KFRI Research Report 149

ANALYSIS OF DATA FROM LONG TERM TRIALS

P. Rugmimi
K. Jayaraman

KERALA FOREST RESEARCH INSTITUTE
PEECHI, THRISSUR

July 1998 Pages: 28

CONTENTS

ABSTRACT

Data arising from repeated measurements of experimental units occur in many occasions in forestry and related fields. Very often such data are analysed without considering their several peculiarities, like correlation between successive measurements and heterogeneity of variances, which may lead to erroneous conclusions. The present study was undertaken with the objective of identifying appropriate methods of analysis of data from long term trials characterised by repeated measurements on experimental units.

In this study, three different methods of analysing repeated measures *viz.,* two way analysis of variance, univariate mixed model analysis of variance and multivariate analysis of variance were discussed with respect to their suitability in different contexts. These three methods were compared using data collected from certain typical situations in forestry and the appropriate method of analysis to be followed in respective cases were identified. Specifically, data collected from a study on several soil properties observed from multiple core samples from 0-15, 15-50 and 50-100 cm layers under six different types of vegetation and another study on annual yield of latex from rubber trees in three years were used for comparing the appropriateness of the methods. The study revealed that multivariate analysis of variance is the most appropriate method of analysis for majority of the soil properties. This was found justifiable because the extent of residual variation in individual soil properties at different depths was not of the same order and also the correlation between values at different layers were not of the same magnitude. Multivariate analysis of variance was found suitable for analysing data on annual yield of latex from rubber trees as well. This was so due to the heterogeneity in variances and covariances of variance-covariance matrix of errors.

1. INTRODUCTION

Repeated measurements of observational units are very frequent in forestry research. The term repeated is used to describe measurements which are made of the same characteristic on the same observational unit but on more than one occasion. In longitudinal studies, individuals may be monitored over a period of time to record the changes occurring in their states. Typical examples are periodical measurements on diameter or height of trees in a silvicultural trial, observations on disease progress on a set of seedlings in a nursery trial, etc. Repeated measures may be spatial rather than temporal. For instance, consider measurements on wood characteristics of several stems at the bottom, middle and top portion of each stem and each set of stems coming from a different species. Another example would be that of soil properties observed from multiple core samples at 0- 15, 15-50 and 50- 100 cm depth from different types of vegetation.

The distinct feature of repeated measurements is the possibility of correlations between successive measurements over space or time. Autocorrelation among the residuals arising on account of repeated measurements on the same experimental units leads to violation of the basic assumption of independence of errors for conducting an ordinary analysis of variance. Quite often, analysis is carried out ignoring the possible correlation among the residuals leading to distorted conclusions. Hence there exists a need to identify suitable methods of analysis considering the peculiarities of the data structure in such cases.

There are different ways of analysing the repeated measurements (Crowder and Hand, 1990). These methods vary in their efficiency and appropriateness depending upon the nature of data. Usually, the repeated measurements over time or space within an individual are considered as levels of a factor and analysis done ignoring the possible within subject correlation. However, there are better alternatives in such contexts and an attempt is made here to compare the different analytical methods possible, taking real life data from certain typical situations in forestry.

2. GENERAL EVALUATION OF METHODS

A common case in many statistical investigations is the collection of data from groups of experimental units each of which is observed under two or more conditions. Such studies are called repeated measurements experiments. Although there exists several methods of analysis of data from repeated measure experiments in various disciplines (Koch *et al.* 1980, Diggle and Donnelly 1989), this has not received much attention in forestry.

Comprehensive studies on the various problems of data analysis in experiments involving repeated measurements are scant. In an experiment with repeated measurements, the measurements have a temporal or spatial sequence; with the consequence, the measurements on the same subject separated in small time or space will in general be highly correlated. A review of some relevant works of past in the topics is furnished in the following.

2.1. Two-way analysis of variance

Yates and Cochran (1938) proposed the analysis of data from a set of experiments involving same or similar treatments carried out at a number of places or in a number of years. They pointed out that the standard analysis of variance procedure suitable for dealing with the results of the single experiments needed modification owing to the lack of equality in the error components and that in the interactions of different groups of treatments with places or time.

Khosla *et al.* (1979) studied the behaviour of experimental errors and presence of treatment x year interaction in the case of groups of experiments, involving single experimental error. Homogeneity of experimental errors was studied with references to different crops, type of experiments and broad soil types. They used the weighted analysis for testing the presence of treatment x year interaction.

Andersen *et al*. (1981) discussed the use of two-way ANOVA. They pointed out that it is inappropriate to use the two-way ANOVA if one of the criteria of partition is the time because of the serial correlation. They suggested that this correlation should be taken into account when the significance of the general trend is evaluated. Arnold (1981) indicated the use of univariate approach when covariance matrix of errors is structured *ie.,* when sphericity is met.

Cullis and McGilchrist (1990) developed a model for growth data from designed experiment. They found that the errors of the model were found to be correlated over time.

2.2. Univariate mixed model analysis of variance

Patterson (1939) considered the problem of field experimentation with perennial crops and suggested the use of split-plot design for the analysis of long term experiments with years assigned to subplots and treatments assigned to main plots.

and Walters (1976) criticised the validity of using split-plot analysis to repeated measurement data. They suggested an alternative analytical approach in which contrasts over time are analysed for examining time and time x treatment interaction effect. They also indicated that, for many situations, orthogonal polynomials (linear, quadratic, etc.) would be appropriate.

Ware (1985) suggested methods that not only take into account the intercorrelation of serial measurements but also accommodate the complexities of typical longitudinal data sets and permit the specification of mean value functions determined by subject matter considerations rather than by constraints introduced by the statistical methodology. The approach could be used for categorical outcomes or for nonlinear modelling of continuous outcome variables. The author discussed three families of covariance functions *viz.,* multivariate, autoregressive and random effects.

