24.1	4.1	47.4
Southland	Altitude (m)	251.1	50.0	625.0
(449)	G (m^2/ha)	51.3	1.1	141.7
	V (m^3/ha)	482.5	0.4	1774.9
	T (years)	26.9	5.0	59.1
	h_{100} (m)	18.8	1.9	37.5
Westland	Altitude (m)	229.0	0.0	330.0
(225)	G (m^2/ha)	29.9	0.01	123.8
	V (m^3/ha)	235.9	0.05	1458.5

Table 3.6: Distribution of Permanent Sample Plots

Region	Number of Permanent Sample Plots	Number of measurements (Number of Growth Periods)
Canterbury	63	241
Nelson	135	929
Southland	112	449
Westland	45	225
Total	355	1844

The screening of the data resulted in 17 plots being disqualified from analysis. The remaining 338 plots consisted of around 1600 measurements for each variable. This quantity of data can be considered more than adequate for modelling growth and yield.

3.2.3 Structure of Stand Level Modelling Data

Permanent sample plot data consist of repeated measurements of several variables taken at different times. The term 'repeated measurements' means that N experimental subjects are observed on each of k successive occasions that possibly correspond to different experimental conditions, the i_{th} subject yielding y_{ij} on the j_{th} occasion. In developing suitable equations by regression methods from such data researchers often treat the measurements as independent, and then use the formulae for independent measurements to determine the standard errors of the parameter estimates. But because these measurements are correlated, residual sums of squares are underestimated and so too are the standard errors of the parameter estimates. Nevertheless, the analysis provides an adequate basis for predicting pre - assumed relationships between the independent and dependent variables (Jensen, 1982), while analysis of residual patterns can help ascertain the goodness of fit.

3.2.4 Sources of Variation in Stand Level Modelling

The main sources of variation considered in the analysis were:

(i) altitude

(ii) locality and factor interactions

(iii) thinning history

(iv) correlated errors

The approach taken was to incorporate factors into equations that reflect these sources of variation for all the available data and to analyze the effects of each of these over time, rather than have separate equations for each behavioural factor. The methods used in incorporating these factors are reported in the next subsections.

3.2.4.1 Altitude

Altitude is important in determining site quality, as it relates to other factors such as soil fertility, temperature and drainage. All these factors are likely to contribute to the growth of trees at a given site , but altitude may reflect all of them in combination to some extent. Woollons and Hayward (1985), for example used altitude as an independent variable in their site index equation for radiata pine in the Central North Island of New Zealand. The range of altitude sampled in the data here was from 0 to 790 m above sea level, which has been consistently found here to have a great influence on mean top height growth and, hence, site index. It was also possible in this study to show that basal area production was related to altitude above sea level. Details of the equations used and results obtained are explained in chapter 4.

Regional distribution of the measurements according to initial stocking, and altitude of the crops are represented in a two way table (altitude classes vs initial stocking), Table 3.7.

3.2.4.2 Locality and Factor Interactions

Local adaptations expressed as locality growth factors have for a long time been considered synonymous to site quality in even aged stands. When the factors which constitute site quality (soil type, soil moisture, soil nutrients elevation, temperature, aspect, distance from sea, and many others) are added up, they amount to a large number even before any interactions are considered. In this study it was shown that locality could be adequately represented by dummy variables rather than equating it to numerous individual factors of site quality. Site index (mean top height at age 40 years), moreover, was significantly less effective than these dummy variables in explaining differences in basal area production trends, although it was found useful for predicting mortality. Locality and site quality are therefore not necessarily interchangeable.

Table 3.7: Initial Stocking vs Altitude Classes

Region	Initial Stocking	Altitude Classes								Total
		≤ 100	200	300	400	500	600	700	800	
		Number of Measurements								
Canterbury	1736				80					80
	2268			6						6
	2315			3				8		11
	3086		36	69	21	18				144
Nelson	1680			12						12
	1736			21	127	81	48			277
	2315				25		3			28
	3086		7	26	194	37	323	25		612
Southland	1543					2				2
	1667					3				3
	1736	47			15	54				116
	2222		2							2
	2315					21		6		8
	2500			5	6	3				14
	2778		5	5						7
	3086	61	60	74	24	7		2		228
	3630		2							2
	3704			3	7					10
	4630	3	13							16
	6944		5	7	7					19
Westland	2314	24								24
	2500		3							3
	3086		38	139	21					198
Total		135	171	370	527	226	374	41		1844

3.2.4.3 Thinning History

Thinning is the most important silvicultural tool used by foresters for moulding the development of even aged crops. Until 1983 it was usual to have separate basal area equations for thinned and unthinned crops (Brender, 1960; Buckman, 1962; Schumacher and Coile, 1960; Pienaar and Turnbull, 1973; Pienaar, 1979; Clutter and Jones, 1980). Bailey and Ware (1983) were the first foresters to have a single equation for thinned and unthinned stands through use of thinning indices. Pienaar and Shiver (1984) later used an index of suppression to develop basal area models for combined unthinned and thinned slash pine stands. They concluded, however, that any basal area projection equation not allowing for growth response in thinned stands relative to unthinned stands might underestimate the yield of thinned stands.
Using data from thinned slash pine plantations, Pienaar et al., (1985) developed a basal area projection equation that incorporated the index of suppression instead of a thinning intensity variable (or thinning index). Murphy and Farrar, (1988) evolved a general technique for introducing thinning variables into basal area projection equations. In their study they concluded that the efficacy of adding a thinning term depends upon the accuracy requirements of the user.

The approach taken in this study is similar to that of Murphy and Farrar (1988), but with some modifications. Data from 171 thinned plots (742 measurements) were available for analysis, the main thinning regimes being shown in Table 3.8. The intervals between thinnings are set out in Table 3.9. Equations and definitions of thinning indices and how they were imposed in the basal area projection equation are described in chapter 4.

Table 3.8: Main Thinning Regimes Analyzed

Region	None	1st	2nd	3rd	4th	5th	6th	Total
	Number of Measurements							
Canterbury	135	19	21	11	24		31	241
Nelson	564	231	54	6	33	41		929
Southland	280	67	36	32	34			449
Westland	170	25	30					225
Total	1149	342	141	49	91	41	31	1844

Table 3.9: Thinning Intervals

Region	Interval between thinnings					
		2nd	3rd	4th	5th	6th
Canterbury	mean	6.1	6.2	9.0	4.0	5.0
	min	3.0	3.2	8.0	3.1	5.0
	max	13.0	11.0	10.1	9.0	5.0
Nelson	mean	3.6	5.0	6.4	6.0	
	min	4.2	6.8	5.0	5.0	
	max	10.8	7.8	7.0	6.0	
Southland	mean	9.8	7.6	8.1		
	min	3.0	1.0	1.1		
	max	29.9	11.2	12.0		
Westland	mean	5.5				
	min	6.1				
	max	3.3				

3.2.4.4 Correlated Errors

As already mentioned Psp data obtained from repeated measurements on an individual plot (in this case), contain correlated errors. In practice, these correlations are frequently ignored and parameter estimates of coefficients assumed to be unbiased, while accepting that the standard error of the mean, ($RMS/\sqrt(N)$ or $\sqrt(MSE/N)$), where RMS is the root mean square, and MSE is the mean square error, is lower than it would be for wholly uncorrelated data (Sullivan and Clutter, 1972). In growth and yield data, least - squares regression is considered to be adequate for parameter estimation, provided that the equations used logically represent the relationships intended between dependent and independent variables and for each parameter estimate the confidence intervals do not include zero. This thesis did not concern itself deeply with correlated errors, but qualification tests initially described in section 3.1.4.3 and later in more detail in section 3.4 were carried out on parameter estimates and residuals to ensure that the correlated errors did not result into equations with biased parameter estimates.

3.3 General Methodology

The methodology presented in this section applies to all three types of modelling: to tree volume, stem taper and stand level growth and yield modelling.

3.3.1 Formation of Data Sets

3.3.1.1 Tree Volume and Taper Data

Values for the following 10 variables formed the basic data set for the various analyses carried out with SAS on the VAX computer at the University of Canterbury.

RE — (Region): region of origin of data item.

FO — (Forest): forest or place of identification of data origin.

TREE — (Tree number): an identification number assigned to the tree within the region and forest.

DBHOB : diameter at breast height over-bark in cm for the tree.

SH: cumulative sectional height in m from the base of the stem to the top of the section.

SEDI: small end diameter inside bark in mm of section of the tree.

LEDI: large end diameter in mm of the section of the tree.

TH : total tree height in m.

V : tree total stem volume m^3 inside bark, computed from the addition of all sectional volumes. The volume of the first 0.15 m section from the ground was calculated as if it were a cylinder, while the volumes of sections above were calculated using the conic integral formula (Whyte, 1970).

One additional variable was derived from the existing variables, z, defined as the height in the tree relative to its total height.

$$z = \frac{(TH - SH)}{TH} \tag{3.2}$$

Using standard IUFRO notation equation 3.2 will be represented as in equation 3.3

$$z = \frac{(h - h')}{h} \tag{3.3}$$

where h is total tree height and h' is the distance from the ground to top diameter d': (h-h') can also be represented as (l), where l is the distance from the tip of the stem to top diameter d'. The tree data master file named TVTP.DAT is fully presented in appendix C.

3.3.1.2 Stand Level modelling Data

Values for 21 principal variables were extracted from PSP data held in F.R.I. archives and were similarly filed on the VAX computer. Because 21 variables amounted to too many for the 80 column system of data processing, the data were punched into three different data sets but under the same file name DFIRS.DAT. These data sets were named DATA A, DATA B, and DATA C, and comprised the following variables:

RE: region of origin of a data item.

FO: forest of origin of data item.

CP: compartment from which data were collected.

P: plot within compartment.

SP: sub-plot, where special treatments were applied (for example fertilization).

AL: altitude of plot in metres above sea level.

T_1: age in years at the beginning of a growth period.

T_2: age in years at the end of a growth period.

HT_1: mean top height in metres at age T_1.

HT_2: mean top height in metres at age T_2.

G_1: net basal area of the stand (m^2/ha) at age T_1.

G_2: net basal area of the stand (m^2/ha) at age T_2.

G_b: net basal area of the stand (m^2/ha) before thinning.

G_a: net basal area of the stand (m^2/ha) after thinning.

N_1: number of stems/ha at age T_1.

N_2: number of stems/ha at age T_2.

N_b: number of stems/ha before thinning.

N_a: number of stems/ha after thinning.

V_1: net volume of the stand inside bark (m^3/ha) at age T_1.

V_2: net volume of the stand inside bark (m^3/ha)at age T_2.

T_t: age of thinning in years.

Four additional variables were algebraically derived from the above original data base. These were:

S: site index in metres at age 40 (if values were available for mean top height at age 40, the mean top heights were recorded as site indices, else the mean top heights were derived by extrapolation using the mean top height projection equation).

d_t: quadratic mean diameter outside bark (cm) of trees removed in thinning.

d_b: quadratic mean diameter of the stand (cm) just before thinning;

X_t : thinning index.

$$X_t = 1 - \frac{d_t}{d_b} \tag{3.4}$$

The definition of thinning term variable and the way it was derived are explained later in chapter 4. The stand level master file data named DFIRS.DAT is fully presented in diskette in appendix C.

3.3.2 Data Format

The two master files TVTP.DAT and DFIRS.DAT were input in FORTRAN format. However, when fitting regression equations to data sets, it is often necessary to change the order, file name and data format to create smaller data sets which are subsets of the master file. This was required to enable the data to be read by analytical packages such as SAS in order to make the processing of the data easier and faster, and use less computer resources. The FORTRAN format of data input was selected because it was compatible with both the operating system, VMS, and the analytical package, SAS, that were used to perform various operations.

Because regression equations were fitted to single variables, separate files for each equation were created from the master files. New variables not in the files were created algebraically using SAS programmes and commands. Variables which could not be derived algebraically from the existing variables were input manually on the keyboard.

The following files can be found in diskette form in appendix C. The definitions of variables applied here are those given in section 3.3.1

VT.SAS

Data and SAS programme used to develop tree volume equation, with variables RE, FO, TREE, DBHOB, TH, v .

TPPD.SAS

SAS data and programme to create permanent SAS data used to develop taper equations with variables RE, FO, TREE, DBHOB, S, SH, SEDI, LEDI, TH, v.
A permanent SAS data file (as described in SAS institute Inc.,1988. Cary, NC., USA) was created for this analysis because this file was too large to be analyzed by conventional procedures.

TPP.SAS

SAS programme used to develop tree taper equation.

HT.SAS

SAS data and programme used to develop mean top height equation, with variables RE, FO, CP, P, HT1, HT2, T1, T2, S.

G.SAS

SAS data and programme used to develop net basal area projection equation with variables RE, FO, CP, P, G1, G2, T1, T2, T_t, S, X_t.

THIN.SAS

SAS data and programme used to develop an equation for predicting net basal area/ha after thinning: with variables RE, FO, CP, P, T1, T2, GB, GA, HT1, HT2, NB, NA.

V.SAS

SAS data and programme used to develop stand volume production equation, with variables, RE, FO, CP, P, V1, V2, T1, T2, HT1, HT2, G1, G2.

M.SAS

SAS data and programme for derivation of stem survival/ha function with variables RE, FO, CP, P, N1, N2, T1 ,T2, S.

3.3.3 Checking Reliability of Data

Analysis of residuals was used to determine the reliability of data. Appropriate equations were first fitted to the data, then all observations having RND (random normal deviate) more than 3.5 were regarded as outliers. More often than not the causes of outliers were blatantly obvious measurement or punching errors, but situations did occur where the obvious causes were unknown, and such measurements were merely categorized as suspicious. It was not possible to check measurement reliability, because the data were collected a long time ago, as long as 80 years ago, for some used in this study. Nevertheless, the importance of having reliable measurements and rigorous data checks during initial processing of information for growth and yield studies was recognized to as great an extent as possible.

3.4 Methods Used for Data Analysis

The two main standard analytical procedures used in this study are linear and non linear ordinary and weighted least - squares regression. Analysis of variance and univariate procedures were used to ascertain the goodness of fit of equations.

These analyses were conducted through use of procedures PROC NLIN, PROC REG, PROC GLM, PROC UNIVARIATE, and PROC MEANS of the SAS package (SAS institute Inc., 1988). Regression equations can be fitted variable sets of any sort, but in this study it was ensured that:

(i) the dependent and independent variables conform to biologically and mathematically sound relationships;

(ii) the functions used are of appropriate form to represent the intended relationship;

(iii) parameter estimates are free of apparent bias.

Various linear and non linear ordinary, and weighted least - squares were fitted to tree volume data, stem taper data and growth and yield data. As pointed out in sections 3.1.4.3 and 3.2.4.4, both tree taper and growth and yield data are associated with correlated errors, which means that conventional statistical analysis cannot be carried out without qualification of the results.

The following subsections explain the main analyses that were carried out in refining the statistical results and in ascertaining the goodness of fit of the equations.

3.4.1 Confidence Intervals

Terms involving parameter estimates were retained only if the parameter estimates were apparently significant at the 5% significance level, meaning that there was less than a 5% nominal chance that the confidence interval could contain zero (i.e., the lower and upper confidence limits are of the same sign). These nominal probabilities do not, of course, reflect true probabilities due to the aforementioned biased estimation of residual variance. They are used, therefore, only indicatively.

3.4.2 Graphical Residual Patterns

Residual charts were used to provide ocular estimates of their normality of errors. If the residuals are normally distributed the residual pattern about the zero reference line shows independent distribution; a bar chart of residuals portrays the shape of the normal curve over the interval of the data set. This visual description is often inadequate on its own, however, because the shape of the chart depends on the scales and class widths used.

3.4.3 Univariate Procedure

The UNIVARIATE procedure was used, therefore, to complement the ocular check of fit through reference to confidence intervals and residual charts. Several statistics were used as indicators in the UNIVARIATE procedure to complete the test of goodness of fit. These were:

- **Mean of Residuals**

The mean should either be zero or at least very close to zero if the equation is to produce a good fit, under the assumption that the residuals are normally distributed with mean zero and standard deviation σ, often represented as NID(0, σ^2)

- **Absolute Mean of Residuals**

The absolute mean of residuals is a measure of the average error prediction of the equation. This should also be very close to zero.

- **Skewness Coefficient**

The skewness of the normal distribution is zero. It is a measure of symmetry, in that it provides inference on the tendency of deviations to be larger in one direction than the other. If the skewness of the equation very much deviated from zero, the fitting was re - assessed. Negative values indicate a distribution with a long tail to the left and positive values indicate a long tail to the right. The unbiased skewness coefficient of the samples is calculated by

$$SK = [\frac{N}{(N-1)(N-2)}][\frac{\sum_{i=1}^{N}(x_i - \bar{x})^3}{s^3}] \tag{3.5}$$

where SK is the skewness coefficient, N is the number of observations and s is the standard deviation of the sample.

- **Kurtosis**

The early understanding of kurtosis was a relative measure of flatness or peaking of the distribution. The larger the value of kurtosis, the more peaked the distribution (Hafley and Schreuder, 1977). More recently, kurtosis has been defined as the heaviness of the tails of a population (SAS Institute Inc., 1988). The heaviness of the tails affects the behaviour of many statistics. Population kurtosis must lie between -2 and positive infinity. However, large values of kurtosis suggest that statistical methods based on normality assumption may be inappropriate. In this study equations with high kurtosis were re-assessed. In most cases, outliers contributed to the high kurtosis, and the corresponding basic data were revised or removed from the data base as described in section 3.3.3. The unbiased sample kurtosis coefficient is calculated as in equation 3.6.

$$K = [\frac{N(N+1)}{(N-1)(N-2)(N-3)}][\frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{(s^4 - 3(N-1)(N-1)/(N-2)(N-3))}] \quad (3.6)$$

- **Extreme Values of the Distribution**

This is a measure of maximum and minimum residuals of the variable being modelled. These values must not diverge unreasonably from the rest of the data. Their absolute values should not differ very much. If, for example the maximum residual value was 20 m^2/ha and the minimum was -2 m^2/ha, then it means that there is at least one excessive outlier on the positive side. The data and equation being modelled should be re-examined.

The above tests were conducted and interpreted jointly, not just on their own. In general an equation will provide a good fit if all the regression parameter estimates of the 95% confidence interval have the same sign, the residual patterns show no or little biased trend, the residuals bar chart show a normal distribution, the average mean of residuals is close to 0, and the absolute mean and extreme values of residuals do not deviate unreasonably from the rest of the data.

Chapter 4

Developing and Fitting the Models

This chapter describes the methodology and development used to derive the equations which form a (i) compatible tree volume and taper prediction system and (ii) whole stand growth and yield model, for Douglas fir grown in the South Island of New Zealand. The approach emphasized in both systems of equations is the need to devise them for as large an overall population as possible, while still allowing for local or regional adaptations through inclusion of dummy and continuous predictor variables such as altitude. Furthermore, in contrast to Garcia (1984b) who developed a method of fitting all three state variables, mean top height (h_{100}), basal area per hectare (G) and stocking (N) simultaneously, the approach adopted here is to fit single equations. The approach here allows the modeller to choose a different functional form for each variable h_{100}, G and N, if needed. This approach also allows greater flexibility to incorporate thinning effects in the model rather than having separate equations for thinned and unthinned stands.

4.1 Development of DfirTree

DfirTree is a compatible tree volume - taper estimation system for Douglas fir [Pseudotsuga menziesii (Mirbel, Franco)] trees from throughout the South Island of New Zealand. A precise tree volume equation which uses pooled data from the three main South Island regions, namely Canterbury, Nelson and Southland was developed. This equation still has the capacity, however, to differentiate attributes due to locality, made possible through incorporating dummy variables among predictors and other sources of variation.

A segmented taper equation compatible to the tree volume equation with two join points was later developed. This taper equation was assessed and compared to existing and other equations and found to conform well with the tree volume equation without the need for additional dummy variables.

4.1.1 Background Information

Douglas fir, [Pseudotsuga menziesii (Mirbel, Franco)], grows in a wide range of localities in the above named main regions of the South Island. There is some Douglas fir growing in Westland, another region, but this sub - population was not included in the analysis because of its relatively minor representation.

The age class distribution of Douglas fir in the South Island as at 01-04-1989 is summarized in Table 4.1.

Table 4.1: Distribution of Douglas fir by Area and Age Class in the S.Island of N.Z.

	Area (ha) planted in Douglas fir				
Age class	Canterbury	Nelson	Southland	Westland	total
0 — 5	1723	1584	2812	0	6119
6 — 10	1171	2535	3564	130	7404
11 — 15	921	2162	513	24	3620
16 — 20	577	1505	578	111	2771
21 — 25	841	1142	793	229	3005
26 — 30	602	825	689	31	2147
31 — 35	401	572	433	24	1430
36 — 40	206	225	261	11	703
41 — 50	254	90	89	6	439
51 — 60	79	430	318	1	828
61 — 80	156	0	155	7	318
Total	6931	11074	10205	574	28784

Source: N.E.F.D. 01-04-1989:

In New Zealand, four volume functions for D.fir have been constructed, namely

(i) T15 — (1958): for all of New Zealand.

(ii) T120 — (1977): for Ashley forest in Canterbury.

(iii) T136 — (1977): for all of New Zealand.

(iv) T228 — (1988): for Longwood forest in Southland.

These four volume functions are still in use, and are in the logarithmic form shown in equation 4.1 (see volume tables, Ministry of Forestry, New Zealand, 1992).

$$\ln(v) = \alpha + \beta \ln(d) + \gamma \ln[\frac{h^2}{h - 1.4}] \qquad (4.1)$$

where v is total stem volume inside bark, d is diameter at breast height over bark, h is total tree height, and α, β and γ are linear least - squares coefficients. Equation 4.1 can also be written in non - linear form as equation 4.2

$$v = \alpha d^{\beta}[\frac{h^2}{h - 1.4}]^{\gamma} \tag{4.2}$$

but has not been computed in this later form.

Today, there has been neither a published review nor sensitivity analysis of these equations. Since construction of the first national volume table for Douglas fir in 1958, the tendency has been either to devise equations for smaller populations or to ignore regional variations and apply a single overall equation. This pattern has also been exhibited in growth and yield modelling. There is an alternative, more sensitive approach, however, that could be adopted, namely to provide additional predictor variables which allow users to disaggregate general trends to reflect local variation with due sensitivity. This is the approach adopted here.

4.1.2 Data Base for Development of Tree Volume Equations

The data used in this study are the same ones used by F.R.I. and the New Zealand Forest Service to develop Tables T15, T120, T136 and T228, plus some others that were measured in Golden Downs forest as part of this study in December 1990. The F.R.I. data were collected between 1948 and 1988.

Before beginning the statistical analyses, the tree profile of each stem was displayed on the screen to ascertain the quality of individual measurements. Suspicious measurements were identified, marked and corrected, but those cases which could not be corrected based on objective evidence, were later discarded. Sectional measurements of stems on 641 stems altogether from Nelson, Canterbury and Southland were selected for analysis, but

44 of these were later discarded from the analysis of stem volume because of obvious measurement errors and deformations that could not be rectified. The data base by regions is shown in Table 4.2.

Table 4.2: Regional Data for Development of Tree Volume Equations

Region	Variable	Mean	minimum	maximum
Canterbury	No. of Trees	359		
	d (cm)	22.10	4.00	82.00
	h (m)	17.80	3.00	42.70
	$d^2h(m^3)$	1.52	0.0048	24.80
	v (m^3)	0.43	0.0030	6.48
Nelson	No. Trees	75		
	d (cm)	29.20	14.0	56.0
	h (m)	24.80	10.40	39.9
	d^2h (m^3)	2.56	2.16	11.69
	v (m^3)	0.73	0.087	3.05
Southland	No. Trees	153		
	d (cm)	24.70	3.00	65.00
	h (m)	18.40	3.00	37.80
	d^2h (m^3)	1.89	0.003	15.97
	v (m^3)	0.54	0.002	4.39
Total Number of Trees	597			

4.1.3 Calculation of Sectional Volumes

Two methods of measuring were applied in generating the individual tree volume data: in the first method, the stems were measured at 0.15 m, 0.75 m, 1.40 m, 3.0 m, followed by 1.50 m intervals to the tree total height. In the second method, stems were measured at 0.15 m, 0.70 m, 1.40 m, 3.0 m, followed by 3.0 m intervals to tree total height. At each measuring point the diameter outside bark by tape and two bark thicknesses by bark gauge were recorded. Inside bark diameters were determined by subtracting twice the average bark thickness from the corresponding outside diameters. The volume v_i of the first segment of each tree stem (0.15 m) was calculated as though it were a cylinder with sectional volume equation shown in equation 4.3

$$v_i = \frac{\pi}{40000} d_i^2 \times 0.15 \tag{4.3}$$

The volumes of each of remaining (N-1) segments of the tree were calculated by applying conoid equation 4.4 (Whyte, 1970).

$$v_i = \frac{\pi}{120000} (d_{il}^2 + d_{is}^2 + d_{il} d_{is}) \times l_i \tag{4.4}$$

Where;

d_{il} large end diameter of section (cm)

d_{is}=small end diameter of section (cm)

Individual under bark total stem volumes, v were calculated through summing the individual sectional volumes as in equation 4.5.

$$v = \sum_{i=1}^{N} v_i \tag{4.5}$$

4.1.4 Analysis of Stem Volume Equations

Various linear and non - linear tree volume equations with the general form *v=f(dbh, height, form factor)*, were fitted to the data. They included:

(i) Schumacher tree volume equation (Schumacher and Hall, 1933)

$$v = \alpha d^{\beta} h^{\gamma} \tag{4.6}$$

(ii) combined variable equation (Spurr, 1952)

$$v = \alpha + \beta d^2 h \tag{4.7}$$

(iii) Meyer's polynomial equation (Meyer, 1953)

$$v = \beta_0 + \beta_1 d + \beta_2 d^2 + \beta_3 dh + \beta_4 d^2 h \tag{4.8}$$

(iv) Honer's transformed variable equation (Honer,1965)

$$v = \frac{d^2}{\alpha + \frac{\beta}{h}} \tag{4.9}$$

(v) constant form factor equation (Gevorkiantz and Olssen, 1955)

$$v = \beta d^2 h \tag{4.10}$$

(vi) non - linear form of N.Z. Forest Service equation (Ministry of Forestry New Zealand, 1992)

$$v = \alpha d^{\beta}(\frac{h^2}{h-1.4})^{\gamma} \tag{4.11}$$

Equations 4.6, 4.9 and 4.11 were fitted using non linear least - squares method, while equations 4.10, 4.8 and 4.7 were fitted using ordinary linear least - squares method. For each equation the resulting residuals (observed - predicted volumes) were plotted against their predicted values. The graphs revealed that the observations had heterogeneous variance, suggesting that it was appropriate to use weighted least - squares regression. To determine an appropriate weight, the independent variable d^2h was subdivided into ordered classes of equal numbers of observations, then the standard deviation of each class was calculated and plotted against the independent variable, d^2h. The standard deviations of d^2h classes were found to be proportional to the corresponding mean values of d^2h classes. This suggested that the reciprocal of d^2h would be an appropriate weight. Standard procedures for coping with heterogeneous variance were followed, as set out, for example, in Clutter et al., (1983); Furnival, (1961); Draper and Smith, (1981). Regression analyses were subsequently performed using the weight shown in expression 4.12.

$$w = \frac{1}{d^2h} \tag{4.12}$$

4.1.5 Criteria for Selection of Stem Volume Equations

The criteria used in selecting the best equation were as listed below.

(i) Residual sum of squares (ESS)

Because all equations (4.6 — 4.11) have the same dependent variable v, it was possible to compare the values of residual sum of squares of each equation directly.

(ii) The Root Mean Square (RMS)

The root mean square (standard deviations) of the equations and the 95 % confidence intervals for all coefficients estimates were determined. The usual index (I) for selection of tree volume equations with different independent variables can be defined as the RMS (σ), divided by the first derivative of the independent variable of the volume equation, $F'(v)$ (Furnival, 1961). This expression is portrayed in equation 4.13.

$$I = \frac{\sigma}{F'(v)} \tag{4.13}$$

Because the dependent variable in all the volume equations analyzed in the study was v, the index was simply calculated as in equation 4.14.

$$I = \sigma \tag{4.14}$$

The smaller the index (I), the more appropriate the equation is.

(iii) Residual Plots and Charts

Plots of residuals against predicted values and frequency charts of residuals were used to ascertain that the residuals for any given equation were normally and independently distributed with mean 0 and standard deviation σ. This analysis went a stage further by allocating the residuals to their associated individual regions, so that the residual patterns for each region could be individually examined.

(iv) Bias

The mean overall bias (B) was calculated according to equation 4.15. Biases were further analysed by individual diameter classes as explained in, for example, Honer (1965), and Hayward (1987).

$$B = \frac{\sum_1^N (vi - \hat{vi})}{N} \tag{4.15}$$

Where v_i is the observed stem volume, $\hat{v}_i$ is the predicted stem volume and N is the number of observations in the class.

This analysis of bias confirmed that it is better to correct for non - homogeneity of variance in linear or non linear equations through the method of weighted least - squares regression (Draper and Smith, 1981; Schumacher and Chapman, 1954; Gedney and Johnson, 1959; Furnival, 1961; Buckman, 1961; *et cetera*), than by logarithmic transformation.

4.1.5.1 Test and Choice of Stem Volume Equation

Residual plots of equations 4.7, 4.10 and 4.11 showed obvious bias. Their fit can be viewed in appendix C in files COMBV.LIS CONSTV.LIS and NZV.LIS respectively. Equation 4.8 fitted the data well, but the parameters β_0, β_1, and β_2 were not significant at the 5% significant level (see the file MEYV.LIS in appendix C).

These four equations were removed from subsequent analyses. Equation 4.9 fitted the data quite well but the standard errors of the parameter estimates were unusually high, and so this equation too was not analysed further (see file HONV.LIS in appendix C). Equation 4.6 fitted the data by far best and was selected (See files SCHUMV.LIS) in appendix C). Table 4.3 summarizes the statistical results for equations 4.6 to 4.11 ran through the same data ($N = 597$). Further analysis was conducted to find whether or not

Table 4.3: Summary Statistics for Stem Volume Equations 4.6 - 4.11

Equation	Parameter	Estimate	SEE	ESS	MSE
4.6	α	0.000090561	0.00000260622	0.000009055656	0.00000015427
	β	2.114685122	0.01075675351		
	γ	0.514830939	0.00344657163		
4.7	α	0.004071	0.00049394	0.00000715604	0.00000012027
	β	0.000028354	0.00000011		
4.8	β_0	0.002104	0.00151585	0.0000458306	0.00000007741649
	β_1	-0.000762	0.00039955		
	β_2	0.000069874	0.000001837		
	β_3	0.011782	0.00154125		
	β_4	0.000023200	0.00000056		
4.9	α	266.19201	14.83665045	0.00005177815	0.00000008702
	β	27857.08622	393.83302031		
4.10	α	0.0000285861	0.0000001312	0.00007973043	0.00000013378
4.11	α	0.000112794	0.000000	0.00378210770	0.00000636718
	β	1.323657982	0.000000		
	γ	0.966724850	0.000000		

regional variation existed which could be incorporated in equation 4.6. The analysis was repeated, therefore, with some modification to allow for two dummy variables for Nelson and Southland, while Canterbury remained the default. The maximal model 4.16 was fitted and parameter estimates were examined as discussed in sections 2.8.5 and 3.3.3.

$$v = (\alpha_1 + \alpha_2 Z_1 + \alpha_3 Z_2) d^{(\beta_1 + Z_1\beta_2 + Z_2\beta_3)} h^{(\gamma_1 + \gamma_2 Z_1 + \gamma_3 Z_2)} \quad (4.16)$$

Where;

α_1, α_2, and α_3 replace α in equation 4.6

β_1, β_2 and β_3 replace β

γ_1, γ_2 and γ_3 replace γ

d, v and h, are as in equation 4.6.

Z_1 and Z_2 are dummy variables set to 1 for Nelson and Southland respectively, otherwise 0.

Equation 4.16 was written in SAS code, such that Z_1 assumes a value of 1 if the region is Nelson, else it assumes a value of 0: similarly, Z_2 assumes a value of 1 if region is Southland, else it assumes a value of 0. Results of the regression analysis showed that, the values of the parameters α_2, α_3, γ_2 and γ_3 were not significantly different from 0. The equation was therefore modified further and re - run without including these variables. The final equation is as shown in equation 4.17.

$$v = \alpha d^{(\beta_1+\beta_2 Z_1+\beta_3 Z_2)} h^{\gamma} \qquad (4.17)$$

The statistics of equation 4.17 are set out in Table 4.4.

Table 4.4: Parameter Estimates and Standard Errors for Modified Schumacher's Stem Volume Equation

Parameter	Estimate	Standard Error	Weighted sum of squares (ESS)	Number of trees	MSE
α	0.000052457	0.000000008550			
β_1	1.910540041	0.00007111878			
β_2	0.000238690	0.00045112873	0.00004590150	597	0.00000007754
β_3	0.000743909	0.00000129919			
γ	0.912310849	0.00000395799			

4.1.6 Goodness of Fit of Modified Schumacher's Volume Equation

Equation 4.17 decreased the error sums of squares (ESS) by about 51%, compared to equation 4.6. The parameter estimates (α, β_1, β_2, β_3 and γ) were all significant at least at the 5% level. There was also an improvement in the regional pattern of residuals and on the overall bias. Inclusion of the dummy variables reduced the mean overall bias (B) from -3.9% to 0.25%. The accuracies of equations 4.6 and equation 4.17 were also analysed by individual diameter classes. Percentage mean biases for each diameter class were calculated as in equation 4.18.

$$E = 100/N \times \sum_{1}^{N}(\frac{(v_i - \hat{v}_i)}{v_i}) \tag{4.18}$$

Where;

v_i = actual volume and $\hat{v}$ is the predicted volume.

E = percentage diameter class mean bias

N = number of observations in the diameter class.

The results of these tests are summarized in Table 4.5.

Table 4.5: Percentage Bias by dbh Classes for Equations 4.6 and 4.17

Dbh class cm	N	Mean Class Volume (M^3)	% Mean Bias Equation 4.6	% Mean Bias Equation 4.17
5.0	33	0.000627	-21.0	16.4
10.0	51	0.04101	-12.2	2.4
15.0	92	0.1159	-2.1	0.17
20.0	150	0.2249	-4.1	-2.4
25.0	109	0.3877	-0.59	0.03
30.0	59	0.5812	-1.25	1.2
35.0	32	0.8000	-5.6	-3.5
40.0	28	1.2589	2.0	0.24
45.0	15	1.5744	0.52	0.80
50.0	7	2.1002	2.7	1.3
55.0	7	2.5079	1.9	1.1
60.0	7	3.2462	-0.09	-2.8
65.0	2	4.6488	9.0	2.9
>65.0	5	5.3600	-0.65	0.00
Overall Bias	597	0.5005	-3.9	0.25

The final tests were standardized through PROC UNIVARIATE, looking at the skewness coefficient, normality (Kolomogorov D statistic) and kurtosis, as well as the mean of residuals. These values are presented in diskette form in the file SHUMVD.LIS in appendix C. These values all showed an acceptable level of conformance with a normal spread of residuals. The fit to the data was enhanced considerably through use of dummy variables in equation 4.17 to characterize locality variations. Figure 4.1 shows the residual pattern when plotted against predicted values. Figure 4.2 shows the residual patterns when plotted against dbhob and Figure 4.3 shows the frequency distribution of residuals. These figures all show no serious bias. Possible biases in the equations were also evaluated by size classes, as explained, for example, in Honer (1965). This method requires that the volume errors should be able to be predicted adequately so that they are independent of tree size and lie within an average of not more than $\pm$ 10% about the size class mean in at least 95% of the classes having more than 5 observations. The percentage volume errors in equation 4.17 showed a bias of 16% in the 5.0 cm diameter class, but this is not considered to be a serious bias as the mean volume of this class is so small, 0.00627 m^3. All other classes showed bias of less than $\pm$ 10.0%.

