KFRI Research Report 65

STATISTICAL TECHNIQUES IN FORESTRY RESEARCH AND FORESTRY

K.Jayaraman
P.Rugmini

KERALA FOREST RESEARCH INSTITUTE
PEECHI, THRISSUR

April 1990 Pages: 46

CONTENTS

ABSTRACT

A comprehensive review and evaluation of the current status of research methods in forestry are made and some specific refinements suggested. The methods in general belong to three categories viz. experiments, surveys and simulations. While these comporients are the essential elements of any scientific research programme, they can be brought under a broader framework of systems analysis designed to arrive at optimal solutions through model building.

In the case of field experiments use of optimum plot size and shape, proper blocking, analysis of covsriance and repeated measurements over time have been found helpful in reducing the error variance. Incomplete block designs could be useful in certain specific cases but the possibility of missing values reduce their utility. Response surface or other optimal fractional factorial designs have limited use in forest field trials unless they are of higher order. Proposition of a multivariate selection index devised to identify superior treatments with balanced expression is an outcome of this study. The magnitude of error variance is usually lesser in the case of nursery trials, when compared to field trials. But this cannot be generalized, since high levels of variability could be observed in the case of a study on growth of bamboo seedlings in bags of different sizes. This indicates the need for choosing appropriate error reducing techniques.

Sample surveys are taken up for a multitude of purposes in forestry like surveys on resources, pests, diseases, industries and wildlife. A review of the current state of affairs indicated the need to work out specific sampling strategies in each of these individual cases. Considering the time and effort involved in ground surveys, remote sensing would be a valid alternative in contexts where it is practicable. A certain extent of redundancy can always be expected within a multivariate system observed through a survey, and this feature can be exploited by developing prediction equations to estimate properties which are hard to measure from easily measurable characteristics.

Simulation techniques are recommended where experiments with real life systems are infeasible. Experimenting on the state of a system with a model over time is termed simulation. Problems like forecasting demand for forest products, projecting future condition of forests under different management regimes etc. are handled by this system of investigation.

The present report also provides a bird's eye-view of the diverse applications of statistical techniques in some specific areas in forestry. Apart from these, statistical expert system was developed for some common situations in forestry research. The system is available as a userfriendly software package and is helpful in choosing an experimental design, a sampling method or any analytical technique suited to the user's specific requirements.

1. INTRODUCTION

It is well known that not only the subject of research but also the method of research is equally important. One is likely to get erroneous results and make faulty conclusions in the absence of proper scientific and objective methods to be employed for the conduct of research.

This project was taken up to evaluate methods of research currently in vogue in forestry and seek possible means of improvement. The nature of forest research and the methodological issues involved are discussed below. This is not intended to be an exposition of elementary statistical methods available in numerous text books on the subject but rather a serious discussion of the nature and problems of forest research methodology. However, both foresters and researchers would be benefited by getting to know the nature of some of their own research efforts. A minimum level of familiarity with statistical methods from the part of readers is assumed. More than suggesting some specific refinements the report examines the scope of the ,discipline of forest biometrics and brings out several problems which merit individual attention in future works. The techniques are discussed under two sections viz. general methods of research in forestry and those of specific interdisciplinary areas. Additionally, in order to make this work practically helpful to the scientists who use statistical methods in their research an attempt was made to develop a statistical expert system covering some of the common situations in forestry research. The details are available in the concerned section.

2. GENERAL METHODS OF RESEARCH IN FORESTRY

Forestry is a synthetic science drawing heavily from physical and biological sciences and entails both basic and applied research. Like in any other science the research methods are in general based on the inductive - deductive approach and involves formulation of hypotheses from observed facts followed by deductions and verification repeated in a cyclical process. Statistical methods are employed for objective verification of hypotheses and to deal with the uncertainty involved in making generalizations. The two major practical aspects of the whole process are the collection of data and interpretation of the collected data. The data may be generated through a designed experiment on a hypothetical population or through a sample survey on a naturally existing population. The collected data are condensed and useful information extracted through techniques of statistical inference. This apart, a method of considerable importance to forestry which has gained wider acceptance in recent times with the advent of computers is simulation. This is particularly useful in forestry because simulation techniques to a great extent can replace large scale field experiments which are extremely costly and time consuming. Mathematical models are developed which capture most of the relevant features of the system under consideration after which experiments on system behaviour can be conducted in computer rather than with real life systems. The above topics are further elaborated below.

2.1. *Experiments*

Experiments serve to test hypotheses under controlled conditions. The basic principles of experimentation are randomization, replication and local control which are the prerequisites for obtaining a valid estimate of error and for reducing its magnitude. The practical considerations in forestry experiments differ with respect to their contexts.

2.1.1. *Field experiments*

These are large scale experiments conducted in the forests and plantations and may address evaluation of many of the management alternatives. Provenance trials, spacing trials, fertilizer trials, regeneration trials, etc. are typical examples. They may be simple experiments involving comparison of a few preselected treatments or complex factorial experiments designed to investigate the responses to combinations of several factors at different levels.

Field experiments are characterized by large error variance caused by a free play of many non-experimental factors and this should be controlled by adopting appropriate field plot techniques. For instance, a study on variability in a 12-year-old teak plantation at Pothuchady showed that coefficient of variation of basal area among individual tree plots was of the order of 96 percent which reduced to 23 percent with a plot size of 36 trees and stabilized thereafter. The point of maximum curvature worked out to be 12 trees which, with the spacing existed in the particular plantation occupied around 100m^2. This shall be applicable for the age group just considered. Optimum plot size and shape of plots and blocks have a role in reducing error variance but quite often these are governed by the nature of treatment and availability of material. Use of guard rows or alleys between plots makes the observations independent, one of the basic requirements in statistical analysis. Analysis of covariance helps in bringing down the error variance as reported by Nair *et al* (1985) in a study on the effect of insect defoliation on volume increments of teak trees. Subsampling within the plots will not help in reducing the error variance though it may sometimes produce an independent estimate of error. But most often it becomes a must when either plot size is large or when the individual responses are difficult to measure. Repeated measurements over time on the same experimental material usually bring down the error variance in the pooled analysis. For instance, in a study on the effect of slash burning on height growth of planted teak (Chacko *et al.*, 1989). the coefficient of variation based on error mean square at the individual measurement period fluctuated around 3.69 percent. A pooled analysis combining the data from nine measurement periods resulted in a coefficient of variation of 2 percent in the log scale.

The traditional design for most of the field experiments in forestry has been randomized complete block design (RCBD). The concept of blocking is to divide the total experimental area into homogeneous groups of plots so that differences among blocks are eliminated in the course of analysis thereby achieving greater precision for treatment comparison. For example Nair *et al.* (1985) reported significant reduction in the error variance due to blocking in a field experiment with teak. But many of the forest field trials require large plot sizes and the blocks become heterogeneous with a fairly large number of treatments. Confounded or incomplete block designs would be valid alternatives but are not as flexible as RCBD with respect to missing value estimation. Highly efficient designs like the class of optimal designs have been proposed in many other fields. But their use is indicated only when the number of experimental units is extremely restricted and higher order interactions are negligible. Response surface designs are suggested when the input factor is quantitative and the object of the experiment is to explore input-response relationship. But in certain cases, higher order designs may be required than the usual second order designs. Yang (1983) found that the quadratic equation specified by the central composite designs was inadequate in approximating the response surface of whiie spruce seedlings to fertilization with N, P and K. Conventional factorial experiments were found useful for testing hypotheses of various effects and also in deriving an equation that adequately approximates the true response surface. Experiments with mixtures like mixed plantations or naturally occurring mixed species stands require special designs like simplex lattice, simplex centroid, etc. (Cornell, 1981). The objectives in such cases will mostly centre around studying the effect of species composition on the stand performance.

Most of the observations coming out of field experiments are quantitative and so are amenable to analysis of variance after due considerations to the assumptions involved. Normality and homoscedasticity are quite often assumed which is not a desirable trend. In cases where these assumptions are not satisfied, appropriate corrective measures like data transformations are to be adopted if gross errors in conclusions are to be avoided. When measurements are repeated over time with the treatment structure intact, an appropriate model for analysis will be that of a standard split plot design (Gomez and Gomez, 1984). In such cases the treatments are taken as main plots and the different stages of observations as subplots. Alternately, if a rate measure is obtained for the changes over time for each replicate value, they could be subjected to a standard analysis of variance dictated by the design concerned. It is a good practice to repeat the experiments in different places to study treatment - environment interaction.

Analysis of variance only reveals the overall differences among the treatments, if any. This is followed by pairwise comparison of the treatments (when qualitative) for identification of the best set of treatments with respect to any particular character. A large number of multiple comparison tests are available in this respect and these vary mainly with respect to their degree of conservation in declaring a specified difference as significant (Federer, 1967). However, when multiple characters are involved, selection of the 'best' treatment poses problems especially when treatments have different ranking for each character. This problem arose in one of the studies on induction. of rooting through application of hormones (Surendran and Seethalakshmi, 1987) wherein the objective was to select the treatments with higher character expression avoiding imbalance in their structure. Factor analysis failed to identify a single factor associated with such a rooting response. An index of the form

$$\text{INDEX} = \text{SUM}\ (z_i) - \text{SD}\ (z)$$

where z_i's are standard normal deviates for each character and SD is the standard deviation of such scores, effectively summarized the notion of 'best' in this respect and helped to identify such treatments, Additional advantage accrued is the information gained on the action of treatments in producing such response through a single ANOVA on the index values.

