

**forest volume estimation
and yield prediction**
vol. 2 - yield prediction

by

d. alder
commonwealth forestry institute, u.k.

The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Food and Agriculture Organization of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

M-35

ISBN 92-5-100923-6

The copyright in this book is vested in the Food and Agriculture Organization of the United Nations. The book may not be reproduced, in whole or in part, by any method or process, without written permission from the copyright holder. Applications for such permission, with a statement of the purpose and extent of the reproduction desired, should be addressed to the Director, Publications Division, Food and Agriculture Organization of the United Nations, Via delle Terme di Caracalla, 00100 Rome, Italy.

© FAO 1980

Table of Contents

	<u>Page</u>
1. <u>INTRODUCTION - THE PROBLEM OF GROWTH AND YIELD PREDICTION</u>	1
1.1 REASONS FOR PREDICTING GROWTH AND YIELD	1
1.1.1 Production Planning	1
1.1.2 Silvicultural Research and Planning	1
1.1.3 Ecological Research and Environmental Management	1
1.2 THE METHODOLOGY OF GROWTH AND YIELD PREDICTION	2
1.2.1 The Estimation of Growth and Yield	2
1.2.2 The Construction of a Mathematical Model and Its Fitting to Growth and Yield Data	3
1.2.3 Testing of the Model for Validity	3
1.2.4 The Application of the Model to the Required End-Use	4
2. <u>DESIGN OF YIELD PREDICTION STUDIES</u>	5
2.1 SPECIAL FEATURES OF REGRESSION PARAMETER ESTIMATION	5
2.2 SAMPLING DESIGN FOR MODEL CONSTRUCTION	5
2.2.1 Temporary Plots	6
2.2.1.1 Forest inventory	6
2.2.1.2 Growth estimation from annual rings	6
2.2.1.3 Sampling for allometric relationships	6
2.2.1.4 Sampling to define parameters of harvesting operations	6
2.2.1.5 Regeneration surveys	7
2.2.2 Permanent Sample Plots	7
2.2.2.1 Number of PSPs required	7
2.2.2.2 Location of PSPs	8
2.2.2.3 Size of PSPs	8
2.2.2.4 Shape of PSPs	9
2.2.2.5 Frequency and timing of remeasurements	9
2.2.2.6 Sampling with partial replacement	10
2.3 EXPERIMENTAL DESIGNS	10
2.3.1 Randomized Designs	11
2.3.1.1 Randomized block experiments	11
2.3.1.2 Factorial experiments	12
2.3.2 Systematic Designs	12
2.3.2.1 Single tree experiments	12
2.3.2.2 Clinal plots	13
2.4 <u>EXAMPLES OF GROWTH AND YIELD EXPERIMENTS</u>	14
2.4.1 Uniform Forests	14
2.4.1.1 Spacing experiments	14
2.4.1.2 Constant basal area thinning experiments	15

2.4.1.3	Thinning experiments using graded thinning treatments	15
2.4.1.4	Factorial experiments with different components of thinning treatment	16
2.4.2	Mixed Forests	16
2.4.2.1	Randomized block design	17
2.4.2.2	Treatment definition	17
2.4.2.3	Measurements and plot design	17
3.	<u>PROCEDURES FOR DATA COLLECTION AND PRIMARY ANALYSIS</u>	19
3.1	SAMPLE PLOT DEMARCATION	19
3.1.1	Location	19
3.1.2	PSP Identification on the Ground	19
3.1.3	Determination of Edge Trees	19
3.1.4	Marking Trees	20
3.1.5	Mapping Trees on the Plot	20
3.1.6	Identity Numbers for Ingrowth	20
3.2	SAMPLE PLOT MEASUREMENT FORMS AND PRIMARY ANALYSIS	21
3.2.1	Uniform Forests	21
3.2.2	Mixed Forest	24
3.2.3	Initial Assessment of Permanent Plots	25
3.3	STEM ANALYSIS PROCEDURE	26
3.4	SPECIAL METHODS OF TREE INCREMENT ESTIMATION	28
3.4.1	Simple Measurements	28
3.4.2	Remeasurement on Buttressed Trees	28
3.4.3	Girth Bands	29
3.4.4	Growth Ring Measurements	30
3.5	INDIRECT TREE AND DOMINANT HEIGHT ESTIMATION	31
4.	<u>DATA STORAGE SYSTEMS</u>	33
4.1	ADVANTAGES OF COMPUTER-BASED DATA STORAGE SYSTEMS	33
4.2	DATA VALIDATION	34
4.3	CONTRACTS FOR THE PREPARATION OF COMPUTER PROGRAMMES	35
4.4	STORAGE SYSTEMS FOR PLOT DATA	36
4.4.1	Introduction	36
4.4.2	File Structures	36
4.4.3	Error Checking and Editing Functions	37
4.4.4	Plot Summaries	38
4.4.5	Other Utilities	38
4.4.6	Data Base Security	38
4.5	DATA TRANSFER BETWEEN COMPUTER SYSTEMS	39

5.	<u>ANALYSIS OF GROWTH AND YIELD DATA FOR UNIFORM FOREST</u>	40
5.1	SITE CLASSIFICATION	40
5.1.1	Use of Dominant Height as an Indicator of Site	40
5.1.2	Construction of Site Index Curves	41
	5.1.2.1 Graphical methods of construction	41
	5.1.2.2 Mathematical methods of fitting site index curves	45
5.1.3	Site Assessment Models Based on Environmental Factors	61
	5.1.3.1 Functional models for site class prediction	61
	5.1.3.2 Construction and selection of environmental variables	62
	5.1.3.3 Problems in the application of site assessment functions	62
5.2	STATIC METHODS OF PREDICTING YIELD	63
5.2.1	Graphical Methods Based on the Diameter/Height Function	65
5.2.2	Direct Statistical Estimation of Mean Diameter Prediction Functions	65
5.2.3	Defining Treatment History in Terms of Stocking and Age	66
5.2.4	Static Yield Functions Predicting Basal Area or Volume	67
5.2.5	Limitations of Static Yield Models	67
5.3	DYNAMIC METHODS OF PREDICTING GROWTH AND YIELD	67
5.3.1	The Basal Area Increment Function	69
	5.3.1.1 Basal area increment as a function of dominant height	69
	5.3.1.2 Other methods of predicting basal area increment	70
	5.3.1.3 Practical problems in analysis of increment data	70
5.3.2	Constructing a Growth Model with Spacing Experiment Data: Marsh's Response Hypothesis	71
5.3.3	Conversion of Growth Models to Yield Models by Integration	73
	5.3.3.1 Introduction	73
	5.3.3.2 Basic theory	73
	5.3.3.3 Application of an integral yield model to different thinning treatments	74
	5.3.3.4 Example of use of an integral yield model	75
	5.3.3.5 Fitting compatible growth and yield models to increment data	77
5.3.4	Use of Growth Models by Simulation	78
	5.3.4.1 Requirements for a simulation model	78
	5.3.4.2 Method of construction of a simulation model	79
	5.3.4.3 Example of a simple simulation model	79
5.4	THINNING	86
5.4.1	The Thinning Ratio	86
5.4.2	Estimating Thinning Yields in Static Models	87
5.4.3	Estimating Thinning Yields in Dynamic Models	88
5.5	MORTALITY	88
5.5.1	Establishment Mortality	88
5.5.2	Density-Dependent Mortality	89
5.5.3	Disease and Pest Mortality	90

5.5.4	Windthrow and Fire Damage	91
5.6	STAND VOLUME PREDICTION	91
5.6.1	Stand Volume Equations Based on Dominant Height and Basal Area	91
5.6.2	Using Tree Volume Equations to Predict Stand Volume	92
5.6.3	Estimation of Volume to a Top Diameter Limit	92
5.6.4	Volumes of Thinning	93
5.7	ADVANCED TECHNIQUES OF GROWTH AND YIELD PREDICTION	93
5.7.1	Size Class Models	93
5.7.2	Tree Position Models	94
6.	<u>ANALYSIS OF GROWTH AND YIELD DATA FOR MIXED FOREST</u>	95
6.1	SITE CLASSIFICATION	96
6.2	STATIC YIELD FUNCTION FOR MIXED FOREST	98
6.2.1	General Principles	98
6.2.2	Data Analysis Procedures	99
6.2.3	Methods of Selecting a Yield Equation	99
6.2.4	Problems with Static Yield Models in Mixed Forest	100
6.2.5	Conclusions regarding Static Yield Models in Mixed Forest	100
6.3	TRANSITION MATRIX MODELS	101
6.3.1	Definition of a Transition Matrix Model	101
6.3.2	Methods of Constructing Transition Matrices	102
	6.3.2.1 Single tree data	103
	6.3.2.2 Size class data	103
6.3.3	Refinements to Transition Models	105
6.3.4	Disadvantages of Transition Models	105
6.4	DISTANCE-INDEPENDENT TREE MODELS BASED ON DIFFERENCE EQUATIONS	106
6.4.1	Definition	106
6.4.2	Allometric and Dynamic Variables	106
6.4.3	Representation of Competition	106
6.4.4	Data Requirements and Approaches to Analysis	107
6.4.5	Basic Model Structure	107
6.4.6	Ingrowth, Mortality and Harvesting	108
6.4.7	Conclusions regarding Tree Models	108
7.	<u>VALIDATION OF GROWTH AND YIELD MODELS</u>	109
7.1	THE ROLE OF VALIDATION	109
7.2	VALIDATION DATA	109
7.3	RESIDUAL ERRORS	110
7.4	GRAPHICAL COMPARISONS	111
7.5	DEFINING THE LIMITS OF MODEL UTILITY	112

8.	<u>THE APPLICATION OF THE MODEL TO THE REQUIRED END USE</u>	113
8.1	INTRODUCTION	113
8.2	EVEN-AGED STANDS	113
8.3	MIXED STANDS	114

FOREWORD

There is probably little argument among forest managers that the ability to estimate the volume of trees and stands and to predict what the forest will produce, on different sites, in response to particular types of silvicultural treatment, is central to all rational planning processes connected with forestry. There is, however, a considerable diversity of opinions over what constitutes "yield", and how it may be estimated and projected into the future.

This manual is an attempt to codify current practices in the field of tree and stand volume estimation and forest yield prediction in a way that is practicable and useful to the person who is charged with the responsibility of producing volume estimations and yield forecasts, but perhaps has not had the benefit of extensive experience in this field.

It must be appreciated, however, that this is a field of human endeavour that is currently in a state of rapid evolution, especially with regard to forests growing in tropical environments. Consequently, all that is said in this manual must be regarded as provisional and subject to future refinement for particular situations that can arise, or new techniques that can be developed, whilst other techniques may exist which are not referred to in this text and which may be superior for particular purposes.

Thus it is not a manual in the true sense; it is rather a set of guidelines for the choice of procedure combined with more precise instructions concerning calculation technique for some specified cases.

The manual is done with special reference to the tropics and applies to natural as well as man made forests. Because of the great difficulties in assessing growth and yield of natural mixed and uneven aged forests, the methods given to construct growth models, however, mainly apply to even aged forests. For mixed forests no specific instructions are given but rather some examples of possible ways of dealing with the problem.

The manual consists of two volumes. The first volume describes techniques of measuring trees and the assessment of volume of trees and stands, and the second volume deals with growth and yield prediction. Descriptions of statistical and mathematical techniques, selected statistical tables, blank copies of calculation and data recording forms and an annotated bibliography are included in a series of appendices.

Volume I of the manual has been written by Francis Gailliez, Centre Technique Forestier Tropical (CTFT), Nogent-sur-Marne, France, and Volume II by Denis Alder, Commonwealth Forestry Institute (CFI), Oxford, Great Britain, who also compiled the appendices. The work of the two authors has been coordinated by Jöran Fries, Swedish University of Agricultural Sciences, Uppsala, Sweden. The work was formulated and guided by Jean-Paul Lanly and Karn Deo Singh of the Forest Resources Division of FAO. Jean Clement (CTFT) was associated at the initial stage of the study.

The first draft of the manual was presented at the meeting of the IUFRO Subject Group S4.01 (Mensuration, Growth and Yield) held in Oxford in September 1979, and was discussed for one full day in detail. Among the participants there were tropical forest mensurationists especially invited by FAO to make a thorough and critical review of the contents of the manual. In addition, the manual was also sent to a number of specialists for comments. Based on these remarks, a revised version of the manual was prepared by the authors concerned.

This manual, being the first of its kind in the field of tropical forestry, has considerable scope for further improvements and additions. Particularly in the case of mixed uneven aged stands further complementary studies are immediately needed. All suggestions in this respect will be very much appreciated.

M.A. Flores Rodas
Assistant Director-General
Forestry Department

1. INTRODUCTION

THE PROBLEM OF GROWTH AND YIELD PREDICTION

1.1 REASONS FOR PREDICTING GROWTH AND YIELD

In order to understand the reasons for the diversity of methods that are being used for growth and yield prediction, it is useful to examine in more detail why growth and yield prediction is necessary.

1.1.1 Production Planning

Effective forest management involves the use of treatment regimes for control of the growing stock in such a way that the increase in the economic and/or social value of the growing stock is more rapid than the interest accumulating on the cost of treatment.

At the same time, all harvesting operations will deplete the future growing stock to a greater or lesser degree. Too heavy a rate of harvesting will ultimately liquidate the forest resource; too low a rate may both deprive a community of immediate resources and reduce potential growth in the forest for the future.

Clearly, rational decisions about treatment and harvesting intensity and timing can only be made if the response of the forest to these operations can be quantified. Growth and yield studies are the means to this end.

1.1.2 Silvicultural Research and Planning

Although the primary objective of growth and yield studies is probably the quantification of forest production in response to treatment and harvesting, there is a strong two-way relationship between growth and yield studies and more qualitative forms of silvicultural research. It is a two-way relationship because:

- (i) The silviculture of a species may determine the type of model that can be used to predict its growth; and may permit a logical basis for grouping species in complex forests. Important silvicultural features of a species will suggest features and relationships that must be included in a quantitative model if it is to be effective.
- (ii) A quantitative model can be used, if it contains suitable relationships, to test silvicultural hypotheses and to suggest experimental designs and treatments that are likely to provide useful results.

1.1.3 Ecological Research and Environmental Management

Quantitative models for growth and yield prediction may interact with the needs of ecological research and environmental planning in several ways. For example:

- (i) A forest model may indicate the amount of light reaching the forest floor at different stages of the growth cycle.
- (ii) A forest model may readily be adapted to show the biomass and rate of production of the tree crop.

- (iii) The duration of the growth cycle during which the forest crop is available to large herbivores is important information for wildlife management.

On the whole, ecological modelling uses rather different techniques from those used in growth and yield studies in forestry. This is because the latter has, of necessity, to focus upon a very precise prediction of the geometric properties of the crop; whilst in ecology, it is possible to deal with populations and levels of an ecosystem as a whole. Furthermore, ecological models tend to concentrate on describing or explaining the main qualitative features of an ecosystem; a high level of precision is rarely either possible or necessary. On the other hand, forest models must be reasonably precise if they are to justify their existence and fulfil their aim.

The techniques of ecological modelling do provide a number of useful points of contact with forest modelling:

- (i) In mixed forest, species composition can be modelled by ecological population dynamics techniques.
- (ii) Where mortality, defect, or loss of growth is attributable to specific diseases or pests of known etiology, then this may be quantified as population dynamics model in which habitat information is derived from a forest growth model and tree growth is influenced by pest population levels.

In the future, these points of contact are likely to enlarge. In particular, the increasing interest in non-wood products from forests and the intractable problems of complex mixed species and age forests may be best accommodated by using modified forms of ecological energy flow/nutrient cycle models.

1.2 THE METHODOLOGY OF GROWTH AND YIELD PREDICTION

The methodology of growth and yield prediction may be thought of as containing four main phases, which are discussed in the following paragraphs.

1.2.1 The Estimation of Growth and Yield

The estimation of growth or yield involves two kinds of problem. One is the problem of definition of what constitutes yield. This may be the timber volume of the crop, or it may be the timber volume of a particular group of species, or it may be some non-timber product, such as bark, foliage, resins, etc.. The most common emphasis in tropical countries is on predicting the volume and assortment of all wood products, including timber, pulpwood, poles and fuelwood. Because the species composition affects the utility of the product, yield cannot be considered apart from species composition in mixed stands.

The measurement of yield and of growth is relatively easy once appropriate definitions have been made. The main difficulties are practical ones associated with access to the forest, demarcation and measurement of plots and the maintenance of permanent plots over long periods of time. These problems are dealt with in Sections 2 and 3.

1.2.2 The Construction of a Mathematical Model and its Fitting to Growth and Yield Data

Once data is available, a mathematical model can be constructed and fitted to this data. A mathematical model consists of sets of equations or graphs showing the quantitative relationships between the variables.

The process of fitting the model may be statistical, using for example linear regression, or it may be subjective, by drawing lines through data plotted on graphs. Such hand-drawn curves can be subsequently expressed as equations if required. Appendix A of this manual gives a number of methods of curve fitting.

The types of curves drawn or equations fitted may be based upon some natural law of growth, or it may be empirical. In the latter case, the function or equation is chosen solely for its ability to represent a particular shape.

At the present time, there is no genuine function for tree growth based upon a natural law, although there are some, such as the Chapman-Richards function (described in Section 5) which are partially representative of growth processes.

1.2.3 Testing of the Model for Validity

Once a model has been constructed and fitted to data, it must be tested to determine its validity and precision. This is best done with a second set of data which was not used to fit any of the functions in the model. The model is used to predict the behaviour of the stands which produced the test data and the results are compared with the actual observations. It is often necessary to repeat this process of validation a number of times, with adjustments or corrections to the model as a result of apparent anomalies showing up at each stage.

There are a number of reasons why models can perform badly when validated:

- (i) The original data set may represent a different pattern of growth behaviour to the test set.
- (ii) Inappropriate methods of fitting the equations may have been used in model construction.
- (iii) Some of the functions may be extrapolated during the test with the validation data into a region where they are inaccurate.
- (iv) If the model involves a system of equations, it may become unstable when treated as a whole, even though each function separately may fit the data adequately.
- (v) There may be various kinds of human error during transcription or application of the various equations or graphs.

These points are spelt out at some length in order to emphasise the importance of thoroughly testing any model before applying it to planning or research. Section 6 deals with model validation in detail.

1.2.4 The Application of the Model to the Required End-Use

Essentially, the growth and yield model may be applied in one of three ways:

- (i) As a simple table or graph or set of tables or graphs. These can be used by forest planners directly or can be fed in in tabular form to a computer for updating a set of inventory data.
- (ii) As a programme for a computer or calculator which can produce a table or graph of growth and yield for a particular set of treatments. This is appropriate when the model has sufficient inherent flexibility so that it is not possible to define all possible predictions in one set of tables.
- (iii) As a computer programme which forms a sub-model within a larger computer programme for forest planning and which will usually incorporate a data base of inventory information and various economic or technical constraints on harvesting and treatment operations.

2. DESIGN OF YIELD PREDICTION STUDIES

2.1 SPECIAL FEATURES OF REGRESSION PARAMETER ESTIMATION

Sampling or experimental designs for growth and yield studies should be conditioned by the type of model to be fitted to the data obtained. This is usually some sort of regression model.

The following points should be borne in mind:

- (i) When the model to be fitted is known to be linear in form, then sampling or experimental treatments should be concentrated at the two extreme ends of the line. For a surface relating three variables, the four extreme corners should be sampled.
- (ii) More usually, the precise shape of the function to be fitted is unknown and likely to be somewhat curved. In this case, a good practice is to break the range of the predictor variable into 5 equal sections and sample each section at the same intensity (subject to (iii) below).
- (iii) Sampling intensity in any part of the range should be proportional to the variance of the samples around the model. This is particularly relevant when predicting volume, as discussed in Part I of this manual.
- (iv) Random or systematic sampling by area, which is most appropriate for forest inventory, is inefficient as a basis for constructing growth and yield models, as it involves too high a sampling intensity in the central part of the range of response and too low an intensity at the extremes of response.
- (v) In experimental designs for growth and yield prediction, extreme treatments should always be incorporated, especially with respect to stand density. This will add greatly to the accuracy of the model which can be fitted to the resultant data.

2.2 SAMPLING DESIGN FOR MODEL CONSTRUCTION

Sampling is an alternative to experimentation in situations where the variables entering the model cannot be controlled by the research worker. In growth and yield studies, this proviso applies principally to site variation. Forest type can be controlled by selection of the experimental area, or by establishment of the desired type of forest; stand density can be controlled by silvicultural and harvesting operations.

Experiments are generally more efficient and hence less expensive for a given accuracy and precision of prediction than sampling. However, both types of data are necessary if site variation is to be effectively included in the model.

The real effects of harvesting operations are also very difficult to simulate experimentally and must usually be determined by a sampling programme, carried out shortly after harvesting.

2.2.1 Temporary Plots

Temporary plots are primarily used for estimation of relationships which are not time dependent. However, this distinction is blurred by the possibility of determining time-dependent relationships from annual ring information in situations where these are present.

2.2.1.1 Forest inventory

Forest inventory designs are primarily determined to give an accurate estimate of forest growing stock in relation to land area. However, much of the information gathered may be useful in growth and yield studies.

The general subject of forest inventory design and analysis is covered in the FAO Manual of Forest Inventory.

It is generally inefficient to require the measurement of parameters on all forest inventory plots that are only required for growth and yield prediction. It is better to select a subset of plots for more detailed measurement.

2.2.1.2 Growth estimation from annual rings

Where clear annual rings are present, then studies on temporary plots can be used in place of permanent sample plots. In general, the use of annual rings for increment estimation is more difficult and more expensive than the use of permanent sample plots. On the other hand, results are obtained much more quickly.

2.2.1.3 Sampling for allometric relationships

An allometric relationship is one between one measurement on a tree and another. For example, the relationship between crown diameter and bole diameter, or between total height and bole length. Allometric relationships may be important in some models. The necessary data is often not available in a suitable form from a forest inventory, so it becomes desirable to carry out a sampling programme to determine the relationship.

The basic sampling unit is usually the single tree, although for convenience plots may be laid out and all trees on the plot measured. The number of samples will depend upon the relationship being studied. A good general procedure is to analyse the data as sampling is being carried out and to terminate sampling once the required degree of accuracy has been obtained.

A tree volume tariff is a particular example of an allometric relationship.

2.2.1.4 Sampling to define parameters of harvesting operations

Most yield prediction models accept as inputs the formal specifications for intermediate or cyclic harvesting operations. It is possible to assume that the operation will be carried out as specified, or it is possible to carry out a sampling programme to examine the relationship between the theoretical specifications and the actual results.

Additionally, most models will require information about the harvesting treatment that are not covered by the specifications. For example, it is usually necessary to know the diameter distribution of removed stems or the relationship between the number of stems removed and the basal area removed.

Sampling carried out shortly after harvesting will provide information on these matters. Alternatively, concealed semi-permanent plots can be set up which are measured before and after the harvesting operation to give an accurate determination of trees removed.

A general feature of these studies is the need to use larger sized plots than other types of study. Typical figures would be:

Uniform forest	0.1 - 0.5 ha
Mixed tropical forest	5 - 10 ha

This is because real harvesting operations tend to have very heterogeneous effects associated with extraction tracks and loading areas.

2.2.1.5 Regeneration surveys

In mixed tropical forests or any other type of mixed age forest, or uniform forests being regenerated by direct seeding, estimates of regeneration may be an important part of a yield prediction model.

Plots for regeneration surveys need to be small. They may be subplots within conventional forest inventory plots or they may be based on a separate sampling scheme carried out 3-5 years after logging. Typically, the plots are subdivided into quadrats; on each quadrat, the presence or absence of species is recorded. Actual counts of trees above a certain diameter or height may also be made, but do not usually add greatly to the usefulness of the information. Typical plot sizes are 0.01 ha (10 x 10 m) or 0.04 ha (20 x 20 m), subdivided in each case into 1 m² or 4 m² quadrats.

2.2.2 Permanent Sample Plots

Most foresters would consider the data obtained from permanent sample plots (PSPs) to be the most important contributor to a growth and yield model. Although this remains true for many situations, experiments must be considered as a necessary adjunct to introduce extremes of treatment that are not found in the forest; whilst measurements on annual rings provide an alternative to PSP measurements.

2.2.2.1 Number of PSPs required

It is not possible to define the number of PSPs required from purely statistical criteria. The precision of a model fitted to PSP data will depend upon the location of plots, and the duration of remeasurement as well as on the covariances of the various predictor variables and coefficients in the fitted model.

Experience suggests however that approximately 100 plots covering the range of site variation and stand history may be sufficient in a given forest type or plantation species, unless there is evidence for distinctively different growth patterns on part of the geographic range.

2.2.2.2 Location of PSPs

Permanent sample plots should be placed with equal frequency in:

- poor sites
- average sites
- good sites

and

- low density stands
- stands of average density
- high density stands

and

- young or recently logged stands
- mid-rotation or midway through felling cycle
- at rotation age or at the end of the felling cycle.

This will probably result in an area distribution for plots which is quite uneven, and appear to be proportionately deficient in the average stands. However, this is the most efficient method of sampling to estimate regression parameters, as discussed in section 2.1.

The type of stratification implied above may not be possible, due to a lack of knowledge of forest growing conditions, in which case plots may be laid down systematically or using a geographical stratification to give equal area coverage. In this case, many more plots will be required than if the more effective type of stratification described above is used.

2.2.2.3 Size of PSPs

In general, the size of permanent sample plots is governed by forest type and the heterogeneity of stocking and species distribution.

In mixed tropical forest, a size of 1 ha is usually appropriate. This may conveniently be divided into 100 10 x 10 m quadrats.

In uniform forest, sizes around 0.05 ha are commonly used.

These figures may be varied a great deal for different circumstances. For experiments, larger sized plots are more normal. The table below gives some details.

PERMANENT PLOT SIZES

Forest type	Mixed	Uniform
Sample plots	1-2 ha	0.04-0.08 ha
Experiments (excluding surrounds)	1-5 ha	0.08-0.12 ha
Studies of real* harvesting operations	5-10 ha	0.1 - 0.5 ha

* As opposed to simulated operations, which come under the category of experiments.

These figures should not be treated too rigidly.

2.2.2.4 Shape of PSPs

Generally speaking, PSPs may be either rectangular or circular. In inventory, other shapes are used, e.g. crosses and clusters of circular plots, which have specific advantages for area sampling but are not particularly useful for PSPs.

Circular plots are faster to lay out than rectangular plots for sizes below 0.1 ha in open stands, or 0.05 ha in dense stands. Their use is also recommended in plantations as effective area is not related to the arrangement of planting rows.

Rectangular plots are more appropriate for plot sizes greater than 0.1 ha.

The ratio of length to breadth for rectangular plots can be altered as required. On steep topography, a high ratio, up to 5 to 1, is better, with the length running up and down the slopes. On level ground, a square plot has a smaller perimeter and will therefore be easier to demarcate and measure.

2.2.2.5 Frequency and timing of remeasurements

The frequency with which PSPs should be remeasured depends upon the growth rate of the trees. It is also useful to remeasure a new plot after a shorter than normal interval in order to benefit as rapidly as possible from the growth data the plot provides.

It should be noted that in general, the longer the interval between remeasurements, the more accurately the tree increments can be determined.

In an organization responsible for a large number of PSPs, it is a good idea to alternate remeasurements so that perhaps only one third of the plots are measured in any one year.

Remeasurement intervals can be given approximately as follows:

<u>Forest type</u>	<u>Measurement interval (yrs)</u>
Young plantations in tropics	1
Older plantations or other uniform forest in tropics	2-4
Mixed tropical forest	3-5
Temperate uniform forest	3-5

The timing of remeasurements should obviously take into account seasonal effects. If there is a definite growing season, measurement should be carried out after the growing season is finished, as timing is less critical. In any case, a given plot should always be measured in the same month when annual measurements are made, to allow exact one year comparisons and increment estimation. With longer remeasurement intervals and less seasonal climates, timing becomes less critical.

In some cases, timing of measurements may be restricted by access considerations or seasonal availability of labour.

2.2.2.6 Sampling with partial replacement

Strictly speaking, sampling with partial replacement is an inventory design where semi-permanent plots are used to supplement information from temporary plots. However, the same general concept can be applied to permanent sample plots.

The more times a PSP is measured, the less information it provides compared with the previous measurement, unless it is growing into an age-site-stand density stratum that has not been well sampled.

For uniform age forests, two basic sampling strategies arise with PSPs:

- (i) Plots are established through all age classes. In this case, sampling is more efficient if a proportion of plots is replaced after the third or fourth remeasurements.
- (ii) Plots are established in young plantations only (because no older age classes exist). In this case, a proportion of the plots, say 30%, must be retained throughout the rotation. The remaining 70% are replaced after 3 or 4 measurements.

In mixed forest, an analogous situation exists, except that instead of age, one is concerned with the number of years since the last harvesting operation.

2.3 EXPERIMENTAL DESIGNS

Experiments are the most efficient and useful source of data for constructing models of growth and yield. However, the usefulness of many experiments that have been laid down in forestry research is limited by a failure to clearly envisage the mathematical model that the experiment is designed to test or parameterize.

Growth and yield studies are not primarily concerned with determining significant differences between treatments, but with constructing response surfaces. It is also a common error to assume that long term forestry experiments can provide useful solutions to problems that have their origins in current and highly fluid economic conditions. It almost always happens that by the time the experiment starts to provide useful data, economic conditions have changed so that the results are irrelevant.

The solution to this problem is always to set up the experiments with the intention of defining general principles, via a mathematical model, rather than to select the 'best' of a set of treatments.

Similarly, the parameters to be measured should never be defined in economic terms, but always in ecological or silvicultural terms.

Short term experiments that are defined in terms of economic parameters are necessary for the costing of silvicultural treatments and determination of utilizable yields in relation to specific harvesting methods, but these matters are outside the scope of this present manual.

The principal experimental designs that are relevant to growth and yield research can be subdivided into randomized or systematic designs. The former can be subject to conventional analysis of variance, whilst the latter are usually satisfactory when regression is the principal method of analysis and may be more economical than randomized designs to set out on the ground. Randomized designs can however also be analysed by regression and are therefore probably to be preferred in all cases except plantation spacing experiments.

2.3.1 Randomized Designs

A principal feature of randomized designs is the allocation of treatments to the plots within an experiment by some random process, usually a table of random numbers.

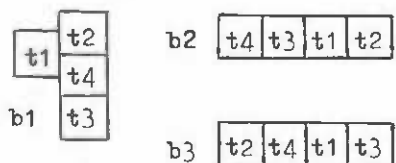
Another feature of randomized designs is the principle of replication. Any treatment must be applied at least twice, on two different plots.

These two features are the common characteristics of a wide diversity of experimental designs, including fully randomized experiments, lattices, latin squares, incomplete blocks, split plots, etc.. Standard textbooks, such as Snedecor (see bibliography) cover the analysis and appropriate usage of such designs. Dawkins, in his well known book of Statforms, gives calculation pro formas for many such designs.

In growth and yield studies, probably only two randomized designs are widely appropriate. These are the randomized block and the factorial experiment.

2.3.1.1 Randomized block experiments

The structure of a typical complete randomized block experiment is illustrated below.



The treatments, of which there may be any number (provided there are at least two), are designated by t1, t2, etc.. These are grouped into blocks, labelled b1, b2, etc.. There may be any number of blocks. Each block contains one replicate of each treatment.

The blocks are laid out so that the variations in site or forest condition are small within the blocks, compared to the variation between blocks.

The plots within a block do not necessarily have to be physically adjacent, as shown above, but usually the plots are relatively close compared with the distance between blocks.

The treatments are assigned to each plot within a block using a table of random numbers.

An incomplete randomized block arises when one or more treatments are not replicated in one or more blocks. This may be a deliberate feature of design, especially where there are large numbers of treatments, or it may be the result of loss of one or more plots through accidents. The analysis of variance of an incomplete randomized block experiment is somewhat more complex than for a complete one, but as far as regression studies are concerned, it does not make a great deal of difference.

In growth and yield experimentation, randomized block experiments are suitable for situations where the treatments do not form a clearly defined dimensional continuum. For example, if the treatments comprise initial spacing for a plantation crop, then this can be varied continuously and factorial or clinal design is more appropriate. But if the treatments are specifications for rainforest harvesting framed in terms of species groups and different cutting limits for different groups, then there is no clear continuum between treatments and a factorial or clinal design cannot be used. A randomized block design is therefore most appropriate.

2.3.1.2 Factorial experiments

Factorial experiments are intended for situations where the treatment consists of two or more interacting factors. For example, in a plantation thinning experiment, one might designate the age of thinning and the intensity of thinning as two separate factors.

Each level of a factor is combined with each level of every other factor. Thus if there are 3 levels of one factor and two levels of a second factor, there will be a total of six treatments.

All the treatments should be replicated at least twice. It is efficient to group the replicates into blocks, as this allows the variation between blocks to be accounted for.

Qualitative treatments such as for example pruning or no pruning, can be included in a factorial experiment as a treatment with two levels - present or absent.

Factorial experiments are well suited to studies in uniform forest involving timing and intensity of thinning, initial spacing, pruning and the use of fertilizers and weed control. They are more difficult to apply meaningfully in mixed forest because of the complex nature of the treatment definition and effects.

2.3.2 Systematic Designs

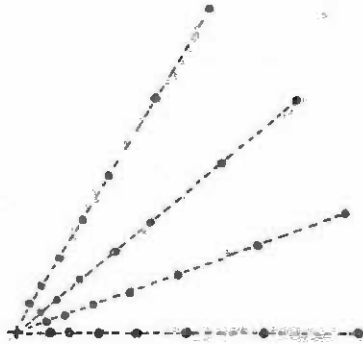
Systematic designs are those in which treatment locations are not randomized, but are laid out in some systematic pattern to economize on the size and cost of the experiment. Systematic experiments cannot be analysed by analysis of variance, but they are very efficient as a means of providing data for regression parameter estimation.

The main application of systematic designs is towards spacing experiments in uniform forest. In any situation where there is the slightest doubt as to the likely outcome of the experiment, e.g. fertilizer experiments, a randomized design should be used.

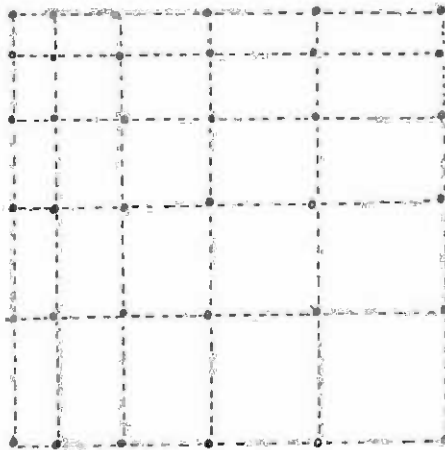
For spacing experiments, two basic design approaches are possible.

2.3.2.1 Single tree experiments

In the single tree systematic spacing design, the spacing varies between each tree in a continuous fashion. One well-known example is the Nelder fan, in which trees are planted along radii going out from a central point, with the distance between trees along radii increasing at the same rate as the distance between radii. The general appearance is illustrated below.



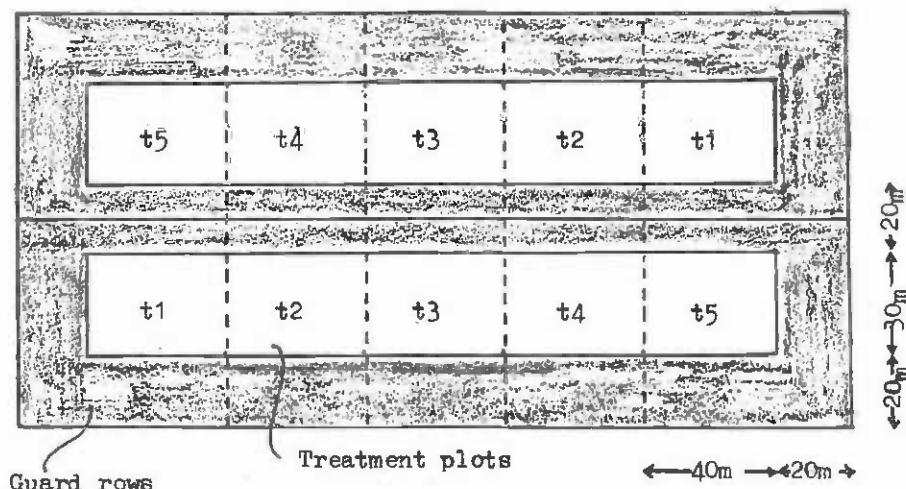
Nelder fans are a little difficult to plan and execute on the ground and some equally effective rectangular designs are possible. The diagram below indicates one in which spacing between trees is increased by half a metre for each tree, in both the vertical and horizontal direction. This plan has the advantage of testing all combinations of square and rectangular spacing and replicating each single combination twice. By using a smaller increment than $\frac{1}{2}$ metre, a larger experiment is obtained, which is less likely to be affected by the loss of single trees.



Single tree spacing experiments provide very useful data on the response of diameter and crown diameter to spacing for uniform forests, but the data is not easily compatible with that from conventional plots unless a single tree modelling strategy has been adopted. Single tree experiments are also very sensitive to the loss of trees, which upset their arrangement and design.

2.3.2.2 Clinal plots

Clinal plots are those in which the treatments are arranged so that successive levels are adjacent. The main advantage of clinal plots is the possibility of eliminating the plot surround except on the outside of the experiment. This is illustrated by the diagram below.



For spacing experiments, it is usually desirable to have from 5 to 8 treatments. The closest spacing should be either 2×2 m or $2\frac{1}{2} \times 2\frac{1}{2}$ m, depending on whether a narrow crowned species (e.g. Eucalypt or Pine) or wide crowned species (e.g. *Gmelina arborea*) is involved. The widest spacing should be 7×7 m or 8×8 m. It is very important to always incorporate these two extremes, as discussed in section 2.1.

Data from clinal plots can be combined with other plot data without difficulty. The plots themselves are less sensitive to the loss of individual trees than single tree experiments.

2.4. EXAMPLES OF GROWTH AND YIELD EXPERIMENTS

There are literally an infinite variety of ways in which experiments may be designed and executed to provide data for growth and yield prediction. The main purpose of this section is to present synthesized examples of the most common types of experiment, with some discussion of their usefulness and the special problems of execution and analysis. There is here a wide divergence in technique between methods for uniform forests of a single age and mixed forests.

2.4.1 Uniform Forests

Growth and yield studies in uniform forest have tended to concentrate on the effects of stand density, fertilization and pruning. Fertilizer and pruning experiments will not be considered directly in this manual as they are rarely relevant to the direct problem of yield prediction. Stand density is of primary importance, as it is the major variable which the forester is able to control during the rotation of a uniform forest. There are four basic ways in which the interaction between stand density and growth may be studied experimentally.

2.4.1.1 Spacing experiments

Spacing experiments may be laid out either as single tree experiments or as clinal plots. The latter is probably a more useful type of experiment. Between 5 and 8 different spacings may be used. It is better to thin the plots to their final spacing in the second

or third year of growth than to plant directly at the final spacing; the latter approach is too sensitive to poor survival following planting. When the thinning to the final spacing is carried out, it should be a non-selective thinning to achieve the desired spacing, not a low or crown thinning.

A spacing experiment should be continued over the whole of a normal rotation of the crop. Once established, it should not be thinned.

It is very easy to analyse spacing experiment data to provide a flexible growth model for the species in question. The details are discussed in section 5 of this manual.

2.4.1.2 Constant basal area thinning experiments

In a constant basal area thinning experiment, plots are laid out systematically or randomly and allowed to grow until each plot reaches its designated basal area. It is then thinned every one or two years to maintain the basal area at the designated level for each plot.

Careful records must be kept of the number of trees removed and the basal area removed if this type of experiment is to be very useful.

Analysis is more difficult than for a spacing experiment and involves the fitting of a model predicting basal area increment as a function of basal area and other variables, for example age and stocking.

As a strategy for constructing a dynamic thinning experiment, this approach has the advantage of simplicity of design, although the execution may not be so easy, as careful and continuous record keeping is required if the results are to be fully useful.

2.4.1.3 Thinning experiments using graded thinning treatments

In this type of experiment, a number of thinnings are defined (typically 4), which differ from each other in terms of both timing and intensity of thinning and possibly in terms of initial spacing also. These treatments can be roughly classified from light thinning to heavy thinning.

This type of experiment is very similar to a constant basal area thinning experiment in terms of analysis and record keeping. Records must be kept of the numbers and basal area of stems removed at each thinning. Analysis can be by several methods, the best of which is the fitting of a predictive model of diameter or basal area increment as a function of basal area, stem numbers, age, etc.

The principal advantage over a constant basal area experiment is that the thinnings used are treatments which might be used in reality, allowing costing of thinnings, thinning damage, windthrow risk and effects on wood quality to be assessed on the experiment. Additionally, thinning need not be performed so frequently as on the constant basal area experiment, simplifying administration.

2.4.1.4 Factorial experiments with different components of thinning treatment

A thinning treatment in a uniform forest can be broken down into a number of components, viz:

- initial spacing
- age of first thinning
- proportion of stocking removed at each thinning
- time to elapse between each thinning.

Other variations are possible, as for example the basal area at first thinning and at subsequent thinnings, or the use of height instead of age.

When the experiment is designed in this way, it can be set out as a factorial experiment by assigning a number of levels to each thinning component.

The result will be a large experiment, with a considerable number of treatment plots, but in the analysis it is possible to separate the different effects of thinning age, thinning intensity, etc.. Obviously the large number of treatments means that very careful administration is required, whilst the factorial nature of the experiment may be easily upset by accidental occurrences such as fire or disease outbreak.

2.4.2 Mixed Forests

The main function of experiments in mixed forest is to provide a controlled degree of disturbance of the forest so that ultimately a model may be fitted relating the increment of trees to the various parameters of the stand following treatment.

A common problem with experiments laid down in mixed forest is that the treatments are specified in terms which are irrelevant to the main parameters of the stand that control growth. This is discussed at length and with numerous examples by Synott (see bibliography). However, a typical case would involve the definition of four treatments as for example:

- 1 - Log all merchantable species down to 30 cm
- 2 - As for 1, but poison or ring-bark all non-merchantable species down to 30 cm
- 3 - As for 2, but killing non-merchantable stems down to 10 cm.
- 4 - As for 3, but removing merchantable species down to 20 cm.

The effect of treatments defined in this way upon the stand will depend entirely upon the condition of the stand prior to logging, its size class and species class distribution. It is possible for the most severe treatment (e.g. 4 above) to have the least effect.

Other problems have arisen by selecting only a subset of trees for increment measurement, on the basis of merchantability criteria which have an unfortunate tendency to change during the intervals between plot measurements. The method of measuring increment has often been unable to cope with the development of buttresses.

At the moment, growth and yield research in mixed tropical forests is developing rapidly. The following recommendations are made therefore simply to help avoid the mistakes of the past and not in order to impose a strait-jacket of unnecessary regimentation on current work.

2.4.2.1 Randomized block design

It is suggested that a randomized design, with replication, should always be adopted in mixed forest experiments. The replicates should be blocked with site and species distribution patterns, as well as past logging history, as uniform as possible within blocks. This implies a careful preliminary survey of the experimental area.

2.4.2.2 Treatment definition

Treatments should be defined in terms of total basal area to be left after logging and/or poisoning or girdling, of trees above a minimum size of 10 cm.

Other definitions of treatment are possible, but should always be made in terms of the remaining stand and not the material to be removed and should be independent of economic criteria.

The treatments adopted should always include two extremes, one being an undisturbed stand and the other being an extremely severe treatment, perhaps removing all material over 10 cm diameter.

2.4.2.3 Measurements and plot design

Large plots are required for experiments involving felling treatments. Typically a 200 x 200 m plot (4 ha) with a 100 m surround is necessary. On the main plot, all large trees (say over 30 cm) should be mapped. Subdivision of the plot into 20 x 20 m quadrats is desirable and allows local competitive effects to be included into models of growth, regeneration and mortality.

Detailed counts of seedlings can be made on a systematic subsample of quadrats.

The methods of increment measurement are discussed in section 3.

3. PROCEDURES FOR DATA COLLECTION AND PRIMARY ANALYSIS

3.1 SAMPLE PLOT DEMARCATION

3.1.1 Location

Permanent sample plots, in particular, need to be accurately located on forest maps and their precise position in the forest determined through the use of survey tape and compass. In addition to recording positions on maps, it is helpful to place stone or concrete markers on nearby forest roads showing the bearing and distance to the plot from the marker.

Temporary plots also need to be located on working maps, but the degree of precision required is not usually so great.

Design principles relating to plot location are discussed in section 2.2.

3.1.2 PSP Identification on the Ground

PSPs must be permanently marked on the ground. Circular plots should be marked at the plot centre with a post of durable wood, concrete or metal bearing the plot identification number. This central point should also be indicated by digging intersecting trenches 50 cm deep and approximately 2.5 metres long, with the point of intersection being the plot centre. This provides a permanent mark on the ground in the event of the plot centre post being lost or stolen.

Rectangular plots should similarly be marked at the four corners with posts, one of which carries the plot identification number. Trenches should be dug to intersect at the corner post positions.

The plot identity should also be prominently painted on a tree near to the centre or a corner post bearing the plot identity.

Rectangular plots subdivided into quadrats may also have smaller posts placed at the quadrat intersections. These should be of a different size to the corner posts to avoid confusion. Alternatively, quadrats may be resurveyed at each measurement.

3.1.3 Determination of Edge Trees

Most trees will be either clearly within the plot or clearly outside it. Some will intersect with the plot edge. These should be included if the estimated centre of the tree is inside the line demarcating the plot and excluded otherwise.

The edge of a circular plot is determined by using a line or rope stretched from the centre post. Care should be taken that the line does not have significant elasticity (many light nylon lines are very elastic) or become wrapped around the centre post whilst working round the plot.

The following table shows radii (i.e. distance from plot centre post to edge) for common plot areas of circular plots.

<u>Plot area (ha)</u>	<u>Radius (m)</u>
0.04	11.28
0.05	12.62
0.08	15.96
0.10	17.84

With small rectangular plots, the edge may be determined by taking a line of sight between two corner posts. With larger plots with intervening vegetation, it will be necessary to insert extra edge markers along a bearing between the corner posts. Small vegetation may be judiciously removed along the plot edge, provided one is not destroying regeneration that forms part of the forest growing stock.

3.1.4 Marking Trees

Trees on permanent plots should if at all possible be permanently identified. There are two ways of doing this:

1. By painting the identity number on the tree
2. By using embossed aluminium tags nailed to the tree.

If the tree number is painted on, then the measurement point for diameter should also be marked by painting a ring around the tree. If an aluminium tag is used it should be nailed to the tree a fixed distance above the diameter measurement point, usually 50 cm above, so that the latter can be exactly relocated. The reference point for diameter measurement is of course normally at 1.3 m initially, except for buttressed rainforest trees, but may change as ground level alters over time.

An aluminium tag should also be nailed to the stump of the tree, close to ground level. This helps to identify cut stumps and gives added security against loss of tags.

Paint markings should not be used as the sole means of identification with species that shed their bark (e.g. many *Eucalyptus* species). Markings should be renewed at each remeasurement where they are becoming worn or have been lost.

3.1.5 Mapping Trees on the Plot

If possible, trees on a plot should be mapped at the initial assessment. For circular plots, record the distance and bearing of each tree from the plot centre. For rectangular plots, subdivide the plot into quadrats, each of which is not more than 10 m by 10 m and then measure the distance of the tree from the two quadrat boundaries. Record these distances as the coordinates of the tree within the quadrat.

Mapping trees is helpful both in resolving the frequent confusion over tree identities that occurs and in the analysis of growth phenomena on the plot.

3.1.6 Identity Numbers for Ingrowth

In natural forests, ingrowth present at each assessment will need to be given an identification number, a quadrat number and coordinates on the plot map. Great care is necessary to ensure that the identity number given is not one previously assigned on that plot, including trees that have died or been removed at an earlier stage. Otherwise great confusion results when the data is processed.

3.2 SAMPLE PLOT MEASUREMENT FORMS AND PRIMARY ANALYSIS

3.2.1 Uniform Forests

Sample plots in even-aged, monospecific forests (normally plantations) require measurement of:

1. Diameter over bark at 1.3 m, using a girth tape calibrated in π cm units, on each tree.
2. Tree heights on a systematic sample of 8 trees, together with heights of dominant trees not included in the sample.

Dominant height is defined as the mean height of the 100 thickest stems per ha. Thus the number of trees required for dominant height estimation is the plot area times 100. For example, on a plot of 0.04 ha, 4 trees are required. Trees with broken or significantly damaged tops are not used in height sampling.

Additional characters can be noted on each tree. Disease problems, dying trees, wind or insect damage and trees marked for thinning can be recorded. Such additional notes should be coded in a rigidly standardized way and entered on the plot record form. The following coding suggestions may be adopted:

<u>Code</u>	<u>Description</u>
no entry	Tree healthy, undamaged, unmarked for thinning, single non-defective stem.
A	Animal damage.
B	Tree broken by wind.
D	Disease problem.
L	Double or multiple leader or stems.
M	Tree marked for thinning (but still standing).
S	Tree dying from suppression.
T	Tree has been felled (may be measured on the ground).
W	Tree leaning or fallen from wind damage.
X	Tree dead.

Following the letters, numeric codes can be placed indicating the degree of severity of the problem. The following scale is suggested:

- 1 Damage/disease present but very light.
- 2 More severe damage - is likely to significantly reduce growth or impede utilization.
- 3 Very severe damage/disease. Likely to kill tree or make it unutilizable.

Other codes can be introduced for specific problems. The important point is that for such a system to be of any use, it must be rigidly adhered to without alteration or omission over many years.

Form 3.1 is shown as a recording form for this type of plot. It is specifically designed to facilitate automatic data processing. Field entries should be made in soft pencil with a rubber being used to remove errors.

The primary variables calculated on a plot are:

1. Stocking per ha (N). Divide the total number of live trees on the plot by plot area. From the example entered in form 3.1:

$$N = 9 / 0.04 = 225 \text{ stems/ha.}$$

2. Diameter of the mean basal area tree (D_g). Sum the diameters squared and divide by the number of trees on the plot. Take the square root of the result. For the example:

$$\begin{aligned} \sum d^2 &= 16237 \\ D_g &= \sqrt{(16237 / 9)} \\ &= 42.5 \text{ cm.} \end{aligned}$$

3. Stand dominant height H_o . This is the mean height of the specified number of dominant trees on the plot. For the example,

$$\begin{aligned} H_o &= (32.1 + 29.6 + 30.8 + 33.1) / 4 \\ &= 31.4 \text{ m.} \end{aligned}$$

4. Stand mean height H . This is the mean of the systematic sample of height trees or as in the example on form 3.1, the mean of all heights. In the example,

$$H = 31.7 \text{ m.}$$

5. Stand volume V . This is usually calculated from an individual tree volume tariff entered by diameter and height. There are two methods:
 - (i) Calculate the volume of the tree of diameter D_g and height H and then multiply by stocking N to give volume in m^3/ha .
 - (ii) Calculate individual tree volumes v from diameter d and height h . Sum these volumes and divide by plot area to give volume per ha.

In this second case, if h is not known for all trees, either estimate it from a height/diameter regression (c.f. section 3) or use H instead.

Both methods introduce an error into volume estimation. The first one has an error that may result from the distribution of diameters and the second one from the estimation of heights. The second method is generally preferable for accuracy, especially if volume is being estimated to a merchantable top diameter limit.

Form 3.1 Sample Plot Assessment Form - Plantations

Sheet ...of...