Figure 4.1: Plot of Residuals vs Predicted Values [m^3]
Equation 4.17

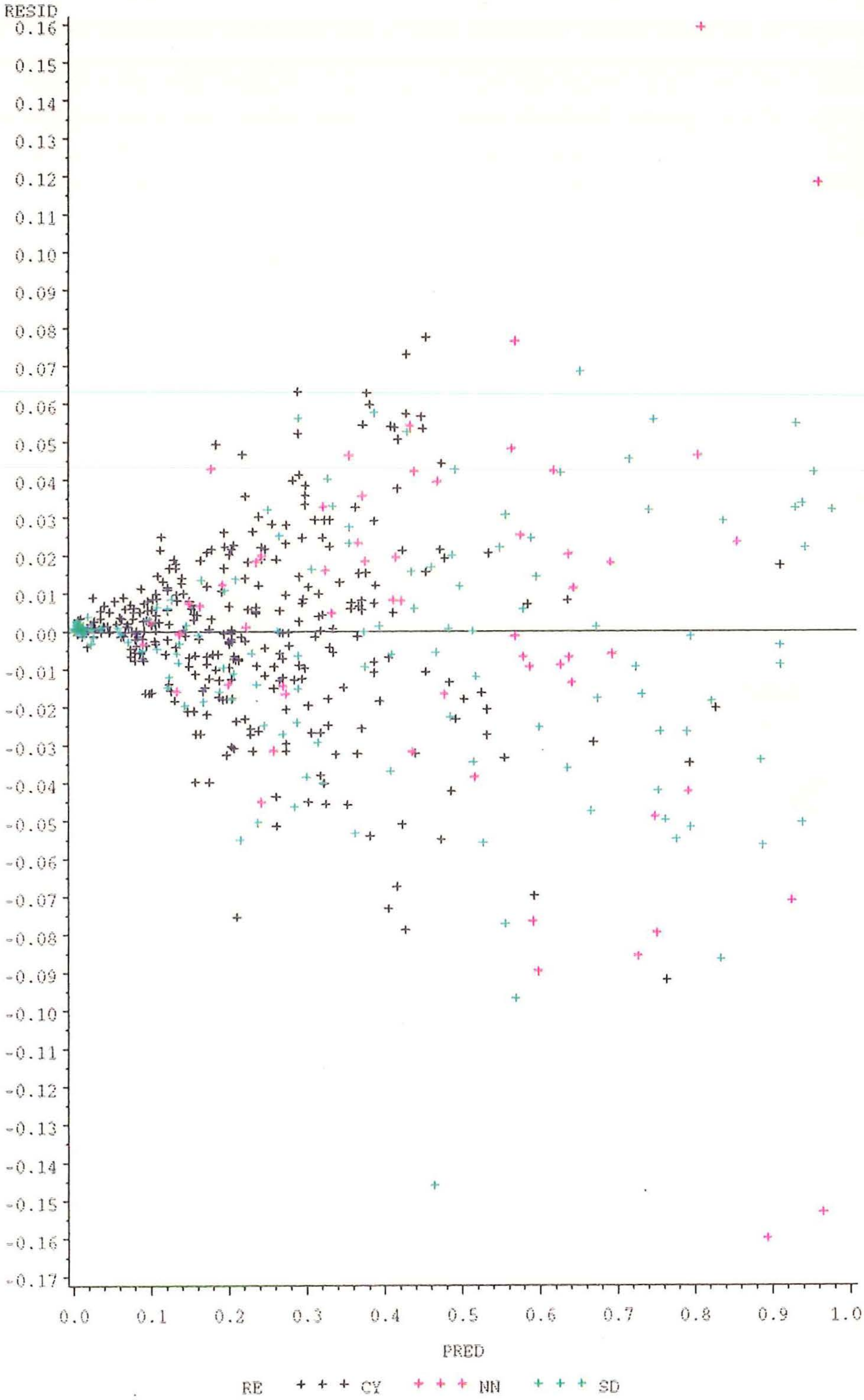


Figure 4.2: Plot of Residuals [m^3] vs DBHOB [cm]
Equation 4.17

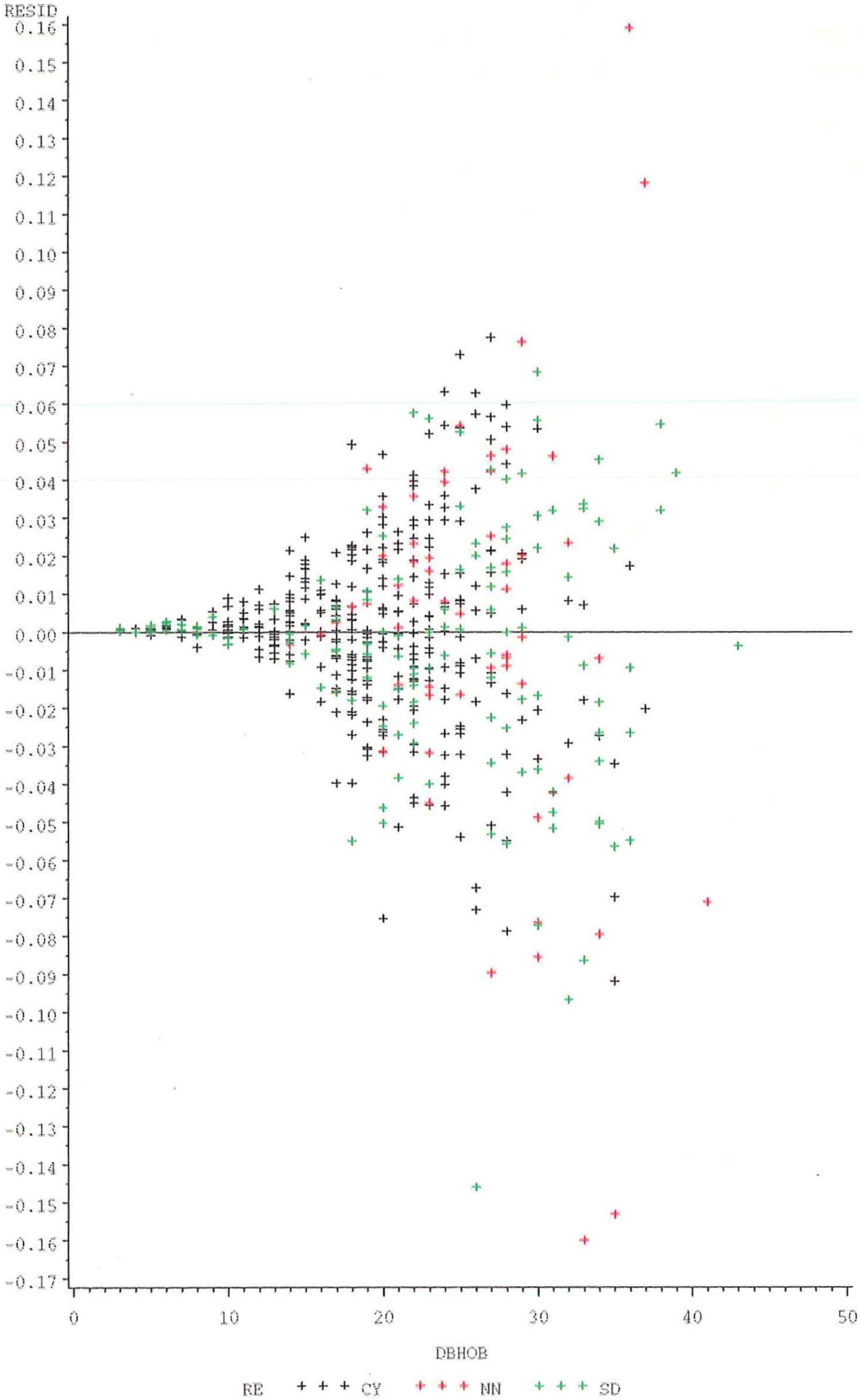
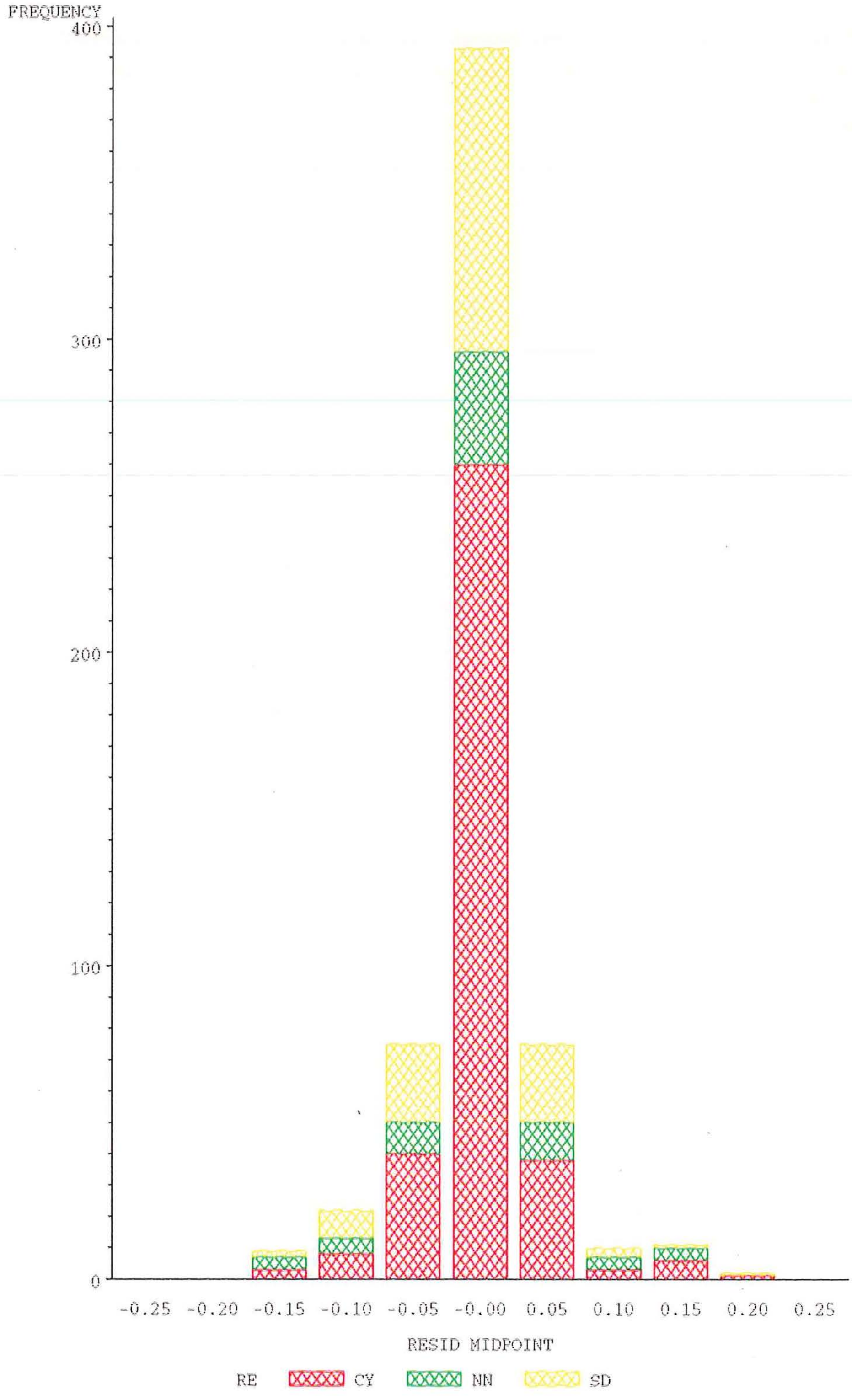


Figure 4.3: Frequency Distribution of Residuals [m^3] classes
Equation 4.17



4.2 Compatible Taper Equations

This section explains the use of enhancements of the previous section in developing a compatible stem taper equation. Compatible taper equations are those when integrated with respect to total tree height will yield the same volume as though it were the tree volume equation.

Two methods for constructing compatible taper equations compatible with the volume equation were employed. These methods are described in the following sub - sections.

4.2.1 Volume Based System

In a volume based system, the volume of the stem inside bark is taken to be that estimated from the tree volume equation. The taper equation can then be derived through its subsequent differentiation, the resulting equations being called volume - based.

4.2.2 Taper Based System

In a taper based system, an equation to describe taper of the stem is first determined and its volume then derived by subsequent integration of the stem profile predicted from the taper equation. The resulting equations are called taper - based. The definitions in 4.2.1 and 4.2.2 above are fully described and discussed in Demaerschalk (1972); Byrne and Reed (1986) and many others.

In this study, the volumes of stems for the volume based system were determined by the tree volume equation 4.17. For the taper based system, an expression of top diameter d'

(taper function) was integrated to get an expression for the tree volume equation.

4.2.2.1 Development of Volume Based and Taper Based Systems

The original taper equations for Douglas fir that are compatible to their volume functions are still maintained and are kept in the library by the Ministry of Forestry. These are:

F136 — for volume function T136, for all of New Zealand,

F228 — for volume function T228, for Longwood forest in Southland.

None — for volume function T15 nor T120.

These equations are full polynomials and have the form shown in 4.19 (see taper tables, FRI-Ministry of Forestry, New Zealand, 1992).

$$\frac{1}{f} = \frac{d'^2 K h}{v} = \beta_1 z + \beta_2 z^2 + ... + \beta_n z^n \tag{4.19}$$

Where;

d'=top diameter inside bark (cm) at h'.

f=form factor;

K=constant to convert cm^2 to m^2 (0.00007854).

h'=distance from the base of the tree to top diameter d';

$$\ell = (h - h').$$

h=total tree height (m).

v=predicted total stem volume by using equation 4.17;

z=relative tree height from the tip: $(h - h')/h$.

β_1, β_2,...β_n are least - squares coefficients to be estimated from the data.

FRI researchers have fitted the full polynomial up to n=9, but this form of polynomial ends up with some terms not significant which are removed from the equation (Gordon, 1983). Katz and Dunningham (1981) modified the taper equation for radiata pine grown under direct saw log regimes so that they have the value of n up to 31 in attempting to explain for butt swell. In this study the following three taper equations were selected, then fitted and compared.

(a) Full polynomial as shown in equation 4.19.

(b) Cao et al., (1980) segmented polynomial taper equation 4.20

$$\frac{d'^2Kh}{v} - 2z = \beta_1(3z^2 - 2z) + \beta_2(z - \alpha_1)^2 I_1 + \beta_3(z - \alpha_2)^2 I_2 \tag{4.20}$$

This equation was modified to equation 4.21 so that the independent variable is consistent with equations 4.19 and 4.23 and can then provide a better comparison of their fits.

$$\frac{d'^2Kh}{v} = \beta_1(3z^2 - 2z) + 2z + \beta_2(z - \alpha_1)^2 I_1 + \beta_3(z - \alpha_2)^2 I_2 \tag{4.21}$$

Equation 4.21 can also be re-parameterized as 4.22

$$\frac{d'^2Kh}{v} = 3\beta_1 z^2 + (2 - \beta_1)z + \beta_2(z - \alpha_1)^2 I_1 + \beta_3(z - \alpha_2)^2 I_2 \tag{4.22}$$

Where;

α_1, and α_2 are the join points.

I_i=1 for z $\geq \alpha_i$

I_i=0 for z < αi

i=1,2.

(c) Proposed taper equation

The proposed segmented polynomial taper equation 4.23 was formulated and tested. This equation fulfills the properties for segmented taper equations (Max and Burkhart, 1976; Cao et al., 1980; McClure and Czaplewski, 1986; Valenti and Cao, 1986; Byrne and Reed, 1986). The methods and conditions used to construct segmented polynomials with defined join points as described for example in Fuller (1969); Hudson (1966); Gallant and Fuller (1973), were adopted in developing the chosen form presented in equation 4.23.

$$\frac{d'^2Kh}{v} = \beta_1 z^2 + \beta_2 z + \beta_3 (z - \alpha_1)^2 I_1 + \beta_4 (z - \alpha_2)^2 I_2 \tag{4.23}$$

Variables and their definitions are as in equation 4.20. The equation has two join points as does equation 4.20 but has 6 parameters to be estimated while equation 4.20 has 5. If the first two terms of equation 4.22 are equated to the first two terms of 4.23 as shown below (* indicates re-parameterization).

$$\beta_1 3z^2 + (2 - 2\beta_1)z = \beta_1^* z^+ \beta_2^* z$$

By virtue of the above re-parameterization the following equality holds

β_1^*=$3\beta_1$ and β_2^*=$(2 - 2\beta_1)$

and so equation 4.20 is mathematically identical to 4.23, but will not give same results as far as modelling is concerned.

Equations 4.19 and 4.20 are the most frequently adopted and are studied widely in the literature. Because of their definition, compatible taper equations require at least an algebraic restriction, so that consistent volume estimates can be obtained by integrating the taper equation as though it were a volume function. Compatibility can be enforced algebraically in a number of ways: by recovering parameter values of the tree volume model using estimated parameters from the fit of the taper model (Byrne and Reed, 1986), or by estimating any free parameters of the taper model given estimated parameters of a particular form of tree volume model (Demaerschalk, 1973; Van Deusen et al., 1982). These two methods are earlier forms. The third method adopts the method of constraining the parameters of, and incorporating estimated tree volume in the taper model. This approach is now widely used (e.g. Goulding and Murray, 1976; McClure and Czaplewski, 1986). The integration and restriction for the proposed equation are demonstrated below.

$$\frac{d'^2 Kh}{v} = \beta_1 z^2 + \beta_2 z + \beta_3 (z-\alpha_1)^2 I_1 + \beta_4 (z-\alpha_2)^2 I_2$$

Merchantable volume, v_m can be obtained from

$$v_m = K \int_0^h d'^2 d\ell \tag{4.24}$$

where,

$$d'^2 = \frac{v}{Kh}[\beta_1 z^2 + \beta_2 z + \beta_3 (z-\alpha_1)^2 I_1 + \beta_4 (z-\alpha_2)^2 I_2]$$

and

$$d\ell = h.dz$$

The integration can then be written as

$$v_m = K \int_0^z \frac{v}{Kh}(\beta_1 z^2 + \beta_2 z + \beta_3 (z-\alpha_1)^2 I_1 + \beta_4 (z-\alpha_2)^2 I_2) h.dz \tag{4.25}$$

or after re - arrangement, as 4.26

$$v_m = v[\frac{\beta_1}{3} + \frac{\beta_2}{2} + \frac{\beta_3}{3}(z-\alpha_1)^3 I_1 + \frac{\beta_4}{3}(z-\alpha_2)^3 I_2] \tag{4.26}$$

Merchantable volume is the same as total volume when the integration is done with respect to z between the limits 0 and 1, and by this condition, I_1=I_2=1. Hence the compatibility restriction is

$$[\frac{\beta_1}{3} + \frac{\beta_2}{2} + \frac{\beta_3}{3}(1-\alpha_1)^3 + \frac{\beta_4}{3}(1-\alpha_2)^3] = 1 \tag{4.27}$$

In equation 4.27 restriction

$$\frac{\beta_1}{3} + \frac{\beta_2}{2} = 1$$

is independent of restriction

$$\frac{\beta_3}{3}(1 - \alpha_1)^3 = \frac{\beta_4}{3}(1 - \alpha_2)^3$$

Restriction 4.27 consists of 4 terms while that of McClure and Czaplewski (1986) has 3 terms. Restriction 4.27 has provided greater flexibility in imposing a second restriction on the parameters of the main restriction as will be shown in subsection 4.2.3.

4.2.3 Data Used for Developing Stem Taper Equations

Table 4.6: Statistics for Regional Sectional Measurements

Region	Variable	Mean	minimum	maximum
Canterbury	No. of trees	394		
	No. of measurements	4351		
	Dbh (cm)	28.3	3.8	88.1
	h (m)	20.7	3.0	46.0
	v (m^3)	0.732	0.0030	7.71
Nelson	No. of trees	82		
	No. of sectional measurements	984		
	Dbh (cm)	30.6	13.7	56.4
	h (m)	25.6	10.4	39.9
	v (m^3)	0.859	0.087	3.526
Southland	No. of trees	164		
	No. of sectional measurements	1661		
	Dbh (cm)	28.0	2.7	93.9
	h (m)	19.3	3.0	38.1
	v (m^3)	0.845	0.0015	7.513
Total no. of trees		641		
Total no. measurements		6996		

Table 4.6 shows the number of sectional measurements and their basic statistics by region. The data shown in Table 4.6 are the same ones used to develop the stem volume equations described in sub - section 4.2.2.1. Sectional measurements were generated from a simple FORTRAN programme (FFSECT - N.Z. Forest Service) which rearranged the data so that the first measurement from the ground level forms the large end diameter for the first section, the second measurement forms the small end diameter of the section and is also the large end diameter of the next section of the tree, and so on. The procedure is repeated up to the tip of the tree where inside bark diameter is conditioned to 0.0 cm .

4.2.4 Fitting of the Selected Equations

4.2.4.1 Volume Based System

This section deals with the statistical analyses of the selected taper equations considered in subsection 4.2.2.1. Equations 4.19, 4.20 and 4.23 were fitted to the same data. The parameters of these equations were restricted as explained in section 4.2.2.1, in order that they can be made compatible. The resulting residuals of the equations were plotted against the predicted values and independent variable z.

(1) **Full Polynomial Taper Equation, 4.19**

Equation 4.19 was fitted to polynomial of order 6, no other terms being significant at the 5% significance level, as shown below.

$$\frac{d'^2Kh}{v} = \beta_6 z^6 + \beta_5 z^5 + \beta_4 z^4 + \beta_3 z^3 + \beta_2 z^2 + \beta_1 z \qquad (4.28)$$

Equation 4.28 was restricted as per equation 4.29, so that, compatible volume estimates can be accrued.

$$\frac{\beta_6 z^7}{7} + \frac{\beta_5 z^6}{6} + \frac{\beta_4 z^5}{5} + \frac{\beta_3 z^4}{4} + \frac{\beta_2 z^3}{3} + \frac{\beta_1 z^2}{2} = 1 \tag{4.29}$$

The statistics of equation 4.28 are presented in Table 4.7. Figures 4.4 and 4.5 show the residuals plotted against predicted values and z. Figure 4.6 shows the frequency distribution of the residuals.

Table 4.7: Parameter Estimates and Standard Errors for Full Polynomial Equation

Parameter	Estimate	SEE	ESS	N	MSE
β_1	2.374066	0.17358559			
β_2	-17.673319	1.35057106			
β_3	71.887964	3.63627716	234.37864	6996	0.033353
β_4	-99.520206	4.05327869			
β_5	45.816741	1.59731925			
β_6	0.085348	0.01768282			

(2) Segmented Taper Equation, 4.21

Table 4.8 presents the parameter estimates and standard errors for this equation. Although

Table 4.8: Parameter Estimates and Standard Errors for equation 4.21

Parameter	Estimate	SEE	ESS	N	MSE
α_1	0.6593284	0.042489478			
α_2	0.9406695	0.001927715			
β_1	0.5338036	0.006759376	211.121241	6996	0.030199
β_2	-1.2130741	0.636468331			
β_3	229.6476472	14.641612424			

the equation fitted the data well, the coefficient β_2 was not significant at the 5% level (see file CAO.LIS in appendix C), and so the equation was not considered for further analysis.

(3) **Proposed Taper Equation**

The parameter estimates and standard errors for the proposed equation 4.23 are summarized in Table 4.9. Figures 4.7 and 4.8 show the plot of residuals plotted against predicted values and z respectively. Figure 4.9 shows the frequency distribution of residuals. The

Table 4.9: Parameter Estimates and Standard Errors for the Proposed Taper equation

Parameter	Estimate	SEE	ESS	N	MSE
α_1	0.6421100	0.036556224			
α_2	0.9430644	0.001799592			
β_1	1.6094235	0.020019514	211.062341	6996	0.030195
β_2	0.9270510	0.013346343			
β_3	-0.9963733	0.361105088			
β_4	247.46764463	15.598570311			

goodness of fit of the equations was assessed on the summary statistics of the equations and their residual patterns (Tables 4.7 to 4.9 and Figures 4.4 to 4.9). Residual plots for equation 4.28 showed irregular trends at most parts of the stem: plots of residuals for equation 4.23 are clearly superior to equation 4.28. Equation 4.23 was identified as superior because it had 10% less ESS than equation 4.28. However, it showed some irregular trends in the middle of the stem. To improve its fit, the technique proposed by Candy (1989), to include the variable d, the (dbhob) at the two join points was imposed, to determine whether or not the fit could be improved. Equation 4.23 was thus modified to equation 4.30.

$$\frac{d'^2Kh}{v} = \beta_1 z^2 + \beta_2 z + d\beta_3(z-\alpha_1)^2 I_1 + d\beta_4(z-\alpha_2)^2 I_2 \tag{4.30}$$

The restriction for equation 4.30 was reinstated through restrictions 4.31 and 4.32 on the parameters without endangering consistency in the error sum of squares.

$$(\frac{\beta_1}{3} + \frac{\beta_2}{2}) = 1 \tag{4.31}$$

and

$$\frac{1}{3}[d\beta_3(1-\alpha_1)^3 + d\beta_4(1-\alpha_2)^3] = 0 \tag{4.32}$$

The main restriction 4.27 remains unchanged, but could be modified to 4.33

$$[\frac{\beta_1}{3} + \frac{\beta_2}{2} + d\frac{\beta_3}{3}(1-\alpha_1)^3 + d\frac{\beta_4}{3}(1-\alpha_2)^3] = 1 \tag{4.33}$$

It was realized in this study that to ensure compatibility is achieved, all three restrictions, 4.31, 4.32 and 4.33 above must be obeyed. Although it may appear that equation 4.33 embeds both equations 4.31 and 4.32, it alone is not sufficient to enforce compatibility, although can be used as a final verification.

- **Assessment of Goodness of Fit of Taper Equations**

The residual pattern for equation 4.30 for most part of the stem was better than without d. The mean residual value for equation 4.30 was lower than in equation 4.23 which does not include d (see files TPPD.LIS and TPP.LIS respectively in appendix C). The residual plots and frequency distribution chart for the equations are shown in Figures 4.10 to 4.12. Parameter estimates and standard errors for equation 4.30 are shown on Table 4.10. Assessment of bias in predicting top diameter d' by z classes for equations 4.23 and 4.30 was conducted so as to quantify the effect of including d in 4.30. Table 4.11 summarizes this information.

Table 4.11 shows that equation 4.23 and 4.30 have more or less the same precision in predicting upper stem diameter inside bark. Equation 4.30 predicts diameter better for z less or equal to 0.3, and when z is greater or equal to 0.8. This implies that it accounts better for the butt swell which is common in mature Douglas fir trees, however, when z is between 0.3 and 0.8 equation 4.23 performs better. Equation 4.30 is therefore recommended, the

Table 4.10: Parameter Estimates and Standard Errors of the adopted Taper Equation

Parameter	Estimate	SEE	ESS	N	MSE
α_1	0.686291182	0.03401590725			
α_2	0.937009056	0.00241022981			
β_1	1.677377987	0.01737692381	237.692642	6996	0.034005
β_2	0.881748009	0.01158461587			
β_3	-0.040497946	0.01530960428			
β_4	5.002404744	0.32911529480			

Table 4.11: Comparison of Bias between equations 4.23 and 4.30

z class	N	Mean volume m^3	Bias in cm	
			Equation 4.23	Equation 4.30
0.1	983	0.9673	-0.72	-0.67
0.2	449	1.1360	-1.42	-1.29
0.3	498	1.0195	-0.38	-0.27
0.4	468	1.1100	0.30	0.38
0.5	514	0.9774	0.55	0.60
0.6	452	1.0356	0.49	0.51
0.7	539	0.9844	0.31	0.28
0.8	533	1.0358	-0.01	-0.08
0.9	1294	0.6264	0.04	-0.007
1.0	1266	1.2072	0.13	0.004
overall bias	6996	0.9834	0.07	-0.09

trade off in loss of precision in the middle section is minimal and is compensated by precision in predicting diameters near the butt, the section more commercially utilized.