One of the most common features of data from field trials is the occurrence of missing values. These are traceable to damages caused by elephants, wind, fire or other natural calamities. Missing values destroy many of the nice features of the designs like balance and also make the analysis complicated. Missing value estimation techniques are available for most of the standard designs when the number of missing observations is few. But a complete least square analysis would be warranted with thoroughly nonorthogonal data with unequal subclass numbers.

Large error variance, possible occurrence of missing values and long time span involved make the field experiments in forestry unattractive and restrict their scope to treatments showing large differences and where no quick answers are anticipated. It is unlikely that the effects of these are nullified by refinements in design or analysis. Also, there are questions on which experiments are not feasible in practice like finding the effects of varying levels of fire on long term yield pattern of forests, etc.

2.1 .2, *Nursery experiments*

Experiments conducted on seedlings in nursery beds, polythene bags, concrete pots or any such media can be referred as nursery trials. The important point is that experimental material can be made relatively homogeneous and several extraneous factors affecting the response can be regulated. For instance, seedlings of uniform size can be selected, the soil mixture made more homogeneous and other management practices can be standardized. Consequently, simple designs like completely randomized design or randomized complete block design are sufficient for such studies. In spite of the uniform conditions provided, high levels of variability can be observed in some nursery trials as in the case of a study on growth of bamboo seedlings in bags of different sizes (Chacko and Jayaraman, 1988). The coefficient of variation varied from 15 to 64 percent over different characters and treatments. The seedlings being grown under identical conditions, this variability might partly be due to genetic differences. In consequence, the experimenter may fail to detect fairly large treatment differences unless enough number of replications are taken or blocking done based on size of seedlings in such experiments. The required number of replications, size of the experimental unit, frequency of observations etc., will need to be worked out for individual species. Most of the other things stated under section 2.1.I apply here as well. White (1984) gives a detailed account of some of the statistical and practical considerations to be made while planning a nursery experiment.

2.1.3. *Laboratory experiments*

These are investigations carried out under well controlled conditions of the laboratory. Many typical examples can be found in the field of entomology and pathology. Simple designs are good enough but nonexperimental error should be taken care of by repeated sampling No additional design or analytical considerations exist other than those mentioned under sections 2.1 .I and 2.1.2.

2.2. Sample surveys

In a broad sense all *in situ* studies involving noninterfering observations on nature can be called field surveys. These may be undertaken for a variety of purposes like estimation of population parameters, comparison of different populations, finding interrelations of variables, studying movement pattern or distribution pattern of organisms etc. Observed relationships from such studies are many times not causative but will have predictive value. Studies in sciences such as economics, ecology and wildlife generally belong to this category. An excellent account of the earlier development of applications of different sampling methods in forestry has been provided by Chacko (1965).

Some of the methodological issues of sample surveys in forestry are discussed below under the different contexts they arise.

2.2.1. *Resource surveys*

Most of the resource surveys start with a stratification of the population into forest tvpes or density classes. Compartments of convenient sizes are then identified and systematic strip or line plot sampling is undertaken within each compartment. Systematic sampling is easier to execute when compared to random sampling patterns. Systematic sampling has got some theoretical disadvantages, the major one being that a precise estimate of variance is not obtainable unless there are at least two random starts. Approximate variance estimates are obtainable through poststratification or by assuming that the samples are drawn at random. But considering the difficulties involved in executing a random sampling pattern in the field, systematic sampling is still considered a valid alternative. Sampling intensity varies from 1 to 5 percent depending on the size and nature of the population.

Presently, ground surveys are increasingly getting replaced by remote sensing techniques and this trend is justifiable when we consider the large effort and time involved in traditional surveys. Sometimes variations in the sampling pattern will be required depending upon the type of resource one is looking for. For instance, sampling method for bamboo and reeds need not be the same as that for canes.

A special type of sampling known as point sampling or variable plot sampling has been in vogue in timber surveys for quite some time though its application is mostly restricted to plantations. This is essentially a probability proportional to size (PPS) sampling scheme. The trees are selected with probability proportional to their basal areas and distances from a fixed sampling point. Sampling is done with a wedge prism or relascope and estimates of stand density are obtainable readily. This is a significant improvement over traditional sampling since actual measurement of the diameters can be avoided for getting some quick estimates of stand density.

Successive inventories are required in forestry for assessing the effectiveness of management and silvicultural practices and for studying the subtle changes over time in a forest ecosystem. Permanent sample plots to some extent serve the above purposes. But, of late, more efficient techniques have been proposed which involve sampling with partial replacement (SPR). The applicability of this method is commonly limited to sampling in two successive inventories keeping a few plots common. But estimates can be developed for all the plots in both the instances by employing regression techniques. This procedure has been found better than taking fresh set of plots every time or observing the same set of plots both the times. Cunia and Chevrou (1969) extended the theory of SPR from two to three or more occasions and showed how it could be applied to continuous forest inventory (CFI). Since then a number of works followed these developments covering different aspects of the method. One notable extension is the one achieved by Cunia and Kyaw (1985) which deals with a transformation of the sample values to make SPR formulae applicable to varying size, shape and structure of sampling units in successive inventories.

2.2.2. *Pest and disease surveys*

Plantations are the domain of most of the economically important pests and diseases. These surveys are comparatively easier and individual plantations. can be conveniently taken as first stage units within a Range or Division. Within a plantation, linear or two dimensional plots can be laid out for estimating the percentage infestation among the trees which is ‘equivalent to a cluster sampling scheme. As a specific case, the sampling intensity required in terms of number of plantations in the case of a survey on the incidence of teak borer in different forest Divisions was found to vary from 5 to 50 percent depending upon the variability in the infestation level existed in these geographic units (Mathew, 1989).

2.2.3. *Socio-economic and industrial surveys*

Stratified multistage sampling plans are especially suitable for socio economic and industrial surveys. The units can be conveniently stratified by size or income levels or even by administrative regions. The advantage is that the sampling frame need be prepared only for the selected subsampling units in a multistage sampling plan. When the sampling units largely differ in their sizes, adjusting the selection probabilities in proportion to the size may result in better estimates.

2.2.4. *Wildlife surveys*

One area where standard methods of sampling fail or are ineffective is in the case of species abundance studies on wildlife. As the population is mostly on the move and sometimes hidden, the observers have to depend on indirect evidences like dung counts or pug marks and so on. Methods vary with the type of animal. Capture release or aerial count techniques may not be applicable in the Indian context for many animals. Transect counts are feasible in places like sanctuaries where some permanent transects can be laid out for periodic observations. Otherwise moving in straight lines in thick jungles is not possible quite often.

Sale and Berkmuller (1988) give some general guidelines to be followed in wildlife surveys. Species occurring in relatively high densities, in habitat in which they are visible when searched for, can be counted by direct sighting methods. This applies to most populations of medium to large ungulates, rhinos and elephants. Species occurring in very low densities, or which are difficult to see because of poor habitat visibility or cryptic behaviour, should be censused either by carefully planned intensive samples, or by indirect methods such as dung or pug mark counts. This applies to most carnivores and small or nocturnal mammals, as well as to some large mammal populations in particularly dense habitat types. Most of the indirect methods are only suitable for obtaining relative indices of population size and only rarely yield a good estimate of actual population numbers.

A method found applicable in the case of small animals like Nilgiri Tahr is the bounded count method (Anon, 1984). The method is based on repeated direct sightings and estimates are developed nonparametrically. Estimation of wildlife abundance is an area where considerable progress is to be achieved but unfortunately there are no general methods available. Specific techniques will need to be developed for individual species considering their individual nature.

Field surveys can in some sense be viewed as observations made on nature's experiments. Nature intrinsically exhibits variation and this feature can be capitalized to gain useful hints on the system behaviour under changing conditions. Many a time, this is the only method available in certain situations.

2.3. *Modelling and simulation*

A mathematical description of a real world system is often referred to as a mathematical model. A system can be formally defined as a set of elements also called components. A set of trees in a forest stand, producers and consumers in an economic system are examples of components. The elements (components) have certain characteristics or attributes and these attributes have numerical or logical values. Among the elements, relationships exist and consequently the elements are interacting. The state of a system is determined by the numerical or logical values of the attributes of the system elements. Experimenting on the state of a system with a model over time is termed simulation (Kleijnen, 1974).

Simulation has multifarious applications in forestry. Some of these applications in specific contexts are discussed below.

2.3.1. *Projection of stand structure through growth models*

Scientific forest management relies to a large measure on the predictions of the future conditions of individual stands. This is achieved by predicting the increment from the current stand structure and updating the current values at each cycle of iteration using a growth model. The structural changes over time can be monitored under different cutting cycles and cutting intensities and optimal management policies can be arrived at based on the results of such simulation runs

Jayaraman and Bailey (1988) proposed a growth model useful for simulating the changes occurring in an uneven aged mixed species stand. The mean annual increment in basal area and number of trees is predicted from the current values of basal area, number of trees, site quality and species composition of the stand and the simulation proceeds by progressive updating of the values of predictor variables in annual cycles. Changes in site quality are carried forward through a linear difference equation. Volume estimates at each time point can be obtained by an appropriate height-diameter relation and a volume table function.

2 3.2. *Forecasting future demand and supply of forest products*

Kumar (1985) reviews the different supply and demand models available in forestry and suggests a new model for a small wood producing country. The model essentially consists of a supply equation, an export function, a home demand equation and ar: identity on the inventories. Functional forms for the equations will have to be deteimined by empirical verification. Parameters can be estimated if data are available on a lengthy time series basis after converting the model to its reduced form. The reduced form expresses each current exogeneous variable as a function of exogeneous and lagged endogeneous variables. Deterministic simulation can then be undertaken by tracing the time path of endogeneous variables by specifying initial values for exogeneous and lagged endogeneous variables.