Forest District Compartment Plot number Species

Plot area Slope Assessment date
 month year

Office use
no. of cards

Tree no.	Diameter bhob	Height	Codes		
<input type="text" value="11"/>	<input type="text" value="39.4"/>	<input type="text" value="28.6"/>			
<input type="text" value="16"/>	<input type="text" value="43.1"/>	<input type="text" value="32.1"/>			
<input type="text" value="13"/>	<input type="text" value="36.4"/>	<input type="text" value=""/>	<input type="text" value="B3"/>	<input type="text" value="X"/>	
<input type="text" value="15"/>	<input type="text" value="48.5"/>	<input type="text" value="29.6"/>			
<input type="text" value="19"/>	<input type="text" value="46.4"/>	<input type="text" value="30.8"/>			
<input type="text" value="21"/>	<input type="text" value="41.2"/>	<input type="text" value="31.3"/>			
<input type="text" value="24"/>	<input type="text" value="38.1"/>	<input type="text" value="34.2"/>			
<input type="text" value="4"/>	<input type="text" value="42.8"/>	<input type="text" value="33.8"/>			
<input type="text" value="9"/>	<input type="text" value="44.9"/>	<input type="text" value="33.1"/>			

Notes *Tree 13 struck by lightning? Broken 1/2 way up and dead*

Assessed by *D. Albin* Date *7/7/79*

USE BACK OF FORM FOR HEIGHT CALCULATIONS

Several different volumes may be calculated, using different volume equations. For example:

- V_{ob} Over bark total volume (i.e. to the tip of the tree).
- V_{ub} Under bark total volume.
- V_7 Volume to 7 cm top diameter limit.
- V_{15} Volume to 15 cm top diameter limit.

3.1. The various symbols and units used for primary stand variables are listed in Table

Table 3.1

SYMBOLS FOR PRIMARY STAND VARIABLES

<u>Symbol</u>	<u>Description</u>	<u>Units</u>
A	Stand age from planting	years
D_g	Diameter of the tree of mean basal area	cm
G	Stand basal area	m^2/ha
H	Stand mean height	m
H_0	Stand dominant height, defined as the mean height of the 100 largest diameter stems/ha	m
N	Stocking - number of live trees/ha	trees/ha
V	Stand volume. Type of volume is denoted by a subscript e.g. V_{ub} - underbark volume V_7 - over bark volume to 7 cm	m^3/ha

Measurements made immediately after thinning may be indicated by a prime ('). E.g. N' stocking after thinning. Removals in thinnings should be denoted by a subscript e. E.g. G_e would be basal area removed in thinning.

3.2.2 Mixed Forest

Primary analysis in mixed forest aims mainly to construct a stand table giving stem numbers grouped by size classes for each species or species group.

Once the stand table has been constructed, then various alternative measures of growing stock can be derived from it using different criteria. For example, one may wish to derive total basal area of trees in a given combination of species groups over a certain size; on another occasion one may repeat the summary using different species groups or size limits.

A stand table can of course be constructed directly as the plot is being measured, by recording only counts by size and species classes. This procedure is not recommended with any type of inventory except the most preliminary resource assessments. Individual tree dimensions and species should always be recorded, even though the intention may be to subsequently summarize the data. This is particularly important on permanent plots, since it allows individual tree increment estimates.

Measurement forms for plots in mixed forest may vary considerably, depending on the characteristics being recorded. The following types of situation are commonly found:

- (a) Mixed tropical montane forest (A few light demanding species of mixed age).

Form 3.1 may be used with the first 'code' column being reserved for species as a two digit or two letter code. A systematic sample of heights should always be measured, to allow a height/diameter curve to be constructed.

- (b) Tropical rainforest (Large numbers of mixed species, climbers, epiphytes, buttresses).

Large plots, over 1 ha, are normally used, subdivided into 10 x 10 m subplots or quadrats. Heights are not normally measurable, but trees may be classified by crown shape and crown position. Two reference diameters should be measured on stems forming buttresses. There are normally several hundred distinct species likely to occur on a plot.

- (c) Sub-humid woodlands (E.g. Miombo forest in eastern Africa).

Here tree form and length of merchantable bole are important characteristics. Height is easily measurable because of the openness of the forest, but only to the nearest metre because of diffuse crown shape. There are likely to be over 100 possible species present.

- (d) Arid-zone woodland

Trees are likely to be multi-stemmed with height and species being the only characters of importance. Height should be measured to the nearest decimetre.

Figure 3.1 shows alternative record formats for tree measurements for these four cases, together with the record format from Form 3.1 for plantations for comparison. A 'record' is assumed to be the amount of data that can be entered on one 80-column punched card, which represents the commonest medium for data input to a computer.

3.2.3 Initial Assessment of Permanent Plots

When a permanent plot is assessed for the first time, the following additional information is required:

- (1) The exact area of the plot, as a plane projection i.e. corrected for slope.
- (2) Basic site information including latitude, longitude, aspect, altitude, slope, slope position, forest history and past land use.
- (3) Meteorological information from the nearest weather station giving monthly precipitation and mean minimum and maximum temperatures.

- (4) The positions of all the trees on the plots. For circular plots, these should be recorded as a bearing and distance from the plot centre. For rectangular plots, these should be given as X and Y coordinates in decimetres from the most south-west corner of the plot.
- (5) Soil profile information including:
 - colour
 - texture
 - pH
 - analysis for N, P, K, Ca, Mg
 - depth
 - bulk density

for each discernible soil horizon. The soil pits or auger holes should be replicated on each plot. Two samples should be sufficient except on very large and variable plots. Analytical techniques may vary somewhat according to methods already in use in a particular country and special conditions encountered, and should be determined in collaboration with soil scientists. However, once a system for soil analysis has been settled upon, it should be modified only for very strong reasons and, at the time of modification, a series of samples (20 or 30) should be analysed by both the old and new methods to determine a regression relation to allow results from the old methods to be compared with the new ones.

3.3 STEM ANALYSIS PROCEDURE

Stem analysis refers to the reconstruction of the growth history of a tree by:

- (a) Felling the tree;
- (b) Cutting discs at intervals of around 2 m along the stem;
- (c) Careful counting and measurement of growth rings on the discs.

A great deal of information about stand dynamics can be gained from stem analysis, but in this manual the main concern is that of reconstructing the dominant height growth history of a stand.

The procedure is only possible in seasonal climates and with species producing clearly defined rings.

For height growth reconstruction, only the height of measurement and ring count need be recorded. With species and climates producing very unambiguous rings this can be done in the field with little difficulty.

Where rings are not so clear, discs must be cut, clearly marked in the field as to their point of origin on the tree and their orientation (i.e. which way up the disc was on the tree), and returned to the laboratory for assessment by one of two methods:

- (1) Planing and polishing the disc, followed by counting of rings along two axes using a microscope on a vernier rack and pinion mounting.
- (2) Cutting of two samples along a cross, with subsequent analysis by X-ray densitometry.

Figure 3.1

RECORD FORMATS FOR TREE MEASUREMENTS IN DIFFERENT FOREST TYPES

Plantation or other single species forest

1 ST TREE				2 ND TREE				3 RD TREE				4 TH TREE				5 TH TREE			
TREE NR.	DIAMETER B.H.O.B.	HEIGHT	CODES	TREE NR.	DIAMETER B.H.O.B.	HEIGHT	CODES	TREE NR.	DIAMETER B.H.O.B.	HEIGHT	CODES	TREE NR.	DIAMETER B.H.O.B.	HEIGHT	CODES	TREE NR.	DIAMETER B.H.O.B.	HEIGHT	CODES

Rainforest with buttressed trees

QUADRAT	1 ST TREE						2 ND TREE						3 RD TREE						NOT USED
	TREE NR.	SPECIES CODE	LOWER REF. DIAM.	UPPER REF. DIAM.	HEIGHT	CROWN SHAPE CROWN POS.	TREE NR.	SPECIES CODE	LOWER REF. DIAM.	UPPER REF. DIAM.	HEIGHT	CROWN SHAPE CROWN POS.	TREE NR.	SPECIES CODE	LOWER REF. DIAM.	UPPER REF. DIAM.	HEIGHT	CROWN SHAPE CROWN POS.	
			REF. HT. DIAM.	REF. HT. DIAM.					REF. HT. DIAM.	REF. HT. DIAM.					REF. HT. DIAM.	REF. HT. DIAM.			

Sub-humid forest (temporary plot format - trees not numbered)

1 ST TREE				2 ND TREE				3 RD TREE				4 TH TREE				5 TH TREE							
SPECIES CODE	DIAMETER B.H.O.B.	TOTAL HEIGHT	MERCANT. MILE HEIGHT	FORM CLASS	DEFECT CODE	SPECIES CODE	DIAMETER B.H.O.B.	TOTAL HEIGHT	MERCANT. MILE HEIGHT	FORM CLASS	DEFECT CODE	SPECIES CODE	DIAMETER B.H.O.B.	TOTAL HEIGHT	MERCANT. MILE HEIGHT	FORM CLASS	DEFECT CODE	SPECIES CODE	DIAMETER B.H.O.B.	TOTAL HEIGHT	MERCANT. MILE HEIGHT	FORM CLASS	DEFECT CODE

Arid-zone woodland. Multi-stemmed trees. (Temporary plot format - trees not numbered).

1 ST TREE			2 ND TREE			3 RD TREE			4 TH TREE			5 TH TREE			6 TH TREE			7 TH TREE			8 TH TREE			9 TH TREE			10 TH TREE		
SPECIES CODE	HEIGHT	NR. STEMS	SPECIES CODE	HEIGHT	NR. STEMS	SPECIES CODE	HEIGHT	NR. STEMS	SPECIES CODE	HEIGHT	NR. STEMS	SPECIES CODE	HEIGHT	NR. STEMS	SPECIES CODE	HEIGHT	NR. STEMS	SPECIES CODE	HEIGHT	NR. STEMS	SPECIES CODE	HEIGHT	NR. STEMS	SPECIES CODE	HEIGHT	NR. STEMS			

In either of cases (1) or (2), false rings or very faint rings should be checked against climatic records. With X-ray densitometry, which produces a numerical estimate of ring width and density, direct correlation with climatic variables is possible.

The age at which the tree reached a given height is given by the number of rings at the base of the tree minus the number of rings at the given height. From this, height can be plotted directly on age for the tree.

If the sample tree is a well-formed dominant, then this height-age curve may be regarded as essentially the same as a height-age curve derived from a permanent sample plot, and may be analysed in the same way.

3.4 SPECIAL METHODS OF TREE INCREMENT ESTIMATION

3.4.1 Simple Measurements

On permanent plots, tree increment is estimated by taking the difference between successive diameter measurements, divided by the measurement interval. For this to be an accurate procedure:

- (1) Individual trees must be clearly and uniquely identified on the plot.
- (2) The point of measurement on the tree must be precisely relocatable. Two alternative methods are possible:
 - (i) paint a ring at the point of breast height measurement. This may be excessively conspicuous under some circumstances;
 - (ii) insert a nail a precise distance above the point of measurement (50 cm is suggested) and relocate the point of measurement with reference to the nail. This nail may also bear the tree identification tag.

3.4.2 Remeasurement on Buttressed Trees.

When trees are developing buttresses, it is customary to measure diameter at a reference point about 1 metre above the buttresses. Since buttresses will extend between remeasurements, it is obviously necessary to move the reference diameter from time to time.

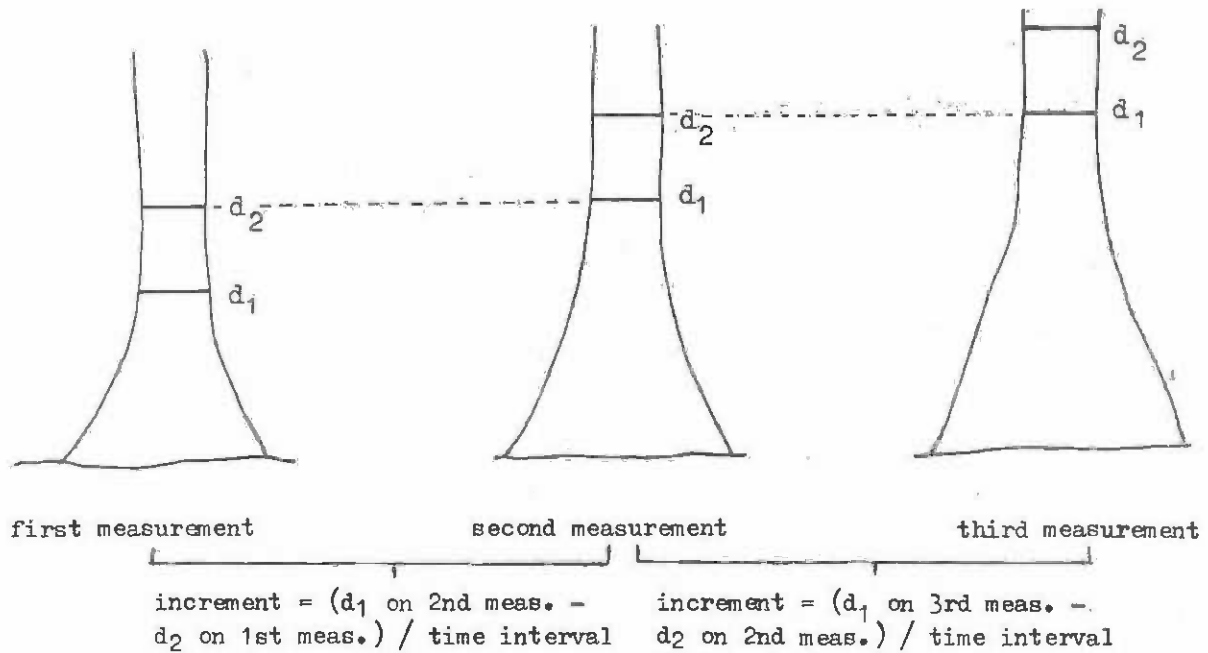
This procedure makes any increment determination from successive remeasurements impossible.

Two approaches may be adopted to counter this problem.

First Use two reference diameters at each remeasurement. Then, if it is necessary to move the lower reference diameter, it replaces the upper one and a new reference diameter is formed above the original top reference diameter. In this way, direct increment estimation is always possible.

It is recommended that the two diameters should be 1.5 m apart, with the lower one 1 metre above the top of the buttresses.

The process is illustrated in the diagram below.



Note that these reference diameters should be marked with paint or nails. It is essential to climb the trees with ladders and measure diameters with girth tapes. Optical instruments such as the Relaskop are not sufficiently accurate for increment estimation.

Second The other approach is to use girth bands, as described below, and make increment estimates over short periods (1-2 years) before buttresses can significantly influence the reference diameter.

3.4.3 Girth Bands

Girth bands may be constructed locally quite cheaply and are particularly useful for intensive research in, for example, thinning experiments. Without a vernier scale, they are accurate to $\pm \frac{1}{2}$ mm, which is quite adequate for increment estimation over 1 year periods on tropical species. With the addition of a vernier scale, the accuracy increases to ± 0.05 mm diameter. At this level, they can indicate physiological responses and seasonal growth fluctuations. (Conventional girth measurements by contrast cannot give increment more accurately than $\pm \frac{1}{2}$ cm diameter even on quite small trees and are much worse on very large trees.)

The basic requirements for construction are as follows:

- (1) Basic materials required are a roll of 1 cm wide stainless steel band, a large length of 1/3 cm coil spring, scales templates for the main and vernier scales, matt-black ceramic paint, a shaped hole punch for the spring fixings and a gas or electric oven.
- (2) Suitable lengths of band are wrapped around empty tins and painted on one side with the black paint. The tin, with the band, is baked to cure the paint.

- (3) A zero mark is scratched in the paint at one end with a metal point. If a vernier scale is to be added, it is included at this point, with the position of the graduations being defined by the vernier template.
- (4) The approximate girth of the tree must be known before adding the 'long' scale, about 10 cm actual length (equivalent to c. 3 cm of diameter), so that the band can be cut to the correct length and the scale put on at the correct point. The scale is scratched onto the band using the appropriate scale template.
- (5) Slots for the retaining spring are punched in the band, one outside the zero mark (& vernier scale), the other inside the long scale. The spring is fitted through one end, the band placed around the tree, taking care to remove loose bark and avoid sag at the back of the tree, and then the free end of the spring is slotted in. The spring should be cut to a length that gives a good positive tension to the band.

Increment is measured as the scale movement between the initial and final reading. The scales are normally in π cm units, giving direct reading of diameter. If the starting point of the long scale is placed an exact number of centimetres from the zero mark, the band also gives an absolute reading of diameter. If the long scale starts at an arbitrary point, then absolute diameter (as opposed to increment) should be measured with a separate girth tape just above or below the band.

Templates for cutting the bands can be purchased from forestry instrument suppliers, or manufactured locally in any well equipped workshop.

Various kinds of sophisticated girth bands, some equipped for telemetry, can be purchased directly. Because of their expense, these should only be used for the most intensive kinds of research and under close supervision of the experimental area against animal or human damage.

3.4.4 Growth Ring Measurements

Where growth rings are present, they can be used to estimate increment. The most reliable method is by the use of stem sections on complete felled trees, taken at d.b.h.

The width of the last three or four annual rings should be measured on two diameters at right angles on the section. These diameters should be along the major and minor axes if the section is elliptical. This gives the periodic underbark increment. It is necessary to construct a regression relating underbark diameter measured directly, with overbark diameter measured with a girth tape to convert the increments underbark onto a common scale with normal d.b.h. measurements.

Increment can also be estimated from samples bored from the tree. This is subject to numerous errors, especially in species with soft timber. The resultant core may be stretched or spirally compressed. It may not be accurately radial.

Because the difficulties in counting growth rings in tropical areas often require resort to calibrated microscopes or X-ray densitometre, complete sections should be regarded as essential. The use of increment borers is not usually a possible option.

3.5 INDIRECT TREE DOMINANT HEIGHT ESTIMATION

Because tree height estimation is a relatively slow procedure, it is not usually desirable to measure more than 8-10 trees on a plot. If the heights of all trees on the plot are required, then a height-diameter curve of the form

$$h = b_0 + b_1d + b_2d^2$$

can be constructed. Some special points should be noted:

- (1) The calculation method for this regression is given in Appendix 2.2, together with an example.
- (2) With more than one plot in a single stand, the sample trees should be pooled to fit the height/diameter regression.
- (3) Samples between stands of varying density or age should never be pooled for determination of the regression unless comparison of separate regressions (Appendix A 2.8) shows there is no significant difference.
- (4) The fitted regression should not be used to predict heights unless three conditions are satisfied:
 - (i) The regression F value is significant at the 95% level.
 - (ii) The b_1 coefficient is positive.
 - (iii) The b_2 coefficient is negative.
- (5) Once the regression has been calculated and tested to see if it is usable, mean height H can be estimated as the height predicted when D_g is substituted for d in the regression.
- (6) Similarly, dominant height H_0 is predicted when the mean diameter of the dominants (i.e. 100 largest diameter trees per ha), symbolised by D_0 , is entered in the regression for d .

Referring back to point (4), it should be noted that when these conditions are not satisfied, it is safer to assume that the individual tree heights h used for volume estimation are equal to the stand mean height H . This situation arises in many tropical species because the variation in tree heights is unrelated to diameter differences and the instrumental error involved in measurement is greater than the effects of diameter on height. Significant regressions and well-developed relationships between tree diameter and height are more likely to occur with more shade-tolerant species and at higher stockings per ha.

In mixed age stands, the regression will always be significant and can always be used for individual tree height determination. However, care should still be taken not to pool regressions for different stands without adequate statistical tests for the homogeneity of the data.

If the height-diameter regression is fitted to an age series of plantations (i.e. a series of stands of different ages) it will also be significant. It should be noted however that this is a different kind of model to the one fitted within a uniform age stand and will not reliably predict individual tree heights within one uniform stand (although it will predict mean height H as a function of mean diameter D_g provided that stand density is constant). Within a single-age stand, the regression reflects dominance differences. Between age classes, it reflects a time-dependent growth relationship.

4. DATA STORAGE SYSTEMS

4.1 ADVANTAGES OF COMPUTER-BASED DATA STORAGE SYSTEMS

A computer-based data storage system will store all permanent sample plot, experimental plot and temporary plot data on magnetic tapes or discs. This data can be accessed quickly for summarization and analysis. It can be subjected to automatic error checking procedures to pin-point doubtful measurements. It can be updated or corrected relatively easily.

Computer information storage (CIS) still requires maintenance of a conventional filing system on each permanent plot or experiment, in which original field sheets and any queries, notes, diagrams and procedural orders are kept.

Until recently, the capital cost of a computer system and the lack of skilled personnel inhibited the use of CIS. With the advent of the microcomputer, costs have fallen drastically and are now of the same order as those for a motor vehicle. Microcomputers usually operate in BASIC, which is a language designed for easy learning. Anybody with a suitably facile brain can teach themselves BASIC, given access to a microcomputer. In addition, some forestry institutions offer specialized training in data processing for graduate or technical staff.^{1/}

Lack of CIS is a significant inhibiting factor in the development and validation of yield models and effective yield planning and control. With manual extraction of data from files, only the simplest kinds of model can be constructed, whilst validation by residual analysis is not possible because of the stupendous amount of work involved. Nor is it possible to examine alternative modelling strategies on the same data or to update the model functions as new data is collected.

Furthermore, the productivity of skilled staff using manual procedures is extremely low. Collection of adequate quantities of data tends to be inhibited, because manual procedures are simply unable to cope with it.

It is strongly recommended that all forestry organizations should either:

- (a) Have access to a large computer facility, with an absolute maximum job turnaround of one day. This job turnaround should be considered from the time the results are returned and should take into consideration periods of the month when the main computer may be totally unavailable (due to priority allocation of computer time to other users), transportation difficulties to or from the computer, etc..

or

- (b) Purchase a microcomputer system with:
 - 32-64k bytes of memory
 - Twin drive diskette or hard disc system
 - A printer

^{1/} E.g. The Commonwealth Forestry Institute at Oxford, U.K.

- A teletype that can be interfaced to the computer, with paper tape reader/punch, for use in on or off-line data preparation
- A BASIC or FORTRAN compiler.

This type of system would cost around \$10 000 in Europe or North America at current prices (mid-1979).

Option (b) gives the forestry organization an in-house, dedicated computer that is likely to provide greater productivity and more rapid personnel training than option (a).

4.2 DATA VALIDATION

Errors arise in magnetically-stored data from the following sources:

- (1) Field measurement errors
- (2) Data entry or keypunching errors
- (3) Programme errors.

The third type of error must be assumed to be eliminated by thorough testing of any computer programmes used to store, correct, update, print or select data. The testing of programmes is the responsibility of the programmer, who is as directly responsible for the errors of his programmes as is the field worker who fails to make correct measurements.

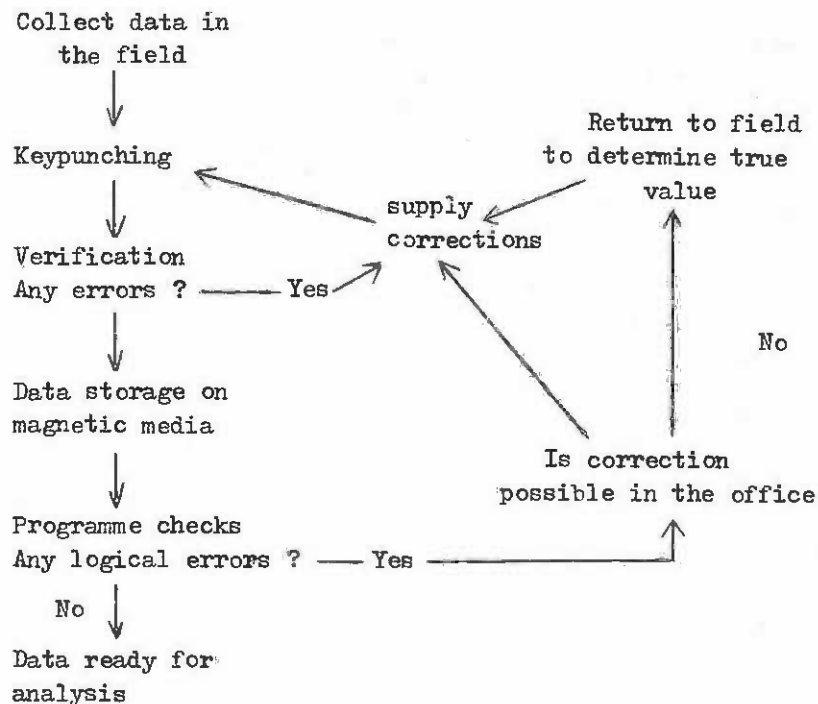
Data entry or keypunching errors are largely eliminated by the process of verification, which should always be used with any large mass of data. Verification involves entering every item of data twice, by two different operators or in two separate runs. The two data sets are then checked against each other, automatically, and any inconsistencies reported to the operator, who can supply an appropriate correction. The actual details of the verification process will depend on the data entry system used. Most data preparation organizations will offer verification as a normal service and this should always be explicitly stipulated when submitting data for keypunching.

Field measurement errors cannot be wholly eliminated, but they can be reduced:

- (a) By attention to the training and morale of field workers and by provision of suitable instruments for their use.
- (b) By running data checking programmes on the magnetically stored data. The data checking is based on examination for logical inconsistencies such as very large or very small measurements, negative increments in diameter or height, missing tree numbers or tree numbers which have reappeared since a previous harvesting, changes in species identification and so on.

Any such logical inconsistency is reported by the data checking programmes and must be examined to determine the likely source of the error. A correction must then be supplied, using a data editing programme, to amend the magnetically-stored data.

The entire process is an ongoing one, as indicated diagrammatically below.



4.3 CONTRACTS FOR THE PREPARATION OF COMPUTER PROGRAMMES

Contracts for computer programme preparation should contain clauses covering the following conditions:

- (i) Full provision by the contractor of source listings (in BASIC, FORTRAN or other language used) of any programmes written.
- (ii) Full documentation of all programmes, including:
 - (a) A dictionary of meanings for identifiers or variables used in the programme.
 - (b) Flow diagrams indicating the sequence of operation of any programmes or subprogrammes written by the contractor and the sequence and nature of data transfers from external media and magnetic storage.
 - (c) Explicit definition of the record structure and usage of all magnetic files and all input and output media.
 - (d) Explanation of theory behind the programme method, together with references to texts or other background material.
- (iii) Liability by the contractor for any programme errors or for any failure of the programmes to operate as specified when used in accordance with the documentation.
- (iv) Instruction by the contractor of some member of the forestry organization staff in the use of the programmes, up to the point where the programmes can be demonstrated to operate to the satisfaction of the forest management without any supervision by the contractor.

- (v) Copyright over all supplied documentation, programmes and reports to be vested in the forestry organization.

It may be possible to reduce the cost of a contract to some degree by relaxing points (iii) and (v), but the other points should always be insisted on if the programmes written are to be of any continuing use to the organization.

4.4 STORAGE SYSTEMS FOR PLOT DATA

4.4.1 Introduction

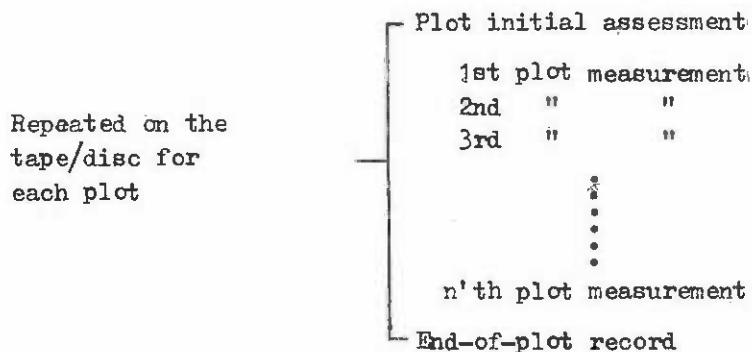
It is not the intention to provide detailed descriptions of programmes for storing and summarizing permanent plot, experimental plot or temporary plot data. These will probably need to vary a good deal according to the computing facilities available. Only the file structures for input and output and the types of functions the programmes can perform, will be described here.

4.4.2 File Structures

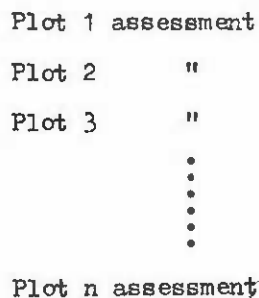
A file, in computing terminology, consists of some body of machine-readable information on magnetic media (tape or disc) or on punched cards or paper tape. Input files provide data for a particular programme and may themselves be output files from another programme.

File structure defines, in the present context, the organization of the different types of information that must be grouped together to represent a single plot.

For permanent plots, the basic information on a single plot may be defined as follows:



For temporary plots, the record structure is much simpler, being only:



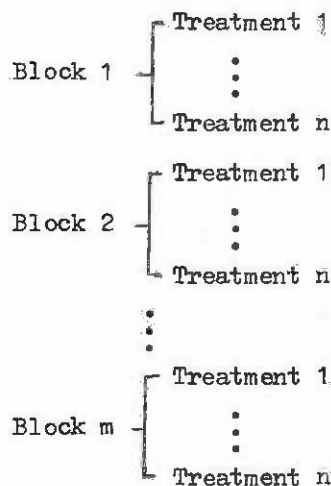
These record structures relate to the permanent data base, which may be on either magnetic tape or disc. For input of data for permanent plots, two types of different input may be possible:

- (i) Initial assessments
- (ii) Remeasurement data.

In case (i), a new plot would be created in the database, but the programme would need to check that plot identification did not conflict with an existing plot.

In case (ii), an existing plot would have to be present (otherwise an error would be reported), with the new measurement being at the appropriate point.

Experimental plots could normally be treated in the same way as permanent plots. However, related treatment plots would be placed adjacent to each other on the tape, for example as:



To retrieve all the data for the experiment from the tape, $n \times m$ plots would be read from the starting position on the tape (or disc) or the first plot in the experiment.

4.4.3 Error Checking and Editing Functions

The following error checks are required for permanent plot information:

- (1) That the plot exists on tape (if a new plot is not being added to the system) or does not yet exist (if a new plot is being added).
- (2) That the remeasurement dates are in a consistent and ascending sequence.
- (3) That trees which have been thinned (i.e. disappeared in past measurements) do not reappear on later measurements.
- (4) That trees do not change species (on mixed forest plots only) or plots do not change species (on plantation plots).

- (5) That diameter increments are positive and not excessively large.
- (6) Checking of the size ranges of parameters such as height, diameter, species code, etc. for excessively large or small values or unrecognised codes.

On temporary plots, only the checks in the 6th category are possible.

Data should be added to the database even though it contains errors; an editing programme may then be used to manipulate and amend the magnetically-stored data. With any significant amount of data, this is usually much more convenient than attempting to correct the original punched cards.

4.4.4 Plot Summaries

The plot summary programme may incorporate the error checking procedures (or they may be in a separate programme). Its main function will be to produce summarized information relevant to the particular forest type in question, for use in further analyses. For plantations, the summaries may include plot mean and dominant height, mean basal area diameter, basal area, stocking and volume. For mixed forest, the summaries will normally also include stand tables of selected characteristics by species and size classes.

The summary programme should be able to produce its output both in a printed form, with titling text to make it comprehensible and in machine readable form, without text, for direct input to data analysis programmes. The machine readable output may be put onto magnetic tape or disc or punched on paper tape or cards, as appropriate. If the output is onto magnetic tape, this should normally be done in 'formatted' or character-encoded form, rather than in machine binary code.

For mixed forests in particular, but also for plantations, it is useful to have a facility so that only selected parts of the output are printed or placed on tape.

4.4.5 Other Utilities

Two other programmes will probably be needed with a sample plot data base:

- A sorting programme, able to reorder the sequence of plots on tape so that they are grouped by forest, district, species, compartment, etc.. Normally, plots will be entered onto the data base in an arbitrary order, whereas the summaries will probably be preferred with some logical sequence.
- Archival and character-encoding programme, to transfer the main data base on disc or tape into a form suitable for archival (see below) or transfer to another computer site. This will involve character-encoding or formatting, the entire data base and writing it onto a magnetic tape.

4.4.6 Data Base Security

A large data base stored on disc or magnetic tape can readily be destroyed by accidents, programme or computer failures. It is essential therefore that after each addition of a significant amount of information, the entire data base is copied onto a spare tape or disc to give a complete second copy. These spare copies can be rotated, so that at any one time there is an up to date working version, an up to date archived version and two or three previous archived versions of the data base.

4.5 DATA TRANSFER BETWEEN COMPUTER SYSTEMS

The necessity frequently arises for transferring data between computer systems, either for cooperative research or to permit research workers from one organization to study elsewhere using their own data.

Large amounts of data are best transferred on industry-standard magnetic computer tapes. These may be 9-track or 7-track tapes. The following information should be ascertained when the tape is written:

- The number of tracks (7 or 9)
- The density (usually 800 or 1 600 bits per inch)
- The parity (even or odd)
- The inter-block gap in mm (or failing this the exact mark and manufacturer of the tape drive mechanism).

Tapes for transfer between computers should always be character-encoded or formatted and should preferably use fixed-length records of moderate size, probably less than 120 characters per record, to facilitate reading the tape. The type of character-encoding used (EBCDIC, ASCII, BCD, etc.) should be determined if possible but is not critical, as transliteration from one to the other is quite simple. A listing of the first and last few hundred lines of the tape should be sent with it to help check that when the tape is read, no records or parts of records have been lost.

Header labels and tapemarks are generally a nuisance when reading strange tapes, so it is better to write the tape as a single file, without a tape label at the start. The end of information is normally indicated by a double tape mark.

Information concerning the mode of the tape (tracks, density, parity, character code), the contents and the address of origin should be attached to the tape with a sticky label.

Tapes can normally be sent easily through the post. However, high frequency metal detectors may erase some or all the information on a tape. The parcel should therefore be clearly marked and easily opened for visual inspection.

5. ANALYSIS OF GROWTH AND YIELD DATA FOR UNIFORM FOREST

Uniform forests are those in which the main crop trees are of a known, uniform age. They are usually composed of a single species or a few ecologically similar species. Uniform forests are, by definition, managed under a clear felling system, with or without intermediate thinnings. Regeneration may be by planting, artificial or natural seeding or from coppice.

In this type of situation, the main parameters of growth and yield prediction are well understood. There is a wide variety of possible models available. The main limiting factor on the effectiveness of a model is usually the availability of data for the forests in question covering a wide range of sites, ages and stand densities.

The growth and yield of a forest can be modelled at three basic levels. These are the whole stand, the size class and the individual tree. For uniform forests, stand models are usually adequate for most purposes. Stand models are also very much simpler to both construct and use. Consequently, this is the modelling strategy dealt with most fully; the other two approaches are discussed briefly in section 5.7.

Even within this single level of stand modelling, there is a wide variety of choice in the particular set of functions to be incorporated in the model. Some alternatives are presented in the different sections which follow, together with an attempt to define situations in which a particular method is most appropriate. Other alternatives have been omitted simply because of the need to keep this manual reasonably concise.

On the whole, techniques are incorporated which are characterized by simplicity, accuracy and flexibility.

Within a particular aspect of stand modelling, such as for example the production of sets of basal area/height/stand density curves, an attempt has been made to include both graphical methods and statistical methods of differing degrees of complexity.

5.1 SITE CLASSIFICATION

The relatively high accuracy possible with growth and yield models for uniform stands results partly from the precision with which it is possible to classify site. This is itself a result of the fact that age is normally known from management records and height of the dominant trees can usually be measured by hypsometers or similar instruments.

5.1.1 Use of Dominant Height as an Indicator of Site

The height of a uniform stand, at a given age, is a good indicator of the potential productivity of that type of forest on that particular site. Hence the construction of height/age curves corresponding to different site classes is the first step in growth and yield model construction.

However, the mean height of a stand is usually sensitive not only to age and site class, but also to stand density. Consequently, dominant height is normally used in defining the height of a stand. Dominant height is almost entirely insensitive to stand density differences.

Dominant height can be defined in various ways, but the definition with the widest currency is that the dominant height of a stand is the mean height of the 100 thickest stems per hectare. Dominant height is also sometimes termed 'top height'.

Under some circumstances encountered in uniform forests in the tropics, dominant height ceases to be a good indicator of site class. This occurs with young stands of very fast growing crops and also with certain species which are notoriously variable in their height growth, such as Pinus caribaea. This situation can be detected by ranking permanent plot data by height within each age class. If the rank position of plots on successive occasions is poorly correlated, then any site class curves constructed must be considered of doubtful value.

The problem arises simply because of the great variability of height growth, relative to the effect of site, on these types of stands. It could be partly overcome by a redefinition of dominant height to require a larger sample of height trees per plot, e.g. equivalent to 200 or 400 stems/ha. An alternative idea is to correlate final productivity with environmental variables and use a site classification based purely on slope, altitude, soil type or other factors which appear to be significant.

5.1.2 Construction of Site Index Curves

The height-age-site index relationship is basic to uniform forest growth prediction. The relationship is usually referred to simply as the site index curves for a species in a given environment.

Construction of site index curves may be by graphical methods or by regression analysis.

5.1.2.1 Graphical methods of construction

Graphical methods of construction proceed as follows:

1. Plot all available height-age data for stands of the species in question. Dominant height should be used, not mean height, as it is much more independent of variations in stand density. Both temporary and permanent plots may be included on the graph. With permanent plots, the points from successive re-measurements should be joined with straight lines. This stage is illustrated in figure 5.1, which shows data from Pinus patula stands in Uganda.
2. Next draw curves by hand through the data. These should attempt to follow the trends of:
 - (i) The plots on the lower edge of the mass of data;
 - (ii) The median tendency through the data;
 - (iii) The upper edge of the data.

In each case the curves should follow as parallel as possible to the tendencies of permanent sample plots on that part of the graph. The drawing of these three curves is shown in figure 5.2.

Figure 5.1

DATA FROM PERMANENT AND TEMPORARY SAMPLE PLOTS READY FOR
CONSTRUCTION OF SITE INDEX CURVES

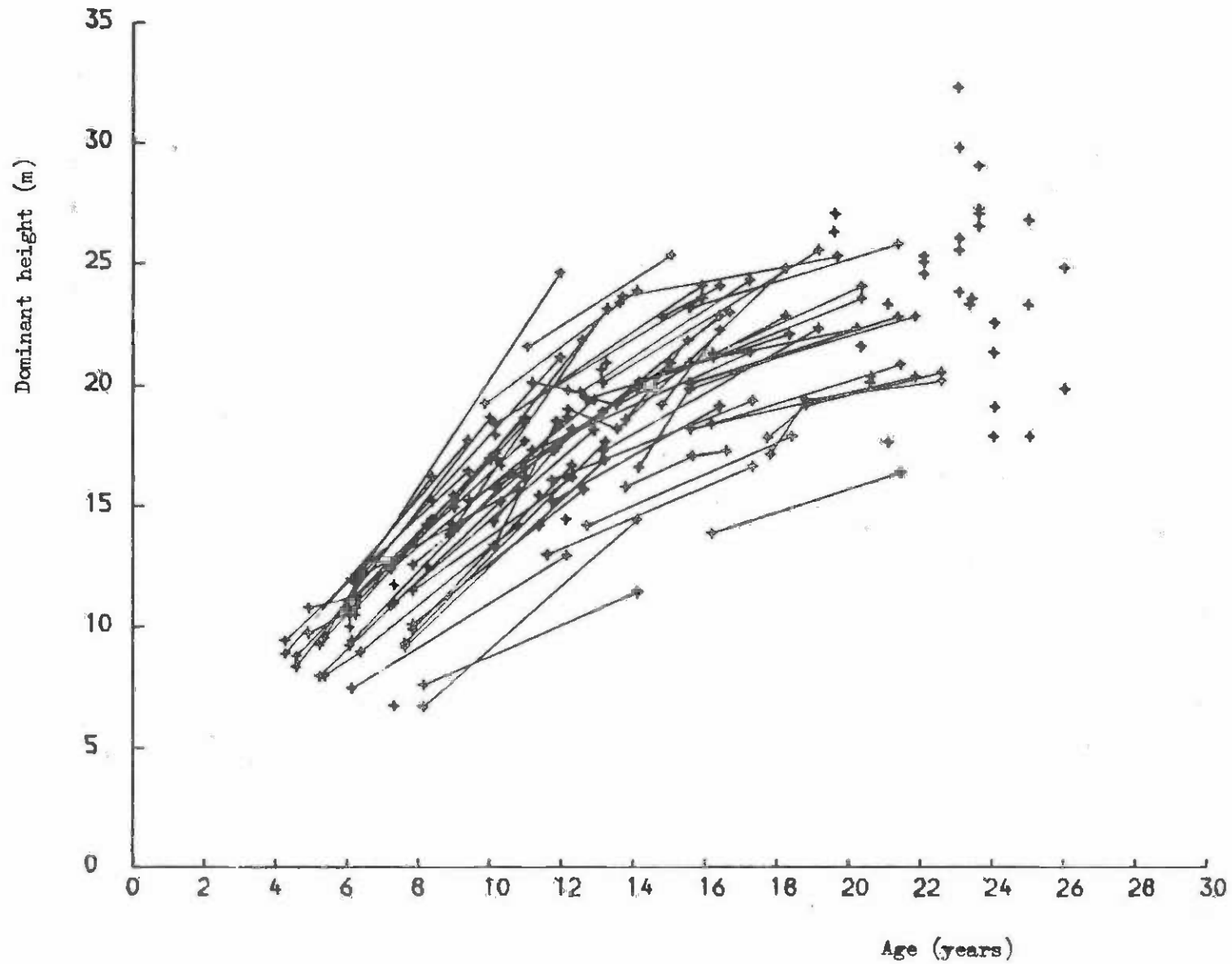


Figure 5.2

HAND-DRAWN SITE INDEX CURVES SKETCHED THROUGH THE DATA TO REPRESENT THE MINIMUM, MEDIAN AND MAXIMUM TRENDS

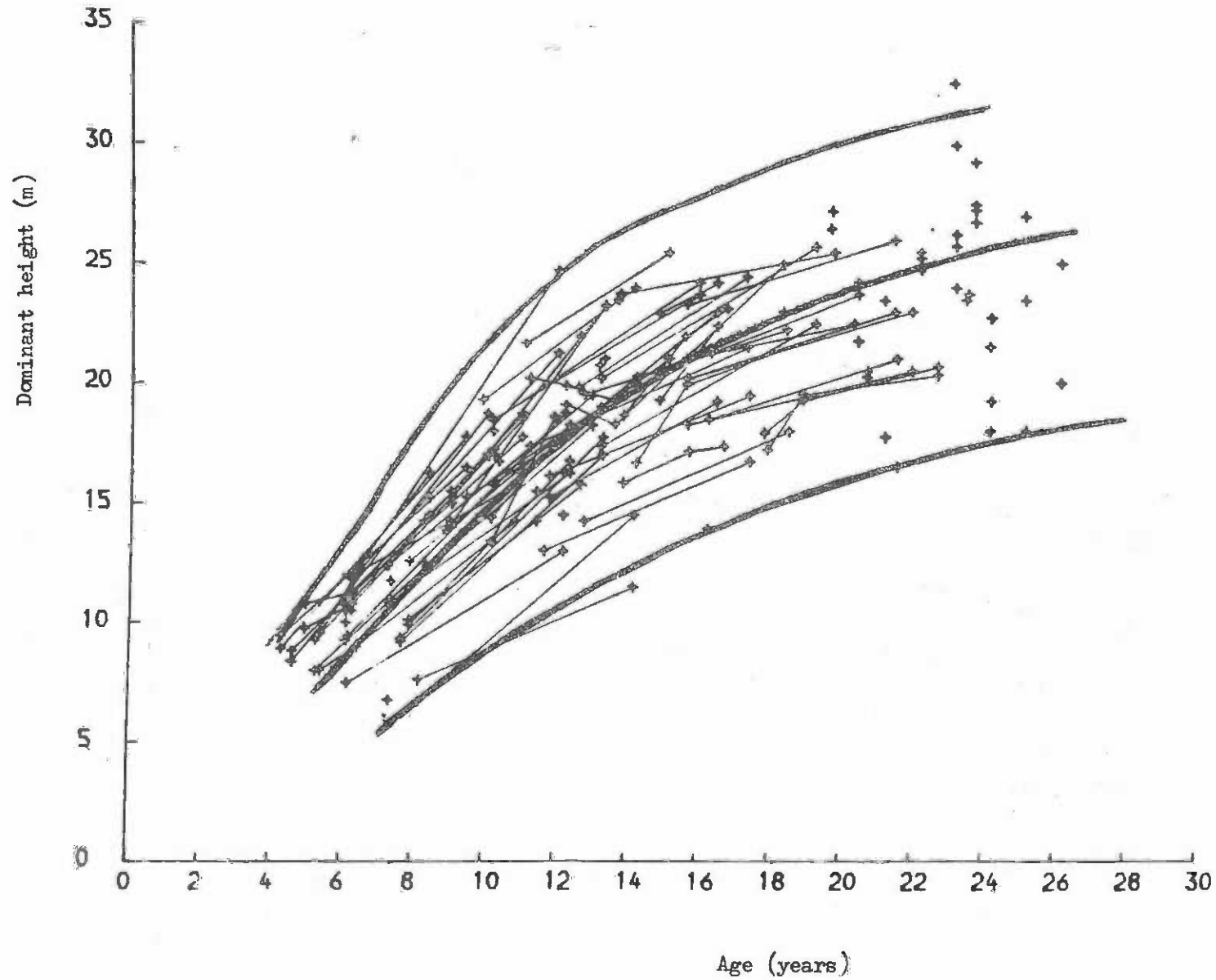
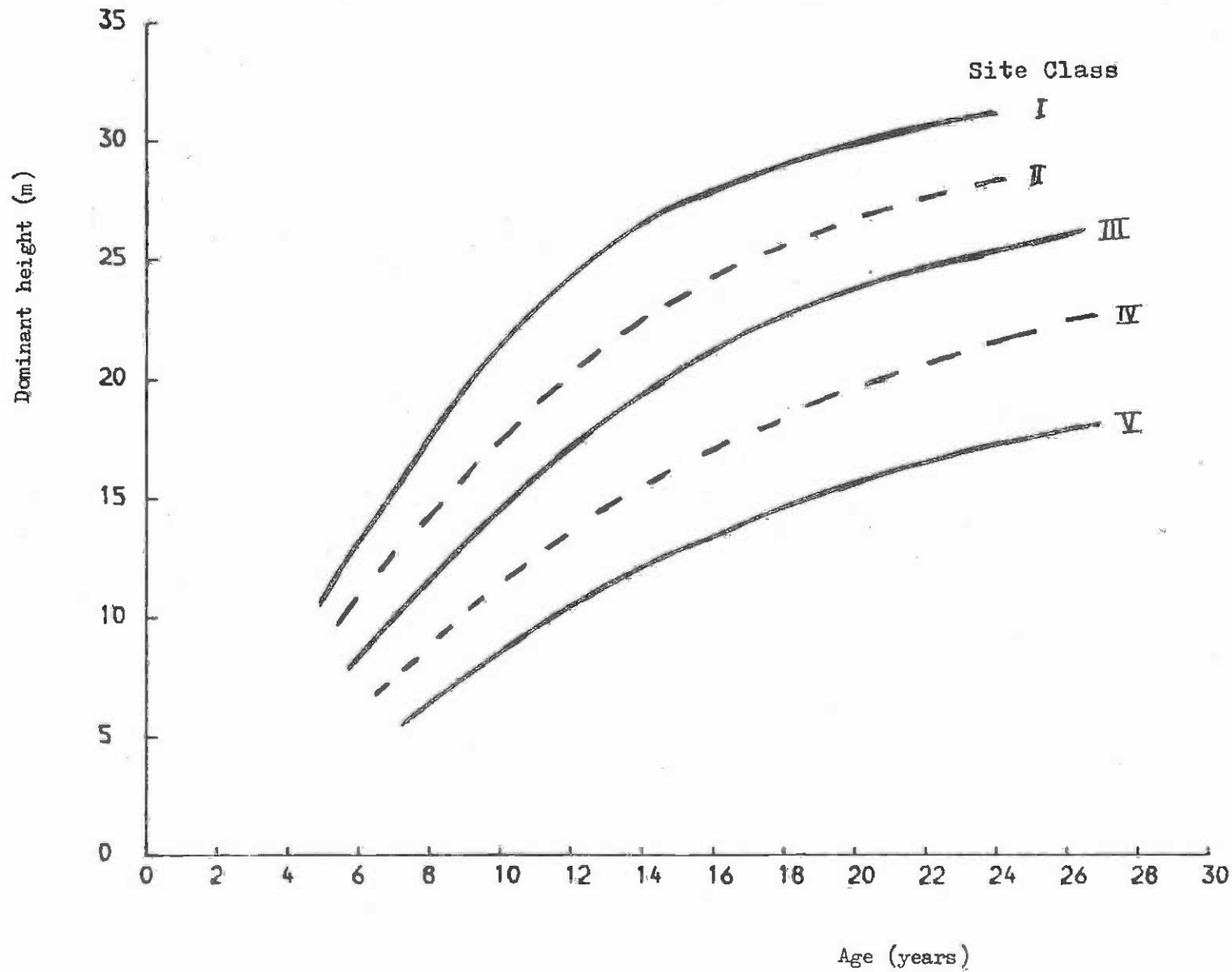


Figure 5.3

TRACING OF THE THREE MAIN SITE INDEX CURVES WITH TWO INTERPOLATED
INTERMEDIATE CURVES ADDED AND EACH CURVE ASSIGNED A SITE CLASS



When all of the data is from temporary plots, the same method can be applied, but there is a possibility of considerable error due to the fact that the plots at different ages may not be equally representative of different sites.

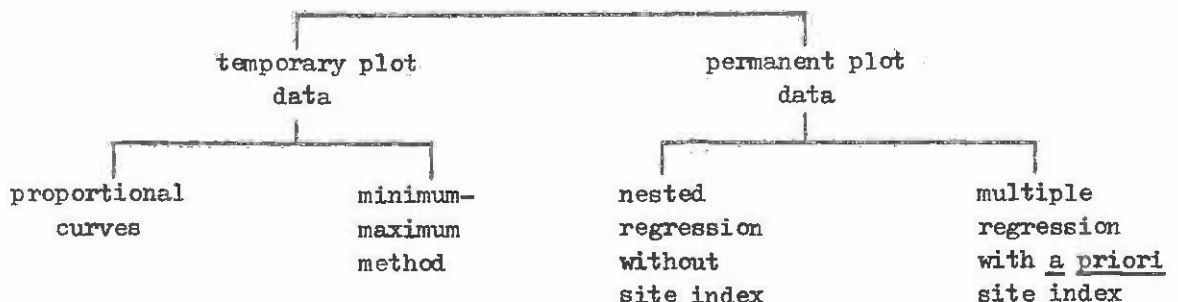
3. Two additional curves can now be interpolated between the upper and central curves and lower and central curves. The system can then be traced onto a separate piece of paper, giving the result shown in figure 5.3.
4. The curve system can be described as an equation using the methods described in Appendix A.1. The curves themselves are usually numbered sequentially and referred to as site class curves (or sometimes yield class, production class, etc.). Thus in figure 5.3 the site classes are numbered from I (most productive) to V (least productive).