Figure 4.4: Plot of Predicted Values vs Residuals
Full Polynomial Taper Equation

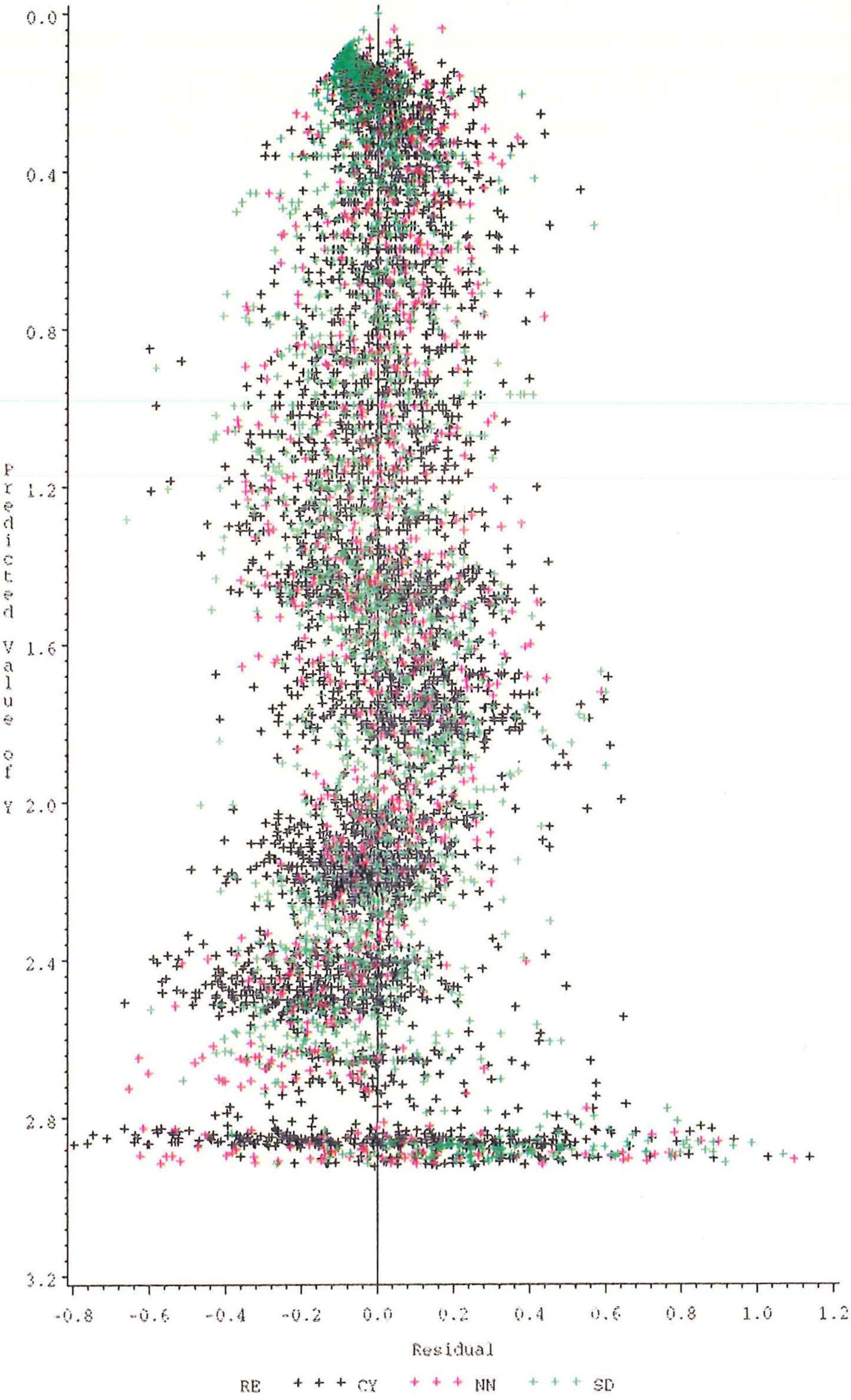


Figure 4.5: Plot of z vs Residuals
Full Polynomial Taper Equation

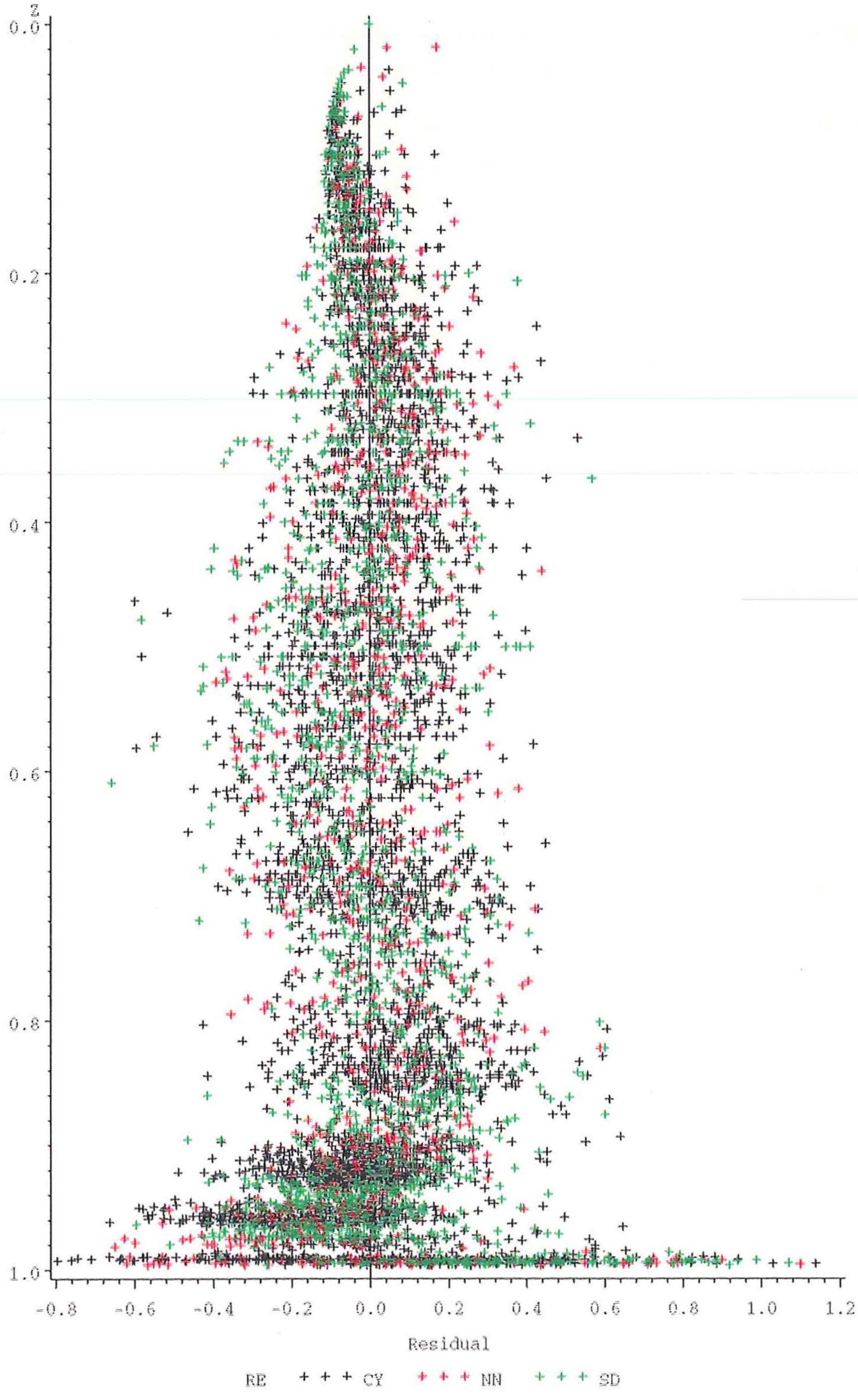


Figure 4.6: Frequency Distribution of Residuals
Full Polynomial Taper Equation

Figure 4.7: Plot of Predicted values vs Residuals
Proposed Taper Equation

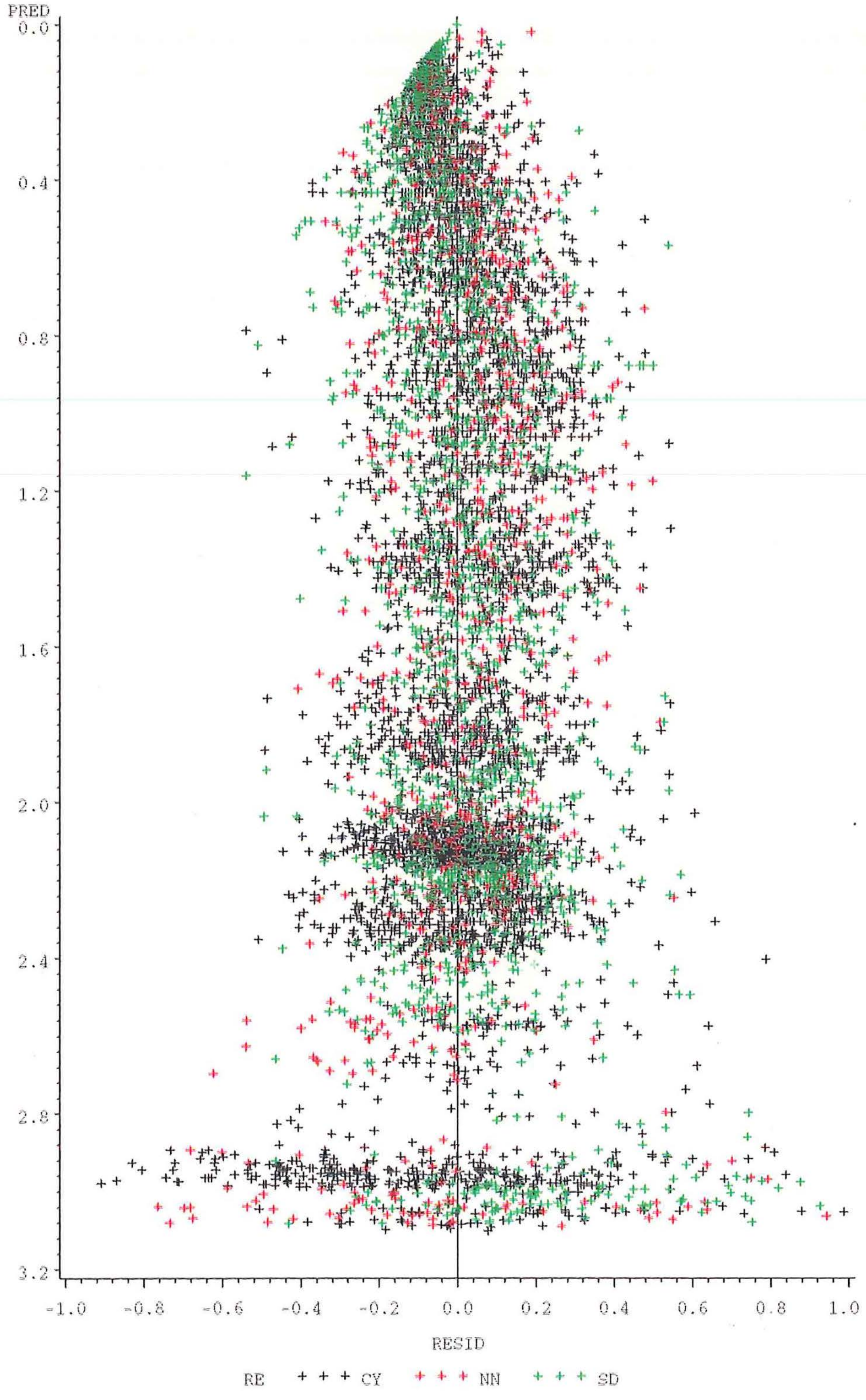


Figure 4.8: Plot of z vs Residuals
Proposed Taper Equation

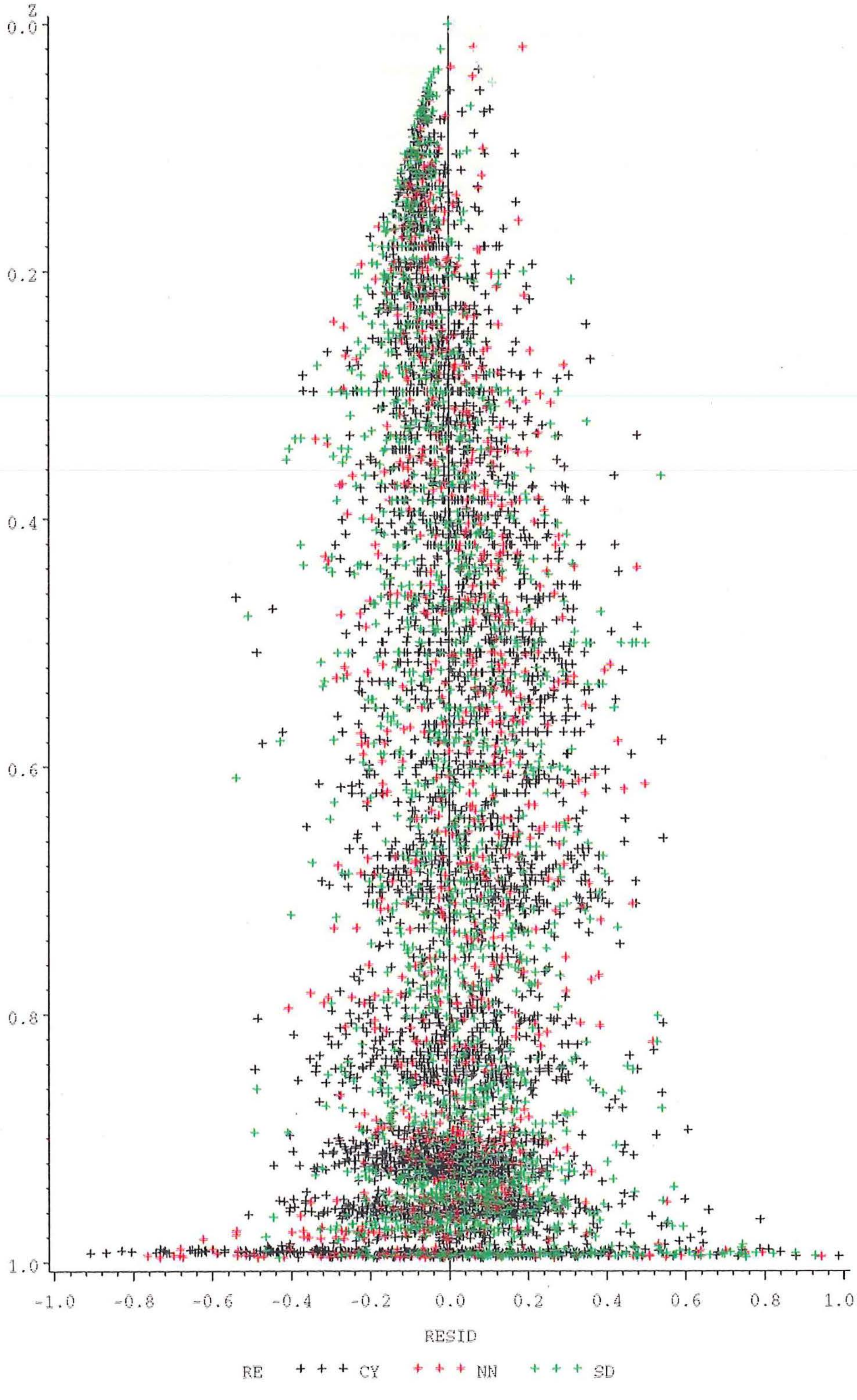


Figure 4.9: Frequency Distribution of Residuals
Proposed Taper Equation

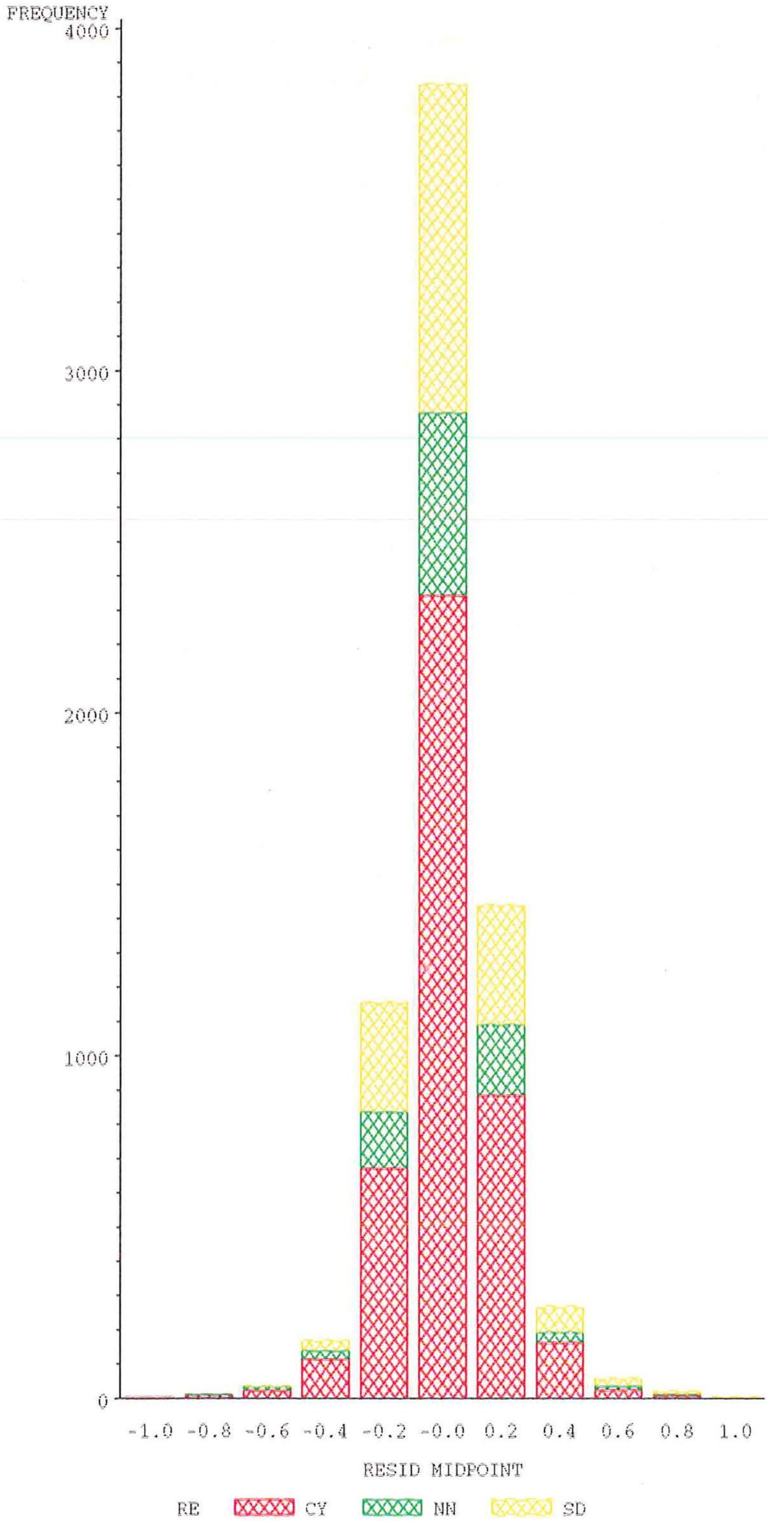


Figure 4.10: Plot of Predicted Values vs Residuals
Adopted Taper Equation

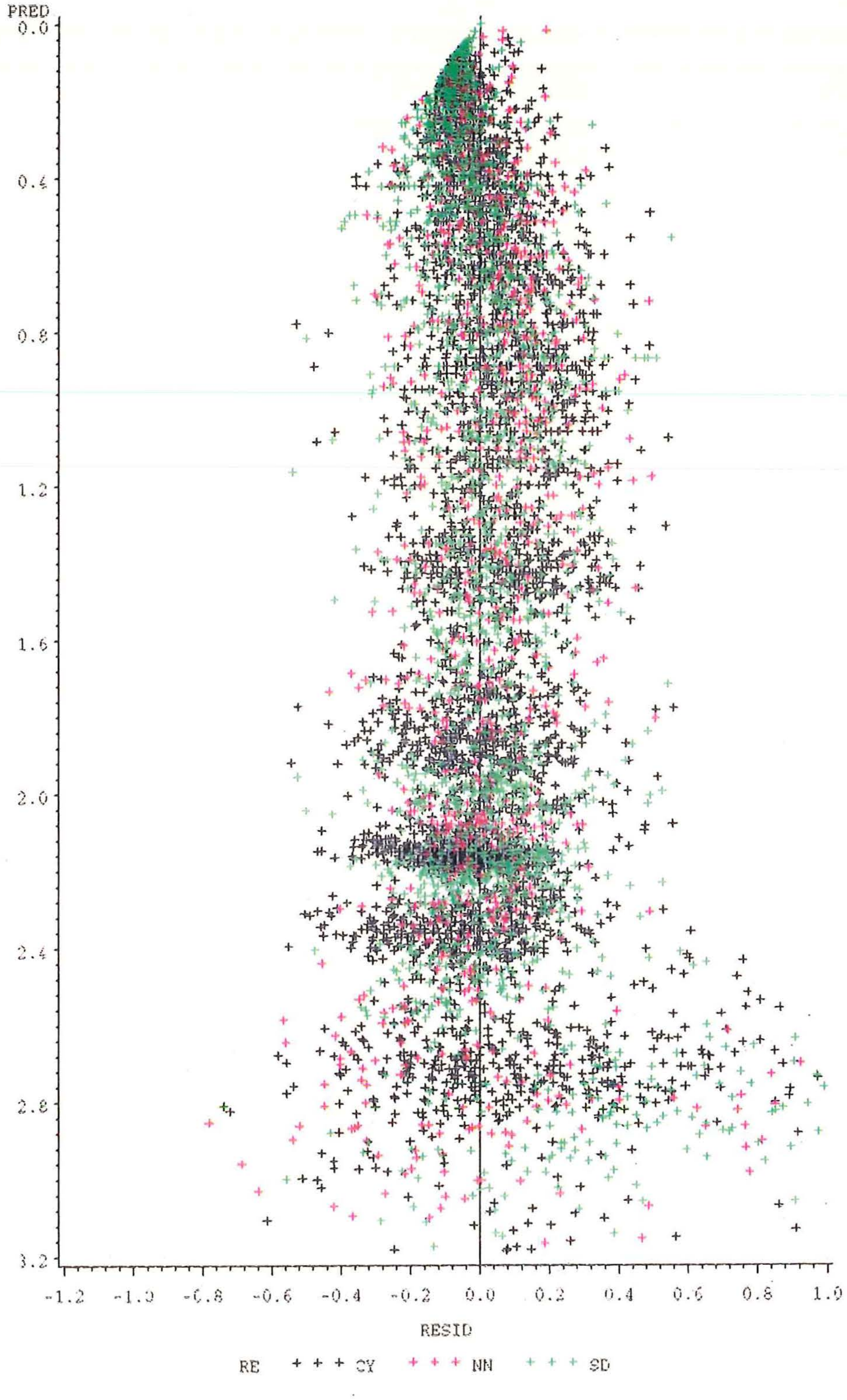


Figure 4.11: Plot of z vs Residuals
Adopted Taper Equation

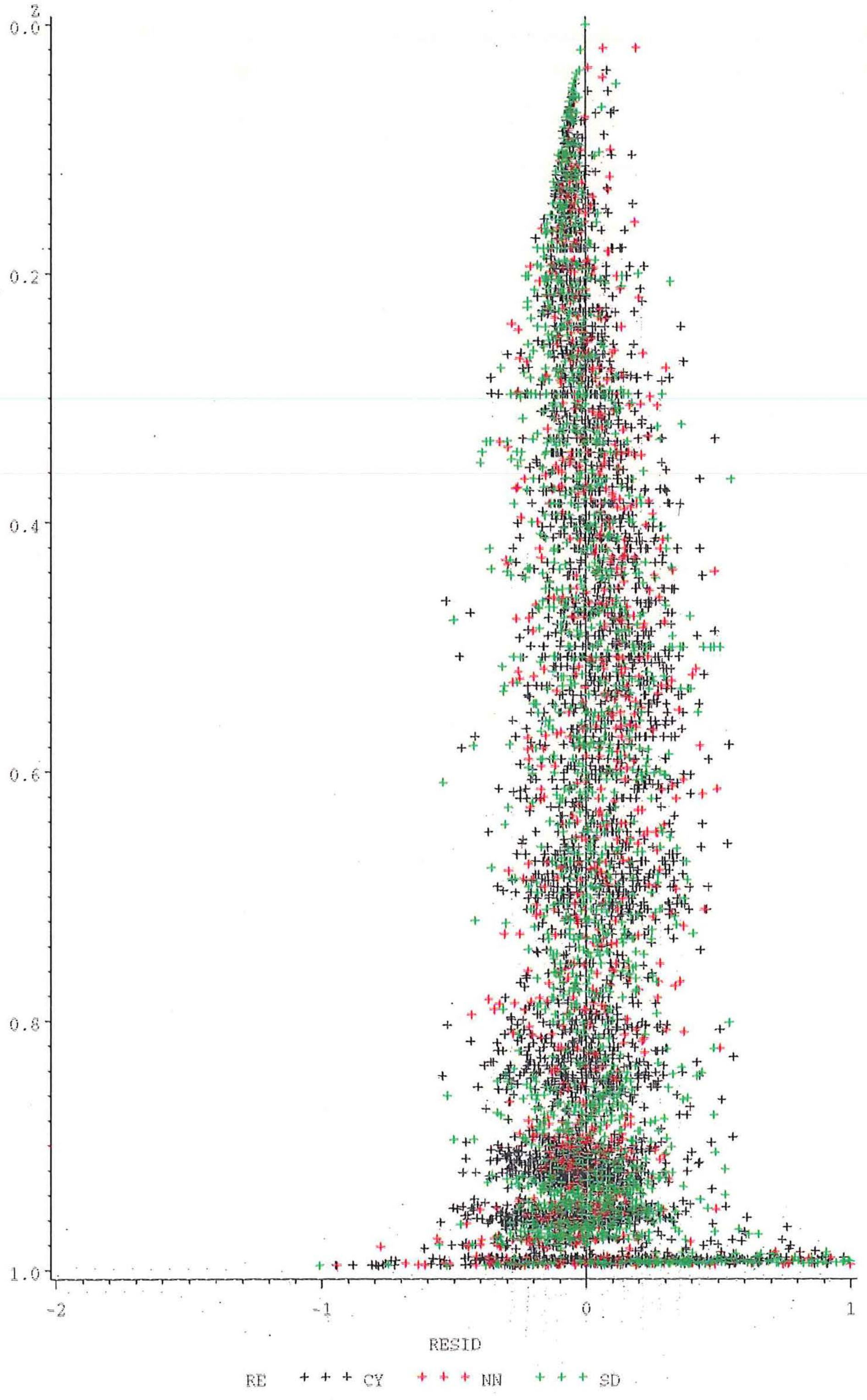
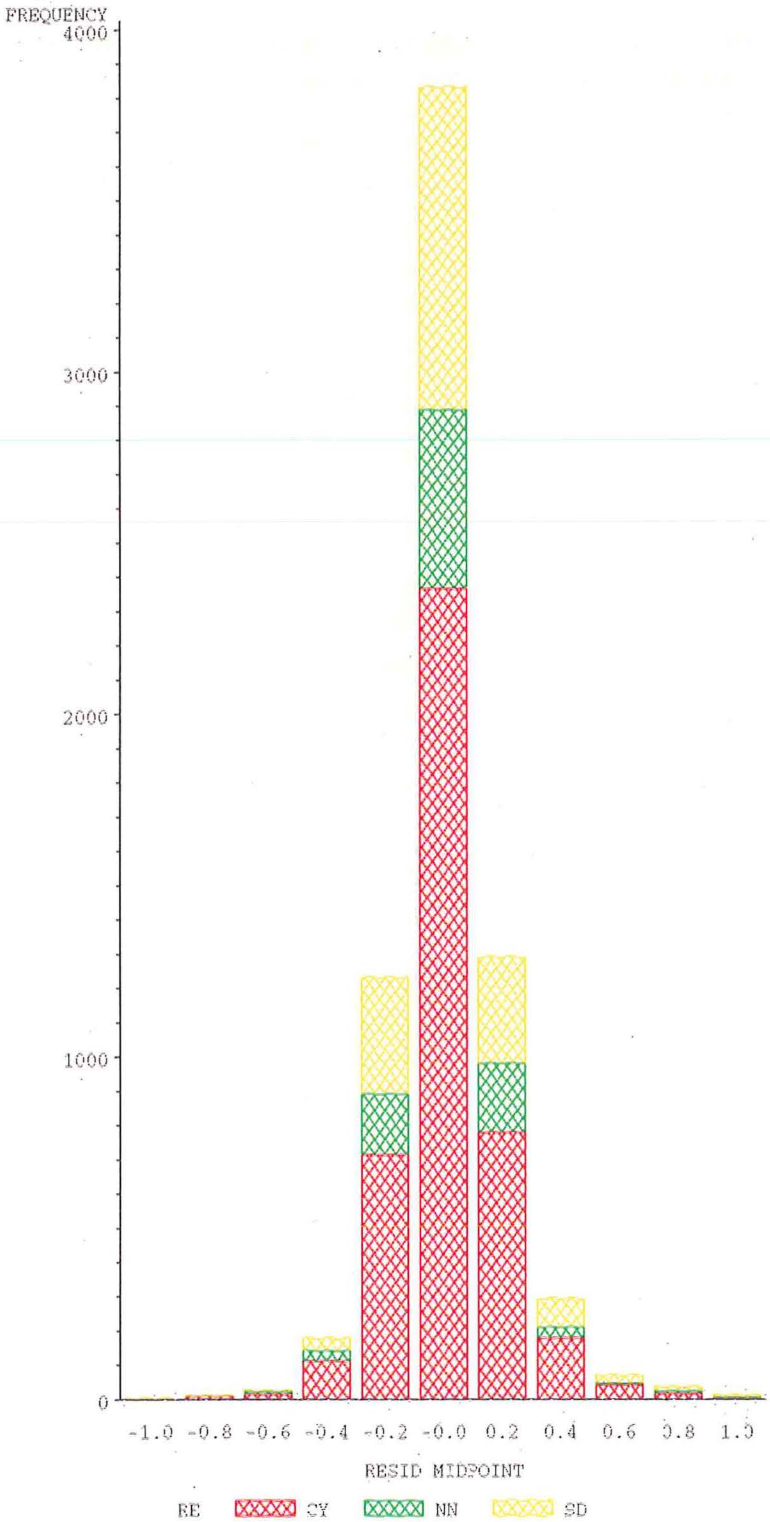


Figure 4.12: Frequency Distribution of Residuals
Adopted Taper Equation



4.2.4.2 Taper Based System

The procedure was repeated, but unlike the volume based system, taper is regarded to be the primary concern, and volume is obtained by integration of the taper equation. The same methodology and equations were retained as for the volume based procedure, but the stem volume v, was substituted as in equation 4.34. The approach is similar to that of Byrne and Reed (1986) but they used Cao et al. (1980) segmented taper equation 4.20.

$$v = Kfd^2h \tag{4.34}$$

After this substitution and rearrangement, equation 4.30 becomes equation 4.35 below.

$$\frac{d'^2}{d^2} = f[\beta_1 z^2 + \beta_2 z + d\beta_3(z - \alpha_1)^2 I_1 + d\beta_4(z - \alpha_2)^2 I_2] \tag{4.35}$$

d' is top diameter (cm), d is dbhob (cm) and f is the form factor, which is simultaneously determined through regression with other parameters of the equation. The integration computed from equation 4.35 is

$$v = \int_0^1 Kfd^2[\beta_1 z^2 + \beta_2 z + d\beta_3(z - \alpha_1)^2 I_1 + d\beta_4(z - \alpha_2)^2 I_2]h.dz \tag{4.36}$$

The results of this integration yields 4.37

$$v = Kfd^2h[\frac{\beta_1}{3} + \frac{\beta_2}{2} + d\frac{\beta_3}{3}(1 - \alpha_1)^3 I_1 + d\frac{\beta_4}{3}(1 - \alpha_2)^3 I_1] \tag{4.37}$$

Traditionally, the total volume of a tree can be obtained by expression 4.38

$$v = Kfd^2h \tag{4.38}$$

Compatibility requires that the total volume of the tree should equal the merchantable volume when integration is performed for z between the limits 0 and 1. Thus, as for equation 4.30 the compatibility restriction is enforced by equations 4.31, 4.32 and 4.33, and v_m equals v for Kfd^2h . Because stem taper depends much on locality, the form factor parameter (f) was estimated for each individual region. Thus, the form factors f_1, f_2 and f_3 for Canterbury, Nelson and Southland regions respectively, are analogous to regional dummy variables in the volume function.

It was noted, however, that by imposing restrictions 4.31 and 4.32 and 4.33 in equation 4.35 the coefficient β_2 was not significantly different from 0. It was possible to derive the parameter estimates of this equation by imposing restriction 4.33 alone, but the compatibility was only satisfied approximately. Table 4.12 summarizes the parameter

Table 4.12: Parameter Estimates and Standard Errors of Taper Based Taper Equation

Parameter	Estimate	Standard Error	ESS	N	MSE
α_1	0.348706716	0.01928597805			
α_2	0.929131830	0.00244793852			
β_1	1.900372163	0.03171623308			
β_2	0.727374098	0.02109987072	42.6656282	6996	0.0061064
β_3	-0.035492183	0.00345130818			
β_4	3.480118250	0.29840727686			
f_1	0.428699044	0.00161329875			
f_2	0.414923399	0.00222663674			
f_3	0.439686110	0.00197526882			

estimates for the approximate compatible taper based taper equation 4.35, also shown in the file TPPF.LIS in appendix C. Figures 4.13 to 4.15 show the plot of residuals against predicted values, plot of residuals against z and a bar chart of the frequency distribution of residuals respectively.

4.2.4.3 Summary - Taper Equations

The volume based system of taper equations was developed successfully (equation 4.30) and is the one recommended for Douglas fir in the South Island. The inclusion of d in the equation has provided two major advantages: (1) the parameter estimates have smaller confidence intervals than a similar equation without d; and (2) the residuals near the butt end are more restricted about the zero line (see Tables 4.9, 4.10 and 4.11, and Figures 4.7 and 4.10). The form of quation 4.30 is also compatible with any volume function regardless of the species. Its applicability, therefore, depends largely on the availability of a volume function and on the evaluation of precision.

The taper based system should only be applied at this time as an approximation when a volume table is unavailable. Tests with this equation showed that a compatibility of 98% can be achieved. This methodology, however, demonstrates a potentially useful analytical procedure, which could be adopted in other sets of data and species.

Figure 4.13: Plot of Predicted Values vs Residuals
Taper Based Taper Equation

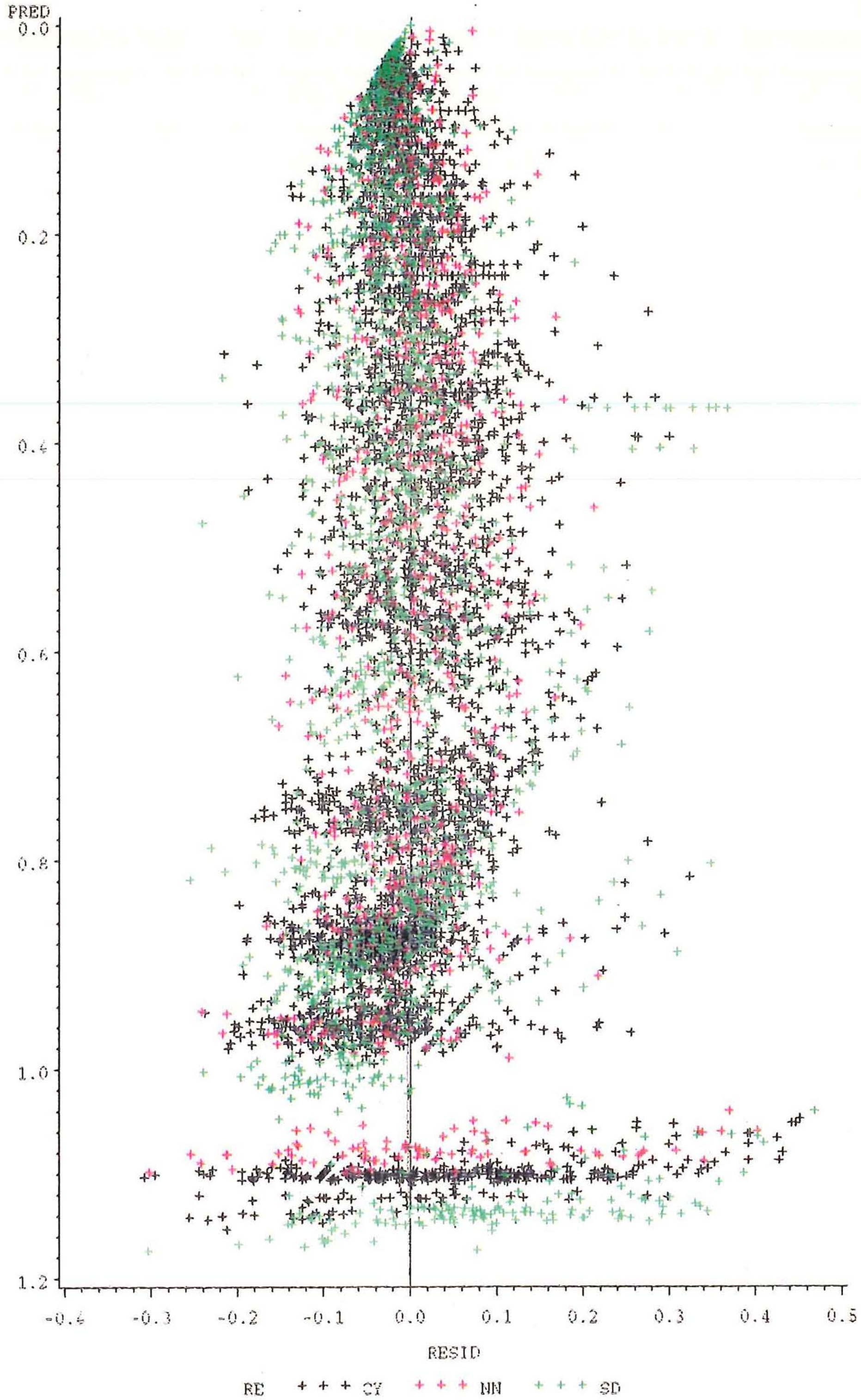


Figure 4.14: Plot of z vs Residuals
Taper Based Taper Equation

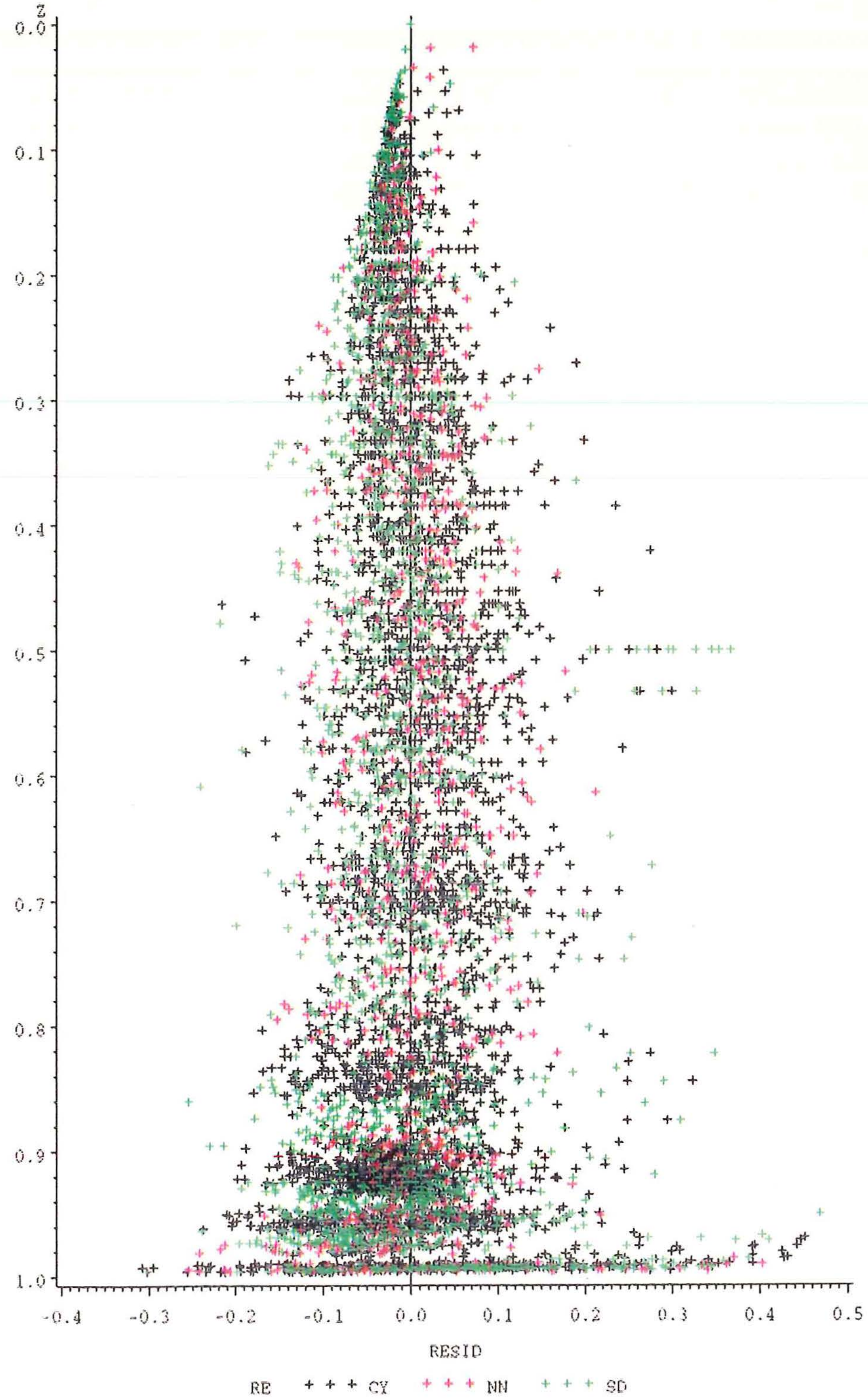
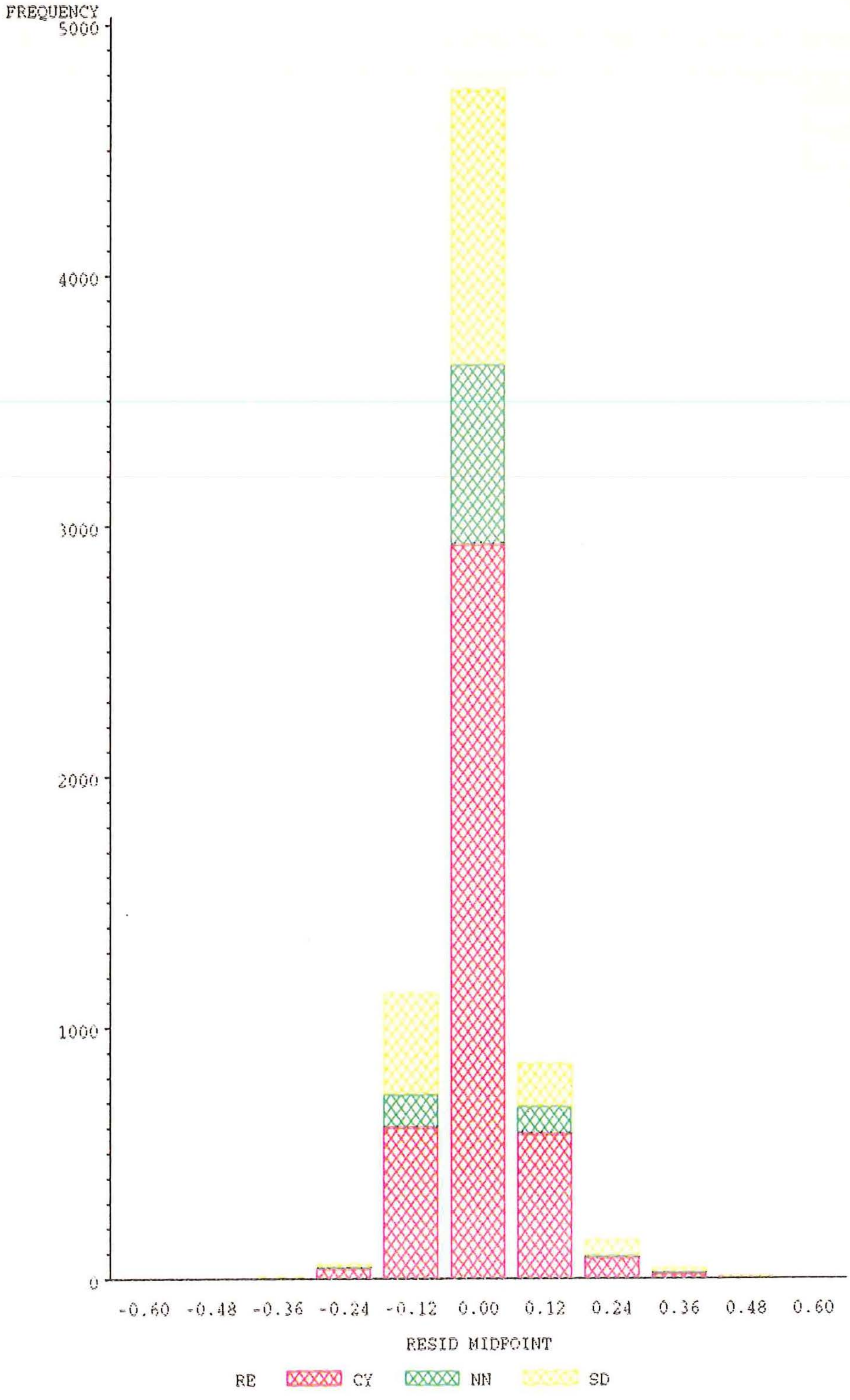


Figure 4.15: Frequency Distribution of Residuals
Taper Based Taper Equation



4.3 Whole Stand Growth and Yield Model [DfirStand]

DfirStand is a whole stand simultaneous growth and yield model applicable to four regions of the South Island in which Douglas fir is grown, namely Canterbury, Nelson, Southland and Westland. Unlike static models (Alder, 1980) DfirStand is a state - space dynamic growth and yield model (Garcia, 1988), capable of predicting rates of change under various conditions. The trajectories over time are then obtained by adding or integrating these rates.

A number of site variables such as altitude, latitude, distance from the sea, annual rainfall and soil type were considered as possible variables to explain regional growth variation.

4.3.1 Development of DfirStand

Polymorphic and anamorphic forms of Gompertz (Nokoe, 1978); Schumacher (Clutter et al., 1983), Hossfeld (Woollons et al., 1990), Levakovich, Weibull, monomolecular, Morgan - Mercer - Flodin, Umemura (Umemura, 1984) and Chapman - Richards (Pienaar and Turnbull, 1973) functions were all fitted to the data in developing individual equations. Comparison of equations was based on ESS and RMS values, parameter estimates and their asymptotic standard errors, PROC UNIVARIATE statistics and plots of residuals for the equations by region as set out in chapter 3 section 3.3.3.

In developing DfirStand, two hypotheses were postulated.

(i) For a large area with large population(s), there exist local biological and environmental adaptations which influence the growth of the crop in diverse ways. Local adaptations can be characterized through use of dummy and other predictor variables. Dummy variables are variables which can assume a value of 0 or 1. When the dummy variable assumes the value of 1 the coefficients which go with it are active, otherwise they are inactive.

(ii) local adaptations referred to in (i) combine with other management induced effects such as thinning and fertilization, which provide a range of influences on crop yield. Local adaptations are specific while the other effects are more general in nature.

These hypotheses were tested to ascertain the potential to develop a single growth and yield model which could be aggregated and disaggregated into individual regions or localities through dummy and other predictor variables reflecting these two kinds of influence.

4.3.2 Equations Employed in Stand Level Modelling

This subsection explains the form of equations used to develop DfirStand. The components of DfirStand are: (1) mean top height equation; (2) site index equation; (3) net basal area/ha equation; (4) equation to predict net basal area/ha after thinning (5) stand volume/ha production equation and (6) stem survival/ha equation. These are explained in the next subsections.

4.3.2.1 Mean Top Height Equation

Table 4.13 presents the data that were used to develop a general mean top height equation. The data are separated into individual regions. The variables used in this equation are age

Table 4.13: Data Used to Develop Mean Top Height and Site Index Equations

Region	N	Variable	mean	minimum	maximum
Canterbury	211	Age (T)	32.9	9.0	57.1
		Top height (m)	22.9	2.9	39.3
		Altitude (m)	327.6	150.0	790.0
Nelson	838	Age (T)	27.4	8.0	58.0
		Top height (m)	22.9	5.6	47.8
		Altitude (m)	439.2	183.0	625.0
Southland	347	Age (T)	33.9	7.0	78.0
		Top height(m)	24.3	4.3	46.2
		Altitude	234.9	0.0	330.0
Westland	189	Age (T)	25.6	5.0	59.1
		Top height (m)	18.4	1.9	37.5
		Altitude (m)	234.0	0.0	330
Total number of observations	1585				

of the crop (T), mean top height (h_{100}), and altitude (AL). The mean top height equation found to fit best was a form of an anamorphic Schumacher shown in 4.39.

$$h_{(100,2)} = h_{(100,1)} e^{(\alpha+\beta\times AL)(\frac{1}{T_1^\gamma}-\frac{1}{T_2^\gamma})} \qquad (4.39)$$

A single equation for all localities was found to be precise and with no regional bias in residual patterns, compared with fitting equations by individual regions or large forest aggregations on their own. Altitude controlled not only the level of asymptote, but also appeared to characterize differences in site quality very well from one locality to another.

Goodness of Fit of the Mean Top Height Equation

Parameter estimates for equation 4.39 are summarized in Table 4.14. All the parameter

Table 4.14: Parameter Estimates and Standard Errors of Mean Top Height Equation

Parameter	Estimate	Standard error	ESS	N	mean Square Error (MSE)
α	9.333951561	0.13347007407			
β	-0.001624622	0.00038075887	521.5633	1585	0.3297
γ	0.316495110	0.01445706630			

estimates were significant at least at the 5% level. Figure 4.16 shows the plot of residuals against predicted values. Figure 4.17 shows the plot of residuals against age (T_i), Figure 4.18 shows the plot of residuals against altitude and Figure 4.19 shows the chart of frequency distribution of the residuals. The precision achieved in this overall equation is better than in any other equation known to exist for this whole population or subset of it, as is also explained elsewhere in Whyte et al. (1992) (see appendix B). The residuals about the predicted values never exceeded ± 1.50 m. Inclusion of latitude, distance from the sea and rainfall was also tested, but none of the variables appeared to contribute any real predictive improvement and so were discarded. Apparently, the height/age curves are wholly anamorphic with respect to altitude and so it is possible to be reasonably confident of predicting mean top height of Douglas fir at any age (or its site index) anywhere in the South Island, given the crop's starting height is above age 5 and altitude within range 0 to 790 m above sea level.

4.3.2.2 Site Index Equation

Site index equations were derived from the mean top height equation by setting T_2 equal to 40 years (see equation 4.40), which is used here as the base age for Douglas fir in New Zealand (Burkhart and Tennent, 1977; Mountfort, 1978).

$$S = h_{(100,1)} e^{(\alpha+\beta \times AL)((\frac{1}{T_1})^{\gamma}-(\frac{1}{40})^{\gamma})} \tag{4.40}$$

The site index curves developed in this study are anamorphic with respect to altitude, they are displayed for the Nelson region at an average altitude of 438 m above sea level in Figure 4.20.

