Embrapa
Empresa Brasileira de Pesquisa Agropecuária
Centro Nacional de Pesquisa de Florestas
Ministério da Agricultura, Pecuária e Abastecimento

ISSN 1517-536X
Dezembro, 2003

Documentos

Multivariate Spatial Statistical Analysis of Multiple Experiments and Longitudinal Data

Marcos Deon Vilela de Resende
Robin Thompson

Colombo, PR
2003

Supervisor editorial: Luciano Javier Montoya Vilcahuman
Normalização bibliográfica: Lidia Woronkoff
Foto(s) da capa: Paulo Gonçalves (IAC)
Editoração eletrônica: Marta de Fátima Vencato

1ª edição
1ª impressão (2003): 300 exemplares

CIP – Brasil. Catalogação na Publicação
Embrapa Florestas

Resende, Marcos Deon Vilela de.
Multivariate spatial statistical analysis of multiple experiments and longitudinal data / Marcos Deon Vilela de Resende, Robin Thompson. – Colombo: Embrapa Florestas, 2003.
126 p. (Embrapa Florestas. Documentos, 90).

ISSN 1517-526X

1. Análise estatística. 2. Planta - Melhoramento. 3. Seleção – Método. 4. REML. 5. BLUP. I. Thompson, Robin. II. Título. III. Série.

CDD 519.535 (21. ed.)

Autores

Marcos Deon Vilela de Resende
Estatistico, Doutor, *Embrapa Florestas,* Estrada da Ribeira, km111 - Colombo,PR, deon@cnpf.embrapa.br.

Robin Thompson
Matematico, Doutor, Biomathematics Unit, Rothamsted Research, Harpenden AL5 2JQ, England, robin.thompson@bbsrc.ac.uk.

Preface

This document reports the work undertaken by Dr. Marcos Deon Vilela de Resende whilst a Fellow of the Rothamsted Research Institute as a Post Doctoral Scientist in the Biomathematics Unit, under the guidance of Dr. Robin Thompson and with financial support of the referred institute, located in London, England. The research project entitled "Spatial Analysis in Perennial Crops" concerned with adapting and extending statistical models for efficient analysis of field experiments. The analytical procedures described are based on the REML method for variance components mixed models. The REML method was invented and improved by Dr. Robin Thompson and co-authors and nowadays is the standard procedure for statistical analysis in a great range of applications. In agricultural field trials the REML method replaced the traditional method of analysis of variance (ANOVA), providing more accurate estimates and predictions.

Chapter 1 considers the spatial statistical analysis of longitudinal data or repeated measures. Practical experiments with several perennial plants generate annually a large amount of data on repeated measures. Improved methods for analysis of such kind of data were developed which will lead to higher efficiency of scientific research in this field.

Chapter 2 considers the analysis of multi-environment trials through the factor analytic multiplicative mixed models (FAMM) which present several advantages over the traditional additive main effects and multiplicative interaction analysis (AMMI). The FAMM models allow for variance heterogeneity, correlated errors

within trials and unbalancing. In addition, provide BLUP of treatments effects, easy choice of the number of multiplicative terms needed and estimates of the full correlation structure among environments.

Chapter 3 deals with competition among plants and its influences on statistical inference from field trials. Several alternative modelling approaches were evaluated for joint consideration of competition and environmental trend or spatial effects. Improved models were found for routine of data analysis in annual and perennial plants.

Several plant species are of great economic and social importance in Brazil. Scientific experiments with these plants are designed with the objectives of providing new technologies which will contribute for the enhancement of production and productivity of the crops. These enhancements will contribute for the economic and social development of the country as well as for the environmental conservation as a result of a reduced pressure over the natural resources. The plant breeding programmes in the country produce annually a huge amount of field data which need to be statistically analysed in an efficient way. In this context, optimal statistical methodology is essential in transforming data in useful scientific information for the rural development. In this sense, the research reported here will bring great impact.

Vitor Afonso Hoeflich
Chefe Geral da *Embrapa Florestas*

Contents

Multivariate Spatial Statistical Analysis of Multiple Experiments and Longitudinal Data

Marcos Deon Vilela de Resende
Robin Thompson

1. Multivariate Spatial Statistical Analysis of Longitudinal Data

1.1 Introduction

Traditional analysis of agricultural field trials considers measures taken from adjacent plants or plots as non-correlated and the spatial positions of the observations are ignored. Hence, the residual covariance matrix is modelled as a diagonal one, with errors assumed as independents. However, the spatial dependence does exist and contributes to the increasing of the residual variance, in a way that is relevant to consider it in the analysis of trials by approaching the correlated error structure through adequate models.