The simplicity of this technique of constructing site index curves by graphical methods is obvious. It has three significant disadvantages:

- (i) The curves produced depend to a great extent upon the judgement of the person doing the work, especially if the data is sparse or largely from temporary plots. Different people will produce different sets of curves, which may be more or less accurate and unbiased in representing the real trend.
- (ii) When there is a large amount of data and it is already stored in a form acceptable to a computer (e.g. on 80 column Hollerith cards or on magnetic tape), then this is a very slow method compared to statistical techniques which can be carried out by computer.
- (iii) When stage 4 above, of describing the curve system as an equation is required (as when the curves are to be used within an inventory or growth projection programme), then the work involved in this step alone may be as great as the entire task of fitting the curves directly by one of the statistical techniques.

5.1.2.2. Mathematical methods of fitting site index curves

Mathematical techniques of fitting site index curves have considerable advantages over graphical methods when a computer is available and the amount of data is large. However, it should not be assumed that the results of these techniques are necessarily more accurate than hand-drawn curves; this will depend very much on the correctness of the height growth model chosen and the validity of the statistical assumptions used in the fitting of the parameters of the model. Mathematical techniques can be classified into four groups in the following way:



Of these four methods, nested regression is statistically the most appropriate and is also amenable to manual calculation. Consequently we shall examine this technique in most detail, discussing the other methods somewhat more briefly.

All four methods can be related to a single model of height growth, which is the Schumacher equation 1/:

$$H_o = H_{max} \cdot \exp(b/A^k) \quad -(1)$$

where H_o is dominant height, H_{max} is a parameter to be fitted and represents the maximum height the species could reach, $\exp()$ is the mathematical notation to indicate that the expression in the bracket is a power of the constant $e = 2.71828$ (i.e. $\exp(2)$ means e^2), b and k are parameters to be fitted and A is the age of the stand.

The shapes given by this equation are illustrated in figure 5.4. By taking logarithms to the base e (\ln) of both sides of equation (1), one gets:

$$\ln H_o = \ln H_{max} + b/A^k \quad -(2)$$

If we let $a = \ln H_{max}$, then a and b can be fitted by linear regression, provided k is known. Appropriate values of k for most species lie between 0.2 and 2 and can be estimated by techniques described later in this section or by nonlinear estimation as described in Appendix A.4. For many species, an assumed value of $k = 1$ will give a satisfactory fit. The b parameter in equation (2) should always be negative; if it is not, check calculations for errors. The a parameter will normally be between 2 and 7; again, check for errors if there is large divergence from this.

For proportional curves fit equation (2) to the set of temporary sample plot data as a whole by linear regression, with the dependent variable Y as $\ln H_o$ and the predictor variable X as $1/A^k$. If k is not known, follow the suggestions in Appendix A.4 to determine it.

This gives the average height growth trend, assuming that in each age class, all sites have an equal likelihood of being represented. If it is known that, for example, older age classes fall on poorer sites and younger ones on the best sites, then do not use this method. Either construct hand-drawn curves or, if PSP data is available, use nested regression.

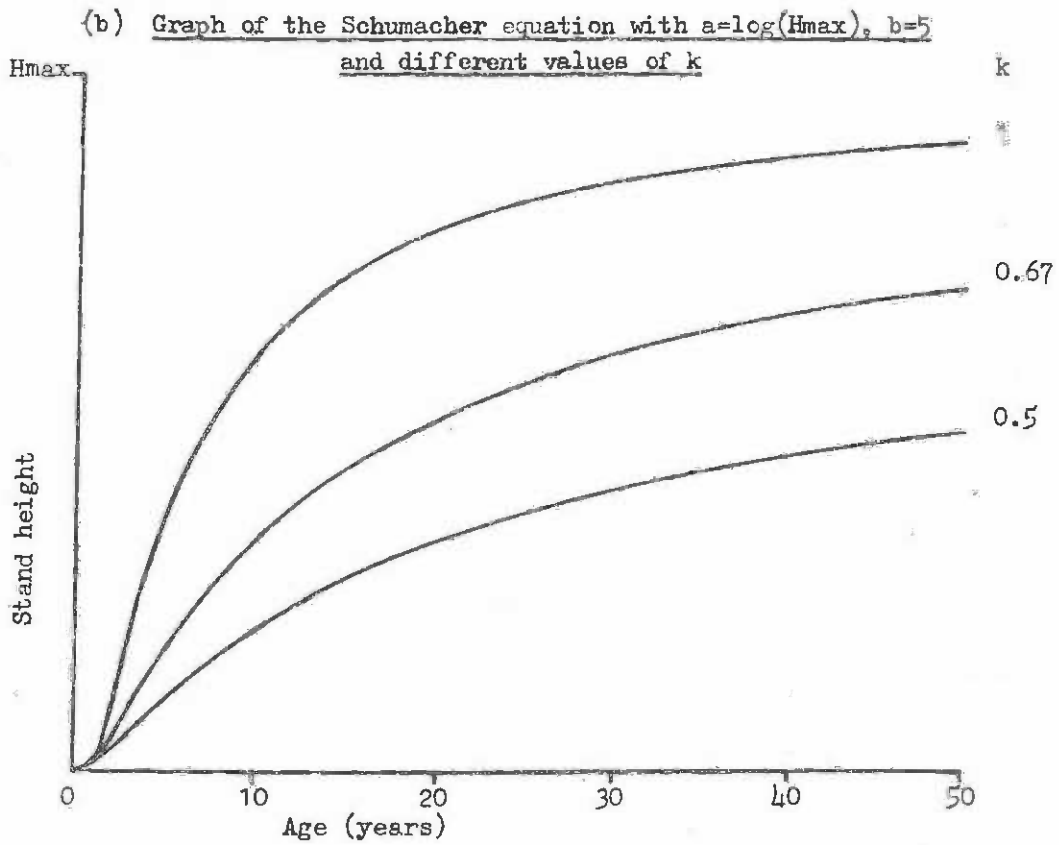
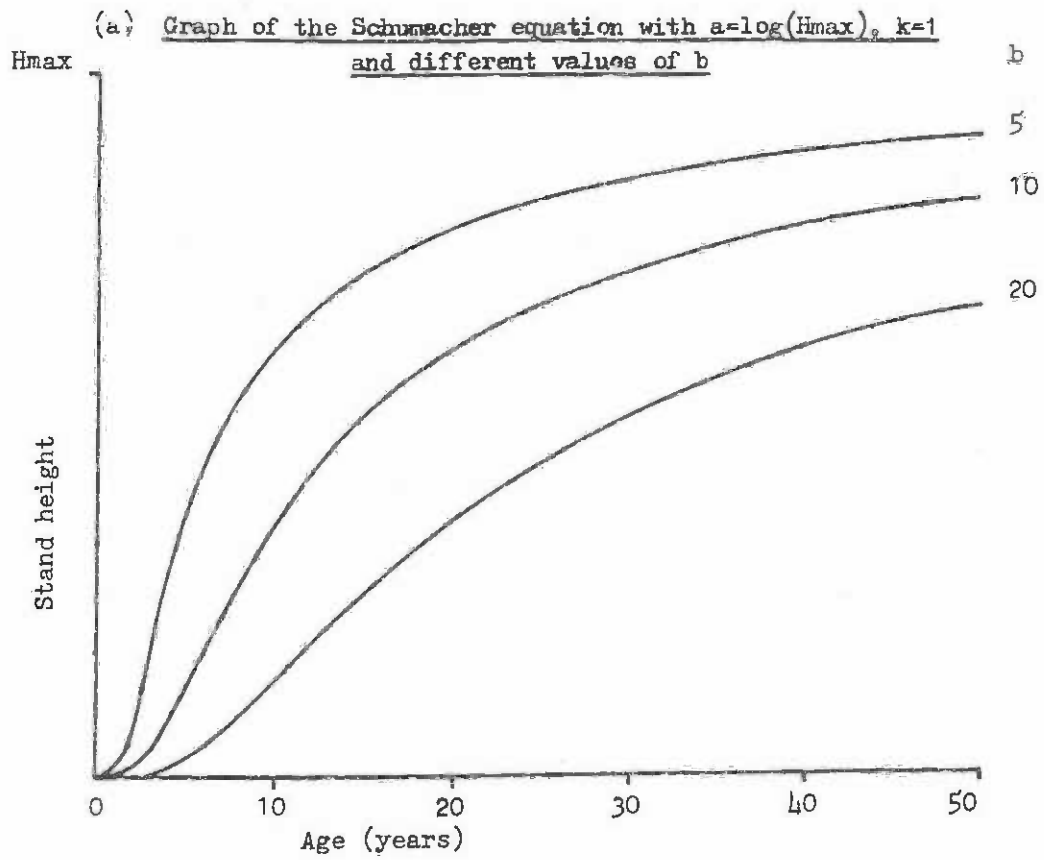
Once the mean height growth curve has been fitted, curves of the same shape can be drawn to pass through different site index values. If the site index S is defined as the dominant height of the stand at an index age A_i , then the a parameter for the curve to pass through this site index, a_i is given by:

$$a_i = \ln S - b/A_i^k \quad -(3)$$

where b and k are from the average curve.

1/ Schumacher, F.X., 1939 A new growth curve and its application to timber yield studies. J. Forestry 37:819-820.

Figure 5.4



The minimum-maximum method is more flexible in the type of curve shapes that result than the proportional curve method, but it requires multiple observations in each age class (at least 3), and hence cannot be used with limited amounts of data. The process proceeds as follows:

- (1) In each age class, calculate the mean H_0 for all plots and the minimum and maximum values of H_0 .
- (2) Fit three separate regressions of the type shown in equation (2) to the maximum, minimum and mean sets of observations. The nonlinear k parameter can be assumed as a constant for all three sets or it can be fitted independently.
- (3) As a final step, the separate coefficients for each of the three lines can, if desired, be harmonized to give a single equation, using the methods given in Appendix A.1.

A more complex variation on this method is justified with large amounts of data in each age class, as might be obtained from a forest inventory. The height observations in each class are sorted into order, from maximum to minimum and each point is assigned a site class S from:

$$S = (i - \frac{1}{2}) / n$$

where i is the plot's position after sorting and n is the number of plots in the age class. Once the plots have been assigned a site class, then the analysis can proceed using multiple regression as in the last method described below.

It should be thoroughly appreciated that the above methods for use with temporary sample plot data should be regarded as producing results that are only of provisional usefulness, as they depend critically upon the assumption that all sites have an equal likelihood of being represented in each age class.

This is rarely the case in reality and hence the curves produced will be in some degree defective. The only solution is to obtain recurrent height-age data from permanent plots or stem analysis trees, which can be analysed by one of the following methods.

Nested regression methods are of two types. There is first of all the use of conditional (or zero-one) variables in multiple regression, as described in the example in Appendix 2.10. This method has not, to the author's knowledge, been used in site index curve construction, probably because with any realistic number of plots, the number of variables involved in the regression would be enormous; but the approach is by no means infeasible, given a specially adapted programme to generate and handle the many zero-one variables. The second method, first described by Bailey & Clutter ^{1/} involves the use of the common slope and common intercept estimators from covariance analysis. This method is well suited to site index curve construction and is sufficiently simple to make manual calculation of the parameters possible.

^{1/} Bailey, R.L. and Clutter, J.L., 1974 Base-Age Invariant Polymorphic Site Curves. Forest Science 20:155-59

The common slope regression model is depicted in figure 5.5(a) and is given by the equation:

$$Y = a_i + b X \quad -(4)$$

where a_i is different for each plot, but b (the slope) is the same for all plots. The common intercept model is shown in figure 4.5(b) and is represented by the equation:

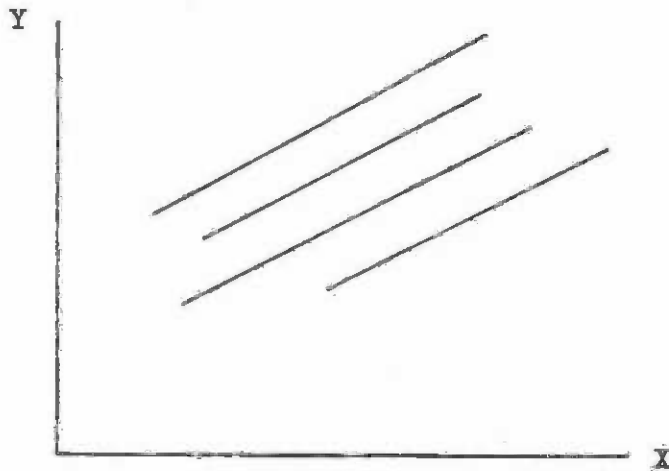
$$Y = a + b_i X \quad -(5)$$

where the intercept a is the same for all plots, but the slopes b_i differ. In terms of the Schumacher equation, either model can be used, with Y as $\ln H_0$ and X as $1/A^k$. The common slope model corresponds in shape to sets of proportional curves, but there is an important distinction between this approach and that for temporary plots, in that the distribution of sites in the different age classes has no effect on this method.

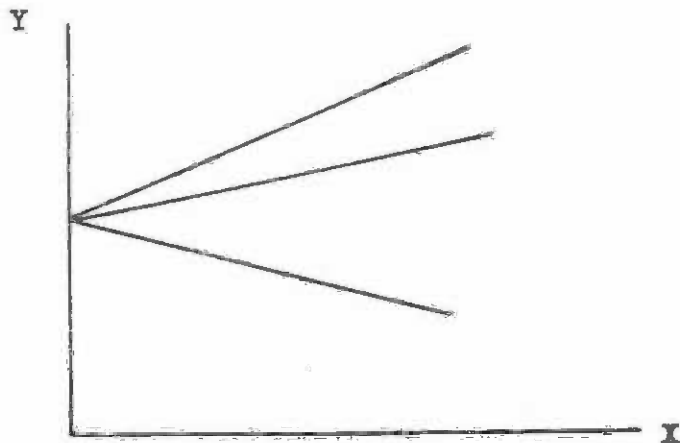
Figure 5.5

REGRESSIONS WITH COMMON SLOPES OR COMMON INTERCEPTS

(a) Common slope regressions for four plots



(b) Common intercept regressions for three plots



The statistical estimators for the common slope and common intercept models are as follows:

$$\text{The common slope } b = \frac{\sum_i \sum_j n_{ij} X_{ij} \cdot Y_{ij} - \sum_j X_{ij} \cdot \sum_j Y_{ij} / n_i}{\sum_i \sum_j n_{ij} X_{ij}^2 - (\sum_j X_{ij})^2 / n_i} \quad (6)$$

$$\text{The common intercept } a = \frac{\sum_i \sum_j n_{ij} Y_{ij} - \sum_i (\sum_j X_{ij} \cdot \sum_j Y_{ij} / \sum_j X_{ij}^2)}{\sum_i n_i - \sum_i ((\sum_j X_{ij})^2 / \sum_j X_{ij}^2)} \quad (7)$$

As these formulae appear rather complex (although in concept they are quite simple) some calculation pro formas are provided with a worked example. These are form 5.1, parts 1 and 2. Part 1 carries out the within-plot summations, corresponding to the Σ for the j subscript from 1 to n_i in the above formulae, whilst part 2 of the form carries out the between-plot summations, corresponding to Σ for i from 1 to m in the formulae.

The example uses the data shown in figure 5.6, from 6 permanent sample plots in Cupressus lusitanica stands in Kenya. A k parameter value of 1 is assumed for illustration purposes. The height-age data are transcribed into the first two columns of part 1 of form 5.1. Two sheets of this form are necessary for the six plots. The transformed X and Y values are entered in columns 3 and 4. X^2 is entered in column 5 and X times Y in column 6. Calculations should be carried out to at least four significant digits. The totals for each plot (within-plot totals) for columns 3 to 6 are entered in the appropriate line. The number of points in each plot is also entered. One then turns to part 2 of form 5.1 to continue the calculations. For each plot, the various within-plot totals (ΣX , ΣY , ΣX^2 and ΣXY) and the number of points n are combined according to the formulae shown at the top of the columns of part 2. These figures are then totalled between-plots to give the items marked (1) to (6) at the bottom. Finally the common slope and common intercept coefficients are calculated as shown in the last two lines.

The final result in this numerical example is that the common slope coefficient is -9.222, whilst the common intercept coefficient is 3.583. If either of these models is plotted as a set of site index curves, they will be found to bend over much more sharply than is indicated by the data in figure 5.6. This arises because the assumed value of k is much too large for this set of data.

Bailey & Clutter, in the paper referred to earlier, show how it is possible to calculate the nonlinear k coefficient directly using a regression model containing this coefficient in linear form, provided that remeasurement data (from PSPs or stem analysis) is available so that height increment can be estimated. The method is as follows:

1. Calculate a set of transformed Y values for the 2nd, 3rd, etc. observations within a plot, from the formula:

$$Y_{ij} = \ln \left\{ \left[\frac{H_{ij} - H_{ij-1}}{A_{ij} - A_{ij-1}} \right] \div \frac{1}{2}(H_{ij} + H_{ij-1}) \right\}$$

There is no Y value corresponding to the first height observation.

2. Calculate a corresponding set of transformed X values from the formula:

$$X_{ij} = \ln(2/(A_{ij} + A_{ij-1}))$$

3. Fit a common slope estimator using form 5.1 or equation (6) using these transformed X and Y values. Note that if form 5.1 is used, only the first two columns and totals (1) and (2) are required on part 2.
4. Subtract 1 from the common slope estimator obtained. The result is the estimate of k required.

When using manual calculation, form 5.2 can be used to carry out the transformations in steps 1 and 2 above. It has been completed for the first plot in the example data to illustrate the usage.

The formulae for the common slope and common intercept estimators, together with the transformation technique for estimating the k parameter, can easily be programmed for small computers or programmable calculators. Any programmable calculator with at least 15 data registers and 200 programme steps should be adequate.

When k parameter is fitted in this way to the data in figure 5.6, the following values are obtained:

$$\begin{aligned} k &= 0.25 \\ b &= -6.638 \quad (\text{common slope model}) \\ a &= 6.311 \quad (\text{common intercept model}) \end{aligned}$$

Let us examine the construction of a set of site index curves from these results, using the common intercept model.

We have:

$$\ln H_o = 6.311 + b_i/A^{0.25} \quad (8)$$

The parameter b_i depends upon site index S. For a selected site index, at index age A_i :

$$\begin{aligned} \ln S &= 6.311 + b_i/A_i^{0.25} \\ \therefore b_i &= (n S - 6.311) \cdot A_i^{0.25} \end{aligned}$$

If we wish to plot curves for site indices 16, 18, 20, 22, 24, 26 using an index age of 20 years, we have:

S	b_i
16	-7.483
18	-7.234
20	-7.011
22	-6.809
24	-6.625
26	-6.456

Then from equation (8), substituting the b_i for each site index curve, values of H_0 can be calculated for selected values of A . The curves that result from the above parameter values are shown in figure 5.7, on the same scale as is used for the data in figure 5.6.

If one wishes to calculate the site index of a stand, given its age and dominant height, then use the formula:

$$\ln S = a + (\ln H_0 - a) \cdot (A/A_i)^k \quad -(9)$$

For example, for the parameter values for a and k given above, suppose we have a stand of 14.5 m at 11.5 years. Then the estimated site index is given by:

$$\begin{aligned} \ln S &= 6.311 + (\ln 14.5 - 6.311) \cdot (11.5/20)^{0.25} \\ &= 3.144 \\ \therefore S &= 23.2 \end{aligned}$$

Hence we can say that the site index of this stand is 23 m.

When the common slope model is used, instead of the common intercept model, then the basic equation is:

$$\ln H_0 = a_i + b/A^k \quad -(10)$$

with a_i being dependent on site index as:

$$a_i = \ln S - b/A_i^k \quad -(11)$$

with site index for a selected height-age observation being given by:

$$\ln S = \ln H_0 + b(1/A^k - 1/A_i^k) \quad -(12)$$

Form 5.1 Common slope and common intercept regression models

Part 1 Plot data summarization. Use as many part 1 sheets as necessary for all plots.

Data transformations used: $X = \frac{1}{AGE}$
 $Y = LOG_e HEIGHT$

Plot 3					
Raw data		Transformed data			
A	H	X	Y	X ²	XY
6.3	9.6	0.1493	2.2618	0.02228	0.3376
7.6	11.0	0.1316	2.3979	0.01731	0.3155
10.6	14.8	0.09434	2.6946	0.00890	0.2542
11.6	15.2	0.08621	2.7213	0.00743	0.2346
12.6	16.3	0.07937	2.7912	0.00630	0.2215
13.5	17.5	0.07407	2.8622	0.00549	0.2120
14.4	18.3	0.06944	2.9089	0.00482	0.2019
Totals	n 7	0.6843	18.6358	0.07253	1.7773

Plot 35					
Raw data		Transformed data			
A	H	X	Y	X ²	XY
6.7	8.1	0.1493	2.0919	0.02228	0.3122
8.5	10.7	0.1176	2.3702	0.01384	0.2789
9.5	10.8	0.1053	2.3795	0.01108	0.2505
10.6	12.5	0.09434	2.5257	0.00890	0.2383
12.6	14.3	0.07937	2.6603	0.00630	0.2111
13.5	15.2	0.07407	2.7213	0.00549	0.2016
14.4	16.5	0.06944	2.8034	0.00482	0.1947
Totals	n 7	0.6894	17.5523	0.07271	1.6872

Plot 36					
Raw data		Transformed data			
A	H	X	Y	X ²	XY
6.7	8.9	0.1493	2.1861	0.02228	0.3263
7.6	11.1	0.1316	2.4069	0.01731	0.3167
10.6	14.6	0.09434	2.6810	0.00890	0.2529
11.6	16.5	0.08621	2.8033	0.00743	0.2417
12.6	18.1	0.07937	2.8959	0.00630	0.2298
13.5	20.0	0.07407	2.9957	0.00549	0.2219
14.4	21.0	0.06944	3.0445	0.00482	0.2114
Totals	n 7	0.6843	19.0135	0.07253	1.8007

Form 5.1 Common slope and common intercept regression models

Part 1 Plot data summarization. Use as many part 1 sheets as necessary for all plots.

Data transformations used: $X = \frac{1}{Age}$
$Y = \text{LOGE HEIGHT}$

Plot 46					
Raw data		Transformed data			
A	H	X	Y	X ²	XY
16.7	21.1	0.05988	3.0493	0.00359	0.1826
17.6	22.4	0.05682	3.1091	0.00323	0.1767
20.6	25.1	0.04854	3.2289	0.00236	0.1564
21.6	25.7	0.04630	3.2465	0.00214	0.1503
22.6	26.3	0.04425	3.2696	0.001959	0.1447
24.4	27.7	0.04098	3.3214	0.001680	0.1361
Totals	n 6	0.2968	19.2189	0.01495	0.9468

Plot 52					
Raw data		Transformed data			
A	H	X	Y	X ²	XY
21.7	18.1	0.04608	2.8959	0.00212	0.1334
22.6	19.4	0.04425	2.9653	0.00196	0.1312
23.5	20.0	0.04255	2.9957	0.00181	0.1275
26.6	22.3	0.03759	3.1046	0.00141	0.1167
27.6	23.3	0.03623	3.1485	0.00131	0.1141
28.5	24.2	0.03509	3.1864	0.00123	0.1118
29.4	24.7	0.03401	3.2068	0.00116	0.1091
Totals	n 7	0.2758	21.5031	0.01100	0.8438

Plot 198					
Raw data		Transformed data			
A	H	X	Y	X ²	XY
22.7	28.3	0.04405	3.3429	0.00194	0.1473
23.5	28.9	0.04255	3.3638	0.00181	0.1431
25.6	31.1	0.03906	3.4372	0.00153	0.1343
26.5	32.5	0.03774	3.4812	0.00142	0.1314
28.6	33.0	0.03497	3.4965	0.00122	0.1223
29.4	33.6	0.03401	3.5145	0.00116	0.1195
Totals	n 6	0.2324	20.6362	0.00908	0.7978

Form 5.1 Common slope and common intercept regression models

Part 2 Totals between plots and coefficient calculation

Plot	$\frac{\Sigma XY - \Sigma X \Sigma Y}{n}$	$\frac{\Sigma X^2 - (\Sigma X)^2}{n}$	$\frac{\Sigma X \Sigma Y}{\Sigma X^2}$	$\frac{(\Sigma X)^2}{\Sigma X^2}$	n	ΣY
3	-0.044483	0.00563479	16.7683	6.45618	7	18.6358
35	-0.041451	0.00481395	15.9972	6.53655	7	17.5523
36	-0.058005	0.0056348	16.9891	6.45618	7	19.0135
46	-0.0038949	0.00026829	18.7967	5.89232	6	19.2189
52	-0.0034221	0.00013348	21.15637	6.91506	7	21.5031
198	-0.0015088	0.00007837	20.4195	5.94821	6	20.6362
Totals	¹ -0.15276	² 0.016564	³ 110.127	⁴ 38.2045	⁵ 40	⁶ 116.560

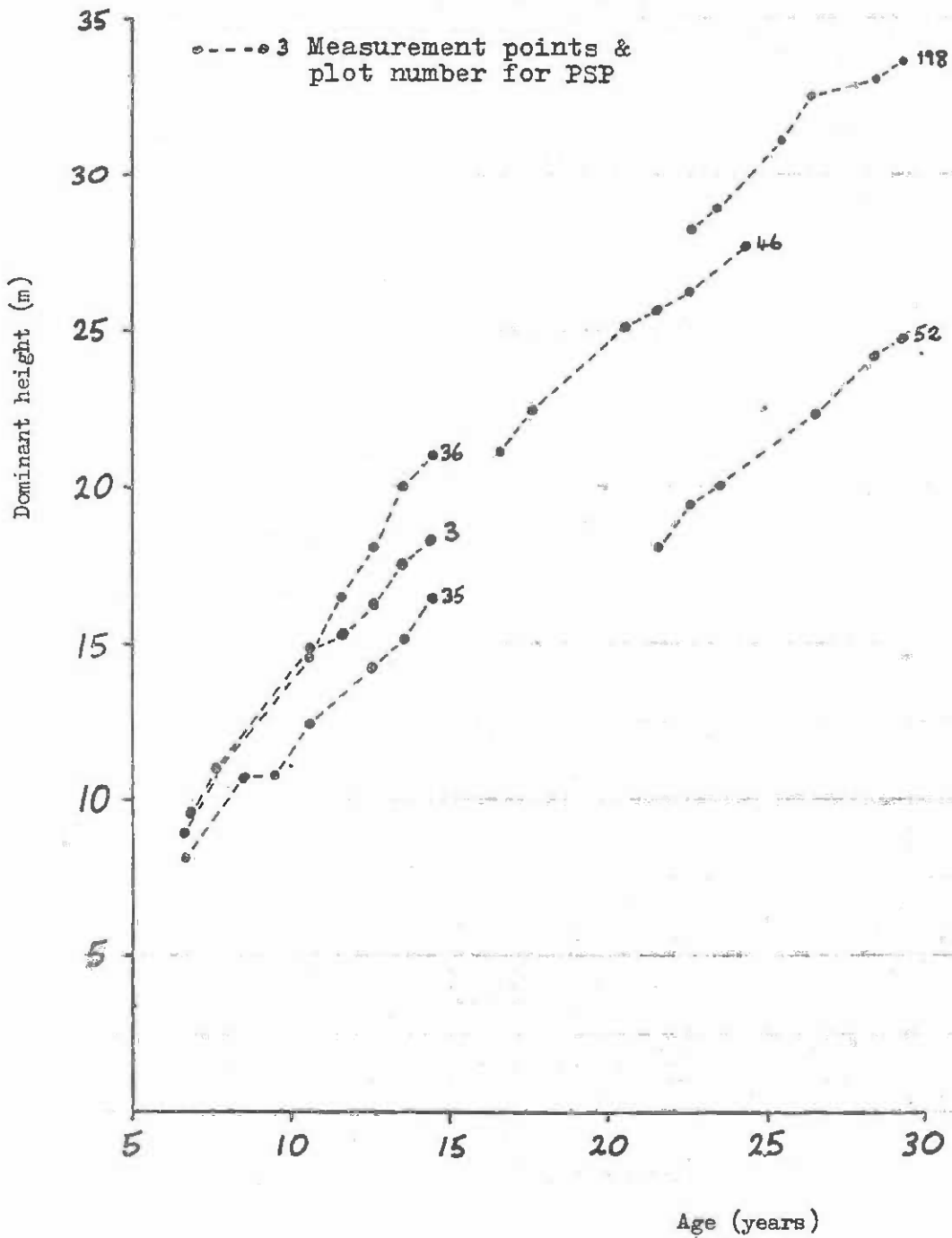
Common slope $b = (1)/(2) = \frac{-0.15276}{0.016564} = -9.222$

Common intercept $a = ((6)-(3))/((5)-(4)) = \frac{(116.56 - 110.127)}{(40 - 38.2045)}$

= 3.583

Figure 5.6

Data from 6 permanent sample plots in Cupressus lusitanica stands in Kenya. Used in text example for fitting site index curves by nested regression



Form 5.2 Transformations to age and height data from PSP's to fit k parameter in Schumacher equation by common slope estimator

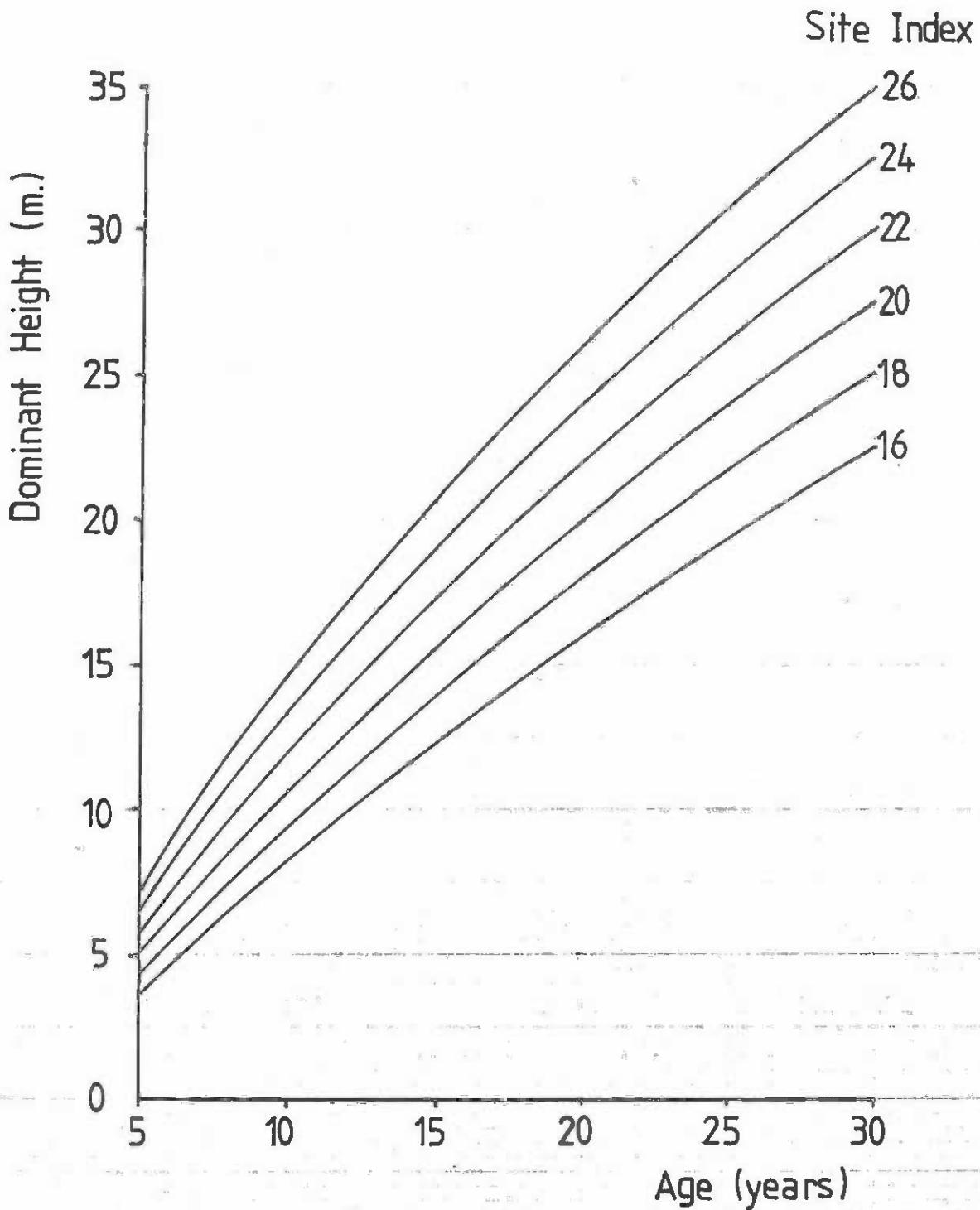
PSP 46

i	j	A_{ij}	a	b	c	X_{ij}	H_{ij}	d	e	f	g	h	Y_{ij}
Plot	Point	Age	Sum	Diff	$2/a$	$\ln c$	Height	Sum	Diff	e/b	$d/2$	f/g	$\ln h$
1	1	16.7					21.1						
1	2	17.6	34.3	0.9	0.05831	-2.842	22.4	43.5	1.3	1.444	21.75	0.06639	-2.712
1	3	20.6	38.2	3.0	0.05236	-2.950	25.1	47.5	2.7	0.9000	23.75	0.03789	-3.273
1	4	21.6	42.2	1.0	0.04739	-3.049	25.7	50.8	0.6	0.6000	25.40	0.02362	-3.746
1	5	22.6	44.2	1.0	0.04525	-3.096	26.3	52.0	0.6	0.6000	26.00	0.02308	-3.769
1	6	24.4	47.0	1.8	0.04255	-3.157	27.7	54.0	1.4	0.7778	27.00	0.05185	-2.959

Transfer
to form
5.1

Figure 5.7

Site index curves produced for the data in figure 5.6 by Bailey & Clutter's nested regression method using the Schumacher equation



Multiple regression methods of fitting site index curves can be used as an alternative to nested regression when the site index or site class of the plot can be determined prior to fitting the model. There are perhaps five ways in which this a priori site class or site index determination can be made:

- (1) For a long series of plot measurements which have passed through the index age, site index can be taken directly as the dominant height at the index age.
- (2) For series of measurements which do not pass through the index age, an equation such as the Schumacher equation can be fitted independently to each plot and used to predict the dominant height at the index age, which is then taken as the site index for the plot.
- (3) For temporary plot data, site class can be assigned on the basis described above under the maximum-minimum method of fitting site curves. This method should not be used with PSP data as it is wasteful of the information inherent in such data.
- (4) An existing set of site curves can be used to assign site class to the plots. Alternatively, hand-drawn curves can be made specially for a particular analysis, to classify plots. This technique is useful when existing curves are almost satisfactory, but of not quite the correct curvature or one wishes to approximate hand-drawn curves by an equation.
- (5) An environmental variable such as altitude, rainfall, etc. can be used as the site indicator variable. This approach is rarely successful due to the poor correlation usually found between a single environmental variable and height growth.

Once the method of assigning site class to the plots has been determined, one has a set of data in which three variables are known for each observation:

Dominant height:	H_0
Age	A
Site class or index	S

Multiple regression can then be used to fit a model which relates H_0 to A and S using various different transformations. Two types of model have been used:

- (1) Constrained models, for use with site index curves, where height is expressed relative to site index and age relative to index age A_1 . The regression fitted is one without an intercept. An example would be the model:

$$(H_0 - S) = b_1(A - A_1) + b_2(A - A_1)^2$$

This kind of model is forced to give a dominant height H_0 equal to site index S when the age is equal to the index age A_1 .

- (2) Unconstrained models, with an intercept term. When used with site index, rather than site class, the curves must be conditioned after fitting to ensure that the dominant height corresponds to the site index at the index age.

Some examples of the types of unconstrained models that can be fitted are:

$$H_o = b_0A + b_1A + b_2S + b_3A.S + b_4A^2 \quad \text{--(13)}$$

$$\log H_o = b_0 + b_1/A + b_2S + b_3S/A \quad \text{--(14)}$$

Equation (13) is based on the quadratic equation. See figure A.2.1(e) in Appendix A for the shapes that a single curve may take. Obviously for site index curves, only one case of the several possible shapes is desired. Equation (14) is based on the Schumacher equation, or figure A.2.1(c).

An infinite variety of other models is possible. In all cases, it is essential to plot the curves after fitting and to overlay them on a graph of the data to ensure that the trends are accurately followed. No reliance can be placed on the statistical parameters (e.g. R^2 , residual standard deviation) associated with multiple regression in determining the suitability of the curves.

When site index is used, as opposed to site class, then it will be found that if the curves are plotted, the predicted dominant height H_o at the index age A_i is generally not equal to the site index S . This is because there is no constraint in the fitting process to bring about this coincidence. Curves for selected real site index values S^* are graphed by calculating the statistical site index S required in the model to give a line passing through S^* at A_i .

Thus from model (13) we have:

$$S^* = b_0 + b_1A_i + b_2S + b_3A_i.S + b_4A_i^2 \quad \text{--(15)}$$

$$S = (S^* - b_0 - b_1A_i - b_4A_i^2)/(b_2 + b_3A_i) \quad \text{--(15)}$$

or from model (14) we have:

$$\log S^* = b_0 + b_1/A_i + b_2S + b_3S/A_i$$

$$S = (\log S^* - b_0 - b_1/A_i)/(b_2 + b_3/A_i) \quad \text{--(16)}$$

The necessity for conditioning site index curves and the mysterious and confusing distinction between statistical site index S and real site index S^* has led many workers to use models of the constrained type described above. Unfortunately, it is often the case that many standard multiple regression programmes do not have a facility for fitting equations without an intercept term, so that this option is not available to the research worker.

With site class curves, the problem does not arise, since there is no requirement for the curves to pass through a particular point.

Multiple regression techniques have the advantage of great flexibility in the type of model adopted. They have the disadvantages of requiring a priori estimation of site index, of being inefficient statistically in not making use of the nested nature of PSP data and of being based on the invalid assumption that the site variable is known to a high degree of precision.

5.1.3 Site Assessment Models Based on Environmental Factors

Site index or site class curves are obviously only useful as a tool for predicting production potential when applied to existing stands. A large class of forest management decisions require some assessment of potential production of a given species before it has been established or at least very early in the life of young stands.

This is only possible in two types of situation:

- Very generalized predictions as to maximum productivity in a given region by comparison with optimal productivity of similar forests growing under similar environmental conditions elsewhere in the world.
- More detailed prediction of site class with reference to a particular set of height/age curves for a given species by construction of a functional relationship for forests already established in the region in question, between site class and environmental factors such as soil nutrients, depth and texture, altitude, aspect and rainfall.

The first approach is particularly relevant for mixed natural forests. As far as man-made uniform forests are concerned, it corresponds more to the phase of species and provenance selection than to that of detailed yield prediction with which this manual is concerned.

5.1.3.1 Functional models for site class prediction

Site class prediction models are constructed in the following stages:

- (1) Construction of site index or site class curves from permanent sample plots in existing forest.
- (2) Collection of environmental data from permanent and/ or temporary plots. Each plot is also assigned a site class on the basis of dominant height and age from the site index curves. The environmental data collected should correspond to the suggestions given in section 3.2.3.
- (3) The various environmental variables are transformed and selected as described in the following section, to produce a predictive model in the form:

$$S = b_0 + b_1.e_1 + b_2.e_2 + \dots + b_n.e_n$$

where b_0, b_1 , etc. are coefficients fitted by multiple regression and e_1 to e_n are relevant transformations of environmental variables.

- (4) Confidence limits for predictions can be defined from the methods given for multiple regression in Appendix A. In general, it is also desirable to test the function by comparison of actual and predicted site classes for an independent set of data to that used to fit the function. This independent data should preferably be from a different region from the main data to test the regional stability of the prediction function.

5.1.3.2 Construction and selection of environmental variables

The total number of environmental variables measured may be quite large, possibly as many as 100 per plot, whilst a usable model should contain the smallest possible number of most easily measurable variables.

Reduction in the total number of variables can be achieved in several ways:

- By synthesis of some items of information. For example, climatic data may be synthesized to give the length of the wet season or growing season with temperatures over 6°C , or average temperatures during the wet season or more complex techniques based on evapotranspiration formulae.
- Statistical selection of the most significant variables. Principal components analysis can be used for this, but it is perhaps simpler to use stepwise multiple regression analysis.
- Graphical analysis of variables considered likely to be most important as limiting factors in growth, selecting only those with definite and obvious relationships with site class.

The actual measured variables should if possible be transformed in order to give values that are likely to be correlated with growth. The most obvious example is aspect, which may be measured in degrees from 0 to 360, where values around zero and 360 both correspond to northerly directions. Taking the sine of the angle divided by 2 gives an aspect code between zero for northerly directions and one for southerly slopes.

5.1.3.3 Problems in the application of site assessment functions

To be effective, a site assessment model should be described in terms of predictor variables that are easily measurable. They should also have a relatively high correlation coefficient, preferably over 0.8 with 20 or more points. Otherwise the relationship may be statistically significant, but of no practical use for prediction purposes because of low precision. The predictive equation should also have as few predictor variables as possible, preferably not more than three or four.

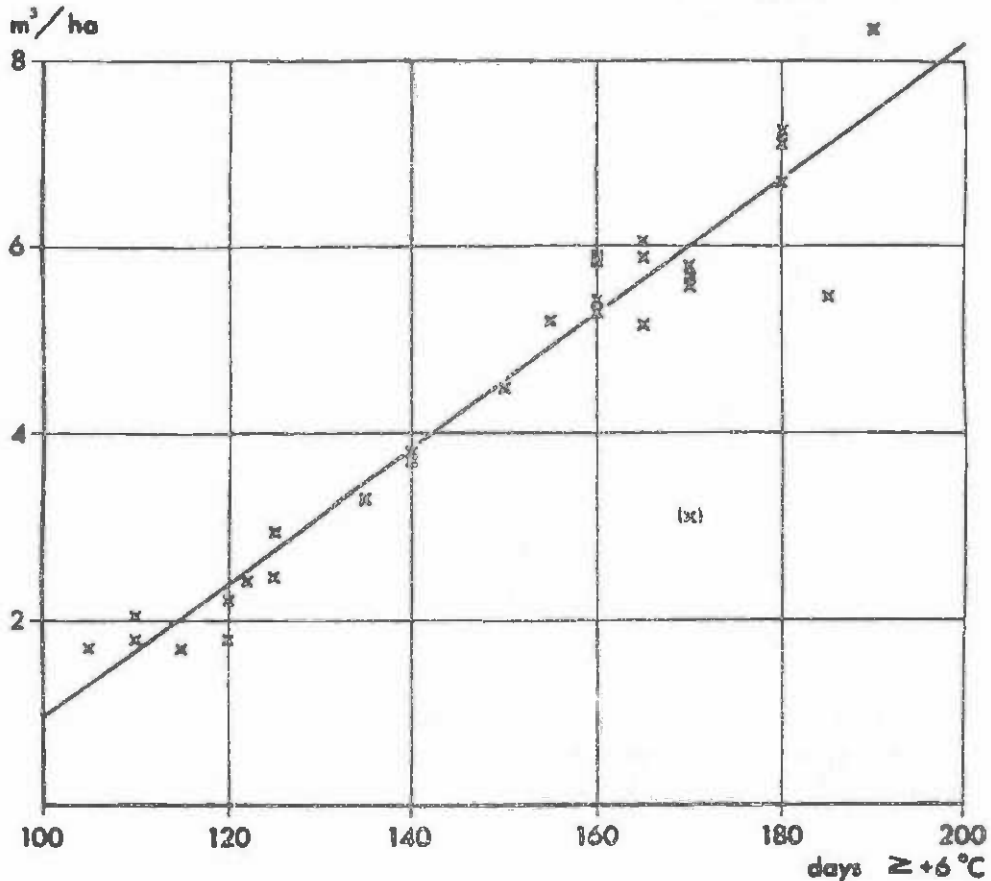
The predictor variables should be quantities which are readily determinable using conventionally available equipment.

An example is shown in figure 5.8 of a predictive model which fulfils these criteria. It is based on a single variable, which is the number of days in the growing season exceeding 6°C . Site index is expressed in terms of mean annual volume increment per annum at its maximum.

The fewer predictor variables are involved in a model, the more likely it is to be usable over a wide region. With many predictor variables, the model becomes very sensitive to the relationships between them, especially if some of the predictor variables are highly correlated. Consequently, regional shifts in the balance between predictor variables can easily invalidate a complex model.

Figure 5.8

Site assessment model for Scotch Pine (*P. sylvestris*) forests in Sweden. Average productivity of regions is plotted against mean number of days exceeding 6°C per year. (Reproduced from Fries, J., 1978 "The assessment of growth and yield and the factors influencing it" Special paper to 8th World Forestry Congress, Djakarta)



5.2 STATIC METHODS OF PREDICTING YIELD

Static yield prediction methods are those in which yield is predicted directly as a function of age, site class and the history of stand density. The methods are static in the sense that the resultant yield functions do not permit any variation in the history of stand treatment, except into broad classes of alternative thinning treatments that are already present in the data.

The components of yield that are of major interest to the forest manager are volume and mean diameter. To determine the volume of stands requires a knowledge of stem numbers and height, as well as mean diameter. Stem numbers are usually the basis for defining thinning treatment, whilst height is the most common means of site classification.

In this section, static methods of predicting mean diameter will be considered.

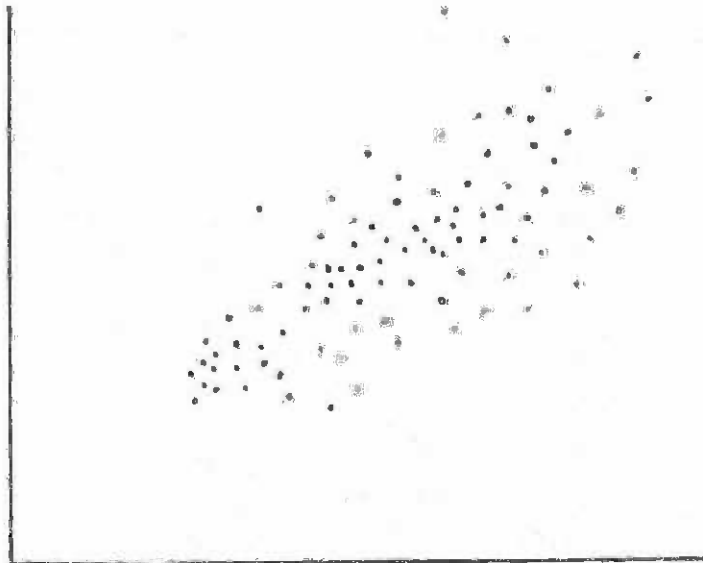
Figure 5.9

Shape of the relationship between mean diameter
and dominant height

- (a) With temporary plot data, a scatter diagram is obtained. The true shape of the diameter/height relationship is difficult to define.

Diameter of tree of
mean basal area

Dg

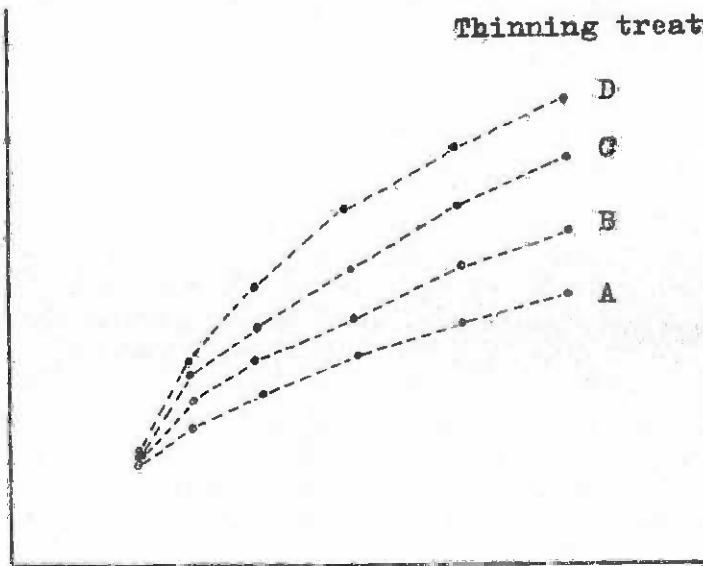


Dominant height Ho

- (b) With thinning experiment data, the shape of the relationship and its dependence on stand density is very clear. A represents the lightest thinning and D the heaviest.

Dg

Thinning treatments



Ho

5.2.1 Graphical Methods based on the Diameter/Height Function

When data from a mass of temporary plots is graphed with mean diameter on the vertical axis and dominant height on the horizontal axis, a diagram is obtained similar to figure 5.9a. If the data is from a thinning experiment, something like figure 5.9b is obtained. Treatment A represents the lightest thinning, i.e. the heaviest stocking, and Treatment D is the heaviest thinning, with the lowest stockings.

It can be seen that the curve is concave with respect to the horizontal axis. With temporary plot data or with permanent plots, a series of hand-drawn curves can be constructed in a manner exactly analogous to that described in section 5.1.2.1 for site index curves, except that here the curves demarcate differences in stand history rather than site class.

With temporary plot data, the shape of the curve may be defined basically from the upper and lower envelopes of the scatter diagram, but obviously the curve shape can be easily misconstrued as a result of a few exceptional or anomolous points and remains largely a matter for subjective judgement. With permanent plot data on the other hand, the main curve trends are readily visible from the slope of the lines joining remeasurements on the same plot.

Height is used on the horizontal axis as a site-dependent indicator of age. Graphs can be plotted using age instead, but in this case a separate set of curves is needed for each site class.

Once a set of curves has been drawn, the stocking history represented by each density class must be determined. This is discussed in section 5.2.3.

Hand-drawn curves can be expressed as equations by using one of the various approximation methods described in Appendix A.

5.2.2 Direct Statistical Estimation of Mean Diameter Prediction Functions

Functions can be fitted directly to mean diameter data by methods analogous to those used for fitting site index curves. The predictor variables may be dominant height and stand treatment history (designated by T), or may include age and site index as separate variables instead of as the combined variable dominant height. The methods that may be used to fit the functions include:

(i) With temporary plot data:

- Sort data by diameter within height classes and assign an order number after sorting. Adjust order numbers within each height class onto a common scale by the transformation:

$$T = i/n$$

where i is the order number within a height class and n is the number of points within the class. T is an index of treatment history. A multiple regression of one of the types discussed below can then be fitted using T as an independent predictor variable.

- Fit a mean trend line by simple regression of mean diameter on dominant height and then construct proportional or parallel (anamorphic) sets of curves.

(ii) With permanent plot data:

- Simple nested regression, as described in section 5.1.2.2 for site index curves. This depends on being able to find a transformation for the data that will result in a figure like figure 5.5a or 5.5b for a common slope or common intercept model. The Schumacher equation may be suitable or a simpler transformation of the type:

$$Dg = a + b.Ho^k$$

where k is a power between 0 and 1.

- Multivariate nested regression, either using the conditional variable technique described in Appendix A2.9 or using more advanced multistage regression techniques.
- Conventional multiple regression, after assigning a treatment history value T to each plot. This may be done in various ways, including a variation on the sorting method described above for temporary plots in which average values of T calculated for each measurement are taken for plots managed according to a consistent thinning schedule.

With thinning experiments, T values may be assigned for each treatment simply as 1, 2, 3 etc. if the treatments can be clearly ordered according to their degree of intensity.

A very common approach to fitting multiple regression models to permanent plots is to use spacing relative to height or even simply stocking, as indices of treatment history. This method can appear to work quite well, but introduces special conceptual difficulties which are discussed below in section 5.2.3.

The actual mathematical function that is used for the basis of the fitting method is not very important, provided it is capable of assuming the correct shape. This will be a curve that is gently concave with respect to the horizontal axis, i.e. bending downwards away from the straight line, as shown in figure 5.9b. Suitable models are:

$$(1) Dg = b_0 + b_1.Ho + b_2.T + b_3.Ho.T + b_4.Ho^2$$

$$(2) Dg = b_0 + b_1.Ho^k + b_2.T + b_3.Ho^k.T$$

The second model can be fitted using trial values of k between 0 and 1 with linear regression methods, selecting a value that gives a maximum coefficient of correlation or it can be fitted directly by nonlinear regression.