Figure 4.16: Plot of Residuals vs Predicted Values[m]
Mean Top Height Equation

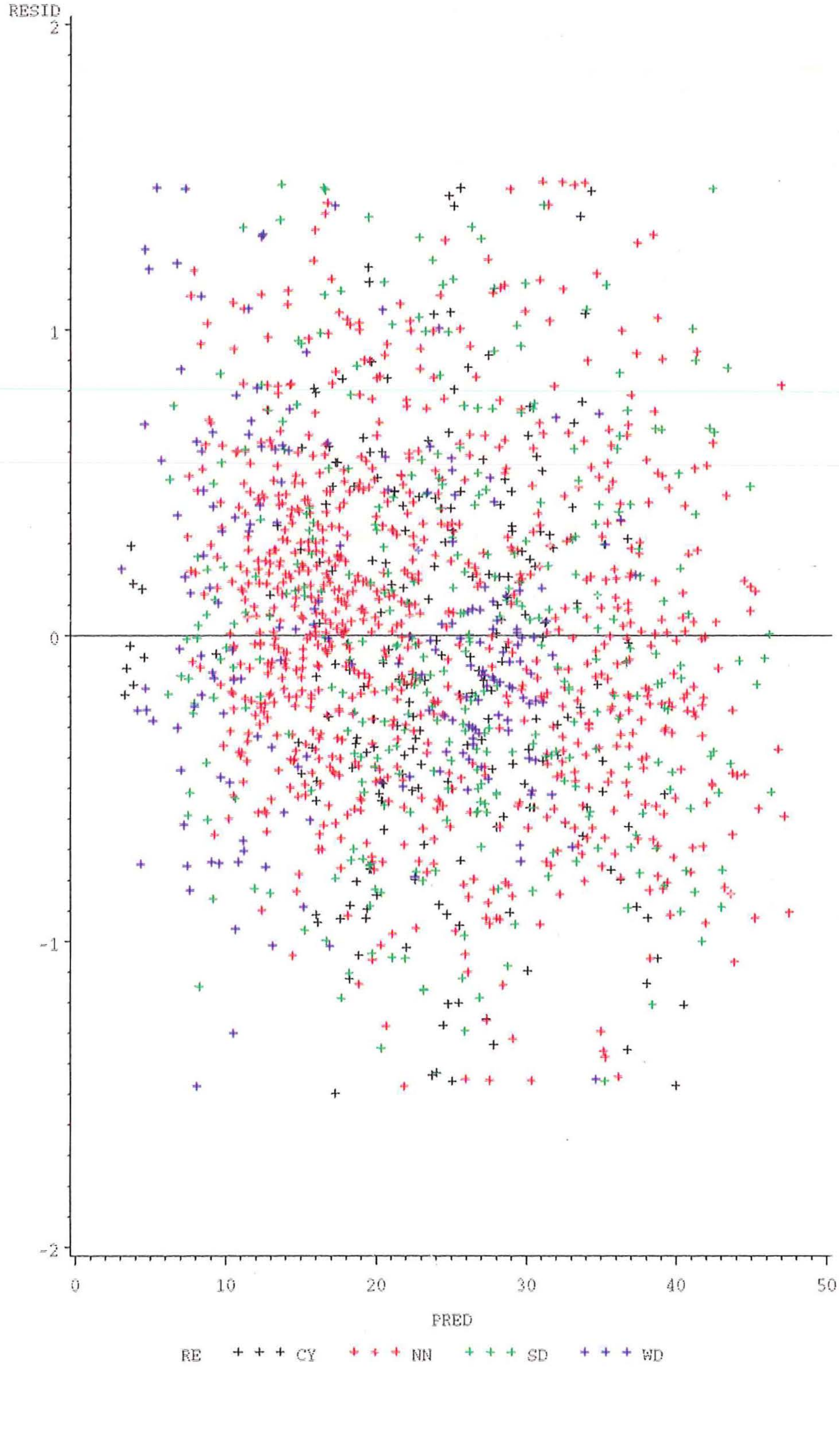


Figure 4.17: Plot of Residuals [m] vs Time [T1 years]
Mean Top Height Equation

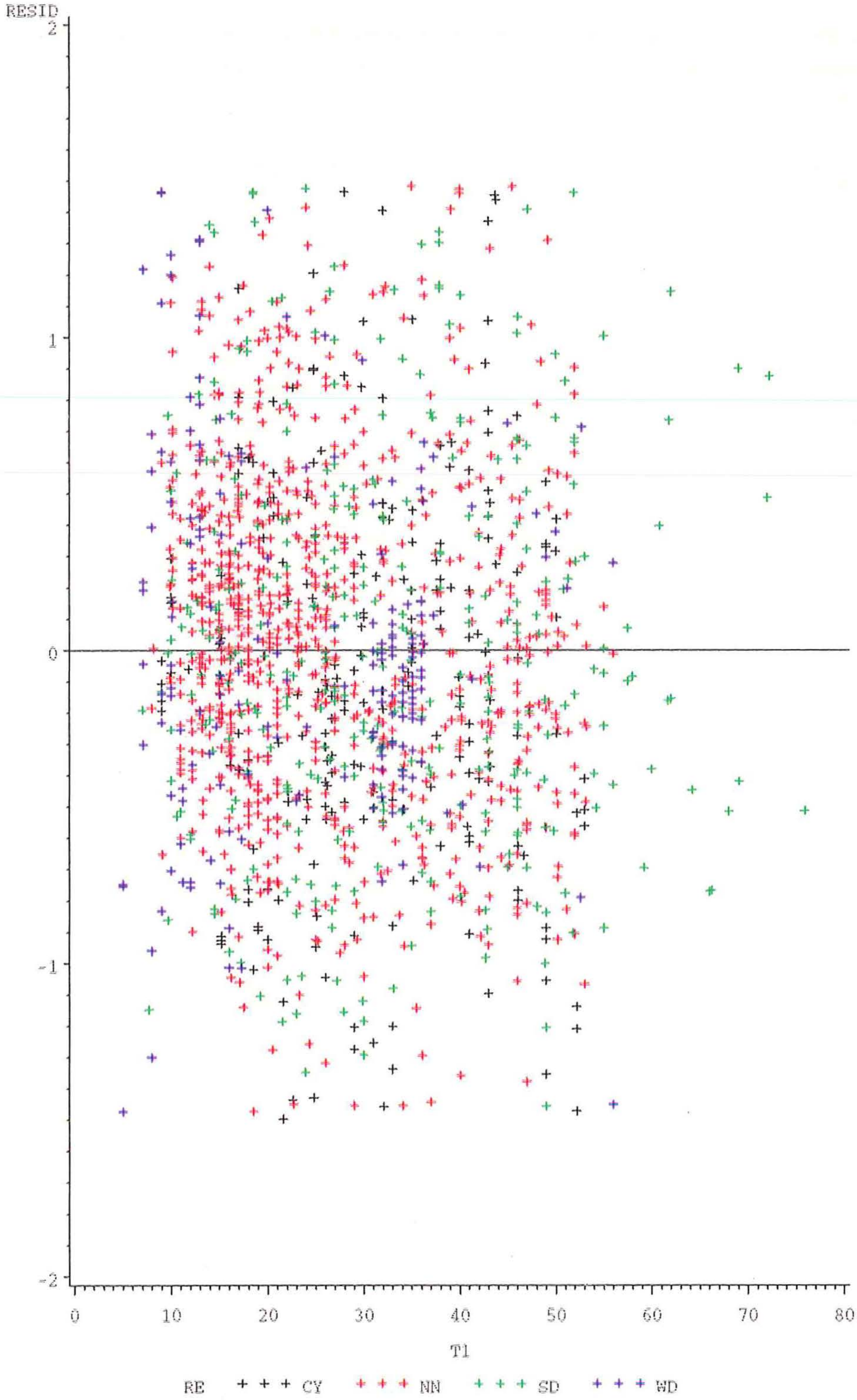


Figure 4.18: Plot of Residuals [m] vs Altitude [m a.s.l]
Mean Top Height Equation

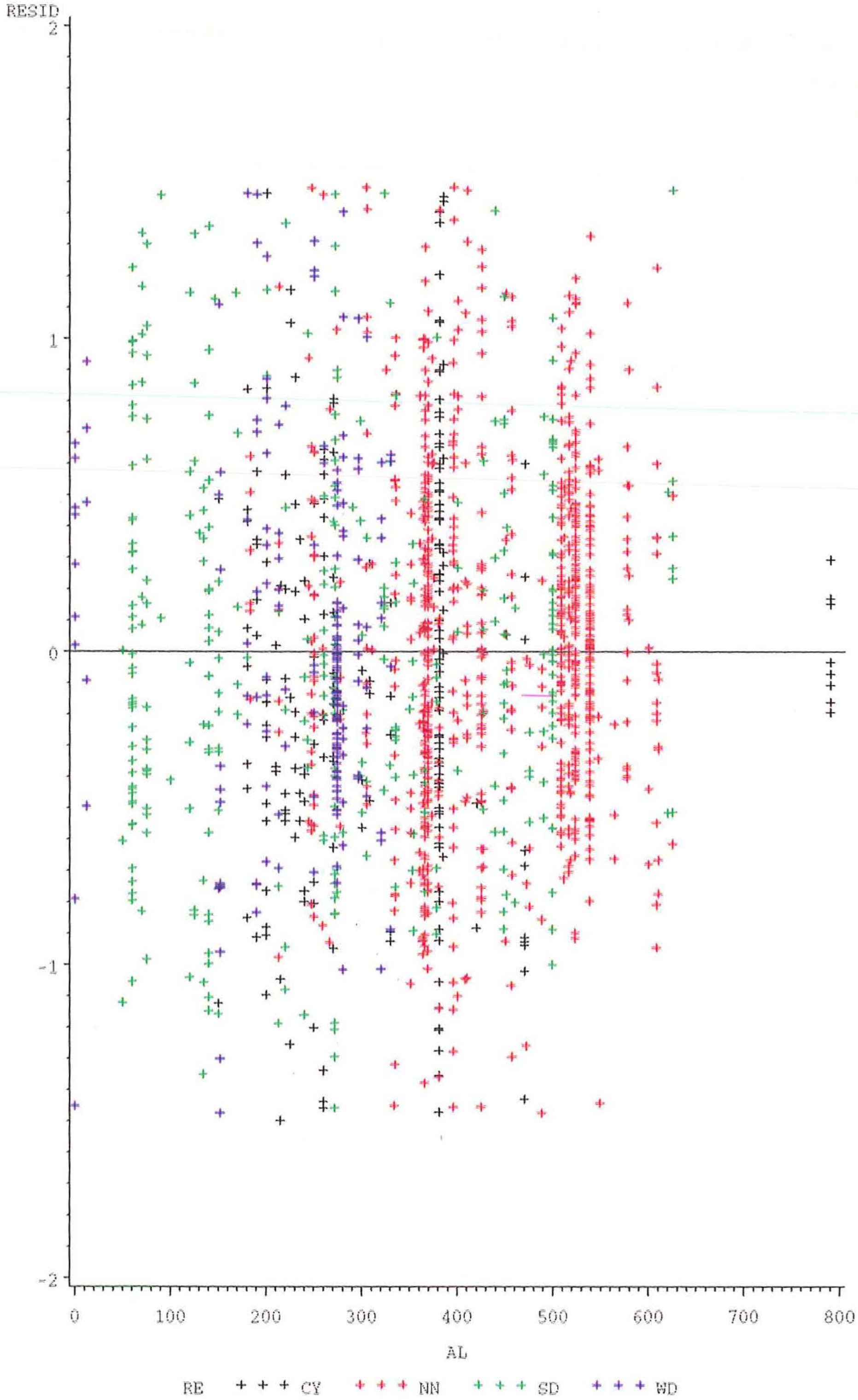


Figure 4.19: Frequency Distribution of Residuals [m classes]
Mean Top Height Equation

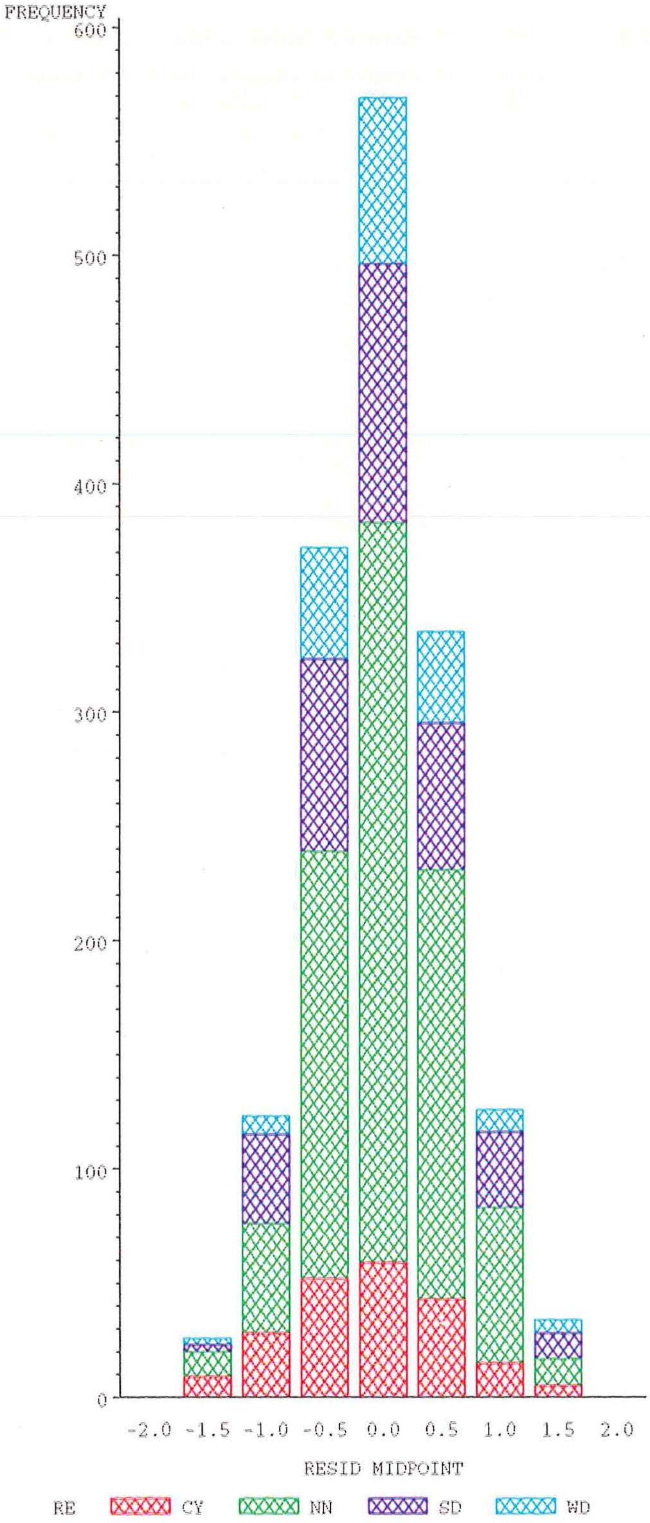
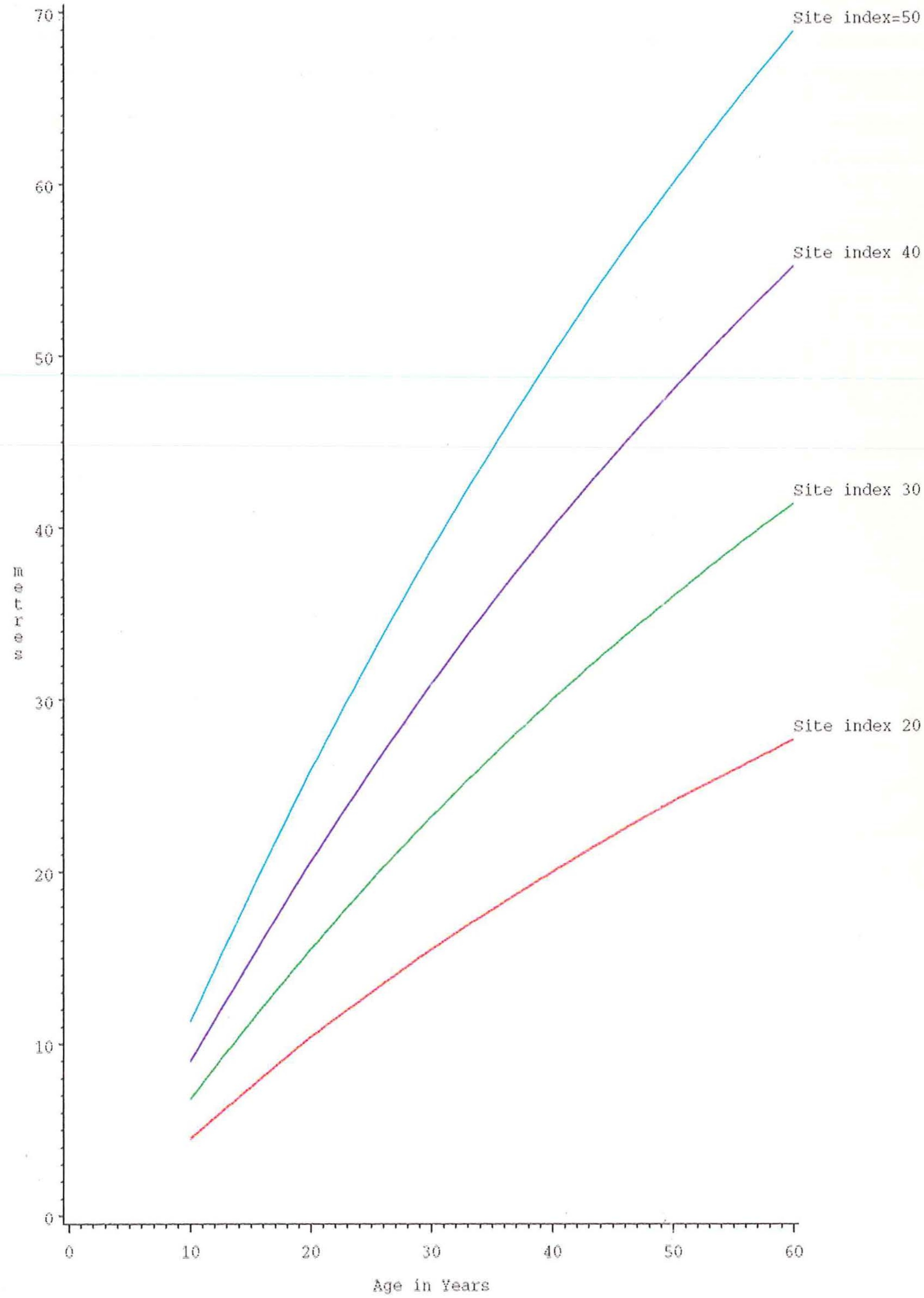


Figure 4.20: Site Index Curves
Douglas fir Nelson Region



4.3.2.3 Net Basal Area Projection Equation

Data Used in Developing of Net Basal Area Equation

The data used in developing a net basal area/ha equation are summarized in Table 4.15 (figures in parenthesis in column 1 are numbers of measurements). The variables used

Table 4.15: Basal Area/ha Data for Canterbury, Nelson, Southland and Westland

Region	Variable	Mean	Minimum	Maximum
Canterbury (189)	Age(yrs)	34.3	11.8	61.0
	G/ha	44.1	4.4	116.2
	Site index (m)	28.1	22.7	35.8
	Altitude (m)	322	150	470
	Thinning index	0.3372	0.07350	1.491900
Nelson (841)	Age(yrs)	27.9	7.0	58.0
	G/ha	41.9	1.2	109.4
	Site index (m)	33.5	22.7	41.9
	Altitude (m)	439	183	625
	Thinning index	0.2577	0.03000	2.3751
Southland (345)	Age(yrs)	33.7	7.0	75.0
	G/ha	50.8	1.1	112.8
	Site index (m)	29.3	18.3	37.6
	Altitude (m)	253	50	625
	Thinning index	0.2349	0.03390	0.6441
Westland (209)	Age(yrs)	26.8	5.0	59.1
	G/ha	29.0	0.0	123.8
	Site index (m)	30.0	10.5	38.1
	Altitude (m)	234	0	330
	Thinning index	07520	0.09900	3.0000

in this equation are net basal area/ha (G), age of crop (T), thinning index (X_t), altitude (AL) and site index (S). In addition three dummy variables K_1, K_2, and K_3 were included to distinguish locality.

The thinning index, X_t, is in this equation defined as

$$X_t = 1 - \frac{d_t}{d_b} \tag{4.41}$$

where d_b is the quadratic mean diameter of the stand just before thinning and d_t is the quadratic mean diameter of trees removed in thinning. Several sigmoidal curves(Schumacher, Chapman-Richards, Gompertz, Hossfeld, Weibull and Morgan-Mercer) were fitted to basal area data, the goodness of fit of each equation was analyzed. A polymorphic form of the Schumacher, equation 4.42 originally used by Clutter and Jones (1980) fitted the data most and was selected to form a base model for subsequent analysis.

$$G_2 = G_1^{(\frac{T_1}{T_2})^{\beta_1}} e^{\alpha(1-(\frac{T_1}{T_2})^{\beta_1})} \tag{4.42}$$

Equation 4.42 fitted the data well, and so, site index was later included as a variable to test whether the fit could be further improved. Equation 4.43 was thus fitted to the data.

$$G_2 = G_1^{(\frac{T_1}{T_2})^{\beta_1}} e^{(\alpha+\beta_2 S)(1-(\frac{T_1}{T_2})^{\beta_1})} \tag{4.43}$$

The impact of site index in the basal area equation was unexpected: not only was it not significant (at 5% level), but the parameter β_2 was negative suggesting that higher site indices had lower net basal area production. This is contrary to what researchers have established for a long time in the context of site index and growth.

A statistical analysis was conducted and it was possible to show that there were interacting effects between altitude and site index particularly in Canterbury data; higher elevations in Canterbury have a higher rainfall and better soils suitable for Douglas fir. It was therefore concluded that altitude could be a more efficient and expressive factor, which could substitute readily for site index. The variable site index (S) was then omitted and altitude above sea level was substituted as in equation 4.44.

$$G_2 = G_1^{\left(\frac{T_1}{T_2}\right)^{\beta_1}} e^{(\alpha+\beta_2 AL)(1-(\frac{T_1}{T_2})^{\beta_1})} \tag{4.44}$$

Locality adaptation was incorporated in the equation through use of three dummy variables K_1, K_2, and K_3, with adaptation coefficients β_3, β_4 and β_5 respectively. A series of nested equations (refer to section 2.8.5) was fitted, allowing the asymptote parameter α to vary for each region, while β_1 and the altitude coefficient, β_2 were assumed to be fixed. A simple SAS program was used to input the dummy variables, such that they assume a value of 1 for the intended region, else a value of 0 is assumed. Out of these nested equations, equation 4.45 fitted the data best. Thus, the dummy variables K_1, K_2, and K_3 assume the value of 1 for Nelson, Southland, and Westland regions respectively, otherwise they assume a value of 0, while Canterbury is the default.

$$G_2 = G_1^{(\frac{T_1}{T_2})^{\beta_1}} e^{(\alpha+\beta_2 AL+K_1\beta_3+K_2\beta_4+K_3\beta_5)(1-(\frac{T_1}{T_2})^{\beta_1})} \tag{4.45}$$

Altitude and dummy variables are being used here for determining the level of asymptote for each region in equation 4.45. The use of local altitude appeals more in terms of localisation rather than if average altitude was used.

Quantification of Amount and Kind of Thinning

Thinning was incorporated in equation 4.45 in such a way that the equation could be used for predictions in both thinned and unthinned stands. The thinning index used in this study is similar to that of Murphy and Farrar (1988) but with some modification, the difference being how the age of thinning T_t is referenced in the equation so that at the same time all biological and logical properties of the equation are retained. Murphy and Farrar (1988), defined their thinning term as

$$X_t \frac{(\frac{1}{T_2} - \frac{1}{T_1})}{(T_t T_2)} \tag{4.46}$$

Where X_t, the thinning index, is defined as

$$X_t = 1 - \frac{d_t}{d_b} \tag{4.47}$$

and

d_t = quadratic mean diameter of trees removed in thinning.

d_b = quadratic mean diameter of the stand just before thinning.

T_t =Thinning age (T_t<=T_1)

T_1 and T_2 are respective ages (years) in a thinning interval.

In their data, thinning was done at equal intervals of 5 years, while the interval of thinning in this study ranged from 3 to 30 years. In this study the thinning term adopted and found to be helpful in predicting the effects of variable thinning intensities and variable thinning intervals was

$$X_t(\frac{1}{T_2} - \frac{1}{T_1})\frac{T_t}{T_2} \tag{4.48}$$

The inclusion of the thinning term 4.48 in the equation has two important properties, namely,

(1) as T_2 approaches ∞ then the thinning term in equation 4.48 approaches 0, so that the ability of stands to respond to thinning diminishes with age,

(2) because $\frac{1}{T_2} - \frac{1}{T_1}$ is always negative, and β_6 was also negative (see Table 4.15), the thinning term is a decreasing function of T_2, but it diminishes at a slower rate than that of Murphy and Farrar (1988) by the factor of $\frac{1}{T_2^2}$.

Because the age of thinning in equation 4.48 is a constant, the overall equation 4.49 retains all the other basic properties, namely, compatibility between growth and yield (Clutter, 1963), path invariance, and consistency, although it is not readily apparent at first glance and without some rearrangement of the last term (thinning term) in the adopted equation 4.49.

$$G_2 = G_1^{(\frac{T_1}{T_2})^{\beta_1}} e^{((\alpha+\beta_2 K_1+\beta_3 K_2+\beta_4 K_3+\beta_5 AL)(1-(\frac{T_1}{T_2})^{\beta_1})+\beta_6 X_t(\frac{1}{T_2}-\frac{1}{T_1})\times\frac{T_t}{T_2})} \tag{4.49}$$

Goodness of Fit of Net Basal Area/ha Projection Equation

Table 4.16 below sets out the parameter estimates and standard errors for equation 4.49 The goodness of fit was evaluated through plots of residuals against predicted values as shown in Figure 4.21, Figure 4.22 shows the plot of residuals against time (T_1), Figure 4.23 shows the plot of residuals against altitude, and Figure 4.24 shows the chart of residuals.

Table 4.16: Parameter Estimates and Standard Errors of Basal Area Equation, 4.48

Parameter	Estimate	SEE	ESS	N	MSE
α	5.096954651	0.02899877466			
β_1	1.064813569	0.01533259541			
β_2	0.053376341	0.01976034060			
β_3	0.276381537	0.02089265203	2674.0500	1592	1.6871
β_4	0.312440622	0.03543866689			
β_5	-0.000244357	0.00005286663			
β_6	-1.834181689	0.83833778987			

There were no apparent regional biases in basal area projection with the overall equation 4.49, whereas there were with the average one as in 4.44. When an average Schumacher fit was made to the data (equation 4.42), the residuals extended $\pm$ 10 m^2/ha about zero, whereas equation 4.49 with its additional explanatory and locality variables was able to contain them all within $\pm$ 4.0 m^2/ha, while 95% of the residuals were contained within $\pm$ 3.0 m^2/ha. Parameter estimates in Table 4.16 can be interpreted as follows: α, β_1, β_5, and β_6 are common coefficients contributing to the asymptote of each region, β_5 is negative indicating that basal area growth declines with increasing altitude. The thinning term coefficient β_6 is also negative, but because $1/T_2 - 1/T_1$ in equation 4.48 is negative, it implies that basal area growth responds positively with thinning. In addition β_2 contributes to the asymptote if region is Nelson, β_3 contributes to the asymptote if region is Southland and β_4 contributes to the asymptote when region is Westland.

Successive Improvement of Net Basal Area Equation

Table 4.17 summarizes the successive improvement of the basal area/ha equation through incorporating different variables and coefficients. The levels of improvement over an

Table 4.17: Successive Improvement of Net Basal Area/ha Equation

Input Variables	ESS	% Reduction in ESS	MSE
α	3765.3352	-	2.3666
α, β_1	3709.3588	1.5	2.3329
α, β_1, Z_1, Z_2, Z_3	2719.4125	26.6	1.7136
α, β_1, Z_1, Z_2, Z_3, AL	2682.3634	1.4	1.6913
α, β_1, Z_1, Z_2, Z_3, AL, X_t	2674.0500	0.31	1.6871

average fit and the precision attained *per se* in equation 4.49 provides a level of sensitivity that managers should be relatively content to work with.

Figure 4.21: Plot of Residuals vs PredictedValues [m^2]
Net Basal area/ha equation

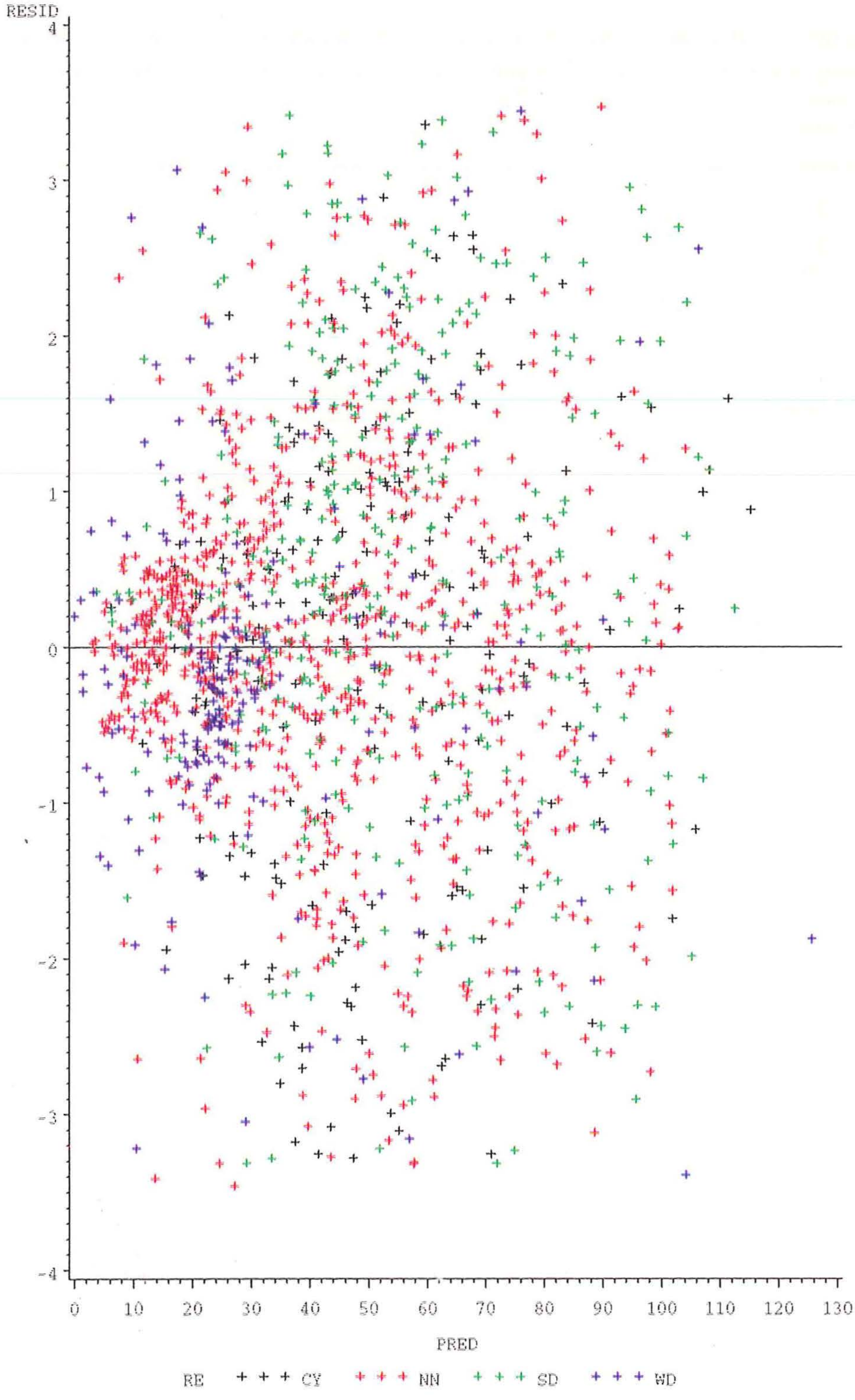


Figure 4.22: Plot of Residuals [m^2] vs Time[T1 years]
Net Basal area/ha equation

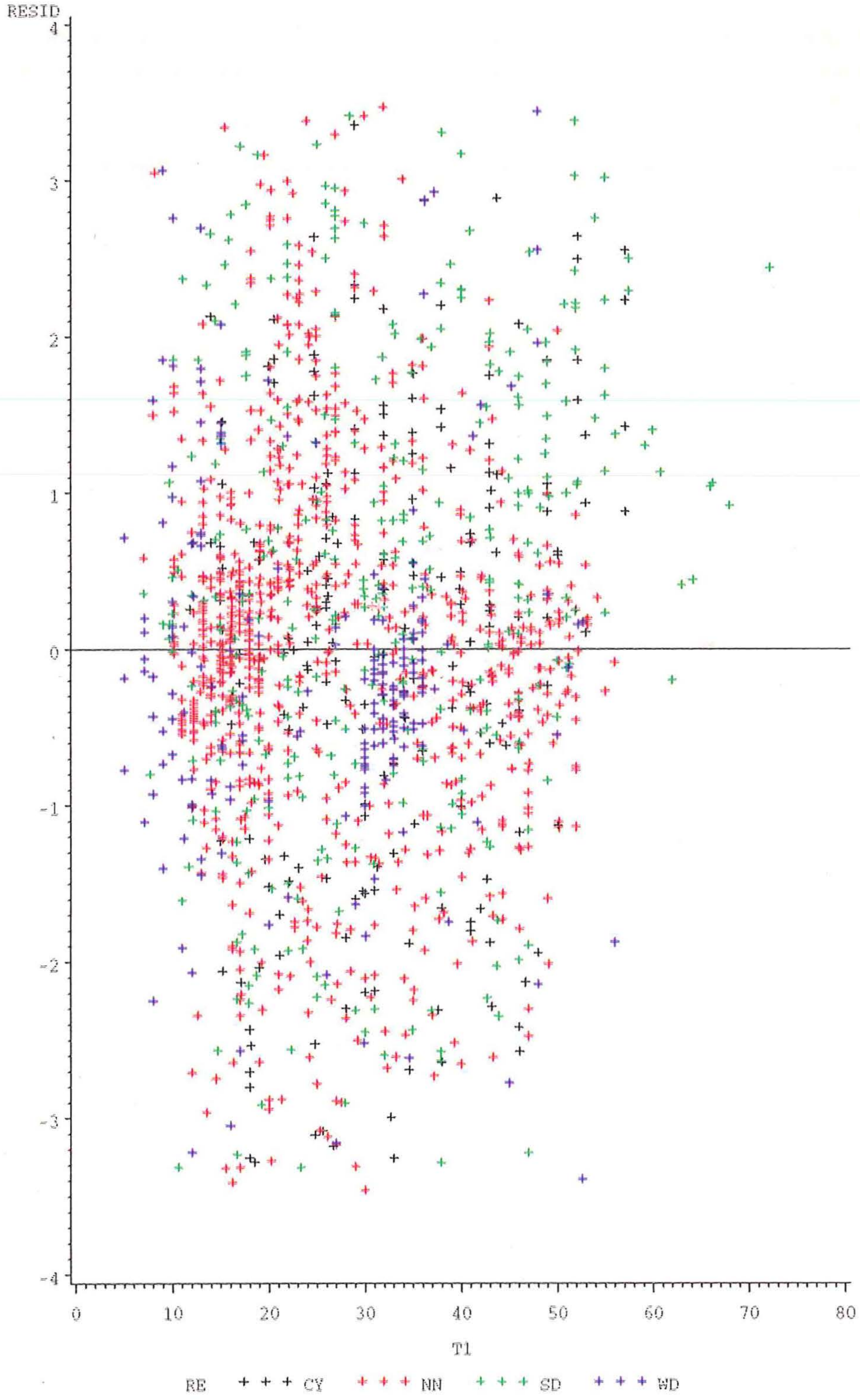


Figure 4.23: Plot of Residuals [m^2] vs Altitude[m a.s.l]
Net Basal area/ha equation

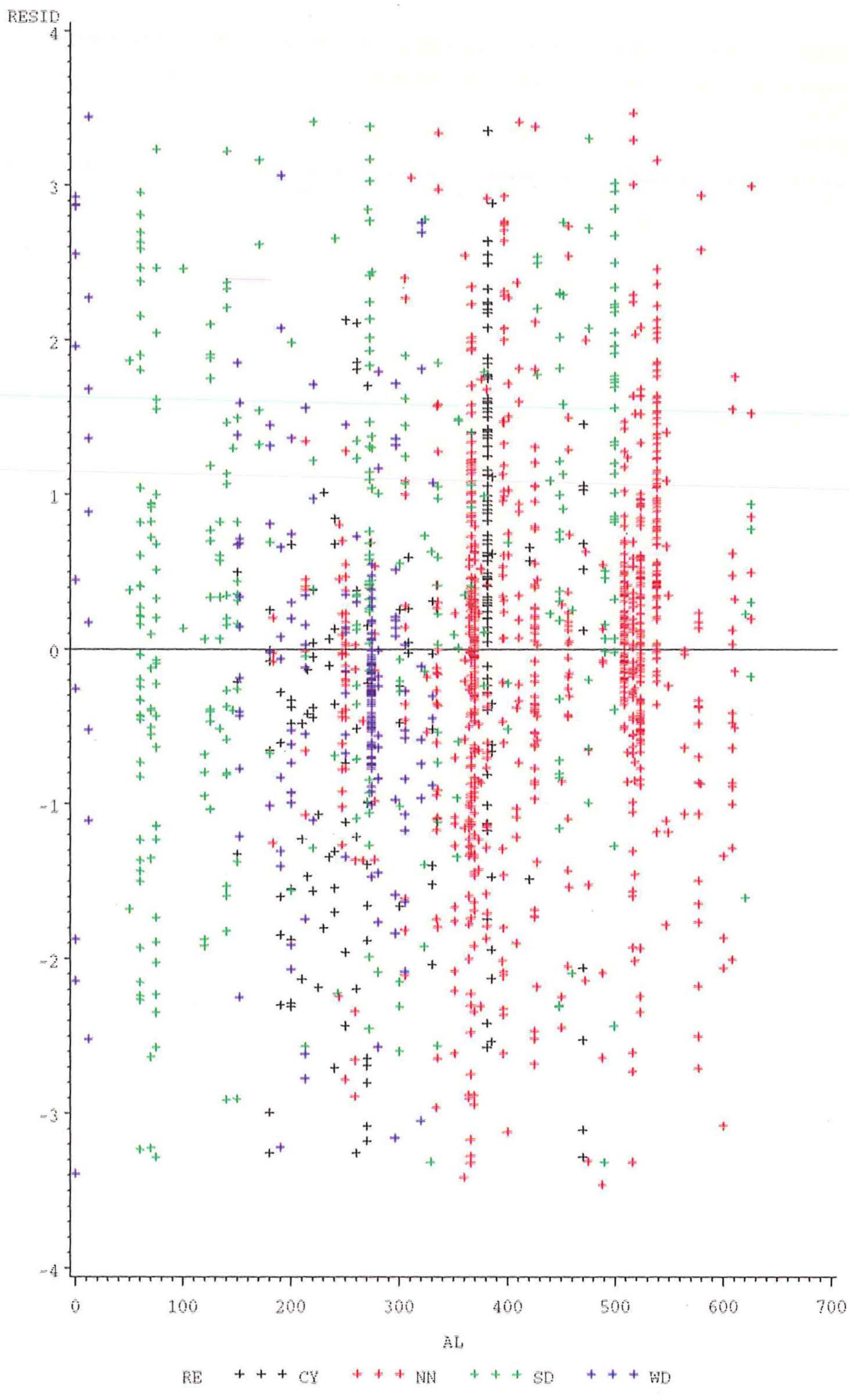
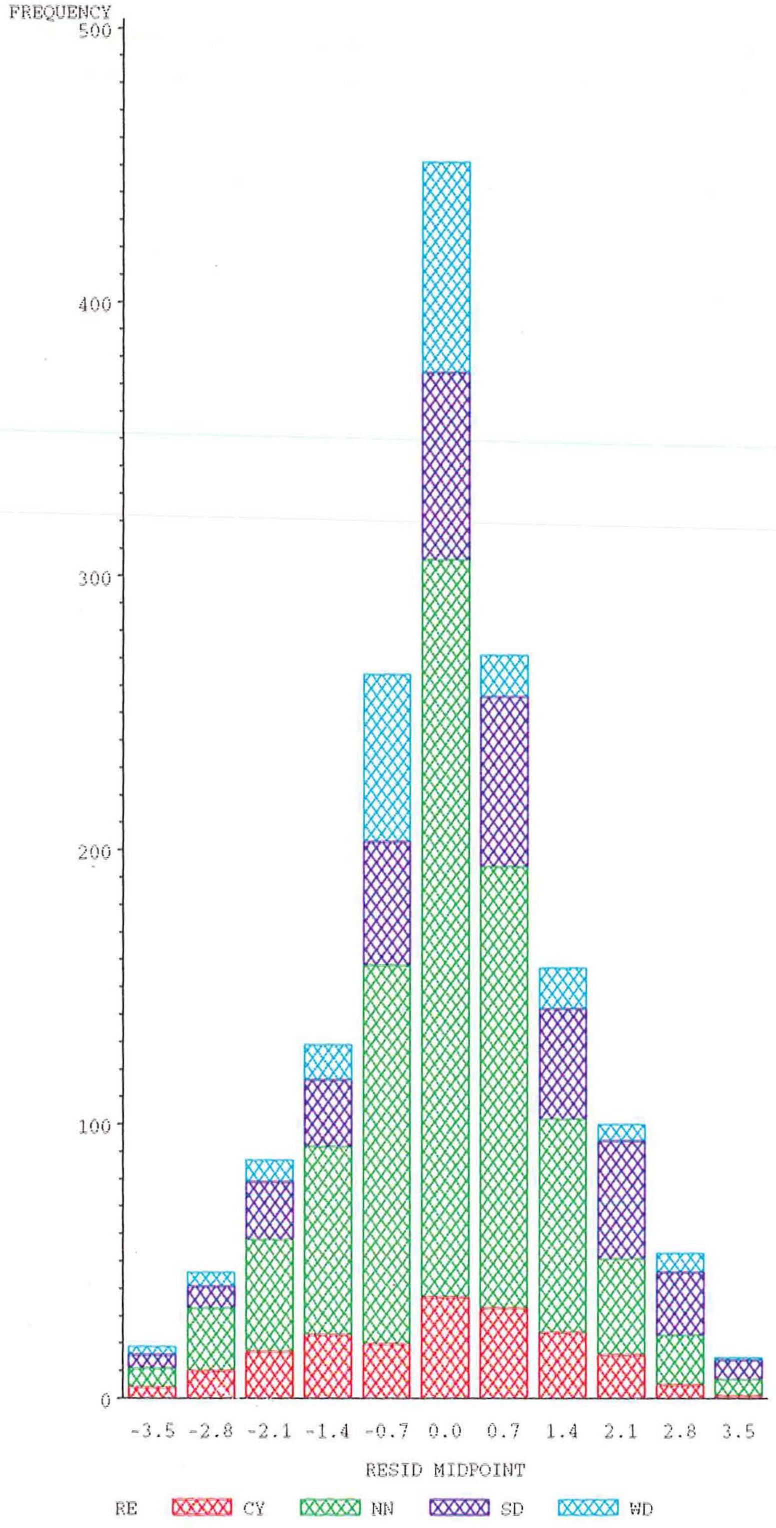


Figure 4.24: Frequency Distribution of Residuals [m^2 classes]
Net Basal area/ha equation



4.3.2.4 Prediction of Net Basal Area/ha After Thinning

Prediction of net basal area/ha and stocking after thinning is not always easy unless these values are specified directly. Nevertheless, if they are not specified an equation to predict basal area after thinning or to estimate the number of trees to be thinned to achieve a given level of basal area needs to be provided. The thinning data for this equation are those summarized in Table 3.9, Altogether 172 measurements were available for this analysis. Equations 4.50 (Matney and Sullivan 1982b) and 4.51 (Garcia, 1984b) were fitted to the thinning data and compared.

$$G_a = \beta_0 G_b^{\beta_1}(1 - (1 - N_a/N_b)^{\beta_2})^{\beta_3} \tag{4.50}$$

$$G_a = G_b e^{\beta_0 (h_{100}^{\beta_1} (N_a^{\beta_2} - N_b^{\beta_2})} \tag{4.51}$$

where G_b , G_a and are net basal area/ha before and after thinning, N_b , N_b are stocking before and after thinning, h_{100} is mean top height and $\beta_{'s}$ are non linear least - squares coefficients. Equation 4.50 fitted the data better and was adopted. Parameter estimates for this equation are summarized in Table 4.18, Figures 4.25 and 4.26 represent the residual plots against predicted values and residual bar chart respectively. In order to make precise projections it is recommended that users should not re-arrange the equation and predict basal/ha before thinning from stocking because:

(1) due to small number of observations (172), the residual bar chart of this equation is slightly biased indicating that the parameter estimates are relatively poor and more bias might be introduced;

(2) basal/ha after thinning estimated from stocking is an average value, while thinnings with different intensities, time and interval between thinning could induce much variation;

(3) stand volume production is more closely related to basal area/ha than to stocking.