2.3 3. *Simulating the effect of fire on long term timber supply*

Wagner (1983) describes a simulation model which can determine the ultimate long-term limit of the annual allowable cut in a forest exposed to a given annual amount of fire. The model's programme carries out the following procedure: (i) It operates an artificial forest with 1000 units of equal but unspecified area. Each stand carries a permanent number label, (ii) It grows the forest according to a specified curve of volume over age, (iii) Each year it burns a specified number of stands selected at random without regard to age, (iv) Later in the same year it cuts a specified number of stands, always selecting the one of highest standing volume, (v) Immediately after burning or cutting, stand age is reset to zero and growth begins again the following year, (vi) Account is kept of ages and volumes at time of fire or cutting, (vii) After each year's operation one year is added to the age array and the process repeats, (viii) After each successive 100 year period the varlous outputs are compiled, averaged and prinred out and (ix) The model's initial state is a rectangular age-class distribution in which 10 stands are alloted to each of ages 0-99. Though several variations on the above pattern are conceivable it illustrates the basic idea and the power of simulation techniques in contexts where other forms of experimentation are infeasible.

While surveys, experiments and simulations are the essential elements of any scientific research programme, they need to be embedded in some larger and more strategic framework if the programme as a whole is to be both efficient and effective. Increasingly it has come to be recognized that systems analysis provides such a framework, designed to help decision makers to choose a desirable course of action or to predict the outcome of one or more courses of action that seem desirable. A more forrnal definition of systems analysis is the orderly and logical organization of data and information into models followed by rigorous testing and exploration of these models necessary for their validation and improvement (Jeffers, 1978).

3. STATISTICAL APPLICATIONS IN SOME SPECIFIC AREAS IN FORESTRY

A number of areas within the realm of forestry can be found wherein substantial statistical applications have been made. The techniques are integrally related to the concepts in the particular subject fields and will require an understanding of both statistics and the concerned disciplines to fully appreciate their implications. Some of these fields are briefly covered in what follows. No attempt is made to improvise the techniques as these require concentrated efforts on single issues and are outside the purview of this project.

3.1. Genetical statistics

In classical or Mendelian genetics the focus of interest centered on the inheritance of qualitative characters and the statistical methods were those applicable to the nominal scale like binomial or chi-square test. Testing the agreement of observed frequency data with those expected by rhe hypothesis of Mendelian segregation was the major task and this included problems such as detection and estimation of linkage. In contrast, the methods of the quantitative genetics dealt with the study of continuous variation caused by action of polygenes, environmental factors and their interaction. Earlier attempts were mostly directed to the estimation of genotype value, genotypic variance and components such as additive and dominance effects. Estimation of more complicated genetic parameters like heritability, genetic correlation, repetability followed and several mating designs suggested for the purpose.

The simplest means of estimating the effects of genetic sources of variation in forestry is through half sibs. For example, if a set of female parents is chosen and a different set of male parents is chosen for each female, then the variation among offspring in different female parent groups is the same as the variation among half sibs. Variance components can be estimated through techniques of analysis of variance and estimates of parameters derivable from these components can be developed. A problem that still remains largely unresolved is the estimation of components of variance with unequal subclass numbers. Not only analysis of variance estimators such as above but also estimates based on regressions of offspring, clones, etc.. on parental performances are available.

Exploitation of geographic variation has continued to hold a prominent role in tree improvement. In provenance research, the emphasis is on undertaking practical experiments to discover regions or provenances capable of producing best trees. Sometimes, provenance trials offer a means to estimate genetic parameters like heritability and genetic correlation and to assess the magnitude of genotype environment interaction.

The next major development in this field was on evolving theories and methods of selection and working out their consequences on genetic parameters. The important parameters were genetic gain and correlated response. Selection of superior individuals based on multiple characters followed by extensive cloning is a quick method of bringing out improvements in individual performance. Where such propagation means are not available family selection followed by intermating among selected families to create superior gene pools becomes useful. Statistical techniques are available to predict the genetic gain under different selection methods with specified selection pressures and restrictive assumptions on the genetic structure of populations involved.

Artificial hybridization has been useful in a few tree species for which several mating designs have been devised. Both general and specific combining abilities are estimable through factorial, full and partial diallel mating systems and by application of analysis of variance techniques.

Yet another interface of statistics in genetics is population genetics which is concerned with studying genetic structure of populations and changes occurring to it over generations, The evolutionary changes occur in the population with forces such as mutation, migration and selection. There are also effects related to small population size and consequent inbreeding. The application of statistical methods have made it possible to trace the consequences of these effects in the populations through abstract models
koong, 1981).

3.2. Numerical taxonomy

Numerical taxonomy deals with grouping of taxonomic units into taxa by numerical methods on the basis of their character states. The term includes the drawing of phylogenetic inferences from the data by statistical or other mathematical methods to the extent to which this is possible (Sneath and Sokal, 1973). The major advantages of using numerical methods for classification are repeatability and objectivity. After having chosen the organisms and the characters to be recorded on them, further steps include working out the resemblances between organisms and constructing taxa based upon these resemblances. Generalizations are then made about taxa such as inference about their phylogeny, choice of discriminatory characters, etc.

Though the logical fundamental taxonomic units in a large majority of instances are individuals, a broader conception is that of operational taxonomic units (OTU) which are the lowest ranking taxa employed in a given study. Certain definitions apply to specifying characters 'as well, such as a unit character. A unit character is a taxonomic character of two or more states which within the study at hand cannot be subdivided logically, except for subdivision brought about by the method of coding. These may be morphological, physiological. behavioural or even ecological in nature.

The estimation of resemblance between pairs of OTU's starts from a matrix whose columns correspond to the different OTU's and the rows specify measurements of unit characters on each of these OTU's. Similarity between two OTU's is generally estimated by means of a similarity coefficient which is a quantification of the resemblance between the elements in the two columns of the data matrix representing the character states of the two OTU's in question. Similarity coefficients are mainly of four groups viz. distance coefficients, association coefficients, correlation coefficients and probabilistic similarity coefficients.

Application of the resemblance measures on the data matrix yields a resemblance matrix, the rows and columns of which refer to the OTU's and the entries are the estimates of the resemblances for every OTU compared with every other OTU. The determination of taxonomic structure is concerned with the recognition of patterns of distribution of OTU's and groups of OTU's (taxa) in a space commonly a hyperdimensional character or attribute space. The most frequently employed strategies for finding clusters, especially in biological material are sequential, agglomorative hierarchic and nonoverlapping in nature. The specific techniques of wide use in this regard are single, complete and average linkage clustering techniques. Certain ordination methods like principal co-ordinate analysis and multidimensional scaling have been found useful in reducing the dimension of the character space and thus to get a glimpse at the relative distribution of the OTU's. Certain graph theoretic concepts have also been applied to represent interrelation among OTU's through two or three dimensional networks. The results of cluster analysis are usually represented by phenograms or ordination plots. Phenons or taxa of any rank can be delineated from these phenograms. Methods of numerical taxonomy have been used in the study of phylogeny as well.

Apart from classification problems identification is another important facet of taxonomic studies. The problem of identification is that of placing an unknown OTU into one of the pre-established taxa based on the character set observed on the specimen at hand. Unlike in classification problems, characters may receive unequal weights in identification tasks. The main methods used in identification are keys and discriminant functions. Gyllenberg (1963) obtains dichotomous characters as the most useful discriminators. For large studies with well-separated taxa, sequential keys are the best. Simultaneous keys are useful for highly polythetic groups without much overlap but discriminant analysis is indicated where there are a few close groups in which identification must be as certain as possible.

Numerical taxonomic methods are also used in fields outside systematic biology, Major applications have been recorded in ecology, biogeography, social and earth sciences. Possible applications of numerical taxonomic methods abound in forestry. The objects of ciassification could be forest or soil types, insects, plants or other organisms related to forests or even wood specimens. Instances of quantitative works in this field have been very few in India.

3.3. Statistical ecology

Most of the mathematical and statistical applications in ecology deaf with the study of temporal and spatial patterns of populations of organisms. The former is often designated as population dynamics and the latter sometimes as statistical ecology (Pielou, 1977). Though population dynamics is largely of deterministic models, stochastic components are increasingly getting incorporated in these models.

Early population growth models dealt with birth and death processes and density dependent population growth of single species. The models were based on differential equations, the integration *of* which resulted in logistic growth curves. Models based on transition matrices featuring age dependent birth and death rates over discrete time points were also proposed during this time. Such matrix models though widely applied were not very successful in forestry. Recent evidences suggest that forest growth processes are not stationary (Binkley, 1980). Continuous time versions of age dependent growth rates were also available which lead to concepts of age/specific life tables. These models took further developments by getting extended to two species and later to the case of multiple species systems.

Considerable insight into the spatial pattern of organisms can be gained by examining the distribution of their numbers per sampling unit. For instance, a clumped spatial pattern is indicated by a negative binomial frequency distribution. But this information by itself may not give any hints on the causes of such spatial patterns because different hypotheses can lead to same type of frequency distribution. When the number of organisms per unit area has a negative binomial distribution, the index parameter k can be used as a measure of aggregation. its low values indicating pronounced clumping. An interesting property of k is that it remains unaltered when a population decreases in size due to random deaths. In more general cases, measures *of* aggregation that (a) does and (b) does not change when some of the population members are removed at random have been proposed by Lloyd (1967). These measures are 'index of mean crowding' and 'index of patchiness' and are expressible directly as functions of observed number of individuals per sampling unit. When the space available to the organisms is an extended continuum instead of being discrete and isolated. units, the aggregation measures are likely to be affected by the size of the sampling unit especially when the spatial pattern is clumped. Special methods devised to test the random mixing of sparse and dense grid cells in such cases are given in Pielou (1977).