Whatever function is used to fit the data, the results should be examined graphically, overlaid on the data, for anomalous behaviour.

5.2.3 Defining Treatment History in Terms of Stocking and Age

The classes of treatment history may be defined graphically or by a numerical method, as discussed in sections 5.2.1 or 5.2.2 respectively. Once this has been done, then the actual stockings involved may be determined by a process of tabulation. If the function or graph relates mean diameter to dominant height, then for each treatment class, construct a

table of height classes and determine the average stocking per ha in each class. The height classes can subsequently be converted into classes of age and site index by reference to the relevant site index curves.

When the mean diameter function has been fitted using stocking or relative spacing as an index of treatment history, then it is not necessary to determine the stocking corresponding to a particular treatment history. However, the coefficients determined for such a function are dependent upon the particular relationship between age and stocking observed in the data. Failure to realise this can mean that the function may be used for age/stocking combinations not represented in the data, giving more or less erroneous results.

5.2.4 Static Yield Functions Predicting Basal Area or Volume

As well as constructing static yield functions to predict mean diameter, it is possible to use identical techniques to predict basal area or volume per hectare. The actual shapes of the functions will be somewhat different and will be as shown in figures 5.10a or b.

In general, use of mean diameter involves certain advantages in terms of overall model simplicity. If basal area or volume are predicted, then it is usually necessary to have some function which will allow mean diameter to be determined subsequently.

Also, with basal area or volume per hectare, a large component of the response to different stand densities is simply the multiplicative effect of different stockings. With the mean diameter function, this influence is removed, so that resultant relationships focus upon the real effects of competition.

5.2.5 Limitations of Static Yield Models

Static yield models have three significant disadvantages:

- (i) It is difficult to combine together data from stands with radically different or variable treatment histories and obtain a consistent and effective yield function.
- (ii) Once the model has been constructed, it cannot be used to predict yields for alternative treatments apart from those represented by the treatment histories incorporated in the model.
- (iii) Determination of thinning yields is difficult, unless accurate records of the diameters of removed stems are kept. Very often with PSPs or even with experiments such data are unavailable.

On the other hand, static models are undeniably easier to construct and use than dynamic models and, in appropriate situations, at least as accurate.

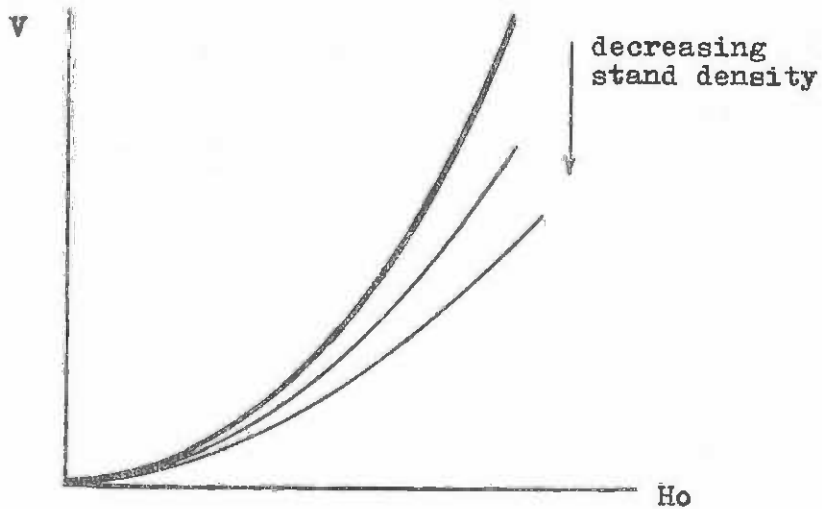
5.3 DYNAMIC METHODS OF PREDICTING GROWTH AND YIELD

A dynamic model is one which models rates of change within a system. As far as forest yield studies are concerned, this means that the basic prediction is of increment in diameter, basal area or volume.

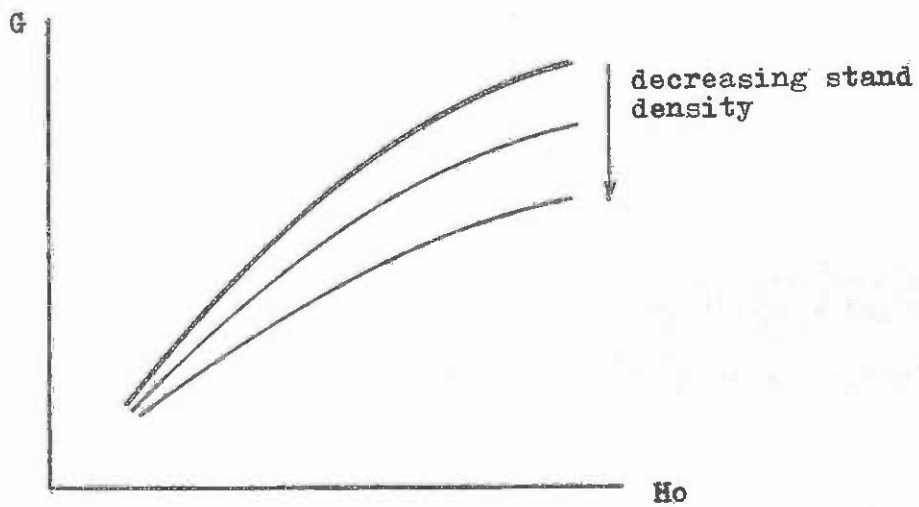
Figure 5.10

Shape of the volume and basal area functions
over dominant height

- (a) Volume per hectare on dominant height. Note that the solid line represents a limit not exceeded at very high stand densities. If logarithmic scales are used on both axes, the relation appears as a straight line.



- (b) Basal area per hectare on dominant height. The solid line is a limit not exceeded at very high stand densities.



Dynamic models have the advantage of being more realistically representative of the true cause and effect dependence between stand density and stand yield than are static yield models. They are free from the limitation that the data should represent a consistent series of stand histories and consequently can be used as the basis for analysing and synthesising data from very diverse types of experiments and permanent sample plots.

Growth models are based around functions that predict increment (of mean diameter, basal area or volume) over short time intervals, as a function of stand density, expressed in terms of basal area and/or age and stand age and site class. The latter two variables may be combined as dominant height as was done for the static yield models.

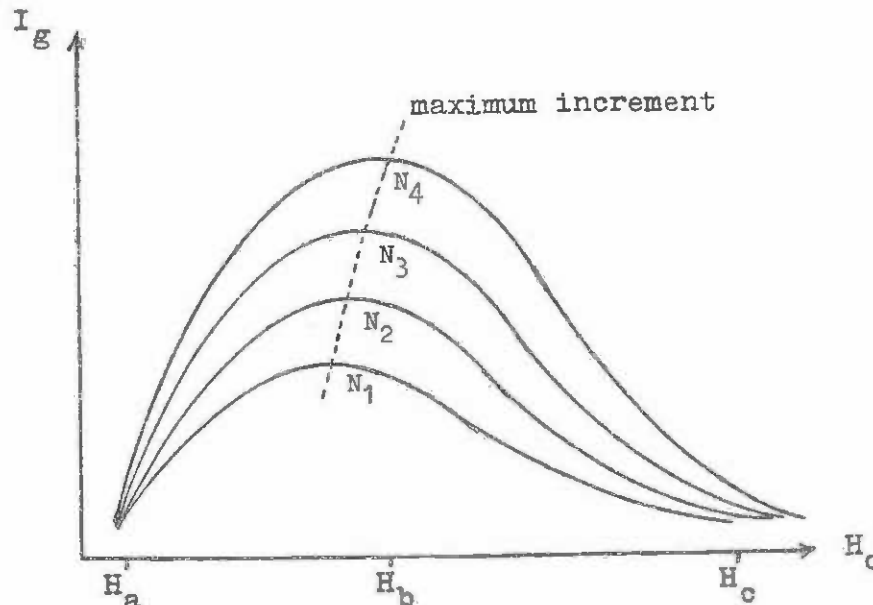
To produce yield predictions, the growth function must either be integrated mathematically or iteratively summed over a succession of years. This latter process is usually achieved by writing a small computer programme, which becomes by definition a computer simulation model of forest growth.

In the sections following, these techniques are discussed in more detail, with particular reference to functions of stand basal area increment. The techniques are not greatly affected by the use of mean diameter or volume as alternative parameters, but basal area increment is perhaps the simplest to model.

5.3.1 The Basal Area Increment Function

5.3.1.1 Basal area increment as a function of dominant height

Stand basal area increment for most plantation species, when plotted over dominant height, shows the following type of trend:



where I_g is current annual basal area increment per ha, H_0 is dominant height and N_1, N_2 , etc. are different levels of constant stocking, with N_4 greater than N_3 , which is greater than N_2 , and so on. The curve has three regions which can be distinguished for the purposes of practical data analysis:

1. The range H_a to H_b shows increment rising sharply to a maximum and then falling again. There are strong differences between stocking classes. This part of the curve is difficult to fit with simple regression models. With some species, it occurs at too early an age to be represented in sample plot data.
2. The range H_b to H_c shows a strong decline in basal area increment with time (increasing H_o) and marked differences between stocking classes. This part of the curve can easily be modelled by several of the regression functions of the type shown in figure A.2.1 in Appendix A. The following model includes the effect of stocking:

$$\ln I_g = b_0 + b_1 H_o^{-k} + b_2 N H_o^{-k} \quad -(1)$$

The coefficient k can be given assumed values (1, 2 etc.) or fitted by nonlinear regression. Natural logarithms (base e) are used to transform I as this simplifies the use of an integral form of this model (see section 5.3.3).

3. Beyond H_c , the differences between stocking classes becomes negligible and the basal area increment will be almost constant. This part of the curve is beyond the culmination of mean annual volume increment in most species, but will be relevant to yield prediction in long rotation sawlog stands. Model (1) above produces a satisfactory type of response in this region.

It is difficult, as has been noted, to find a well behaved linear function that fits the entire range of this curve without bias over any portion of it. Because the point H_b usually occurs early in the life of the stand, it is better to fit a function such as (1) above only to data beyond the maximum I_g , and to predict standing basal area G at H_b from a yield function in terms of stocking. For example:

$$G_o = a + b_1 N + b_2 N^2 \quad -(2)$$

where G_o is the initial basal area, at a defined dominant height H_b , and depends only on the stocking at that point in time, regardless of previous stand history.

5.3.1.2 Other methods of predicting basal area increment

Basal area increment can be predicted as a function of age or of standing basal area. When age is used, different sets of increment curves are required for each site index class. With standing basal area, satisfactory models cannot be obtained for stands with very variable histories. Consequently, the prediction of basal area increment from dominant height and stocking, as described above, is recommended.

5.3.1.3 Practical problems in analysis of increment data

Increment data is always very variable. This is the combined effect of:

1. Year to year variation in increment due to climatic fluctuations.
2. Instrumental error. Simple measurement systems such as diameter tapes may have the same order of error as the increment being measured.

A consequence of this is that large amounts of data are required to obtain satisfactory estimates of the regression coefficients. Furthermore, because the effects of different stand densities show out only weakly through the general variation, extreme differences in stand density must be present in the data if a density-dependent model is to be constructed.

The best increment functions may only have coefficients of determination of between 0.7 and 0.8. To test the suitability of a particular model under these circumstances, it must be used in its integral form (or as part of a simulation model) to predict final yields. Small biases in the increment function which do not show up graphically or by direct residual analysis can produce large errors in yield prediction because error in the increment function is always cumulative.

5.3.2 Constructing a Growth Model with Spacing Experiment Data: Marsh's Response Hypothesis

It is possible to develop growth models based on spacing experiments of the type described in section 2, using simplified graphical techniques, which can give accurate predictions for the growth of stands subject to a variety of different thinning treatments. The basis of this method is Marsh's hypothesis:

"The increment of thinned stands is equivalent to that of unthinned stands of the same stocking (measured in stems per unit area) and density (measured in basal area or volume per unit area), but of younger age (i.e. the age at which they had the same basal area or volume per unit area)". 1/

This hypothesis has been validated with several sub-tropical species and appears to give accurate and unbiased estimates of growth following thinning, provided that the growth is measured over not less than 3 years. The method proceeds as follows:

- (1) First graph the spacing experiment results using smoothed means for each stocking level, using basal area over age (figure 5.11).
- (2) Define the thinning treatment to be used in terms of both the basal area at which thinning takes place and the basal area and stocking to be removed.

Suppose for example we have a treatment schedule as follows:

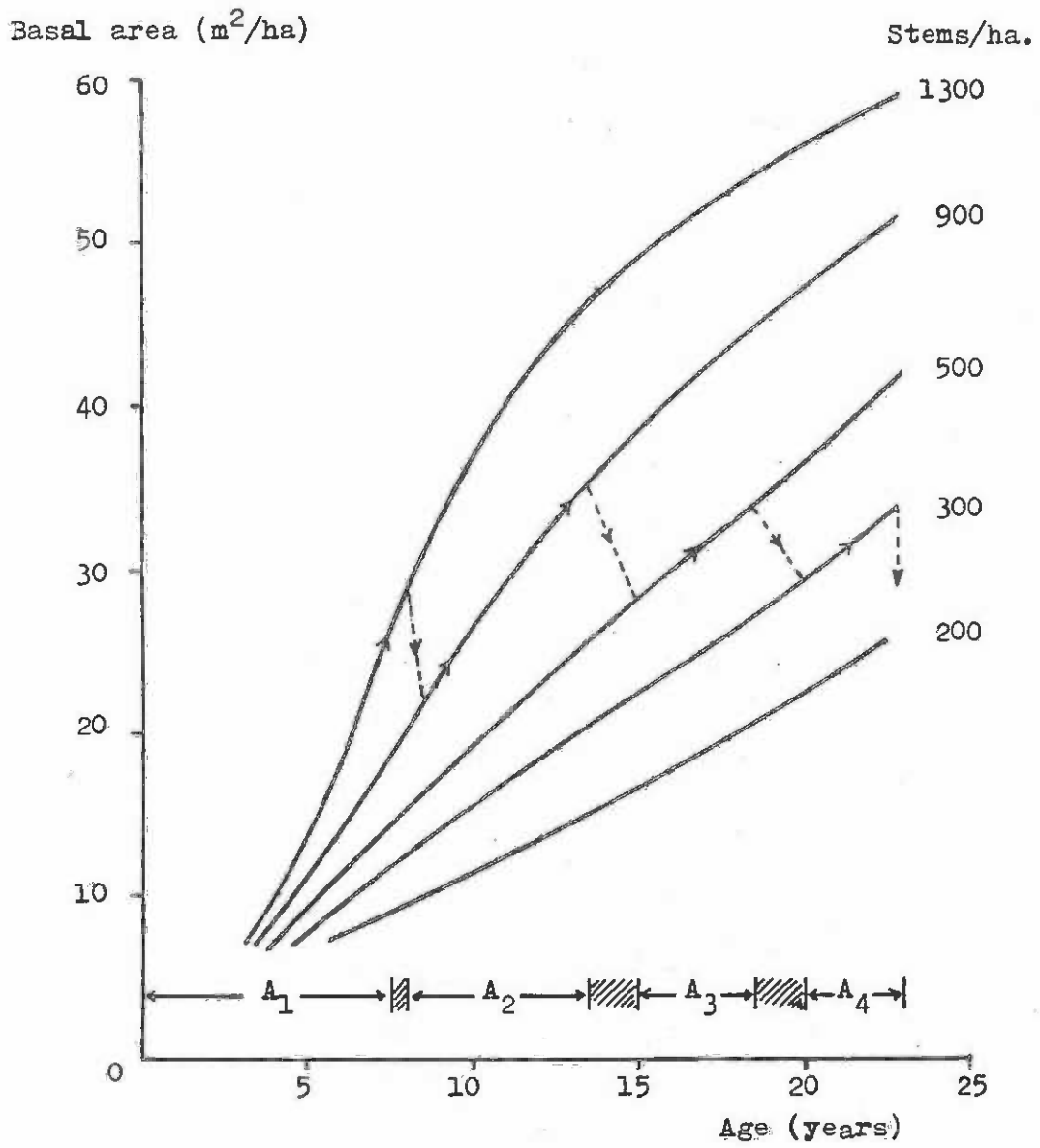
Plant at 1 300 trees/ha.

<u>Thin. no.</u>	<u>Thin at Basal Area</u>	<u>G after Thinning</u>	<u>N after Thinning</u>
1	28	22	900
2	35	28	500
3	35	30	300
CF	35	clearfell	

1/ Marsh, E.K. and Burgers, T.F. 1973. The Response of Even-aged Pine Stands to Thinning. Forestry in S. Africa, 14: 103-111.

Figure 5.11

Analysis of spacing experiment
using Marsh's hypothesis



- (3) Draw the treatment schedule on the graph (broken lines on figure 5.11) and from Marsh's hypothesis determine the time interval between thinnings.

<u>Thin no.</u>	<u>Apparent Age</u>	<u>Interval</u>	<u>Real Age at Time of Thinning</u>
1	0-8	8	8
2	8.5-13.5	5	13
3	14.5-18.5	4	17
CP	20-23	3	20

- (4) Reconstruct the basal area/age relationship using the correct age axis for the thinned stand. It is then possible to construct the various other yield statistics for the stand in relation to age.

A variation on this method allows a single diagram to be used for constructing yield curves for different site classes. This involves using dominant height, instead of age, as the x-axis of the graph. Height is then used as a site-dependent transformation of age with the age intervals being determined via the site index curves. This method can also be used where the spacing experiment data do not adequately cover a range of sites, to provide hypothetical yield tables for such sites. Note however that the substitution of dominant height for separate site-age combinations may not always give a sufficiently accurate representation of site-dependent responses.

Increment functions obtained from permanent sample plots of uncontrolled spacing or from tree increment cores can be used to construct curves for stand basal area of unthinned stands at different stockings, which may then be analysed graphically as above. This is convenient when the users of a particular growth model do not have access to computing equipment.

5.3.3 Conversion of Growth Models to Yield Models by Integration

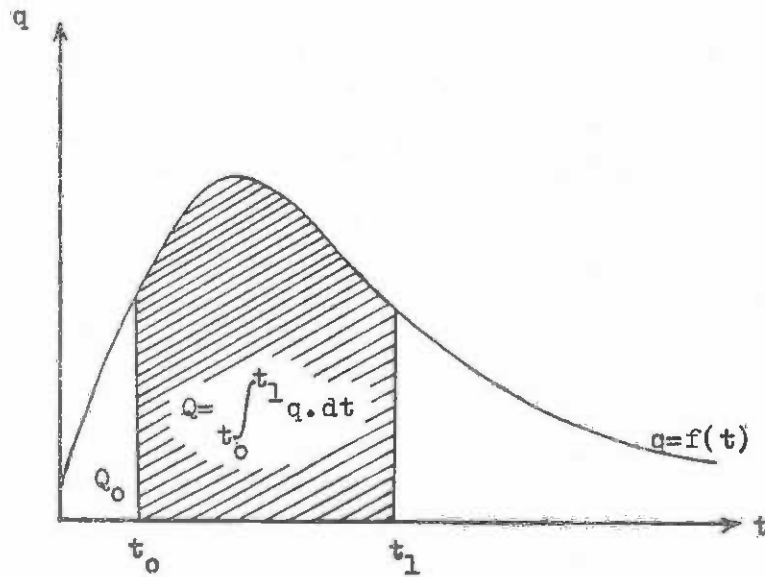
5.3.3.1 Introduction

The mathematical integration of a basal area increment function for a given series of thinning operations results in a basal area yield function for that thinning series. Mathematical integration has advantages over simulation as a way of using a growth model in that no special computing equipment is needed. On the other hand, many deceptively simple growth models (e.g. equation (1), section 5.3.1.1) may be very difficult to integrate.

This method assumes an elementary knowledge of differential and integral calculus on the part of the research worker.

5.3.3.2 Basic theory

If we have an increment function of the type shown below:



where q is any growth rate (basal area, diameter or volume growth) and t is the time axis, then the growth rate equation is represented very generally by:

$$q = f(t) \tag{1}$$

whilst the area under this increment curve is the total accumulated yield. Thus the accumulated yield up to time t , is given by:

$$Q = Q_0 + \int_{t_0}^{t_1} q \cdot dt \tag{2}$$

where Q_0 is the accumulated yield at the start of the period, t_0 .

For an unthinned stand, the basal area/age curve represents the integral function equivalent to equation (2) above with $Q_0 = 0$ and $t_0 = 0$. Hence any function fitted to data from unthinned stands provides an already integrated form of growth model. Consequently, if any data from a spacing experiment is fitted with a yield function dependent on stand density, then that yield function can be used to accurately predict the growth of thinned stands. This is a generalization of Marsh's hypothesis to any type of growth function using any variables.^{1/}

5.3.3.3 Application of an integral yield model to different thinning treatments

If an integral yield model has been fitted to data from unthinned stands which predicts basal area from dominant height and stocking, thus:

$$G = f(H_0, N) \tag{1}$$

^{1/} Marsh's hypothesis as discussed earlier is equivalent to integration of a function of the form:

$$I_g = f(G, N)$$

Then if the stand is thinned at a dominant height H_t to a stocking N_t , the yield up to the time of thinning will be:

$$G_t = f(H_t, N_0) \quad (2)$$

where N_0 is the initial stocking (after early mortality is accounted for). If the thinning ratio (C.f. section 5.4) is:

$$T_r = \frac{N'/N}{G'/G}$$

$$\therefore G' = \frac{N'/N}{T_r/G}$$

then for this particular thinning the standing crop basal area after thinning will be:

$$G'_t = \frac{N_t/N_0}{T_r/G_t} \quad (3)$$

whilst the yield at an interval $t + i$ after the thinning will be:

$$G_{t+i} = G'_t + f(H_{t+i}, N_t) - f(H_t, N_t) \quad (4)$$

5.3.3.4 Example of use of an integral yield model

Figure 5.12 shows data from a spacing experiment in Pinus patula stands at Kwira, Tanzania, designated Experiment 345. In all there were 8 treatments, laid out systematically, with 2 replicates for each treatment using rectangular plots of 0.08 ha. Figure 5.12 shows only the data from 4 plots at stockings of 173, 347, 694 and 1 388 stems/ha. A model to predict standing basal area was fitted to the whole set of data (192 observations) by weighted multiple linear regression, using the model:

$$\ln G = b_0 + b_1 H^* + b_2 E + b_3 EH^* \quad (5)$$

where $H^* = 1/(H_0 - 1.3)$

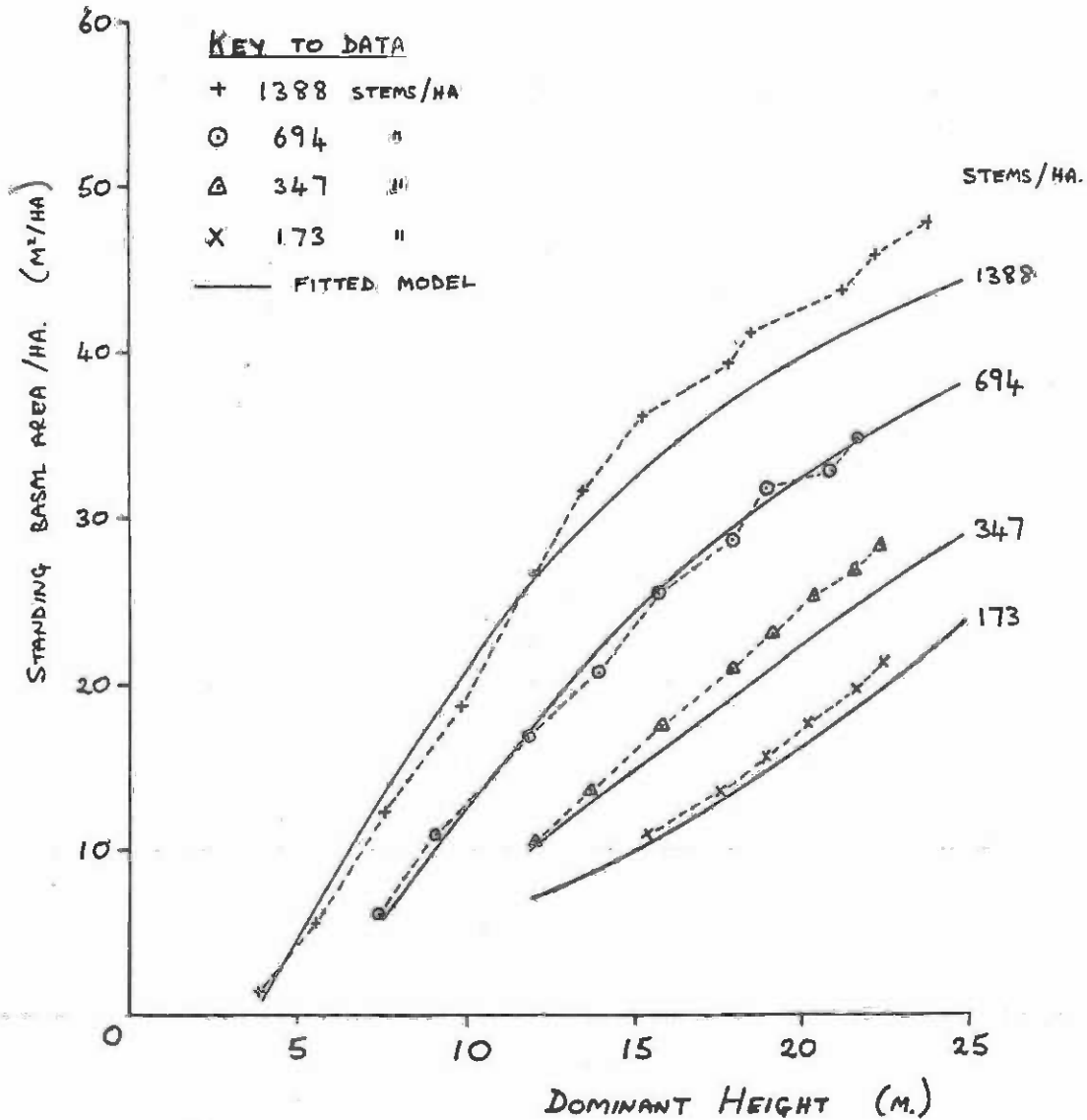
and $E = 100/\sqrt{N}$

The model was weighted by G to avoid the excessive bias towards lower values that often results when a logarithmic transformation is used. An R^2 of 0.95 was obtained with coefficient values:

b_0	4.0865
b_1	1.5991
b_2	0.047838
b_3	-4.3063

The model is shown on figure 5.12 overlaid on data from 4 treatments.

Figure 5.12 Spacing experiment 345 at Kwira, Tanzania, in Pinus patula stands



It was desired to construct a yield table for stands planted at 1 100 stems/ha, thinned once at 9 years using systematic thinning to 700 stems/ha and clear felled at 16 years.

For average sites, 9 years corresponded to a dominant height of 17 m, and the rotation age of 16 years to 24 m. The basal area at 9 years, prior to thinning, can be calculated directly from equation (5):

$$\begin{aligned}
 H^* &= 1/(17 - 1.3) \\
 &= 0.0637 \\
 E &= 100/\sqrt{1\ 100} \\
 &= 3.02 \\
 \ln G_9 &= 4.0865 + 1.5991 \times 0.0637 + 0.047838 \times 3.02 - 4.3063 \times 3.02 \times 0.0637 \\
 &= 3.5043 \\
 G_9 &= e^{3.5043} = 33.26 \text{ m}^2/\text{ha}
 \end{aligned}$$

Since the thinning from 1 100 to 700 stems/ha was to be systematic, the proportion of basal area removed would be the same as the proportion of stocking removed. From equation (3) above, with $T_x = 1.0$ by definition:

$$\begin{aligned}
 G'_9 &= \frac{700/1\ 100}{1/33.26} \\
 &= 21.17 \text{ m}^2/\text{ha}
 \end{aligned}$$

Therefore the basal area extracted in thinning was:

$$\begin{aligned}
 G_e &= 33.26 - 21.17 \\
 &= 12.1 \text{ m}^2/\text{ha}
 \end{aligned}$$

The yield at clearfelling is obtained from equation (4), where $f(H_{t+i}, N_t)$ corresponds to equation (5) entered with $H = 24$ m, and $N = 700$ stems/ha, giving:

$$f(H_{t+i}, N_t) = 35.98$$

and $f(H_t, N_t)$ is equation (5) entered using $H = 17$ m:

$$f(H_t, N_t) = 26.97$$

so that basal area yield at clearfelling is given by:

$$\begin{aligned}
 G_{CF} &= 21.17 + (35.98 - 26.97) \\
 &= 30.18 \text{ m}^2/\text{ha}
 \end{aligned}$$

5.3.3.5 Fitting compatible growth and yield models to increment data

When no suitable spacing experiment data is available for direct fitting of the integrated form of a yield model, then a growth model can be fitted to increment data from permanent plots or stem analyses which can subsequently be integrated to give a yield model. Clutter^{1/} proposed a growth model of the type:

^{1/} Clutter, J.L. 1963. Compatible growth and yield models for Loblolly Pine. Forest Sci. 9 (3): 354-371

$$I_g = G(a + bS - \ln G)A^{-1} \quad (6)$$

where I_g is basal area increment/ha/year, S is site index and a and b are coefficients which can be estimated by linear regression by fitting the function:

$$\left(\frac{AI_g}{G} + \ln G\right) = a + bS \quad (7)$$

where the expression on the left hand side is the dependent Y variable and site index S is the predictor variable. This model integrates to give the yield equation:

$$\ln G_t = a + bS - \frac{A_o}{A_t} (a + bS - \ln G_o) \quad (8)$$

where G_t is the yield at the end of the period and G_o is the yield at the start of the period. No thinnings may occur during the prediction interval, but the basal area after thinning becomes G_o for predicting the basal area immediately prior to the next thinning.

5.3.4 Use of Growth Models by Simulation

5.3.4.1 Requirements for a simulation model

Simulation modelling is a much easier method of using a growth model than mathematical integration in the majority of cases, but it does normally require access to a small computer. To use a growth model by simulation, the following points must be defined from permanent sample plots or experimental data:

- (1) The height/age/site index function:

$$H_o = f_1(A, S)$$

- (2) The basal area increment function, in terms of

height
or age
and site index

and

basal area
and / or
stocking/ha

$$I_g = f_2(H_o, N)$$

or $I_g = f_2(H_o, G)$

or $I_g = f_2(A, N, S)$

or $I_g = f_2(A, G, S)$

c.f. Clutter's model,
section 4.1.1.5 for an example

or other combinations of A, G, N, S, H_o on the right hand side.

- (3) The relationship between stocking removed and basal area removed for each thinning type:

$$N/N = f_3(G/G)$$

- (4) The volume function from height and basal area:

$$V = f_4(H_o, G)$$

- (5) The initial basal area of the stand at a certain reference age or height at which the simulation commences, as a function of the planted stocking:

$$G_o = f_5(N_o)$$

These functions need not be continuous mathematical relationships. Each function may be a set of equations, one of which is selected by a conditional process; or it may be a set of tabulated values, as will be seen in the example below.

5.3.4.2 Method of construction of a simulation model

Simulation models are constructed by coding the functional relationships and the logical interconnections between them, as statements in a computer language, usually FORTRAN, BASIC, or ALGOL. This 'source language' model is then 'compiled' by the computer into its own internal machine code and run in the computer. As it runs, it will require data defining:

1. The site index and initial spacing of the stand
2. The timing and intensity of thinnings.

The programme will then generate output, in the form of a stand table or graph, as designed by the computer programmer. The simulation can be divided into two logically distinct phases:

- (1) The initialization phase, where the initial values of stand basal area, height and stocking are defined and table headings are printed out.
- (2) The dynamic phase in which the growth of the stand over A is computed and added to existing growing stock; any required harvesting operations are carried out; and the age of the stand is incremented by B and the process repeated. During the dynamic phase, summaries of growing stock are printed out. The dynamic phase is terminated when the stand is clearfelled or a previously set time limit is reached. (A is a period of time (usually one to five years) and B is the length of the period.)

5.3.4.3 Example of a simple simulation model

In this example, we construct a simple simulation model in BASIC for even-aged Pinus patula stands. The functions in the model are as follows:

- (1) The height-age function is taken from Alder 1/ and is:

$$\ln H = b_0 + b_1/A + ((\ln S - c_1)/c_2) (b_2 + b_3/A)$$

where b_0, b_1, b_2, b_3 are obtained from regression analysis from PSP data and for P. patula which have the values:

$$\begin{aligned} b_0 &= 3.6068 \\ b_1 &= -17.513 \\ b_2 &= 0.008057 \\ b_3 &= 0.3308 \end{aligned}$$

whilst c_1, c_2 are defined as:

$$\begin{aligned} c_1 &= b_0 + b_1/15 \\ c_2 &= b_2 + b_3/15 \end{aligned}$$

The index age for site index is 15 years.

- (2) The basal area increment function is derived from Experiment 345, Kwira, Tanzania, by tabulating mean basal area increment by classes of stocking and dominant height, giving the result shown in Table 5.1. This table is used directly in the programme without further analysis.
- (3) The thinning ratios are defined by the following simple assumptions:
- (a) Systematic thinning: $T_r = 1.0$
- (b) Selective low thinning: First thinning $T_r = 0.7$
Later thinnings $T_r = 0.9$
- (4) The volume function is the Pinus patula stand volume equation for Kenya:

$$V_t = N(-0.0072 + 0.00002887 Dg^2 + 0.00002077 H D_g + 0.000032765 DgH)$$

This gives total volume. The merchantable volume to 20 cm diameter top is calculated by:

$$V_m = V_t (0.97352 - 21.9737 \exp(-0.15407 Dg))$$

- (5) Initialization of basal area is done by using a stand basal area of zero at age 3 years.

1/ Alder, D. 1977. A Growth and Management Model for Coniferous Plantations in East Africa. D. Phil. thesis, Oxford University.

Table 5.1

VALUES OF BASAL AREA INCREMENT TABULATED BY STOCKING AND HEIGHT
FOR EXPERIMENT 345, KWIRA, TANZANIA

Dominant Height, m.	Spacing class, m.						
	2-3	3-4	4-5	5-6	6-7	7-8	8-9
3-5	3.89	2.87					
5-7	6.59	3.69					
7-9	7.22	4.49					
9-11	7.33	6.10	3.83				
11-13	5.90	5.39	4.55	3.56	3.17		
13-15	4.59	4.59	4.41	3.50	2.84	1.91	
15-17	2.82	3.30	4.35	2.93	2.34	2.18	1.51
17-19	2.24	2.22	3.34	2.50	2.00	2.11	1.54
19-21	2.09	2.03	2.40	1.73	1.42	1.72	1.43
21-23	2.01	1.57	1.99	1.32	1.28	1.37	1.27
23-25	1.86	1.93	1.59				

Values within the table are basal area increment per ha per annum.
E.g. at a spacing of 4.5 m and a dominant height of 12 m, increment
is 4.55 m²/ha/yr.

The programme listing is given below:

```

5 SELECT PRINT 005(64)
10 REM EXAMPLE OF STAND GROWTH SIMULATION MODEL
20 REM INITIALIZATION PHASE
30 PRINT "SIMULATION MODEL FOR PINUS PATULA PLANTATIONS"
40 PRINT "GIVE INITIAL STOCKING AND SITE INDEX, PLEASE"
50 INPUT N,S
60 REM READ TABULATED INCREMENT VALUES
70 DIM T(7,11)
80 MAT READ T
90 DATA 3.89,6.59,7.22,7.33,5.90,4.59,2.82,2.24,2.09,2.01,1.86
100 DATA 2.87,3.69,4.49,6.10,5.39,4.59,3.30,2.22,2.03,1.57,1.93
110 DATA 0,0,0,3.83,4.55,4.41,4.35,3.34,2.40,1.99,1.59
120 DATA 0,0,0,0,3.56,3.50,2.93,2.50,1.73,1.32,0
130 DATA 0,0,0,0,3.17,2.84,2.34,2.00,1.42,1.28,0
140 DATA 0,0,0,0,0,1.91,2.18,2.11,1.72,1.37,0
150 DATA 0,0,0,0,0,0,1.51,1.54,1.43,1.27,0
152 REM READ SITE INDEX CURVE COEFFICIENTS
153 READ B0,B1,B2,B3
154 DATA 3.6068,-17.513,0.008057,0.3308
    
```



```
155 LET C1=B0+B1/15
156 LET C2=B2+B3/15
160 REM SET STAND BASAL AREA TO ZERO, AGE TO 3
170 LET G=0
180 LET A=5
190 REM READ DETAILS OF STAND TREATMENT
200 PRINT "HOW LONG IS THE ROTATION, PLEASE"
210 INPUT R
220 DIM A1(10),N1(10),T$(10)
230 PRINT "HOW MANY THINNINGS, PLEASE"
240 INPUT M
250 IF M = 0 THEN 300
260 FOR I=1 TO M
270 PRINT "GIVE AGE,STEMS/HA LEFT, AND TYPE(S/L) FOR THINNING ";I
280 INPUT A1(I),N1(I),T$(I)
290 NEXT I
300 LET M=1
305 SELECT PRINT 215(120)
310 REM PRINT TABLE HEADINGS
320 PRINT
330 PRINT TAB(20);"GROWING STOCK";TAB(80);"THINNINGS"
340 PRINT
350 PRINT USING 360
360 % AGE HDOM STEMS/HA DIAM(G) BA/HA VOL/HA VOL/HA(20CM)
VOL/HA VOL/HA(20CM)
370 REM DYNAMIC PHASE OF SIMULATION
380 REM CALCULATE DOMINANT HEIGHT
390 LET H=EXP(B0+B1/A+((LOG(S)-C1)/C2)*(B2+B3/A))
400 REM SELECT HEIGHT AND STOCKING CLASS FOR GROWTH INCRMENT
410 LET H1=(H-1)/2
411 IF H1<12 THEN 420
412 LET H1=11
420 LET E1=SQR(1/N)*100-1
430 IF T(E1,H1)=0 THEN 920
440 REM ADD INCREMENT TO CURRENT BASAL AREA
450 LET G=G+T(E1,H1)
460 REM TEST IF THINNING REQUIRED IN CURRENT YEAR
470 IF A=A1(M) THEN 670
480 REM CALCULATE MEAN BA DIAMETER, & VOLUMES
490 LET D =SQR(G/(N*0.00007854))
500 LET V=N*(-0.0072+0.00002887*D!2+0.00002077*H*D+0.00003276*D!2*H)
510 LET V1=V*(0.97352-21.9737*EXP(-0.15407*D))
515 IF V1 > 0 THEN 520
516 LET V1 = 0
520 REM PRINT GROWING STOCK DETAILS
530 PRINT USING 540,A,H,N,D,G,V,V1;
540 % ### ##.# ##### ##.# ##.# ##.# ##.# ##.#
550 REM PRINT THINNING DETAILS IF A THINNING WAS PERFORMED
560 IF T2=0 THEN 600
570 PRINT TAB(64);
580 PRINT USING 590,N2,D1,G1,V2,V3;
590 % ### ##.# ##.# ##.# ##.# ##.#
600 REM INCREASE AGE BY 1 YEAR, & REPEAT DYNAMIC PHASE
605 PRINT
610 LET A = A+1
620 LET T2=0
630 IF A(= R THEN 370
640 PRINT
650 PRINT TAB(20);"END OF SIMULATION"
660 STOP
```

```
670 REM SECTION TO SIMULATE LOW OR SYSTEMATIC THINNING
680 REM DETERMINE THINNING RATIO FROM THINNING TYPE
690 IF T$(M)="L" THEN 720
700 LET T3=1
710 GOTO 760
720 IF M)1 THEN 750
730 LET T3=0.7
740 GOTO 760
750 LET T3=0.9
760 REM CALCULATE STOCKING REMOVED N2 & BA REMOVED G1
770 LET N2=N-N1(M)
780 LET G2=N1(M)/N/(T3/G)
790 LET G1=G-G2
800 REM ADJUST STOCKING & BA OF RESIDUAL STAND
810 LET G=G2
820 LET N=N1(M)
830 REM COMPUTE MEAN DIAMETER AND VOLUMES OF THINNINGS
840 LET D1=SQR(G1/(N2*0.00007854))
850 LET V2=N2*(-0.0072+0.00002887*D1!2+0.00002077*H*D1!2+0.00003276*D1!2*H)
860 LET V3=V2*(0.97352-21.9737*EXP(-0.15407*D1!2))
861 IF V3 ) 0 THEN 870
862 LET V3 = 0
870 REM INCREMENT THINNING NUMBER AND RETURN TO MAIN PART OF
880 REM SIMULATION
890 LET M=M+1
900 LET T2=1
910 GOTO 480
920 REM SECTION TO ABANDON PROGRAM WHEN SIMULATIONS GOES
930 REM OUTSIDE LIMITS OF AVAILABLE GROWTH DATA
940 PRINT TAB(20);"SIMULATION EXCEEDS RANGE OF GROWTH FUNCTION"
950 PRINT TAB(20);"RUN ABANDONED"
960 END
```

The output produced by one run is shown in figure 5.13. Table 5.2 lists the definitions of each of the variables used in the programme. This simulation programme is intended to be an example of the flexibility of the technique and is not intended to suggest that the specific functions used are in any sense the best or most preferable methods.

Figure 5.13

Computer printout produced by stand growth
simulation model

SIMULATION MODEL FOR PINUS PATULA PLANTATIONS
 GIVE INITIAL STOCKING AND SITE INDEX, PLEASE
 1700 23
 HOW LONG IS THE ROTATION, PLEASE
 16
 HOW MANY THINNINGS, PLEASE
 1
 GIVE AGE,STEMS/HA LEFT, AND TYPE(S/L) FOR THINNING 1
 8 700 S

GROWING STOCK							THINNINGS				
AGE	HDOM	STEMS/HA	DIAM(G)	BA/HA	VOL/HA	VOL/HA(20CM)	STEMS/HA	DIAM(G)	BA/HA	VOL/HA	VOL/HA(20CM)
5	6.1	1700	7.0	6.5	8.6	0.0					
6	8.5	1700	10.1	13.8	45.3	0.0					
7	10.8	1700	12.5	21.1	96.0	0.0					
8	12.9	700	14.2	11.1	61.8	0.0	1000	14.2	15.9	138.8	0.0
9	14.8	700	16.9	15.7	101.6	0.0					
10	16.5	700	18.6	19.0	137.7	0.0					
11	18.1	700	19.6	21.2	168.4	0.0					
12	19.5	700	20.5	23.2	198.7	10.0					
13	20.7	700	21.4	25.3	230.1	38.5					
14	21.9	700	22.1	26.8	257.8	63.1					
15	22.9	700	22.7	28.4	285.9	89.4					
16	23.9	700	23.5	30.3	317.9	122.6					

END OF SIMULATION

Table 5.2

DEFINITION OF VARIABLES USED IN STAND GROWTH
SIMULATION PROGRAMME

<u>Variable</u>	<u>Size</u>	<u>Description of use</u>
A		Stand age, years
A1	10	List of ages for thinnings
B0]	Coefficients for height growth function, representing b_0, b_1, b_2, b_3 in text
B1		
B2		
B3		
C1]	Coefficients for height growth function, representing c_1, c_2 in text
C2		
D		Mean basal area diameter of growing stock
D1		Mean basal area diameter of trees removed in thinning
E1		The stand density class, going from 1 for spacings of 2-3 m to 7 for spacings of 8 m or more
G		The basal area/ha of the growing stock
G1		The basal area/ha of trees removed in thinnings
G2		The basal area of growing stock immediately after thinning
H		The dominant height of the stand
H1		The dominant height class, going from 1 for heights 3-5 m, to 11 for heights over 23 m
I		A counter for thinning number during the reading in of thinning details
M		Initially, the number of thinnings. During the dynamic phase, represents the next thinning number due
N		The stocking/ha of the stand
N1	10	The list of residual stockings to be left after each thinning
N2		The number of trees per ha removed in a thinning
R		The rotation age for the stand, in years
S		The site index of the stand, in m dominant height at age 15
T	7, 11	The table of diameter increments corresponding to the spacing class (1st dimension) and height class (2nd dimen- sion). Values as per table 5.1 in text.

<u>Variable</u>	<u>Size</u>	<u>Description of use</u>
T2		Set to 1 if the stand has been thinned in the current year of simulated time, otherwise set to 0. Controls the format of printing.
T3		Gives the thinning ratio T_r , as defined in the text for each type of thinning. Set to 1 for systematic thinning, 0.7 for 1st low thinning, 0.9 for other low thinnings.
T\$	10	A list of thinning types, for each thinning. "S" represents systematic thinning and "L" thinning from below.
V		The total volume over bark per ha of the growing stock.
V1		The volume o.b./ha to a top diameter limit of 20 cm, for the growing stock.
V2		Total volume/ha of thinnings.
V3		Volume/ha to 20 cm limit of thinnings.

5.4 THINNING

The economic component of growth in uniform stands managed for timber will be partly removed in commercial thinnings. In addition, non-commercial thinnings may be carried out to reduce stand density and give a corresponding greater stand mean diameter in subsequent thinnings and the final crop.

Obviously, the modelling of thinning is central to any exercise in yield prediction. A thinning operation is very largely characterized by the number of stems per ha and the basal area per ha removed. In single tree models, it is necessary to describe the distribution of removed trees; but this is not necessary for stand models, which are the main subject of this section.

5.4.1 The Thinning Ratio

A useful way of characterizing thinnings is in terms of the thinning ratio:

$$\text{Thinning ratio} = \frac{\text{No. of stems left/No. stems before thinning}}{\text{Basal area left/Basal area before thinning}}$$

With a little algebra, the thinning ratio can be seen as being equivalent also to:

$$\text{Thinning ratio} = \frac{(\text{Mean BA diameter before thinning})^2}{(\text{Mean BA diameter after thinning})^2}$$

Typical values for the thinning ratio for different types of thinning are:

<u>Thinning type</u>	<u>Value of thinning ratio</u>
Early low thinnings	0.6-0.8
Later low thinnings	0.8-1.0
Non-selective thinning	1.0
Crown thinning	1.1-1.3

The thinning ratio can be determined empirically using temporary plots on which thinnings are carried out by contractors or which are marked for thinning without actually felling the trees. The first technique gives a more realistic assessment, but may not be appropriate for experimental types of thinning which one simply wishes to evaluate on a dynamic model.

Estimates of thinning ratios obtained from permanent sample plots may not be reliable, as PSPs are normally clearly marked in the forest and hence are unlikely to receive typical treatment.

5.4.2 Estimating Thinning Yields in Static Models

In order to estimate the thinning yield for a particular static yield function, then the actual thinning intensity, defined in terms of stocking removed or basal area removed, and the thinning ratio, as defined above, must be known. It is also necessary to know whether the yield function is based on measurements made immediately before thinning, immediately after thinning or, as is most normal, an uncontrolled mixture of measurements before, after and between thinnings.

Assuming that the yield function predicts mean basal area diameter and it is based on an uncontrolled mixture of data, then the diameter predicted at the time of thinning may be assumed to be a mean of before and after thinning diameters, viz:

$$D_p = (D_b + D_a)/2 \quad (1)$$

where D_p is diameter predicted from the static yield function and D_b and D_a are the before and after thinning diameters at that point in time, which have to be determined.

The thinning ratio also relates the unknown diameters D_a and D_b , giving two simultaneous equations:

$$Tr = D_b^2 / D_a^2 \quad (2)$$

These two equations can be solved to give D_a and D_b :

$$D_a = 2D_p / (1 + \sqrt{Tr}) \quad (3)$$

$$D_b = 2D_p - D_a \quad (4)$$

Knowing the before and after thinning mean basal area diameters, then it is possible to work out the basal area removed knowing the stocking removed or the stocking removed knowing the basal area removed.

In the case where thinning intensity is specified in terms of stocking, then the mean basal area diameter of the trees removed in thinnings is given by:

$$D_t = D_a \sqrt{\left\{ \frac{Tr \cdot M_b - N_a}{N_b - N_a} \right\}} \quad (5)$$

or

$$D_t = D_b \sqrt{\left\{ \frac{M_b - N_a / Tr}{N_b - N_a} \right\}} \quad (6)$$

Note that equations (5) and (6) are independent of the assumption (1) that predicted diameters are averages of values immediately before and after thinning. Thus either equation (5) or (6) can be used to determine mean diameter of thinnings if the functions are defined in terms of predictions of after thinning diameters or before thinning diameters respectively.

All the algebra leading to the above formulae is based on the simple relation:

$$G = N.k.D_g^2$$

where G is basal area, N is stocking, k is the constant 0.00007854 for diameters in centimetres, basal area in m²/ha and stocking in numbers/ha, and D_g is the mean basal area diameter. Once the mean diameter of thinnings is known, then total and merchantable volumes can be readily calculated from a suitable volume equation.

5.4.3 Estimating Thinning Yields in Dynamic Models

As with static yield models, both the thinning intensity and the thinning ratio must be known. From these two statistics, the basal area removed for a given stocking or alternatively the stocking removed for a given basal area can be readily determined and both parameters of the growing stock adjusted for the thinning, with thinning removals being calculated directly. Examples have already been discussed in sections 5.3.3 and 5.3.4.

An essential feature of thinning in dynamic models is that the subsequent growth is influenced by the intensity of thinning, as stand density will be modified following treatment. In a static model this does not happen and, hence, the danger of obtaining incorrect results with static models by applying thinning treatments that do not correspond to the fitted functions.

5.5 MORTALITY

Mortality can often be regarded as negligible in many types intensively managed-uniform forest, once a particular crop has become established. In other cases, however, there is a marked reduction of stem numbers over the passage of time which must in some way be accounted for by the growth and yield model. There are several types of mortality which must be considered.

5.5.1 Establishment Mortality

Establishment mortality defines the percentage of viable seedlings which fail to survive the first year. In the case of plantations, it is obviously quite easy to define in the field, but in the case of stands established by artificial or natural seeding, it is probably better to consider the absolute number of seedlings that become established, rather than to consider a ratio defining percentage survival or percentage stocking.

Survival can also be defined in terms of percentage of the area that is fully stocked, as discussed in section 6.

The factors that most strongly influence survival and must be considered in any model constructed to predict it are:

- The method and intensity of site preparation.
- The weather at the time of establishment and for the months following.

- The aspect of the site.
- The degree of exposure of the site.
- The nature of the top 10 cm of the soil, in particular, and of the soil generally.

For plantations, the following must also be considered:

- The age of the seedlings.
- Procedures for handling the seedlings from nursery to the forest.
- The season of the planting in relation to dormancy.

For seeding by artificial or natural methods, one must consider:

- Pesticidal coatings or treatments to reduce seed predation.
- Seed pre-treatment to stimulate or improve germination.

Obviously, determination of all these parameters and their inclusion in a useful quantitative model is virtually impossible. It is better to construct simple and robust models that can easily be related to practices followed in a particular area or point of time.

The model will usually be in the form of a multiple regression between survival (relative or absolute) and two or three of the above parameters, coded in appropriate ways.

5.5.2 Density-Dependent Mortality

Density-dependent mortality may be a direct result of suppression, but more usually is indirect, in that the less vigorous trees, with crowns lower in the canopy, become susceptible to disease or animal damage to a much greater degree than more vigorous trees.

This type of mortality can usually be omitted from models of stands planted at wide spacings and grown on short rotations or subject to adequate thinning. However, many growth and yield predictions relate to unthinned stands or to stands in which thinning has been delayed and hence some effort must be made to include this type of mortality.