Gill (1986) proposed some modifications in split-plot analysis for the repeated measurements when the number of individuals per treatment is not more than five or six. He partitioned the treatment x period interaction of the univariate split-plot analysis to permit sensitive comparison of treatments. Modifications for the procedure were given for the case of heterogeneous variance and covariance.

Gill (1988) criticised the analysis of increment of the original data with split-plot model. He was of the opinion that reduction of inter-period correlation by using first order difference did not necessarily eliminate problems with heterogeneity of variance-covariance matrix over time. For the homogeneous conditions, the expected variance of a simple trend contrast was shown to be the same for either analysis.

Verbyla and Cullis (1990) presented an approach for the, analysis of the repeated measures data, when an additional level of dependence viz., spatial correlation also was incorporated in the model. They used residual maximum likelihood method for estimation purposes.

2.3. Multivariate analysis of variance

Steel (1955) and Cole and Grizzle (1966) proposed multivariate analysis of variance thereby eliminating the problems of univariate analysis of variance model. According to them, this procedure provides a unified approach to the analysis of repeated measures data with all the power, scope and flexibility of the univariate analysis of variance.

Danford *et al.* (1960) studied an experiment involving repeated measures data. They indicated that if the compound symmetry condition is not satisfied, the use of multivariate approach is suggestive.

Davidson (1972) reported that the multivariate tests were nearly as powerful as the univariate tests when the sample sizes exceeded the number of variables by at least 20.

Mendoza *et al.* (1974) found that the multivariate tests appeared more powerful than univariate tests but had a higher Type I error rate. LaTour and Miniard (1983) and 0' Brien and Kaiser (1985) also recommended the multivariate approach to repeated measures analysis.

Moser *et al.* (1990) discussed the issues involved in analyzing the repeated measures data. They discussed the model construction, univariate versus multivariate solution, statistical assumptions behind the models and criteria for selecting a particular approach. The authors concluded that although no single approach is consistently best, the multivariate approach is always appropriate and provides the same interpretation as the univariate approach. They also pointed out that when appropriate assumptions such as spherecity is met, univariate analysis is more suitable for such data.

3. MATERIALS AND METHODS

Considering the array of methods and their applications discussed in the literature, certain selected methods were compared in order to understand their appropriateness in analysing two sets of data that were available. A description of the methods and the data used are given below.

3.1. Statistical analysis

Consider the situation where each of n individuals are observed on p occasions. The combined vector of responses of order (np x 1) may be denoted by **y**. Assume that **y** follows the model.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{1}$$

where **X** is the (np x p) design matrix, $\boldsymbol{\beta}$ is a px1 vector of unknown regression parameters and **e** is a vector of random errors with

$$E(\mathbf{e}) = \mathbf{0}, D(\mathbf{e}) = \mathbf{V} \tag{2}$$

where D stands for dispersion operator and E for expectation.

Let the observations be grouped by individuals and the repeated measurements on an individual be ordered by space or time. Then the dispersion matrix of **e** will have the following structure

$$D(\mathbf{e}) = \mathbf{V} = \text{diag}(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \Sigma n) \tag{3}$$

where Σ_i = covariance matrix of errors of repeated measurements for the ith individual, diag $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \Sigma_n)$ indicates a block diagonal matrix with Σ_i's in the principal axis and null matrices in the other places. Further let $\Sigma_i = \Sigma$ for i =1, …, n. The matrix Σ can take several forms. In the more restrictive case, Σ can be represented as

$$\Sigma = \sigma^2 \mathbf{I}_\mathbf{p} + \sigma^2_\alpha \mathbf{J}_\mathbf{p}$$

where $\mathbf{I_p}$ is a p x p identity matrix and $\mathbf{J_p}$ is a p x p matrix with all elements one. Σ in this case is said to have compound symmetry or Σ is said to be structured. A test of compound symmetry is given by Rouanet and Lepine (1970).

The simplest form for the Σ is one that arises from independent observations of constant variance, i. e.,

$$\Sigma = \sigma^2 \mathbf{I_p} \tag{5}$$

Σ in this case is said to follow the sphericity assumption. A test of sphericity due to' Mauchly (1940) can be performed which tests whether Σ has the structure given in equation (5).

If the variance-covariance matrix of the repeated measurements on individuals satisfy condition (5), then the data can be analysed through an ordinary two-way analysis of variance (ANOVA). This condition will be indicated by the nonsignificance of the compound symmetry and sphericity tests. If the variance-covariance matrix satisfies condition (4) and does not satisfy the condition (5) as portended by the nonsignificance of compound symmetry test and significance of sphericity test, then the data can be analysed through univariate mixed model ANOVA. If the variance-covariance matrix does not satisfy (4) as implied by the significance of the compound symmetry test, *i.e.,* is unstructured, then the data has to be analysed through multivariate analysis of variance (MANOVA).

The multivariate approach considers the measurements on a subject to be a sample from a multivariate normal distribution and makes no assumption about the characteristics of the variance-covariance matrix and it is always a legitimate procedure to adopt. The univariate approach requires certain assumptions about the variance-covariance matrix. If these conditions are met, especially for small sample sizes, the univariate approach is more powerful than the multivariate approach. However, Rouanet and Lepine (1970) pointed out that the multivariate approach is less powerful when sphericity holds.

The details of the three methods referred above are given below under a simplified observational set up.

The general layout here is that of n individuals x p occasions with the individuals divided into G groups of size n_g (g = 1, 2, ..., G). Let the hypothesis to be tested involve a comparison among the groups.