A different sampling method for studying spatial pattern is distance sampling wherein the distance to the nearest neighbouring individual is measured from a randomly located point or individual. But this does not provide an estimate of density unless the pattern is random and so has to be used in conjunction with quadrat sampling to measure and test the degrees of randomness. Spatial patterns arising from the process of dispersal are amenable to be studied through diffusion theory and some of these models possess remarkable predictive ability.

In the case of vegetatively reproducing plants which commonly occur as extensive clumps of shoots it is natural to treat the clumps rather than the shoots as entities whose pattern is to be studied. When the pattern of a clumped species can be mapped the result is a two-phase mosaic with a patch phase (where the plant occurs) and a gap phase (where it is absent). Other examples are maps or aerial photos showing land versus water or forest versus grassland. A random line mosaic can be taken as a standard of randomness. It has the following specification. If the mosaic is sampled along a row of equidistant points a unit distance apart, then the sequence of phases encountered (gap or patch) is a realization of a two-state Markov chain. The pattern of the mosaic is completely specified when any two independent transition probabilities are given. This also offers a method of testing whether a natural mosaic can be regarded as random.

Association or lack of it among pairs of groups of species is of obvious ecological interest. When the units for observations are discrete a simple test for association of two species or groups within the sample is provided by the chi-square value computed out of the contingency table. With more than two species the tests are not straightforward and may need appropriate grouping of the frequency classes (Pielou, 1977). When similar tests of association are made on individuals in a continuum one has to see that the sampling units are widely separated and do not share common patches arising from chance factors.

The two major aspects of interest with many-species populations are species abundance relations and measurement of diversity. The species abundance relations are studied through distributions like' lognormal, negative binomial and geometric distributions. These functions provide satisfactory approximations to relationship between number of species and number of individuals and make it possible to predict the total number of species in the whole population from the observed number of species in the sample. In cases where it is difficult to count the number of individuals species-area curves are constructed and similar information can be obtained in certain specific cases. Species diversity is possible to be measured by Shannon's information measure which has several desirable features. But Brilloun's function is more suitable with small and censused communities.

Some additional problems dealt by ecologists are concerned with classification and discrimination of communities which are most appropriately tackled by numerical taxonomic methods. When the entities to be classified do not form distinct groups as it usually occurs, ordination has been found useful in understanding their distribution pattern. Ordination amounts to projecting the original hyperspace onto a space of fewer dimensions in such a way that the arrangement of the points suffers the least possible distortion. Dimension reducing methods like principal component analysis, principal co-ordinate analysis are applicable to linear data structures. But data having nonlinear structures need catenation for dimension reduction. Catenation preserves the relative position of points even in the reduced space.

Ecology is usually defined as the study of organisms in relation to their environments. Hence ecologists frequently measure the environmental variables along with the observations on the organisms in their habitats. Canonical correlation analysis has been found useful in such situations to study the interrelations between these two sets of variables. Studies on vegetation versus environment are amenable to such analysis. An extension of canonical correlation analysis called multiple discriminant analysis operates on several sets of qualitatively similar variates. In ecological contexts, each set usually consists of the amount of each species in each unit of a sample and each of k sets comes from a different population, for instance a different geographical locality. The purpose of multiple discriminant analysis is to find what linear combination of the variates will give the greatest differences among set means.

3.4. Quantitative ethology

Like in the case of many other disciplines the need for quantification in ethology arises from the presence of individual differences in animal behaviour and the need for an objective criterion to decide between cases where marginal differences exist. Behavioural studies in the usual case yield time series data. This could be a complete record of the events or observation at fixed intervals of time. Alternatively, data could be collected on the sequence of activities without a time base on focal animals or their groups. The utilities of these methods are different and particular methods are td be chosen as per one's need.

A simple repertoire analysis involves studying the relationship between the type of acts versus the total number of acts performed and this is most effectively done by a logarithmic regression function. An important parameter: repertoire fraction, is estimated by analysing the observed behavioural abundi ance distribution. This is done by extrapolating the frequency of a zero-abundance class from a theoretical probability distribution fitted to such data. The lognormal Poisson distribution is found especially suited for the purpose.

Information theory is applied to the measurement of communication between individuals. A typical data set may be in the form of a sequence of signals versus responses. Losey (1978) lists a number of information measures for such data and discusses the problems in their application.

Duration, interval and latency of behavioural acts are of interest to ethologists. A study of the distribution of interval between successive acts is many a time revealing. These are most closely approximated by negative exponential or mixed exponential distribution. If events in a time series are independent of one another, there will be no significant correlations among the times of their occurrences. Autocorrelation coefficients may be used to determine the existence of such correlations. Behavioural sequences in discrete time units are amenable to be studied through Markov chains. Sequences in continuous time can be subjected to naive and descriptive techniques of time series analysis or structured and explanatory techniques of stochastic processes.

Hierarchical cluster analysis is particularly well suited for ethological work. The observation of a single individual for any length of time often generates data on a number of variables across a series of time units. This can be conceived of as a simple (behaviour X time) raw data matrix. Similar observations on several individuals can be collapsed through some summary statistics. Alternatively a behaviour X individual data matrix can also be thought of across the time units. Cluster analysis proceeds through computing similarity measures between behaviour types and grouping the entities by application of one of several algorithms available for the purpose. Cluster analysis is helpful in identifying common causal factors in behaviour but is to be applied with caution.

When frequency data are available, classified in the form of a multiway contingency table, log-linear models are of use in analysing such data provided the sample size is large enough. As a general rule, a table with fewer than two observations per cell on average is likely to present difficulties with the analysis. The fit of a particular model is tested by estimating the expected cell frequencies in a way that conforms with that model and by comparing these estimates with the observed cell frequencies by means of a chi-square statistic. Yet another method of dealing with frequency data is the multidimensional scaling which can be used to group related behavioural types. The idea is to reconstruct a set of coordinates for the data points from a given distance matrix and several algorithms are available for the purpose.

Much insight can be gained from multivariate data through principal Component and factor analyses. Principal component analysis is just an orthogonal transformation of the original variables into independent linear combinations (components) and practical experience with the technique suggests that a valid interpretation may usually be expected for those components which account for a significant part of total variation. Factor analysis has an underlying model which expresses each original variable as functions of a set of common factors and specific factors. Factors are again linear combinations of original variables but they need not be made independent and can be oriented in such a way that they are meaningful biologically. The major objectives in both the cases could be dimension reduction and identification of common causes for individual behavioural acts.

In principal component analysis and factor analysis the purpose is to examine variation within a single population. However, animal behaviour research often investigates differences or similarities in patterns of variation between populations. Multivariate analysis of variance and discriminant analysis techniques provide powerful qualitative tools to examine the pattern of intergroup behavioural phenomena. The former is mostly used to investigate the intergroup differences and the latter for classification of individuals of unknown identity to one of the pre-established groups. These also provide information on the character set most responsible for the group differences,

3.5. Mensuration

Mensuration is concerned with measurements in forestry. Measurements arise in terms of billets stacked or otherwise, individual trees or whole stands and the features measured will quite often be length, area or volume and weight in a few cases. Obviously certain features are easily measured and some with difficulty. The prediction of those attributes which are difficult to measure directly in terms of easily measurable Characteristics is the crux of mensuration problems. Since the prediction is probabilistic in nature the Science of statistics plays a major role in this field. Some of the important concepts in the field of mensuration are briefly described here.

3.5.1. Site quality

Site quality is supposed to be a measure of the productive capacity of a designated land area as measured from the existing condition of a stand growing there. Naturally it is a post planting concept and is highly related to the planting material and the associated envlronment 'including management. The concept is useful in the estimation of growing stock in even-aged stands and also in deciding the kind of treatment to be given to a stand based on its current status.

Although some indirect methods are available which utilize information on climatic and edaphic factors and also lesser vegetation characteristics, site quality estimation is usually based on stand top height which is found to be less affected by stand density. As found described in the Working Plan reports, if a complete quality class mapping is intended, then top height is determined for every square chain plot in that area. Top height is obtained as the average height of dominant trees in a square chain plot. It is also found that very often stock mapping is done based on partial enumeration to save time and effort. Chaturvedi and Khanna (1982) have noted the following, 'In India. however upto 1930 average height of all trees was used but after 1930 the site quality classes are based on the top height which is the height corresponding to the mean diameter (calculated from basal area) of the 250 biggest diameters per hectare as read from height-diameter curve'. This seems to imply that the basic unit to be taken for evaluation of top height should be of size one hectare. The definition offers no guidelines as to the number of trees to be taken when the required number is not available in the site.

Most height based methods of site quality evaluation involve the use of site index curves. Any set of site index curves is simply a family of height development patterns with qualitative symbols or numbers associated with the curves for referencing purposes. The most common method of referencing uses the height achieved at some specified reference age which is selected to lie close to the average rotation age.

Site index curves are of three types. First is an anamorphic type, for any two curves of which the height of one at any age is a constant proportion of the height of the other at the same age. In a polymorphic-disjoint curve family, this proportionality relation does not hold but the curves do not cross within the range of interest. For curves in the polymorphic nondisjoint curve family, at least some of the curves intersect within the age range of interest Remeasurement height / age data or stem analysis height / age data can be used to develop all the above three types of site index equations. With temporary plot height / age data, it is only possible to produce anamorphic types. Fitting site index curves is accomplished through one of the following method viz. the guide curve method, the difference equation method or the parameter prediction method (Clutter *et al.,* 1983.)