A useful model is the Reineke line, in which stocking is plotted against dominant height, using logarithmic scales on both axes. This gives a diagram of the type shown in figure 5.15. The Reineke line defines the maximum stocking that can be sustained at any given dominant height.

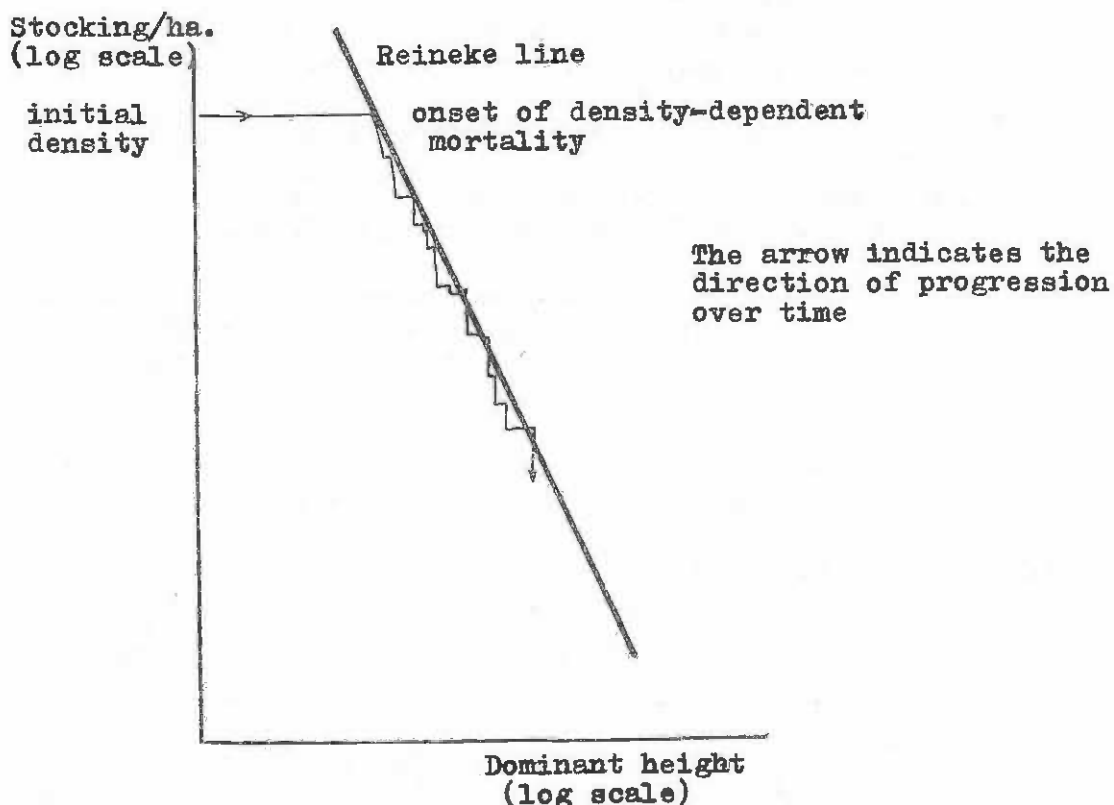
If stands are normally well thinned, it may be difficult to obtain the necessary data to construct this line. There are advantages in establishing in a given plantation a series of about 10 plots of forest established at 1 x 1 or 2 x 2 m on widely different sites to provide this data. This obviously does not apply in cases where there are extensive unthinned stands.

In static yield models, density-dependent mortality is implicit in the mean diameter/dominant height functions and need not be explicitly defined.

In dynamic growth models, the reduction in stocking caused by mortality will be effected by a simulated low thinning, with a thinning ratio around 0.6, whenever the stocking/height relationship moves to a point to the right of the Reineke line.

Figure 5.15

Relationship between stocking and dominant height over time
in an unthinned, heavily stocked uniform forest



Note that although the Reineke line will be very well defined for a single stand, its position and slope may be dependent on site. The lowest part of the line may also tend to bend downwards, giving a slight curve.

5.5.3 Disease and Pest Mortality

As mentioned in section 5.5.2, some aspects of disease and pest related mortality can be implicitly included in a simple density-dependent model. However, there are many cases where the disease or pest incidence is epidemic in nature and tends to occur as outbreaks following particular patterns of weather and moving outwards from epicentres of infection. Techniques for modelling such problems are somewhat beyond the scope of this manual. Such models are often stochastic, in that only the probability of an outbreak can be predicted.

The probability of an outbreak may be a function of weather pattern, condition of the forest growing stock or of distance from another outbreak area. From a managerial point of view, the probability of an outbreak can be used to assess the expected value of the crop at some future point in time and hence the opportunity cost of pre-emptive felling or of taking or failing to take control measures.

5.5.4 Windthrow and Fire Damage

As with pest and disease problems, both windthrow and fire damage effects have density-dependent and density-independent effects, with the latter tending to be catastrophic in nature, i.e. destroying the whole stand. Again, the catastrophic component can only be predicted as a probability of occurrence, dependent on climatic and site factors, as well as the condition of the growing stock. With fire, additional factors, such as the time since the last controlled burning and the use of low pruning, may be important variables in determining the probability of catastrophic damage.

The successful prediction of the probability of a catastrophic event is extremely difficult where such events are rare. Where fire, wind damage or certain pests or diseases are normal, such models may be constructed; but it is still likely to require a careful process of data gathering and evaluation over about 20 years before anything useful may emerge. It should also be noted that the parameters for such models must be continually revised, as they are likely to be very sensitive to climatic fluctuations.

5.6 STAND VOLUME PREDICTION

In the foregoing parts of section 5, various methods of predicting dominant height, basal area, stocking and mean basal area diameter have been discussed. The final stage in a yield prediction system for a single stand is to use these variables to predict total and merchantable volume.

Part I of this manual has dealt exhaustively with the techniques of measuring and modelling the volume of individual trees. This section only concerns itself with a resumé of methods particularly appropriate to models of stand volume from the parameters normally calculated in a yield model.

5.6.1 Stand Volume Equations Based on Dominant Height and Basal Area

Total stand volume, either to the tip or to some small diameter limit such as 10 cm, can be accurately predicted by an equation using basal area and dominant height. Typical equations are:

- (1) $V = b_0 + b_1.G.H$ (linear)
- (2) $V/G = b_0 + b_1.H$ (linear, weighted by G)
- (3) $\log V = b_0 + b_1.\log G + b_2.\log H$ (logarithmic)
- (4) $V/G = b_0 + b_1.H + b_2.H^2$ (quadratic, weighted by G)

A wide variety of other models is possible. As has been discussed in Part I, some form of weighting is normally desirable for fitting volume equations, as the residual error tends to be approximately proportional to the volume. Equations (2), (3) and (4) all achieve this. Model (2) is very simple to fit graphically.

The quantity V/G is widely known as form height and has the dimension metres.

The degree of fit obtained when stand total volume is regressed with basal area and dominant height with models such as the above is usually very high, often with correlation coefficients exceeding 0.99.

5.6.2 Using Tree Volume Equations to Predict Stand Volume

It is a fairly commonplace practice to use a tree volume equation to predict stand volume by entering the equation with mean basal area diameter and Lorey's mean height. The latter is the mean of a systematic sample of tree heights weighted by tree basal area. Alternatively, arithmetic mean height or dominant height are used.

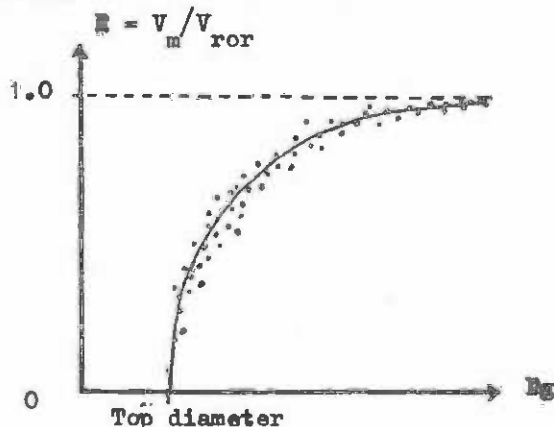
This method gives a biased result in most cases because the equation coefficients have not been influenced by the changing diameter distribution and form of trees with stand parameters.

In many cases the bias, which is normally towards underestimation of stand volume, is small enough to be regarded as negligible. The method will tend towards serious underestimation of volume however if the equation is predicting volume to a cut-off diameter and the mean diameter is close to this limit.

5.6.3 Estimation of Volume to a Top Diameter Limit

It is often the case in uniform forests that volume yields are required to be estimated for two or three top diameter limits. This may be done as follows:

- (1) Fit a stand volume equation as described above (s. 5.6.1.), either overbark or underbark, as required.
- (2) For each sample plot, calculate the ratio of volume to the selected top diameter limit to total volume.
- (3) A graph of these ratios against stand mean diameter will produce a diagram of the type shown below.



- (4) This shape can be fitted by the function:

$$R = 1 - a \cdot \exp(b \cdot D_g)$$

which can be transformed into:

$$\log(1 - R) = a^* + b \cdot D_g$$

and fitted by linear regression. Note that a is $\exp(a^*)$.

With several merchantable ratios, an attempt can be made to harmonize the coefficients so that they are themselves functions of the top diameter limit. Alternatively, one could attempt to construct a multivariable extension of the above model incorporating merchantable diameter.

When fitting the model, it is important that data points with zero ratios are excluded. When the function is actually used, if a negative ratio is predicted, the merchantable volume should be taken as zero.

5.6.4 Volumes of Thinnings

Various methods of calculating thinning volumes can be adopted. The stand volume equation may be entered using the basal area of thinnings. Alternatively, a tree volume table may be entered with thinning mean diameter and estimated mean height (perhaps from a height/diameter regression). The most unbiased method is to construct a stand volume equation using only thinnings, i.e. a regression of thinned volume on thinned basal area and stand dominant height. This regression can be tested to see if it differs significantly from the general stand volume equation and, if not, the latter may be used.

Because thinning volumes are often not measured, it may not be possible to construct such a model. It is suggested that in such cases, thinned volume is calculated as the difference between stand volume immediately before and after thinning. This technique may be applied to both total and merchantable volumes.

5.7 ADVANCED TECHNIQUES OF GROWTH AND YIELD PREDICTION

There are two types of model which can be applied to uniform stands which have not been referred to directly in the foregoing sections, in spite of the fact that they are in use in a number of forestry organizations. These are size class models and tree position models.

5.7.1 Size Class Models

In a size class model of a uniform stand, the growing stock at any point in time is described by a frequency distribution of tree sizes. The increment on each size class is calculated separately, usually as a function of site, age and stand density and the relation between the size class diameter and the mean or dominant diameter.

Thinnings and mortality must be described in terms of frequency distributions also, with the growing stock distribution being modified for each thinning or mortality event.

The advantage of this type of model for uniform stands is that it gives more detailed information about the size assortment of yield, especially if a stem taper function is also used for volume calculation. The disadvantage is that the analysis of the growth data is more complex than for whole stand models and requires consideration of additional parameters for diameter distributions; whilst the use of the model necessitates access to an electronic computer.

Models of this type are not necessarily more accurate in predicting whole stand parameters such as basal area or mean diameter; but they do provide more detailed information about the stand.

5.7.2 Tree Position Models

Tree position models are those in which the actual spatial relationship between all the trees in a simulated plot are represented by the model. Inter-tree competition is dependent on the relative sizes and positions of neighbouring trees. This information is used to provide a competitive index for each tree which is used to reduce the actual growth of the tree, relative to that which it would have if it were growing free of any competition. The latter is dependent on the age and site.

Tree position models require a formidable number of calculations and stretch the resources of even quite large computers. As a technique for modelling the growth of uniform stands, they are unnecessarily complex, since they do not provide significantly more useful information about stand growth than a size class model, whilst requiring probably two orders of magnitude more calculations. The principle advantages of tree position models are in relation to physiological research and to growth and yield prediction in mixed stands.

For readers interested in pursuing the literature on these more advanced types of model, the bibliography published by the Commonwealth Agricultural Bureau entitled "Computerized Methods in Forest Planning and Forecasting" is recommended. Details are in Appendix D.

6. ANALYSIS OF GROWTH AND YIELD DATA FOR MIXED FOREST

Mixed forest, in the context of this manual, implies any forest in which the individual stands that compose the basic mensurational unit contain a mixture of age classes. Typically, mixed forest will contain also a variety of species, which may be ecologically quite similar or composed of several ecological groups, with each group tending to be dominant in a particular stratum of the canopy or on particular microsites or in different successional phases following gap formation or harvesting operations.

The distinction between mixed and uniform forests is not absolute. There are cases in which a mixed stand consists essentially of only two age classes - very small understorey regeneration and a dominant overstorey - and where harvesting is essentially a clear felling operation. In these cases, the techniques of uniform stand yield prediction may be largely applicable. Similarly, stands of uniform age, but of mixed species, may need to be analysed using methods appropriate for mixed stands, in order to allow for different proportions of species mixtures within a single model.

Mixed forests generally involve much more severe mensurational and sampling problems than uniform forests. The value per unit area of a mixed forest will normally be lower than that for a uniform forest at maturity and, hence, once cannot easily justify an equivalent sampling intensity. At the same time, the dispersion of valuable species and the variability of the forest, implies that the sampling intensity required to obtain the same predictive power as for a uniform stand model must be much greater.

The mensurational problems imposed by poor access, dense understorey, buttresses and the near impossibility of tree total height measurement in many cases, means that measurement costs are much higher for a given level of precision. This tends to further reduce the amount of sampling that can be undertaken within a fixed cost budget.

Consequently, most attempts at yield prediction in mixed forest have been based upon relatively small quantities of data; the resultant models not unnaturally have much lower precision than would be considered adequate in uniform forest management. There are some exceptions to this, as for example the yield prediction work in the Philippines reported by Revilla ^{1/}, which cites results from Dipterocarp forest involving over 240 permanent sample plots.

Because of the variety of species composition, floristic structure, ecological situations and silvicultural practices that are possible in mixed forest, there are an even greater diversity of modelling strategies available than for uniform stands. Many of the published methods are essentially untested except on the basis of very small numbers of plots. The techniques selected for discussion in this section are perhaps the most robust methods and fall into three groups:

- Yield functions applicable to simpler types of mixed forest.
- Transition matrix models.
- Distance-independent tree models based on difference equations.

^{1/} Revilla, A.R. 1979 Yield prediction in cut-over Dipterocarp stands in the Philippines. Paper to Seminar on Management of Dipterocarp Forests, Metro Manila. 20 pp.

It is probable that in the long term, a fourth group of models, the tree position models, will be the most effective basis for yield prediction in mixed stands. However, such models require more powerful computers than are normally available today and considerable investment into research into appropriate sampling techniques before they become a viable proposition.

In this chapter, we do not attempt to give any specific instructions concerning the construction of growth models and yield forecasting for mixed forests but rather some examples of possible ways of dealing with these problems. Extensive work is however going on in the field of modelling for mixed forests and it is hoped that viable methods will shortly be developed.

SITE CLASSIFICATION

Because mixed forests normally comprise many species, the concept of site index must be related either to a species within an individual tree model or to a species association in a yield model. There are many specific techniques for determining site index that have been proposed. They are all equivalent however in their origins. Site indicators are either:

- # Mensurational parameters in various combinations. (E.g. dominant height and age in the conventional uniform stand site index curve).
- Environmental factors, such as altitude, soil type, rainfall, etc.; or indirect environmental indicators such as indicator species or geographic groupings.

The first class of site indicator represents in fact the observed residual from a particular model, in which site is not included. Thus if one devises a model, such as for example, a transition matrix model, for a mixed stand, and applies it to a number of plots, one will find a different residual error for each plot. That error can be coded on a scale such as 1, 2, 3, 4, 5 etc., with a median value representing a zero error and one then has a mensurational site class system applicable to those plots.

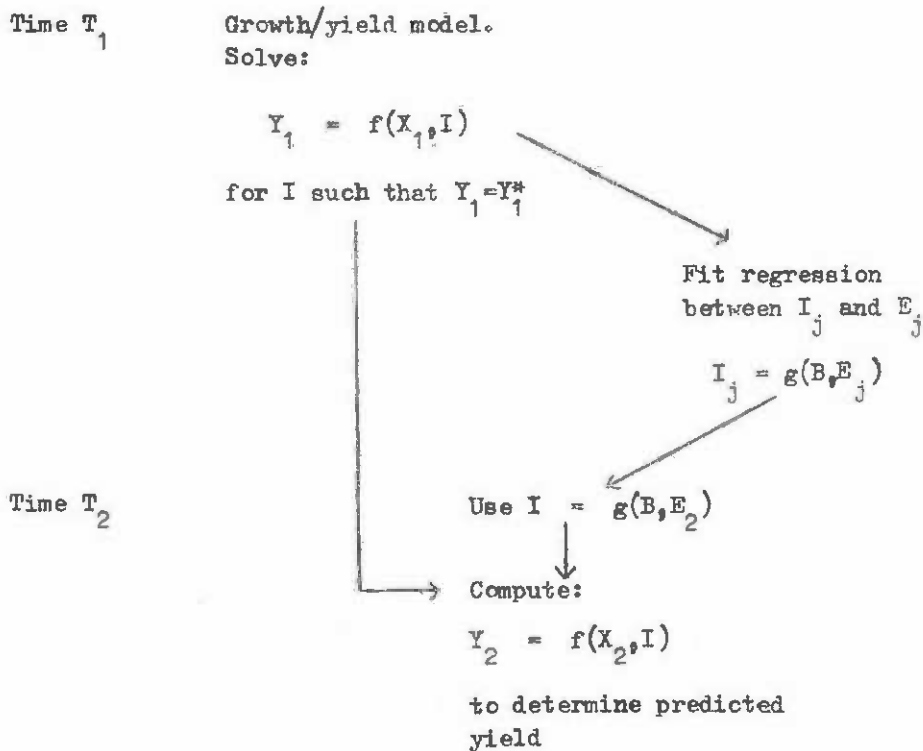
To use such a site class system in practice, the errors (i.e. site indices) must be determined at one point in time and then used as a compensating input to the model for a prediction at a second point in time. This is precisely what is done with a conventional set of site index curves for a plantation. From a known height-age input, site class is determined, which is then fed into the model at a second point in time in order to obtain a much more accurate prediction of subsequent height than would be possible from the mean height-age curve.

Once a set of site class values has been assigned by determining residuals from a site-independent model, then the residuals can be correlated with environmental factors, to determine which are most effective in explaining the residual variation of the model. Indicator species are included by using a zero-one variable for each species: zero for absence from the plot and one for presence on the plot. Other qualitative factors, such as soil type, can be coded in the same way. Appendix A discusses the use of zero-one variables in multiple regression.

The use of mensurational variables as site indicators is less easy with a mixed stand than with uniform stands and can only effectively be done with a static yield model. However, environmental factors can be used as site indicators with any type of model; with mixed forests, environmental site indicators will thus be more important than with uniform stands. In mixed forests, environmental parameters are a necessary part of an effective model, whereas with uniform stands using height as a basis for site determination, environmental parameters are an optional extra.

It should be appreciated from the above discussion that the conventional terms site index, site class, etc. are to some degree misnomers. What one is really dealing with are self-calibrating models, using an error estimate at one point in time to reduce the prediction error for a later estimate. There is nothing inherently one-dimensional about the effect of the environment on a forest growth model. In addition, some part of the error that self-calibration can effectively compensate for may be unrelated to the environment and may be due to historical factors (past stand treatment), stand density, if this is not adequately incorporated in the model, and genetic effects (provenance or species variation).

These points can be summarised symbolically as follows:



Symbols

- B Coefficients in the regression between site index and environmental parameters.
- E_j Set of environmental parameters on a series of plots, used to fit regression.
- E_2 Environmental parameters on plot at time T_2
- $f()$ Any mathematical model used as the growth model.

- $g()$ Any mathematical function (linear or nonlinear) used as the site index-environment regression.
- I Estimated site index value for a plot.
- I_j Set of site index values on the same plots as E_j were determined.
- X_1) Independent variables used in the growth or yield model at times 1 and 2.
- X_2)
- Y_1) Yield estimated from the model at times 1 and 2.
- Y_2)
- Y_1^* Actual yield on plot on which X_1 were determined.

6.2 STATIC YIELD FUNCTIONS FOR MIXED FOREST

6.2.1 General Principles

Static yield functions have been discussed in section 5 with reference to uniform forest. The principal feature of a static function is that time is included in the model as a total elapsed time from some reference point. In a uniform forest, the reference point is usually the date of planting or establishment and time is the age of the forest. In a mixed forest model, the time base may be the last harvesting operation or the last forest inventory or it may be some other point.

The model must take the general form:

$$Y = f(x_1, x_2, \dots, x_n, t)$$

where Y is the measure of yield of interest. This may be timber volume of merchantable species, basal area, fuelwood tonnage/ha, or non-timber yield such as fruit, seed or bark production.

The variables x_i are any variables which fulfil two criteria:

- (1) They are useful in predicting Y . That is, they add significantly in a statistical sense to the goodness of fit of the function $f()$ to the yield.
- (2) They can be determined without requiring remeasurement of the forest at time t . They may be inventory statistics at the base time t_0 , or they may be regional site indices derived from soil, topographic or climatic data. They may be qualitative variables denoting the particular type of harvesting treatment the forest has received.

The time elapsed from the time base is t . This must be present if the function is to be useful as a planning tool.

The form of the function will vary tremendously depending upon the yield being predicted, the data involved and, to some extent, the sophistication of the analytical tools available.

6.2.2 Data Analysis Procedures

The data may be from temporary or permanent plots or from experiments. The simplest type of data analysis that one is likely to undertake will be multiple linear regression, coupled with graphical analysis of data and of residuals. An electronic computer is therefore essential, although it could be quite a small one.

The data from the plots must be summarized prior to determining the yield equation. The summary will derive the following types of information for each plot, on each occasion of measurement.

- (1) Plot identity; for experiments this will include block number, treatment number and replicate number.
- (2) Yields; there may be only one type of yield to consider (e.g. total merchantable volume) but more usually yield will be expressed by species groups and in size/quality classes, so there may be eight or more types of yield.
- (3) Possible predictor variables. These include:
 - Basal area/ha by species groups at time t_0 .
 - Stocking % by species groups at time t_0 .
 - Logging intensity at time t_0 .
 - Silvicultural class.
 - Forest type code.
 - Environmental indicators (soil type, altitude, rainfall, etc.).

Once the plot summaries have been derived, they are maintained in computer accessible form for the various subsequent analyses.

For analysis purposes, various procedures such as principal components analysis, stepwise regression, combinatorial regression can be used to give the automatic selection of the predictor variables which 'best' predict the yields. However, a priori selection of predictor variables in a relatively simple equation is generally preferable, combined with a careful graphical analysis of residuals and preferably some commonsense relationship between the form of the function and the reality of the biological situation predicted.

6.2.3 Methods of Selecting a Yield Equation

Section 5 has considered the various forms of equation that can be used to model asymptotic growth processes. In Appendix A, figure A.2.1. shows a variety of basic functions which can represent other shapes besides the asymptotic. In developing a curve or function to model yield development in natural forest, it is best first to graph the data, plotting yield against time from the time base (whether a felling or inventory). The data should be classified by basal area of the stands at the base time; separate graphs may be drawn for different sites or forests of different species composition.

On the basis of such graphs, a suitable equation can be selected from the various forms described elsewhere in this manual. The data may be fitted directly, using multiple regression or nested regression; or a hand-drawn line may be sketched through the data and subsequently approximated by an equation if required.

In many instances where only light selection felling is carried out and the forest is a complex mixture of many species and ecological types, there may be no direct relationship discernible over time between yield, basal area and site class. In such cases, prediction of growth with a simple yield model is not an effective strategy.

6.2.4 Problems with Static Yield Models in Mixed Forest

Because static yield models are relatively straightforward to construct and use, it might seem that they offer something of a panacea for the problems of yield prediction in mixed forests. Unfortunately, this is far from the case. The problems are essentially of the same kind as those discussed in section 5.2.5 regarding yield models for uniform forests, but are rather more severe in degree.

There are two basic kinds of problem:

- (1) A particular model of yield has implicit in it a historical sequence of events corresponding to those occurring in the data set used to construct the model. The model cannot be applied reliably to data which has experienced a different history.
- (2) There is a problem of compatibility. Suppose for example one predicts three types of volume; V_m for merchantable species; V_p for partially merchantable species; and V_u for unused species. The total of these three should logically represent the total volume of the forest. However, if one compares $(V_m+V_p+V_u)$ with total volume, either from actual data or from a fourth function fitted directly to total volume, one will find a considerable bias.

Neither of these problems is insurmountable in principle. With adequate quantities of data from permanent plots and long term experiments, historical factors representing different sequences of treatment can be incorporated as additional qualitative variables. Compatibility problems may be overcome by fitting constrained regressions which are forced to satisfy particular requirements.

Static yield models above all suffer from the basic limitation that they cannot utilize data from varied historical sources (e.g. different types of experiments, permanent sample plots and short term tree increment plots); nor can they be adapted to predict yield for historical regimes (i.e. sequences of treatment) other than those implicit in the data used to construct them.

6.2.5 Conclusions regarding Static Yield Models in Mixed Forest

- Static yield models are relatively simple to construct, given adequate data.
- In many types of mixed forest, clear relationships between time, treatment intensity and site are not observable. In such cases, static models cannot be used.

- Care must be taken with a static model to recognize that it is tied to predefined historical treatment sequences implicit in the data used to construct it.
- When predicting several types of yield, either constrained regressions should be fitted or it must be accepted that the different predictions cannot be regarded as additive to an unbiased total.

6.3 TRANSITION MATRIX MODELS

Transition matrix models provide a method for making short term predictions of forest growing stock on the basis of large quantities of poor quality data from recurrent measurements. They can be developed from, for example, continuous forest inventory data based on permanent plots or temporary plots near the same locations, in which trees have been measured only by diameter and species class. The method can be used with data that contains significant numbers of erroneous measurements. Short term predictions, in this context, implies periods up to forty or fifty years or one cutting cycle.

The method is objective, but not necessarily very accurate. Its principal advantage is that the data base can be analysed automatically and may be strictly conventional inventory data. Its principal disadvantage is that it is not usually possible to construct adequately refined transition matrices unless large quantities of data are available.

6.3.1 Definition of a Transition Matrix Model

A transition matrix model requires that a system be represented by a row of variables which may be called a state vector. One can imagine this as, for example:

$$x_m = (x_{m1}, x_{m2}, \dots, x_{mn})$$

where each x_{mi} represents the number of trees in the i th diameter class. The first subscript (m) denotes the time period. The x 's do not need to represent size classes. They could, for example, be total biomass in different trophic levels of an ecosystem; or development stages in an animal population. But in growth models for mixed forests, the state vector elements will normally be size classes.

The transition matrix itself, denoted by T , consists of an $n \times n$ table of elements, like this:

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \dots & t_{1n} \\ t_{21} & t_{22} & & \dots & t_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ t_{n1} & t_{n2} & \dots & \dots & t_{nn} \end{bmatrix}$$

Each element t_{ij} represents the proportion of element x_{mi} that becomes $x_{m+1,j}$ over the interval m to $m+1$. The total value of $x_{m+1,j}$ will be the sum of all the t_{ij} times the x_{mi} . That is:

$$x_{m+1,j} = \sum_i^n t_{ij} \cdot x_{mi} \tag{1}$$

By repeating equation (1) for each value of j from 1 to n , one calculates an entire new set of values for X_{m+1} from X_m .

The following example may clarify matters.

The table below assumes a state vector (X) of diameter classes, as indicated:

		Diameter class at end of period					Total
		<20	20-40	40-60	60-80	80+	
Diameter at start of period	< 20	0.6	0.39	0.01	0	0	1.0
	20-40	0	0.68	0.30	0.02	0	1.0
	40-60	0	0	0.75	0.21	0.04	1.0
	60-80	0	0	0	0.9	0.1	1.0
	80+	0	0	0	0	1.0	1.0

The transition matrix probabilities are given within the table. Note that the zero values of the lower diagonal are indicative of the fact that trees do not move into smaller size classes. The addition of the transition probabilities to 1.0 along the rows is necessary if all the states of the system are truly defined by the state vector.

Now assume that the diameter class distribution at the start of the period is:

diameter	<20	20-40	40-60	60-80	80+	
frequency	1037	219	88	22	7	trees/ha

Construct a table as shown below, with the class total in the right hand column and the other figures obtained by multiplying the total by the corresponding transition probability. Then:

		Diameter class at end of period					Total
		<20	20-40	40-60	60-80	80+	
Diameter at start of period	< 20	623	404	10			1037
	20-40		149	66	4		219
	40-60			66	19	3	88
	60-80				20	2	22
	80+					7	7
Total		623	553	142	43	12	

add the columns vertically to get the new frequency distribution for the stand.

This example ignores ingrowth, mortality and harvesting. These are considered in the following sections.

6.3.2 Methods of Constructing Transition Matrices

The methods for constructing transition matrices may be differentiated according to whether single tree, permanent plot remeasurements are available or whether one is dealing only with diameter class measurements.

6.3.2.1 Single tree data

		Second measurement										
		I	A	B	C	D	H	Total
First measurement	I											
	A											
	B											
	C											
	D											
	.											
	.											
	M											
H												

The framework of the transition matrix is shown above. The letters A,B,C,D, ... denote successive diameter classes. M denotes mortality and H harvested stems. I denotes ingrowth.

The transition matrix is constructed as follows:

- (1) Define the initial class and final class, over 1 time period, of a tree. Tally that tree in the appropriate row/column intersection and in the total column on the right. The latter is based on the initial class value. Ingrowth trees are defined as those which appear in a class at the second measurement, but were not present in any class at the first measurement. Mortality trees are those which are dead at the second measurement; and harvested trees those which have been removed.
- (2) Once all the trees in the entire data set have been tallied, divide the transition tallies by the total in the right hand column to give a proportion, rounded to two decimal places.
- (3) Check that the transitions along each row add to 1.0. If there is a small discrepancy (± 0.01 or 0.02), it will be a rounding error; adjust some figures at random until they total 1.0. If there is a large discrepancy, there has been an arithmetic error in that row.

If one of the totals columns is zero, that class must be amalgamated with an adjacent size class, as there are no data defining transitions relating to it.

6.3.2.2 Size class data

The diagram below illustrates a transition matrix as it must be constructed if only size class data is available. It must be assumed that:

- All ingrowth occurs into the lowest size class.

- Outgrowth from one class can only occur into the next higher class.
- The number of stems harvested in each class is known from an independent source or deduced from the treatment prescription.

The shaded blocks in the table are not used; they are zeros within the main transition matrix.

	I			H	A	
I						?
			P_o			
			P_r			
H					1	
B						

Diameter classes in the stand at the first measurement are entered in column A. Diameter classes at the second measurement go in row B. Harvested stems are entered in column H. The diagonals P_o and P_r are calculated from the following formulae:

$$P_o = B - A + P_o^* + H \quad \text{---(2)}$$

$$P_r = A - P_o \quad \text{---(3)}$$

In equation (2), A is the row total and B the column total for a particular P_o ; P_o^* is the value of P_o in the next row down. P_o^* is zero by definition for the largest size class. H is the harvested stems for that row. P_o is calculated recursively, starting at the bottom right corner with $P_o^* = 0$, and then working up to the left.

After the values of P_o and P_r have been calculated in stems/ha, they are converted to proportions by dividing by the corresponding row total A.

The statistics P_o and P_r represent real processes as follows:

- (1) In the ingrowth row, I, P_r is the proportion of seedlings which does not enter the lowest diameter class. P_o is the proportion which does and is accounted as ingrowth.
- (2) In the other rows, representing successive size classes, P_r is the proportion of stems in that class which remain there during any one time period.

- (3) P_0 is the proportion of stems growing out of one class and into the next higher class during any time period.

The figure shows the value of A corresponding to the ingrowth row as a λ . Any arbitrarily large figure may be used, as this row represents the total number of seedlings forming the potential ingrowth pool. However, if regeneration surveys are conducted, then this figure may be estimated absolutely or modified in proportion to a percentage stocking.

Note also that the transition H/H is shown as 1.0 because 100% of the stems harvested in period A (the earlier period) will remain harvested at period B.

Mortality can be added to the matrix in a way exactly similar to harvesting. Like harvesting, it must be measured, which means that the number of dead trees in each size class must be assessed at period B. It is also added to equation (2) in the same way as H .

6.3.3 Refinements to transition models

Two basic refinements must be added to a transition matrix model for mixed forest before it becomes a workable tool:

- (a) Species must be grouped and separate matrices constructed for each group. The number of groups should not be too large, otherwise one may have too little data for many transitions.
- (b) Data should be grouped into basal area classes for different forests, to allow a different transition matrix to be used for different stand densities.

6.3.4 Disadvantages of Transition Models

Transition matrix models have several disadvantages.

- It is difficult and tedious to represent dynamic interactions. For example, the interaction between stand density and growth rate. It can only be done by having a separate transition matrix for each level of the controlling variable. This may be tolerable with only one significant interaction, but with two or more, the number of matrices required is the product of the number of levels of each interaction.
- The precision of a transition model is limited by the need to work with broadly defined classes; otherwise too many undefined transitions occur or the matrices become too large for easy computation.
- Transition models are inefficient in terms of the number of parameters required to define a growth process. A 9×9 matrix (81 parameters) might only correspond to a single tree parameter growth equation.
- Because of their inefficiency and lack of dynamic interaction, transition matrix models offer little scope for improved understanding of forest growth processes.

6.4 DISTANCE-INDEPENDENT TREE MODELS BASED ON DIFFERENCE EQUATIONS

6.4.1 Definition

A tree model is one in which each tree in a stand is individually represented by a set of variables, describing, for example, tree species, tree diameter, height and crown condition. In a distance-dependent model, the tree position in the stand is also represented, typically as x-y coordinates. In a distance-independent model, tree position is not represented. Distance-independent models are generally much more economical in terms of computer resources than tree position models; but give a less realistic and detailed representation of inter-tree competitive processes. In this manual, only the distance-independent model is considered, as experience has suggested that for gross projections in yield, there is no specific advantage in going for the more complex and computationally demanding tree position model.

6.4.2 Allometric and Dynamic Variables

As has been said, each tree in the stand is represented by a set of variables. In the simplest case, a tree will be represented by only one variable, its diameter. Usually, species will also be identified and perhaps height or crown class. Other tree variables, such as crown diameter, volume or biomass, may be derived by allometric relationships with tree diameter and height.

Dynamic variables are those which are predicted from the state of the tree at a previous time period. Allometric variables are those which represent static relationships between different dimensions of the tree at the same point in time.

6.4.3 Representation of Competition

The individual tree dynamic variables, such as diameter, will be predicted from an equation incorporating tree diameter at the last time period and also some measure of stand density. Stand density may be represented in a number of ways, as an absolute measure such as number of trees over a certain size limit or the basal area of trees in the stand or as a relative measure such as basal area divided by the maximum basal area possible on that site; or it could be measured in some novel way by considering, for example, total leaf biomass on a given area.

However stand density is measured, it will retain the property that it is in some sense the summation of the individual tree variables in the stand. Thus basal area is the sum of the individual tree sectional areas.

Stand density obviously varies from place to place in a stand. If the model represents a large plot, of say 1 or 10 ha, then the overall stand density will not necessarily reflect very accurately important variations such as the occurrence of gaps. This difficulty can be avoided by dividing the simulated plot into quadrats, of say 10 x 10 m, and calculating the stand density as it affects any one quadrat as the average density of that quadrat and its eight neighbouring quadrats. For edge quadrats, the neighbours can include the quadrats on the opposite side of the plot in order to avoid edge effects.

6.4.4 Data Requirements and Approaches to Analysis

Data must be available from permanent or temporary sample or experimental plots in order to define all the dynamic and allometric relationships in a model. Any variables included in the model must have been actually measured on sample trees. If competition is to be defined in terms of quadrat stand densities, as discussed above (section 6.4.3), then the sample plots or experiments must also have been measured so that trees could be placed in quadrats, and a quadrat by quadrat competition index calculated.

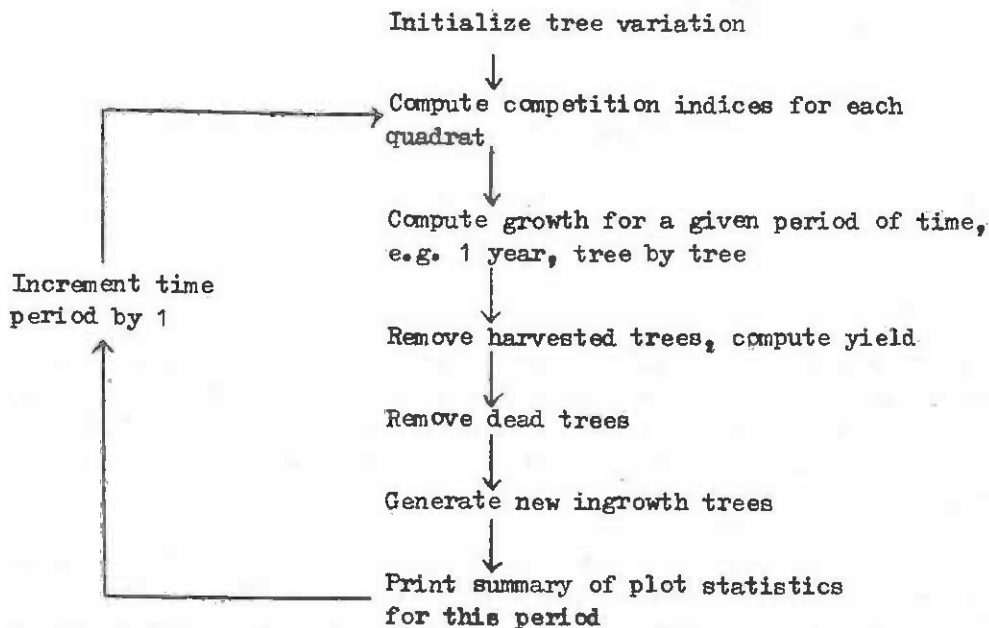
With mixed species forests, it may well be that rare species will not be represented by many points. In these cases it is better to group the rarer species together and provide them with a common set of growth and allometric functions.

Unlike a transition matrix model, the process of fitting the various equations in the model to the field data requires considerable skill. The various problems are discussed elsewhere in this manual, notably in section 5 and Appendix A, but the research worker has to consider:

- the problem of selecting a function which produces the appropriate shape
- fitting the data using a suitable and adequate statistical and/or graphical technique, together with an examination of residuals for bias
- what will happen if in applying the model, an extrapolation will be demanded from a particular function and if the result it gives under such circumstances is reasonable, if not necessarily accurate.

6.4.5 Basic Model Structure

The overall model structure is very similar to that for dynamic stand models, discussed in section 5.3.4, and is illustrated in the diagram below. There is first of all an initialization phase, where the state of the stand at the start of the simulation is given. This could be derived from actual inventory data or it could be generated randomly or from a set of functions.



Then the model enters a cyclic or iterative phase, where each cycle represents one unit of the basic growth period, typically 1 or 5 years. In each iteration, a series of calculations are repeated following the main stages indicated. Generally, the order in which these stages are carried out is not important; so that one could, for example, remove dead or harvested trees before, rather than after, calculating growth on the residual stand.

6.4.6 Ingrowth, Mortality and Harvesting

From the point of view of tree model construction, ingrowth, mortality and harvesting have in common the following:

- they involve the creation or removal of trees from the list of trees in the model
- they are inherently random processes and cannot easily be treated as simple functional relationships
- adequate descriptive data is often lacking on these parameters.

Ingrowth and mortality are often both partially density dependent; so that the number of ingrowth trees or the number of trees dying may both be functions of stand density. Once this number has been determined, the actual selection of individuals may involve a variety of ad hoc rules. With ingrowth, one typically creates trees of random sizes and species with limits determined perhaps by a probability distribution based on data, or perhaps on an arbitrary maximum size that appears reasonable. With mortality, the likelihood of any individual tree dying may be a function of its size and species.

Stochastic or probabilistic processes of this kind can be modelled very easily in a computer simulation by using a uniform random number generator that produces randomly distributed numbers between zero and one. Such functions are often standard library functions in FORTRAN programming and are part of the standard language in BASIC.

Suppose for example that a function indicates that a particular tree in the model has a 0.7 probability of dying in the current year. A random number between 0 and 1 is generated; if it is less than 0.7, then the tree is presumed to have died and a mortality routine is entered which records the details of the tree for the plot summary and removes it from the table of trees alive in the model. If the random number is greater than 0.7, then the tree survives and grows until the next period.

Harvesting processes are similar to mortality in that they involve the removal of trees, but usually in accordance with definite rules relating to the species and size.

6.4.7 Conclusions Regarding Tree Models

Individual tree models are highly flexible, but require considerable skill in both computer programming and in data analysis on the part of the research worker. They also require the use of a large computer. There is no single stereotyped method of constructing such models. The interested reader should study references such as those given in the CAB (1977) Bibliography listed in Appendix B.

Individual tree models are not a panacea. Like all other forecasting methods they require an extensive data base of remeasurements on carefully maintained and assessed permanent sample and experimental plots. Also, like other methods discussed earlier, their sophistication is no guarantee of their accuracy.

7. VALIDATION OF GROWTH AND YIELD MODELS

7.1 THE ROLE OF VALIDATION

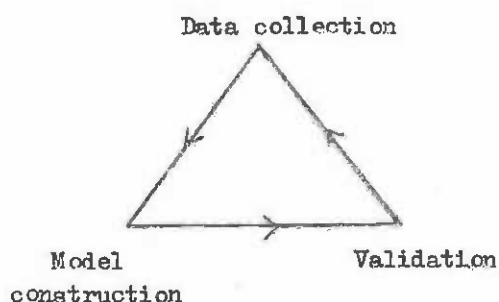
Validation is, literally, the process of determining whether or not a model truly represents reality. However, the following points must be borne in mind:

- Model predictions approach reality as a set of successive approximations as more and more effort is put into data collection and model construction.
- At some point, the effort involved in obtaining more data or producing a more statistically sound model becomes more expensive than the marginal improvement in model predictions warrant.
- Forest models do not have a uniform degree of error over their whole range of predictive behaviour, but will be more or less accurate over different ranges of predictor variables.

The process of validation may be as time consuming and expensive as that of model construction. Special experiments may need to be constructed to test the model; data processing systems may need to be set up to allow models to be validated by comparison with large numbers of plots; considerable statistical analysis may be required to estimate the covariances of residual errors.

This input of effort into the validation process is in no sense wasted. Science differs from philosophy and religion in being subject always to the criterion of empirical validation of theories and models. A model which is not validated is simply speculation and guesswork.

Furthermore, the nature of model construction implies that it must interact with validation as a cyclic process.



Validation will show weaknesses in model behaviour which will lead to improved model structure or to the necessity of collecting more field data. Experiments on models, even those which are quite invalid may lead to alternative concepts of experimental design and data capture.

7.2 VALIDATION DATA

In order to validate a model, its behaviour must be compared with observations from real situations whose history and treatment are precisely known. This data can be called validation data and may be obtained from experiments, permanent sample plots or temporary

sample plots. Two distinct situations usually exist with respect to validation data:

- The data used to validate the model is the same as that used to construct the various functions forming part of the model. We can call this self validation.
- The data used to validate the model has not been used to estimate any of the function parameters in the model. This situation may be called independent validation.

There is nothing extraordinary about the concept of self validation. It is the normal procedure for regression analysis, for example, where all statistical estimators are based on the errors between model predictions and the observed values used to fit the model.

Self validation can however, be dangerously misleading. Regression analysis provides a good example: if the assumptions of the statistical model used (uniform error variance, normally distributed uncorrelated errors, predictor variables known without error) are incorrect then the statistical estimators and parameter values obtained will also be incorrect.

Self validation is particularly dangerous with small numbers of observations and with very complex models. With simple models (perhaps only one or two functions) and with large amounts of comparison data it can be quite acceptable.

Independent validation is a much more satisfactory procedure from every point of view. It gives an absolute reflection of how effective a model is as a predictive tool. The main difficulty with independent validation is that a considerable body of data may need to be ignored when constructing the model. Probably the best procedure, and one widely adopted by the systems modelling community, is that of halving the data set by a random or systematic process and using half the data for model construction and half for validation.

Unfortunately, there is a certain grey area between self validation and independent validation. It is possible to construct a model using one type of information from a set of plot or experimental data and validate it using another, independent statistic from the same plots. Such partially independent validation may be regarded as less satisfactory than fully independent validation, but more indicative of a model's true validity than self validation.

7.3 RESIDUAL ERRORS

Validation of models is usually based upon an analysis of residual errors. These are defined as:

$$\text{Residual error} = \text{Observed value} - \text{predicted value}$$

This is exactly analogous to the residual error used in regression analysis. It is suggested that the reader refer to the notes in Appendix A.2.6; identical techniques can be applied to error analysis from models.

If the residual error is to give a true indication of a model's performance, it must be assumed that the model can be set up in such a way that the independent variables giving rise to the observed value are identical to those for the predicted value. If this is not the case, then the error of the model will tend to be exaggerated.

For example, a dynamic stand model requires information about the planting density and timing and intensity of thinnings if it is to make an accurate prediction. Such a model requires long term permanent sample plots or thinning experiment data for accurate validation.

A growth and yield model will generally predict several statistics, for example mean diameter, dominant height, total volume and merchantable volume. Residual errors generally increase (relative to the statistic predicted) from height to diameter, to total volume to merchantable volume. Consequently, it is usually sufficient to carry out validation studies on total volume or merchantable volume; unless there is a special need to know the precision of diameter and height estimates, these need not be the subject of separate validation studies.

Residual errors can be summarized in several ways:

- Graphically, as plots of residual errors against predicted values or predictor variables. The comments in Appendix A.2.6 apply in this case.
- As a coefficient of determination, analogous to R^2 in regression studies. This is calculated as:

$$r^2 = \frac{\text{Sum of squares of residual errors}}{\text{Sum of squares of observed values}}$$

- As a residual standard deviation, calculated as:

$$\frac{\text{Sum of squares of residual errors}}{\text{No. of validation samples} - \text{No. of predictor variables}}$$

- As a percentage residual standard deviation (equivalent to a coefficient of variation) calculated as:

$$\frac{\text{Residual standard deviation}}{\text{Mean predicted value}}$$

In quoting or using these statistics, the reader should appreciate that they are approximations which cannot necessarily be related to specific confidence intervals or levels of significance.

7.4 GRAPHICAL COMPARISONS

Analysis of residual errors is a somewhat abstract technique. An alternative approach to model validation is that in which a graph is plotted of the statistic of interest against some predictor variable for both real stands and the yield model. For example, one could plot the development of volume against time for actual stands and for a model.

This type of approach gives a more concrete appreciation of the strengths and weaknesses of a model, than residual analysis. However, it cannot effectively be used to summarize the behaviour of a model involving many predictor variables: nor can it easily be used to represent results with large amounts of validation data.

It is suggested that both types of validation are usually necessary. Residual analysis can be used with large sets of data and to summarize results into a small number of statistics, over the whole range of predictor variable values. Graphical comparisons are effective for presenting the key aspects of model behaviour for publication or communication.

7.5 DEFINING THE LIMITS OF MODEL UTILITY

The residual errors from a model will generally tend to increase as one moves towards more extreme values of predictor variables. This is likely to be especially true if one's validation data covers a broader range of sites, ages and growing conditions than the data used to construct the model.

However, there will be limits to the validation data, as there will have been to the data used in model construction. It is important, as part of the validation process, that one examines how the model behaves outside these limits and estimates, for each type of predictor variables, a range of values outside which the model predictions become obviously inaccurate and unusable.

This is important because many applications of growth and yield models may be in situations where absurd values are not immediately apparent. For example, as part of an economic forest sector model or as a component in a programme for calculating cutting plans.

The limiting range of values within which a model may be reasonably precise (to quoted specifications) should therefore be explicitly defined during the validation process.

8. THE APPLICATION OF THE MODEL TO THE REQUIRED END USE

8.1 INTRODUCTION

Essentially, the growth and yield model may be applied in one of three ways:

- (i) As a simple table or graph or set of tables or graphs. These can be used by forest planners directly or can be fed in tabular form to a computer for updating a set of inventory data.
- (ii) As a programme for a computer or calculator which can produce a table or graph of growth and yield for a particular set of treatments. This is appropriate when the model has sufficient inherent flexibility so that it is not possible to define all possible predictions in one set of tables.
- (iii) As a computer programme which forms a sub-model within a larger computer programme for forest planning and which will usually incorporate a data base of inventory information and various economic or technical constraints on harvesting and treatment operations.

Alternatives i and ii have been sufficiently dealt with in 5.3.4. Remains to comment on alternative iii where the model is used in connection with inventory data in forest planning.

A necessary prerequisite in this case is that the variables included in the model as parameters (predictor variables) also are included in the inventory data.

8.2 EVEN-AGED STANDS

In the case when the forest consists of even-aged stands the following three functions might be included in the model:

$$H_0 = f(S_p, S, A)$$

$$I_g = f(S_p, S, A, G) \text{ or } I_g = f(S_p, H_0, N)$$

$$H_f = f(S_p, H_0, N) \text{ or } H_f = f(S_p, H_0, N, G)$$

where S_p denotes species, S site quality class, H_0 dominant height, A age, N number of trees per hectare, G basal area per hectare, I_g basal area increment per hectare and year and H_f form height, defined by the function $V = G \cdot H_f$ where V is volume per hectare. In this case, the variables S_p , A , H_0 , N and G have to be known for the inventory data while site class, S , may be calculated by means of the first function.

We may assume that the model is calculated by means of observations on growth and yield plots. We may further assume that the inventory data, to which the model will be applied, is derived from observations on sample plots. If the plot size is the same in both cases the model may directly be applicable to the inventory data. If, however, there is a difference in size in that the inventory plots for example are smaller than the growth and yield plots, the model may give biased results. There are two reasons for this. One is that the site class will be somewhat underestimated on small plots (0.01-0.02 hectares)

compared to larger plots (0.05-0.1 hectares). This bias is, however, usually only a few percent. The other reason for bias is that the effect of competition from trees outside the plot will increase with decreasing plot size. This in turn means that a growth function containing competition or density variables like N or G is correct only for plots of the size for which it was calculated. The size of this bias depends on the difference in size between the growth plots and the inventory plots, the type of growth function in the model and finally on the variation within the inventoried stands. Nothing can therefore be said in general about the size of this error.

Provided that the growth plots and the inventory plots are of more or less equal size, the application of the model to the inventory data should be done by forecasting the development for every single inventory plot. At the end of the forecasting period, the plots may be sorted into strata according to species, site class and age and the sums and means for each stratum calculated.

If, however, there is a great difference in plot size in that the growth plots are much bigger than the inventory plots it might be advisable firstly to sort the inventory data according to species, site class and age and then to apply the model to these strata instead of to each separate inventory plot.

8.3 MIXED STANDS

In the case of a forest consisting of mixed (uneven aged) stands the following functions might be included in the forecasting model:

$$I_v = f(S, F, G(1), G(2), \dots G(n), N(1), N(2), \dots N(n), L(1), L(2), \dots L(n))$$

$$V = f(S, F, G(1), G(2), \dots G(n), N(1), N(2), \dots N(n))$$

where I_v is growth in merchantable volume per hectare and year, S is site class defined by environmental indicators (soil type, altitude, rainfall, etc.), F is forest type, $G(1), G(2), \dots G(n)$ is basal area after logging by species group per hectare, $N(1), N(2) \dots N(n)$ is the number of trees after logging per species group and hectare, $L(1), L(2) \dots L(n)$ is logging intensity per species group and finally V is merchantable volume after logging per hectare.