The structure of two-way ANOVA, univariate mixed model ANOVA and one-way MANOVA shall take the following forms :

Two - way ANOVA

The model for the two-way classification with interaction is

$$y_{gjk} = \mu + \alpha_g + \beta_j + \gamma_{gj} + e_{gjk} \tag{6}$$

where y_{gik} is the observation on kth individual in the gth level of factor group andjth level of factor occasion; g=1, ..., G, j=1, ..., p, k=l, ..., n_g . In model (6), μ is the general mean, α_g is the effect of gth level of factor group, β_j is the effect ofjth level of factor occasion, γ_{gj} is the interaction effect for the gth level of factor group andjth level of factor occasion and e_{gjk} is the random error component which is independently and normally distributed with mean zero and variance σ_w^2. In the model, α_g 's and β_j's are assumed to be fixed.

Let $y_{g..}$ denote the total of all observations under the gth level of factor groups, $y_{.j.}$ denote the total of all observations under thejth level of factor occasion, $y_{gi.}$ denote the total of all observations in the (*gj*)th cell, y,,, denote the grand total of all the observations. These notations are expressed mathematically as

$$y_{g..} = \sum_j \sum_k y_{gjk}\ ,\ y_{.j.} = \sum_g \sum_k y_{gjk}\ ,\ y_{gj.} = \sum_k y_{gjk}\ ,\ y_{...} = \sum_g \sum_j \sum_k y_{gjk}$$

The two-way ANOVA is shown in the following.

Sources of variation	Degrees of freedom	Sum of squares	Mean sum of squares	F- ratio
Groups	G-1	SS_G	$MS, = \frac{SS_G}{G-1}$	$\frac{MS_G}{MS_E}$
Occasions	p-1	SS_O	$MS, = \frac{SS_O}{p-1}$	$\frac{MS_O}{MS_E}$
Occasions X Groups	(G-1) (p-1)	SS_{OG}	$MS_{OG} = \frac{SS_{OG}}{(G-1)(p-1)}$	$\frac{MS_{OG}}{MS_F}$
Error	$p\sum_g (n_g - 1)$	SS_E	$MS_E = \frac{SS_E}{p\sum (n_g - 1)}$	
Total	$p\sum_g n_g - 1$	SS_T		

The computational formulae for the sum of squares in the above table are as follows,

$$SS_T = \sum_g \sum_j \sum_k y_{gjk}^2 - \frac{y_{...}^2}{p\sum_g n_g}$$

$$SS_G = \sum_g \frac{y_{g..}^2}{pn_g} - \frac{y_{...}^2}{p\sum_g n_g}$$

$$SS_O = \sum_j \frac{y_{.j.}^2}{\sum_g n_g} - \frac{y_{...}^2}{p\sum_g n_g}$$

$$SS_{OG} = \sum_g \sum_j \frac{y_{gj}^2}{n_g} - \sum_g \frac{y_{g...}^2}{pn_g} - \sum_j \frac{y_{.j.}^2}{\sum_g n_g} + \frac{y_{...}^2}{p\sum_g n}$$

$$SS_E = SS_T - SS_G - SS_O - SS_{OG}$$

Univariate mixed model ANOVA

The model used is

$$y_{gjk} = \mu + \alpha_g + e_{gj} + \beta_j + \gamma_{gj} + e_{gjk} \tag{7}$$

where y_{gjk}, μ, α_g, β_j and γ_{gj} are same as defined in the case of two-way ANOVA ; $g=1, \ldots, G, j=1, \ldots, p, k=1, \ldots, ng$. In model (7),the random component e_{gj} is assumed to be independently and normally distributed with mean zero and variance σ_e^2 and e_{gjk} is the random error component which is also assumed to be independently and normally distributed with mean zero and variance σ_w^2. In the model, α_g 's and β_j's are assumed to be fixed. The notations $y_{g..}$, $y_{.j.}$, $y_{gj.}$ and $y_{...}$ are same as defined in the two-way ANOVA.

The univariate mixed model ANOVA is shown below.

Sources of variation	Degrees of freedom	Sum of squares	Mean sum of squares	F-ratio
Groups	G-1	SS_G	$MS_G = \frac{SS_G}{G-1}$	$\frac{MS_G}{MS_{E_a}}$
Individuals within groups (Ea)	$\sum_g (n_g - 1)$	SS_{E_a}	$MS_{E_a} = \frac{SS_{E_a}}{\sum_g (n_g - 1)}$	
Occasions	P-1	SS_O	$MS_O = \frac{SS_O}{p-1}$	$\frac{MS_O}{MS_{E_b}}$
Occasions X Groups	(G-1)(p-1)	SS_{OG}	$MS_{OG} = \frac{SS_{OG}}{(G-1)(p-1)}$	$\frac{MS_{OG}}{MS_{E_b}}$
Occasions X individuals within groups (Eb)	$(p-1)\sum_g (n_g - 1)$	SS_{E_b}	$MS_{E_b} = \frac{SS_{E_b}}{(p-1)\sum_g (n_g - 1)}$	
Total	$p\sum_g n_g - 1$	SS_T		

The computational formulae for the sum of squares in the above table *viz.,* SS_T, SS_G, SS_O and SS_{OG} are same as defined in two-way ANOVA. The computational formulae for the sum of squares SS_{E_a} and SS_{E_b} are given below.