3.5.2. Yield table

Yield table is a tabular statement which summarises all the essential data relating to the development of a fully stocked and regularlythinned, even: aged crop at periodic intervals covering greater part of its useful life. Yield tables have multiple uses in management such as estimating growing stock, fixing rotation age, assessing site quality and suggesting thinning levels. The standard regression techniques have gained wide acceptance for the construction of yield equations. The set of regression equations underlying a yield table is termed a stand model.

Stand models are mainly of three forms : (i) whole stand models (ii) diameter distribution models and models based on individual trees. Whole stand models are based on stand features like total volume, crop diameter, crop height, stand density, etc. expressed as functions of age. The model based on Richard's functions considered by Franz and Rawat (1979) belong to this category. Diameter distribution models on the other hand approximate an observed diameter distribution of a stand with some theoretical frequency distributions and establish age-dependent relations with parameters of such distributions. The predicted diameter distributions are then condensed to stand volume figures by appropriate functions later. Some of the theoretical distributions found useful for the purpose are weibull, beta and gamma. Jayaraman and Rugmini (1988) have indicated the suitability of beta distribution in representing the diameter distribution for even-aged teak. Individual tree models are described by Arney (1979). They essentiallydescribe the development of individual trees in relation to the competition stress the trees suffer, and later sum up individual volume to whole stand levels, Of late, models based on multivariate allometric relations have been introduced by Garcia (1984). There are also cases where yield tables are established through generalized least squares (Ferguson and Beech, 1978).

3.5.3. *Volume table*

Volume table is essentially the output of a regression function with tree volume as the dependent variable and diameter and/or height as independent variable (s). In practice the best suited function out of a set of polynomial or exponential models is selected based on some goodness of fit criterion. Adjusted coefficient of multiple determination is the most widely used one. But when the form of dependent variable is different in different equations Furnival Index (Furnival, 1961) may be used for comparison. Recent research indicates the use of data splitting techniques involving PRESS statistic (Montgomery and Peck, 1982) for model selection. With unequal error variances, either weighted least squares (Cunia, 1964) or power transformation suggested by Box and Cox (1964) are of use. While using a volume table. care has to be made not to extrapolate the results. Moreover cumulative error in a large number of predictions is usually much lesser than that occurs with an individual case.

3.5.4. *Taper functions*

Taper is the decrease in diameter of a tree stem or a log from base upwards and the rate of taper is termed form. Taper varies not only between but also within trees. The segments of a tree bole approximate to various geometrical frustrums. The volumes of individual segments of a tree can be calculated from taper tables which exist for different diameter and height or form classes. Max and Burkhart (1976) have presented the method and application of segmented polynomial regression models to describe the tree taper. Husch *et al.* (1972) suggest that three regression models are required to describe taper; one model each for the lower, middle and upper segments, A number of other authors also have proposed various types of models useful in describing stem taper. All these models, like that of volume and yield table models, use regression equations in various forms.

3.5.5. *Biomass estimation*

Biomass refers to the total mass of living material in a location and its estimation has assumed considerable importance in recent years. Foresters are mostly concerned with tree biomass and the common procedure of estimation is through the use of regression equations and stand tables. A few stems are destructively sampled and weight of each component is determined and related by regression to some dimensions of the standing tree. A stand table which classifies stems per unit area by units of the dimension used in the regression (usually diameter) is then expanded to an estimate of biomass by multiplying the number of stems in each dimension class by weight. In the literature on biomass estimation in forestry many of the theoretical refinements in regression like generalized least squares, stepwise regression, nonlinear regression, etc. are found used extensively.

The above review would indicate that the subject field of mensuration has a strong footing on the theory of regression and sampling. Hence, any theoretical advancements happening in these areas will be of direct relevance to mensurationists.

3.6. Forest management

Forest management in the broad sense, integrates all of the biological, social, economic and other factors that affect management decisions about the forest. Historically, forest management has dealt primarily with silviculture and the biological management of the forests. In a narrow sense forest management refers to the study and application of analytical techniques to aid in choosing those management alternatives that contribute most to the organizational objectives (Leuschner, 1984).

3.6.1. Forest valuation

Other than the objectives and constraints themselves a prerequisite for decision making is an appraisal of the alternatives. The alternatives can be assessed by placing a value on forest production. Though market prices are the best value estimates available, the time value of money is to be taken into consideration while appraising many of the forestry investments. Foresters have formulated several financial criteria like Present Net Worth (PNW), Internal Rate of Return (IRR) and Land Expectation Value (Le) towards this end and also devised methods of adjusting their values for inflation and taxes. The statistical considerations come in while adjusting for the risk involved in running many of the forestry enterprises.

There are three different conditions under which decisions are made. These are certainty, risk and uncertainty. Certainty exists if there is only one outcome for each alternative. Stated differently, the outcome for each alternative is known. Risk exists if a probability distribution can be attached to the different states of nature and hence to the different outcomes. Uncertainty exists if there is no information about the probability distributions of the states of nature. Under uncertainty minimax rule is the most conservative one as minimizes the maximum loss due to any decision. The riskiness of an investment is the amount of certainty with which the return on that investment, including recouping the initial investment, can be predicted. Riskiness, then is defined as the variability of returns from a proposed investment. It is measured by either the variance or standard deviation of a probability distribution of the returns on that investment. In general investments with lesser variability in returns are preferred. Though there are a number of other techniques available for use under risky environment like Decision Trees and CAPM technique these are less frequently put to actual use.

The case of certainty of outcomes strictly never occurs in real life situations. But decision making with certainty is probably the most widely used model, or assumption, despite its unrealism. All cash flow projections that have only a single point estimate of yield projections using results from a single table or equation are implicitly using the certainty model. Here the point estimate and the year in which it occurs are analysed as if they undoubtedly will occur. Decisions are then made ~~based~~ on the basis of these analyses. The widespread use of the certainty model is probably because it is the easiest assumption to make and the easiest to implement. However the decision makers should remember that these decisions are almost always being made in a risky world.

3.6.2. Forest regulation

Traditional forest regulation has centered around the concepts of sustained yield and normal forest. The implicit objectives in most historical regulation schemes are to maximize the volume harvested and to maintain an increasing or eventually even wood flow. The normal forest provides maximum yield once it is attained. Most of the regulation schemes strive to attain this ideal by manipulating the cutting schedules.

Linear programming (LP) and associated techniques like integer, non-linear and dynamic programming techniques are widely used in forestry. The general linear programme either maximizes or minimizes a linear function called objective function subject to a set of linear equalities and / or inequalities called the constraints. A typical problem for linear programming would run like this. Suppose there are two tracts of land. Tract 1 has 20 hectares and an allowable cut of 100m^3/ha. Tract 2 has 40 hectares and 120m^3/ha allowable cut. It costs Rs. 40ha to prepare and administer a timber sale on either tract. The total appropriation for preparation and administration is Rs. 1500. At least 10000 visitor days of recreation must be provided. Tract 1 can provide 200 visitor and Tract 2 can provide 500 visitor days per hectare. Of course recreation is lost when an area is cut. The problem is to determine how many hectares to cut in each tract to maximise timber cut. LP is extensively used in harvest scheduling as well. Here, the objective shall be to identify the management regime in terms of the cutting schedule which maximizes the PNW or harvest volume over a specified planning horizon and cutting period length subject to certain constraints on the magnitude of wood flow. Several ready made packages like MAXIMILLION, Timber RAM, ECHO and TREES are now available for executing general LP scheduling models.

Yet another use of LP is found in the multiple use management problems. Optimum allocation of the resources for multiple uses, which brings forth maximum returns again subject to the usual management constraints, is achieved through the same algorithm. Goal Programming, a variation of the LP which incorporates multiple objectives in the optimization has been found specially suited for such situations.

3.7. *Econometrics*

Econometrics deals with the application of statistical techniques in the field of economics. Since the major sets of techniques applied here do not largely differ from that of many other interdisciplinary areas and have been referred to many times. the discussion is limited to a few cursory remarks.

Forest economists are concerned with allocating scarce resources among competing means to satisfy human wants for forestry products. The resources of interest here are the products themselves and the resources used to produce these products. Any economic study on these has to take care of their particular nature like immobility of the production units, long production period, flexibility with harvesting options, derived demand from finished products, the various externalities and nature of ownership. The major aspects of interest with forest products are their market systems, demand analysis, benefit cost analysis and pricing all of which involve mathematical concepts. Economics of timber production involves timber harvesting decisions which have to take into account costs and returns that are incurred or received at different times and other external pressures put on these resources. Of late economists have started recognizing the role of several nontimber like forage, water, wildlife, recreation and a multitude of minor forest products which had led to the theory of multiple use based on operations research techniques. It has also become important to realize that economists cannot work in isolation and have *to* pay attention to the factors operating at the national or international levels (Gregory, 1987).

With regard to the techniques as such both time series and cross section data are utilized by economists and regression and multivariate techniques have dominated this area. Most of the studies with historical data utilize time-series analysis (Box and Jenkins, 1970). Further, a large part of the theory deals with econometric modelling based on simultaneous equation methods in the case of macroeconomics. Some concepts and techniques pertaining to microeconomics can be found under the section on forest management.

4. A STATISTICAL EXPERT SYSTEM FOR SOME COMMON SITUATIONS IN FORESTRY RESEARCH

Expert systems are intelligent computer programmes capable of performing at the level of human experts in narrow domains of solving problems which require significant human expertise. This is a branch of the fast growing field of artificial intelligence (Al).