To enable forecasting the same variables (predictor variables) have to be recorded in the inventory as those included in the model. The discussion on plot size in relation to forecasting foreven-aged stands applies in principle also to mixed stands. It must be admitted, however, that very little so far has been done in the field of growth models for tropical mixed forests and that this field still is in a stage of research. It is thus not possible to give any precise instructions on how to make forecasts concerning the development of tropical, mixed forests after logging. For temperate, mixed forests, models have been developed by which reliable forecasts may be done (see e.g. Monserud 1980). It might be worthwhile trying these models also for tropical mixed forests, but this has so far not been done.

Monserud, A, Ek, A, 1980, Comparison of two stand growth models for northern hardwoods - Wright H, (editor) 1980. Planning, performance and evaluation of growth and yield studies. Meeting of IUFRO 54.01. Commonwealth Forestry Institute, Oxford, Great Britain, pp.8.

Appendix A

STATISTICAL AND MATHEMATICAL TECHNIQUES FOR YIELD STUDIES

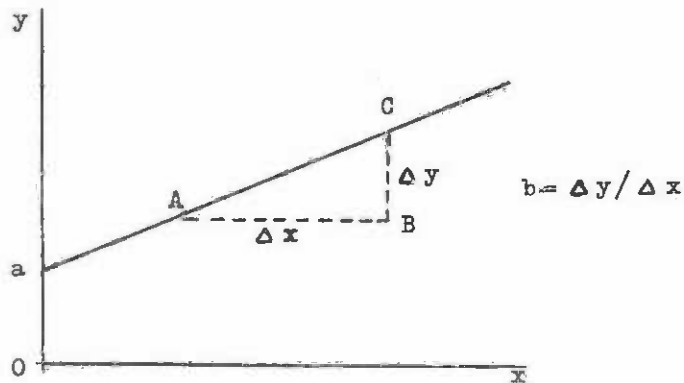
	<u>Page</u>
1. EQUATIONS AND GRAPHS	
1.1 The Straight Line	117
1.2 Transformations for Curve Fitting	118
1.3 Example of Approximation of a Curve Using Transformation to a Straight Line	119
1.4 Polynomials for Curve Approximation	121
1.5 Graphs Involving Three Variables	125
1.6 Example: Fitting an equation to a system of curves by harmonization.	126
2. CURVE FITTING BY LINEAR LEAST SQUARES ANALYSIS	129
2.1 Simple Linear Regression	129
2.2 Regression with Two Predictor Variables	132
2.3 Data Transformations and Curve Fitting	133
2.4 Multiple Regression Analysis	140
2.5 Calculations for Estimating the Parameters of a Multiple Regression Model	141
2.6 Residual Analysis	150
2.7 Weighted Regression	152
2.8 Comparison of Regression Lines	153
2.9 Nested Regression Using Conditional Variables	160
3. SOLUTION OF EQUATIONS	163
3.1 Solution of the Quadratic Equation	163
3.2 Graphical Solution of Equations	163
3.3 Numerical Solution of Equations	165
4. FITTING NON-LINEAR MODELS	166

1. EQUATIONS AND GRAPHS

This section concerns methods of expressing graphical relationships as equations.

1.1 The Straight Line

The simplest type of graph is the straight line:



This can be represented as an equation by:

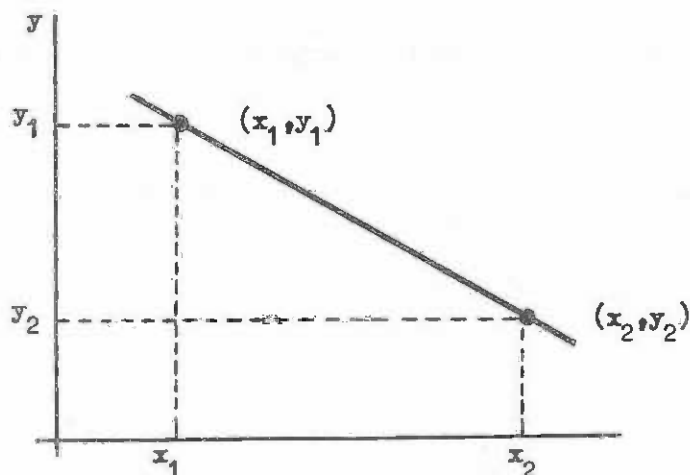
$$y = a + b \cdot x$$

a is the value of y at the point where the line crosses the y axis (i.e. at $x=0$). b is the slope of the line, and can be conveniently worked out from a graph as follows:

- (i) Take any convenient interval Δx (say for example, 10 units of x), and draw a line AB of that length parallel to the x axis (see diagram above).
- (ii) Measure the length Δy of the line parallel to the y axis from B to the graph line. If BC is measured upwards, Δy is positive. If C is below B then Δy is negative.
- (iii) The slope b is given by $\Delta y / \Delta x$. Note that Δy and Δx are measured in the units of the y and x axes, and not in actual distance on the paper in centimetres etc.

An alternative algebraic method of calculating the a and b coefficients for a straight line is as follows:

- (i) Take any two convenient points on the line, designated by (x_1, y_1) and (x_2, y_2) , as shown on the diagram below.



(ii) Calculate the slope b as:

$$b = (y_2 - y_1) / (x_2 - x_1)$$

(iii) Calculate the intercept a as:

$$a = y_1 - b \cdot x_1$$

1.2 Transformations for Curve Fitting

Many relationships between variables in forest yield prediction are in the form of curves when represented graphically. The model, or particular equation that represents the curve, must be known if an exact representation of the curve is to be worked out. In practice, the correct model is not usually known and only an approximation is possible.

Perhaps the most easily fitted and used approximating equations are those which involve transformations of the x or y units, but leave the general equation in linear form. Common transformations are:

$\log_{10} x$	logarithm to the base 10
\log_e or \ln	" " " " e (2.71828....)
$1/x$	reciprocal
\sqrt{x}	square root
x^2	square

The transformations can be applied to both the x and y variables to give various combinations and resultant curves.

Data that is obviously not linear in form can be plotted using various transformations until a linear shape is obtained. The work is reduced if one has some idea in advance of the transformations likely to be successful. In the main part of this manual examples for particular types of data are suggested.

1.3 Example of Approximation of a Curve Using Transformation to a Straight Line

Figure A.1.1(a) shows a curve representing a relationship between tree volume (V) and diameter (D). We assume that the curve has been drawn by hand and it is desired to express it as an equation.

Points at convenient intervals on the x-axis are tabulated from the graph:

Points from Original Function		Transformations		
D	V	D ²	log D	log V
10	0.05	100	1.000	-1.301
20	0.20	400	1.301	-0.699
30	0.45	900	1.477	-0.347
40	0.85	1 600	1.602	-0.071
50	1.50	2 500	1.699	0.176
55	1.90	3 025	1.740	0.279

The first transformation tried is shown in figure A.1.1(b), where D² is plotted against V. The points from the original curve almost follow a straight line, but there is still a slight curvature apparent.

Another transformation is shown in figure A.1.1(c). Here we have taken logarithms to base 10 of both axes. The transformed values are listed for the 6 selected points in the table above. It can be seen that this transformation gives an almost perfect fit, except for the smallest point which is slightly above the line. From the figure we can calculate the slope and intercept for this line. Taking the two points (1.301, -0.699) and (1.740, 0.279) shown by the double circles on the figure and as D at 20 and 55 cm in the table above, we have:

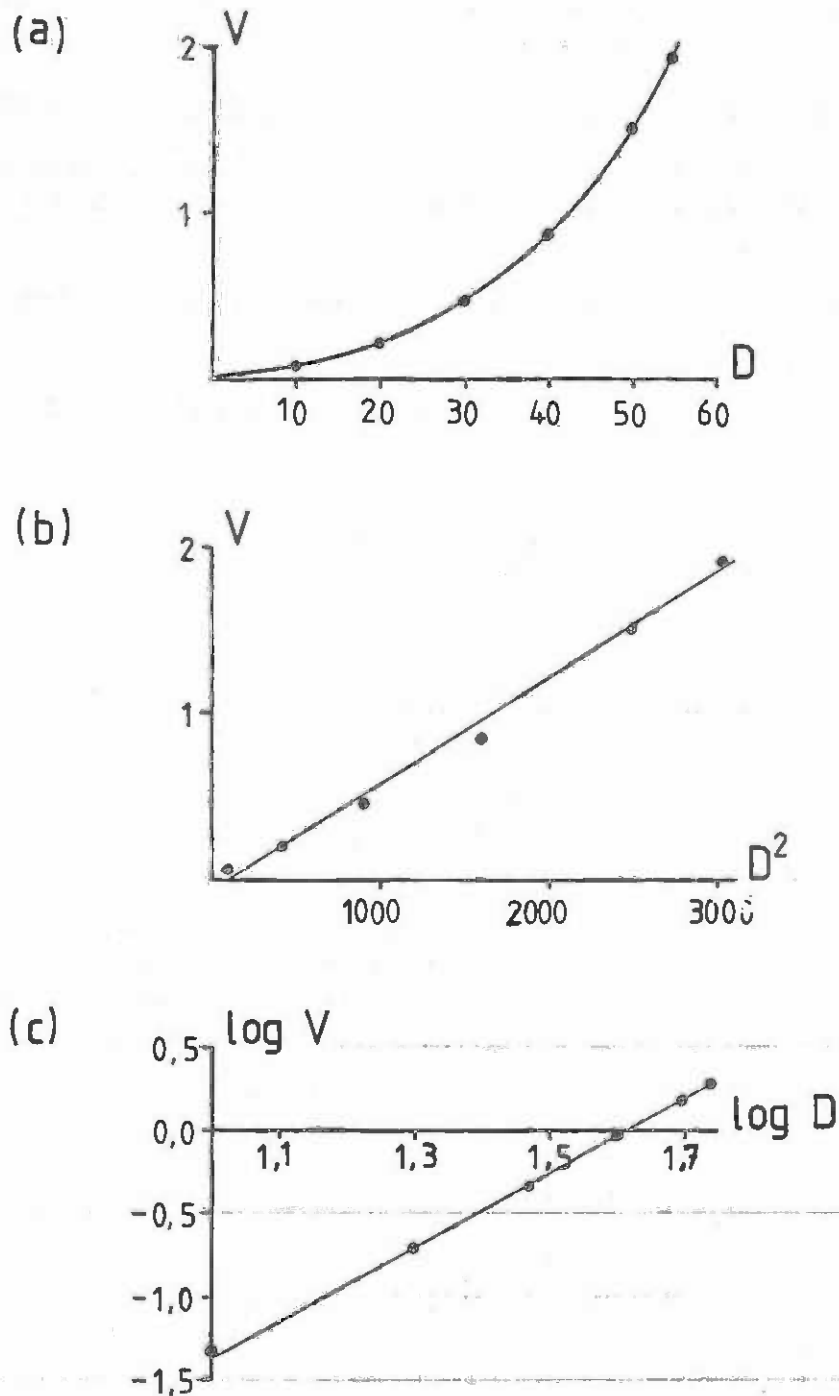
$$\begin{aligned}
 b &= (0.279 - (-0.699)) / (1.740 - 1.301) \\
 &= 0.978 / 0.439 \\
 &= 2.228
 \end{aligned}$$

$$\begin{aligned}
 a &= -0.699 - 2.228 \times 1.301 \\
 &= -3.597
 \end{aligned}$$

So our equation approximating the curve in figure A.1.1(a) is:

$$\log_{10} V = -3.597 + 2.228 \log_{10} D$$

Figure A.1.1



Transformation of a hand-drawn curve to give a straight line approximation.
(a) Original free-hand curve with selected points for calculation of transformed values.
(b) x-axis transformed by taking square of selected points. (c) x and y axes transformed by taking logarithms to base 10 of the selected points.

1.4 Polynomials for Curve Approximation

Equations involving only two coefficients are convenient for the approximation of functions because they can be drawn as straight lines with suitable transformations. However, some shapes of curve cannot be treated in this way. In this case the technique of fitting a quadratic or cubic polynomial to the function may yield good results.

A polynomial is an equation of the form:

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 + \dots + b_nx^n$$

where n is the order and the b_i are coefficients. The higher the order the greater the flexibility of the function. On the other hand, high order polynomials are difficult to fit by manual techniques.

For quadratic (2nd order) and cubic (3rd order) polynomials, manual calculation of the coefficients is possible in about 30 minutes with the help of a 4-function (+, -, x, ÷) calculator. A form is provided to assist the user (form A.1) which shows in this case the fitting of a cubic equation to the line in figure A.1.2. Four points are taken from the function to be fitted and entered in table (1) of the form. The points need not be in any order, but it is desirable that:

- (i) One point each should represent the two extremes of the function. (In this case, X_1 and X_3 .)
- (ii) X_4 should not be an extreme point if the quadratic equation coefficients are to be calculated.
- (iii) The other points should be more or less evenly spaced.
- (iv) When only the quadratic equation is to be calculated, three points are sufficient.

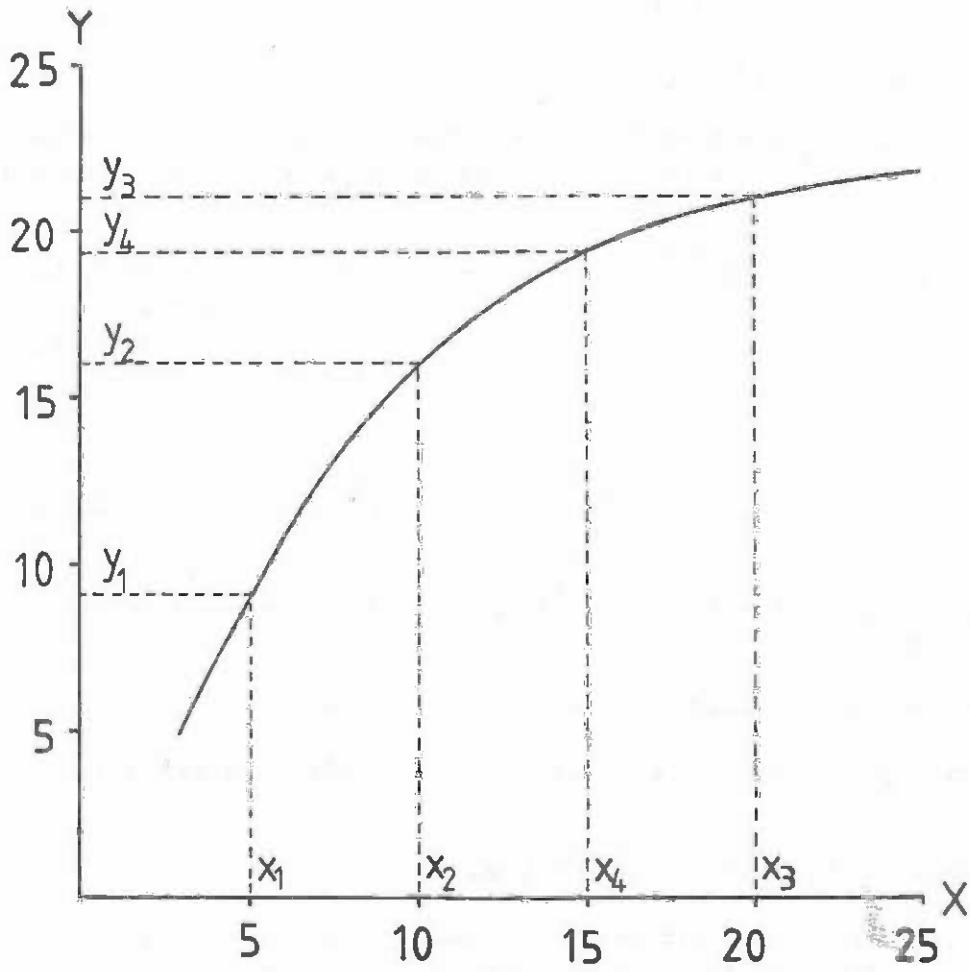
Form A.1 explains the necessary calculations.

Quadratic and cubic curves provide a useful means of approximating hand-drawn curves by an equation. When using them the following details should be noted:

- (i) Always check the calculations and the suitability of these curves by drawing the function on top of the original curve. It is usually preferable to start by fitting the quadratic equation. Then, if this is not sufficiently accurate, compute the cubic equation coefficients.
- (ii) Never use a polynomial equation to extrapolate beyond the two extreme points used in fitting it. If extrapolation is necessary, extend the hand-drawn curve first and then fit a new equation with a new extreme point.

Figure A.1.2

Hand-drawn curve approximation by a cubic polynomial. (see form A.1 for calculations)



Form A.1 Coefficients for an approximating quadratic or cubic curve

Objective: To calculate the coefficients b_i in either of the equations:

$$Y = b_0 + b_1 \cdot X + b_2 \cdot X^2 + b_3 \cdot X^3 \quad (\text{cubic curve})$$

or

$$Y = b_0 + b_1 \cdot X + b_2 \cdot X^2 \quad (\text{quadratic curve})$$

Items marked \diamond are not required for the quadratic curve calculations.

(1) Tabulate 3 or 4 data points for the function and compute squares and cubes.

i	1	2	3	4
Y_i	9.2	16.0	21.0	\diamond 19.4
X_i	5	10	20	\diamond 15
X_i^2	25	100	400	\diamond 225
X_i^3	\diamond 125	\diamond 1000	\diamond 8000	\diamond 3375

(2) From table (1) calculate differences as shown:

i	$Y_{i+1} - Y_i$	$X_{i+1} - X_i$	$X_{i+1}^2 - X_i^2$	$X_{i+1}^3 - X_i^3$
1	6.8	5	75	\diamond 875
2	5	10	300	\diamond 7000
3	\diamond -1.6	\diamond -5	\diamond -175	\diamond -4625

(3) From table (2) construct the terms shown in the table below according to the following definitions:

i	U_i	V_i	W_i
1	1.36	15	\diamond 175
2	0.5	30	\diamond 700
3	\diamond 0.32	\diamond 35	\diamond 925

$$u_i = (Y_{i+1} - Y_i) / (X_{i+1} - X_i)$$

$$v_i = (X_{i+1}^2 - X_i^2) / (X_{i+1} - X_i)$$

$$w_i = (X_{i+1}^3 - X_i^3) / (X_{i+1} - X_i)$$

/...

Form A.1 (continued...)

- (4) From table (3) calculate the terms in the following table from these definitions:

i	$v_{i+1} - v_i$	p_i	q_i
1	15	-0.05733	35
2	5	-0.036	45

$$p_i = (u_{i+1} - u_i) / (v_{i+1} - v_i)$$

$$q_i = (w_{i+1} - w_i) / (v_{i+1} - v_i)$$

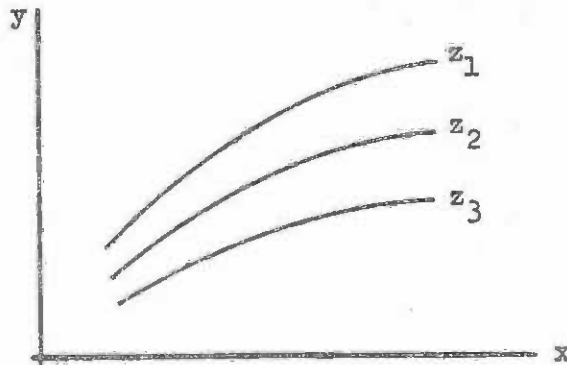
- (5) Calculate the coefficients according to the following formulae:

<u>Coefficient</u>	<u>Cubic eqn.</u>	<u>Quadratic eqn.</u>
b_3	$(p_2 - p_1) / (q_2 - q_1)$	None
	= <u>0.002133</u>	
b_2	$p_1 - b_3 \cdot q_1$	p_1
	= <u>-0.1320</u>	= _____
b_1	$u_1 - b_2 \cdot v_1 - b_3 \cdot w_1$	$u_1 - b_2 \cdot v_1$
	= <u>2.967</u>	= _____
b_0	$Y_1 - b_1 \cdot X_1 - b_2 \cdot X_1^2 - b_3 \cdot X_1^3$	$Y_1 - b_1 \cdot X_1 - b_2 \cdot X_1^2$
	= <u>-2.5995</u>	= _____

* * * * *

1.5. Graphs Involving Three Variables

All the graphs considered in the above paragraphs relate a y variable with an x variable. Many relationships involve a third variable which we may call a z variable. The graph may appear something like this:



Here, each level of z (z_1, z_2, z_3 , etc.) forms a distinct relationship between x and y . Such cases are common with site index curves where height (y) depends on age (x) for different levels of site index (z), or volume or basal area curves depending on height (x) and stand density (z).

Such systems of curves having been drawn by hand through data may need to be expressed as equations for calculation or tabulation purposes, for use in a computer programme, or simply for convenience of communication and analysis.

The system of curves can be expressed either:

- (i) as a series of separate equations with a different relationship between x and y for each level of z ; or
- (ii) as a single equation in which z enters as a distinct variable.

The first method can lead to the second by the process called harmonization. Sets of equations of the first type have two disadvantages:

- (i) There is no method of determining a value of y given a z value intermediate between the levels chosen.
- (ii) A large number of coefficients are required to describe the curve system,

Equations of the second type are therefore to be preferred; they are more concise and can be used to calculate y for any given values of x and z (within the range of the curve system).

The example below shows how a system of harmonized curves can be constructed. The general principle involves graphing the coefficient values against the z variable levels and for each coefficient, deriving a new expression to predict the coefficient value given a particular z .

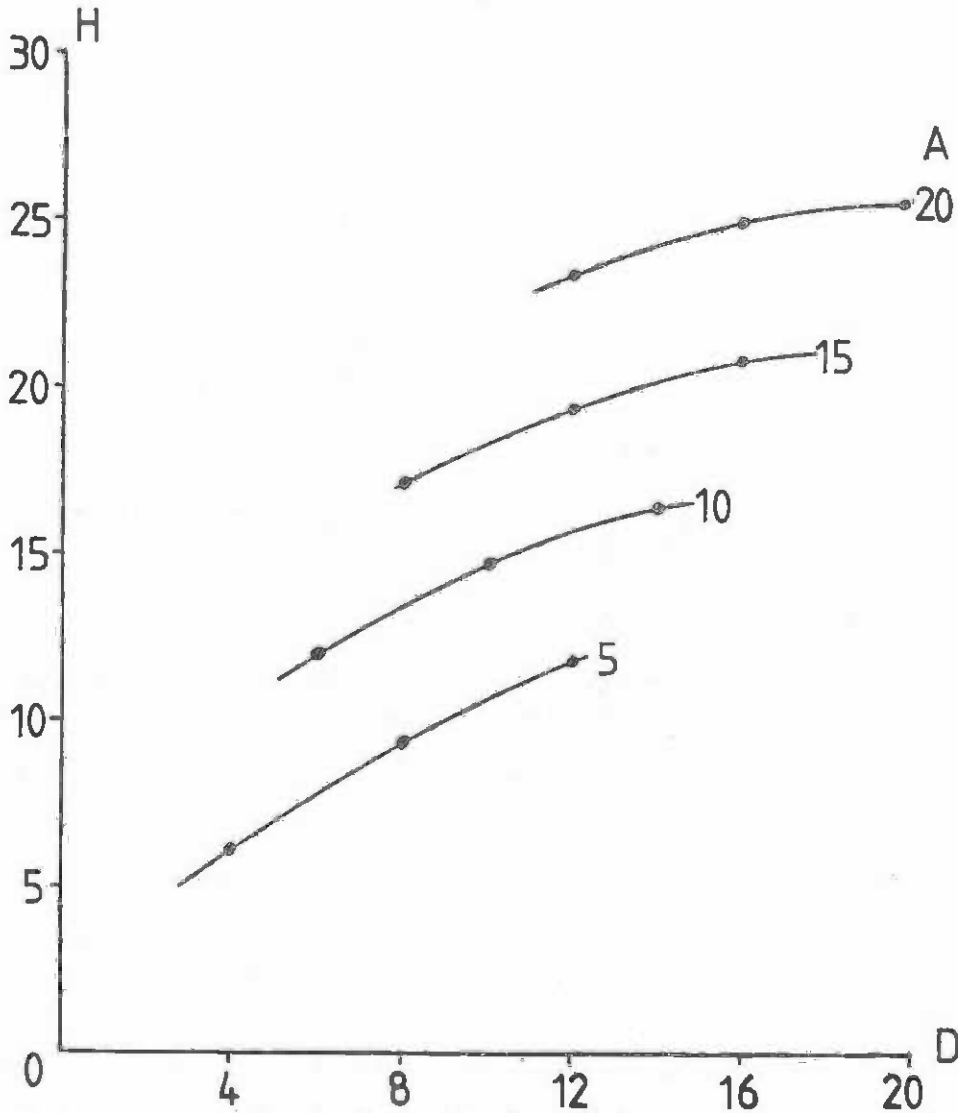
1.6 Example: Fitting an equation to a system of curves by harmonization

Figure A.1.3 shows a system of height (H) over diameter (D) curves for individual trees in four different age classes (A). The curves were originally drawn by hand through data.

The first stage is to fit a separate equation to each line. It is decided to use a quadratic function as the approximating equation.

Figure A.1.3

Hand-drawn curves of height on diameter for different age classes, to be approximated by harmonized quadratic equations (see example in text).



The calculations, which are not shown, involve selecting 3 points from each line and calculating the coefficients according to the method of form A.1. The selected points are shown on each curve in the figure by * . The table below gives the coefficient values for each age class:

Age Class	Coefficients		
	b_0	b_1	b_2
5	2.2	1.0625	-0.0219
10	6.6	1.0500	-0.0250
15	9.8	1.1375	-0.0281
20	12.5	1.2750	-0.0313

In order to reduce the system of four quadratic equations to a single system, the coefficients b_1 must be made to depend on age. The form of this dependence can best be seen by graphing the coefficients against age class (figure A.1.4).

From that figure it can be seen that the relation between b_0 and age follows a gentle curve, whilst b_1 versus age is a steeper, more asymmetric curve; b_2 versus age appears almost exactly linear. The equations were calculated (again using form A.1) for these three curves as follows:

$$b_0 = -0.0187 + 1.153 A - 3.100 A^2 \quad -(1)$$

$$b_1 = 1.225 - 0.0508 A + 0.00400 A^2 - 6.67 \times 10^{-5} A^3 \quad -(2)$$

$$b_2 = -0.0188 - 0.000627 A \quad -(3)$$

The general equation for the lines in figure A.1.4 is the quadratic equation, as selected earlier, with the form:

$$H = b_0 + b_1 D + b_2 D^2 \quad -(4)$$

The coefficient values in equation (4) can now be derived for any age between the limits 5 and 20 using equations (1) to (3).

This is a fairly complex example of harmonization intended to illustrate the full scope of the principles involved. In practice some simplification is often acceptable. For example, in the above case a quite close approximation to the original hand-drawn function can be obtained if average values of b_1 and b_2 are used and only b_0 depends upon age as a linear function:

$$b_0 = a_0 + a_1 A$$

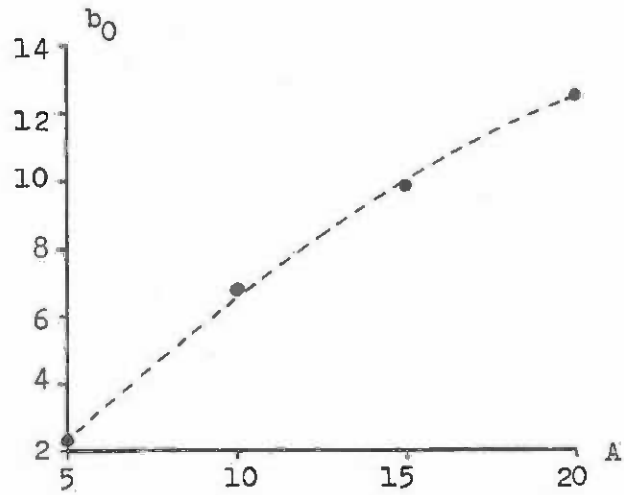
the overall equation, substituting for b_0 will be:

$$H = a_0 + a_1 A + b_1 D + b_2 D^2$$

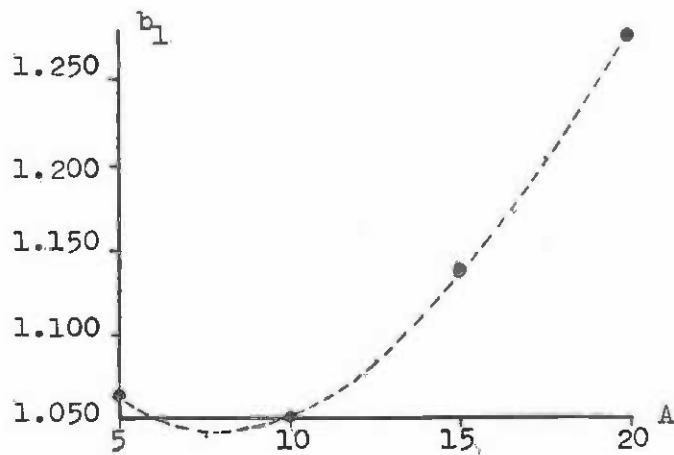
Figure A.1.4

Coefficient values plotted against age class A for the quadratic equations fitted to the lines of figure A.1.3

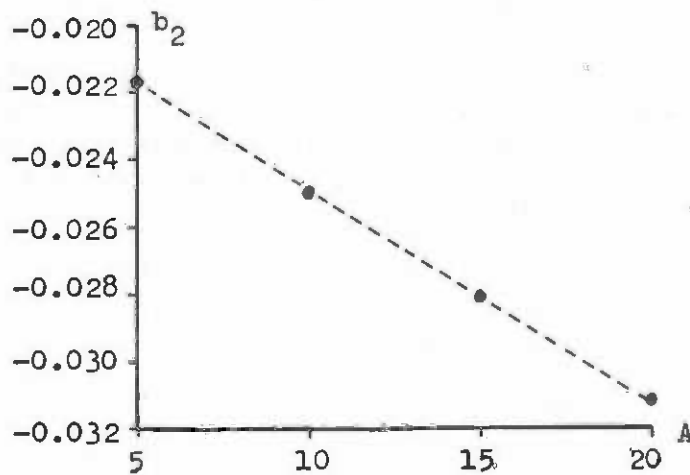
(a)



(b)



(c)



This model is not a perfect fit (as is the one worked out above) but the error over the range of the hand-drawn function is within $\pm \frac{1}{2}$ m of predicted height which may well be acceptable if the original data was well scattered and consequently the hand-drawn lines subject to some uncertainty.

As with polynomials, harmonized functions should never be used for extrapolation! If extrapolation is necessary, draw the extended curve set by hand to cover the range of intended use and recalculate all the coefficients of the approximating function.

2. CURVE FITTING BY LINEAR LEAST SQUARES ANALYSIS

2.1 Simple Linear Regression

Linear regression analysis is dealt with in many text books in great detail and with numerous examples (c.f. Snedecor, Draper & Smith and Seber in the bibliography). It is proposed here only to give the barest outline of the fundamental ideas, basic calculations and most essential statistical parameters.

Linear least squares analysis is a slightly more precise name for linear regression analysis. It refers to a technique for fitting predictive equations to raw data (i.e. unsmoothed observations) based on the principle of minimizing the squares of the deviations from a straight line through the points. When the assumptions, which are listed below, are fully satisfied the parameters (coefficients) for the fitted equation will be those which are most likely to be correct in a statistical sense. When the assumptions are not satisfied, then the method will still give parameter estimates, but they will no longer be the best ones obtainable; better parameter estimates could be obtained by deriving from first principles correct "maximum likelihood equations" for the particular class of problem involved. This latter subject is outside the scope of this appendix.

The definition of the simplest type of regression problem can be shown with the help of figure A.2.1. It involves a relationship between two variables, x and y. The x variable is assumed to be the one that is to be used to predict the y variable. The statistical model is:

$$y = a + b \cdot x + e_i \quad \text{---(1)}$$

where the e_i are the random deviations of each point from a line. In figure A.2.1 the solid line represents the part

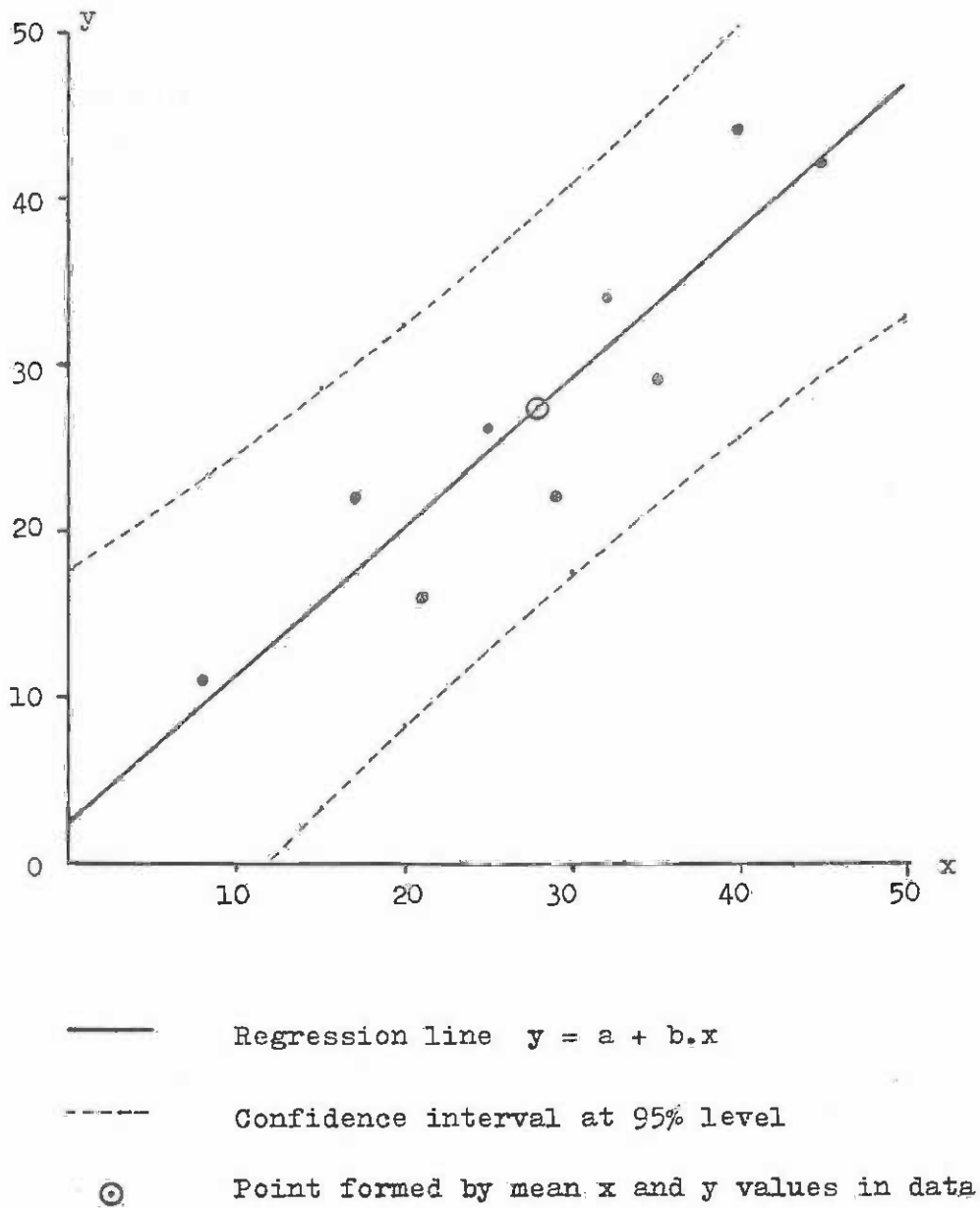
$$y = a + b \cdot x$$

usually called the regression equation. a and b are the coefficients of this equation whose estimation from the data is the primary purpose of the analysis. The e_i are called 'residuals' and represent the vertical distances between the points and the line.

Form A.2 Calculation of statistics for simple linear regression using Statform 11 from Dawkins, 1968 (reproduced with permission). The data is that shown in figure A.2.1.

LINEAR REGRESSION and CORRELATION							
Line	1	Nature of the observations: Hypothetical height-diameter observations on individual trees.					
	2	Dependent or left-hand variable, y: Height, m.					
	3						
	4	Independent or right-hand variable, x: Diameter, cm.					
	5						
	6	Sums of squares and products. n = 9 pairs of observations					
	7	$(\Sigma y)^2 = 60516$	$\Sigma y \Sigma x = 61992$	$(\Sigma x)^2 = 63504$	\bar{y}	\bar{x}	
	8	$\Sigma y^2 = 7718$	$\Sigma yx = 7839$	$\Sigma x^2 = 8134$	11	8	
	9	$(\Sigma y)^2/n = 6724$	$\Sigma y \Sigma x/n = 6888$	$(\Sigma x)^2/n = 7056$	22	17	
	10	$SSy = 994$	$SPyx = 951$	$SSx = 1078$	16	21	
	11	Linear regression	Coefficient; $SPyx/SSx = 0.88219 = b$		22	29	
	12		Constant; $\bar{y} - b\bar{x} = 2.63178 = a$		29	35	
	13	Regression SS = $b \times SPyx = 838.96$					34
	14	Total SS is the SSy. Total d.f. = $n-1 = 8$					26
	15	Residual SS and d.f. obtained by subtraction.					26
	16	ANOVAR				42	45
	17	Source	SS	d.f.	MS	Variance ratio	Table F
	18	Regression	838.96	1	837.0	37.9	29.25
	19	Residual	155.0	7	22.14	@ $p=0.1$	
	20	Total	994	8			44
	21	Coefficient of determination, r^2 ,					
	22	Regression SS/Total SS = $839/994 = 0.844 = r^2$					
	23	Coefficient of linear correlation, r,					
	24	$r = \sqrt{r^2} = 0.919$ with sign as for b					
	25	Residual standard deviation = $\sqrt{\text{Resid. MS}} = 4.71 = \text{SD. resid.}$					
	26	Coefficient of residual variation % = $100 \times \text{SD resid.}/\bar{y} = 19.1\% = \text{CV}\%$					Σy
	27	Standard error for random samples = $\text{SD resid.}/\sqrt{n} = 1.57 = \text{SE. resid.}$					Σx
	28	Student's t for $n-2$ d.f. and P.05 is 2.36					\bar{y}
	29	Sampling error % = $100 \times \text{SE. resid.}/\bar{y} = 0.6\% = E\%$					\bar{x}
	30	Standard error of the coefficient b is $\sqrt{\text{Resid. MS}/SSx} = 0.1433 = \text{SE}_b$					246
	31	Its t-ratio is b/SE_b , or $\sqrt{\text{variance ratio}}$, = 6.156 = t-ratio					252
	32	Confidence limits for estimation of y from observations of x; if m observations are taken of x, having a mean of \bar{x}' , there are m possible estimates of y. Confidence limits for their mean y' ($= \Sigma y/m$) are:					273
	33						28
	34	$y' \pm t \times \text{SD. resid.} \times \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(\bar{x}' - \bar{x})^2}{SSx}}$					
	35						

Figure A.2.1 A simple linear regression model fitted to data



The assumptions made in simple linear regression are that:

- (i) The model is truly linear, as represented by equation (1) above, and not curved in any sense.
- (ii) The residuals are normally distributed with a constant variance over the range of x values under consideration.
- (iii) The residuals are independent of each other; they are not correlated or grouped in any way.
- (iv) The values of the sample points on the x axis can be determined exactly.

The calculation of the various statistics required for linear regression analysis are shown in form A.2, which reproduces Statform 11 from Dawkins (see bibliography for details) with the working of the data shown in figure A.2.1. The most important statistics are the coefficients (lines 11 and 12), the variance ratio (line 18), the coefficient of determination r^2 (line 22), the residual standard deviation (line 25) and the confidence limits for predictions (lines 32-35).

The r^2 values and the residual standard deviation are useful for the comparison of regressions. The higher the r^2 the more precise the relation. When r^2 is one, the residuals are zero and the data fall exactly on the line. Similarly, the closer the residual standard deviation approaches to zero, the more precisely the regression predicts the y values. The confidence intervals of the regression line show the limits within which the true mean of a selected number of y values should lie.

It will be noted that the regression line will always pass through the point formed by the mean of the x and y values. The confidence bands are also curved in form and become rapidly wider as one moves outside the range of the data contributing to the regression.

The x variable in a regression analysis is called the independent or predictor variable; it provides the basis on which the predictions will be made. The y variable is called the dependent or response variable; it is assumed to be controlled in some degree by the level of the predictor variables.

2.2 Regression with Two Predictor Variables

Regression with two predictor variables assumes the model

$$Y = b_0 + b_1X + b_2Z$$

where the X and Z are known predictor variables and Y is the dependent variable; the b_i are the coefficients to be determined. The assumptions and general principles are the same as for simple linear regression. The coefficients and statistics for this model can be calculated by hand although the method is rather more involved than for simple linear regression. It is a useful model for fitting a variety of curves, as will become apparent in section 2.3.

Form A.3 shows how the calculations proceed. The example uses the data from figure A.2.1 and form A.2 for the height-diameter relationship of 9 sample trees, but this time with the addition of an extra variable (diameter)², so that the equation being fitted becomes:

$$H = b_0 + b_1D + b_2D^2$$

After working through the form the coefficient values calculated are:

$$b_0 = 8.31177$$

$$b_1 = 0.371583$$

$$b_2 = 0.00953469$$

all to 6 significant digits. It is important when carrying out the calculations to work to 6 or more significant digits and not to round small numbers (such as g_1 , g_2 , g_3 in the example) to a few decimal places; otherwise considerable loss of accuracy may result.

The lower variance ratio and t-values for the coefficients obtained with this regression as compared with the earlier linear one (figure A.2.1, form A.2) reflects the fact that the addition of the extra variable increases the uncertainty of the parameter estimates with respect to the population from which this sample of points was drawn. In this case neither of the t-values is significant for the two variable coefficients b_1 and b_2 ; this indicates that the model would be just as efficient for prediction purposes if one or other variable were omitted.

2.3 Data Transformations and Curve Fitting

Various types of curvilinear functions may be fitted using linear regression techniques by making appropriate transformations of the dependent and predictor variables. It is desirable when making transformations of data to proceed through the following stages:

- (i) Plot the untransformed data on normal graph paper and examine it to see whether any curvature in the trend line is apparent.
- (ii) If a curvature is apparent, select an appropriate transformation for the dependent and/or predictor variables and calculate transformed values for each data point.
- (iii) Plot the transformed values on normal graph paper and examine the points to see whether the trend is now linear. If it is, fit a regression using the transformed data values.

The selection of a suitable transformation can be aided by the diagrams in figure A.2.1, which illustrate some commonly used curve shapes. Functions (a) to (d) can be fitted by simple regression. Functions (e) and (f) require two predictor variables; consequently the transformed values cannot be examined graphically as in (iii) above.

Form A.3 Coefficients and statistics of a regression with two predictor variables (part 1)

Description of problem: Height-Diameter data from example in section A.2.1, but with $Z = X^2$

(1) Regression data		
Y	X	Z
11	8	64
22	17	289
16	21	441
22	21	841
29	35	1225
34	32	1024
26	25	625
42	45	2025
44	40	1600

(4) Corrected products
$\Sigma y^2 = \Sigma Y^2 - (\Sigma Y)^2/n = 994.000$
$\Sigma xy = \Sigma XY - \Sigma X \cdot \Sigma Y/n = 751.000$
$\Sigma zy = \Sigma ZY - \Sigma Z \cdot \Sigma Y/n = 52331.667$
$\Sigma x^2 = \Sigma X^2 - (\Sigma X)^2/n = 1078.000$
$\Sigma xz = \Sigma XZ - \Sigma X \cdot \Sigma Z/n = 57716.000$
$\Sigma z^2 = \Sigma Z^2 - (\Sigma Z)^2/n = 3238501.556$

(5) Determinant & Gauss multipliers
$D = \Sigma x^2 \cdot \Sigma z^2 - (\Sigma xz)^2 = 159968,020.9$
$g_1 = \Sigma z^2/D = 0.0202447$
$g_2 = -\Sigma xz/D = -0.000360797$
$g_3 = \Sigma x^2/D = 6.7388 \times 10^{-6}$

(6) Regression coefficients
$b_1 = g_1 \cdot \Sigma xy + g_2 \cdot \Sigma zy = 0.371583$
$b_2 = g_2 \cdot \Sigma xy + g_3 \cdot \Sigma zy = 0.00953469$
$b_0 = \bar{Y} - b_1 \cdot \bar{X} - b_2 \cdot \bar{Z} = 8.31177$

(7) Regression sum of squares
$\Sigma \hat{y}^2 = b_1 \cdot \Sigma xy + b_2 \cdot \Sigma zy = 852.34$

(8) Residual sum of squares
$\Sigma d^2 = \Sigma y^2 - \Sigma \hat{y}^2 = 141.66$

(9) Residual degrees of freedom
$df = n - 3 = 6$

(2) Totals and products			
ΣY^2 7718	ΣY 246	ΣXY 7839	ΣZY 274661
	n 9	ΣX 252	ΣZ 8134
		ΣX^2 8134	ΣXZ 285468
			ΣZ^2 10,589830

(3) Means
$\bar{Y} = \Sigma Y/n = 27.3$
$\bar{X} = \Sigma X/n = 28$
$\bar{Z} = \Sigma Z/n = 903.7$

Part 2

Form A.3 (...Part 2)

(10) Correlations between variables
$r_{yx} = \Sigma xy / \sqrt{(\Sigma x^2 \cdot \Sigma y^2)} = 0.9187$
$r_{yz} = \Sigma yz / \sqrt{(\Sigma z^2 \cdot \Sigma y^2)} = 0.9224$
$r_{xz} = \Sigma xz / \sqrt{(\Sigma x^2 \cdot \Sigma z^2)} = 0.9768$

(11) Multiple correlation coefficient R and R ²
$R^2 = \Sigma \hat{y}^2 / \Sigma y^2 = 0.8575$
$R = \sqrt{R^2} = 0.9260$

(12) Analysis of variance				
	sum of squares	d.f.	mean square	variance ratio
Regression	$\Sigma \hat{y}^2$ 852.34	2	426.17	18.05
Residual	Σd^2 141.66	df 6	23.61 = s ²	
Total	Σy^2 994.00	n-1 8		

(13) Residual standard deviation
$s = \sqrt{s^2} = 4.86$

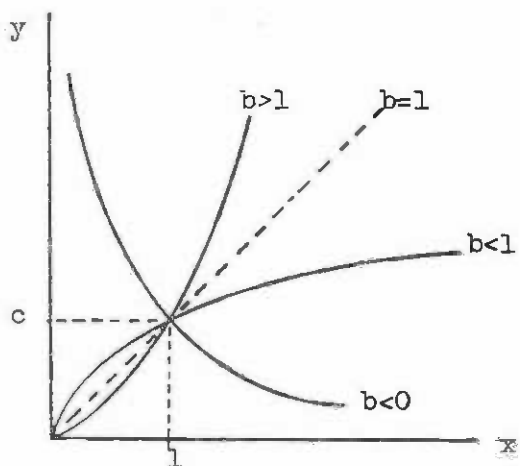
(14) Standard error and t-value of coefficients		
Coefficient	b_1	b_2
Standard error s_b	$s \cdot \sqrt{g_1} = 0.6915$	$s \cdot \sqrt{g_3} = 0.01262$
$t = b/s_b$	0.537	0.756

(15) Confidence intervals for predictions
For a series of m estimates of Y at a given level of X and Z, the standard error of the mean prediction is given by:
$s_y = s \cdot \sqrt{(1/m + 1/n + g_1 \cdot x^2 + 2 \cdot g_2 \cdot x \cdot z + g_3 \cdot z^2)}$
where $x = (X - \bar{X})$ and $z = (Z - \bar{Z})$. For a large sample, 1/m will be zero; for a single point it will be unity.

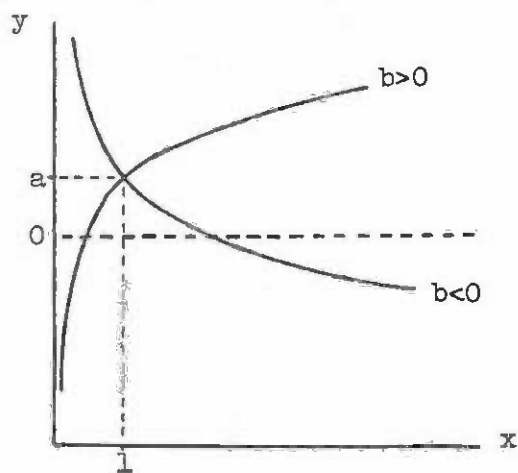
Figure A.2.1 Curve shapes resulting from different functions that can be fitted by linear regression.

($c = 10^a$ or e^a depending on whether logarithms to base 10 or base e are used.)

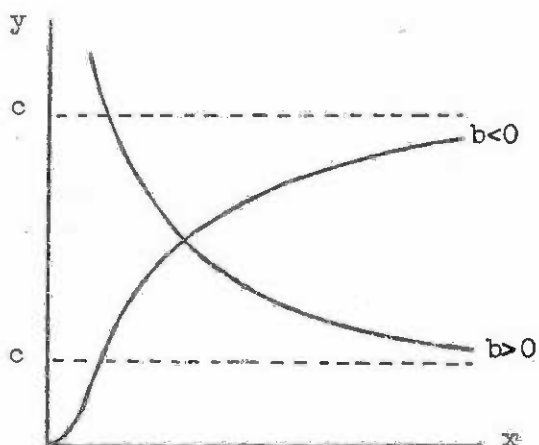
(a) $\log y = a + b \log x$



(b) $y = a + b \log x$



(c) $\log y = a + b/x$



(d) $\log y = a + b \cdot x$

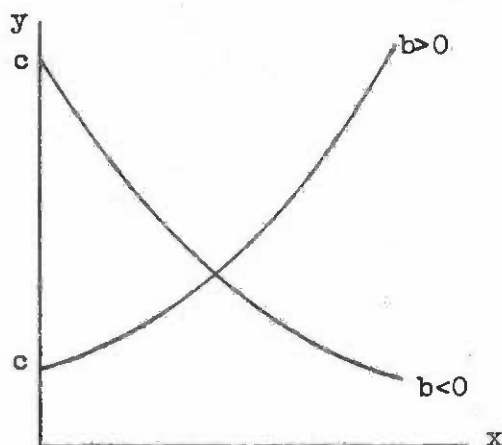
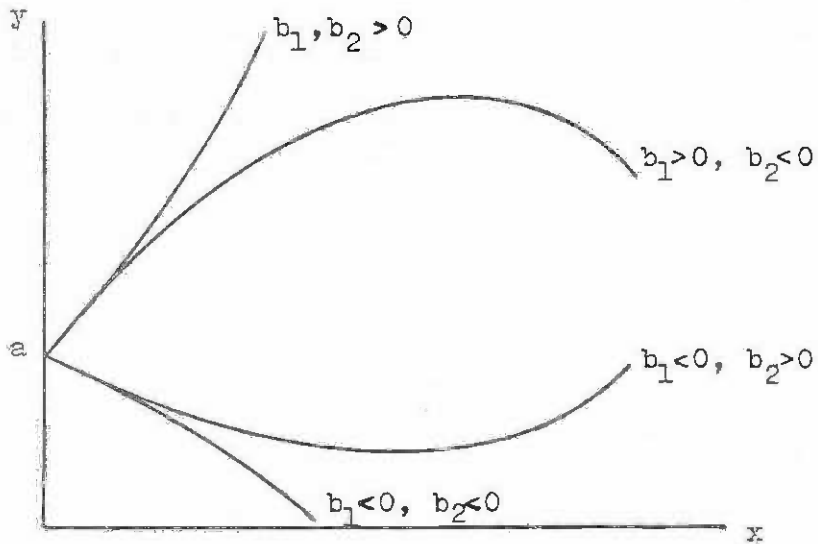
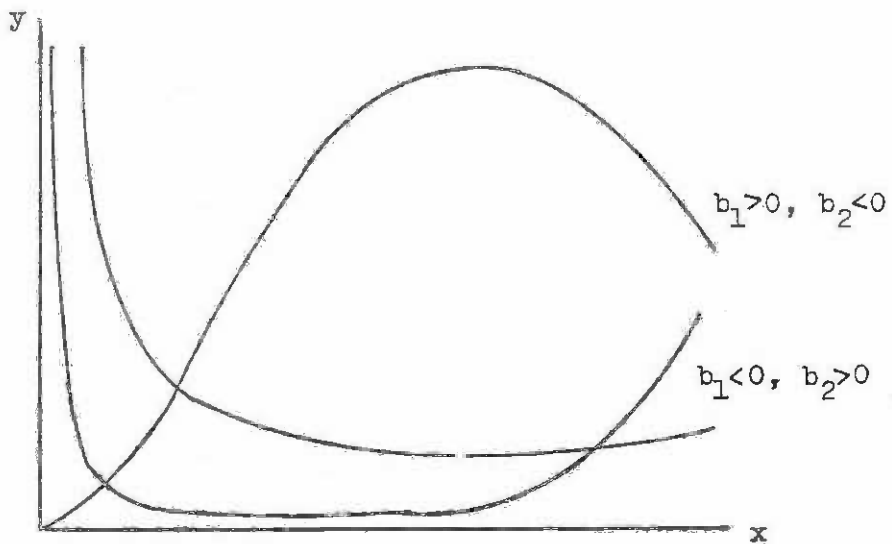


Figure A.2.1 (continued....)