$$SS_{E_a} = \sum_g \sum_k \frac{y_{g.k}^2}{p} - \sum_g \frac{y_{g..}^2}{pn_g}$$

where $y_{g.k} = \sum_j y_{gjk}$

$SS_{E_b} = SS_T - SS_G - SS_{E_a} - SS_O - SS_{OG}$

One - way MANOVA

The general model is

$$\mathbf{y}_i = \mu_i + \mathbf{e}_i \qquad \textbf{(8)}$$

for individual i, (i = 1, ..., n) where $\mathbf{y_i}$ is the vector of p measurements on an individual, μ_i is the corresponding mean vector *i.e.*, $\mu_i = E(\mathbf{y}_i)$ and $\mathbf{e_i}$ is a vector of random error with mean **0** and variance-covariance matrices $V(\mathbf{e_i}) = \mathbf{I}$;; thus $\mathbf{I}$; is of order p x p. Note the implication here that $V(\mathbf{y_i}) = V(\mathbf{e_i}) = \mathbf{\Sigma}$ is the same for all i. In the case of an one-way setup, the MANOVA table can be constructed as follows.

One-way MANOVA is shown below.

Sources of variation	Degrees of freedom	S.S.P. matrix
Between groups	G-1	$\mathbf{T_B} = \sum_g n_g(\overline{\mathbf{Y}}_g - \overline{\mathbf{Y}})(\overline{\mathbf{Y}}_g - \overline{\mathbf{Y}})'$
Within groups	$\sum_g (n_g - 1)$	$\mathbf{T}_w = \sum_g \sum_i (\mathbf{Y}_{gi} - \overline{\mathbf{Y}}_g)(\mathbf{Y}_{gi} - \overline{\mathbf{Y}}_g)'$
Total	$\sum_g n_g - 1$	$\mathbf{T} = \mathbf{T_B} + \mathbf{T_W}$

Σ

As in univariate ANOVA, strong differences between groups will be established if $\mathbf{T_B}$ is much larger than $\mathbf{T_W}$, or equivalently if $\mathbf{T_W}$ makes a much smaller contribution to $\mathbf{T}$ than does $\mathbf{T_B}$. The common tests employed are Pillai's trace, Wilks's lambda and Hotelling's trace (Morrison, 1976).

where $y_{g.k} = \sum_j y_{gjk}$

$SS_{E_b} = SS_T - SS_G - SS_{E_a} - SS_O - SS_{OG}$

One - way MANOVA

The general model is

$$y_i = Pi + e_i \quad (8)$$

for individual i, (i = 1, ..., n) where $\mathbf{y}_i$ is the vector of p measurements on an individual, μ_i is the corresponding mean vector *i.e.*, $\boldsymbol{\mu}_i = E(\mathbf{y}_i)$ and $\mathbf{e}_i$ is a vector of random error with mean **0** and variance-covariance matrices $\mathbf{V}(\mathbf{e}_i) = \boldsymbol{\Sigma}$; thus Σ is of order p x p. Note the implication here that $V(\mathbf{y}_i) = \mathbf{V}(\mathbf{e}_i) = \boldsymbol{\Sigma}$ is the same for all i. In the case of an one-way setup, the MANOVA table can be constructed as follows.

One-way MANOVA is shown below.

Sources of variation	Degrees of freedom	S.S.P. matrix
Between groups	G- 1	$\mathbf{T_B} = \sum_g n_g(\overline{\mathbf{Y}}_g - \overline{\mathbf{Y}})(\overline{\mathbf{Y}}_g - \overline{\mathbf{Y}})'$
Within groups	$\sum_g (n_g - 1)$	$\mathbf{T_w} = \sum_g \sum_i (\mathbf{Y}_{gi} - \overline{\mathbf{Y}}_g)(\mathbf{Y}_{gi} - \overline{\mathbf{Y}}_g)'$
Total	$\sum_g n_g - 1$	$\mathbf{T = T_B + T_W}$

$\overline{\mathbf{Y}}$ is the overall mean vector $= \sum_g n_g \overline{\mathbf{Y}}_g / \sum_g n_g$

$\mathbf{Y}_{gi}$ is the vector of observations of *i*th individual in the gth group.

As in univariate ANOVA, strong differences between groups will be established if **TB** is much larger than **Tw,** or equivalently if **Tw** makes a much smaller contribution to **T** than does **Tg.** The common tests employed are Pillai's trace, Wilks's lambda and Hotelling's trace (Morrison, 1976).

Pillai's trace	$V = \text{trace}(\mathbf{T_W}\,\mathbf{T}^{-1})$
Wilks's lambda	$W = \det(\mathbf{T_W})/\det(\mathbf{T})$
Hotellings' trace	$T = \text{trace}(\mathbf{T_B}\,\mathbf{T_W}^{-1})$

The hypothesis that the within subjects variance-covariance matrices are equal across all levels of the between-subjects factor can be examined using the multivariate generalization of Box's M test. It is based on the determinants of the variance-covariance matrices for all between-subjects cells in the design. However, Box's M test is very sensitive to departures from normality. The test seems to be satisfactory when G and p are less than 5 and with each n_g greater than 20 (Crowder and Hand, 1990).

It may be noted that variance-covariance matrix of repeated measurements need not be the same for all individuals or over the different treatment groups. The problem can be alleviated to some extent by resorting to data transformations (Montgomery and Peck, 1982).

The above procedures were tested using two data sets in order to find out the type of Σ matrices that are obtained in practice and the type of analysis required in respective cases. All the above analysis were carried out using 'MANOVA : Repeated Measures' procedure of SPSS described in Norusis (1988) on two data sets which are described in the following. The use of the three methods *viz.*, two-way ANOVA, univariate mixed model ANOVA and MANOVA, is illustrated in the Appendix 1 and the computer programmes used for illustration are given in the Appendix 2.