Table 4.18: Parameter Estimates for Basal Area/ha After Thinning

Parameter	Estimate	SEE	ESS	N	MSE
β_0	1.285979977	0.08614742159			
β_1	0.925156873	0.01576363838			
β_2	1.104682651	0.09263427419	917.27792	172	5.45999
β_3	0.732974116	0.03283246952			

Figure 4.25: Plot of Residuals vs Predicted Values[m^2]
Basal area/ha After thinning Equation

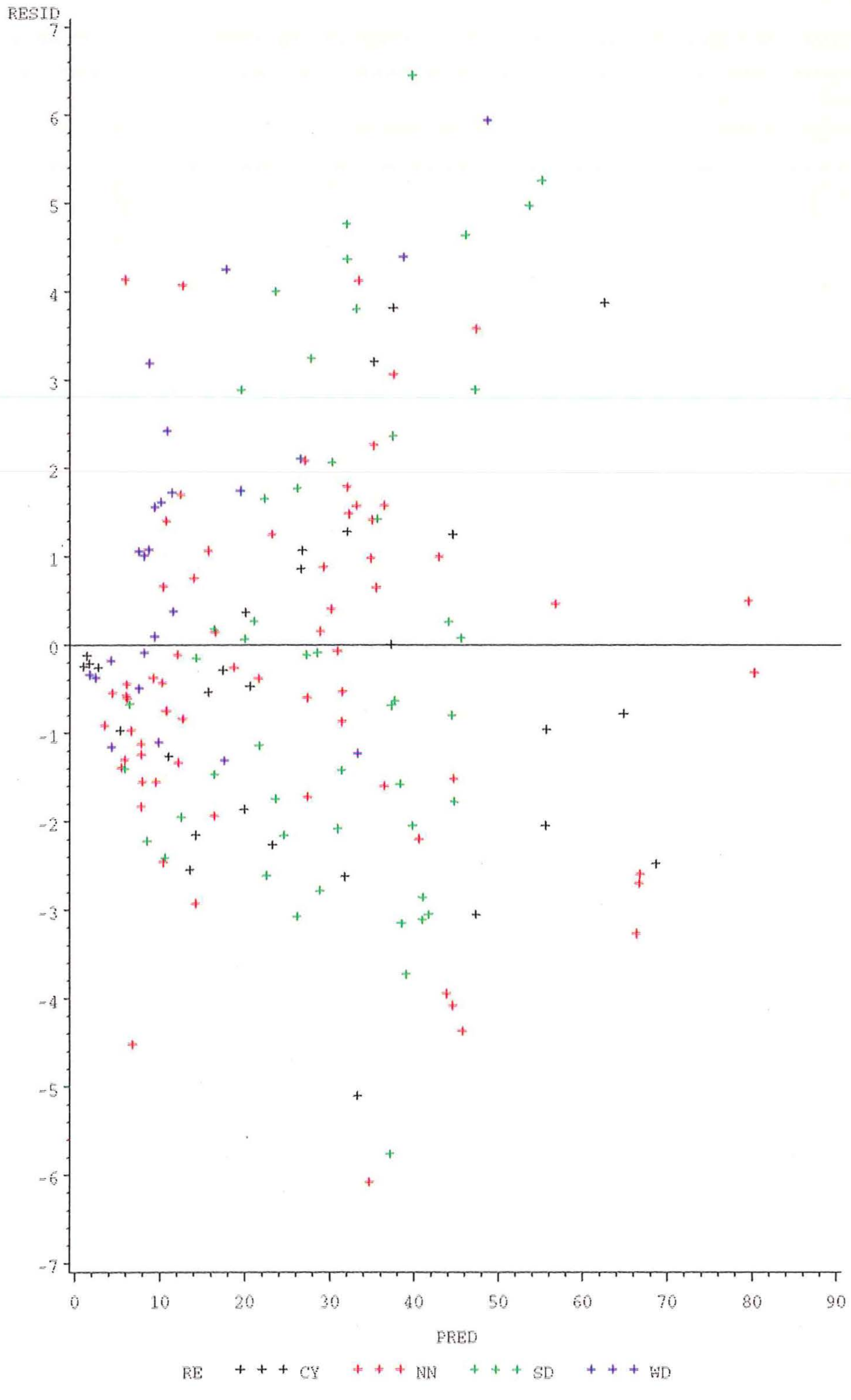
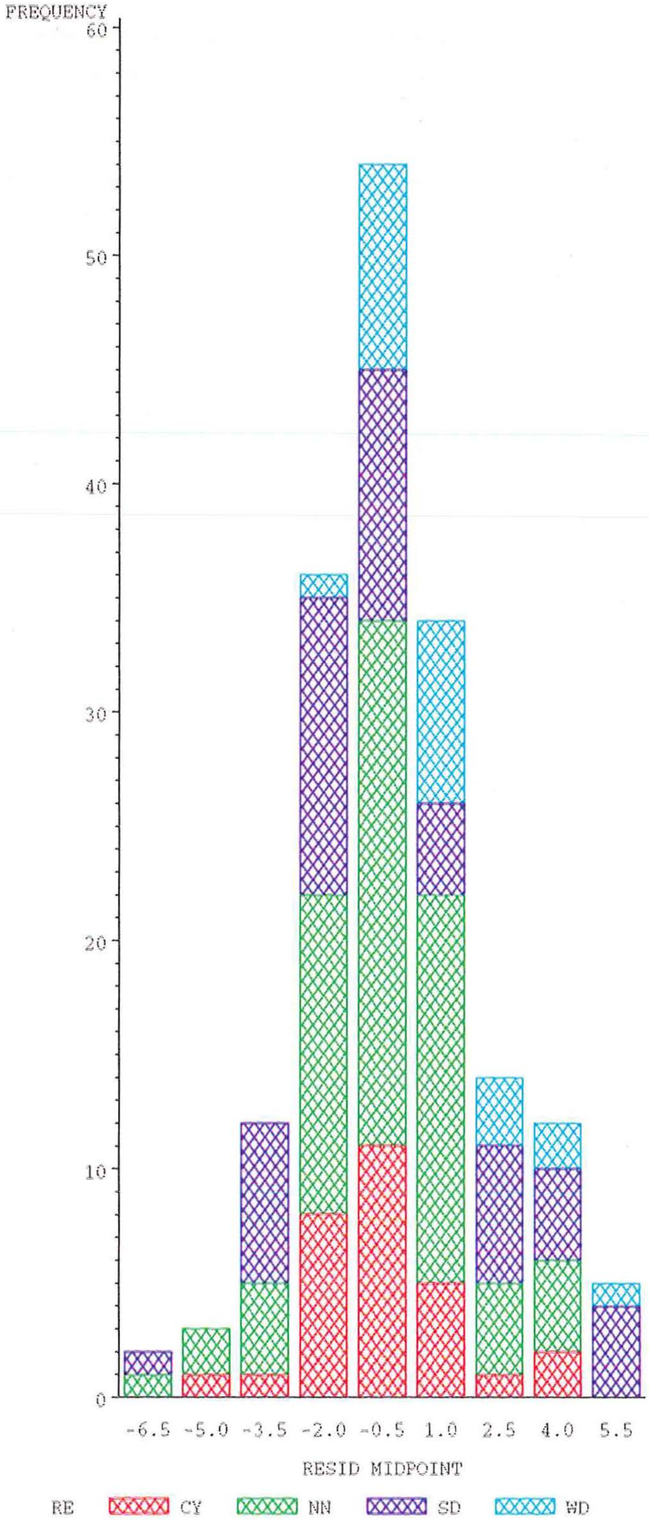


Figure 4.26: Frequency Distribution of Residuals[m^2 classes]
Basal area/ha After thinning Equation



4.3.2.5 Stand Volume Equation

The plot volumes were recalculated using the tree volume equation explained in section 4.1.4 The data were then used to develop a new stand volume equation. The statistics of the basic data used for developing the stand volume production equation are summarized in Table 4.19. The combined variable of form in equation 4.52 fitted the data best.

Table 4.19: Regional Data for Development of Stand Volume Equation

Region	Variable	N	Mean	minimum	maximum
Canterbury	160	G	49.3	6.5	108.2
		h_{100} (m)	24.7	9.3	39.3
		V	445.8	24.5	1278.8
Nelson	762	G	43.1	3.2	109.4
		h_{100} (m)	23.7	8.6	46.6
		V	399.6	11.7	1451.2
Southland	293	G	53.7	7.6	109.5
		h_{100} (m)	26.3	6.9	46.2
		V	445.8	24.5	1278.8
Total Number of Measurements	1215				

$$V = \alpha G^{\beta} h_{100}^{\gamma} \tag{4.52}$$

The equation was weighted by a weight w shown in equation 4.53 to counteract heterogeneity of variance of volume per hectare with respect to net basal area per hectare.

$$w = \frac{1}{G^2 h_{100}} \tag{4.53}$$

Parameter estimates and standard errors of equation 4.52 are set out in Table 4.20.

Table 4.20: Parameter Estimates and Standard Errors of Stand Volume Equation

Parameter	Estimate	SEE	ESS	N	MSE
α	0.5665215579	0.00455205065			
β	0.9776534563	0.00192235190	3.5835043	1215	0.0029567
γ	0.8790153619	0.00295599158			

Table 4.21: Regional biases in Predicting Volume/ha

Region	N	Residual		
		Mean	minimum	maximum
Canterbury	160	-2.0	-35.9	25.8
Nelson	762	-0.35	-45.5	47.9
Southland	293	3.0	-37.9	47.4
Overall bias	1215	0.25	-45.5	47.9

Goodness of Fit of The Stand Volume Equation

The variables used in developing the stand volume equations (basal area/ha, and mean top height) provide an indirect indication of the inherent regional variability. This advantage is considerable in fitting the stand volume equation, because the residual patterns (see Figures 4.27 and 4.28) showed no biases for either the overall data or individual regions. Table 4.21 sets out the mean, minimum and maximum residual values for individual regions. These values show no serious bias for any region.

Figure 4.27: Plot of Residuals vs Predicted Values[m^3]
Stand Volume/ha Equation

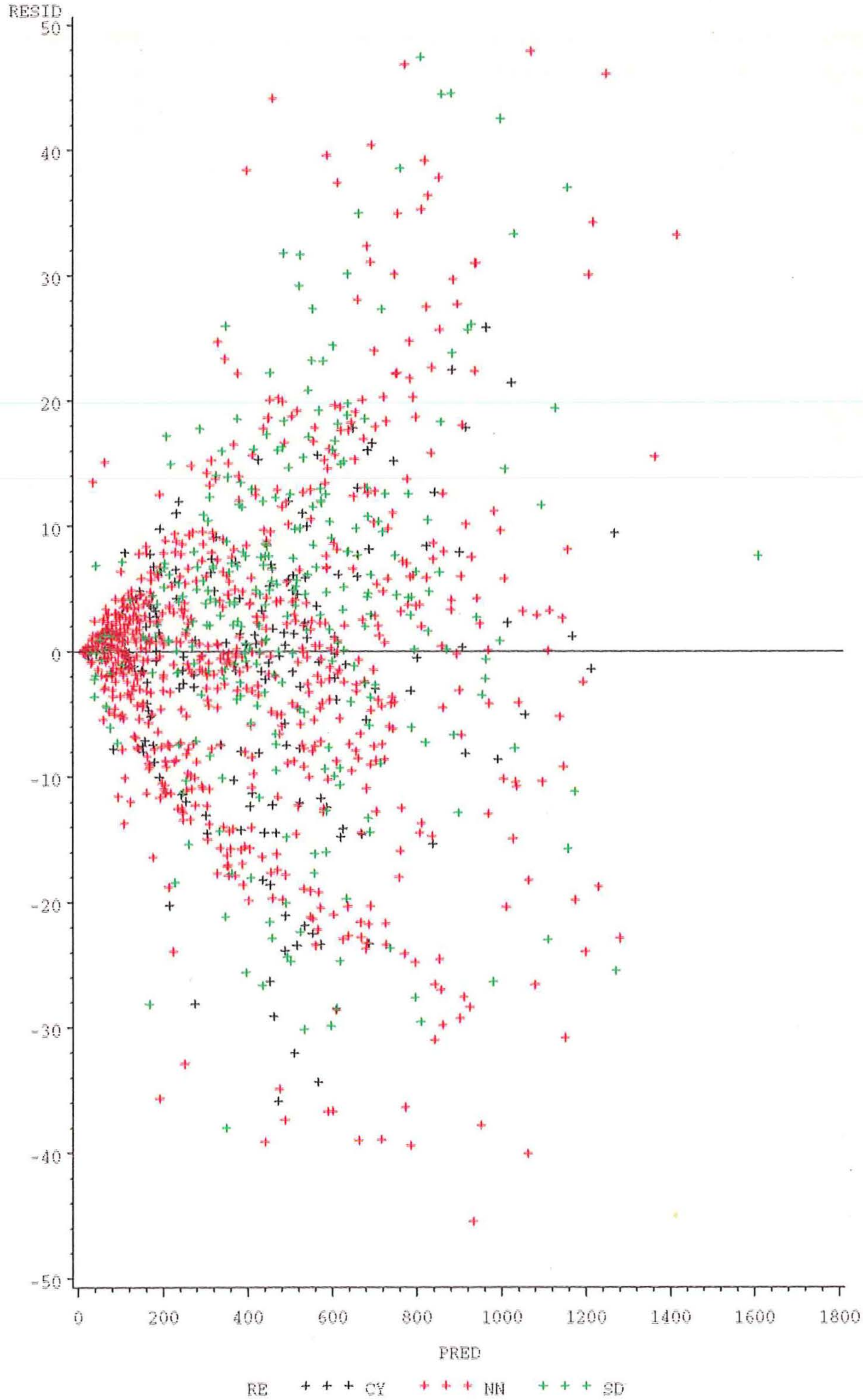
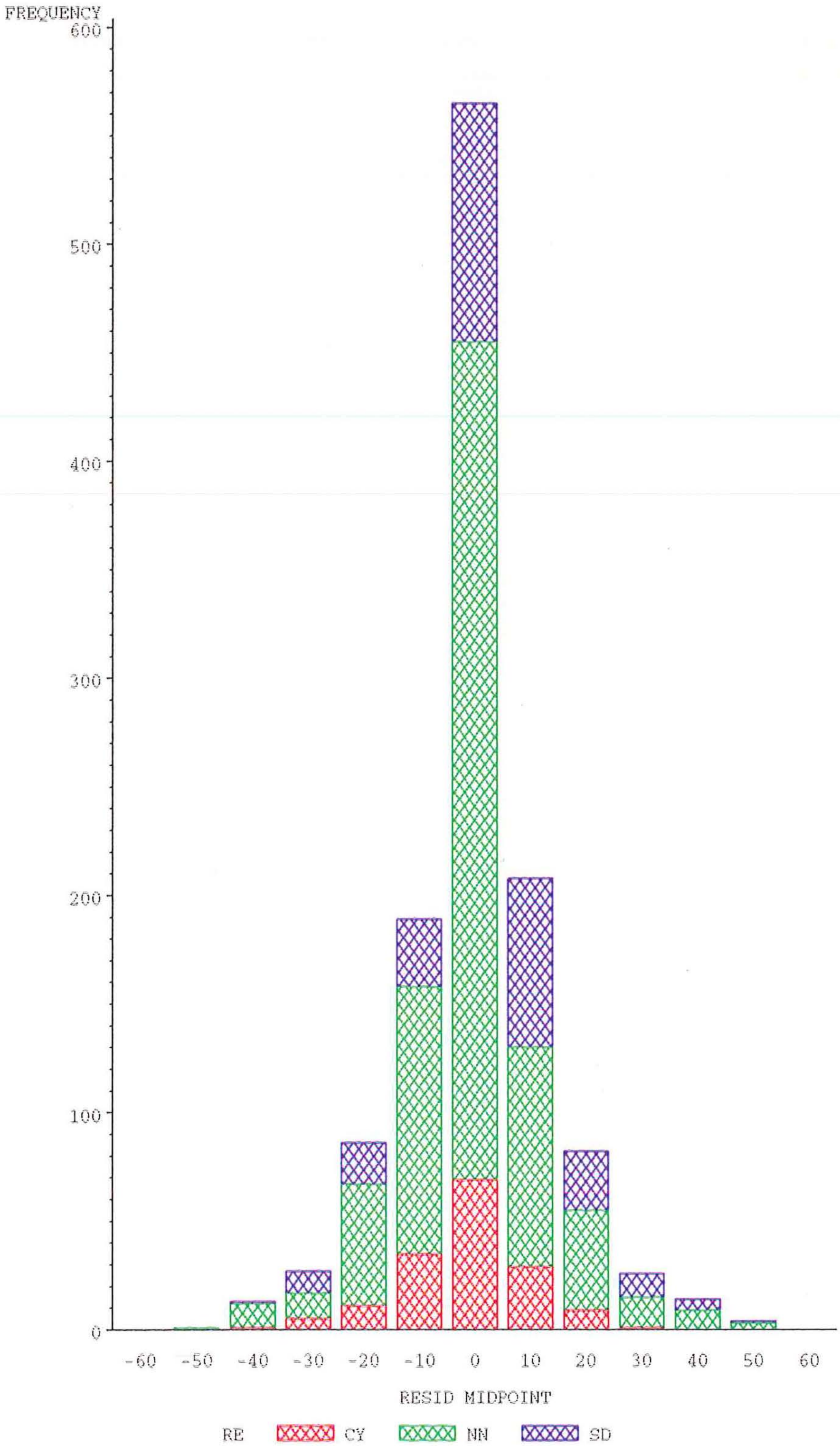


Figure 4.28: Frequency Distribution of Residuals [m^3 classes]
Stand Volume/ha Equation



4.3.2.6 Stem Survival/ha Equation

Screening of Mortality Data

Mortality data used in this analysis were carefully screened to ensure that only those intervals in which actual mortality occurred were included. Frequently, there is no mortality from one interval to the next and such individual data provide misleading estimates of the model. Moreover, excessive mortality due to windstorms were also discarded and so only measurements which reflected regular mortality were used to develop the stem survival/ha equation. These data are summarized in Table 4.22. A modified Gompertz function of the form presented in equation 4.54 was found to fit stocking survival best, with site index featuring as an explanatory environmental variable.

$$N_2 = N_1^{e^{(\beta_1(T_2^2-T_1^2))}} \times e^{(\alpha+\beta_2 S)(1-e^{(\beta_1(T_2^2-T_1^2)})} \quad (4.54)$$

Table 4.22: Data Used for Developing Stem Survival/ha Equation

Region	N	AGE(T)			Stems/ha		
		mean	min	max	mean	min	max
Canterbury	51	33.4	15.0	61.0	1141	74	2426
Nelson	217	28.2	8.0	53.5	1323	128	3517
Southland	96	30.5	7.7	63.0	1239	130	2822
Westland	63	28.0	5.0	59.1	991	222	3901
Total	427						

Goodness of Fit of the Stem Survival/ha Equation

Parameter estimates for stem survival/ha equation are set out in Table 4.23. Plots of residuals against predicted values, age of crop and site index, (Figures 4.29, 4.30 and 4.31), and the frequency distribution of residuals in Figure 4.32 of the mortality equation show no apparent bias, but asymptotic values were clearly related to site index: crops on higher site indices tended to suffer more mortality than those on lower ones. As always, the mortality function was the weakest link in the whole set of functions, but residuals for equation 4.54 did not exceed $\pm$ 100 stems/ha, while, without the site index term they were within $\pm$ 200 stems/ha. In this case the inclusion of site index was enough to sharpen the fit for all South Island sites considerably.

Table 4.23: Parameter Estimates and Standard Errors of Stem Survival/ha equation

Parameter	Estimate	SEE	ESS	N	MSE
α	7.504278342	0.33296588527			
β_1	-0.000148262	0.00001936157	504509.01	427	1189.88
β_2	-0.069398588	0.00081560793			

Table 4.24: Regional Biases in Predicting Stems/ha

Region	N	Residual		
		Mean	minimum	maximum
Canterbury	51	-6.9	-90.0	59.0
Nelson	217	4.4	-89.0	93.0
Southland	96	-1.6	-92.0	99
Westland	63	-9.6	-78.0	53.0
Overall bias	427	-0.34	-92.0	99.0

Table 4.24 sets out the regional biases in predicting stem survival/ha. Due to a small

number of observations in Canterbury, Southland, and Westland, these regions show a slight imbalance of residuals, but well balanced in the overall equation.

Figure 4.29: Plot of Residuals vs PredictedValues [N/ha]
Stem Survival/ha Equation

Figure 4.30: Plot of Residuals vs Time[T1 years]
Stem Survival/ha Equation

Figure 4.31: Plot of Residuals [N/ha] vs SiteIndex [m]
Stem Survival/ha Equation

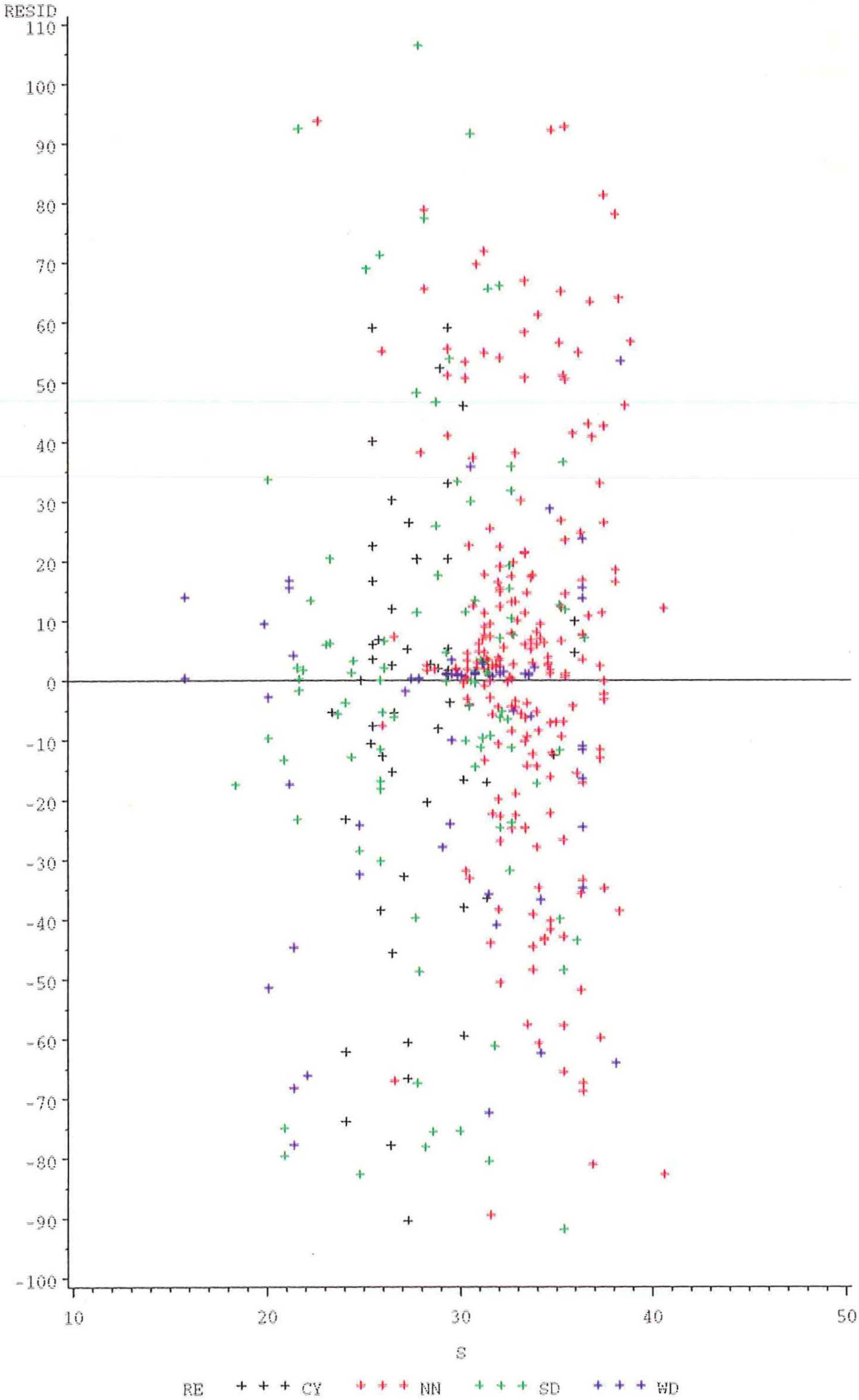
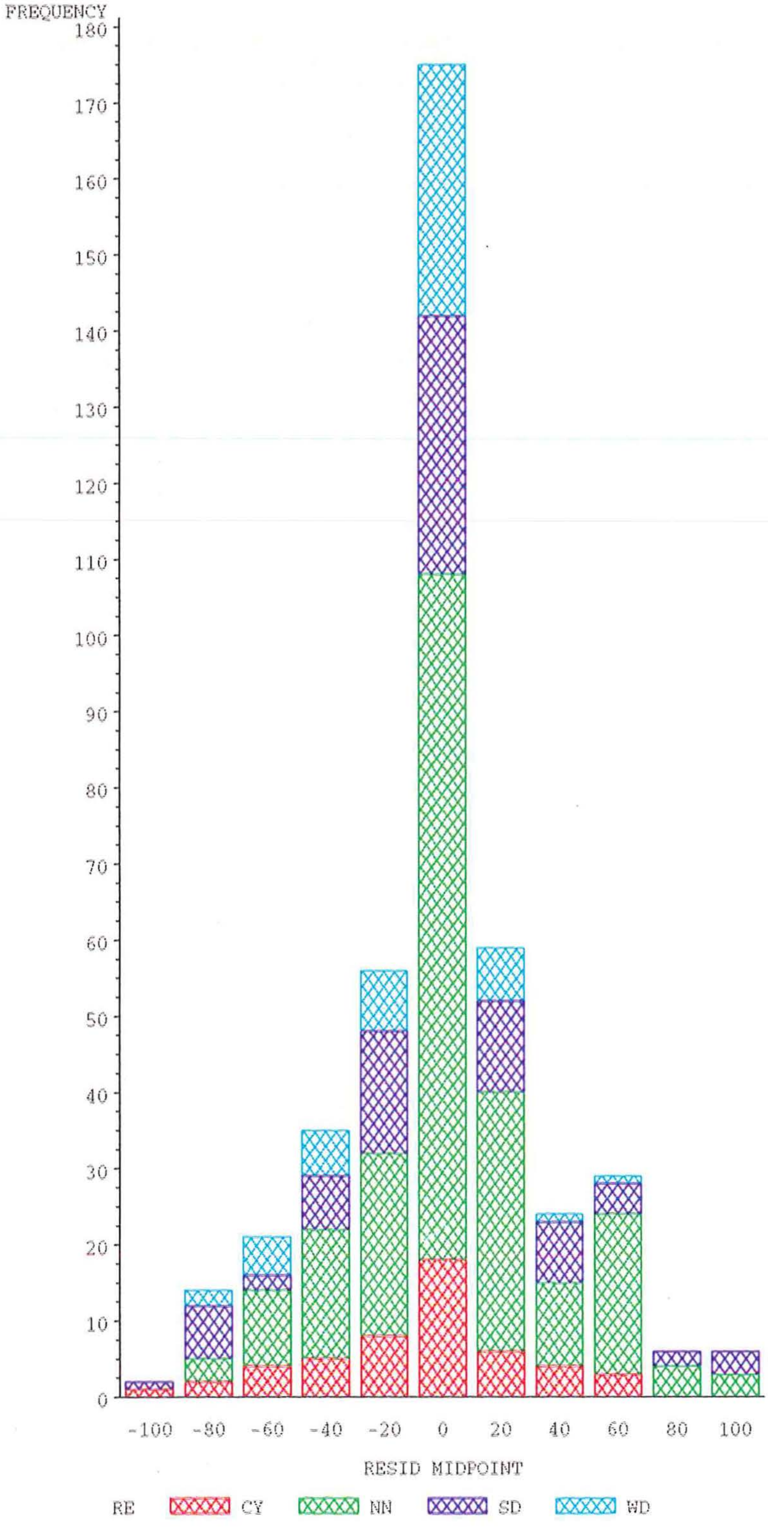


Figure 4.32: Frequency Distribution of Residuals [N/ha classes]
Stem Survival/ha Equation



Chapter 5

Verification and Validation

5.1 Verification of Equations Employed

The stem volume and taper equations and each of the stand level growth and yield model equations were verified in terms of their theoretical soundness and estimated reliability of prediction. This was accomplished through examination and analysis of (1) residual patterns, (2) frequency distribution of residuals and (3) asymptotic standard errors of coefficients. It was not possible to examine the success of applying them in everyday practice, as, unlike DFCNIGM and SIDFIR which have been routinely applied, there has not been any routine implementation for DfirTree and DfirStand. Nevertheless, by comparing outputs for DfirTree and DfirStand, with DFCNIGM and SIDFIR, some indications of their reliability in practice can be gauged.

(1) Residual Patterns

The resulting residuals of each equation were graphically plotted against predicted values and other variables included in the equation and some that were not: for example, the residuals for the mean top height and basal area projection equations were plotted against predicted values and the independent variables, age and altitude, to ascertain whether or not the trends over age and altitude were normally distributed about zero. Other variables not included in the equation such as distance from the sea were also employed and tests were made separately for each locality. The visual assessments about the goodness of fit of equations were reinforced with analysis and interpretation of univariate statistics about the distributions of residuals.

(2) Frequency Distribution of Residuals

Frequency distributions of residuals in the form of bar charts were also used to assess the degree of normality of their distribution.

Analyses of residuals in terms of both their patterns and frequency distributions were performed overall and also at disaggregated regional levels to find out whether or not there was conformance to normality in their distribution. The results of these tests showed no serious bias in any region. This ocular method of assessment can be misleading, however, because of the effect of choice of size class and so other statistics were used to ascertain the goodness of fit more objectively as set out below.

- **Asymptotic Standard Errors of Parameter Estimates**

Asymptotic standard errors of parameter estimates expressed as confidence limits are, of course, biased, as explained in chapter 3 section 3.3.3. But they do provide a measure of acceptability, provided that the existence of bias is recognized.

- **The Standard Deviation of Residuals**

The standard deviation of residuals provides a measure of dispersion of residuals about their mean. Random normal deviate (RND) computed through use of equation 5.1 was used to measure the degree of dispersion of each observation.

$$RND = [ABS(RESID)/RMS] \tag{5.1}$$

where RMS is the root mean square of residuals, and $ABS(RESID)$ is an absolute value of the residual for any one observation. All observations with RND of more than 3.5 were regarded as outliers and were subjected to further scrutiny.

- **Extreme Values of Residuals**

For each equation the extreme values of residuals (minimum and maximum) were assessed to check that their absolute values were in balance and that they were within practical limits for the variable. For example, an extreme residual value of 5 m for mean top height is too high and signals the presence of at least an outlier on the positive side: on the other hand a maximum of 1.5 m and a minimum of -5.0 m indicate an imbalance of residuals, with possible outliers on the negative side.

- **Mean and Mean Absolute Residual**

The mean of the residuals should equal zero, but in reality, estimated mean residuals of an equation are rarely exactly zero. They should, however, be close to it. The value of the mean absolute residual is more important than its sign, being a measure of average of the departure from the true values, but both assist in interpreting residual trends.

- **PROC UNIVARIATE Statistical Tests**

Finally, PROC UNIVARIATE statistical tests were used to ascertain the normality of distribution of residuals. For each equation, values for kurtosis, skewness, Kolomogorov - D statistic and the normal probability plot were carefully assessed to find out whether or not they were within practical limits for the normal distribution as set out for example in SAS Institute Inc., (1988). The skewness of the normal distribution is 0 by definition, while the value of kurtosis is supposed to lie between -2 and $+\infty$. For any one equation where the value of skewness was not close to 0, then the equation was re-assessed. Similarly, high values of kurtosis (5.0 and above) signals the presence of major outliers. Any one equation which had such a value of kurtosis was subjected to further rigorous examination. Table 5.1 below sets out the summary statistics of each equation in DfirTree, DfirStand, SIDFIR and DFCNIGM, Table 5.2 compares the statistical tests for the existing tree volume and taper equations with those of DfirTree.

Table 5.1: Statistical Comparisons for DfirTree, DfirStand, SIDFIR and DFCNIGM

Model	Equation	Residual Parameters					
		N	Average Mean	Absolute Mean	RMS	Min	Max
DfirTree	Stem volume (m^3)	597	0.000035	0.027	0.000278	-0.17	0.20
	Stem taper (cm)	6996	0.009	0.1268	0.18	-1.000	0.997
DfirStand	Top height (m)	1585	0.000676	0.4529	0.573	-1.4975	1.4838
	Basal area (m^2/ha)	1592	0.04135	0.9792	1.299	-3.9822	3.7566
	Stand volume (m^3/ha)	1215	0.1215	8.9799	0.054	-45.5	47.9
	Mortality (N/ha)	427	-0.32206	25.0	34.5	-95	99
SIDFIR	Top height (m)	1415	0.04	-	0.695	-	
	Basal Area (m^2/ha)	1415	-0.09	-	2.63	-	
	Stand volume (m^3/ha)	1415	-4.17	-	85.9	-	
	Mortality (N/ha)	1415	-1.14	-	57.4	-	
DFCNIGM	Top height (m)	1949	0.00874	0.42	-	-1.9	2.1
	Basal area (m^2/ha)	1746	0.02563	0.53	-	-2.37	2.5
	Stand volume (m^3/ha)	1649	-0.01136	6.74	-	-43.01	41.50
	Mortality (N/ha)	790	-2.76359	16.0	-	-117	121

The values for DfirStand presented in Table 5.1 were compared with similar values for DFCNIGM and SIDFIR and were found to be of the same order as DFCNIGM and a considerable improvement over SIDFIR. These values can be interpreted for each variable as follows.