(e) $y = a + b_1x + b_2x^2$



(f) $\log y = a + b_1 \log x + b_2 x$



The effect of data transformations upon the basic assumptions of linear regression involves three important points:

- (i) Regressions in which the same dependent variable has been subject to different transformations cannot be compared directly for goodness of fit using the correlation coefficient R (or R^2).
- (ii) The distribution of the residuals will be influenced by any transformation of the dependent variable.
- (iii) The regression may be biased by a transformation of the dependent variable.

For comparison of regressions for goodness of fit when several transformations of the dependent variable are involved, the Furnival Index must be used^{1/}. This is calculated as:

$$FI = s. (\prod f'(y)^{-1})^{1/n}$$

or more simply:

$$FI = s. \text{antilog} ((\sum \log f'(y)^{-1})/n)$$

Here s is the residual standard deviation from the fitted regression; n is the number of data points; and $f'(y)^{-1}$ is the reciprocal of the derivative of the transformation applied to the y variable with respect to y.

Since the non-mathematical reader will not be familiar with derivatives, appropriate forms for the commonest transformations are given below:

<u>Transformation</u>	<u>$f'(y)^{-1}$</u>
None	1
$\log_{10} y$	2.3026 y
$\log_e y$	y
y/w	w
1/y	$-y^2$
y^k	$1/(ky^{k-1})$

In the above y is the original data variable, w is any weight to be used in fitting the y to normalise residuals (see later section on weighted regression for more details), k is any constant that is used to transform y by raising it to a constant power. E.g. if y^2 were the dependent variable, k would be 2. Note that when no transformation is applied the Furnival Index is identical with the residual standard deviation.

^{1/} See Furnival, G.M., 1961 An index for comparing equations used in constructing volume tables. Forest Science 7(4)337-341.

Given a set of regressions for which Furnival Indices have been computed, all involving the same basic dependent variable, the equation which fits best will be that with the smallest index.

The effect of transformations upon the residuals' distribution and upon the bias of the regression will be considered in detail in the section upon residual analysis and weighting. For simple usage of the transformation principle in association with manual calculation, the correction factor for bias resulting from a logarithmic transformation of the dependent variable should be mentioned.

This correction factor for bias, due to Meyer^y, assumes that the model being calculated has the form:

$$\log_a y = f(x) + e_i$$

where the errors e_i are normally distributed and homogeneous with a standard deviation s , which is the residual standard deviation calculated in the regression analysis. The model that is to be actually applied is however:

$$y = a^{(f(x))}$$

where a is either 10 or e depending on whether common or natural logarithms are to be used. If the assumptions about error distribution are correct, then a systematic error will occur which can be compensated for by using a correction factor C so that the above formula becomes:

$$y = C \cdot a^{(f(x))}$$

The correction factors for common and natural logs are:

- (i) Common (base 10) logs:
 $C = 10 \cdot 1.1513 s^2$
- (ii) Natural (base e) logs:
 $C = e^{s^2/2}$

where s is the residual standard deviation obtained from the original regression with log y .

The validity of the correction factor is dependent upon assumptions about error distribution which need not be correct. Hence, it is not possible to state in any fixed or dogmatic fashion that this correction should or should not be used in a particular case; preferably the distribution of the residuals should be examined graphically as described in the section on residual analysis before arriving at any decision.

In practice, when the degree of fit obtained is high (i.e. R^2 over about 0.9) then the various arguments about alternative fitting methods, error distributions, etc., are essentially academic; the fitted function may be safely manipulated and transformed as if it were a deterministic algebraic function.

^y See Meyer, H.A., 1944. A correction for systematic error occurring in the application of the logarithmic volume equation. Pennsylvania State University Forest Research Paper 7.

2.4 Multiple Regression Analysis

Simple linear regression and regression with two predictor variables represent models, respectively, of the type:

$$Y = a + b.X \tag{1}$$

$$Y = a + b_1.X + b_2.Z \tag{2}$$

where a, b, b₁ and b₂ are the coefficients to be estimated and X, Z and Y are variables. Both of these are examples of the general multivariate linear regression model which has the form:

$$Y = b_0 + b_1.X_1 + b_2.X_2 + b_3.X_3 + \dots + b_m.X_m \tag{3}$$

where the b_i are the coefficients to be estimated by regression and the X_i are different predictor variables. This can be written in short form as:

$$Y = b_0 + \sum_{i=1}^m b_i.X_i \tag{4}$$

The various predictor variables can be of the following different sorts:

- (i) Each X_i may be a separate independent variable. For example, in a particular application X₁ may be height and X₂ diameter. The independent variables may be more or less correlated among themselves, provided the correlation is not perfect (i.e. R = 1). In the latter case one of the variables involved must be omitted.
- (ii) Some variables may be combinations and/or transformations of other variables. For example, X₃ might be log(X₂) and X₄ might be X₁.X₂. However, additive combinations (e.g. X₃ = X₁ + X₂ or X₄ = X₁ - X₂) are not permitted; they result in perfect correlation among variables.
- (iii) Some variables may be conditional variables having values of only zero or one. For example, in a particular regression, X₁ might be zero for data from one seed source and one for data from a different source. This is discussed further in the section on comparison of regressions and covariance analysis.

In most applications in forest yield prediction only a few basic independent variables are involved, often only 2 or 3, but with perhaps 2 or 3 other variables constructed as transformations in order to provide particular curve shapes. For example, the site index model:

$$\log H = b_0 + b_1/A + b_2.S + b_3.S/A$$

involves three basic variables: height (H), age (A) and site index (S). These are transformed and combined to give the regression model:

$$Y = b_0 + b_1.X_1 + b_2.X_2 + b_3.X_3$$

where Y is log(H), X₁ is 1/A, X₂ is S and X₃ is S/A.

2.5 Calculations for Estimating the Parameters of a Multiple Regression Model

As in the previous cases of regression with one and two predictor variables, in multiple regression one is primarily concerned to calculate the coefficient values, the coefficient of determination (R^2), the residual standard deviation, the variance ratio of the residuals to the regression mean square, the standard errors of the coefficients and the statistics necessary for constructing confidence intervals for predicted values.

These parameters are usually estimated using a computer programme. Manual calculations are very tedious and it is very easy for errors to slip in. However, the manual calculation procedures are given here both for the benefit of those who wish to use them as they stand and as an algorithm that will provide the basis for a computer programme if it is necessary to write one from scratch.

It is best to consider the calculations as following a number of stages:

Stage 1 Define the model to be fitted and tabulate the transformed variable values.

The general model is:

$$Y_j = b_0 + \sum_{i=1}^m b_i \cdot X_{ij} + e_j$$

where Y_j are the dependent variable observations ($j = 1$ to n); X_{ij} are the predictor variable observations with m variables per observation ($i = 1$ to m) and n observations. The b_i are the coefficients to be estimated, the e_j are the residuals between the observed Y_s and the Y_s which will be estimated from the linear model.

In the calculation example we wish to fit the model:

$$Y = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + b_3 \cdot X_3$$

given the data:

Y	X_1	X_2	X_3
1	4	9	4
3	3	9	9
7	3	6	6
9	5	8	3
14	4	7	1
18	6	4	5
19	5	3	7
21	7	1	9
24	6	3	4
26	7	2	5

where n is 10 (number of observations) and m is 3 (number of predictor variables).

Stage 2 Form the totals and cross-products of the data defined algebraically as:

(i) Totals

$$T_i = \sum_j^n X_{ij} \quad \text{for } i = 1 \text{ to } m$$

$$T_y = \sum_j Y_j$$

(ii) Cross-products

$$PX_{ik} = \sum_j^n X_{ij} \cdot X_{kj} \quad \text{for } i \text{ and } k = 1 \text{ to } m$$

$$PY_i = \sum_j^n Y_j \cdot X_{ij} \quad \text{for } i = 1 \text{ to } m$$

For the example we have the following totals and products:

k	y	1	2	3
T_k		142	50	52
			52	53

PX_{ik}	k	1	2	3
	i	1	2	3
	1	270	227	267
	2	-	350	256
	3	-	-	339

k	1	2	3
PY_k	811	522	752

The total sum of squares of the Ys must also be calculated:

$$\Sigma Y^2 = 2714$$

Stage 3 Correct the sums of products for the means. From the values PX_{ik} , PY_i and Y^2 the following corrected values are formed:

$$QX_{ik} = PX_{ik} - T_i \cdot T_k / n$$

$$QY_i = PY_i - T_i \cdot T_y / n$$

$$\Sigma Y^2 = \Sigma Y^2 - T_y^2 / n$$

The calculated values for the example are:

QX_{ik}	$i \backslash k$	1	2	3
	1	20	-33	2
	2	-	79.6	-19.6
	3	-	-	58.1
	k	1	2	3
	QY_k	101	-216.4	-0.6

$$\Sigma Y^2 = 697.6$$

Stage 4 Solve for the coefficient values and at the same time invert the cross-product matrix using Jordan elimination.

QX and QY are arranged in a table (which we will call C) in the following way:

1	2	3	...	k	...	m	1
2							2
.							.
.		QX_{ik}					QY_i
.							.
i							.
.							.
.							.
.							.
m							m

At the same time, the values QX_{21} , QX_{31} , QX_{32} , etc. are filled in as having the same values as QX_{12} , QX_{13} , etc. For the example we have therefore:

C_{ij}	$i \backslash j$	1	2	3	4
	1	20	-33	2	101
	2	-33	79.6	-19.6	-216.4
	3	2	-19.6	58.1	-0.6

Having drawn up the table or matrix C we are now ready to proceed with the steps of the Jordan elimination process.

Jordan elimination

k'th column			
c_{ij}	c_{ik}	c_{ij}	
c_{kj}	c_{kk}	c_{kj}	k'th row
c_{ij}	c_{ik}	c_{ij}	

Perform m reductions of the table. On the k 'th reduction replace each element c_{ij} of the table by a new value c_{ij}^* according to the following sequence:

(i) let $c_{kk}^* = 1/c_{kk}$ (c_{kk} is called the 'pivot')

(ii) let each element of the k 'th column, excluding the pivot, be replaced by:

$$c_{ik}^* = -c_{ik} \cdot c_{kk}^*$$

(iii) let all the elements c_{ij} , but excluding the k 'th row or the k 'th column, be replaced by:

$$c_{ij}^* = c_{ij} + c_{ik}^* \cdot c_{kj}$$

(iv) let the elements of the k 'th row, excluding the pivot, be replaced by:

$$c_{kj}^* = c_{kj} \cdot c_{kk}^*$$

After the last reduction then the elements c_{11} to c_{mm} comprise the 'inverse matrix' of the regression problem. These elements, as we shall see, are important in calculating the variances and covariances of the coefficients and hence the confidence limits of the regression estimates. The elements in the $m+1$ 'th column are the regression coefficients b_1 to b_m . For the numerical example the three reductions are shown below:

REDUCTION 1

0.05000000	-1.65000000	0.10000000	5.05000000
1.65000000	25.15000000	-16.30000000	-49.75000000
-0.10000000	-16.30000000	57.90000000	-10.70000000

REDUCTION 2

0.15825049	0.06560636	-0.96938363	1.78608349
0.06560636	0.03976143	-0.64811133	-1.97813121
0.96938369	0.64811133	47.33578528	-42.94353876

REDUCTION 3

0.17810238	0.07887895	0.02047887	0.90664807
0.07887895	0.04863523	0.01369178	-2.56610485
0.02047887	0.01369178	0.02112566	-0.90721086

Although Jordan elimination sounds fairly complex when explained in full, and is certainly very tedious to carry out manually, it can be programmed very easily for a computer. Below is a section of code in BASIC which inverts and solves the regression problem in the matrix C. The variable letters used correspond exactly to the explanation above.

Code in BASIC for Jordan Eliminations

```
100 FOR K = 1 TO M
110 LET C(K,K) = 1/C(K,K)
120 FOR I = 1 TO M
130 IF I = K THEN 190
140 LET C(I,K) = -C(I,K)*C(K,K)
150 FOR J = 1 TO M+1
160 IF J=K THEN 180
170 LET C(I,J)=C(I,J)+C(I,K)*C(K,J)
180 NEXT J
190 NEXT I
200 FOR J = 1 TO M+1
210 IF J=K THEN 230
220 LET C(K,J) = C(K,J)*C(K,K)
230 NEXT J
240 NEXT K
```

The coefficients b_1 to b_m have now been calculated and are stored in $c_{1,m+1}$ to $c_{m,m+1}$. Thus, for the numerical example we have:

$b_1 = 0.9066$
 $b_2 = -2.566$
 $b_3 = -0.9072$

rounding to 4 significant figures. The intercept b_0 is calculated from:

$$b_0 = (T_y - \sum_{i=1}^m b_i \cdot T_i) / n$$

For the numerical example we have:

$$b_0 = (142 - (50 \times 0.9066 + 52 \times (-2.566) + 53 \times (-0.9072)))/10 \\ = 27.82 \text{ (to 4 sig. figures)}$$

The elements c_{11} to c_{mm} of the final table after the last Jordan reduction we will refer to as the inverse matrix c .

Stage 5 Calculate total, regression and residual sum of squares and hence coefficient of determination (R^2), variance ratio (F) and residual standard deviation.

The total sum of squares (TSS) is $\sum y^2$, as calculated at stage 3.

The regression sum of squares (RSS) is given by:

$$RSS = \sum_{i=1}^m b_i \cdot QY_i$$

The residual sum of squares (DSS) is obtained as the difference between these two:

$$DSS = TSS - RSS$$

The degrees of freedom for the total sum of squares is $n-1$. For the regression it is m . For the residuals it is $n-m-1$.

The mean square of the regression (RMS) is:

$$RMS = RSS/m$$

The mean square of the residuals (DMS) is:

$$DMS = DSS/(n-m-1)$$

The variance ratio is:

$$F = RMS/DMS \text{ with } n-m-1 \text{ and } m \text{ degrees of freedom.}$$

This F value can be used to test the statistical significance of the regression.

For the numerical example we have therefore:

$$TSS = 697.6 \quad (\text{see } \sum y^2 \text{ in stage 3})$$

$$RSS = 0.9066 \times 101 + (-2.566) \times (-216.4) + (-0.9072) \times (-0.6) \\ = 647.4$$

$$DSS = 697.6 - 647.4 \\ = 50.2$$

$$RMS = 647.4/3 \\ = 215.8$$

$$DMS = 50.2/(10-3-1) \\ = 8.367$$

$$F = 215.8/8.367$$

$$= 25.79 \text{ with 6 and 3 d.f.}$$

The F values from tables with these d.f. and for P=0.1% is 23.70. Hence we may say that there is less than 1 chance in 1 000 of the relation between the dependent and predictor variables being due purely to chance factors outside the regression model.

The coefficient of determination R^2 is given by:

$$R^2 = \text{RSS/TSS}$$

The multiple correlation coefficient $R = \sqrt{R^2}$.

For the example we have:

$$\begin{aligned} R^2 &= 647.4/697.6 \\ &= 0.9280 \\ R &= 0.9633 \end{aligned}$$

The residual standard deviation s is the square root of the residual mean square:

$$s = \sqrt{\text{DMS}}$$

which for the example is:

$$s = 2.893$$

Stage 6 Calculate standard errors and t-statistics for the regression coefficients.

The standard errors of the coefficients b_1 to b_m are given by:

$$s(b_i) = s \cdot \sqrt{c_{ii}}$$

The t-values of the b_i are given by:

$$t(b_i) = b_i / s(b_i)$$

These t-values can be tested for significance with $n-m-1$ d.f. For the numerical example we have:

Coefficient	Standard Error	t	Significance
b_1 0.9066	1.221	0.742	-
b_2 -2.566	0.6380	-4.022	**
b_3 -0.9072	0.4205	-2.157	-

** indicates significance at the 1% level ($t=3.71$ with 6 d.f.). Coefficients b_1 and b_3 are not significant. This does not mean that both can be removed from the regression. One or the other (preferably the least significant, here b_1) could be removed and the results recalculated by the procedure explained later. The t-values of the remaining coefficients would be found to have increased and both would probably be significant in the new regression.

Stage 7 Standard errors of predictions from the model are derived from the formula:

$$S_y = s \sqrt{1/k + 1/n + \sum_i \sum_j c_{ij} x_i x_j}$$

where $x_i = (X_i - \bar{X}_i)$, X_i is a given value of predictor variable i , \bar{X} is the mean value for the regression model data and k is the number of repeated estimates of Y to be made. S_y is the standard error of the mean estimate of Y . For a single prediction, $k = 1$; for the population standard error, $1/k = 0$.

For the example, suppose we wish to calculate the standard error for a single estimate of Y at $X_1 = 8$, $X_2 = 4$ and $X_3 = 6$.

First calculate the x_i :

$$\begin{aligned}x_1 &= 8 - 5.0 = 3 \\x_2 &= 4 - 5.2 = -1.2 \\x_3 &= 6 - 5.3 = 0.7\end{aligned}$$

Then calculate the sum $\sum_i \sum_j c_{ij} x_i x_j$:

$$\begin{aligned}& 3(3 \times 0.1781 + (-1.2) \times 0.07888 + 0.7 \times 0.02048) \\& + (-1.2) \times (3 \times 0.07888 + (-1.2) \times 0.04864 + 0.7 \times 0.01369) \\& + 0.7(3 \times 0.02048 + (-1.2) \times 0.01369 + 0.7 \times 0.02113) \\& = 1.178\end{aligned}$$

$$\begin{aligned}\text{Then } S_y &= 2.893 \sqrt{1 + 1/10 + 1.178} \\& = 4.366\end{aligned}$$

Thus, the standard error for a single estimate of Y from this model at $X_1 = 8$, $X_2 = 4$, $X_3 = 6$ is ± 4.366 .

This formula will also give the standard error of the intercept b_0 , if required, using $1/k = 0$ and $x_i = \bar{X}_i$.

This completes the main calculations for multiple regression. Two other types of calculation are commonly performed. These are the computation of the correlation matrix between variables and the shortcut calculation for removal of variables from a regression.

The correlation matrix This is a table showing the correlation between any pair of variables. It is constructed by arranging a matrix V as follows:

	1	2	3	4	m + 1
1	y^2	QY				
2	QY	QX				
3						
.....						
.....						
.....						
m + 1						

and then noting that the correlation between variable i and variable j is:

$$r_{ij} = v_{ij} / \sqrt{(v_{ii} \cdot v_{jj})}$$

This can be done at any time after stage 3 in the regression calculation.

Since the matrix V is symmetrical about the diagonal V_{11} to V_{mm} , it is not necessary to calculate the lower half, as, for example, $r_{23} = r_{32}$, and so on.

Removing a variable from the regression When a variable is found to be non-significant it may be desirable to recalculate the regression statistics with that variable omitted.

To do this the values of the inverse matrix c and the coefficients are adjusted according to the following formula where u denotes the subscript of the unwanted variable:

$$b_i^* = b_i - c_{iu} \cdot b_u / c_{uu}$$

$$c_{ij}^* = c_{ij} - c_{iu} \cdot c_{ju} / c_{uu}$$

For the numerical example suppose we wish to omit the effect of variable X_1 . The original inverse matrix and coefficients were (to 6 significant figures):

	1	2	3	b
1	0.178102	0.0788790	0.0204789	0.906648
2	0.0788790	0.0486352	0.0136918	-2.56610
3	0.0204789	0.0136918	0.0211257	-0.907211

Eliminating variable 1 we have:

$$b_2^* = -2.56610 - 0.0788790 \times 0.906648 / 0.178102$$

$$= -2.96764$$

$$b_3^* = -0.907211 - 0.0204789 \times 0.906648 / 0.178102$$

$$= -1.01146$$

The revised inverse matrix is:

	2	3
2	0.0137007	0.00462197
3	0.00462197	0.0187710

The various other statistics from stage 5 on can then be recalculated for the new regression. A second, third, etc. variable may be successively eliminated in this way.

This completes the summary of multiple regression calculations. This work is usually performed by computer but can be done by hand for small problems. Great care should be taken not to lose significant digits by rounding to a fixed number of decimal places during the matrix inversion process.

2.6 Residual Analysis

Residual analysis is a tool that should be associated with the intelligent use of multiple regression. Its purpose is threefold:

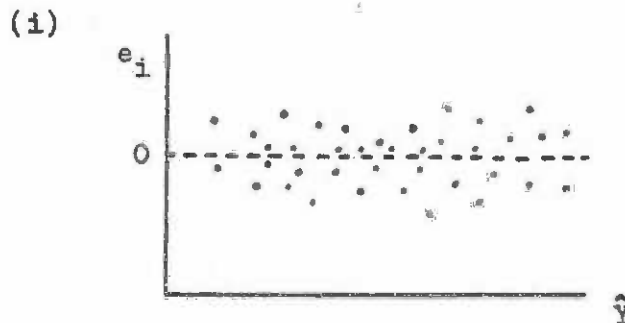
- (i) To determine whether the residuals from the regression conform to the assumptions of the model, i.e. are uncorrelated, normally distributed and of uniform variance.
- (ii) To assess 'lack of fit' in the model from systematic trends in the residuals.
- (iii) To examine visually the shape of relationships between residuals and possible predictor variables not yet introduced into the regression model.

Residual analysis (like most other aspects of regression calculations!) is very tedious to perform by hand. It should however be part of every good computer programme for multiple regression analysis.

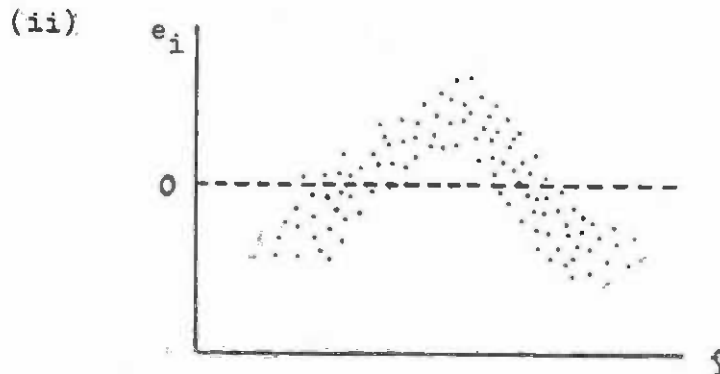
The residual e is defined as the difference between the actual value of the dependent variable Y and the value predicted from the regression model \hat{Y} . That is:

$$e_i = Y_i - \hat{Y}_i$$

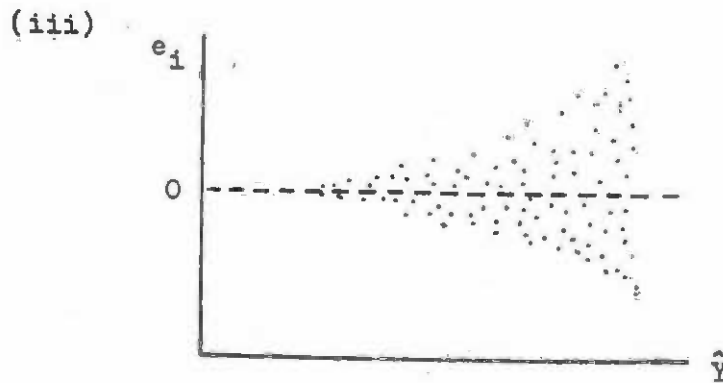
When the e_i are plotted against the \hat{Y}_i then several possible kinds of graphs may be obtained:



This is the ideal situation with the variation in the e_i uniform with respect to \hat{Y} . The model appears to have no undesirable characteristics.



This indicates systematic behaviour of the e_i with respect to Y and shows lack of fit. In other words, there is a better model than the current one, using the same basic predictor variables but with different or additional transformations that could be fitted and would remove the systematic trend in the e_i .



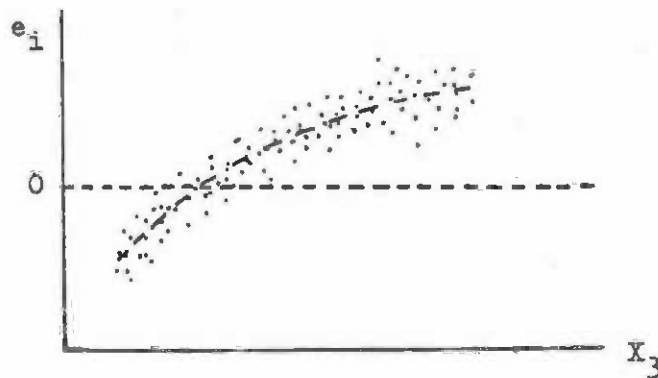
Here the variances of the e_i are not uniform with respect to Y . In this case, better estimates of the coefficients can be obtained either by using a transformation of the Y variable or through the use of weighted regression (see next section 2.7).

When the e_i are plotted against X variables not yet included in the model then one may have either a random scatter or some systematic pattern similar to (ii) above. In the latter case the X variable could be included in the regression. The shape of the general trend in the pattern may not be linear, but following one of the trends shown in figure A.2.1. In this case the appropriate transformation of the X variable can be used.

For example, suppose we have the model:

$$Y = b_0 + b_1 X_1 + b_2 X_2$$

where the residuals plotted against an additional variable X_3 give the following pattern:



The dotted line (---) appears to follow a slight curve which could be allowed for by adding X_3 and X_3^2 to the regression (c.f. figure A.2.1(e)) to give the model:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_3^2$$

After fitting this second model the trend of the residuals from the new model against X_3 would appear similar to the first residual pattern shown in (i) above.

2.7 Weighted Regression

Weighted regression is used when the residuals do not have uniform variance with respect to the Y values. This situation commonly arises when complex transformations of the Y variable are used to obtain particular curve shapes. In yield prediction it is also commonly found with volume data where the variance tends to be proportional to the volume.

Associated with each observation is a particular weight w_i . The value of this can be determined empirically or it may be derived from some theoretical reasoning. For the best fit the weights w_i should be proportional to $1/s_i^2$ where s_i is the residual standard deviation at Y_i .

Empirical weights can be calculated as follows:

(1) Fit the unweighted regression model using the procedures described in the preceding section.

(2) Calculate the squared residuals:

$$e_i^2 = (Y_i - \hat{Y}_i)^2$$

(3) Then either:

(i) Tabulate mean values of w_i ($= 1/e_i^2$) for classes of Y_i ;

or

(ii) Fit a regression relating e_i^2 to Y_i (possibly a simple regression:

$$e^2 = a + b.Y$$

will be adequate for most cases). Use this regression to predict a weight w_i ($= 1/e_i^2$) for each Y_i .

In the case of theoretical weights then one uses whatever formulation is involved to calculate the w_i for each Y_i .

For weighted linear regression with one predictor variable, the calculations are as follows:

- (1) Compute weighted totals, sums of squares and products:

$$\begin{array}{lll} A = \sum w_i & B = \sum w_i \cdot x_i & C = \sum w_i \cdot x_i^2 \\ D = \sum w_i \cdot y_i & E = \sum w_i \cdot y_i^2 & F = \sum w_i \cdot x_i \cdot y_i \end{array}$$

- (2) Calculate weighted means and corrected products and squares:

$$\begin{array}{ll} \bar{x} = B/A & \bar{y} = D/A \\ SS_x = C - B^2/A & SP_{xy} = F - B \cdot D/A \end{array}$$

- (3) All other calculations then proceed as from line 11 of form A.2 (Statform 11 of Dawkins, 1968).

For multiple weighted regression the same general principles apply. Most advanced regression programmes have a facility for weighting.

Weighted regression should not be used unless there is strong a priori theoretical justification for its use, or alternatively, a clear indicated trend in the squared residuals when plotted against \bar{Y} that may provide the basis for empirical weighting.

For well defined functions with high R^2 , weighting is unlikely to result in much practical improvement in the degree of fit. It is a technique that may be regarded as a refinement to analysis to be used once preliminary analyses of a set of data have been completed.

2.8 Comparison of Regression Lines

It is frequently necessary to try and decide whether or not two or more regression lines are so similar that they can effectively be replaced by a single line. This case may arise, for example, when it is desired to pool data from different species with similar growth habits or from different forests, regions, site types or provenances.

The regressions to be compared may be simple regressions with a single predictor variable or multivariate regressions. Two basic techniques exist for the comparison:

- (i) Covariance analysis. This is perhaps more suited to manual computation and simple regression, although it can equally well be extended to the multivariate case.
- (ii) Significance tests on conditional variables. This is quite suitable for complex models and interactions, but leads to large multivariate problems. It is a useful method of adapting standard computer multiple regression packages to regression comparison problems.

Covariance analysis is a large subject. Here, only the techniques necessary for comparing the slopes and intercepts between simple regressions are considered. The calculations are explained in relation to an example representing hypothetical data for the volume-height line from two geographically distinct plantations (called I and II), using 10 plots on each. The data is illustrated in figure A.2.2 and set out in table A.2.1. The problem is to determine whether it is reasonable to combine the data to fit a common regression line based on all 20 plots. The model to be fitted is the logarithmic volume line with the form:

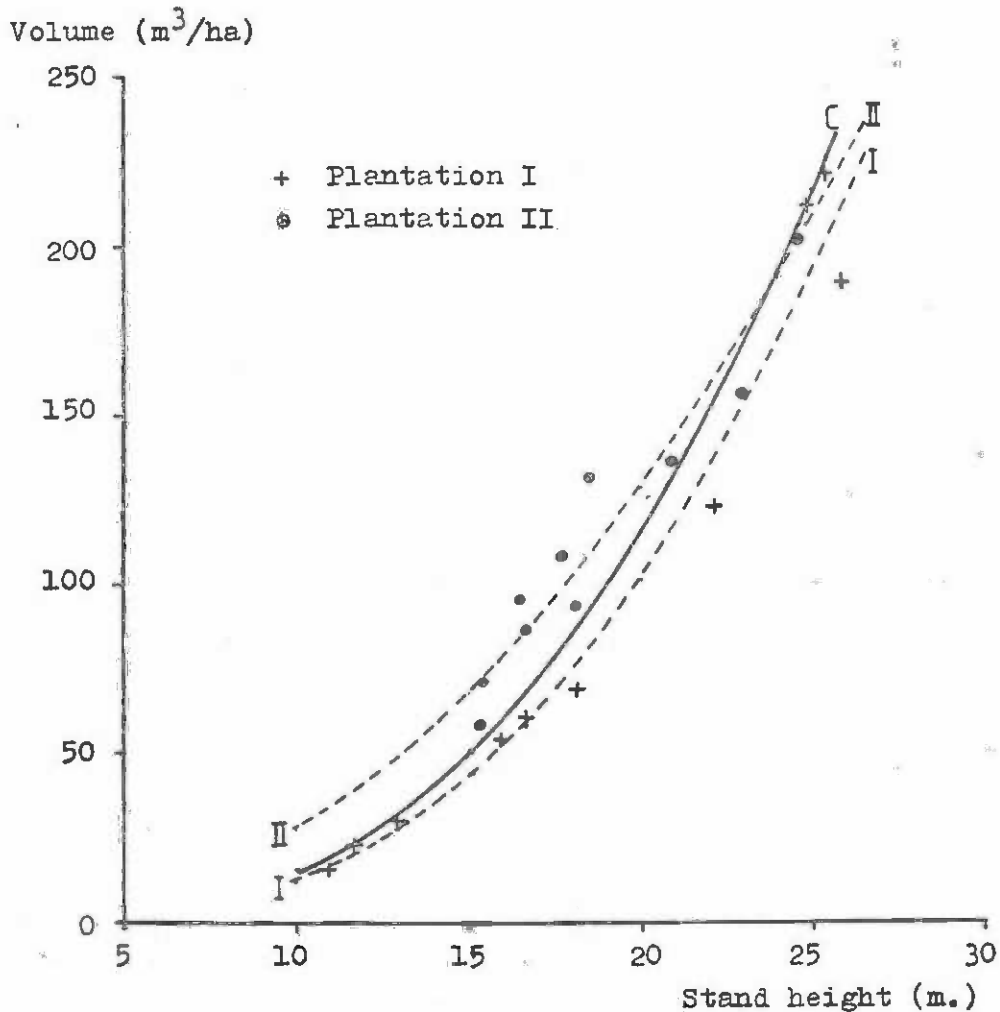
$$\log V = a + b \cdot \log H \quad \text{---(1)}$$

This is the linear form of the model with the parameters a and b being estimated by simple regression. By taking antilogs to both sides the equation becomes:

$$V = A \cdot H^b \quad \text{---(2)}$$

where $A = 10^a$. In this form points from the fitted line can be calculated for plotting on normal graph paper.

Figure A.2.2 Hypothetical volume-height data used for comparison of regression lines example, section 2.9.



--- individual regression lines; — combined regression

Table A.2.1. Original and transformed data from two plantations used in the regression comparison example.

Plot No.	Volume m ³ /ha	Height m.	log. V y	log. H x
Plantation I				
1	30.0	12.9	1.4771	1.1106
2	23.3	11.6	1.3674	1.0645
3	212.6	24.9	2.3276	1.3962
4	15.1	10.8	1.1790	1.0334
5	50.4	15.8	1.7024	1.1987
6	189.4	25.8	2.2774	1.4116
7	68.7	18.2	1.8370	1.2601
8	221.2	25.4	2.3448	1.4048
9	123.9	22.1	2.0931	1.3444
10	60.8	16.7	1.7839	1.2227
Plantation II				
11	70.6	15.4	1.8488	1.1875
12	131.3	18.5	2.1183	1.2672
13	201.7	24.6	2.3047	1.3909
14	156.6	22.9	2.1948	1.3598
15	136.2	20.8	2.1342	1.3181
16	87.0	16.7	1.9395	1.2227
17	95.0	16.5	1.9777	1.2175
18	108.4	17.7	2.0350	1.2480
19	93.6	18.0	1.9713	1.2553
20	57.7	15.3	1.7612	1.1847

The sums, sums of squares and products of the data (table A.2.1) are calculated both for the separate sets of data and for the two sets combined as follows:

<u>Statistic</u>	<u>Source of Data</u>		
	I	II	Combined (C)
ΣY	18.38954	20.28547	38.67501
ΣY^2	35.37950	41.39075	76.77025
ΣX	12.44696	12.65166	25.09862
ΣX^2	15.67703	16.05102	31.72805
ΣXY	23.43337	25.76286	49.18623
n	10	10	20
$SS_y = \Sigma Y^2 - (\Sigma Y)^2/n$	1.56198	0.24072	1.98243
$SS_x = \Sigma X^2 - (\Sigma X)^2/n$	0.18435	0.04457	0.23101
$SP_{xy} = \Sigma XY - \Sigma X \cdot \Sigma Y/n$	0.53398	0.09837	0.65176
coefficients:			
$b = SP_{xy}/SS_x$	2.89657	2.20716	2.82135
$a = (\Sigma Y - b \cdot \Sigma X)/n$	-1.76640	-0.76388	-1.60685
regression sums of squares:			
$RSS = b \cdot SP_{xy}$	1.54671	0.21712	1.83884

From these basic calculations the analysis of variance can be performed. The following additional quantities are calculated:

- (i) Sums of squares between b coefficients:

$$SS_b = \Sigma RSS - (\Sigma SP_{xy})^2 / \Sigma SS_x$$

For the example this is:

$$\begin{aligned} SS_b &= (1.54671 + 0.21712) - (0.53398 + 0.09837)^2 / \\ &\quad (0.18435 + 0.04457) \\ &= 0.01708 \end{aligned}$$

- (ii) The sums of squares between a coefficient:

$$SS_a = \Sigma SS_y - \Sigma RSS$$

$$\begin{aligned} &= (1.56198 + 0.24072) - (1.54671 + 0.21712) \\ &= 0.03887 \end{aligned}$$

- (iii) The residual sum of squares:

$$DSS = SS_{y_c} - (SS_a + SS_b + RSS_c)$$

where the C subscript denotes quantities from the combined regression statistics on the table above.

For the example, the deviation sum of squares is:

$$\begin{aligned} \text{DSS} &= 1.98243 - (0.03887 + 0.01708 + 1.83884) \\ &= 0.08764 \end{aligned}$$

- (iv) The degrees of freedom for the above quantities are as follows where n_I and n_{II} are the number of points in the separate regressions, r is the number of regressions being compared and n_C is the total number of points:

<u>statistic</u>	<u>d.f.</u>
SSy_C	$n_C - 1$
DSS	$n_I + n_{II} - 2r$
$\left. \begin{array}{l} \text{SSb} \\ \text{SSa} \end{array} \right\}$	$r - 1$
RSS_C	1

The analysis of variance table for the comparison of regressions may be set out as follows:

<u>Source of Variation</u>		<u>Sum of Squares</u>	<u>d.f.</u>	<u>Mean Square</u>	<u>Variance Ratio</u>
Combined regression					
	RSS_C	1.83884	1	1.83884	335.7***
Between slopes	SSb	0.01708	1	0.01708	3.1
Between intercepts					
	SSa	0.03887	1	0.03887	7.1*
Residuals	DSS	0.08764	16	0.005478	-
Total	SSy_C	1.98243	19	-	-

The asterisks denote the significance level of the different variance ratios, determined by looking up F values in statistical tables with 1 and 16 d.f. at the 5% (*), 1% (**) and 0.1% (***) probability levels. In the above the regression is very highly significant (i.e. not due to chance). The difference between slopes in the two separate regressions is not significant, i.e. is most likely due to chance sampling effects in the two sets of data. On the other hand, the intercepts do differ significantly at the 5% level, suggesting that the regressions are in fact distinct although they have the same slope. Consequently, we may conclude for the example that the data from the two plantations cannot be pooled to give a combined regression for prediction purposes without a consequent loss of accuracy.

This comparison method can be extended to any number of simple regressions by noting that in the above calculations, wherever separate quantities for the two regressions are added together, then any number of quantities may be added. Dawkins (see bibliography) presents a calculation pro forma for comparing up to 4 regressions.

The use of conditional variables for comparison of regression lines. Conditional variables are those which can only take values of zero or one, depending on whether a particular observation is to enter into some part of a regression model or not. They can be used for the comparison of regression lines and for fitting sets of nested data (see section 2.10).

Consider the case described in the example above, comparing volume height lines for two simple regressions from plantations I and II. For each plantation we have a model:

$$\log V = a + b \cdot \log H \tag{1}$$

We can form a single model for both sets of data if we introduce an extra variable Z which is zero for data from forest I and one for data from forest II. This combined model is:

$$\log V = a_1 (1 + a_2 \cdot Z) + b_1 (1 + b_2 \cdot Z) \log H \tag{3}$$

For observations with Z = 0 (forest I) then this amounts to:

$$\log V = a_1 + b_1 \log H \tag{4}$$

But when Z = 1 (forest II) the coefficients are:

$$\log V = (a_1 + a_2) + (b_1 + b_2) \log H \tag{5}$$

To fit equation (3) by linear regression the brackets are removed giving:

$$\log V = a_1 + a_1 a_2 Z + b_1 \log H + b_1 b_2 Z \cdot \log H \tag{6}$$

which is equivalent to the regression equation:

$$Y = c_0 + c_1 X_1 + c_2 X_2 + c_3 X_3$$

Using a multiple regression programme to fit the coefficients, the values obtained were:

		t
c_0	-1.76640	-12.3
c_1	1.00179	3.1
c_2	2.89657	25.2
c_3	-0.68884	-2.7

The t values for the coefficients with 16 d.f. indicate a similar but not identical result to the covariance analysis. The c_1 coefficient, corresponding to the term Z in equation (6), indicates a highly significant additional term for the intercept for the second set of data. The c_3 coefficient, corresponding to $Z \cdot \log H$, is also significant at the 5% level, indicating that the slope of the two sets of data differs. In the covariance analysis it will be remembered, the slopes did not differ significantly although the intercepts did. This difference in the results of the two methods is due to the difference in the statistical models and the hypotheses being tested. The covariance analysis asks the question:

"Do either or both of the regressions differ significantly from a pooled regression?"

whereas, this technique asks:

"Does the second regression intercept (c_1) or slope (c_3) differ from that of the first regression?"

It will be appreciated that the second type of test will be more sensitive to differences than the first, but on the otherhand, it is not so directly relevant if one wishes to know whether one can safely pool regressions.

The coefficient values for the two separate lines can be worked out from the above coefficients and will be found to be identical to the coefficients for the separate lines fitted in the covariance analysis. They are:

Forest I:

Intercept	c_0	-1.76640
Slope	c_2	2.89657

Forest II

Intercept	$c_0 + c_1$	-0.76461
Slope	$c_2 + c_3$	2.20773

In fact, the figures appear to differ in the third decimal place from those calculated earlier, but this is an effect of rounding error, as in the computer programme all calculations, including the log. transformations were carried out to 15 significant figures, whereas with the manual calculation used in the covariance analysis, only 5 to 8 significant figures were used.

The use of conditional variables can become quite complex and is also quite flexible. For example, given any dependent variable Y and any predictor variable X with a conditional variable Z, we can have a model with a common slope and only a different intercept for the two sets of data:

$$Y = c_0 + c_1 Z + c_2 X$$

With three sets of data, two conditional variables are needed. They are;

- Z_1 0 for data from set 1, 1 for sets 2 and 3.
- Z_2 0 " " " " 1 and 2, 1 for set 3.

And the model to test for different slopes and intercepts is:

$$Y = c_0 + c_1 Z_1 + c_2 Z_2 + c_3 X + c_4 Z_1 X + c_5 Z_2 X$$

Further information on the ways conditional variables may be used is given in the next section.

2.9 Nested Regression Using Conditional Variables

Conditional variables whose value may be either zero or one have already been alluded to in the previous section. Apart from their use for comparing regressions they can be used to fit regression models to nested data.

Nested data arises when the measurements are grouped into sampling units or plots where the within-plot regression may be different from the between-plot regression. The commonest situation in forest yield studies is with permanent sample plot measurements where one has data grouped by plots. A typical problem will illustrate the point:

Example Figure A.2.3 shows hypothetical height-age data from 5 permanent sample plots each of which has been remeasured 3 or 4 times. The objective is to estimate the mean slope of the height-age relationship, using a quadratic model of the form:

$$H = b_0 + b_1A + b_2A^2 \tag{1}$$

where b_1 and b_2 are assumed to be the same for all plots but b_0 may differ (expressing a site effect).

If a regression analysis is made of the data in figure A.2.3 using equation (1) as the model and treating each observation as separate and independent, then the coefficient values obtained are:

- b_0 7.807
- b_1 0.6939
- b_2 -0.008652

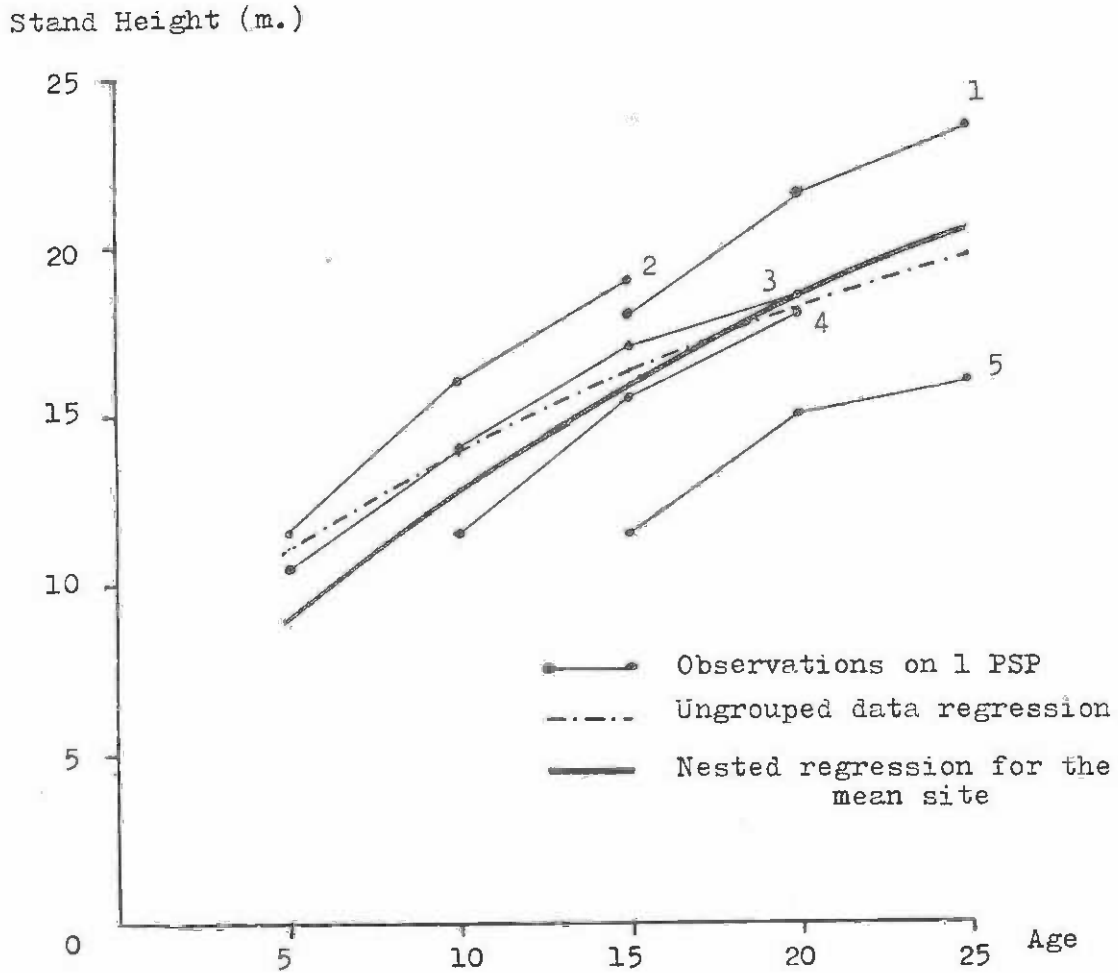
with a coefficient of determination (R^2) of 0.53 and an F value of 4.94. The significance level of the regression is 0.2%.

It can be seen from figure A.2.3 that this regression, drawn as a broken line, obviously underestimates the mean slope of the plots.

To fit the nested model then 4 artificial variables are introduced which can be called P_2 to P_5 . These have the following values:

- P_2 - 1 for data from plot 2, zero for all other data
- P_3 - 1 " " " " 3, " " " " "
- P_4 - 1 " " " " 4, " " " " "
- P_5 - 1 " " " " 5, " " " " "

Figure A.2.3 Hypothetical height-age data to exemplify effect of fitting nested regression



The model fitted is:

$$H = a_1 + a_2 P_2 + a_3 P_3 + a_4 P_4 + a_5 P_5 + b_1 A + b_2 A^2 \quad -(2)$$

where the a_i and b_i are the regression coefficients. This is the quadratic model as in equation (1), but with a different intercept for every plot. The intercept values equivalent to b_0 in equation (1) are:

<u>Plot</u>	<u>Intercept</u>
1	a_1
2	$a_1 + a_2$
3	$a_1 + a_3$
4	$a_1 + a_4$
5	$a_1 + a_5$

When fitted to the data from figure A.2.3 the coefficient values obtained were:

a_1	7.432
a_2	-0.3915
a_3	-1.755
a_4	-3.427
a_5	-6.833
b_1	0.9192
b_2	-0.01157

with a coefficient of determination of 0.98 and an F value of 70.88 with 6 and 9 degrees of freedom. This is significant at less than 0.0001%. The slope of this line, given by b_1 and b_2 , is drawn on figure A.2.3 for the mean intercept values of the plots as a heavy solid line. It can be seen that this gives a much better representation of the mean slope of the plots than the regression fitted to ungrouped data.

This technique can be extended to give different intercepts and different slopes for each plot or a common intercept and different slopes. The latter case is most likely to be useful with models that result in asymptotic graphs such as the Schumacher equation (see main text). For a single X and Y variable and with 3 plots, for example, we have:

Common slope, different intercepts:

$$Y = a_1 + a_2 P_2 + a_3 P_3 + b_1 X$$

Different slope and intercepts:

$$Y = a_1 + a_2 P_2 + a_3 P_3 + b_1 X + b_2 P_2 X + b_3 P_3 X$$

Different slope, common intercept:

$$Y = a_1 + b_1 X + b_2 P_2 X + b_3 P_3 X$$

In the above the a_i and b_i are intercept and slope coefficients to be fitted by regression; the X and Y are normal variables and the P_j are variables which are 1 for plot j and zero for other plots.

These ideas can be extended to multivariate models, although the number of coefficients becomes fairly large. Nested regressions of this type can be calculated by hand, as although the number of variables may be large, the 0-1 nature of most of the variables means that many short-cuts are possible in the calculations. If this is done, then the Jordan elimination technique for inverting the corrected cross-product matrix, given in section 2.5, is not the fastest or easiest method; Seber (1977 - see bibliography) gives details of more efficient methods for those acquainted with matrix algebra. However, for computer programmes the relative inefficiency of conventional calculation techniques is unimportant since the processing times involved are in any case very short.

3. SOLUTION OF EQUATIONS.

3.1 Solution of the Quadratic Equation

The quadratic equation is widely used as a regression model for data exhibiting a slight curvature in the X-Y trend. Having fitted an equation to predict Y, it may sometimes be necessary to solve for X given a known value for Y. To do this, one uses the standard formula for the roots of a quadratic equation. If the regression model is:

$$Y = b_0 + b_1X + b_2X^2$$

then rewrite it as:

$$aX^2 + bX + c = 0$$

where $a = b_2$, $b = b_1$ and $c = b_0 - Y$, and obtain X from:

$$X = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

There are normally 2 solutions, depending on whether a + or a - sign is taken before the square root sign. It will usually be obvious that only one of the two solutions calculated in a particular case is possible. For example, one might derive diameters of +20 and -15 as solutions to a problem; only the positive diameter has any meaning.