3.2. Data set I

Data on soil properties reported by Balagopalan (1991) formed the first data set. The data were gathered from five plots of 50 m x 50 m laid out randomly in each of six vegetational types *viz.*, plantations of eucalypt, rubber and teak and natural forests of evergreen, semi-evergreen and moist deciduous types. Each plot was separated from the other by about 50 m within each vegetation type. One soil pit was taken from every plot and samples of soil from 0-15, 15-50 and 50-100 cm layers were collected. The soil poperties studied were gravel, sand, silt, clay, bulk density (BD), particle density (PD), porespace, maximum water holding capacity (WHC), volume expansion, soil pH, loss on ignition, acid insoluble, electrical conductivity (EC), total N, exchangeable ammonium nitrogen, nitrate nitrogen, organic carbon (OC), ferric oxide (Fe_2O_3), aluminium oxide (Al_2O_3), potassium oxide (K_2O), calcium oxide (CaO) and magnesium oxide (MgO). The original data with respect to each property were transformed to appropriate scale using the procedures described by Montgomery and Peck (1982).

3.3. Data set I1

The data on total yield of latex from rubber trees were obtained from a clonal trial conducted at the Central Experimental Station, Chethackal of the Rubber Research Institute of India, Kottayam. The field trial consisted of eight clones of *Hevea brasiliensis* GTI, RRII 161, RRII 162, RRII 168, RRII 174, RRII 176, RRII 177 and RRII 178. The experiment was laid out in a completely randomized design with forty trees per clone. Each tree constituted a replication. The trial was started in 1977 and the observations on the monthly total yield of rubber latex were taken from January 1986 through December 1988. For the purpose of this study, data on annual total yield of latex in three years were used. The data were transformed to logarithmic scale in order to avoid heterogeneity in variances (Montgomery and Peck, 1982)

4. RESULTS AND DISCUSSION

4.1. Data set I

Firstly, residuals of ANOVA (through ordinary least squares) were computed from the data on soil properties measured in soils under different types of vegetations. The effects included in the ANOVA were vegetation, depth and vegetation x depth. Variances of residuals at various depth levels were computed for both original and transformed data set. This was done to see the effect of transformation on homogeneity or otherwise of variances. Homogeneity of variance of residuals was tested through Bartlett's test statistic. The test statistic was found significant in original and transformed data sets in the case of many soil properties (Table 1). Transformation was found to reduce the heterogeneity only in the case of two soil properties *viz.,* PD and EC. The reverse effect was also observed in the case of few soil properties.

Correlation between the residuals at different depth levels was estimated. The above exercise was done both with original and transformed data set to see the effect of transformation on the correlation between depth levels. Significant correlation was obtained between the surface and subsurface layers in the case of many soil properties (Table 2). Transformation was found to break this correlation in many cases as well. Occasional cases of the reverse effect were also observed. Nonsignificant correlation between the residuals of three layers in the transformed scale was observed in the case of soil properties *viz.*, gravel and sand. For gravel and sand, compound symmetry and sphericity statistics became nonsignificant (Table 3). The correlation between the residuals at three layers was found nonsignificant for porespace, for which the compound symmetry statistic alone turned out nonsignificant. Significant correlation between the residuals of three levels was observed in the case of rest of the soil properties, for which the compound symmetry statistic became significant.

Based on the above results, the appropriate method of analysis to be followed in the case of each variable is indicated in Table 3.

Table 1. Variances of the residuals at different depth levels for the soil properties observed in different vegetation types.

Soil property	Transformation used	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	χ_B^2
Gravel	None	5.92 (5.92)	11.69 (11.69)	74.99 (74.99)	49.57** (49.57**)
Sand	y^2	11.81 (458.25)	14.46 (350.25)	77.07 (739.77E+02)	32.37** (207.87**)
Porespace	y^2	6.77 (123.90)	6.87 (127.60)	10.89 (371.50)	2.18ns (12.06**)
Silt	ln(y)	10.28 (0.36)	9.06 (0.24)	6.84 (0.19)	1.21ns (3.00ns)
K_2O	y^{-1}	3.8E-04 (0.00)	4.8E-04 (0.00)	3.4E-04 (0.00)	3.74ns (0.00ns)
Soil pH	$y^{0.5}$	0.05 (0.01)	0.04 (0.01)	0.02 (0.01)	3.51ns (4.38ns)
Clay	$y^{0.5}$	`9.68 (0.29)	13.59 (0.53)	14.29 (0.45)	1.23ns (2.70ns)
BD	y^4	3.2E-03 (0.00)	5.0E-03 (0.00)	6.3E-03 (0.00)	3.40ns (0.00ns)
PD	y^2	1.8E-03 (0.00)	5.2E-03 (0.00)	0.6E-03 (0.00)	29.64** (0.00ns)
WHC	$y^{0.5}$	45.43 (0.83)	13.88 (0.14)	27.23 (0.36)	9.57** (21.32**)
Volume expansion	None	0.35 (0.35)	0.90 (0.90)	0.40 (0.40)	8.11* (8.11*)
Loss on ignition	ln(y)	3.67 (1.01)	131.51 (6.60)	61.98 (6.97)	64.29** (25.81**)
Acid insoluble	$y^{0.5}$	6.66 (0.27)	6.65 (0.35)	2.37 (0.09)	8.79* (13.18**)
EC	$y^{0.5}$	8.3E-04 (0.00)	23.4E-04 (0.00)	3.4E-04 (0.00)	25.31** (0.00ns)
Total N	ln(y)	43.77E+03 (1.59)	43.38E+03 (0.12)	17.44E+03 (0.05)	7.05* (85.09**)
Available N	ln(y)	15.54E+02 (0.11)	36.20E+02 (1.54)	577.48 (0.05)	22.12** (90.83**)
Exchangeable ammonium nitrogen	ln(y)	457.35 (0.47)	901.94 (1.16)	130.14 (0.06)	23.54** (49.41**)
Nitrate nitrogen	ln(y)	99.90 (0.22)	109.02 (1.02)	84.43 (0.12)	0.48ns (36.00**)

(Contd...)