A statistical expert system essentially tries to replace a statistical consultant. The knowledge base of a statistician, once stored in a computer can be accessed in an interactive manner with an appropriate software package. A sample of such a knowledge base relevant to forestry research is given below. This is made operational with a software package by name 'IITM RULE' developed at the Indian Institute of Technology, Madras.

The knowledge base acts as a key to choose a statistical technique suited to a specific situation. Certain general comments are in order here. The system is open-ended and may need extension at many of the terminal points. Presently only suggestions as to which class of techniques is to be adopted when, are given. In course of time this can be linked up with the actual execution of the analysis suggested. In some cases though the same test is specified in different contexts it should be understood that the test criterion or the formula for computation could differ. Although not categorically declared a minimum sample size will be needed for all the procedures to be effective. Some of the terms that shall be encountered while using the key are explained below.

Nominal scale refers to measurement at its weakest level when numbers or other symbols are used simply to classify an object, person or characteristic.

Ordinal scale is one wherein given a group of equivalence classes, the relation 'greater than' holds for all pairs of classes so that a complete rank ordering of classes is possible.

When a scale has the characteristics of an ordinal scale, and when in addition the distances between any two numbers on the scale are of known size, interval scale is achieved.

An interval scale with a true zero point as its origin forms a ratio scale. In a ratio scale, the ratio of any two scale points is independent of the unit of measurernent.

A response surface is the geometrical configuration of a factor- response relationship. It depicts how the response variable is related to the levels of input variables.

From a layman's point of view, normal distribution refers to the typical symmetrical bell shaped frequency distribution for the observations. In stricter terms it has to follow a particular probability density function.

A paired sample arises when observations are generated on the same sampling units under two different conditions like before and after some treatment or when subjects are grouped by matching characteristics and assigned different treatments within each group. The word related is used when such a set arises from several combinations instead of just two.

.

Code	Dialogue	Action/Suggestion
1	*Purpose of consulting*	
	a) to plan a study	Go to 11
	b) to analyse data	Go to 12
11	*Type of investigation*	
	a) experiment	Go to 111
	b) survey	Go to 112
111	*Advice required on*	
	a) choosing a design	Go to 111 1
	b) others	Consult a statistician
111 1	*Number of factors*	
	a) single factor	Goto 111 11
	b) multiple factors	Goto 111 12
I11 11	*Variability in the experimental material*	
	a) homogeneous	Completely randomized design
	b) heterogeneous	Go to 111 112
111 112	*Treatments form subgroups*	
	a)	Go to 111 112 1
	b) no	Go 111 1.12 2

Code	Dialogue	Action / Suggestion
111 112 1	*Reason for forming subgroups*	
	a) less experimental material for some treatments	Augmented designs
	b) treatments within subgroups related	Compact family block design
111 112 2	*Number of treatments*	
	a) less than or equal to 10	Randomized complete block design
	b) more than 10	Go to 111 112 22
111 112 22	*Number of treatments*	
	a) perfect square	Lattice designs
	b) not a perfect square	Balanced incomplete block designs or partially balanced incomplete block designs
111 12	*Nature of factors*	
	a) quantitative	Go to 111 121
	b) qualitative	Go to 111 122
111 121	*Objective of the experiment*	
	a) to study the response surface	Go to 111 121 1
	b) to study main effects and interactions	Go to 111 122
111 121 1	*Nature of factor levels*	
	a) proportions	Mixture designs
	b) others	Response surface designs
111 122	*Number of factors*	
	a) two	Go to 111 122 1
	b) three	Go to 111 122 2
	c) more than three	Go to 111 122 3
111 122 1	*Variability in the experimental material*	
	a) homogeneous	Completely randomized design (Factorial)
	b) heterogeneous	Go to 111 122 12

Code	Dialogue	Action / Suggestion
111 122 12	*Factor levels require large plot size*	
	a) for one factor	Split plot design
	b) for two factors	Strip plot design
	c) not required	Randomized complete block design (Factorial)
111 122 2	*Variability in the experimental material*	
	a) homogeneous	Completely randomized design (Factorial)
	b) heterogeneous	Go to111 122 22
111 122 22	*Factor levels require large plot size*	
	a) for one factor	Split plot design
	b) for two factors	Split - split plot design
	c) not required	Confounded designs
111 122 3	*Information needed*	
	a) main effects only	Main effect orthogonal plans
	b) main effects and some interactions	Fractional factorial designs
112	*Advice required on*	
	a) sampling scheme	Go to 112 1
	b) others	Consult a statistician
112 1	*Variability in the population*	
	a) high	Go to 112 11
	b) low	Go to 112 12
112 11	*Classification possible by source of variability*	
	a) Yes	Go to 112 111
	b) no	Go to 112 112

Code	Dialogue	Action / Suggestion
112 111	*Sampling units large and divisible further*	
	a) Yes	Stratified multistage sampling (Go to 112 12 for sampling at different stages within strata)
	b) no	Stratified sampling (Go to 112 12 for sampling within strata)
112 112	*Sampling units large and divisible further*	
	a) Yes	Multistage sampling (Go to 112 12 for sampling at different stages)
	b) no	Go to 112 12
112 12	*Effort in locating a selected unit*	
	a) easy	Go to112 121
	b) difficult	Systematic sampling
112 121	*Nature of the sampling units*	
	a) forms natural clusters	Go to112 121 1
	b) independent	Go to 112 121 2
121 121 1	*Size of the sampling units*	
	a) nearly equal	Go to 112 121 11
	b) largely unequal	Go to 112 121 12
112 121 11	*Effort in measuring some characters*	
	a) easy	Cluster sampling
	b) difficult	Multiphase cluster sampling
112 121 12	*Effort in measuring some characters*	
	a) easy	Cluster sampling with PPS
	b) difficult	Multiphase cluster sampling with PPS
112 121 2	*Effort in measuring some characters*	
	a) easy	Simple random sampling
	b) difficult	Multiphase sampling

Code	Dialogue	Action / Suggestion
12	*Objective of analysis*	
	a) to estimate	Go to 121
	b) to compare (location)	Go to 122
	c) to study relations	Go to 123
	d) others	Consult a statistician
121	*Sample from*	
	a) finite population	Estimators as per the sampling scheme and the property estimated
	b) infinite population	Goto 121 2
121 2	*Observations on sampling units*	
	a) independent of each other	Go to 121 21
	b) not independent	Estimation not possible
121 21	*Number of variables*	
	a) univariate	Go to 121 211
	b) multivariate	Go to 121 212
121 211	*Scale of measurement*	
	a) ratio	Go to 121 211 1
	b) interval	Go to 121 211 1
	c) ordinal	Go to 121 211 3
	d) nominal	Go to 121 211 4
121 211 1	*Distribution of the variable* (Use the goodness of fit test to confirm)	
	a) normal	Go to 121 211 11
	b) unknown	Go to 121 211.3
121 211 11	*Property estimated*	
	a) location	Sample mean
	b) dispersion	Sample standard deviation
	c) skewness	Coefficient of skewness
	d) kurtosis	Coefficient of kurtosis
121 211 3	*Property estimated*	
	a) location	Sample median
121 211 4	*Property estimated*	
	a) location	Sample proportion

Code	Dialogue	Action/Suggestion
121 212	*Scale of measurement*	
	a) ratio	Go to 121 212 1
	b) interval	Go to 121 212 1
	c) ordinal	Go to 121 212 3
	d) nominal	Go to 121 212 4
121 212 1	*Distribution of the variables* (Use the goodness of fit test to confirrn)	
	a) multivariate normal	Go to121 212 11
	b) unknown	Go to 121 212 3
121 212 11	*Property estimated*	
	a) location	Vector of sample means
	b) dispersion	.Matrix of sample variances and covariances
121 212 3	*Property estimated*	
	a) location	Vector of sample medians
121 212 4	*Property estimated*	
	a) location	Vector of sample proportion
122	*Observations on sampling units*	
	a) independent of each other	Go to 122 1
	b) not independent.	No test available
122 1	*Number of variables*	
	a) univariate	Go to 122 11
	b) multivariate	Break up to univariate cases or Go to 122 12
122 11	*Number of populalions to compare*	
	a) one	Go to 122 111
	b) two	Go to 122 112
	c) more than two	Go to 122 113
122 111	*Scale of measurement*	
	a) ratio	Go to 122 111 1
	b) interval	Go to 122 111 1
	c) ordinal	Kolrnogorov - Smirnov one sample test
	d) nominal	Chi-square test for goodness of fit

Code	Dialogue	Action / Suggestion
122 111 1	*Sample size*	
	a) less than or equal to 30	Go to 122 111 11
	b) more than 30	Standard normal deviate test for mean
122 111 11	*Distribution of the variable*	
	a) normal	Students't test for mean
	b) unknown	Kolmogorov-Smirnov one sample test
122 112	*Scale of measurement*	
	a) ratio	Go to 122 112 1
	b) interval	Go to 122 112 2
	c) ordinal	Go to 122 112 3
	d) nominal	Go to 122 112 4
122 112 I	*Sample size*	
	a) less than or equal to 30	Go to 122 112 11
	b) more than 30	Go to.122 112 12
122 112 11	*Distribution of the variable*	
	a) normal	Go to 122 112 111
	b) unknown	Go to 122 112 3
122 112 113	*Independence of samples*	
	a) independent	Go to 122 112 111 1
	b) paired	Paired t test
122 112 111 1	*Equality of variances* (Use variance ratio test to confirm)	
	a) equal	Student's t test for means
	b) unequal	Cochran's weighted t test for means
122 112 12	*Independence of samples*	
	a) independent	Go to 122 112 121
	b) paired	Paired t test
112 112 121	*Equality of variances* (Use variance ratio test to confirm)	
	a) equal	Standard normal deviate test for means
	b) unequal	Cochran's weighted t test