When b^2 is less than $4ac$, the quantity under the square root sign is negative and there is no real solution to the equation. For most forestry applications this is likely to imply an error in the coefficient values or a Y value that is too large or too small. Suppose, for example, that one has a height-age function that reaches a maximum at 42 m height. If one tries to solve it to see at what age the stand will reach 50 m, then this condition will arise; there is no solution.

3.2 Graphical Solution of Equations

There are quite a number of complex models that cannot readily be manipulated by algebra and which must be solved either by a graphical or a numerical method. The graphical method of solution is self-evident and can best be illustrated by an example. It consists simply of plotting the value of Y for an equation of the type:

$$Y = f(X)$$

for selected values of X; joining the points by a smooth curve and then estimating the value of X for the desired Y at which a solution is required.

Suppose, for example, that one has fitted the equation:

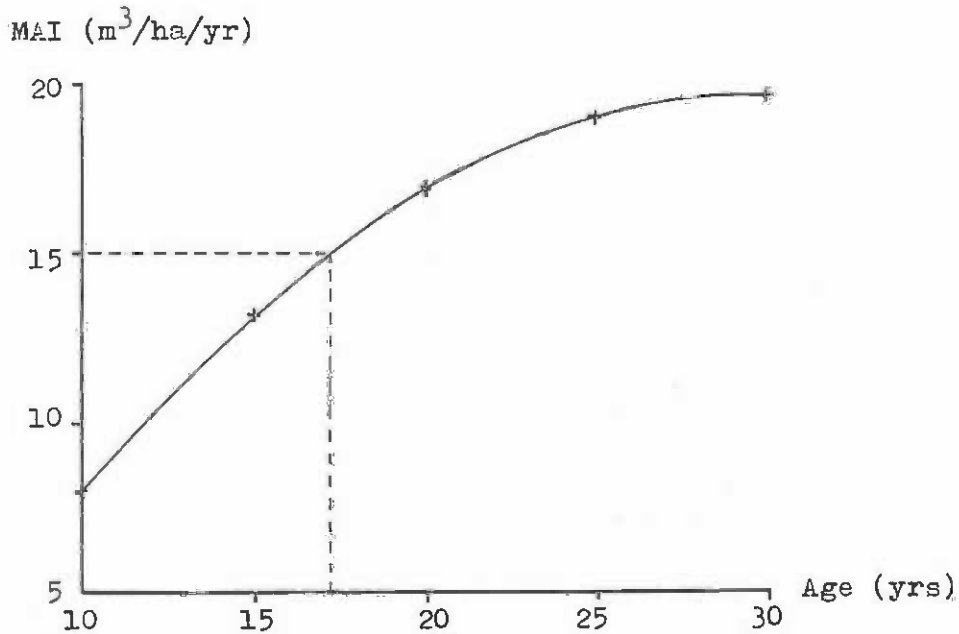
$$\log \text{MAI} = -0.8892 - 0.03055 A + 2.097 \log A$$

relating mean annual volume increment to age for a plantation and it is necessary to solve it to determine the age when MAI first reaches $15 \text{ m}^3/\text{ha}/\text{yr}$. For a graphical solution use

the equation to calculate MAI for a series of ages which it is hoped will bracket the desired answer. Below are tabulated a series of MAIs calculated for ages 10 to 30:

<u>Age</u>	<u>MAI</u>
10	7.99
15	13.15
20	16.91
25	18.99
30	19.58

These are then plotted, as shown by the +s in the graph below, and connected by a smooth curve. It is then possible to read off the x value corresponding to the known y value which in this case is, for an MAI of 25 m³/ha/yr, an age of 17.2 years.



The main disadvantages of graphical solution methods are:

- (i) The accuracy is restricted to perhaps $\pm \frac{1}{2}\%$ of the range of the x scale.
- (ii) The method is slow and essentially manual. It cannot be used within a computer programme or be programmed into a calculator.

3.3 Numerical Solution of Equations

There are a number of numerical solution methods for equations, but one of the simplest and most reliable is the so-called 'Bisection Method'. This proceeds through the following stages:

- (i) Define the function to be solved as:

$$Y = f(x)$$

where y is a known value, but x is the unknown to be determined by the method. There is a range of values within which x must be known to lie, from x_a to x_b , such that:

- (1) $f(x_a)$ is less than $f(x_b)$;
- (2) $f(x_a) - y$ has the opposite sign to $f(x_b) - y$.

- (ii) Calculate a trial value x_i from:

$$x_i = \frac{1}{2}(x_a + x_b)$$

- (iii) Calculate the value y_i for x_i :

$$y_i = f(x_i)$$

- (iv) If: y_i is greater than y , let $x_b = x_i$;
 y_i is less than y , let $x_a = x_i$;
 y_i is exactly equal to y , then the solution is x_i ;
this case is unlikely to occur in practice.

- (v) Go back and repeat steps (ii) to (iv) n times, where n is determined by the accuracy to which one wishes to approximate the correct solution. Given initial values of x_a and x_b then the accuracy of the solution will be:

$$\text{accuracy} = 2^{-n}(x_a - x_b)$$

After the n iterations have been completed, then x_i is the desired solution to the accuracy stipulated above. In practice, since the iterations are carried out at high speed by computer or programmable calculator, it is usually sufficient to use a fixed value of n , say 15 or 20 for a range of different problems. We have for example:

<u>No. of Iterations</u>	<u>Accuracy as % of Range</u>
5	3%
8	0.3%
10	0.1%
15	0.003%
20	0.0001%
30	$10^{-7}\%$

For most forestry problems 15 iterations are sufficient to give a realistic degree of accuracy.

As an example of the bisection method, suppose that we have, as in the previous section, an equation relating age (A) and mean annual increment (MAI) of the form:

$$\log \text{MAI} = -0.8892 - 0.03055 A + 2.097 \log A$$

which we wish to solve to determine the age at which MAI reaches $15 \text{ m}^3/\text{ha}/\text{yr}$. Taking x_a as 1 year, x_b as 30 years and y as 15 years, then applying stages (ii) to (iv) enumerated above, the solution goes as follows:

i.	x_a	x_b	x_i	y_i
1	1.000	30.000	15.500	13.595
2	15.500	30.000	22.750	18.254
3	15.500	22.750	19.125	16.370
4	15.500	19.125	17.312	15.091
5	15.500	17.312	16.406	14.370
6	16.406	17.312	16.859	14.737
7	16.859	17.312	17.085	14.916
8	17.085	17.312	17.199	15.004
9	17.085	17.199	17.142	14.960
10	17.142	17.199	17.170	14.982
11	17.170	17.199	17.185	14.993
12	17.185	17.199	17.192	14.999
13	17.192	17.199	17.195	15.001
14	17.192	17.195	17.193	15.000
15	17.192	17.193	17.193	14.999

The main disadvantages of the bisection method is that compared to some numerical solution methods (notably Newton's method), the convergence with each iteration is quite slow; the method also fails when there are two (or any other even number) solutions within the initial range given, and of course, one must be able to stipulate a range for the solution. For most forestry work these disadvantages seldom become important. If they do appear to be critical, then the worker should refer to some textbook on numerical methods, such as, for example, Stark (1970) for more detailed guidance.

4. FITTING NONLINEAR MODELS

A nonlinear model is any model which cannot be transformed into a form whose parameters can be estimated directly by linear least squares analysis, as described in section 2.

For example, the equation:

$$Y = a.X^b$$

can be transformed by taking logarithms of both sides to give the linear model:

$$\log Y = \log A + b \log X$$

so that $\log A$ and b can be directly estimated by linear regression. On the other hand:

$$Y = a + b.X^c$$

where a , b and c are coefficients to be determined, cannot be directly transformed into such a linear form and hence it is said to be a nonlinear model or equation.

In this manual it is only possible to deal with fitting methods for a limited range of nonlinear functions in common use in growth and yield studies characterised by the following features:

- (i) There is only one predictor (X) and one dependent (Y) variable;
- (ii) There are three parameters to be estimated (which we will call a , b and k) by the fitting process;
- (iii) Provided that parameter k is assumed to be a known value, a and b can be estimated, using appropriate transformations, by simple linear regression.

The types of equations which fall within these restrictions include the following widely used forms:

$$Y = a + b.X^k$$

$$Y = a.\exp(b.X^k)$$

$$Y = a(1 - \exp(-k.X))^b$$

There are a number of other models besides these that also can be fitted, but these are less commonly used in forestry work and hence are not listed here.

The main problem is the estimation of the parameter k . For a known value of k we can derive linear forms for the three equations as follows:

$$Y = a + b.(X^k)$$

where X^k is the predictor variable.

$$\log_e Y = \log_e a + b.(X^k)$$

where again the predictor variable is X^k and the dependent variable is $\log_e Y$.

$$\log_e Y = \log_e a + b.\log_e (1 - \exp(-k.X))$$

where the predictor variable is $\log_e (1 - \exp(-k.X))$ and the dependent variable is $\log_e Y$.

These three linearized forms can all be fitted using simple linear regression except for the k parameter. The value of k which gives the best fit can be estimated graphically.

The essential principal of the graphical approach is to calculate the residual sum of squares for the linear forms, using the conventional methods described in section 2.1, for a series of trial values of k . These residual sums of squares are plotted graphically

against the trial values of k and a smooth curve is drawn through the points. The value of k at which a minimum sum of squares is observed to occur provides the best estimate of k ; the regression is recalculated using this value of k to give the corresponding best estimates of the a and b parameters.

For example, suppose we have the data relating plot age and height shown on figure A.4.1 to which we wish to fit the model (3):

$$Y = a(1 - \exp(-k.X))^b$$

where plot height is Y and age is X . We use the linear form:

$$\log Y = \log a + b.\log(1 - \exp(-k.X))$$

and estimate $\log a$ as the intercept and b as the slope in a simple linear regression of $\log Y$ on $\log(1 - \exp(-k.X))$. In this regression we are mainly concerned at this stage with the residuals or deviations, sum of squares, symbolised as DSS. This can be obtained as shown on Form A.2 in section 2.1. With the data from figure A.4.1 and using the trial values of k shown below we obtain the deviation sums of squares:

<u>k</u>	<u>Deviation Sums of Squares</u>
0.05	0.2802
0.15	0.0302
0.25	0.0732
0.35	0.2657
0.45	0.4885

The best fit (minimum DSS) appears to be near a k value of 0.15. Taking some additional trial values we have:

<u>k</u>	<u>DSS</u>
0.11	0.0916
0.13	0.0543
0.17	0.0187
0.19	0.0187
0.21	0.0287
0.23	0.0474

Plotting these values graphically (figure A.4.2) we can see that the minimum appears to be close to $k = 0.18$. When the coefficient values are calculated for this value of k , we find:

$$\log_e a = 3.290 \quad (k = 0.18 \quad a = 26.83) \quad b = 5.199$$

and hence our resultant equation is:

$$H = 26.83(1 - \exp(-0.18.A))^{5.199}$$

which is plotted on figure A.4.1 as a broken line.

Figure A.4.1 Hypothetical height-age data for a plot, together with a model (broken line) fitted by the nonlinear regression method described in the text

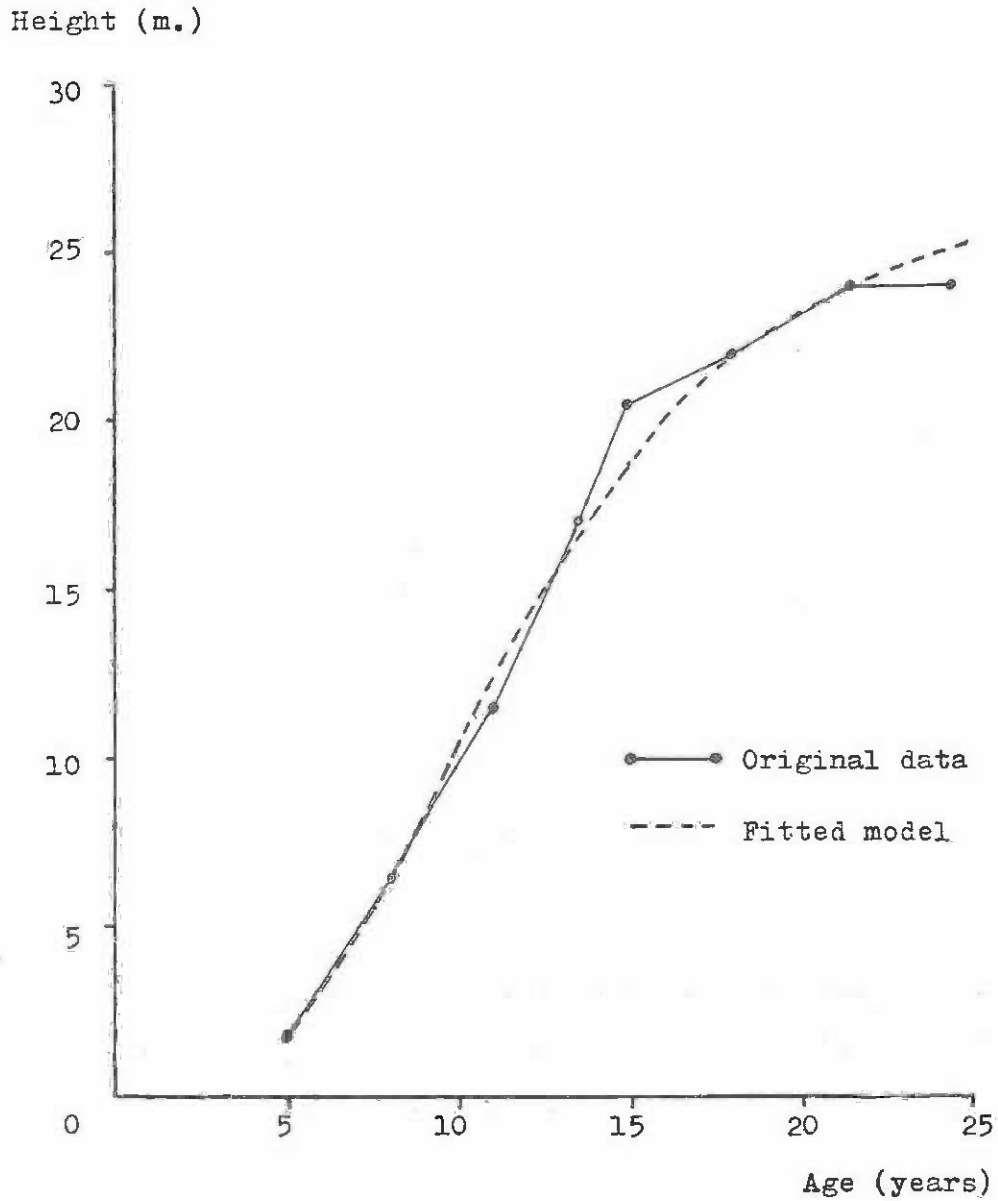
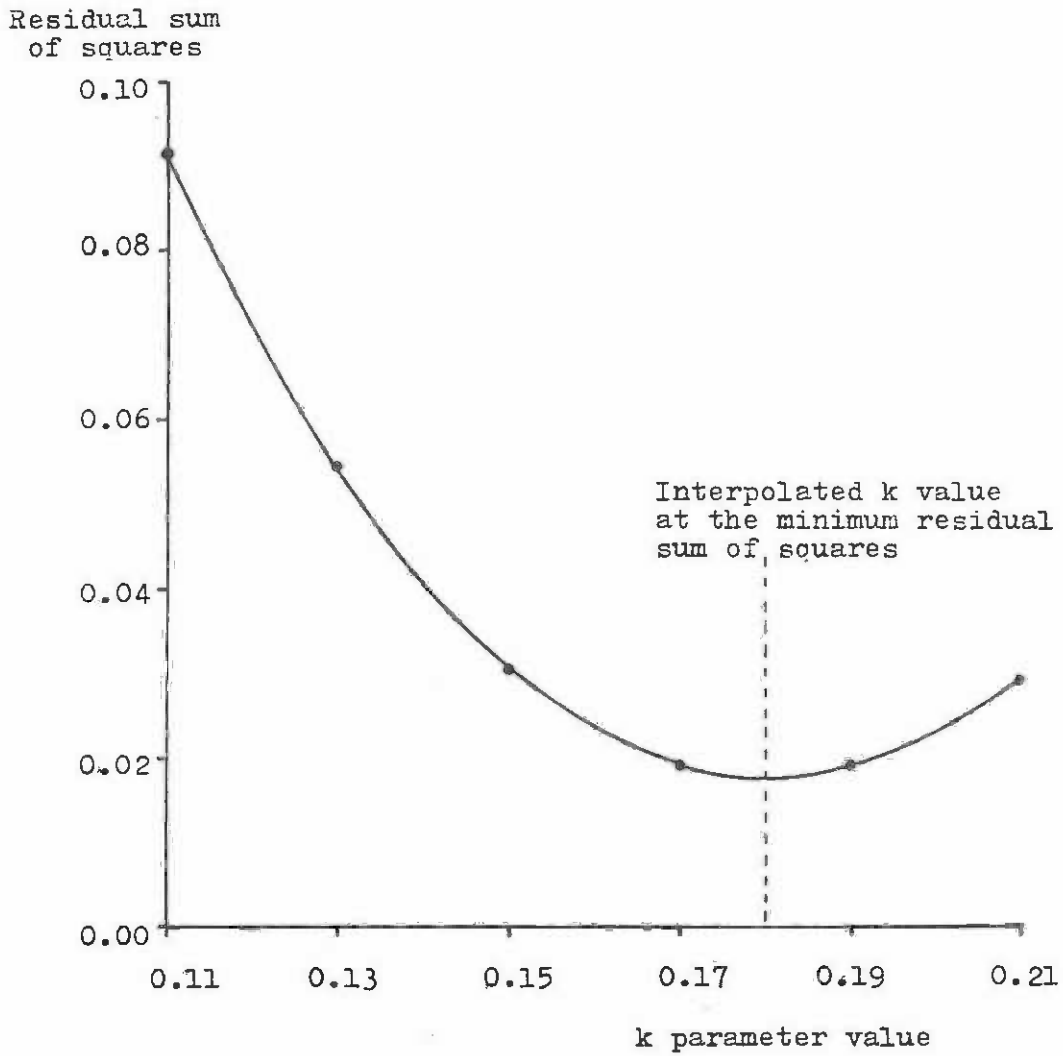


Figure A.4.2 Graphical determination of k parameter value which minimizes residual sums of squares in a nonlinear regression



This approach can be easily extended in two ways, although a detailed discussion lies outside the scope of this manual.

- (i) Any model in which only one of the parameters to be fitted is nonlinear can be fitted in an analogous manner to that described above. For example, we could have:

$$Y = b_0 + b_1 X_1^k + b_2 X_2 + b_3 X_1^k X_2$$

and fit the b_i using multiple linear regression with a series of trial values of k ; plot the resultant residual sums of squares for each trial; estimate the resultant k at the minimum; and then fit the b_i using that value of k .

- (ii) The search for the value of k which minimizes the residual sum of squares can be automated so that all the parameters including the nonlinear parameter can be calculated by a computer programme. This is quite easy to do with a single nonlinear parameter with the resultant programmes being compact enough even for some programmable calculators. Suitable minimization methods are suggested in Sadler (1975).

The above methods are designed for the worker who does not have access to a large or medium sized computer. Given good computing facilities, the best and most flexible approach to nonlinear curve fitting is by one of the modifications of the so-called Gauss-Newton procedure, as discussed in such standard textbooks as Draper & Smith (1966). These methods are quite difficult to understand, although the ultimate result, namely a set of fitted coefficients, may be easy enough to use. Consequently, most workers will need access to a computer programme that can be adapted to their machine. There are a number of such programmes available, including for example, SNIFTA (Small Nonlinear Fitting Algorithm) which is written in ANSI 66 FORTRAN and is available from:

Biometrics Section
Unit of Tropical Silviculture
Commonwealth Forestry Institute
South Parks Road, Oxford OX1 3RB
United Kingdom.

Appendix B

TABLES OF COMMON TRANSFORMATIONS USED FOR
REGRESSION MODELS DESCRIBED IN THE TEXT

Table 1 Natural logarithms (base e) for values 1 to 100

X	ln X	X	ln X	X	ln X
1	0.00000	36	3.58351	71	4.26267
2	0.69314	37	3.61091	72	4.27666
3	1.09861	38	3.63758	73	4.29045
4	1.38629	39	3.66356	74	4.30406
5	1.60943	40	3.68887	75	4.31748
6	1.79175	41	3.71357	76	4.33073
7	1.94591	42	3.73766	77	4.34380
8	2.07944	43	3.76120	78	4.35670
9	2.19722	44	3.78418	79	4.36944
10	2.30258	45	3.80666	80	4.38202
11	2.39789	46	3.82864	81	4.39444
12	2.48490	47	3.85014	82	4.40671
13	2.56494	48	3.87120	83	4.41884
14	2.63905	49	3.89182	84	4.43081
15	2.70805	50	3.91202	85	4.44265
16	2.77258	51	3.93182	86	4.45434
17	2.83321	52	3.95124	87	4.46590
18	2.89037	53	3.97029	88	4.47733
19	2.94443	54	3.98898	89	4.48863
20	2.99573	55	4.00733	90	4.49980
21	3.04452	56	4.02535	91	4.51085
22	3.09104	57	4.04305	92	4.52178
23	3.13549	58	4.06044	93	4.53259
24	3.17805	59	4.07753	94	4.54329
25	3.21887	60	4.09434	95	4.55387
26	3.25809	61	4.11087	96	4.56434
27	3.29583	62	4.12713	97	4.57471
28	3.33220	63	4.14313	98	4.58496
29	3.36729	64	4.15888	99	4.59511
30	3.40119	65	4.17438	100	4.60517
31	3.43398	66	4.18965	101	4.61512
32	3.46573	67	4.20469	102	4.62497
33	3.49650	68	4.21950	103	4.63472
34	3.52636	69	4.23410	104	4.64439
35	3.55534	70	4.24849	105	4.65396

Table 2 Exponential function exp(X) for values 0 to 5

X	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	1.00	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09
0.1	1.10	1.11	1.12	1.13	1.15	1.16	1.17	1.18	1.19	1.20
0.2	1.22	1.23	1.24	1.25	1.27	1.28	1.29	1.30	1.32	1.33
0.3	1.34	1.36	1.37	1.39	1.40	1.41	1.43	1.44	1.46	1.47
0.4	1.49	1.50	1.52	1.53	1.55	1.56	1.58	1.59	1.61	1.63
0.5	1.64	1.66	1.68	1.69	1.71	1.73	1.75	1.76	1.78	1.80
0.6	1.82	1.84	1.85	1.87	1.89	1.91	1.93	1.95	1.97	1.99
0.7	2.01	2.03	2.05	2.07	2.09	2.11	2.13	2.15	2.18	2.20
0.8	2.22	2.24	2.27	2.29	2.31	2.33	2.36	2.38	2.41	2.43
0.9	2.45	2.48	2.50	2.53	2.55	2.58	2.61	2.63	2.66	2.69
1.0	2.71	2.74	2.77	2.80	2.82	2.85	2.88	2.91	2.94	2.97
1.1	3.00	3.03	3.06	3.09	3.12	3.15	3.18	3.22	3.25	3.28
1.2	3.32	3.35	3.38	3.42	3.45	3.49	3.52	3.56	3.59	3.63
1.3	3.66	3.70	3.74	3.78	3.81	3.85	3.89	3.93	3.97	4.01
1.4	4.05	4.09	4.13	4.17	4.22	4.26	4.30	4.34	4.39	4.43
1.5	4.48	4.52	4.57	4.61	4.66	4.71	4.75	4.80	4.85	4.90
1.6	4.95	5.00	5.05	5.10	5.15	5.20	5.25	5.31	5.36	5.41
1.7	5.47	5.52	5.58	5.64	5.69	5.75	5.81	5.87	5.92	5.98
1.8	6.04	6.11	6.17	6.23	6.29	6.35	6.42	6.48	6.55	6.61
1.9	6.68	6.75	6.82	6.88	6.95	7.02	7.09	7.17	7.24	7.31
2.0	7.38	7.46	7.53	7.61	7.69	7.76	7.84	7.92	8.00	8.08
2.1	8.16	8.24	8.33	8.41	8.49	8.58	8.67	8.75	8.84	8.93
2.2	9.02	9.11	9.20	9.29	9.39	9.48	9.58	9.67	9.77	9.87
2.3	9.97	10.07	10.17	10.27	10.38	10.48	10.59	10.69	10.80	10.91
2.4	11.02	11.13	11.24	11.35	11.47	11.58	11.70	11.82	11.94	12.06
2.5	12.18	12.30	12.42	12.55	12.67	12.80	12.93	13.06	13.19	13.32
2.6	13.46	13.59	13.73	13.87	14.01	14.15	14.29	14.43	14.58	14.73
2.7	14.87	15.02	15.18	15.33	15.48	15.64	15.79	15.95	16.11	16.28
2.8	16.44	16.60	16.77	16.94	17.11	17.28	17.46	17.63	17.81	17.99
2.9	18.17	18.35	18.54	18.72	18.91	19.10	19.29	19.49	19.68	19.88
3.0	20.08	20.28	20.49	20.69	20.90	21.11	21.32	21.54	21.75	21.97
3.1	22.19	22.42	22.64	22.87	23.10	23.33	23.57	23.80	24.04	24.28
3.2	24.53	24.77	25.02	25.27	25.53	25.79	26.04	26.31	26.57	26.84
3.3	27.11	27.38	27.66	27.93	28.21	28.50	28.78	29.07	29.37	29.66
3.4	29.96	30.26	30.56	30.87	31.18	31.50	31.81	32.13	32.45	32.78
3.5	33.11	33.44	33.78	34.12	34.46	34.81	35.16	35.51	35.87	36.23
3.6	36.59	36.96	37.33	37.71	38.09	38.47	38.86	39.25	39.64	40.04
3.7	40.44	40.85	41.26	41.67	42.09	42.52	42.94	43.38	43.81	44.25
3.8	44.70	45.15	45.60	46.06	46.52	46.99	47.46	47.94	48.42	48.91
3.9	49.40	49.89	50.40	50.90	51.41	51.93	52.45	52.98	53.51	54.05
4.0	54.59	55.14	55.70	56.26	56.82	57.39	57.97	58.55	59.14	59.73
4.1	60.34	60.94	61.55	62.17	62.80	63.43	64.07	64.71	65.36	66.02
4.2	66.68	67.35	68.03	68.71	69.40	70.10	70.80	71.52	72.24	72.96
4.3	73.69	74.44	75.18	75.94	76.70	77.47	78.25	79.04	79.83	80.64
4.4	81.45	82.26	83.09	83.93	84.77	85.62	86.48	87.35	88.23	89.12
4.5	90.01	90.92	91.83	92.75	93.69	94.63	95.58	96.54	97.51	98.49
4.6	99.48	100.48	101.49	102.51	103.54	104.58	105.63	106.69	107.77	108.85
4.7	109.94	111.05	112.16	113.29	114.43	115.58	116.74	117.91	119.10	120.30
4.8	121.51	122.73	123.96	125.21	126.46	127.74	129.02	130.32	131.63	132.95
4.9	134.28	135.63	137.00	138.37	139.77	141.17	142.59	144.02	145.47	146.93

Table 3 Powers of reciprocals for values 2 to 50

X	Value of Power									
	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
2	.87055	.75785	.65975	.57434	.50000	.43527	.37892	.32987	.28717	.25000
3	.80274	.64439	.51728	.41524	.33333	.26758	.21479	.17242	.13841	.11111
4	.75785	.57434	.43527	.32987	.25000	.18946	.14358	.10881	.08246	.06250
5	.72477	.52530	.38073	.27594	.20000	.14495	.10506	.07614	.05518	.04000
6	.69882	.48835	.34127	.23849	.16666	.11647	.08139	.05687	.03974	.02777
7	.67761	.45915	.31112	.21082	.14285	.09680	.06559	.04444	.03011	.02040
8	.65975	.43527	.28717	.18946	.12500	.08246	.05440	.03589	.02368	.01562
9	.64439	.41524	.26758	.17242	.11111	.07159	.04613	.02973	.01915	.01234
10	.63095	.39810	.25118	.15848	.10000	.06309	.03981	.02511	.01584	.01000
11	.61904	.38321	.23722	.14685	.09090	.05627	.03483	.02156	.01335	.00826
12	.60836	.37010	.22516	.13697	.08333	.05069	.03084	.01876	.01141	.00694
13	.59870	.35844	.21460	.12848	.07692	.04605	.02757	.01650	.00988	.00591
14	.58989	.34797	.20526	.12108	.07142	.04213	.02485	.01466	.00864	.00510
15	.58181	.33850	.19694	.11458	.06666	.03878	.02256	.01312	.00763	.00444
16	.57434	.32987	.18946	.10881	.06250	.03589	.02061	.01184	.00680	.00390
17	.56742	.32197	.18269	.10366	.05882	.03337	.01893	.01074	.00609	.00346
18	.56097	.31469	.17653	.09903	.05555	.03116	.01748	.00980	.00550	.00308
19	.55494	.30796	.17090	.09484	.05263	.02920	.01620	.00899	.00499	.00277
20	.54928	.30170	.16572	.09102	.05000	.02746	.01508	.00828	.00455	.00250
21	.54394	.29587	.16094	.08754	.04761	.02590	.01408	.00766	.00416	.00226
22	.53890	.29042	.15651	.08434	.04545	.02449	.01320	.00711	.00383	.00206
23	.53413	.28530	.15239	.08139	.04347	.02322	.01240	.00662	.00353	.00189
24	.52961	.28048	.14855	.07867	.04166	.02206	.01168	.00618	.00327	.00173
25	.52530	.27594	.14495	.07614	.04000	.02101	.01103	.00579	.00304	.00160
26	.52120	.27165	.14158	.07379	.03846	.02004	.01044	.00544	.00283	.00147
27	.51728	.26758	.13841	.07159	.03703	.01915	.00991	.00512	.00265	.00137
28	.51353	.26371	.13542	.06954	.03571	.01834	.00941	.00483	.00248	.00127
29	.50994	.26004	.13260	.06762	.03448	.01758	.00896	.00457	.00233	.00118
30	.50649	.25653	.12993	.06581	.03333	.01688	.00855	.00433	.00219	.00111
31	.50318	.25319	.12740	.06410	.03225	.01623	.00816	.00410	.00206	.00104
32	.50000	.25000	.12500	.06250	.03125	.01562	.00781	.00390	.00195	.00097
33	.49693	.24694	.12271	.06098	.03030	.01505	.00748	.00371	.00184	.00091
34	.49397	.24401	.12053	.05954	.02941	.01452	.00717	.00354	.00175	.00086
35	.49111	.24119	.11845	.05817	.02857	.01403	.00689	.00338	.00166	.00081
36	.48835	.23849	.11647	.05687	.02777	.01356	.00662	.00323	.00157	.00077
37	.48569	.23589	.11457	.05564	.02702	.01312	.00637	.00309	.00150	.00073
38	.48310	.23339	.11275	.05447	.02631	.01271	.00614	.00296	.00143	.00069
39	.48060	.23097	.11100	.05335	.02564	.01232	.00592	.00284	.00136	.00065
40	.47817	.22865	.10933	.05228	.02500	.01195	.00571	.00273	.00130	.00062
41	.47582	.22640	.10772	.05125	.02439	.01160	.00552	.00262	.00125	.00059
42	.47353	.22423	.10618	.05028	.02380	.01127	.00533	.00252	.00119	.00056
43	.47130	.22213	.10469	.04934	.02325	.01096	.00516	.00243	.00114	.00054
44	.46914	.22009	.10325	.04844	.02272	.01066	.00500	.00234	.00110	.00051
45	.46704	.21812	.10187	.04758	.02222	.01037	.00484	.00226	.00105	.00049
46	.46499	.21622	.10054	.04675	.02173	.01010	.00470	.00218	.00101	.00047
47	.46299	.21436	.09925	.04595	.02127	.00985	.00456	.00211	.00097	.00045
48	.46105	.21257	.09800	.04518	.02083	.00960	.00442	.00204	.00094	.00043
49	.45915	.21082	.09680	.04444	.02040	.00937	.00430	.00197	.00090	.00041
50	.45730	.20912	.09563	.04373	.02000	.00914	.00418	.00191	.00087	.00040

Table 4 Spacing and relative spacing % from height and stocking

Spacing	2.0	2.5	3.0	3.5	4.0	4.5	5.0	6.0	7.0	8.0 m.
N/ha.	2500	1600	1111	816	625	494	400	278	204	256
Height	----- Relative spacing % -----									
5	40	50	60	70	80	89	100	119	140	160
6	33	41	50	58	66	74	83	99	116	133
7	28	35	42	50	57	64	71	85	100	114
8	25	31	37	43	50	56	62	74	87	100
9	22	27	33	38	44	49	55	66	77	88
10	20	25	30	35	40	44	50	59	70	80
11	18	22	27	31	36	40	45	54	63	72
12	16	20	25	29	33	37	41	49	58	66
13	15	19	23	26	30	34	38	46	53	61
14	14	17	21	25	28	32	35	42	50	57
15	13	16	20	23	26	29	33	39	46	53
16	12	15	18	21	25	28	31	37	43	50
17	11	14	17	20	23	26	29	35	41	47
18	11	13	16	19	22	24	27	33	38	44
19	10	13	15	18	21	23	26	31	36	42
20	10	12	15	17	20	22	25	29	35	40
21	9	11	14	16	19	21	23	28	33	38
22	9	11	13	15	18	20	22	27	31	36
23	8	10	13	15	17	19	21	26	30	34
24	8	10	12	14	16	18	20	24	29	33
25	8	10	12	14	16	17	20	23	28	32
26	7	9	11	13	15	17	19	23	26	30
27	7	9	11	12	14	16	18	22	25	29
28	7	8	10	12	14	16	17	21	25	28
29	6	8	10	12	13	15	17	20	24	27
30	6	8	10	11	13	14	16	19	23	26
31	6	8	9	11	12	14	16	19	22	25
32	6	7	9	10	12	14	15	18	21	25
33	6	7	9	10	12	13	15	18	21	24
34	5	7	8	10	11	13	14	17	20	23
35	5	7	8	10	11	12	14	17	20	22
36	5	6	8	9	11	12	13	16	19	22
37	5	6	8	9	10	12	13	16	18	21
38	5	6	7	9	10	11	13	15	18	21
39	5	6	7	8	10	11	12	15	17	20
40	5	6	7	8	10	11	12	14	17	20
41	4	6	7	8	9	10	12	14	17	19
42	4	5	7	8	9	10	11	14	16	19
43	4	5	6	8	9	10	11	13	16	18
44	4	5	6	7	9	10	11	13	15	18
45	4	5	6	7	8	9	11	13	15	17

Appendix C

BLANK COPIES OF CALCULATION AND
DATA RECORDING FORMS DESCRIBED IN THE MAIN TEXT

Form 3.1 Sample Plot Assessment Form - Plantations

Sheet ...of...

Forest	District	Compartment	Plot number	Species	Office use
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	
Plot area	<input type="text"/>	Slope	Assessment date	month	year

no. of cards

Tree no.	Diameter bhob	Height	Codes		
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Notes

Assessed by Date / /

USE BACK OF FORM FOR HEIGHT CALCULATIONS

Form 4.1 Common slope and common intercept regression models

Part 1 Plot data summarization. Use as many part 1 sheets as necessary for all plots.

Data transformations used: X = _____
Y = _____

Plot					
Raw data		Transformed data			
		X	Y	X ²	XY
Totals	n				

Plot					
Raw data		Transformed data			
		X	Y	X ²	XY
Totals	n				

Plot					
Raw data		Transformed data			
		X	Y	X ²	XY
Totals	n				

Form 4.1 Common slope and common intercept regression models

Part 2 Totals between plots and coefficient calculation

Plot	$\frac{\sum XY - \frac{\sum X \sum Y}{n}}{n}$	$\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n}$	$\frac{\sum X \sum XY}{\sum X^2}$	$\frac{(\sum X)^2}{\sum X^2}$	n	$\sum Y$
Totals	1	2	3	4	5	6

Common slope $b = (1)/(2) = \underline{\hspace{2cm}} = \underline{\hspace{2cm}}$

Common intercept $a = ((6)-(3))/((5)-(4)) = \underline{\hspace{2cm}}$
 $= \underline{\hspace{2cm}}$

Form 4.2 Transformations to age and height data from PSP's to fit k parameter in Schumacher equation by common slope estimator

i	j	A_{ij}	a	b	c	X_{ij}	H_{ij}	d	e	f	g	h	Y_{ij}
Plot	Point	Age	Sum	Diff	z/a	$\ln c$	Height	Sum	Diff	e/b	$d/2$	f/g	$\ln h$



Transfer
to form 4.1

Form A.1 Coefficients for an approximating quadratic or cubic curve

Objective To calculate the coefficients b_i in either of the equations:

or $Y = b_0 + b_1 \cdot X + b_2 \cdot X^2 + b_3 \cdot X^3$ (cubic curve)

$Y = b_0 + b_1 \cdot X + b_2 \cdot X^2$ (quadratic curve)

Items marked \diamond are not required for the quadratic curve calculations.

- (1) Tabulate 3 or 4 data points for the function and compute squares and cubes.

i	1	2	3	4
Y_i				\diamond
X_i				\diamond
X_i^2				\diamond
X_i^3	\diamond	\diamond	\diamond	\diamond

- (2) From table (1) calculate differences as shown:

i	$Y_{i+1} - Y_i$	$X_{i+1} - X_i$	$X_{i+1}^2 - X_i^2$	$X_{i+1}^3 - X_i^3$
1				\diamond
2				\diamond
3	\diamond	\diamond	\diamond	\diamond

- (3) From table (2) construct the terms shown in the table below, according to the following definitions:

i	u_i	v_i	w_i
1			\diamond
2			\diamond
3	\diamond	\diamond	\diamond

$$u_i = (Y_{i+1} - Y_i) / (X_{i+1} - X_i)$$

$$v_i = (X_{i+1}^2 - X_i^2) / (X_{i+1} - X_i)$$

$$w_i = (X_{i+1}^3 - X_i^3) / (X_{i+1} - X_i)$$

Form A.1 (continued....)

(4) From table (3) calculate the terms in the following table, from these definitions:

i	$v_{i+1} - v_i$	p_i	q_i
1			◊
2	◊	◊	◊

$$p_i = (u_{i+1} - u_i) / (v_{i+1} - v_i)$$

$$q_i = (w_{i+1} - w_i) / (v_{i+1} - v_i)$$

(5) Calculate the coefficients according to the following formulae:

<u>Coefficient</u>	<u>Cubic eqn.</u>	<u>Quadratic eqn.</u>
b_3	$(p_2 - p_1) / (q_2 - q_1)$	None
	= _____	
b_2	$p_1 - b_3 \cdot q_1$	p_1
	= _____	= _____
b_1	$u_1 - b_2 \cdot v_1 - b_3 \cdot w_1$	$u_1 - b_2 \cdot v_1$
	= _____	= _____
b_0	$Y_1 - b_1 \cdot X_1 - b_2 \cdot X_1^2 - b_3 \cdot X_1^3$	$Y_1 - b_1 \cdot X_1 - b_2 \cdot X_1^2$
	= _____	= _____

* * * * *

Form A.3 Coefficients and statistics of a regression with two predictor variables (part 1)

Description of problem: _____

<p>(1) Regression data</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 33%;">Y</th> <th style="width: 33%;">X</th> <th style="width: 33%;">Z</th> </tr> </thead> <tbody> <tr> <td style="height: 150px;"> </td> <td> </td> <td> </td> </tr> </tbody> </table>	Y	X	Z				<p>(4) Corrected products</p> $\Sigma y^2 = \Sigma Y^2 - (\Sigma Y)^2/n = \underline{\hspace{2cm}}$ $\Sigma xy = \Sigma XY - \Sigma X \cdot \Sigma Y/n = \underline{\hspace{2cm}}$ $\Sigma zy = \Sigma ZY - \Sigma Z \cdot \Sigma Y/n = \underline{\hspace{2cm}}$ $\Sigma x^2 = \Sigma X^2 - (\Sigma X)^2/n = \underline{\hspace{2cm}}$ $\Sigma xz = \Sigma XZ - \Sigma X \cdot \Sigma Z/n = \underline{\hspace{2cm}}$ $\Sigma z^2 = \Sigma Z^2 - (\Sigma Z)^2/n = \underline{\hspace{2cm}}$ <p>(5) Determinant & Gauss multipliers</p> $D = \Sigma x^2 \cdot \Sigma z^2 - (\Sigma xz)^2 = \underline{\hspace{2cm}}$ $g_1 = \Sigma z^2 / D = \underline{\hspace{2cm}}$ $g_2 = -\Sigma xz / D = \underline{\hspace{2cm}}$ $g_3 = \Sigma x^2 / D = \underline{\hspace{2cm}}$ <p>(6) Regression coefficients</p> $b_1 = g_1 \cdot \Sigma xy + g_2 \cdot \Sigma zy = \underline{\hspace{2cm}}$ $b_2 = g_2 \cdot \Sigma xy + g_3 \cdot \Sigma zy = \underline{\hspace{2cm}}$ $b_0 = \bar{Y} - b_1 \cdot \bar{X} - b_2 \cdot \bar{Z} = \underline{\hspace{2cm}}$ <p>(7) Regression sum of squares</p> $\Sigma \hat{y}^2 = b_1 \cdot \Sigma xy + b_2 \cdot \Sigma zy = \underline{\hspace{2cm}}$ <p>(8) Residual sum of squares</p> $\Sigma d^2 = \Sigma y^2 - \Sigma \hat{y}^2 = \underline{\hspace{2cm}}$ <p>(9) Residual degrees of freedom</p> $df = n - 3 = \underline{\hspace{2cm}}$
Y	X	Z					

(2) Totals and products				df = n - 3 =
ΣY^2	ΣY	ΣXY	ΣZY	
	n	ΣX	ΣZ	
		ΣX^2	ΣXZ	
			ΣZ^2	

(3) Means	
$\bar{Y} = \Sigma Y/n =$	$\underline{\hspace{2cm}}$
$\bar{X} = \Sigma X/n =$	$\underline{\hspace{2cm}}$
$\bar{Z} = \Sigma Z/n =$	$\underline{\hspace{2cm}}$

Part 2

Form A.3 (...Part 2)

(10) Correlations between variables
$r_{yx} = \Sigma xy / \sqrt{(\Sigma x^2 \cdot \Sigma y^2)} =$
$r_{yz} = \Sigma yz / \sqrt{(\Sigma z^2 \cdot \Sigma y^2)} =$
$r_{xz} = \Sigma xz / \sqrt{(\Sigma x^2 \cdot \Sigma z^2)} =$

(11) Multiple correlation coefficient R and R ²
$R^2 = \Sigma \hat{y} / \Sigma y =$
$R = \sqrt{R^2} =$

(12) Analysis of variance				
	sum of squares	d.f.	mean square	variance ratio
Regression	$\Sigma \hat{y}^2$	2		
Residual	Σd^2	df	= s ²	
Total	Σy^2	n-1		

(13) Residual standard deviation
$s = \sqrt{s^2} =$

(14) Standard error and t-value of coefficients		
Coefficient	b_1	b_2
Standard error s_b	$s \cdot \sqrt{g_1} =$	$s \cdot \sqrt{g_3} =$
$t = b/s_b$		

(15) Confidence intervals for predictions
For a series of m estimates of Y at a given level of X and Z, the standard error of the mean prediction is given by:
$s_y = s \cdot \sqrt{(1/m + 1/n + g_1 \cdot x^2 + 2 \cdot g_2 \cdot x \cdot z + g_3 \cdot z^2)}$
where $x = (X - \bar{X})$ and $z = (Z - \bar{Z})$. For a large sample, 1/m will be zero; for a single point it will be unity.

Appendix D

ANNOTATED BIBLIOGRAPHY

This bibliography is not a comprehensive bibliography of forest mensuration literature. It simply suggests a set of reference books to accompany this manual, which may be used to research or extend the techniques suggested.

Technical articles referred to in the text are given as footnotes and are not represented here.

Assmann, E., The Principles of Forest Yield Study. Pergamon Press. pp 506
1970

A very comprehensive treaty on the theory of forest growth, involving discussion of numerous examples. Most of the concepts are generalized from North European, even-aged stands, and are not necessarily valid in the tropics. Only static yield and allometric models are discussed.

Avery, T.E., Forest Measurements. McGraw-Hill. pp 290.
1967

Emphasises practical aspects of field work and primary analysis of data, as opposed to model construction. Oriented to North American practices and terminology, so not entirely suited for the tropics, but still a useful textbook.

Burley J. and Wood, P.J. (Editors), A Manual on Species and Provenance Research with Particular Reference to the Tropics. Tropical Forestry Paper 10, Commonwealth Forestry Institute. Oxford.
1976

A useful companion volume to the present work, covering practical aspects of research requirements for the exploratory phases of the introduction of exotic plantation species. Practical in orientation and simply written. Separate special appendix (by J.F. Hughes and R.A. Pluntre) covers wood quality studies.

Cab, Computerized Methods in Forest Planning and Forecasting. Annotated Bibliography FL4.
1977 Commonwealth Agricultural Bureau, UK.

Covers published world literature 1973-1976. Gives a full abstract on each reference. Most entries are directly relevant to the field of yield prediction and control.

Carron, L.T., An Outline of Forest Mensuration. Australian National University Press.
1968 pp 224.

Better oriented to yield and volume table construction in widely spaced plantations than American or European texts. Brief and clear summary of methods, but does not discuss problems of model fitting or construction.

Daniel, C. and Wood, F.S., Fitting Equations to Data. Wiley-Interscience. pp 342.
1971

Covers practical use of multiple regression for model fitting, including residual analysis, nested regression models, stepwise regression and non-linear fitting. Includes examples and documentation for a computer package, but no source listing.

Day, A.C., FORTRAN Techniques. Cambridge University Press. pp 96.
1972

A good supplement to any introductory FORTRAN course. Covers methods of drawing graphs, sorting, pointers, stacks, open subroutines, recursion and other techniques.

Draper, N. and Smith, H., Applied Regression Analysis. Wiley-Interscience. pp 407.
1966

Now almost a classical textbook on multiple and non-linear regression. Covers theory of simple regression, matrix algebra for multiple regression, residual analysis, stepwise regression, use of regression to analyse controlled experiments, and introductory non-linear regression. Elementary matrix algebra is required, together with differential calculus.

FAO, Manual of Forest Inventory with Special Reference to Mixed Tropical Forests. Food
1973 and Agriculture Organization, Rome. pp 200.

An essential reference for yield study work. Much discussion of design of forest sampling and appropriate formulae. Practical guidelines for executing a forest inventory and data recording.

Fries, J. (Editor), Growth Models for Tree and Stand Simulation. Research Note No. 30.
1974 Department of Forest Yield Research, Royal College of Forestry, Stockholm, Sweden,
pp 379. (Available from J. Fries, University of Agricultural Sciences, S-750 07
Uppsala, Sweden)

Collection of papers from IUFRO, Working Party 54.01-4 meetings. Many useful ideas.

Fries, J., Brukhart, H. and Max, T., (Editors), Growth Models for Long Term Forecasting of
1978 Timber Yields. -FWS-1-78, School of Forestry and Wildlife Resources, Virginia
Polytechnic Institute and State University, Blacksburg, VA 24061, USA. pp 249.

Collection of papers from IUFRO, Subject Group 54.01 meeting. Many useful ideas.

Green, P.E. and Carroll, J.D., Mathematical Tools for Applied Multivariate Analysis.
1976 Academic Press. pp 376.

A good self-teaching text for matrix algebra as it relates to multiple regression, principal component analysis, factor analysis and ordination analysis.

Husch, B., Miller, C.I. and Beers, T.W., Forest Mensuration.
1972

A 'classical' forest mensuration text, emphasising measurement rather than models, and with only very limited treatment of growth prediction. Oriented to North American conditions, so some practices are inappropriate for tropical use, while specifically tropical problems are not considered. A useful textbook of basic mensuration.

Land, A. and Powell, S., FORTRAN Codes for Mathematical Programming. Wiley. pp 249.
1973

Gives complete programmes and full documentation for linear, integer and quadratic programming using FORTRAN, with a brief explanation of underlying theory.

Parde, J., Dendrometrie.
1961

A classical text of forest mensuration, but with rather more emphasis on yield table construction than Husch *et al.* Very good on instrumental measurements, especially Relaskop, and on the development of simple models for the inter-relationship of stand variables. Probably the best current text book on forest mensuration. Not oriented to tropical problems. In French.

Pielou, E.C., Mathematical Ecology. Wiley-Interscience. pp 385.
1977

An excellent reference book for mathematical modelling of population dynamics, population dispersion and association analysis. Of more interest to the ecologist/silviculturist than to the forest manager. The association analysis section is important for studies in tropical rainforests.

Poole, R.W., An Introduction to Quantitative Ecology. McGraw-Hill. pp 532.
1974

Covers a wide range of techniques in ecology with an emphasis on model building. Less theoretical, and with more examples and more basic and introductory information than Pielou (see above). Assumes a little calculus and matrix algebra.

Prodan, M., Forest Biometrics. Pergamon Press. pp 447.
1968

Very useful treatment of the statistical background to forest growth studies. Requires moderate mathematical competence. Good treatment of fitting non-linear growth functions, but the book is not directly oriented to practical application. Mainly influenced by European forestry practices.

Royce-Sadler, D., Numerical Methods for Non-linear Regression. University of Queensland
1975 Press. pp 89.

A summary of the main practical methods for fitting the coefficients of non-linear models. Assumes some calculus and matrix algebra. Very clear and brief exposition of the essential algorithms.

Seber, G.A.F., Linear Regression Analysis. Wiley-Interscience. pp 465.
1977

More theoretical, more comprehensive and more up to date than Draper & Smith. Good understanding of matrix algebra required, but a very useful reference book for the statistician.

Shannon, R.E., Systems Simulation: The Art and Science. Prentice-Hall. pp 387.
1975

A useful reference book for workers interested in the construction of simulation models. Covers random number generation, model fitting, systems analysis, experiments on models, decision-making using models and gives several case studies.

Snedecor, G.W. and Cochran, W.G., Statistical Methods. Iowa State UP.
1967

Excellent reference book for statistical methods. Does not require an excessively high level of mathematical ability (calculus and matrix algebra not needed). Subjects covered include linear and non-linear regression, multiple regression, experimental and sampling design and analysis. All techniques are well illustrated with worked examples and there are numerous test questions on each section.

Stark, P.A., Introduction to Numerical Methods. Macmillan. pp 334.
1970

Basic textbook on numerical methods for computers. Includes solution of equations, simultaneous equations, numerical integration and differentiation, and interpolation with polynomials. Most methods are illustrated with short FORTRAN programmes. Also gives introduction to matrix algebra. Requires mathematical ability to about university entrance standard. Explains matrix algebra, but assumes knowledge of calculus.

Universal Encyclopaedia of Mathematics. Pan Books. pp 715.

An inexpensive (paperback) and extremely useful reference book for formulae in algebra, trigonometry, geometry, calculus, etc., mathematical tables and terminology.

Wagner, H.M., Principles of Management Science. Prentice-Hall. pp 612.
1975

An excellent textbook of quantitative decision-making methods for managers, including linear and mathematical programming of all types, inventory control, simulation and queueing models. Does not assume much mathematical ability (no calculus or matrix algebra).

Wright, H. (Editor), Planning, Performance and Evaluation of Growth and Yield Studies.
1980 Commonwealth Forestry Institute, Oxford, Great Britain (in print).

Collection of papers from IUFRO, Subject Group 54.01 meeting. Many useful ideas.