Table 1. Contd.

Soil property	Transformation used	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	χ_B^2
OC	$y^{0.5}$	0.03 (0.01)	0.10 (0.03)	0.03 (0.01)	14.49** (10.93**)
Fe_2O_3	$y^{0.5}$	3.13 (0.27)	3.81 (0.31)	1.66 (0.12)	5.02ns (6.28*)
Al_2O_3	$y^{0.5}$	9.97 (0.43)	9.10 (0.27)	2.52 (0.06)	14.22** (22.70**)
CaO	None	4.1E-04 (4.1E-04)	5.2E-04 (5.2E-04)	0.7E-04 (0.7E-04)	25.11** (25.11**)
MgO	None	2.8E-04 (2.8E-04)	6.2E-04 (6.2E-04)	0.3E-04 (0.3E-04)	49.83** (49.83**)

Note : Figures in parenthesis represent variances in the transformed scale.

** - significant at P = 0.01

* - significant at P = 0.05

ns - non significant

$\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ are estimates of residual variance at the three depth levels starting from the surface

χ_B^2 - Bartlett's test statistic

Table 2. Correlation between the residuals at different depth levels for the soil properties observed in different vegetation types.

Soil property	Transformation used	Original data			Transformed data		
		r_{12}	r_{13}	r_{23}	r_{12}	r_{13}	r_{23}
Gravel	None	0.24	0.24	-0.05	0.24	0.24	-0.05
Sand	y^2	0.12	-0.03	-0.01	0.17	-0.19	0.09
Porespace	y^2	0.47**	-0.26	0.01	-0.10	0.21	-0.32
Silt	ln(y)	0.62**	0.35	0.07	0.84**	0.55**	0.45*
K_2O	y^{-1}	0.07	0.54**	-0.03	0.06	0.54**	-0.04
Soil pH	$y^{0.5}$	-0.03	-0.02	0.19	0.01	-0.03	0.19
Clay	$y^{0.5}$	0.56**	0.39*	0.37*	0.43*	0.28*	0.24
BD	y^4	0.07	0.18	-0.10	-0.16	-0.29	0.02
PD	y^2	-0.02	-0.05	-0.20	0.44*	-0.22	0.01
WHC	$y^{0.5}$	0.21	0.34	-0.13	0.18	0.27	-0.07
Volume expansion	None	-0.20	0.35	0.50**	-0.20	0.35	-0.50**
Loss on ignition	ln(y)	-0.02	0.04	-0.21	0.10	0.07	-0.27
Acid insoluble	$y^{0.5}$	0.43*	0.45*	-0.03	0.50**	0.46*	0.04
EC	$y^{0.5}$	-0.33	-0.11	-0.35	0.37	0.17	0.13
Total N	ln(y)	-0.54**	-0.06	-0.24	-0.44*	0.10	-0.27
Available N	ln(y)	-0.60**	0.32	-0.16	-0.36	0.28	-0.13
Exchangeable ammonium nitrogen	ln(y)	-0.33	0.52**	-0.33	0.06	0.35	-0.14
Nitrate nitrogen	ln(y)	-0.22	0.48**	-0.00	-0.33	0.37*	-0.07
OC	$y^{0.5}$	-0.49**	0.47**	-0.27	-0.46*	0.43*	-0.26
Fe_2O3	$y^{0.5}$	0.83**	0.37*	0.38*	0.84**	0.30	0.34
Al_2O3	$y^{0.5}$	0.41*	0.43*	0.35	0.43*	0.38*	0.38*
CaO	None	0.13	0.41*	-0.03	0.13	0.41*	-0.03
MgO	None	0.14	0.43*	0.06	0.14	0.43*	-0.06

** - significant at P = 0.01, * - significant at P = 0.05

r_{12}, r_{13} and r_{23} - correlation between different layers

Table 3. Results of compound symmetry and sphericity tests with respect to different soil properties

Soil properties	Transformation used	χ_c^2	χ_s^2	Method selected
Gravel	None	2.34(ns)	3.65(ns)	M1
Sand	y^2	4.54E-07(ns)	4.48(ns)	M1
Porespace	y^2	7.14E-07(ns)	7.68*	M2
Silt	ln(y)	350.02**	-	M3
K_2O	y^{-1}	671.15**	-	,,
Soil pH	$y^{0.5}$	531.01**	-	,,
Clay	$y^{0.5}$	210.43**	-	,,
BD	y^4	190.11E+04**	-	,,
PD	y^2	124.79**	-	,,
WHC	$y^{0.5}$	197.14**	-	,,
Volume expansion	None	92.07**	-	,,
Loss on ignition	ln(y)	339.94**	-	,,
Acid insoluble	$y^{0.5}$	326.57**	-	,,
EC	$y^{0.5}$	690.48**	-	,,
Total N	ln(y)	327.60**	-	,,
Available N	ln(y)	352.40**	-	,,
Exchangeable ammonium nitrogen	ln(y)	302.88**	-	,,
Nitrate nitrogen	ln(y)	305.10**	-	,,
OC	$y^{0.5}$	434.40**	-	,,
Fe_2O3	$y^{0.5}$	302.31**	-	,,
Al_2O3	$y^{0.5}$	265.82**	-	,,
CaO	None	656.53**	-	,,
MgO	None	665.69**	-	,,

** - significant at P=0.0l

* - significant at P=0.05

ns - non significant

χ_c^2 - compound symmetry test statistic, χ_s^2 - sphericity test statistic

M1 - two-way ANOVA, M2 - univariate mixed model ANOVA, M3 -

4.2. Data set II

Residuals of ANOVA (through ordinary least squares) were computed from the data on total yield of latex from rubber trees in three years. The sources of variations included in the ANOVA were clones, years and clones x years. Variances of residuals at each year for both original and transformed data were computed. This was done to see the effect of transformation on homogeneity or otherwise of variances. Bartlett's test for homogeneity of variance was used for this purpose. The test statistic was found significant in original as well as in transformed data (Table 4). Thus transformation could not avoid the heterogeneity of variance of residuals of different years.