According to Fisher (1925) and Steel and Torrie (1980), randomisation of treatment plots across replications can provides neutralisation of the effects of spatial correlation, leading to a valid analysis of variance. However, although the randomisation theory emphasises this kind of neutralisation, that is more efficient when spatial models are used. Besides, the local control schemes relying on blocking can be inefficient in accounting of all environmental gradients and trends and even the incomplete blocks do not provide a complete evaluation of the environmental effects. Once blocking is made before the establishment of the trials, the presence of patchy and environmental gradients within blocks is frequently observed (mainly in perennial crops) by the occasion of the data

collecting. This reveals that blocks were not adequately designed a priori. In such a situation, only the spatial analysis techniques can circumvent the estimation problems and provide efficient analysis.

The main procedures aimed at the control and account of spatial correlation among neighbouring observations are **time series models** (Gleeson and Cullis, 1987; Martin, 1990; Cullis and Gleeson, 1991; Gilmour, Cullis and Verbyla, 1997; Gilmour et al., 1998; Cullis et al. 1998; Smith, Cullis and Thompson, 2001) using ARIMA models and REML estimates of variance components (Cooper and Thompson, 1977) and **geostatistical models** (Cressie, 1993; Grondona and Cressie, 1991; Zimmerman and Harville, 1991).

The time series models were first used by Gleeson and Cullis (1987) that considered the errors through an autoregressive integrated moving average process (ARIMA (p, q, d)) in one direction. This model was considered inefficient and Martin (1990) and Cullis and Gleeson (1991) extended such model to two directions: rows and columns. The extended model is of the form ARIMA (p_1, d_1, q_1) x ARIMA (p_2, d_2, q_2). This class of models is called error in variables and account for a tendency effect (ξ) plus an independent error η. In annual crops experiments and in knowledge absence of the correct correlation structure, Gilmour, Cullis and Verbyla (1997) suggested the modelling of ξ as a first order separable autoregressive process (AR1 x AR1). This auto-regressive process in two directions has shown efficiency in a gamma of situations (Grondona et al., 1996; Gilmour, Cullis and Verbyla, 1997; Cullis et al., 1998; Apiolaza, Gilmour and Garrick, 2000; Gilmour, 2000; Qiao et al., 2000; Costa e Silva et al., 2001; Resende and Sturion, 2001; Smith, Cullis and Thompson, 2001; Stringer and Cullis, 2002; Resende, Thompson and Welham, 2003). A process (AR1 x AR1) is flexible and permits to model local and global tendencies as well as extraneous variations, taking into account the three major sources of spatial variation, according to Gilmour, Cullis and Verbyla (1997). The ARIMA methods of Gleeson and Cullis (1987), Martin (1990) and Cullis and Gleeson (1991) encompass the nearest-neighbour (NN) methods of Papadakis (1937), the Papadakis' iterated NN method (Papadakis, 1970; Bartlett, 1978) and other previous methods (Papadakis, 1984; Bartlett, 1938; Atkinson, 1969; Wilkinson et al., 1983; Green et al., 1985; Besag and Kempton, 1986; Williams, 1986) of neighbour analysis.

The geo-statistical procedures consider directly the spatial heterogeneity through the inclusion of the tendency effects and error correlation in modelling the residual covariance matrix (Duarte, 2000). They search for a general covariance function estimate, which is used directly in estimation and prediction procedures. They permit the evaluation of the spatial variability pattern in the experimental area through the adjusted semivariance matrix. This matrix is used as weighting in the generalised least square equations. The semivariance matrix can be adjusted by several models such as spherical, exponential and Gaussian. The standard model for fitting a function to the experimental variogram in field trials is the exponential one. Grondona and Cressie (1991) and Cressie (1993) tried to fit several classes of models to the experimental variogram in several field trials and concluded that none achieved a better fit (according to the weighted-least-square criterion) than the exponential model. Other results based on experimental data have also shown that the exponential spatial model mostly explained the sample variogram (Joyce et al., 2002). And because the covariogram model is exponential, residuals can be interpreted as a realisation of a first-order autoregressive process. This makes sense since the AR1 model projects the autocorrelation to distant lags as a power function of the distance apart. The exponential model does about the same. According to Gleeson (1997), the geostatistical approach of Zimmerman and Harville (1991) called random field linear model is equivalent to fitting separable ARIMA processes and according to Gilmour, Thompson and Cullis (1995) it is equivalent to a first order separable autoregressive process (AR1 x AR1) without the independent error. However, the geostatistical models are often isotropic and Cullis and Gleeson (1991) and Gilmour, Cullis and Verbyla (1997) have shown that anisotropic models are often preferred for modelling the variance structure in field trials. Furthermore, the assumption of separability results in significant savings in computer time. Based on these facts, the ARIMA time series models should be preferred as they encompass all the other main approaches.