Code	Dialogue	Action / Suggestion
122 112 2	*Independence of samples*	
	a) independent	Randomization test for two independent samples
	b) paired	Randomization test for matched pairs
122 112 3	*Independence of samples*	
	a) independent	Mann-whitney U test
	b) paired	Wilcoxon matched-pairs signed rank test
122 112 4	*Independence of samples*	
	a) independent	Chi-square test for two. independent samples
	b) paired	McNemar test for the significance of changes
122 113	*Scale of measurement*	
	a) ratio	Go to 122 113 1
	b) interval	Go to 122 113 1
	c) ordinal	Go to 122 113 3
	d) nominal	Go to 122 113 4
122 113 1	*Distribution of the variable*	
	a) normal	Go to 122 113 11
	b) unknown	Go to 122 113 12
122 113 11	*Independence of samples*	
	a) independent	Go to 122 113 111
	b) related	Paired t test,
122 113 111	*Equality of variances* (Use Bartlett's test for homogeneity of variances to confirm)	
	a) equal	Analysis of variance as per the experimental design followed by multiple comparison tests if necessary
	b) unequal	Analysis of variance as per the experimental design after transformations followed by multiple comparison tests if necessary

Code	Dialogue	Action/Suggestion
122 113 12	*Independence of samples*	
	a) independenl	Analysis of variance as per the experimental design after transformations followed by multiple comparison tests if necessary
	b) related	Friedman two-way analysis of variance by ranks
122 113 3	*Independence of samples*	
	a) independent	Kruskal-wallis one-way . analysis of variance
	b) related	Friedman two-way analysis of variance by ranks
122 113 4	*Independence of samples*	
	a) independent	Chi-.square test for k independent samples
	b) related	Cochran *Q* test
122 12	*Scale of measurement*	
	a) ratio	Go to 122 121
	b) interval	Go to 122 121
	c) ordinal	Multivariate nonparametric tests
	d) nominal	Multivariate nonparametric tests
122 121	*Distribution of the variables*	
	a) multivariate normal	Goto 122 121 1
	b) unknown	Multivariate nonparametric tests
122 121 1	*Number of populations to compare*	
	a) one	Goto 122 121 11
	b) two	Goto 122 121 12
	c) more than two	Go to 122 121 13
122 121 11	*Sample size*	
	a) more than the number of variables	F test based on Hotelling's T^2
	b) less than or equal to the number of variables	No test available

Code	Dialogue	Action / Suggestion
122 121 12	*independence of samples*	
	a) independent	Go to 122 121 121
	b) paired	Use the corresponding univarate test
122 121 121	*Equality of covariance matrices* (Use likelihood ratio test to confirm)	
	a) equal	Go to 122 121 121 1
	b) unequal	Go to 122 121 121 2
122 121 121 1	*Sample size of the two populations*	
	a) combined sample size less one greater than the number of variables	F test based on Hotelling's T^2
	b) Combined sample size less one less than or equal to the number of variables	No test available
122 121 121 2	*Sample sizes of the two populations*	
	a) equal	Go to 122 121 121 21
	b) unequal	Go to 122 121 121 22
122 121 121 21	*Sample size of **any** one population*	
	a) greater than .the number of variables	F test based on Hotelling's T^2
	b) less than or equal to the number of variables	No test available
122 121 121 22	*Smaller of* the *sample sizes of two populations*	
	a) greater than the number of variables	F test based on Hotelling's T^2
	b) less than or equal to the number of variables	No test available

Code	Dialogue	Action / Suggestion
122 121 13	independence of samples	
	a) independent	Go to 122'121 131
	b) related	Use the corresponding univariate tests
122 121 131	Equality *of* covariance matrices (Use likelihood ratio test to confirm)	
	a) equal	Go to 122 121 131 1
	b) unequal	Consult a statistician
122 121 131 1	Sampiesize	
	a) combined sample size greater than the number of variables	Multivariate analysis of variance
	b) combined sample size less than or equal to the number of variables	No test available
123	Relation between	
	a) variables	Go to 123 1
	b) entities	Go to 123 2
123 1	Narure *of* relation	
	a) Interrelation	Go to 123 11
	b) dependence	Go to 123 12
123 11	Scale *of* measurement	
	a) ratio	Go to 123 111
	b) interval	Go to 123 111
	c) ordinal	Go to 123 113
	d) nominal	Go to 123 114
123 111	*Type of* relation	
	a) between two variables ignoring the effects of others	Go to 123 111 1
	b) between two variables eliminating the influence of others	Go to 123 111 2
	c) joint effect of many **on** one variable	Go to 123 111 3
	d) between sets of variables	Go to 123 111 4

Code	Dialogue	Action / Suggestion
123 111 1	*Nature of relation* (Use scatter diagrammes to confirm)	
	a) linear	Go to 123 111 11
	b) nonlinear	Go to 123 111 12
123 111 11	*Distribution of the variables*	
	a) bivariate normal	Go to 123 111 111
	b) unknown	Go to 123 113 1 after transforming to ordinal scale
123 111 111	*Inference needed*	
	a) estimation	Pearson's product moment correlation coefficient
	b) test of significance	Go to 123 111 111 2
123 111 111 2	*Nature of null hypothesis*	
	a) $P = O$	Student's t test for testing correlations
	b) $P = P_o$	Standard normal deviate test after Fisher's z transformation
	c) $P_1 = P_2$	
	d) $P_1 = P_2 = \dots = P_k$	Chi - square test for homogeneity of correlation coefficients
123 111 12	*Inference needed*	
	a) estimation	Correlation ratio
	b) test of significance $(P=O)$	F test for correlation ratio
123 111 2	*Nature of relation*	
	a) linear	Go to 123 112 1
	b) nonlinear	Go for regression analysis given in 123 121 13
123 112 1	*Distribution of the variables*	
	a) multivariate normal	Go to 123 112
	b) unknown	Go to 123 113

Code	Dialogue	Action / Suggestion
123 112 11	*Inference needed*	
	a) estimation	Partial correlation coefficient
	b) test of significance (P=O)	Student's t test for partial correlation coefficient
123 111 3	*Nature of relation*	
	a) linear	Go to 123 111 31
	b) nonlinear	Go for regression analysis given in 123 121 13
123 111 31	*Distribution of the variables*	
	a) multivariate normal	Go to 123 111 311
	b) unknown	Nonparametric regression analysis
123 111 311	*Inference needed*	
	a) estimation	Multiple correlation coefficient
	b) test of significance (P=O)	F test for multiple correlation coefficient
123 111 4	*Type of relation*	
	a) linear	Go to 123 111 41
	b) nonlinear	Go for multivariate regression analysis
123 111 41	*Distribution of the variables*	
	a) multivariate normal	Canonical correlations
	b) unknown	Multivariate nonpara-metric case
123 113	*Type of relation*	
	a) between two variables ignoring the effects of others	Go to 113 1
	b) between two variables eliminating the influence of others	Kendall partial rank correlation coefficient
	c) among several variables	Kendall coefficient of concordance

Code	Dialogue	Action/Suggestion
123 113 1	*Inference needed*	
	a) estimation	Kendall rank correlation coefficient
	b) test of significance	Go to 123 113 12
123 113 12	*Sample size*	
	a) less than 10	Exact probability test
	b) more than or equal to 10	Standard normal deviate test
123 114	*Inference needed*	
	a) estimation	Contingency coefficient
	b) test *of* significance	Chi-square lest for contingencycoefficient
123 12	*Scale of measurement*	
	a) ratio	*Go* to 123121
	b) interval	Go to 123 123
	c) ordinal	Nonparametric regression analysis
	d) nominal	Go to 123 124
123 121	*Form of the model*	
	a) single equation	Goto 123 121 1
	b) simultaneous:equation model	Consult a statistician
123 121 1	*Nature of relation*	
	a) linear	Go to 123 121 I1
	b) linearizable	Linearize through transformation and Go to 123 121 11
	c) nonlinear	Go to 123 121 13
123 121 1I	*Predictor variables* (Avoid highly correlated variables among predictors)	
	a) already selected	Go to 123 121 111
	b) to be selected from a larger set	Do stepwise regression and Go to 123 121 111

Code	Dialogue	Action / Suggestion
123 121 111	*Residuals of ordinary least squares*	
	a) outliers present	Eliminate outliers and go to123 121 111 2
	b) no outliers	Go to 123 121 111 2
123 121 111 2	*Nature of residuals of ordinary least squares* (Use Durbin-Watson test to confirm)	
	a) independent	Go to 123 121 111 21
	b) correlated	Go to 123 121 111 21 after reparameterizing throygh Cochrane and Orcutt procedure
123 121 112 21	*Variance of residuals* (Examine the residuals of ordinary least squares)	
	a) equal	Go to 123 121 111 211
	b) proportional to some factor	Same as 123 121 111 211 except that estimation is by weighted least squares
	c) unequal without any pattern	Same as 123 121 111 211 except that estimation is by generalized least squ-ares
123 121 111 211	*Inference needed*	
	a) estimation	Ordinary least squares
	b) test of significance	Go to 123121 111 211 2
	c) model comparison	Go to 123 121 111 211
	d) model validation	Use 'PRESS' statistic
123 121 111 211 2	*Distribution of the residuals*	
	a) normal	F test through analysis of variance
	b) unknown	Repeat estimation after normalizing transforma-tions