Correlation between the residuals at different years was computed. The above exercise was done both with original and transformed data set to see the effect of transformation on the correlation between residuals at different years. Significant correlation was observed between years in original as well as in transformed data (Table 5). Transformation could not break the corrleation between the residuals at different years.

The compound symmetry statistic was found significant when there were significant correlations between the residuals at three years. Based on the above result, the appropriate method of analysis to be followed is indicated in Table 6.

Original data				Transformed data [ln(y)]			
$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	σ_3^2	χ_B^2	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_3^2$	χ_B^2
39.78E+02	62.13E+02	20.73E+03	56.63**	0.04	0.06	0.87	213.13**

$\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ are estimates of residual variance of different years

χ_B^2 - Bartlett's test statistic

** - significant at P = 0.01

Table 5. Correlation between the residuals at different years for annual yield of latex from rubber trees of different clones.

Original data			Transformed data [ln(y)]		
r12	r13	r23	r12	r13	r23
0.66**	0.39**	0.59**	0.65**	0.49**	0.63**

** - significant at P = 0.01

r12,r13,r23 - correlation between three years

Table 6. The results of the compound symmetry and sphericity tests with respect to data on total yield of rubber latex (transformed data).

χ_c^2	χ_s^2	Method selected
351.49**	-	M3

** - significant at P= 0.01, χ_c^2 - Compound symmetry test statistic, χ_s^2 - Sphericity test statistic, M3 - MANOVA

Although the three methods specified here serve to test similar hypothesis, there are subtle differences in the interpretation of the results of each method. Both the univariate procedures portend similar'effects although the testing procedures for the effects are slightly different. The interaction effect *viz.*, between subjects factor x within subject factor is extracted as a separate component in addition to the main effects. In multivariate analysis, this interaction effect is implicit in the testing of the main effects. For instance, in the present case the different vegetations are compared with respect to any particular soil property at different layers simultaneously by treating the values at different depths as a vector admitting the possible correlations between depths at the same time.

5. CONCLUSIONS

The study on soil properties observed from multiple core samples from 0 - 15, 15-50 and 50-100 cm layers under six different types of vegetation indicated the following. The data pertaining to the soil properties *viz.*, gravel and sand can be analysed through ordinary two-way ANOVA; porespace can be analysed through univariate ANOVA; silt, clay, BD, PD, WHC, volume expansion, loss on ignition, acid insoluble, EC, total N, available N, exchangeable ammonium nitrogen, nitrate nitrogen, OC, K_2O, soil pH, Fe_2O_3, Al_2O_3, CaO and MgO can be analysed through MANOVA. Since majority of the soil properties required use of MANOVA, this analysis is recommended for data sets of the type described here. This was found justifiable because the extent of residual variation in individual soil properties at different depths was not of the same order and also the correlations between values at different layers were not of the same magnitude.

The residual variation in latex yield in different years were found significantly different and also the correlations between latex yields in different years were different implying the use of MANOVA for analysing data of such nature.

6. REFERENCES

Andersen, A. H., Jensen, E. B. and Schou, G. 1981. Two-way analysis of variance with correlated errors. *International Statistical Review,* 49, 153-167.

Arnold, S. F. 1981. *The Theory of Linear Models and Multivariate Analysis.* Wiley, New York.

Balagopalan, M. 1991. *Studies on Soil Organic Matter.* Ph. D. thesis. University of Calicut, Calicut, 247p.

Cole, J. W. L. and Grizzle, J. E. 1966. Application of multivariate analysis of variance to repeated measurements experiments. *Biornetrics,* 22, 810-828.

Crowder, M. J. and Hand, D. J. 1990. *Analysis of Repeated Measures.* Chapman and Hall, New York. 257p.

Cullis, B. R. and McGilchrist, C. A. 1990. A model for the analysis of growth data for designed experiments. *Biornetrics,* 46, 131-142.

Danford, M. B., Hughes, H. M. and McNee, R. C. 1960. On the analysis of repeated measurements experiments. *Biornetrics,* 16,547-565.

Davidson, M. L. 1972. Univariate versus multivariate tests in repeated-measures experiments. *Psychological Bulletin,* 77, 446-452.

Diggle. P. J. and Donnelly, J. B. 1989. A selected bibliography on the analysis of repeated measurements and related areas. *Australian Journal of Statistics,* 31(l), 183-193.

Gill, J. L. 1986. Repeated measurements : sensitive test for experiments with few animals. *Journal of Animal Science,* 63,943-954.

Gill, J. L. 1988. Repeated measurements. Split-plot trend analysis versus analysis of first differences. *Biometrics,* 44, 289-297.

Khosla, R. K., Rao, P. P. and Das, M. N. 1979. A note on the study of experimental errors in groups of agricultural field experiments conducted in different years. *Journal of Indian Society of Agricultural Statistics,* 31,65-68.