(1) Stem Volume Equation

The average mean residual for the stem volume equation of 0.000035 m^3 and the mean absolute residual of 0.027 m^3 imply that there is a slight over-prediction with mean departure of 0.027 m^3 from the true values. The maximum expected error in the equation is 0.17 m^3 and the minimum error is -0.20 m^3. The new stem volume function has a value of 0.06% for the coefficient of variation compared to 8.9% in T136, which means that the new equation has a higher level of precision than volume Table T136.

Table 5.2: Comparison of Stem Volume and Taper Tables - D.fir New Zealand

	Table (Equation)	Locality	Standard error of the mean
Stem Volume Equations	DfirTree	South Island	0.06%
	T120	Ashley Forest	6.7%
	T136	All of N.Z	8.9%
	T228	Longwood	7.9%
Stem Taper Equations	DfirTree	All N.Z	0.0 cm
	F136	All N.Z	0.0 cm
	F228	Longwood	0.12 cm
Source: FRI, Ministry of Forestry N.Z. (1992)			

(2) **Stem Taper Equation**

With an average mean error of 0.009 cm and a mean absolute residual of 0.1268 cm the stem taper equation has a good fit, though with a slight over-prediction compared with the true values. The maximum error expected in the equation is 0.99 cm which is very well balanced with the minimum of -1.0 cm and the RMS of 0.18 cm.

(3) **Mean Top Height Equation**

The mean average residual, mean absolute residual and residual mean square error values for mean top height are 0.000676 m, 0.45 m and 0.329 m^2 respectively, indicating that the equation slightly over-predicts tree heights with an average deviation of 0.45 m. These values are similar to, but all slightly better than those for DFCNIGM and SIDFIR. The maximum error, moreover, expected from using this equation is 1.48 m, the minimum is -1.49 m while the maximum error in DFCNIGM is 2.1 m and minimum error is -1.9 m.

The RMS of this equation is 0.573 m while the corresponding value for SIDFIR is 0.695 m.

(4) **Net Basal Area/ha Equation**

The net basal area/ha equation has a mean average residual of 0.041 m^2/ha, implying a slight over-prediction, with an absolute mean departure of 0.97 m^2/ha from true values. The RMS value of this equation is 1.299 m^2/ha compared with 2.63 m^2/ha for SIDFIR. The equation has an expected maximum error of 3.75 m^2/ha and a minimum error of -3.98 m^2/ha, again well balanced,

(5) **Stand Volume Equation**

The residuals of the stand volume equation have a maximum value at 47.9 m^3/ha and a minimum value of -45.5 m^3/ha, corresponding values for DFCNIGM being 41.5 m^3/ha and -43.0 m^3/ha, respectively. The mean residual for DfirStand is 0.1215 m^3/ha with an average departure of 8.9 m^3/ha from the true values; corresponding values for DFCNIGM are -0.1126 m^3/ha, and 6.7 m^3/ha respectively. That is, the level of precision of the stand volume equation for DfirStand is about the same as that DFCNIGM. The RMS for DfirStand is 0.054 m^3/ha, which is much superior to that of SIDFIR, 85.9 m^3/ha.

(6) **Stem survival/ha Equation**

DfirStand shows an average mean of -0.32206 stems/ha in predicting stem survival/ha, indicative of a slight over-prediction, while the average deviation from the true values is 25 stems/ha. The corresponding values for DFCNIGM are -2.76 and 16, respectively, while SIDFIR has a mean stocking residual of -1.14 stems/ha. Thus, DFCNIGM predicts slightly more stems/ha than SIDFIR and DfirStand and DfirStand predicts slightly fewer stems/ha than SIDFIR. DfirStand is slightly better than SIDFIR and DFCNIGM, therefore. The maximum and minimum residuals for DfirStand are 99 and -95 stems/ha, DFCNIGM has a correspondingly wider range of extreme values with a maximum and minimum of 121 and -117 stems/ha. The RMS value for DfirStand is 34.5 stems/ha and 57.4 for SIDFIR.

The overall statistical comparison between DfirTree and the existing stem volume and taper equations shows that DfirTree to be slightly more precise. Similarly, DfirStand shows an overall improvement in the statistics over SIDFIR for all equations, particularly in stand basal area/ha and stand volume/ha equations; this is attributed to the inclusion of dummy variables in both DfirTree and DfirStand.

5.2 Validation

Validation of models is usually best done by an independent data set. Such data sets are, however, not usually available in most situations, as is the case here for Douglas fir in the South Island, because modellers are reluctant to siphon off basic measurements for just such a purpose. The overall data set here was large, but, according to the objectives of the study, the modelling was analysed in separate regions from which the data originated. Because of unequal distribution of these data, it was not possible to set aside a validation data set for each region without undermining the objectives of the study. Canterbury was represented by 13% of the stand modelling data and 60% of sectional measurement data. Nelson was represented by 50% of the stand modelling data, while it was represented by only 14% of sectional measurement data. Southland was represented by 24% of stand modelling data and 25% of sectional measurement data. Westland was represented by 12% of stand modelling data, while it was not included in the stem volume and taper analysis because of insufficient representation. Therefore, setting aside a validation data set from within the whole data set, would have resulted in too small numbers of sectional measurements for Nelson and Southland, and too few data for stand modelling for Westland and Southland. The whole data set was, therefore, used for developing and fitting the models, as well as their evaluation. As more permanent plot measurement data accumulate, an independent data set for validation can be set aside and validation tests can be conducted more objectively.

5.2.1 Validation of DfirTree

5.2.1.1 Stem Volume Equation

Stem volume prediction was validated in terms of percentage error prediction by diameter classes for individual regions. The volume equation used for validation is T136 (FRI, Ministry of Forestry N.Z., 1992) which was developed to cater for all Douglas fir grown in New Zealand. The new equation and that of T136 were run through the same data, and then their errors in stem volume prediction by diameter classes were compared. The formula used to predict this percentage error is equation 4.18 as defined in chapter 4 section 4.1.4. Table 5.3, shows the percentage volume prediction errors by dbh classes for Canterbury, Nelson and Southland. Table 5.3 shows that the overall biases for the new stem volume equations for Canterbury, Nelson and Southland are 0.24%, -0.26% and 0.51% compared with -5.7%, -6.1% and -8.7% in volume Table T136. However, the Canterbury and Southland ones under-predicts volume of trees in the very smallest, 5 cm dbh class, while T136 over-predicts. The general trend for both Table T136 and the new equations is over-prediction, but, bias tests by dbh classes show better results for DfirTree in comparison to Table T136.

Table 5.3: Bias in Stem Volume Prediction

Region	dbh Class (cm)	Mean height (m)	N	% Volume bias	
				DfirTree	T136
Canterbury	5	4.2	4	12.0	-18.3
	10	4.9	9	13.8	-6.7
	15	11.5	38	3.4	-4.9
	20	15.5	75	0.8	-5.6
	25	17.9	109	-2.4	-7.8
	30	19.4	69	0.4	-3.8
	35	19.2	17	-1.2	-4.1
	40	19.2	8	-5.0	-6.2
	45	27.6	5	0.78	-3.4
	50	28.9	7	-2.5	-1.5
	>50	34.1	16	-0.09	-3.5
	No. of trees		359		
	Mean bias			0.24	-5.7
Nelson	15	13.9	4	-3.7	-8.9
	20	19.8	14	4.2	-1.4
	25	23.3	17	-0.18	-6.4
	30	26.2	17	-1.1	-7.5
	35	25.1	7	-2.4	-6.7
	40	30.3	16	-1.3	-6.9
	>40	36.2	4	-2.1	-8.8
	No. of trees		75		
	Mean bias			-0.26	-6.1
Southland	5.0	3.0	11	17.2	-28.9
	10.0	3.8	9	20.2	-7.5
	15.0	6.4	13	-0.45	-13.3
	20.0	15.4	13	-2.1	-8.4
	25.0	17.8	27	-6.1	-11.3
	30.0	21.8	23	-1.3	-6.4
	35.0	23.1	25	-1.2	-5.6
	40.0	25.1	17	-3.1	-7.4
	45.0	25.0	11	1.7	-1.2
	50.0	23.3	6	0.92	-0.1
	>50.0	30.9	9	0.04	-2.9
	No. of trees		163		
	Mean bias			0.51	-8.7

5.2.1.2 Stem Taper Equation

The prediction of taper was validated for independent regions by z classes, where z is defined as the relative height from the top of the tree (see section 3.3, equation 3.2). Table 5.4 summarizes this error prediction for individual regions. The precision of the new equation was compared to Taper Table F136 (FRI, Ministry of Forestry N.Z, 1992) which is compatible with volume Table T136. The two models were run through the data, then their biases of diameter prediction were compared. The formula used to calculate the bias in taper prediction is equation 5.2.

$$E = \sum_{1}^{N}(d' - \hat{d}')/N \tag{5.2}$$

where E is mean bias of prediction in cm, d' is the observed diameter, $\hat{d}'$ is the predicted diameter and N is number of observations in the class. Table 5.4 shows that only those parts of the tree where z is less than 0.20 (i.e., near the top of the tree) is the taper prediction for Table F136 more precise than the new taper equation. For most parts of the tree, where z is greater than 0.20 and in the commercially realizable part of the stem, the new equation performs with much less bias than Table F136. An important feature is that, as z approaches 1.0 the precision of the new equation increases, while that of Table F136 declines. The overall performance of the new equation for each region shows an improvement of precision over Table F136.

Table 5.4: Validation of Stem Taper Prediction

Region	z Class	Mean Volume (m^3)	N	Bias (cm) DfirTree	Bias (cm) F136
Canterbury	0.10	0.9424	594	-0.67	- 0.07
	0.20	1.0992	266	-1.0	-0.14
	0.30	1.0020	292	-0.08	-0.42
	0.40	1.1099	270	0.45	-0.89
	0.50	0.9938	305	0.69	-0.97
	0.60	1.0430	249	0.53	-0.79
	0.70	1.0011	317	-0.32	-0.11
	0.80	0.9389	345	-0.16	0.21
	0.90	0.5600	820	-0.13	0.33
	1.00	1.1785	783	-0.08	-0.34
	N total		4241		
	Mean bias			-0.08	-0.34
Nelson	0.10	1.0042	125	-0.24	0.35
	0.20	1.0342	67	-0.99	-0.09
	0.30	0.9564	77	0.20	-0.24
	0.40	0.9844	76	0.53	-1.0
	0.50	0.9367	79	0.70	-1.2
	0.60	0.8995	78	0.68	-0.87
	0.70	0.9372	74	0.23	-0.41
	0.80	1.0780	79	-0.10	0.12
	0.90	0.6876	150	-0.15	-0.47
	1.00	1.0074	179	-0.42	-1.0
	N total		984		
	Mean bias			-0.02	-0.50
Southland	0.10	1.0061	264	-0.87	- 0.22
	0.20	1.2793	116	-2.0	-0.97
	0.30	1.0967	129	-0.97	-1.4
	0.40	1.2136	122	0.13	-1.4
	0.50	0.9638	130	0.29	-1.4
	0.60	1.1058	125	0.35	-1.1
	0.70	0.9724	148	0.22	-0.13
	0.80	1.3103	109	0.19	0.68
	0.90	0.7652	324	0.09	-0.11
	1.00	1.3999	304	0.49	-0.46
	N total		1771		
	Mean bias			-0.15	-0.38

5.2.2 Validation of DfirStand

Growth and yield predictions by three different regimes have been compared by Law (1990) for SIDFIR and DFCNIGM. These regimes are:

(1) initial stocking 1600/ha, unthinned;

(2) initial stocking 1600/ha, thinned at 10 m to 400 stems/ha;

(3) initial stocking 1600/ha, thinned at 10 m to 600 stems/ha, thinned at 20 m to 250 stems/ha.

Table 5.5: Comparison Between SIDFIR DFCNIGM and DfirStand

Regime	model	Top height	Stocking	basal area	Mean dbh	Volume
1	SIDFIR	33.3	1023	93.5	34.1	1117
	DFCNIGM	32.5	1176	75.1	28.5	868
	DfirStand (CY)	33.1	1005	78.2	31.5	876
	DfirStand (NN)	33.0	1005	80.0	31.8	895
	DfirStand (SD)	33.2	1005	100.9	35.8	1124
	DfirStand (WD)	33.2	1005	104.6	36.4	1664
2	SIDFIR	33.3	354	72.0	50.9	859
	DFCNIGM	32.5	379	70.2	48.6	816
	DfirStand (CY)	33.1	350	65.0	48.6	731
	DfirStand (NN)	33.0	350	66.2	49.1	744
	DfirStand (SD)	33.2	350	80.1	54.0	897
	DfirStand (WD)	33.2	350	82.5	54.8	923
3	SIDFIR	33.3	243	55.2	53.9	659
	DFCNIGM	32.5	242	56.4	54.5	656
	DfirStand (CY)	33.1	245	55.3	53.6	624
	DfirStand (NN)	33.0	245	56.0	54.0	632
	DfirStand (SD)	33.2	245	63.5	57.5	714
	DfirStand (WD)	33.2	245	64.7	58.0	728

The simulations were done at site index 30.0 m grown from age 0 to 45 years. Because DfirStand depends on altitude for its prediction, the mean elevation for each region was used (Canterbury 326 m, Nelson 438 m, Southland 251 m, Westland 229 m). Table 5.5 sets out the outputs of SIDFIR, DFCNIGM and DfirStand for the chosen regimes. The symbols CY , NN , SD , and WD refer to Canterbury, Nelson, Southland and Westland respectively. It can be deduced from Table 5.5 that, DfirStand predicts virtually the same mean top height at age 45 years as DFCNIGM, and SIDFIR for site index 30 m. The predicted basal areas/ha for SIDFIR and DFCNIGM are average values, DfirStand separates estimates for predicted basal area/ha for each region through use of dummy variables for locality and through altitude. Thus, at age 45 years Westland has the highest production of basal area/ha, followed by Southland, then Nelson and Canterbury the least. Stand volume/ha production and mean stand diameter of the stand follow a similar pattern. The predicted stem survival/ha for DFCNIGM for unthinned regime 1 is higher than for SIDFIR and DfirStand, but is very similar for all models for regimes 2 and 3. Given initial condition of the stand, DfirStand will predict approximately equal values of mean top height and stem survival/ha as for SIDFIR and DFCNIGM. Essentially, the three models differ in basal area/ha production brought about by dummy variables, similarly, values derived from it such as volume/ha and mean dbhob follow the same pattern.

5.2.3 Desirable Properties of the Employed Equations

5.2.3.1 Stem Volume Equation

The stem volume equation implicitly has inputs for region through dummy variables Z_1 and Z_2, dbh (d), and height (h), and outputs stem volume by region.

$$v = \alpha d^{(\beta_1+\beta_2 Z_1+\beta_3 Z_2)} h^{\gamma}$$

Biologically, this is appealing because usually trees growing at different sites or regions may need to be classified by their own stem volume equations. The use of dummy variables eliminates this need and it ensures that biases due to locality on parameter estimates are minimized, even if the regional differences in stem volume outputs by the same inputs are small.

5.2.3.2 Stem Taper Equation

The biological aspects of stem form were not considered in this thesis, but, even without such consideration, it has been possible to define adequately the stem taper in a mathematical equation. The segmented taper function which was used here divides the stem into three segments, then assigns an appropriate function which describes the stem form in that section.

5.2.3.3 Mean Top Height Equation

The mean top height equation incorporates altitude (AL) in metres above sea level.

$$h_{(100,2)} = h_{(100,1)} e^{(\alpha+\beta\times AL)(\frac{1}{T_1^\gamma}-\frac{1}{T_2^\gamma})}$$

It was found that the mean top height for a given age decreased with altitude above sea level, a biologically realistic trend. As T_2 approaches infinity, $h_{(100,2)}$ approaches the upper asymptote, $(\beta_1+\beta_2 AL)$, when T_1 equals T_2, then $h_{(100,1)}$ equals $h_{(100,2)}$ *(consistency property)*, and the projection from T_1 to T_3 yields the same result as the projection from T_1 to T_2 followed by projection from T_2 to T_3, *(path invariance property)*. This equation predicts the same mean top height as the ones for SIDFIR and DFCNIGM, but is slightly more precise, as altitude represents variations due to locality better.

5.2.3.4 Net Basal Area/ha Equation

Unlike SIDFIR and DFCNIGM, DfirStand predicts different values of G/ha for each region. Consequently, it also predicts different values for V/ha and quadratic mean stand diameter for a given set of inputs.

$$G_2 = G_1^{(\frac{T_1}{T_2})^{\beta_1}} e^{((\alpha+\beta_2 AL+K_1\beta_3+K_2\beta_4+K_3\beta_5)(1-(\frac{T_1}{T_2})^{\beta_1})+\beta_6 X_t(\frac{1}{T_2}-\frac{1}{T_1})\times\frac{T_t}{T_2})}$$

When site index was included in this equation instead of altitude, the trend was a decline of net basal area/ha with increase in site index, a biologically unreasonable pattern. Altitude was found to explain the variability in site quality better, there being a decrease in the level of the asymptote $[(\alpha + \beta_2 AL + K_1\beta_3 + K_2\beta_4 + K_3\beta_5)]$ with increasing altitude.

This is more realistic and so altitude was the variable retained. The asymptote also varied from one region to another, the variations being induced through regional dummy variable coefficients. Other desirable properties include, *consistency*; thus, G_2=G_1 when $T_1 = T_2$ and *path invariance*, that is, the projection from T_1 to T_3 gives the same result as the projection from T_1 to T_2 followed by the projection from T_2 to T_3.

5.2.3.5 Stand Volume/ha Equation

$$V = \alpha G^{\beta} h_{100}^{\gamma}$$

There was no need to disaggregate the stand volume/ha equation regionally. Differences from region to region are induced through separate effects in the stand net basal area/ha equation, the stem volume equation, and impact of altitude in height. The stand volume equation transforms the state variables, basal area/ha and mean top height, into volume/ha at specified crop ages.

5.2.3.6 Stem Survival/ha Equation

DfirStand predicts basically the same stocking over-time as do SIDFIR and DFCNIGM. Although site index proved not significant in predicting net basal area/ha, its inclusion did improve the prediction of stocking.

$$N_2 = N_1^{e^{(\beta_1(T_2^2 - T_1^2))}} \times e^{(\alpha+\beta_2 S)(1-e^{(\beta_1(T_2^2-T_1^2)})}$$

The equation has a lower asymptote $(\alpha+\beta_2 S)$ which is controlled by the value of site index (S). The coefficients indicate that, the higher the site index the lower the level of the asymptote. The equation is *consistent*, in that $N_1 = N_2$ when $T_1 = T_2$, and it is *path invariant*, that is the projection from T_1 to T_3 is the same as the projection from T_1 to T_2 followed by the projection from T_2 to T_3.

Table 5.6: Summary Output - DfirStand

Region	Age years	Top Height m	Basal area/ha m^2/ha	Stocking stems/ha	Volume/ha m^3/ha
Canterbury Mean altitude 326 m.	20	15.2	20.0	1000	113
	25	19.2	31.2	952	216
	30	23.0	41.8	890	340
	35	26.6	51.4	843	475
	40	30.0	59.9	785	616
	45	33.2	67.9	727	765
	50	36.2	74.5	671	908
	55	39.0	80.3	619	1049
	60	41.7	85.5	570	1187
Nelson Mean altitude 438 m.	20	15.5	20.0	1000	113
	25	19.5	31.5	952	219
	30	23.2	42.5	890	346
	35	26.7	52.5	843	486
	40	30.0	61.5	785	632
	45	33.1	69.5	727	781
	50	36.0	76.4	671	930
	55	38.7	82.6	619	1078
	60	41.3	88.2	570	1223
Southland Mean altitude 229 m.	20	15.1	20.0	1000	113
	25	19.2	33.1	952	230
	30	23.0	46.1	890	374
	35	26.6	58.2	843	537
	40	30.0	69.2	785	709
	45	33.2	79.1	727	888
	50	36.3	87.9	671	1066
	55	39.2	95.8	619	1246
	60	41.7	102.9	570	1422

Table 5.6 summarizes simulation results of DfirStand with initial starting conditions: site index of 30 m, basal area of 20 m^2/ha, 1000 trees/ha, all at age 20 and a mean regional altitude. Figures 5.1 and 5.2 are graphical representations of net basal area/ha and volume/ha development in Canterbury Nelson and Southland.

5.2.4 Limitations to Applicability of the Models

(1) **Stem Volume and Taper Model**

The dbh range in the data set used for modelling stem volume and taper was between 5 and 95 cm. The equations should be applied to trees of dbh outside this range only with extreme caution. Trees smaller than 12 cm in dbh are also not predicted precisely in relative terms. The equations reflect their applicability to Canterbury, Nelson and Southland separately.

(2) **Stand Model**

The stand model data ranged in age from 5 to 78 years, with firm confirmation of trends older than 60 years. Separation of growth trends in the Canterbury, Nelson and Southland regions is very distinct.

Figure 5.1: Net Basal Area/ha Production [m^2]
Nelson, Canterbury, and Southland

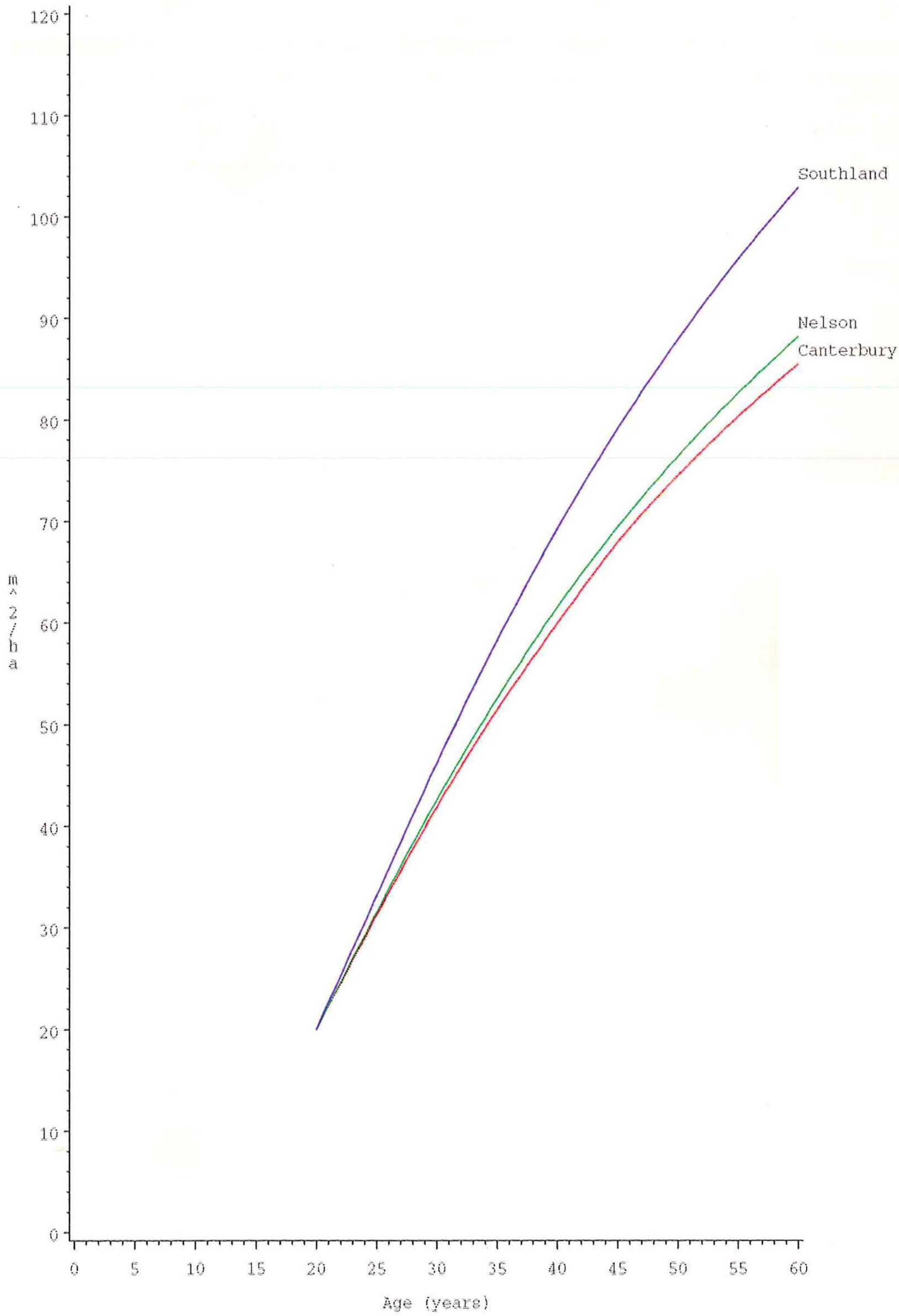
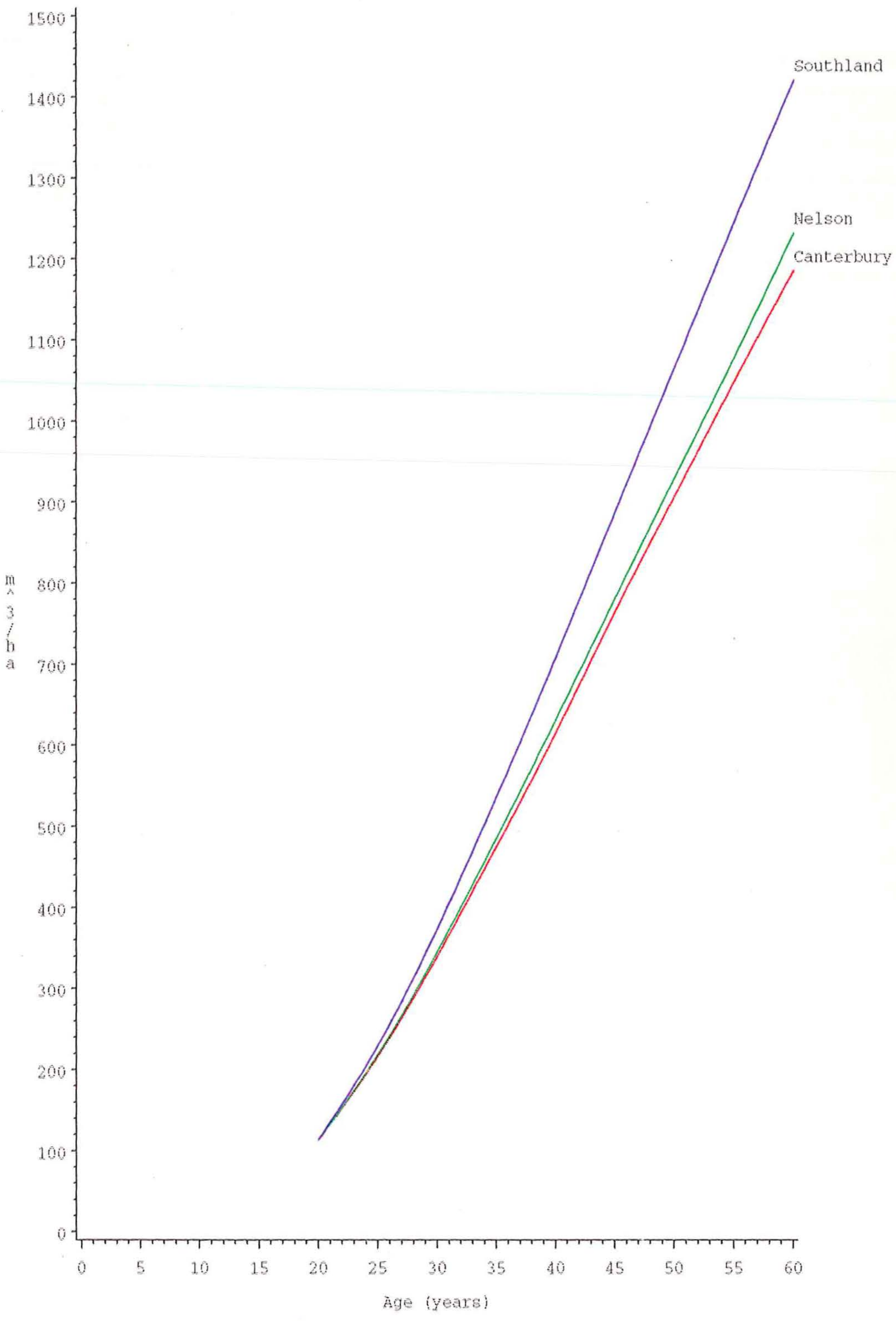


Figure 5.2: Net Volume/ha Production [m^3 ib]
Nelson, Canterbury, and Southland



Chapter 6

Discussion

This study has produced (1) a tree stem volume and taper model and (2) a stand growth and yield model for Douglas fir grown throughout the South Island of New Zealand. DfirTree is a tree based model representing one equation for predicting stem volumes for trees grown in Canterbury, Nelson and Southland regions separately, but in one equation, and another which is a volume based compatible taper equation. DfirStand is a whole stand growth and yield model consisting of equations for predicting mean top height or site index, net basal area/ha, volume/ha, and stem survival/ha for trees grown in Canterbury, Nelson, Southland and Westland regions separately. All equations show a precision that is at the same level as, or better than similar existing models. Comparison of predictions indicated also that they can be put into use by managers to produce as reliable or better yield forecasts than those that have previously existed.

6.1 Computational Methods - Non Linear Regression

In fitting non linear regression equations using the SAS package (SAS Institute Inc. 1988), four iterative methods (algorithms) were available for computing the parameter estimates. These methods are

(1) Gauss - Newton (Hartley, 1961).

(2) Marquardt (Marquardt, 1963).

(3) Gradient (Bard, 1970).

(4) DUD (Ralston and Jennrich, (1971)

Since non linear least - squares estimation is an iterative procedure, it is necessary to provide the regression functional form, declare parameter names and their initial approximate values, and possibly specify the derivatives of the model with respect to the parameters. Methods 1, 2 and 3 require the user to supply the partial derivatives while method four determines the derivatives during iterations by default. There is no one method superior to the other, but, the success in fitting an equation depends on (i) providing satisfactory initial values of parameter estimates, (ii) program statements to direct the computation (e.g. restrictions, grid search techniques, bounds, *et cetera*), (iii) nature of the equation being modelled and (iv) the skill and experience of the researcher in the field of modelling. In many instances, the initial estimates were obtained by grid search techniques, where several starting values were imposed and SAS evaluates their error sum of squares and begins iterations with those values that conferred the minimum error sum of squares. Finally, if this technique was not practical an *intelligent guess* of starting values was executed.

There was no one algorithm that was superior to the other, experience with this modelling has shown that the efficiency of convergence depended largely upon the adequacy of the starting values. Nevertheless, parameter estimates did not differ significantly through using either algorithm, but, the ones in "DUD" were found to converge better than in the other methods. Therefore, all equations adopted in this thesis were standardized in "DUD", as this offered not only the minimum error sum of squares compared to other methods, but also consistency in parameter estimates for both the stem volume and taper and the stand growth equations.

6.2 New Features of the Study

Equations employed by DfirTree and DfirStand have shown an acceptable level of precision, but, unlike other regional models they do not give overall average estimates. In contrast to the approach adopted previously in SIDFIR (Law, 1990), which predicts overall average values for the whole of South Island of New Zealand, DfirTree and DfirStand endeavour to provide flexibility for aggregation and disaggregation by employing dummy and local site variables, along with other predictor variables to improve the sensitivity of estimation. The subsections below summarizes some of the major new features which have arisen from this study.

6.2.1 Stem Volume Equation

6.2.1.1 Use of Dummy Variables

The stem volume equation employs Schumacher's tree volume model, but the practicality of disaggregating regional adjustments through use of dummy variables on the inputs to provide distinct stem volume estimates by region is amply demonstrated. Comparison with a similar volume table, namely T136 which is the one currently being used has shown that DfirTree can provide more precise volume estimates. The overall regional biases by diameter classes for DfirTree were lower than those of Table T136 (refer to chapter 5 section 5.2.1.1 Table 5.4).

6.2.1.2 Use of Weighted Non linear-least Squares Regression

Existing volume tables in New Zealand are either logarithmic or arithmetic. This study has provided an alternative modelling approach by producing similar equations through weighted non linear-least squares regression. The improvement in stem volume biases in absolute terms over Table T136 were 96% in Canterbury and Nelson and 94% in Southland. The improvements can be attributed to: (1) the practicality of including dummy variables, even though they do not confer much change in the outputs of stem volumes regionally, but balance the residuals in the equation instead; (2) weighted non linear least - squares regression, which also corrected the heterogeneity of the variance in stem volume shown to be a major form of improvement.

6.2.2 Stem Taper Equation

6.2.2.1 Use of Polynomial Segmented Taper Equation

(a) Volume - based compatible taper equation

The use of a segmented volume - based taper function achieved an overall higher precision of diameter inside bark estimation than the existing taper Table T136. Bias tests between Table F136 and the new equation showed that, the new equation over-predicted inside bark diameters for the 20% top portion of the tree stem. The predictions for this part of the tree were poorer than those of T136. But, DfirTree predicted inside bark diameters for the bottom 80% of the stem with higher precision than Table F136 did. This is of great advantage since the bottom portion of the stem is the most utilized and commercially valuable part of the tree. The segmented taper function showed an advantage in describing the tree shape by allowing changes in basic form to an appropriate equation, yet allowing for a continuous and smooth transition across the join points. The overall improvement in regional biases in absolute terms were 76% in Canterbury, 96% in Nelson and 60% in Southland.

(b) Taper - Based Taper Equation

As outlined earlier, it was not possible to impose all compatibility restrictions on the parameters of this equation as in the counterpart volume-based system. An exactly compatible model could not be obtained without at least one of the parameter estimates not being significantly different from zero. Partial restriction of the parameters (as explained in chapter 4 section 4.2.3) resulted in approximately 98% compatibility. This equation could in future be tried in another set of data or species to see if full compatibility could be achieved.

6.2.2.2 Utility of New Taper Function

The stem taper equations for Douglas fir in New Zealand have utilized a full polynomial approach as shown in equation 6.1.

$$\frac{d'^2 K h}{v} = \beta_1 z + \beta_2 z^2 + ... + \beta_n z^n \tag{6.1}$$

The equations (F136, F226 and F228) have produced good results, but over-prediction of diameters in the top 20% of large trees has also been reported (Goulding and Murray, 1976). These equations can be integrated then rearranged as shown in 6.5 to predict sectional volume (v_{sm}) along the stem at two given height limits. Merchantable volume for section 1, v_{m1} , can be obtained by substituting the value of z in the equation as $(h - h'/h)$ or (l/h), as defined earlier, even without the need for top diameters (d') as demonstrated here below. Volume for section 1, (v_{m1}) , is predicted by

$$v_{m1} = v[\frac{\beta_1}{2} z_1^2 + \frac{\beta_2}{3} z_1^3 + ... + \frac{\beta_n}{(n+1)} z_1^{(n+1)}] \tag{6.2}$$

and for section 2, (v_{m2}), as

$$v_{m2} = v[\frac{\beta_1}{2} z_2^2 + \frac{\beta_2}{3} z_2^3 + ... + \frac{\beta_n}{(n+1)} z_2^{(n+1)}] \tag{6.3}$$

Thus, the volume between the two sections above, (v_{m1} and v_{m2}), is obtained subtracting 6.2 from 6.3, assuming that z_2 is larger than z_1.

$$v_{sm} = v_{m2} - v_{m1} \tag{6.4}$$

which can be evaluated as equation 6.5.

$$v_{sm} = v[\frac{\beta_1}{3}(z_2^2 - z_1^2) + \frac{\beta_2}{3}(z_2^3 - z_1^3) + ... + \frac{\beta_n}{(n+1)}(z_2^{(n+1)} - z_1^{(n+1)})] \quad (6.5)$$

Due to the nature of equation 6.1, it is not possible to solve and predict sectional volumes at two given top diameter limits d'_1 , and d'_2.

DfirTree adapts a quadratic segmented polynomial taper equation 6.6 (all variables and coefficients as defined before).

$$\frac{d'^2Kh}{v} = \beta_1 z^2 + \beta_2 z + d\beta_3(z - \alpha_1)^2 I_1 + d\beta_4(z - \alpha_2)^2 I_2 \quad (6.6)$$

This equation can also be integrated and rearranged to obtain an expression for sectional volume between two height limits in a similar method as in full polynomial taper equations. Volume for section 1, (v_{m1}), is predicted by

$$v_{m1} = v[\frac{\beta_1}{3}z_1^3 + \frac{\beta_2}{2}z_1^2 + (d\frac{\beta_3}{3})(z_1 - \alpha_1)^3 I_1 + (\frac{d\beta_4}{3})(z_1 - \alpha_2)^3 I_2] \quad (6.7)$$

and volume for section 2, (v_{m2}), as

$$v_{m2} = v[\frac{\beta_1}{3}z_2^3 + \frac{\beta_2}{2}z_2^2 + (\frac{d\beta_3}{3})(z_2 - \alpha_1)^3 I_1 + (\frac{d\beta_4}{3})(z_2 - \alpha_2)^3 I_2] \quad (6.8)$$

Thus, the volume of the section between z_1 and z_2, v_{sm}, is obtained by applying equation 6.4, which evaluates to 6.9

$$\begin{aligned} v_{sm} = v[(\frac{\beta_1}{3})(z_2^3 - z_1^3) + (\frac{\beta_2}{2})(z_2^2 - z_1^2) + (\frac{d\beta_3}{3})((z_2 - \alpha_1)^3 I_1 \\ -(z_1 - \alpha_1)^3 I_1) + (\frac{d\beta_4}{3})((z_2 - \alpha_2)^3 I_2 - (z_1 - \alpha_2)^3 I_2))] \quad (6.9) \end{aligned}$$

DfirTree also produces an added capability, being able to derive and solve for an expression for computing sectional volume at two given diameter limits more easily than has been possible previously. McClure and Czaplewsk (1986) have applied the same technique but theirs involved 5 parameters while here it is demonstrated for six parameters. Given two top diameters, d'_1 and d'_2, z_1 and z_2 can be solved by applying the general solution for quadratic equations as illustrated below.

$$z = \frac{(-B \pm \sqrt{(B^2 - 4AC)})}{2A} \tag{6.10}$$

Only the non negative values are considered in the solution of the above equation (0 <= z <= 1.0), and the coefficients can be equated such that

$$A = \beta_1 + d(\beta_3 I_1 + \beta_4 I_2) \tag{6.11}$$

$$B = \beta_2 - 2d(\alpha_1 \beta_3 I_1 + \alpha_2 \beta_4 I_2) \tag{6.12}$$

$$C = d(\alpha_1^2 \beta_3 I_1 + \alpha_2^2 \beta_4 I_2) - \frac{d'^2 Kh}{v} \tag{6.13}$$

After the solution for z has been found, the procedure for solving for sectional volume between two given top diameters is the same as that shown in equation 6.9. The methodology is applicable to both taper based and volume based taper functions.