Code	Dialogue	Action / Suggestion
123 121 111 211 3	*Form of the dependent variable*	
	a) same	Adjusted coefficient of multiple determination
	b) different	Furnival index
123 121 13	*Nature of function*	
	a) polynomials	Go to 123 121 11
	b) others	Use maximum likelihood estimation procedures
123 124	*Nominal scale for*	
	a) dependent variable	Go to 123 124 1
	b) independent variable	Generalized least squares
123 124 1	*Nature of the dependent variable*	
	a) dichotomous	Use. maximum likelihood estimation procedure
	b) proportions	Go to 123 124 12
123 124 12	*Nature of the experiment*	
	a) bioassay	Probit analysis
	b) others	Use normalizing transformations and go to 123 121
123 2	*Number of variables*	
	a) univariate	Intraclass correlation coefficient
	b) multivariate	Go to 123 22
123 22	*Purpose of study*	
	a) classification	Cluster analysis
	b) identification	Discriminant analysis

5. CONCLUSIONS

Research in forestry has progressed through scientific method. The levels of organization that were dealt with varied from micro organisms to world forests. Some of these are amenable to experimentation but some others are just observable in their natural states. The long time span involved in many forestry issues makes it imperative to seek their solutions through modelling and simulation. Thus experiments, surveys and simulation constitute three broad strategies of investigations in forestry. Simple experimental designs like randomized block design and lattice designs coupled with error reducing techniques such as use of optimum plot size, analysis of covariance and data transformations seem to be appropriate for most of the forestry trials. There is a need *to* develop methods for estimating growing stock of minor forest products and for conducting wildlife censuses. Future forestry trials have to take advantage of the powerful simulation techniques which can to some extent replace costly and time consuming field trials.

Applications of statistical and mathematical techniques in forestry have not remained restricted to the use of certain experimental. designs or sampling schemes. Their uses arise in a wide variety of contexts. Estimation of variance components in genetics, classification and discrimination problems in taxonomy, study of spatial and temporal distributions in ecology, analysis of behavioural patterns in ethology, prediction problems in mensuration, optimization problems in forest management and trend analysis in econometrics are just a few such instances. But perhaps the most important among them all would be that of obtaining a measure of uncertainty attached to our conclusions!

It is not uncommon to find that statistical techniques are misused more than used. Insufficient attention is given to the assumptions involved in many of the statistical tests and estimation procedures. Statistical expert systems have tremendous potential as an educational tool and as a practical aid for the scientists in choosing the specific techniques that are needed in their research. A beginning in this direction has been made here by developing a statistical expert system for some common situations in forestry research.

6. REFERENCES

Anonymous 1984. Statistical survey of wildlife at Eravikulam national park. Census of Nilgiri Tahr. Project Report. Department of Mathematics, Union Christian College, Alwaye. 32p.

Arney, J. D. 1979. An individual tree model for stand simulation in Douglas fir. In Fries, J. (ed.) Growth Models for Tree and Stand Simulation. Royal College of Forestry, Sweden. 379 p.

Binkley, C. S. 1980. Is succession in hardwood forests a stationary Markov process? For. Sci., 26(4):566-570.

Box, G.E.P. and Cox, D. R. 1964. An analysis of transformation. J. R. Stat. Soc. Ser. B. 26:211-243.

Box, G. E P. and Jenkins, G. M. 1970 (2nd ed,). Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco. 575p.

Chacko, V. J. 1965. A Manual on Sampling Techniques for Forest Surveys. The Manager of Publications, Delhi. 172p.

Chacko, K. C. and Jayaraman. M. S. 1988. Effect of Container size on growth of *Bambusa arundinacea* Willd. seedlings in nursery. Paper presented in the International Bamboo Workshop Cochin. November, 1988.

Chacko. K C., Sankar, S., Pandalai. R. C. and Nandakumar, U. N. 1989. Studies on the effect of slash burning on planting site for teak. KFRI Research Report 61. Kerala Forest Research Institute, Peechi, Kerala, India. 21p.

Chaturvedi, A. N. and Khanna, E. S. 1982. Forest Mensuration. International Book Distributors, India. 406p.

Clutter, J. L., Fortson, J. C., Pienaar, L. V., Brister, G. H., and Bailey, R. L. 1983. Timber Management : A Quantitative Approach. John Wiley and Sons, New York. 333p.

Cornell, J. A. 1981. Experiments with Mixtures : Design, Models and the Analysis of Mixture Data. John Wiley and Sons, New York. 305 p.

Cunia. T. 1964. Weighted least squares method and construction of volume tables. For. Sci., 10(2) : 180-191.

Cunia, T. and Chevrou, R. B. 1969. SPR on three or more occasions. For. Sci., 15(2) : 204-224.

Cunia, T. and Kyaw, U. S. 1985. Sampling with partial replacement with different types of samples. Can. J. For. Res., 15(1) : 191 - 194.

Federer, W. T. 1967. Experimental Design : Theory and Application. Oxford and IBH Publishing Co., New Delhi. 544p.

Ferguson, I. S. and Beech, J. W. 1978. Generalized least squares estimation of yield function. For. Sci., 24(1) : 27-42.

Franz, F. and Rawat, A. S. 1979. Detailed nonlinear asymptotic regression studies on tree and stand growth with particular reference to forest yield research in India. In Fries, J. (ed.) Growth Models for Tree and Stand Simulation. Royal College of Forestry, Swedon. 379p.

Furnival, G. M. 1961. An index for comparing equations used in constructing volume table. For. Sci., 7(4) : 337-341.

Garcia, O. 1984. New class of growth models for even aged stands. *Pinus radiata* in Golden downs forest. New Zealand J. For. Sci., 14(1) : 65-88.

Gomez, K.A. and Gomez, A. A. 1984 (2nd ed.). Statistical Procedures for Agricultural Research. John Wiley and Sons, New York. 680p.

Gregory, G. R. 1987. Resource Economics for Foresters. John Wiley and Sons, New York. 477p.

Gyllenberg, H.G. 1963. A general method for deriving determination schemes for random collections of microbial isolates. Ann. Acad. Sci. Fenn., Ser. A. IV. Biol., No. 69. 23p.

Husch, B., Miller, C. I. and Beers, T. W. 1972 (2nd ed.). Forest Mensuration. Ronald Press Company, New York. 410p.

Jayaraman, K. and Bailey, R. L. 1988. Modelling mixed species stands with Scheffe's Canonical polynomials. In Ek. A. R. *et a/* (eds.). Forest Growth Modelling and Prediction. Vol. 1. Proceedings of the IUFRO Conference Aug. 23-27, 1987, Minneapolis, Minnesota. USDA For. Serv. North Central Forest Experiment Station. General Technical Report, No - 120, 201-208.

Jayaraman, K. and Rugmini, P. 1988. Diameter distribution for even-aged teak. Ind. J. For., 11(2) : 145-147.

Jeffers, J.N.R. 1978. An Introduction to Systems Analysis : with Ecological Applications. Edward Arnold. London. 198p.

Kleijnen, J.P.C. 1974. Statistical Techniques in Simulation, Part 1. Marcel Dekker, Inc., New York. 285p.

Kumar, R. 1985. Supply and demand models in forestry - a survey and a suggested model. Journal of World Forest Resource Management, l(2) : 133-150.

Leuschner, W. A. 1984. Introduction to Forest Resource Management. John Wiley and Sons, New York. 298p.

Lloyd, M. 1967. Mean Crowding. J. Anim. Ecol., 36 : 1-30.

Losey, G. S. Jr. 1978. Information Theory and Communication. In Colgan, P. W. (ed.) 'Quantitative Ethology'. John Wiley and Sons, New York. pp 43-78.

Mathew, G. 1989. Biology and ecology of the teak trunk borer *Cossus-cadambae* Moore and its possible control. KFRl 'Research Report (under preparation). Kerala Forest Research Institute, Peechi, Kerala, India. 78p.

Max, T. A. and Burkhart. H. E. 1976. Segmented polynomial regression applied to taper equations. For. Sci., 22(3) : 283-289.

Montgomery, D. C. and Peck, E. A. 1982. Introduction to Linear Regression Analysis. John Wiley and Sons, New York. 504p.

Nair, K. S. S., Sudheendrakumar, V. V.. Varma, R. V., and Chacko, K. C. 1985. Studies on the seasonal incidence of defoliators and the effect of defoliation on volume increment of teak. KFRl Research Report 30. Kerala Forest Research Institute, Peechi, Kerala, India. 78p.

Namkoong, G. 1981. Introduction to Quantitative Genetics in Forestry. Castle House Publications, U. K. 342p.

Pielou, E. C. 1977. Mathematical Ecology. John Wiley and Sons, New York. 385p.

Sale, J. B. and Berkmuller, K. (eds.) 1988. Manual of wildlife techniques for India. Field document No.11. Wildlife Institute of India, Dehra Dun.

Sneath, P.H.A. and Sokal, R. R. 1973. Numerical Taxonomy. W. H. Freeman and Company, San Francisco.

Surendran, T. and Seethalakshmi, K. K. 1987. Vegetative propagation of some important tree species by rooting cuttings. KFRI Research Report 50. Kerala Forest Research Institute, Peechi, Kerala, India. 24 p.

Wagner van, C. E. 1983. Simulating the effect of forest fire on longterm annual timber supply. Can. J. For. Res., 13(3): 451-457.

White, T. L. 1984. Designing nursery experiments, In Duryea, M.L. and Landis. T. M. (eds.) Forest Nursery Manual: Production of Barefoot seedlings. Forest Research Laboratory, Oregon State University. 291-306.

Yang, R. C. 1983. Composite design versus factorial experiments in forest fertilization trials. Can. J. For. Res., 13(3): 438-444.