Koch, G. G., Amara, I. A., Stokes, M. E. and Gillings, D. B. 1980. Some views on parametric and non-parametric analysis for repeated measurements and selected bibliography. *International Statistical Review,* 48,249-265.

LaTour, S. A. and Miniard, P. W. 1983. The misuse of repeated measures analysis in marketing research. *Journal of Marketing Research,* 20,45-57.

Mauchly, J. W. 1940. Significance test for sphericity of a normal n-variate distribution. *Annals of Mathematical Statistics,* 29, 204-209.

Mendoza, J. L., Toothaker, L. E. and Nicewander, W. A. 1974. A Monte Carlo comparison of the univariate and multivariate methods for the groups by trials repeated measures design. *Multivariate Behavioural Research,* 9, 165-178.

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

Morrison, D. F. 1976. *Multivariate Statistical Methods.* McGraw Hill, New York.

Moser, E. B., Saxton, A. M. and Pezeshki, S. R. 1990. Repeated measures analysis of variance : application to tree research. *Canadian Journal of Forest Research,* 20, 524-535.

Norusis, M. J. 1988. SPSS/PC + Advanced Statistics Version 2.0 for the IBM PC/XT/AT and PS/2. SPSS Inc. Chicago.

O' Brien, R. G. and Kaiser, M. K. 1985. MANOVA method for analyzing repeated measures designs : an extensive primer. *Psychological Bulletin*, 97, 316-333.

Patterson, H. D. 1939. *Statistical Techniques in Agricultural Research.* McGraw Hill and Co., New York.

Rouanet, H. and Lepine, D. 1970. Comparison between treatments in a repeated-measures design : ANOVA and multivariate methods. *British Journal of Mathematical and Statistical Psychology,* 23, 147-163.

Rowell, J. G. and Walters, D. E. 1976. Analysing data with repeated observations on each experimental units. *Journal of Agricultural Science,* Camb, 87,423-432.

Steel, R. D. G., 1955. An analysis of perennial crop data. *Biometrics,* 11, 201-212.

Verbyla, A. P. and Cullis, B. R. 1990. Modelling in repeated measures experiments. *Applied Statistics,* 39(3), 341-356.

Ware, J. H. 1985. Linear models for the analysis of longitudinal studies. *The American Statistician,* 39(2), 95-101.

Yates, F. and Cochran, W. G. 1938. The analysis of groups of experiments. *Journal of Agricultural Science,* 28, 556-580.

7. APPENDICES

Appendix 1. Illustration

The use of the three methods *viz.*, two-way ANOVA (Ml), univariate mixed-model ANOVA (M2) and MANOVA (M3), is illustrated using data on soil property *viz.*, silt content taken from the data set I. The corresponding ANOVA tables are given in Table Al.

Table Al. Illustrations of methods MI, M2 and M3 using data on silt content in logarithmic scale.

Two-way ANOVA

Source of variation	Sum of squares	df	Mean sum of squares	F- ratio
Vegetation	5.01	5	1.00	46.77**
Depth	0.09	2	0.05	2.19ns
Vegetation X depth	0.25	10	0.03	1.16ns
Error	1.54	72	0.02	--

Univariate mixed model ANOVA

Source of variation	Sum of squares	df	Mean sum of squares	F- ratio
Vegetation	5.01	5	1.00	23.08**
Individuals within vegetation (E_a)	1.04	24	0.04	--
Depth	0.09	2	0.05	4.50*
Vegetation x depth	0.25	10	0.03	2.37*
Depth x individuals within vegetation (E_b)	0.50	48	0.01	--

** - significant at P = 0.01
* - significant at P = 0.05 ns - nonsignificant

MANOVA

Statistic	Value of the test statistic	Approx. F	Hyp.df	Error df	Prob. of F
Pillai's trace	1.25	3.40	15	72	<0.05
Hotelling's trace	7.97	10.98	15	62	<0.05
Wilks's lambda	0.08	6.20	15	61	<0.05

Appendix 2. Computer Programmes

Computer programmes used for illustration are given below.

Two-way ANOVA

```
Set disk = 'Temp.rst'.
Translate from 'Dat.dbf'.
Compute silt = Ln(silt+l0).
Anova variables = silt by veg(1,6) depth(1,3)
/option=4.
finish.
```

Univariate mixed model ANOVA

```
Set disk = 'Temp1.rst'.
Translate from 'Dat.dbf'.
Compute silt = Ln(silt+l0).
Manova silt by veg(1,6) depth(1,3) ind(1,5)
/method = sstype(unique) modeltype(observations) estimation(qr)
/design= ind within veg=l veg vs 1 depth vs within+residual veg by depth vs
within+residual.
Finish.
```

MANOVA

```
Set disk = 'Temp2.rst'.
Translate from 'Dat1.dbf'.
Compute siltl = Ln(siltl+l0).
Compute silt2 = Ln(silt2+10).
Compute silt3 = Ln(silt3+10).
Manova siltl silt2 silt3 by veg(l,6)
/design.
```

The structure of the Dat.dbf is :

Field	Field name	Type	Width	Dec
1	VEG	Numeric	3	
2	DEPTH	Numeric	3	
3	SILT	Numeric	7	2
Total			14	

The structure of the Datl. dbf is :

Field	Field Name	Type	Width	Dec
1	VEG	Numeric	3	
2	DEPTH	Numeric	3	
3	IND	Numeric	3	
4	SILT1	Numeric	6	2
5	SILT2	Numeric	6	2
6	SILT3	Numeric	6	2
Total			28	

VEG - vegetation codes ; DEPTH - depth codes ; SILT - silt content ;
IND - individual codes ; SILT1, SILT2 and SILT3 - silt content at three depth levels starting from the surface.