Determination of sectional volume between two diameter limits is often necessary in allocating wood by assortments, rather than having merchantability limits according to height classes, DfirTree provides a flexibility which is more appealing because the assortments can be set according to top diameter limits; for example, peeler logs, saw logs and pulp wood, are more meaningful if top diameters are also specified. This capability is not feasible in higher orders polynomial taper equations of the form shown in 6.1 because of the high powers that need to be solved.

6.2.3 Net Basal Area Projection Equation

6.2.3.1 Use of Dummy Variables

Before development of DfirStand, the tendency was to develop models for small populations or at least models that give average yield values for basal area/ha and volume/ha, for an entire region. DfirStand provides an alternative approach, by using the pooled data while at the same time providing the capability of estimating yields for specific regions and even for a locality. The approach adopted has used dummy variables to predict the asymptote of basal area/ha for regions (K_1, K_2 and K_3 for Nelson, Southland and Westland respectively), while altitude contributed to levelling the asymptotes for different localities within regions. The advantages that accrue by using this approach are set out below.

(1) Greater local sensitivity is provided.

(2) The methodology could well be extended to cater for further stratification and more refined crop typing.

(3) A better understanding of the underlying nature of growth projection and yield forecasts is obtained.

This approach has recognised that there are existing differences in basal area/ha production in Canterbury, Nelson, Southland and Westland regions.

6.2.3.2 Substitution of Altitude for Site Index

This study has shown that site index as a variable is not a good indicator of basal area production for Douglas fir grown in South Island. It has been shown in this study that, altitude can substitute for site index, and that increasing altitude above sea level in metres has a negative effect on basal area/ha production, which is biologically sensible. Site index, however, did not contribute positively but suggested that higher site indices have a lower basal area/ha as asymptote, not an easily reconcilable result. On the other hand, stem survival/ha, could not be related readily to altitude directly, but a higher site index provided a statistically significant lower level of asymptote.

6.2.3.3 Reference to Thinning Age

The basal area projection equation includes age of thinning as a component of the thinning index; thus, basal area/ha development depends on the kind of thinning and the time elapsed since the last thinning. Because the equation is also path invariant, it provides managers with a tool to simulate yield according to different thinning regimes and ages of thinning more realistically.

6.3 Possible Refinements to DfirTree and DfirStand

6.3.1 DfirTree

(1) Further improvement in the stem volume equation may be needed for small dbh classes (<15 cm), if predictions for these small size classes are considered important.

(2) Not enough information about the silvicultural history of trees sampled for stem volume and taper analysis (e.g pruning, disease and thinning) was available for analysis. Such information could be vital in reviewing and refining the stem volume and taper equations through, for example, inclusion of additional variables such as age, thinning history and other such variables.

(3) Due to the small number of sectional measurements in Nelson, the stem taper and volume equations for that region may need to be revised further as more sectional measurements become available.

6.3.2 DfirStand

(1) The differences in basal area/ha and yield in Canterbury, Nelson Southland and Westland need to be further evaluated in terms of diseases, soil factors, frost burns and other climatic factors, data which again was lacking for the analysis.

(2) As site index proved to be of less relevance in yield prediction in all regions, it would appear that alternatives such as altitude and similar site variables should be further investigated. The potential is apparent from this study and, continuance could confer a greater understanding of how basal area growth is influenced.

(3) Further data from Westland are needed, especially to revise the net basal area/ha equation, if forecasts for that region are to be as reliable as elsewhere in the South Island.

(4) The success of applying a thinning index was small, probably because it is less precisely determined than the other predictor variables. Numerous researchers have

found the same, with whatever index they have defined. In future more research is needed in this area.

6.4 Recommendations

6.4.1 Applicability of Equations

The stem volume and taper equations apply to Douglas fir grown in Canterbury, Nelson and Southland (see appendix A.1), but other regional breakdowns could also be examined. The stand volume/ha equations apply to Canterbury, Nelson and Southland but they too could be further disaggregated. The age range of 5 to 78 years is very wide, but could be further be strengthened for Westland. Users wishing to employ the models to make production forecasts for Douglas fir plantations should collect local site and crop production data to allow further independent verification.

Chapter 7

Summary and Conclusions

This study has shown that it is possible to derive precise tree volume equations, stem taper equations and stand growth functions for Pseudotsuga menziesii *(Mirbel, Franco)* that are representative of crops located throughout the South Island of New Zealand. The equations are not overall averages, but ones which can be disaggregated through using *dummy* and other variables to reflect locality variation. Using standard IUFRO notation and the definitions adopted earlier, the best fitting equations selected were as summarized in the following sub - sections.

7.1 Stem Volume and Taper Equations (DfirTree)

7.1.1 Stem Volume Equation

A modified Schumacher tree volume equation fitted the tree volume data best.

$$v = \alpha d^{(\beta_1 + Z_1\beta_2 + Z_2\beta_3)} h^{\gamma} \tag{7.1}$$

The stem volume equation contained two dummy variables, one for Nelson and another one for Southland while Canterbury was represented by default. Heterogeneity of the variance of stem volume was counteracted by a weight, w, shown in equation 7.2.

$$w = \frac{1}{d^2 h} \tag{7.2}$$

The equation has a high precision , while the minimum and maximum residuals range from -0.17 and 0.20 m^3 respectively. The weighting reduced the total sum of squares by about 45%, after which a further 5% reduction in ESS was realized through the inclusion of dummy variables, but, a greater benefit was a better balance in the distribution of residuals by region when compared to a similar equation without dummy variables.

7.1.2 Stem Taper Equations

Stem taper was found to be best described with a segmented polynomial of order 2 with two join points.

$$\frac{d'^2 K h}{v} = \beta_1 z^2 + \beta_2 z + d\beta_3 (z - \alpha_1)^2 I_1 + d\beta_4 (z - \alpha_2)^2 I_2 \tag{7.3}$$

This algebraically compatible (volume based) segmented taper equation combined well with the Schumacher's tree volume equation 7.1. No further locality representation in addition to the tree volume characterization was needed. The equation has a standard deviation of of ± 0.18 cm for the predicted diameter inside bark. The maximum and minimum biases in predicting mean diameter inside bark for all the data are 0.99 and -1.0 cm respectively. Compared to the segmented taper equation devised by Cao et al. (1980) which consists of five parameters, the equation form used here consisting of six parameters, provided a greater flexibility to constrain the parameters to obtain compatible volume estimates. This equation also incorporated at the join points, diameter at breast height over bark, d (as was adopted by Candy, 1989) which improved the fit by controlling predictions of the butt swell. The necessity for an order three polynomial segmented model, even for small trees with little or no butt swell for which a two segment model without butt swell might be considered adequate, was clearly apparent when d was excluded as a variable.

The full polynomial taper equation suggested by Goulding and Murray (1976), has sometimes led to inclusion of high powers of a predicting variable (e.g. polynomials of order 31 in Katz and Dunnigham, 1981 and order 41 in Gordon, 1983) in attempting to describe the butt swell, but these high powers have no biological relevance and seem unnecessary with the present solution. The adoption of such high order polynomial taper equations also fails to allow the user to derive merchantable height at a specified inside bark diameter, and hence merchantable volume between two specified inside bark diameters, without undue complexity. This disadvantage is also overcome through the adoption of the simpler quadratic formulation in the segmented equation proposed here.

Although there was slightly more bias in diameter near the tip of the tree as a consequence of forcing compatibility than without such restriction, the trade off was a substantial gain in precision in diameter in the more commercially valuable two lower segments of the stem. Stand variables such as site index and management effects such as thinning, pruning *et cetera* could not be evaluated, because of a lack of adequate information. Their indirect effects, however, could well be represented reasonably through the inclusion of d. This volume based system of volume and compatible taper equations is shown to generate consistent volume estimates for specified merchantability limits.

This study has also recommended a methodology for analysis and derivation of compatible taper - based systems for estimating stem volume and taper, based on a similar approach, but, instead of having v estimated from a volume equation, it is substituted by a constant form factor equation, as shown in equation 7.4.

$$v = fKd^2h \tag{7.4}$$

The compatible taper based taper equation is of the form set out in equation 7.5.

$$\frac{d'^2}{d^2} = f[\beta_1 z^2 + \beta_2 z + d\beta_3(z - \alpha_1)^2 I_1 + d\beta_4(z - \alpha_2)^2 I_2] \tag{7.5}$$

where f is the form factor estimated simultaneously with other parameters and individually for each region. This equation is empirically appealing because it models both the stem volume and taper at the same time. Moreover, inclusion of separate values of f, namely f_1, f_2 and f_3, can reflect regional stem taper characteristics, which eventually affect the stem volume, and are thus analogous to the regional *dummy variables* used in the volume based system. This equation system could not be made exactly compatible as was the case for the volume based one, but the consequences of this inability are small.

Nevertheless, the same methodology can be used to derive the parameter estimates for other sets of data and species.

7.2 DfirStand

7.2.1 Mean Top Height Equation

A general mean top height of Schumacher's projection equation 7.6 was found to fit the data best for the entire of South Island without the need for local dummy variables.

$$h_{(100,2)} = h_{(100,1)} e^{(\alpha+\beta\times AL)(\frac{1}{T_1^\gamma}-\frac{1}{T_2^\gamma})} \tag{7.6}$$

Altitude (AL), however, was found to be a variable that set the overall asymptotic value for the equation and which adequately represented differences of site quality from one locality to another. For every 100 m rise in altitude, the asymptote was lowered by about 2%, which gave rise to an anamorphic series of site index curves that were altitude based. Site index, the mean top height at age 40 years, can be derived from setting T_2=40 years in equation 7.6. The level of precision achieved by this equation is the best so far available for Douglas fir grown in New Zealand, with the RMS value of 0.573 and the standard error of the mean being $\pm$2.6%. The extreme values of residuals were -1.49 and 1.48 m for the minimum and maximum respectively.

7.2.2 Net Basal Area/ha Equation

Equations that include site index have been used extensively in net basal area/ha and volume prediction equations (Schumacher, 1939; Clutter, 1963; Pienaar and Shiver, 1985;

Bailey and Ware, 1983; Clutter and Jones, 1980; Murphy and Farrar, 1988, *et cetera*). These equations have most often been in the form of a modified Schumacher equation such as the one found best in this study and set out in 7.7 below except that in this study altitude was a better predictor than site index,

$$G_2 = G_1^{(\frac{T_1}{T_2})} e^{(\alpha+\beta S)(1-(\frac{T_1}{T_2}))} \tag{7.7}$$

where α represent the asymptote, β is a coefficient and S is site index. Equations such as this have either been general averages for, or specific to a prescribed population. In this study a deeper analysis was conducted to explore the existence of local variations through including also *dummy variables* for the population of Douglas fir in the South Island of New Zealand. The form of equation finally found to be most appropriate in this study was 7.8 below,

$$G_2 = G_1^{(\frac{T_1}{T_2})^{\beta_1}} e^{((\alpha+\beta_2 AL+K_1\beta_3+K_2\beta_4+K_3\beta_5)(1-(\frac{T_1}{T_2})^{\beta_1})+\beta_6 X_t(\frac{1}{T_2}-\frac{1}{T_1})\times\frac{T_t}{T_2})} \tag{7.8}$$

where X_t is the thinning index, defined in this study as the ratio of quadratic mean diameter of trees removed in thinning to quadratic mean diameter of the stand before thinning. There was a 26.7% improvement in the error sum of squares by including the dummy variables. The equation has higher precision than existing ones; minimum and maximum residuals being -3.98 and 3.75 m^2/ha respectively. The stand net basal area/ha equation accounted for most of the local variations that existed, while the regional dummy variables and altitude in the equation determined the asymptote separately for each locality, thus providing what is referred to here as *local adaptation*. The variable site index, which has been traditionally adopted in assisting to predict growth and yield, did not contribute effectively to asymptotic growth because of its relatively weak interaction with altitude, leaving altitude alone to be a dominant predicting factor. The thinning index

(X_t defined above) was imposed in the basal area/ha equation, so that it was possible to simulate the response of growth with respect to the kind of thinning and the time elapsed since last thinning. Despite the addition of all these adaptations, the equation preserved the basic growth modelling requirements for consistency and path invariance.

7.2.3 Stand Volume Equation

Stand volume relationships were derived from sample plot measurements of basal area/ha and mean top height together with revised tree volume equations. The combined variable equation 7.9 was found to conform best with the data.

$$V = \alpha G^{\beta} h_{(100)}^{\gamma} \tag{7.9}$$

Heterogeneity in variance of stand volume was counteracted with a weight, w, shown in equation 7.10.

$$w = \frac{1}{G^2 h_{(100)}} \tag{7.10}$$

The minimum and maximum residual in predicting stand volume/ha were -45.5 and 47.9 m^3/ha respectively. High precision for predicting volume/ha was achieved.

7.2.4 Stem Survival/ha Equation

A modified Gompertz function shown below in 7.11 fitted the mortality data best.

$$N_2 = N_1^{e^{(\beta_1(T_2^2 - T_1^2))}} \times e^{(\alpha+\beta_2 S)(1-e^{(\beta_1(T_2^2-T_1^2)})} \tag{7.11}$$

There were no significant differences in mortality trends among regions, but the values of asymptotes were clearly related to site indices. Trees on stands with higher site indices tended to suffer more mortality, which is biologically consistent with other laws of growth, namely that trees grow faster on sites of higher site quality until they compete for supportive resources. The reliability of predicting regular mortality is within $\pm 2.8\%$ of the standard error of the predicted mean stocking. The maximum residual in predicting mortality is 99 stems/ha, the minimum being -95 stems/ha.

Forecasting growth and yield over a large area through use of general sets of equations that can be disaggregated is intuitively appealing. It makes good use of all the data and avoids the endless and unrealistic demands for very specialized models, leading to proliferation of models, sometimes for even very small populations. The use of *dummy variables* where appropriate was examined in terms of their biological significance, and as a potential methodology for characterizing *locality adaptations*, if other forms of predictor variables were insufficiently sensitive. This approach was found to be eminently satisfactory, especially because of (1) overall minimization of the error sum of squares, though there may not be large differences between regional outputs, and (2) a balanced distribution of residuals. Sensitivity analysis of behaviour in relation to local factors, individually and their interactions, was also helpful in explaining biological variability. The applicability of this generalized approach to growth modelling with the ability to disaggregate, demonstrated that traditional yield table concepts can be enhanced further today through use of modern technological capabilities which allow the acquisition, storage, processing and communication of information in a much more elegant form than has been possible in the past.

Acknowledgements

I wish to thank my supervisor Dr. A.G.D. Whyte for his guidance, especially for his tireless efforts in the preparation of the final draft.

Thanks are also due to my associate supervisor Dr. R.C. Woollons of the School of Forestry, University of Canterbury, Christchurch, New Zealand from whom I received much statistical advice.

Timberlands South (N. Z.) supplied me with the permanent sample plot data for Douglas fir in the South Island. Tasman Forestry, New Zealand allowed me to collect part of the data for use in the study from their forest plantations. Their contribution is greatly appreciated.

I also much appreciated the friendly attitude and moral support of Professor G. Sweet and his wife Margaret Sweet of the School of Forestry, University of Canterbury, Christchurch, New Zealand.

Special thanks are due to the Commonwealth Scholarship Committee, New Zealand, for providing me with the opportunity to study in New Zealand and for the financial support to my family.

I owe thanks to my parents Emil and Flora who sacrificed in so many ways to make this thesis a reality.

Lastly, (but not the least) I thank my wife Grace and son Mcha who suffered in the first hand the pains of preparation of this thesis and remained helpful, their support is invaluable.

Bibliography

[1] Adams, D.M., and R.W. Hynes. (1980). The 1980 softwood timber assessment market model: structure, projections, and policy simulations. For. Sci. Monograph 22. 64 p.

[2] Alban, D.H. (1976). Estimating site index in Northern Minnesota. USDA For. Serv. Res. Pap. NC - 130. 13 p.

[3] Allison, B.J. (1987). Forest Resource Maturity Simulator - 1987. New Zealand Forest Products Ltd., Tokoroa. 175 p.

[4] Alder, D. 1979. "A Distance - Independent Tree Model for Exotic Conifer Plantations in East Africa". Forest Sci. 25: 59 - 71.

[5] Alder, D. (1980). Forest volume estimation and yield prediction. Vol. 2 - Yield Prediction. FOA Forestry Paper 22/2.

[6] Amidon, L.E. (1984). A general taper functional form to predict bole volume for five mixed - conifer species in California. Forest Sci., Vol. 30 (1): 166 - 171.

[7] Anon. (1948). Form - class volume tables (2nd edition). Dominion Forest Service, Dept. of Mines and Resources, Ottawa, Canada.

[8] Arney, J.D. (1972). Computer simulation of Douglas fir growth. Unpublished thesis. Oregon State University. 79 p.

[9] Avery, T.E., and H.E. Burkhart. (1983). Forest measurements. McGraw - Hill Book Co., New York.

[10] Bates, C.G. (1918). Concerning site. J. For. 16: 383 - 388.

[11] Bailey, R.L. (1972). Development of unthinned stands of *Pinus radiata* in New Zealand. Ph.D. thesis, Univ. of Ga, Athens. 73 p.

[12] Bailey, R.L., and R.R. Dell. (1973). Quantifying diameter distributions with the Weibull function. For. Sci. 19: 97 - 105.

[13] Bailey, R.L., and J.P. Clutter. (1974). Base - age invariant polymorphic site index curves. For. Sci. 20(2): 155 - 159.

[14] Bailey, R.L., N.C. Abernathy, and E.P. Jones. Jr. (1981). Diameter distribution model for repeated repeatedly thinned slash pine plantations. *In* proceedings of the First Biennial Southern Silvicultural Research Conference (J. P. Barnett, ed). USDA For. Serv. Tech. Rep. SO - 34: 115 - 126.

[15] Bailey, R.L., and K.D. Ware. (1983). Compatible basal - area growth and yield model for thinned and unthinned stands. Can. J. For. Res. 13: 563 - 571.

[16] Bailey, R.L., T.M. Burgan, and E.J. Jokda. (1989). Fertilized mid - rotation aged slash pine plantations: stand structure and yield prediction models. S.J.A.F. 13: 76 - 80.

[17] Bard, J. (1970). Comparison of gradient methods for solution of the nonlinear parameter estimation problem. SIAM Journal of Numerical Analysis 7: 157 - 186.

[18] Beck, D.E. (1971). Polymorphic site index curves for white pine in the South Appalachians. USDA For. Serv. Res. Pap. SE - 80.

[19] Beekhuis, J. (1978). Growth decline in Douglas fir. FRI Symposium no. 15.0.

[20] Behre, C.E. (1923). Preliminary notes on studies of tree form. Jour. For. 21: 507 - 511.

[21] Behre, C. E. (1927). Form class taper tables and volume tables and their application. J. Agric. Res. 35: 673 - 744.

[22] Behre, C.E. (1935). Factors involved in the application of form class volume tables. J. Agric. Res. 51: 669 - 713.

[23] Bennett, F.A., C.E. MacGee, and J.L. Clutter. (1959). Yield of old - field slash pine plantations. USDA For. Serv. S.E. Exp. Stn.. Res. Pap. 107: 19 p.

[24] Bennett, F.A., and B.F. Swindel. (1972). Taper curves for planted slash pine. USDA Forest Serv Res Note SE - 179: 4 p.

[25] Berkey, C.S. (1982a). Comparison of two longitudinal models for Preschool children. Biometrics 38: 221 - 234.

[26] Berkey, C.S. (1982b). Bayesian approach for nonlinear growth model. Biometrics 38: 953 - 961.

[27] Bickford, C.A., F.S. Baker, and F.G. Wilson. (1957). Stocking, normality, and measurement of stand density. J. For. 55: 99 - 104

[28] Bigin, G.S. (1984). Taper equations for second - growth mixed conifers of Northern California. For. Sci. 30: 1103 - 1117.

[29] Boardes, B.E., R., R.L. Bailey, and K.D. Ware. (1984). Slash pine site index from a polymorphic model by joining (splining) nonpolynomial segments with an algebraic difference method. For. Sci. 30(2): 411 - 423.

[30] Boardes, B.E., R.A. Souter, R.L. Bailey, and K.D. Ware. (1987). Percentile - based distributions characterize forest stand tables. For. Sci. 33(2): 570 - 576.

[31] Brackett, M. (1973). Notes on tariff tree volume computation. State of Washington, Dept. of Nat. Res. Resource Mgt. Rpt. No. 24.

[32] Brickell, J.E. (1968). A method of constructing site index curves from measurements of tree age and height: its application to inland Douglas - fir. USDA For. Serv. Res. Pap. INT - 47.

[33] Bruce, D. (1926). A method of preparing timber yield tables. J. Agric. Res. 32: 543 - 557.

[34] Bruce, D., and F.X. Schumacher. (1950). Forest mensuration. McGraw Hill, New York. 484 p.

[35] Bruce, D., R.O. Curtis and C. Vancoevering. (1968). Development of a system of taper and volume tables for red alder. Forest Sci. 14: 339 - 350.

[36] Bruce, D., D.J. Dermors, and D.L. Reukema. (1977). Douglas fir managed yield simulator: DFTT user's guide. USDA For. Service. Ge. Tech. Rep. PNW 57: 26 p.

[37] Bruce, D. (1977). Yield differences between research plots and managed forests. J. For. 75: 14 - 22.

[38] Bruce D., and L.C. Wensel. (1987). Modelling forest growth: Approaches, and problems. USDA For. Serv. Gen. Tech. Rep. NC -120: 1 - 8.

[39] Buckman, R.E. (1961). Development and use of three stand volume equations for Minnesota. J. For. 59: 573 -575.

[40] Buckman, R.E. (1962). Growth and yield of red pine in Minnesota. U. S. Dept. Agric. Tech. Bull. 1272. 50 p.

[41] Brender, E.V. (1960). Growth predictions for natural stands of loblolly pine in the lower Piedmont. Ga. Forest Res. Council Rpt. 6. 7 p.

[42] Brender, E.V., and J.L. Clutter. (1970). Yield of even - aged natural stands of loblolly pine. Ga. For. Res. Counc. Rep. no. 23.

[43] Burkhart, H.E., and R.B. Tennent. (1977). Site index equations for Douglas fir in Kaingaroa Forest. N.Z. J. For. Sci. 7(3): 417 - 419.

[44] Burkhart, H.E. (1977). Cubic - foot volume of loblolly pine to any merchantable top limit. Southern J. Appl. For. 1: 7 - 9.

[45] Byrne, J.C., and D.D. Reed. (1986). Complex compatible taper and volume estimation systems for red and loblolly pine. Forest Sci. 32: 423 - 443.

[46] Candy, S.G. (1989). Compatible tree volume and variable - form stem form stem taper models for *Pinus radiata* in Tasmania. N.Z. J. For. Sci. 19(1): 97 - 111.

[47] Carron, L.T. (1971). Volume tariff systems. Forestry XLIV: 145 - 150.

[48] Cao, Q.V., Harold E. Burkhart and Timothy A. Max. (1980). Evaluation of two methods for cubic - volume prediction of loblolly pine to any merchantable limit. Forest Sci. Vol. 26 (1); 71 - 80.

[49] Cao, V.Q., and H.E. Burkhart. (1984). A segmented distribution approach for modelling diameter frequency data. For. Sci. 3: 129 - 137.

[50] Carmean, W.H. (1970). Tree height - growth patterns in relation to soil and site. *In* Chester, T. Youngberg, and Charles, B. Davey (ed's). Tree growth and forest soils. Oregon State Univ. Press, Corvallis.

[51] Carmean, W.H. (1972). Site index curves for upland oaks in the central states. For. Sci. 18: 109 - 120.

[52] Cajander, A.K. (1926). The theory of forest types. Acta For. Fenn. 29: 108.

[53] Clutter, J.L. (1963). Compatible growth and yield models for loblolly pine. Forest Sci. 9: 354 - 371.

[54] Clutter, J.L. and F.A. Bennett. (1965). Diameter distributions in old - slash pine plantations. Ga. For. Res. counc. Rep. 23. 9 p.

[55] Clutter J.L., and Allison B.J. (1974). A growth and yield model for *Pinus radiata* in New Zealand. *In* J. Fries (ed), growth models for tree and stand simulation. Royal coll. For., Stockholm, Research Notes 30. 136 - 160 p.

[56] Clutter, J.P. and D.M. Belcher. (1978). Yield of site prepared slash pine plantation in the lower coastal plain of Georgia and Florida. *In* Growth and yield for long - term forecasting of timber yields (J. Fries, H. E. Burkhart, and T. E. Max., Ed's). Va. Polytech. Inst. and State Univ. Sch. For. and Wild. Resour. Publ. FSW - 1 - 78: 53 - 70.

[57] Clutter, J.L. (1980). Development of taper functions from variable - top merchantable volume equations. For. Sci. 26(1): 117 - 120.

[58] Clutter, J.P and E.P., Jr. Jones. (1980). Prediction of growth after thinning in old - field slash pine plantations. USDA For. Serv. Res. Pap. SE - 217.

[59] Clutter, J.L., C. Forison, L.V. Pienaar, G.H. Brister, and R.L. Bailey. (1983). Timber management: a quantitative approach. John Wiley and Sons, New York. 333 p.

[60] Chapman, H.H., and W.H. Meyer. (1949). Forest mensuration. McGraw - Hill Book Co., Inc. New York. 522 p.

[61] Coile, T.S. (1948). Relation of soil characteristics to site index of loblolly and shortleaf pines in the lower piedmont region of North Carolina. Duke Univ. School of For. Bull. 13.

[62] Coile, T.S. (1952). Soil and growth of forests. Adv. in Agron. 4: 329 - 398.

[63] Coile, T.S. and F.X. Schumacher. (1964). Soil - site relations, stand structure, and yields of slash and loblolly pine plantations in Southern United States. T. S. Coile Inc., Durham. NC.

[64] Cunia, T. (1962). Weighted least squares method and construction of volume tables.

Preliminary report. Canadian International paper Co. Montreal, Quebec. Canada. 54 p.

[65] Cunia, T. (1964). Weighted least squares method and construction of volume tables. Forest Sci. 10: 180 - 191.

[66] Curtis, R.O. (1972). Yield tables past and present. J. For. 70(1):28 - 32.

[67] Curtis, R.O., G.W. Clendenen, and D.J. DeMars. (1981). A new stand simulator for coast Douglas - fir: DFSIM user's guide. U.S. For. Serv. Gen. Tech. Rep. PNW - 128.

[68] Daubenmire, R.F. (1961). Vegetative indicators of rate of height growth in ponderosa pine. For. Sci. 7: 24 - 34.

[69] Daubenmire, R.F., and J.B. Daubenmire. (1968). Forest vegetation of Eastern Washington and North Idaho. Wash. Agric. Exp. Stn. Tech. Bull. 60.

[70] Davis, A.W., and P.W. West. (1981). "Remarks on generalized least squares estimation of yield functions": by I. S. Ferguson and J. W. Leech. For. Sci. 27(2): 233 - 239.

[71] Deadman, M.W., and C.J. Goulding. (1979). A method for the assessment of recoverable volume by log types. N.Z. J. For. Sci. 9(2): 225 - 239.

[72] Demaerschalk, J.P. (1971). Taper equations can be converted to volume equations and point sampling factors. For. Chron. 47: 352 - 354.

[73] Demaerschalk, J.P. (1972). Converting volume equations to compatible taper equations. Forest Sci. 18: 241 - 245.

[74] Demaerschalk, J.P. (1973a). Derivation and analysis of compatible tree taper and volume estimating systems. ph.D. thesis, University of British Columbia, Vancouver.

[75] Demaerschalk, J.P. (1973b). Integrated systems for the estimation of tree taper and volume. Can. J. Forest Res. 3: 90 - 94

[76] Demaerschalk, J.P., and A. Kozak. (1977). The whole bole system: a conditioned dual - equation system for precise prediction of tree profiles. Can. J. Forest Res. 7: 488 - 497.

[77] Draper, N.R., and H.Smith. (1981). Applied regression analysis. John Wiley and Sons, Inc., New York. 407 p.

[78] Dress, P.E. (1970). A system for the stochastic simulation of even - aged forest of pure species composition. Thesis for the ph.D. degree, Purdue University, Indiana. 253 p.

[79] Dunningham, A.G., and M.E. Lawrence. (1987). An "interim stand growth model for radiata pine on the central North Island plateau. N. Z., F. R. I. Report no. 3 (unpubl).

[80] Dykstra, D.P., and M. Kallio. (1986). Introduction to IIASA forest sector model. *In* Proc. of the 18th IUFRO World Congress, Division 4: 124 - 135.

[81] Edwards, L., J. Haynes, and B. Strand. (1987). Improving TMP pulp properties by sulphonation process alternatives. Paper presented to Canterbury Chemical Engineering seminar.

[82] Ellis, J.C. (1979). Tree volume tables for major indigenous species in New Zealand. New Zealand Forest Service. FRI. Tech. Pap. 67.

[83] Eng, G., H.G. Daellenbach, and A.G.D. Whyte. (1986). Bucking tree - length stems optimally. Can. J. For. Res. 16: 1030 - 1035.

[84] Farrar, R.M.,Jr. (1979). Growth and yield predictions for thinned stands of even - aged natural longleaf pine. USDA For. Serv. Res. Pap. SO - 156.

[85] Faustmann, M. (1849). Calculation of the value which forest land and immature stands possess for forestry. Translated by W. Linnard (1968). Commonwealth Forest Institute paper no. 42. University of Oxford.

[86] Ferguson, I.S., and J.W. Leech. (1978). Generalized least squares estimation of yield functions. Forest Sci. 24: 27 - 42.

[87] Ferguson, I.S. (1979). Growth functions for radiata pine plantations. *In* H.L. Wright. ed. Proceedings IUFRO s4.01, Oxford. Sept. 1979: 25 - 45.

[88] Fisher, R.A. (1947). The design of experiments, 4th ed., Oliver and Boyd, Edinburgh.

[89] FFSECT - Programme to analyze sectional measurement information. Ministry of Forestry, New Zealand.

[90] Freese, F. (1962). Elementary forest sampling. USDA Dept. Agric. handbook no. 232.

[91] FRI - Ministry of Forestry New Zealand. (1992). Volume tables. General release.

[92] FRI - Ministry of Forestry New Zealand. (1992). Taper tables. General release.

[93] Friedly, J.L., and R.A. Tufts (1989). Analysis of loblolly pine tree centre of mass and mass of inertia. Forest. Sci. 35: 126 - 136.

[94] Fries, J. (1967). Mathematical problems in the construction of yield tables. *In* Communications of the Federal Forest Res. Inst. Vienna. Translated from German. Publish. for USDA For. Serv. Nat. Sci. Foundation, Washington D.C. by the Indiana Nat. Scientific Documentation Centre. New Delhi: 77 - 95.

[95] Furnival, G.M. (1961). An Index for comparing equations used in constructing volume tables. Forest Sci. 7: 337 - 341.

[96] Fuller, W.A. (1969). Grafted Polynomials as approximating functions. The Australian Journal of Agricultural Economics vol. 13: 35 - .46

[97] Gallant, A.R., and W.A. Fuller. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. J. Am. Stat. Assoc. 68: 144 - 147.

[98] Garcia, O. (1979). Modelling stand development with stochastic differential equations. *In* Elliot, D. A. (ed). Mensuration for management planning of exotic forest plantations, N.Z. Forest Service, Forest Research Institute symposium no. 20.

[99] Garcia, O. (1981). An Interactive Forest Simulator for long range planning. N.Z. J. For. Sci. 11(1): 8 - 22.

[100] Garcia, O. (1984a). FOLPI, a Forestry - Oriented Linear Programming Interpreter. *In* Nagumo, H. *et al.* (ed's), proceedings IUFRO symposium on forest management planning and managerial economics, Tokyo University: 293 - 305.

[101] Garcia, O. (1984b). New growth models for even - aged stands: *Pinus radiata* in Golden Downs forest. N.Z. J. For. Sci. 14: 65 - 88

[102] Garcia, O. (1987). Experience with an advanced growth modelling methodology. Proceedings IUFRO Forest Growth Modelling and Prediction Conference. Minneapolis, MN, Aug 24 - 28.

[103] Garcia, O. (1988). Growth modelling - a (Re)view. N.Z. Forestry Article.

[104] Gedney, D. R., and F.A. Johnson. (1959). Weighting factors for computing the relation between tree volume and d. b. h. in Pacific North West. Pac. N. W. Forest and Range Exp. Stn. Res. Note 174. 5 p.

[105] Gertner, G.Z. (1984). Localizing a diameter increment model with a sequential Bayesian procedure. For. Sci. 30(4): 851 - 864.

[106] Gevorkiantz, S.R., and L.P. Olsen. (1955). Composite volume tables for timber and their application in the Lake States. U. S. Dept. Agric. Tech. Bull. no. 1104.

[107] Gilmour, J.W. (1966). The pathology of forest trees in New Zealand. N.Z. For. Res. Inst. Tech. Pap. 48.

[108] Girard, J.W., and D. Bruce.(1946). Board foot volume tables for 32 - foot logs. Amer. Bank Bldg., Portland, Oregon: 40 p.

[109] Goulding, C.J., and J.C. Murray. (1976). Polynomial taper equations that are compatible with tree volume equations. N Z J Forest Sci. 5: 313 - 322.

[110] Goulding, C.J., and J.W. Shirley. (1979). A method to predict the yield of log assortments for long - term planning. *In* Elliott, D. A. (ed) Mensuration for Management of Exotic Forest Plantations. New Zealand Forest Service, Forest Research Institute symposium no. 20: 301 - 314.

[111] Golding, D.L., and O.F. Hall. (1961). Tests of precision of cubic foot tree volume equations on aspen, jack pine and white spruce. For. Chron. 37:123 - 132.

[112] Gordon, A. (1983). Comparison of compatible polynomial taper equations. N.Z. J. For. Sci. 13: 146 - 155.

[113] Gordon, A. (1985). A manual for instruction for planning and executing an inventory of tree volume and taper. N.Z. Forest Service. Unpub.

[114] Gray, H.R. (1966). Principles of forest tree and crop volume growth. Bull. For. and Timb. Bur., Canberra. no. 42.

[115] Graney, D.L. and H.E. Burkhart. (1973). Polymorphic site index curves for shortleaf pine in Ouachita Mountains. USDA For. Serv. Res. Pap. SO - 85.

[116] Green, E.J., W.E. Strawdermann, and C.E. Thomas. (1992). Empirical Bayes development of Honduran pine yield models. Forest Sci. 38(1): 21 - 33.

[117] Gujarat, D. (1970). Use of dummy variables in testing for equality between sets of coefficients in linear regression: a generalization. American statistician 25 (4): 18 - 22.

[118] Hellawell, C.R. (1978). Douglas fir as a source of timber for engineering purpose. FRI symposium no. 15: 240 - 252.

[119] Harrison, W.C., H.E. Burkhart, T.E. Burk, and D.E. Beck (1986). Growth and yield of Appalachian mixed hardwoods after thinning. Sch. of For. and Wildl. Resour. VPI and SU. B. V. Publ. FWS - 1 - 86.

[120] Heijbell, I. (1928). (A system of equations for determining stem form in pine). Svensk. SkogasForen. Tidskr. 3 - 4: 393 - 422. (in Swedish, summary in English).

[121] Hayward, W.J. (1987). Volume and taper of *Eucalyptus regnans* grown in Central North Island of New Zealand. N.Z. J. For. Sci. 17(1): 109 - 120.

[122] Hafley, W.L., and H.T. Schreuder. (1977). Statistical distribution for fitting diameter and height data in even aged stands. Can. J. For. Res. 7: 481 - 487.

[123] Hartley, H.O. (1961). The modified Gauss - Newton method for the fitting of no - linear regression functions by least squares. Technometrics 3: 269 - 280.

[124] Hodkings, E.J. (1961). Estimating site - index for longleaf pine through quantitative evaluation of associated vegetation. Proc. Soc. Am. For. 1960: 28 - 32.

[125] Hoel, P.G. (1964). Methods of comparing growth type curves. Biometrics 20: 859 - 872.

[126] Hoyer, G.E. (1972). Measuring and interpreting Douglas fir management practices (Explanation of simulation technique, its results and meaning). State of Washington, Department of Natural Resources, Report no. 24.

[127] Hudson, D.J. (1966). Fitting segmented curves whose join points have to be estimated. American Statistical Association Journal 61: 1097 - 1129.

[128] Hojer, A.G. (1903). Tallens och granens tillvaxt. Bihang till Fr. Loven. Om Vara barrskogar. Stockholm, 1903. (in Swedish).

[129] Hood, I.A., and K.J. Kershaw. (1973). Distribution and life history of *Phaeocryptopus gaeumannii* on Douglas fir in New Zealand. N.Z. For. Serv. For. Res. Inst. For. pathology Rep. 37 (unpub).

[130] Hood, I.A., and K.J. Kershaw. (1975). Distribution and infection period of *Phaeocryptopus gaeumannii* in New Zealand. For. Sci. 5(2). 201 - 208.

[131] Honer, T.G. (1965). A new total cubic - foot function. For. Chron. 41(4): 476 - 493.

[132] Hummel, F.C. (1955). The volume/basal area line; a study in forest mensuration. U.K. For. Comm. Bull. no. 24. 84 p.

[133] Hummel, F.C., G.M. Locke, and J.P. Verel. (1962). Tarif tables. For. Rec. For. Comm. no. 31. (London).

[134] Husch, B. (1963). Forest mensuration and statistics. Ronald Press Co., New York.

[135] Husch, B., C.I. Miller, and T.W. Beers. (1972). Forest mensuration. Ronald Press Co., New York. 410 p.

[136] Jensen, D.R. (1982). Efficiency and Robustness in use of repeated measurements. Biometrics 38: 813 - 825.

[137] Jenss, R.M., and N. Bayley. (1937). A mathematical method for studying the growth of a child. Human biology 9: 556 - 563.

[138] James, R.N., and Bunn, E.H. (ed's). (1978). Editors summary. FRI symposium No. 15.

[139] Johnson, N.L., and S. Kotz. (1970). Continuous univariate distribution —- Houghton Miffling Co., Poston.

[140] Johnson, K.N., and E.L. Scheurman. (1977). Techniques for prescribing optimal timber harvest and investment under different objectives - discussion and synthesis. For. Sci. Monograph 18.

[141] Johnson, S. (1989). Modelling regional forest industry development in New Zealand. Ph.D. thesis. Sch. of For. Univ. of Canterbury. Christchurch, N.Z.

[142] Johnston, D.R., and R.T. Bradley. (1963). Forest management tables. Comm. For. Rev.,42: 217 - 227.

[143] Jolly, N.W. (1951). The volume line theory in relation to the measurement of the standing volume of a forest (with particular reference to *Pinus radiata*). S. Australia Woods and Forest Dept.

[144] Katz, A., and A.G. Dunningham. (1981). Compatible volume and taper functions for New Zealand radiata pine grown under direct sawlog regime. Forest mensuration report no. 59. 12 p.

[145] King, J.E. (1966). Site index curves for Douglas - fir in the Pacific Northwest. Weyerhauser Forestry Paper 8.

[146] Kirkland, A. (1969). Notes on the Establishment and thinning of old crop Douglas fir in Kaingaroa forest. N.Z. J. For. 14: 25 - 37.

[147] Kozak, A. and J.H.G. Smith. (1966). Critical analysis of multivariate techniques for estimating tree taper suggests that simpler methods are best. For. Chron. 45: 458 - 463.

[148] Kozak, A., D.D. Munro and J.H. Smith. (1969). Taper functions and their applications in forest inventory. For Chron 45: 278 - 283.

[149] Kozak, A. (1988). A variable - exponent taper equation. Can. J. For. Res. 18: 1363 - 1368.

[150] Knoebel, B.R., H.E. Burkhart, and E.B. Donald. (1986). A growth and yield model for thinned stand of yellow poplar. For. Sci. Monograph 27.

[151] Law, K.R.N. (1990). A growth model for Douglas fir grown in the South Island of New Zealand. FRI/Industry Research Cooperatives: Stand Growth Modelling Cooperative report no. 18. Forest Research Institute, Private Bag, Rotorua.

[152] Lee, Y. (1967). Stand models for lodgepole pine and limits to their application. Unpublished thesis. University of British Columbia. 322 p.

[153] Lemmon, P.E. and F.X. Schumacher. (1962). Volume and diameter growth of Ponderosa pine trees as influenced by site index, density, age, and size. For. Sci. 8:236 - 249.

[154] Liu, Xiu. (1990). Growth and yield of Douglas fir plantations in the central North Island of New Zealand. Ph.D. thesis. Sch. of For. Univ. of Canterbury, Christchurch, N.Z.

[155] Lin, J.Y.(1970). Growing space index and stand simulation of young Western hemlock in Oregon. Thesis for the ph.D. degree, Duke univ. 182 p.

[156] Lewis, N.B. and J.H. Harding. (1963). Soil factors in relation to pine growth in South Australia. Austal. For. 27. 27 - 34.

[157] Lewis, N.B., A. Keeves, and J.W. Leech. (1976). Yield regulation in South Australian *Pinus radiata* plantations. South Australia woods and forest department. Bulletin no. 23.

[158] Loetsch, F. and K.E. Haller. (1973). Forest inventory vol. 2. BLV Verlagsesellschft Munchen Bern Wien. 469 p.

[159] Lundgren, A.L., and W.L. Dolid. (1970). Biological growth functions describe published site index curves for Lake States timber species. USDA For. Serv. Res. Pap. NC - 36.

[160] McClure, J.P., and R.L. Czaplewski. (1986). Compatible taper equation for loblolly pine. Can. J. For. Res. 16: 1272 - 1277.

[161] MacGee, C.E. (1961). Soil site index for Georgia slash pine. USDA For. Serv. S.E. For. Exp. Stn. Paper 119.

[162] MacKinney, A.L., F.X. Schumacher, and L. E. Chaiken. (1937). Construction of yield tables for nonnormal loblolly pine stands. J. Agric. Res. 54: 531 - 545.

[163] MacKinney, A.L. and L.E. Chaiken. (1939). Volume, yield and growth of loblolly pine in the mid - Atlantic coastal region. Appalachian Forest Exp. Stn. Tech. Note 33: 30 p.

[164] McGuigan, B.N. (1984). A log resource allocation model to assist forest industry managers in process selection and allocation, wood allocation and transportation and production planning. Appita 37(4): 289 - 296.

[165] McGuigan, B.N. (1992). REGRAM - A flexible forest management optimization tool. *In* A. G. D. Whyte (ed). Proceedings of the IUFRO s3.04.01 conference on Integrated Decision Making and Control of Forest Operations. Sch. of For. Univ. Canterbury, Christchurch, N.Z. Jan. 1992: 34 - 38.

[166] Matte, L. (1949). The taper of coniferous species with special reference to loblolly pine. For. Chron. 25: 21 - 31.

[167] Marquardt, D.W. (1963). An algorithm for least - squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11: 431 - 441.

[168] Martin, A.J. (1981). Taper and volume equations for selected Appalachian hardwood species. USDA For. Serv. Res. Pap. NE - 490. 22 p.

[169] Matney, T.G., and A.D. Sullivan. (1982a). Variable top volume and height predictors for slash pine trees. For. Sci. 28(2): 274 - 282.

[170] Matney, T.G. and A.D. Sullivan. (1982b). Compatible stand and stock tables for thinned loblolly pine stands. Forest Sci. 28: 161 - 171.

[171] Max, T.A., and H.E. Burkhart. (1976). Segmented polynomial regression applied to taper equations. Forest Sci. 22: 283 - 289.

[172] Meyer, W.H. (1930). Diameter distribution series in even - aged forest stands. Yale Forestry School Bull. 28.

[173] Meyer, H.A. (1941). A correction for a systematic error occurring in the application of the logarithmic volume equation. Pa. State Forestry Sch. Res. Paper 7: 3 p.

[174] Meyer, H.A., and D.D. Stevenson. (1943). The structure and growth of virgin beech - birch - maple - hemlock forest in Northern Pennsylvania. J. Agric. Res. 667: 465 - 484.

[175] Meyer, H.A. (1953). Forest Mensuration. Penn Valley Publishers, State coloedge, Pa. 357 p.

[176] Mesavage C., and J.W. Girard. (1946). Tables for estimating board foot volume of timber. Washington, D. C. U.S. Forest Service.

[177] Merrell, M. (1931). The relationship of individual growth to average growth. Human Biology 3: 36 - 70.

[178] Mitchell, K.J. (1969). Simulation of the growth of even - aged stand of white spruce. School of For. Yale Univ. Bull. No. 75. 48 p.

[179] Moser, J.W. and T.W. Beers. (1969). Parameter estimation in nonlinear volume equations. Journal of Forestry: 878 - 879.

[180] Mountfort, C.J. (1978). Growth of Douglas fir. N.Z. Forest Service. For. Res. Inst., Symposium no. 15: 375 - 411.

[181] Murphy, P.A. and Robert M. Farrar, Jr. (1988). Basal - area projection equations for thinned natural even - aged forest stands. Can J. For. Res. 18: 827 - 832.

[182] Monserud, R.A. (1984a). Problems with site indexes: An opinionated review. P. 167 - 180 *In* Forest land proc., J. Bockheim (ed). Dep. Soil Sci., univ. Wisconsin, Madison.

[183] Monserud, R.A. (1984b). Height growth and site index curves for inland Douglas - fir based on stem analysis data and forest habitat type. Forest Sci. 30: 943 - 965.

[184] Monserud, R.A. (1985a). Comparison of Douglas fir site index and height growth curves in Northwest. Can. J. For. Res. 15(4): 673 - 679.

[185] Monserud, R.A. (1985b). Applying height growth and site index curves for inland Douglas fir. USDA For. Serv. Res. Pap. INT - 357. 22 p.

[186] Monserud, R.A. (1987). Variations on a theme of site index. USDA For. Serv. Tech. report no. NC - 120: 419 - 427.

[187] Munro, D.D., and J.P. Demaerschalk. (1974). Taper - based versus volume - based compatible estimating systems. For. Chron. 50(5): 1 - 3.

[188] Munro, D.D. (1974). Forest growth models — a prognosis. *In* Growth models for tree and stand simulation (J. Friees, ed.): Royal College of Forestry, Stockholm, Sweden.

[189] Myers, C.A., and J.S. Van Deusen. (1960). Site index of Ponderosa pine in Black Hills from soil and topography. J. For. 58: 548 - 555.

[190] Myers, C.A. (1971). Field and computer procedures for managed stand yield tables. USDA Res. Pap. RM - 79: 24 p.

[191] Navon, D.I. (1971). Timber RAM: A long - range planning method for commercial timberlands under multiple - use management. USDA, Pacific Southwest Forest and Range Exp. Stn. Res. Pap. PSW - 70. 22 p.

[192] Nelson, T.C. (1964). Diameter distribution and growth of loblolly pine. For. Sci. 10: 105 - 115.

[193] Newberry, J.D., and L.V. Pienaar. (1978). Dominant height growth models and site index curves for site - prepared slash pine plantations in the lower coastal plain of Georgia and North Florida. Univ. of Ga. Plantation Mgt. Res. Coop. Res. Pap. no. 4.

[194] Newberry, J.D., and T.E. Burk. (1985). S_b distribution - based models for individual tree merchantable volume - total volume ratios. For. Sci. 31: 389 - 398.

[195] Newham, R.M., and J.H.G. Smith. (1964). Development and testing of stand models for Douglas fir and lodgepole pine. For. Chron. 40. 492 - 502.

[196] Newham, R.M. (1964). The development of a stand model for Douglas - fir. Unpublished thesis. University of British Columbia. 201 p.

[197] Newham, R.M. (1967). A modification to the combined variable formula for computing tree volumes. J. For. 65: 719 - 720.

[198] New Zealand national exotic forest directory, (NEFD). (1989).

[199] Nokoe, S. (1978). Demonstrating the flexibility of the Gompertz function as a yield model using mature species data. Comm. For. Rev. 57. 35: - 42.

[200] Ogweno, D.C. (1992). Planning tactical and operational log allocation by stem quality grades. *In* A. G. D. Whyte (ed). Proceedings of the IUFRO s3.04.01 conference on Integrated Decision Making and Control of Forest Operations. Sch. of For. Univ. Canterbury, Christchurch, N.Z. Jan. 1992: 77 - 81.

[201] Olson, D.F. Jr., and L. Della - Bianca. (1959). Site index comparisons for several tree species in Virginia - California Piedmont. USDA For. Serv. S.E. Exp. Stn. Pap. 104.

[202] Ormerod, D.W. (1973). A simple bole model. For Chron 49: 136 - 138.

[203] Osumi, S. (1959). Studies on stem form of the forest trees. I. J. Jpn. For. Soc. 41: 471 - 479.

[204] Osborne, J.G., and F.X. Schumacher. (1935). The construction of normal - yield and stand tables for even - aged timber stands. J. Agric. Res. 51: 547 - 564.

[205] Pienaar, L.V., and K.J. Turnbull. (1973). The Chapman - Richards generalizaton of von Bertalanffy's model for basal area growth and yield in even - aged stands. Forest Sci. 19: 2 - 22.

[206] Pienaar, L.V. (1979). An approximation of basal area growth after thinning based on growth in unthinned plantations. Forest Sci. 25: 223 - 232.

[207] Pienaar, L.V., and B.D. Shiver. (1984). An analysis of basal area growth in 45 - year - old unthinned and thinned slash pine plantation plots. Forest Sci. 30: 933 - 942.

[208] Pienaar, L.V., B.D. Shiver, and G. E. Grider (1985). Predicting basal area growth in thinned slash pine plantations. For. Sci. 31(3): 731 - 741.

[209] Pnevmaticos, S.M., and S.H. Mann. (1972). Dynamic programming in tree bucking. For. Prod. J. 22(2): 26 - 30.

[210] Ralston, A., and C.L. Meek. (1976). Encyclopaedia of computer science. First edition. Petrocelli - Charter. New York.

[211] Ralston, M.L., and R.I. Jenrich. (1979). Dud. a derivative - free algorithm for nonlinear least - squares. Technometrics 1: 7 - 14.

[212] Reineke, L.H. (1933). Perfecting a stand density index for even - aged forests. J. Agric. Res. 46: 627 - 638.

[213] Romancier, R.M. (1961). Weight and volume of plantation grown loblolly pine. USDA for. Serv. S.E. Exp. Stn. Res. Note No. 161.

[214] Roth, F. (1916). Concerning site. Forestry quarterly 15: 3 - 13.

[215] SAS Institute Inc. (1988). SAS/Language reference, Version 6, Fourth Edition. SAS institute. Inc., Cary, NC., USA.

[216] SAS Institute Inc. (1988). SAS/Procedures guide, Version 6, Fourth Edition. SAS institute. Inc., Cary, NC., USA.

[217] SAS Institute Inc. (1988). SAS/STAT user's guide, Version 6, Fourth Edition. SAS institute. Inc., Cary, NC., USA.

[218] SAS Institute Inc. (1988). SAS/GRAPH user's guide, Version 6, Fourth Edition. SAS institute. Inc., Cary, NC., USA.

[219] Schumacher, F.X., and F. Hall (1933). Logarithmic expression of tree volume. J. Agric. Res. 47: 719 - 734.

[220] Schumacher, F.X. (1939). A new growth curve and its application to timber - yield studies. J. For. 37: 819 - 820.

[221] Schumacher, F.X, and R.A. Chapman. (1954). Sampling methods in forestry and range management. Duke Univ., School of For. Bull. 7, revised. 222 p.

[222] Schumacher, F.X. and T.S. Coile. (1960). Growth and yield of natural stands of the Southern pines. T. S. Coile, Inc., Durham, N. C. 115 p.

[223] Shifley, S., and S. Fairweather. (1983). Individual-tree diameter growth and mortality models for western Oregon. *In* Proceedings of renewable resource inventories for monitoring changes and trends. Corvallis, Oregon. 737 p.

[224] Smalley, G.W. and R.L. Bailey. (1974). Yield tables and stand structure for loblolly pine plantations in Tennessee, Alabama, and Georgia highlands. U.S. Dep. Agric. For. Serv. Pap. SO - 96

[225] Smith, W.B. (1981). Adjusting the STEMS regional forest growth model to improve local predictions. USDA Forest Res. Note NC - 297: 5p.

[226] Spurr, S. H. (1952). Forest Inventory. Ronald Press Co., New York. 476 p.

[227] Spurr, S.H., and B.V. Barnes. (1980). Forest ecology (3rd ed). John Wiley and Sons. New York. 687 p.

[228] Stage, A.R. (1973). Prognosis model for stand development. USDA For. Serv. Res. Pap. INT - 137. 32 p.

[229] Stage, A. (1981). Use of self calibration procedures to adjust general regional yield models to local conditions. *In* Proceedings of forest resource inventory, growth models, management planning, and remote sensing. IUFRO World Congress, Kyoto, Japan: 365 - 375.

[230] Steel, R.G.D., and J.H. Torrie. (1980). Principles and procedures of statistics: a biometrical approach. McGraw - Hill Book Co., Inc., 633 p.

[231] Steinbrenner, E.C. (1975). Mapping forest soils on Weyerhauser lands *In* the pacific Northwest, In B. Bernier and C. H. Winget (ed's), Forest soils and forest land management. Les Presses de L'univer

[232] Sullivan, A.D., and J.L. Clutter. (1972). A simultaneous growth and yield model for loblolly pine. For. Sci. 18: 76 - 86.

[233] Threadgill, J., and A. A. Twaddle. (1985). AVIS - assessment of value by individual stems. FRI Bulletin no.110.

[234] Troudsell, K.B., D.E. Beck, and F.T. Lloyd. (1974). Site index for loblolly pine in Atlantic Coastal Plain of the Carolinas and Virginia. USDA For. Serv. Res. Pap. SE - 115.

[235] Turnbull, K.J., and Hoyer G.E. (1965). Construction and analysis of comprehensive tree - volume tarif tables. Resource Management Report no. 8. State of Wash., Dept. of Nat. Resources, Olympia, Wash.

[236] Umemura, T. (1984). An entirely new growth curves based on a second order linear differential equation. Proc IUFRO Symp. For. Mang. Plan and Mang. Econ. Uni. Tokyo, Japan, Oct. 15 - 19. 1984: 591 - 597.

[237] Ure, J. (1950). The natural vegetation of the Kaingaroa plains as indicator of site quality for exotic conifers. N.Z. J. For. 6: 112 - 123.

[238] Van Deusen, P.C., T.G. Matney, and A.D. Sullivan. (1982). A compatible system for predicting the volume and diameter of sweetgum trees to any height. South. J. Appl. For. 6: 159 - 163.

[239] van Wyk, J.L. (1983). A management decision support system for process selection and log and wood residue allocation for the new crop radiata pine in New Zealand. Appita 37(3): 219 - 222.

[240] van Wyk, J.L., and G. Eng. (1987). Computer - aided sawmill design and evaluation. *In* Kininmonth, J.A. (compiler). Proceedings of the Conversion Planning Conference. FRI Bulletin no. 128: 83 - 96.

[241] Ware, K.D., B.E. Boardes, and R.L. Bailey. (1988). Estimating growth and yield of thinned slash pine plantations on old fields. *In* Forest Growth Modelling and Prediction. Vol. 1. USDA For. Serv. North Central For. Exp. Stn. Gen. Tech. Rep. NC - 120: 33 - 36.

[242] Ward, N.H. (1987). Modelling Japan - South Seas trade in forest products. USDA For. Serv., PNW. Res. Stn. Gen. Tech. Rep. PNW - GTR - 210. 28 p.

[243] Whyte, A.G.D. (1970). Sectional Measurement of Trees: A Rationalised Method. New Zealand Forest Service. ODC. 524.1: 74 - 79.

[244] Whyte, A.G.D., M.J. Temu, and R.C. Woollons. (1992). Improving yield forecasting reliability through aggregated modelling. *In* G. B. Wood, and B. J. Turner (ed's). Proceedings IUFRO - Integrating Information over Space and Time; Australian National University Canberra, Jan. 13 - 17, 1992: 81 - 88.

[245] Wirshat, J. (1938). Growth rate determination in nutrition studies with the Bacon pig, and their analysis. Biometrica 30: 16 - 28.

[246] 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.

[247] Woollons, R.C., A.G.D. Whyte, and Liu Xiu. (1990). The Hossfeld function: an alternative model for depicting stand growth and yield. Journal of Japanese Forestry Society.

[248] Whiteside, I.D., and W.R.J. Sutton. (1983). A silvicultural stand model: implications for radiata pine management. N.Z. J. For. 28(3): 300 - 313.

[249] Whiteside, I.D., McGregor, M.J., and B.R. Manley. (1987). Prediction of radiata pine log grades. *In* Kininmonth, J. A., (compiler) Proceedings of Conversion Planning Conference. FRI Bulletin no. 128: 55 - 69.

[250] Wykoff, W.R., N.L. Crookston, and A.R. Stage. (1982). User's guide to stand prognosis model. USDA For. Serv. General technical report INT - 133. 112 p.

[251] Wykoff, W.R. and R.A. Monserud. (1987). Representing site quality in increment models: a comparison of methods. Paper presented at IUFRO Forest Growth Modelling and Prediction Conference, August 23 - 27, 1987, Minneapolis, MN. 8 p.

[252] Zahner, R. (1962). Loblolly pine site curves by soil groups. For. Sci. 8: 104 - 110.

[253] Zon, R. (1913). Quality classes and forest types. Proceedings of the society of American foresters 8: 100 - 104.

Appendix A

Data Summary

A.1 Sectional Measurement Data by Forests

Forest	Number of Trees sampled
NELSON REGION	
Golden Downs	82

Forest	Number of trees Sampled
CANTERBURY REGION	
Ashley	264
Albury Park	12
Darfield	11
Hamner	39
Holme Station	9
Homebush	9
Kaiwara Station	2
Methven	1
Mayfield	4
Peel	7
Pusey Station	4
Raincliff	7
S.P.B Centennial block	9
Winscale	6
Total	384

Forest	Number of Trees Sampled
SOUTHLAND REGION	
Berwick	11
Conical Hill	21
Dunedin City Council	7
Dusky	4
Herbert	3
Naseby	11
Longwood	59
Pebby Hills	10
Dusky	5
Tapanui	20
Total	175

A.2 Schedule of Permanent Sample Plots

Forest	Number of Permanent Sample Plots
CANTERBURY	
Ashley	33
Geraldine	7
Hamner	20
Omihi	3
Total	65

Forest	Number of Permanent Sample Plots
NELSON	
Golden Downs	127
Wairau	1
Rai Valley	7
Total	135

Forest	Number of Permanent Sample Plots
SOUTHLAND	
Berwick	15
Beaumont	1
Herbert	12
Hokonui	9
Naseby	4
Longwood	28
Pomohaka	28
Rankleburn	9
Silverpeaks	6
Total	112

Forest	Number of Permanent Sample Plots
WESTLAND	
Granville	2
Hochstetter	13
Mawhera	12
Mahinapua	4
Victoria SFP	14
Total	45

Appendix B

Research Paper

Reprint of Whyte, A.G.D., M.J. Temu, and R.C. Woollons. (1992). Improving yield forecasting reliability through aggregated modelling. *In* G.B. Wood and B.J. Turner (ed's). Proceedings IUFRO, Integrating Information over Space and Time. Australian National University Canberra. Jan 13 - 17, 1992: 81 -88.

IMPROVING YIELD FORECASTING RELIABILITY THROUGH AGGREGATED MODELLING

A.G.D.Whyte, M.J.Temu & R.C.Woollons,
School of Forestry, University of Canterbury,
Christchurch, New Zealand

SUMMARY

There has been a tendency in some countries to develop and use growth and yield models that are specific to increasingly restricted sub-populations. In New Zealand, there are currently about 16 models in use for one species, radiata pine alone, when 20 years ago there were two. One way to reduce possible unnecessary proliferation is to fit average curves or functions for large populations. This approach, however, may not satisfy the level of sensitivity that is considered vital for the intensive plantation management which now exists in New Zealand. Another possibility is to provide additional predictor variables which allow users to disaggregate general trends so that production forecasts can be tailored to local variations. An analysis is presented here to demonstrate the potential of this strategy to reduce the number of growth and yield models while still providing adequate local sensitivity. The main advantages are:

(1) a larger data base is available than would eventuate if functions were developed locally or even regionally;

(2) there is better consistency in growth predictions across regions and over time; and

(3) new plantation areas with no existing growth functions have a better potential to be appropriately accommodated.

Permanent sample plot data pertaining to Douglas fir throughout the whole of the South Island of New Zealand have been analysed by way of example to demonstrate the feasibility of accounting for local variations in stand height, basal area per hectare, stocking and in tree volume and taper functions.

INTRODUCTION

This paper sets out to argue the case for a return to the original object of yield forecasting, namely to characterise general cumulative growth trends broadly and to cater for later disaggregation and adaptation to local conditions if necessary. To some extent this approach was forced on early researchers, particularly when only graphical solutions were available, because of the limitations of their tools. Today, with an ever increasing computer and statistical capability, it is relatively easy to devise a set of growth functions and prepare a computer model for individual sub-populations as the need arises.

In New Zealand today there are 16 existing growth models for radiata pine alone reflecting differences mainly in locality. This number does, admittedly, include three pairs of models for the same locality (one out of each pair catering for fertilised stands and the other for unfertilised stands), but it does not include separation of crops into genetically improved planting stock or not, a feature of several other models requiring different functions and completely separate paths through the computer programs, so that in effect they too are different models. We know, moreover, of at least two other radiata pine growth models that are currently under construction in New Zealand, and there could well be others.

There are two basic concerns that must be addressed in this regard:

(1) is such proliferation warranted statistically and biologically?
(2) even if separation were warranted, is it advisable?

The answer to the first of these questions is probably no, if prior studies are anything to judge from. Until the late 1960's, for example, production forecasting for radiata pine throughout New Zealand was conducted with tables derived by Lewis (1954) for unthinned stands and by Beekhuis (1966) for thinned stands. The only part of the prediction system that differed from locality to locality was for the thinned stands model, in which regional height on age and stand volume functions could be substituted for the national average ones. Curiously, predictions of gross and net basal area per hectare growth from the unthinned model have been found to be almost exactly the same as from the thinned one between any two times of thinning (School of Forestry, University of Canterbury, unpublished student exercises). García (1990), moreover, inferred from an analysis of a very large database that choice of stocking regime was likely to be more important than locality in influencing final yield. Furthermore, there has been widespread adoption of a Central North Island model (KGM3) to represent growth of radiata pine from the far North to the deep South of New Zealand in preference to local regional equations, because the former has been found to mirror the actual yields of certain stocking regimes better than the latter.

The second question is, in our opinion, probably the more crucial one, because it raises all sorts of impracticalities and imponderables. For example, should a production forecaster change growth models from time to time, as apparently more reliable ones materialise? Should that forecaster calibrate the full range of **all** model options against actual realisations? How can one be sure that what appears to be best locally based on prior monitoring, will remain so in the future? If a change is made in the growth model, should all or just some previous forecasts be completely revised to maintain consistency over time?

There is, however, another slant to the problem. Plantation managers are seeking to make decisions with increasing sensitivity in regard to the cropping regimes to adopt. The growth models they resort to could well provide them with illusory accuracy and sometimes misleading predictive capability. The whole estimating process is further confused by the confounding use of tree or stand volume functions that have historically become part and parcel of the computation of plot volumes that subsequently became basic data inputs used to derive growth model predictions of volume per hectare. Again, one needs to examine carefully whether or not all previously determined plot volumes should be revised when apparently better fitting basic volume functions materialise from time to time.

A more satisfactory solution is to develop crop growth functions and also stem taper and volume functions which use predictor variables that help characterise not only past but which could also represent likely future variability. If that objective can be achieved, the problem of choice of model to reflect variation in space and time consistently would largely be overcome. The following case study indicates the potential there might be to aspire to achieve such a goal.

DOUGLAS FIR CASE STUDY

Investigations have been carried out with a large data-base of Douglas fir permanent sample plot measurements from all parts of the South Island of New Zealand for a post-graduate research study by the second-named author. Altogether, 368 permanent sample plots of between 0.02 and 0.04 ha and 2137 sets of measurements were used in the analysis. A number of site variables such as altitude, latitude, distance from the sea, annual rainfall, soil type, and so on were also considered as possible variables to explain regional growth variation. As altitude and locality were later found to be the most promising predictors, a breakdown of plots by region, altitude and age is shown in Table 1.

Table 1 *Plot frequency by locality, altitude and age*

Region	Altitude (m - asl)	Age Range (years)	No. of Plots	Total No. of Measurements
Nelson	0 - 200	12 - 29	3	10
	201 - 400	15 - 54	63	472
	401 - 600	8 - 58	70	570
	601 - 800	25 - 36	3	27
Canterbury	0 - 200	28 - 47	10	47
	201 - 400	13 - 61	43	224
	401 - 600	26 - 40	6	24
	601 - 800	1 - 12	4	12
Southland	0 - 200	22 - 65	59	261
	201 - 400	17 - 78	43	200
	401 - 600	13 - 64	13	12
Westland	0 - 200	12 - 59	18	86
	201 - 400	15 - 54	27	183
Total			368	2137

The data also cover a wide range of growing conditions, with annual rainfalls ranging from 500 to 5600 mm, mean annual temperatures from 10°C to 13°C, low to very high soil fertility, all of which factors individually and in various combinations could lead to greatly differing levels of crop productivity. Analysis of crop production data were conducted with a full knowledge of these sorts of environmental variable for all plots.

This region was also chosen because another Douglas fir growth model had just been constructed for it, which differs from the one reported here in that it consists of production functions that represent averages throughout the whole of the South Island. While we have full knowledge of its predictive capability and can compare results from it with our own, it is not possible to quote these comparisons directly here, because it is expressly forbidden to do so by the Research Modelling Cooperative that commissioned that other study. We can only infer broadly and indirectly, therefore, what such comparisons might reveal.

Stand Height Growth

Polymorphic and anamorphic forms of the Schumacher, Chapman-Richards, Gompertz, Hossfeld, Levakovic and Weibull functions were fitted to the data. The mean top height equation found to fit best was a form of the Schumacher equation (see Schumacher, 1939),

$$\bar{h}_{100,2} = \bar{h}_{100,1} \exp((\beta_1 + \beta_2 \text{ Altitude})(1/T_1^{\beta_3} - 1/T_2^{\beta_3})) \qquad (1)$$

where $\bar{h}_{100,i}$ represents mean top height at age T_i and β_1, β_2 and β_3 are coefficients estimated by non-linear least-squares and altitude is height above sea level in m.

Site index, the mean top height of Douglas fir at age 40 years, is derived by setting $T_2 = 40$ in equation (1).

Comparison of equations was based mainly on various statistical tests of the model residuals by region. A single equation for all localities was found to be more precise and with no regional bias in the residual patterns, compared with fitting equations by individual regions or large forest aggregations on their own. Altitude controlled not only the level of the asymptote but also appeared to characterise differences in site quality very well from one locality to another.

Indeed, the precision achieved in this overall equation is better than in any other existing equation known to exist for this whole population or subset of it. The residuals about the predicted values never exceed 1.50 m. Latitude, distance from the sea and rainfall variables were also tested, but did not appear to contribute to any real predictive ability and so were discarded. Apparently, the height/age curves are wholly anamorphic with respect to altitude and so it is possible to be reasonably confident of predicting mean top height of Douglas fir at any age (or its site index, if such a measure has any utility) anywhere in the South Island, given the crop's starting height above age 10 and altitude within the range 50 to 750 m above sea level.

Net Basal Area per Hectare Projection

Several environmental, crop regime and dummy variables were examined within the framework of various functional forms of basal area projection equation. Again a modified Schumacher proved best and the inclusion of altitude improved the fit considerably. The addition of a thinning index and dummy regional variables also had a major impact on predictability. The final form found most appropriate was:

$$G_2 = G_1^{(T_1/T_2)^{\beta_0}} \exp(\beta_1 + \beta_2 K_1 + \beta_3 K_2 + \beta_4 K_3 + \beta_5 \text{ Altitude})(1-(T_1/T_2)^{\beta_0}) + \beta_6 X_t \left(\frac{1/T_1 - 1/T_2}{T_2 * T_t}\right) \quad (2)$$

where G_i = net basal area/ha at age T_i

T_t = age of thinning

K_1, K_2, K_3 represent dummy variables for the Nelson, Southland and Westland regions respectively, else Canterbury

β_i are coefficients estimated by non-linear least-squares

X_t, thinning index,

$$= \left(1 - \frac{\text{quadratic mean dbhob of trees removed in thinning}}{\text{quadratic mean dbhob of all trees just before thinning}}\right).$$

When an average Schumacher fit was made to the data (i.e. without the three dummy variables and X_t) the residuals extended ±14 m^2/ha about zero, whereas equation (2) above with its additional explanatory and locality variables was able to contain them all within ± 2.9 m^2/ha and 99 per cent of them within ± 2.2 m^2/ha. As can be seen from Table 2, precision for the individual regional equations was no better or even poorer than for the overall equation with dummy variables for region. There were also no apparent regional biases in basal area projection with the overall equation, whereas there were with the average one. The coefficients in the individual regional equations were less precise than the corresponding overall ones, and some were not statistically significant. The levels of improvement over an average fit and the absolute precision attained *per se* in the overall equation provide a level of sensitivity that managers should be relatively content to work with.

Table 2 *Regional and Overall Basal Area Growth Equation Statistics*

Region	β_0	β_1	β_2	β_3	β_4	β_5	β_6	RMS
Canterbury	0.976	5.222	-	-	-	0.000 25	0.201	1.550
Nelson	1.084	5.254	-	-	-	-0.000 29	0.116	1.191
Southland	1.076	5.562	-	-	-	-0.000 32	0.193	1.931
Westland	1.088	5.619	-	-	-	-0.000 90	0.148	1.076
Average[1]	1.058	5.576	-	-	-	-0.000 78	0.168	1.633
Overall[2]	1.059	5.272	0.064	0.285	0.312	-0.000 30	0.169	1.398

1 no dummy variables
2 with locality dummy variables

A common justification for the proliferation of models is that localities represent different growing conditions that lead to different site indices and different basal area growth trajectories. Several earlier research studies, therefore, have used site index as an independent variable with which to predict future basal area: e.g. Schumacher (1939), Clutter (1963), Bailey & Ware (1983), Murphy & Farrar (1988). Some of these also found that the Schumacher functional form proved best, generally

$$G_2 = G_1^{(T_1/T_2)} \exp (\beta_1(1 - T_1/T_2) + \beta_2(1 - T_1/T_2)^* \text{ site index}) \quad (3)$$

Our study has shown that altitude was a much more discerning predictor variable than site index and completely subsumed any impact that site index might have on predictions. Equation (2), even without the X_t variable, was clearly superior to (3). For any similar stocking regime, the basal area growth trends for single regions were reasonably parallel. If they had not been, then there might have been some justification for modelling regions separately, but that was not the case for this population; hence, it seemed better to model basal area growth overall.

The necessity for dummy variables to reflect different basal area growth trajectories for each of the four main regions suggests that there is an opportunity to find better environmental variables to explain them: the use of dummy variables meanwhile is an acceptable expedient. The standard errors for their coefficients were of the order of ± 3.5 per cent, precise enough to be categorical about confirming locality adaptations for stand basal area trends.

The addition of the thinning index, X_t, had a major impact on predicting future basal area per hectare, thus confirming the observations by García (1990) already mentioned. For intensively managed plantations in New Zealand, there is usually inventory information about basal area and stocking per hectare before thinning to allow X_t to be derived easily and routinely. While stocking density regimes for radiata pine in New Zealand are often specified only in terms of numbers of stems/ha, they are usually done in terms of basal area/ha for Douglas fir. The addition of the last term in equation (2), incidentally, did not contravene the important properties of consistency and path invariance that such growth models should possess, although it is not readily apparent at first glance and without some rearrangement of the last term.

Mortality

A modified Gompertz function was found to fit stocking survival best, with site index featuri as an explanatory environmental variable.

$$N_2 = N_1 \exp(\beta_1(T_2^2 - T_1^2)) + (\beta_0 + \beta_2 \text{ site index})(1 - \exp(\beta_1(T_2^2 - T_1^2))) \quad ($$

where N_i = no. of live stems/ha at age T_i

β_i are non-linear regression coefficients

No significant differences in mortality patterns were evident among regions, but the asymptc values were clearly related to site index: crops on higher site indices tended to suffer m mortality than on lower ones. As always, the mortality function was the weakest link in : whole set of functions, but residuals for equation (4) ranged between ± 30 stems/ha, wh: without the site index term, it was only within 200 stems/ha in 95 per cent of plots but went high as 400/ha at the very extreme. In this case, the addition of site index was enough sharpen the fit for all South Island sites considerably. Restricting the fits to individual regi resulted again in considerable loss of precision (see Temu, 1992), largely because of a smal range of site index in each sub-set of the data.

Stem Volume and Taper

A preliminary analysis of volume/ha was compromised by there having been a single taper a volume equation used in the sample plot data-processing system for calculating plot volun throughout the whole South Island. Casual observations and inferences that can be dra from considering equations (1) and (2) would suggest that this assumption was inappropri Sectional measurements used to derive the present tree volume and taper equations w supplemented with additional measurements, then analysed further for regional variati There were about 600 trees altogether and 7000 measurements of diameter and height in analysis. Volume was found to be best predicted from

$$v = \alpha\, d^{\beta_1 + \beta_2 K_1 + \beta_3 K_2} h^{\gamma} \quad ($$

where v = volume inside bark
d = diameter at breast height outside bark
h = total height
β_i are non-linear least-squares regression coefficients
K_1 and K_2 represent dummy variables for Nelson and Southland respectively, else Canterbury or Westland

The taper equation found best was a segmented polynomial function with 2 join poi compatible with (5) above,

$$d'^2 \left(\frac{kh}{v}\right) = \beta_1 Z^2 + \beta_2 Z + \beta_3\, d(Z - a_1)^2 I_1 + \beta_4\, d\,(Z - a_2)^2 I_2$$

where $Z = (h - h')/h$
h' = height at an intermediate point in the stem
d' = diameter inside bark at h'
$I_i = 1$ if $Z \geq a_i$, or else $= 0$, for $i=1,2$
a_1 and a_2 are join points
β_i are linear least-squares regression coefficients

The regional differences in the volume equation were sufficient, because of the volume taper compatibility constraints, to represent regional differences in taper; thus, locality dummy variables also in the taper equation were found to be redundant.

Residuals for the volume equation lay within ±0.20 m^3 while all predicted upper stem diameters were within ± 10 mm of the actual values and within ± 7 mm for 99 per cent of the observations. Again, this level of accuracy and precision was not achieved from regional equations on their own nor from the aggregate population without any allowance for locality (see Temu, 1992).

It has not yet been possible at time of writing to assess the benefits of using the revised volume and taper system for predicting plot volumes and the consequences of incorporating these revised estimates on the stand volume function that would be subsequently incorporated in the growth model. Preliminary indications are that they will be considerable.

CONCLUSIONS

This case study has shown that more reliable forecasting of growth and yield for large populations, using pooled data and environmental or dummy variables to provide locality adaptations that provide realistic disaggregations, is entirely feasible. It makes good use of all relevant data and avoids the limited perspective and inadequate coverage that is a feature of modelling restricted sub-populations. There is still a need for further research to find biologically sensible variables rather than locality dummy variables, though the precision of the latter is surprisingly good. The approach demonstrated and recommended here should lead to better consistency in yield forecasting over space and time than a strategy based on proliferation of available models. While the case study was for Douglas fir in the South Island of New Zealand, evidence is also now emerging which indicates that there is at least as good, if not better potential with this recommended approach to model the growth of radiata pine successfully.

ACKNOWLEDGEMENTS

This analysis could not have been attempted without the cooperation of the Forest Research Institute, Rotorua which stores and maintains the national sample plot and sectional measurement data base. Thanks are due also to Timberlands South for providing hard copies of the data. Financial support for the second author with a Commonwealth Scholarship is gratefully acknowledged.

REFERENCES

Bailey, R.L. and Ware, K.D. 1983. Compatible basal area growth and yield model for thinned and unthinned stands. Canadian Journal of Forest Research 13 (4): 563-571.

Beekhuis, J. 1966. Prediction of yield and increment in *Pinus radiata* stands in New Zealand. NZFS Technical Report No. 49. 40 p.

Clutter, J.L. 1963. Compatible growth and yield models for loblolly pine. Forest Science 9: 354-371.

García, O. 1990. Growth of thinned and pruned stands. NZ FRI Bulletin No. 151: 84-97.

Lewis, E.R. 1954. Yield of unthinned *Pinus radiata* in New Zealand. NZFS Fores Research Note 1 (No. 10). 26 p.

Murphy, P.A. and Farrar, R.M. 1988. Basal area projection equation for thinned natura even-age stands. Canadian Journal of Forest Research 18 (4): 827-32.

Schumacher, F.X. 1939. A new growth curve and its application to timber yield studies Journal of Forestry 37: 819-820.

Temu, M.J. 1992. Ph.D. thesis, University of Canterbury, Christchurch, N.Z. (*I, preparation*).

Whyte, A.G.D. 1990. Forecasting and monitoring yields form low-density plantations. N FRI Bulletin No.151: 77-83.

Appendix C

Diskette - Data Files

The attached diskette consist of original data files, SAS files used to develop the equations and their outputs. The names of these files and variables are as described in chapters 3 and 4, and elsewhere in this thesis.

The files are compressed into an executable file called **TEMU.EXE**. To decompress the files create a working directory and then type **TEMU** at prompt, the files will be copied into that directory.

Appendix D

Diskette - Yield Simulation model

The yield simulation model consists of two files namely, DFIRTREE.EXE and DFIRSTAN.EXE.

DFIRTREE.EXE is a simulation model for the volume based compatible taper estimation system for Canterbury, Nelson and Southland regions (DfirTree), to begin simulation exit to DOS and type DFIRTREE.

DFIRSTAN.EXE is a simulation model for the whole stand growth and yield model for Douglas fir in Canterbury, Nelson, Southland and Westland (DfirStand), to begin simulation exit to DOS and type DFIRSTAN.