The reports attesting the efficiency of spatial analysis are referring to annual crops or forest trees with a single measure per plant. The spatial analysis concerning to repeated measure data or multivariate data on each plant has not been accounted yet, despite the great number of crops that generates a large amount of this sort of data. Perennial plants are of extreme importance in several tropical and subtropical parts of the world and include crops such as coffee, tea, cashew, coconut, cocoa, rubber tree, oil palm, sugar cane and several fruit trees and forage plants. The application of spatial analysis in these categories of

plants involves modelling of the several random effects through different structures for each one, providing an adequate account of the repeated measures together with the spatial variation. Also, these repeated measures are commonly obtained from multi-environment trials and analyses involving several repeated measures in several sites with spatial variation are demanded.

In modelling longitudinal or repeated measures data arising from perennial individuals, several approaches can be used such as repeatability, multivariate, random regression, spline, character process and ante-dependence models. The simplest (repeatability) and the more complete and parameterised (multivariate) models are not likely to be useful in practice. Parsimonious approaches such as random regression or covariance functions (Kirkpatrick, Hill and Thompson, 1994), smoothing cubic splines (White, Thompson and Brotherstone, 1999; Verbyla et al., 1999), character process models (Jaffrezic, White and Thompson, 2003) and structured ante-dependence models (Jaffrezic et al., 2002) should be tried for the sake of practical efficiency. Character process and structured ante-dependence models have proved efficiency in a number of situations (Jaffrezic et al., 2002).

We analysed a total of 26,370 observations from 3 trials of tea plant concerning to 8 yield annual measures, through different spatial and non-spatial models. The classes of methods applied were: (1) univariate spatial models for individual annual measures on each trial; (2) longitudinal non-spatial models for the several measures on each trial; (3) longitudinal and spatial models simultaneously for repeated measures in each trial. These situations are mandatory in any breeding program of a perennial crop and all data should be analysed simultaneously in sake for maximum efficiency of the improvement program. The adequate modelling and computing are critical for obtaining reliable estimates and satisfactory practical results.

1.2 Description of Models

1.2.1 General Linear Mixed Model and REML Estimation

A general linear mixed model has the form (Henderson, 1984; Searle et al. 1992; Gilmour et al., 2002; Thompson, 2002; Thompson et al., 2003):

$$y = X\beta + Z\tau + \varepsilon \qquad (1)$$

with the following distributions and structures of means and variances:

$$\tau \sim N(0,\ G) \qquad E(y) = X\beta$$

$$\varepsilon \sim N(0,\ R) \qquad Var(y) = V = ZGZ' + R$$

where:

y: known vector of observations.

β: parametric vector of fixed effects, with incidence matrix X.

τ: parametric vector of random effects, with incidence matrix Z.

ε: unknown vector of errors.

G: variance-covariance matrix of random effects.

R: variance-covariance matrix of errors.

0: null vector.

Assuming G and R as known the simultaneous estimation of fixed effects and the prediction of the random effects can be obtained through the mixed model equations given by:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \tilde{\tau} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

The solution to this system of equations for $\hat{\beta}$ and $\tilde{\tau}$ leads to identical results as that obtained by:

$\hat{\beta} = (X'V^{-1}X)^{-}X'V^{-1}y$: generalised least square estimator (GLS) or best linear unbiased estimator (BLUE) of β;

$\tilde{\tau} = GZ'V^{-1}(y - X\hat{\beta}) = C'V^{-1}(y - X\hat{\beta})$: best linear unbiased predictor (BLUP) of τ ; where $C' = GZ'$: covariance matrix between τ and y.

When G and R are not known, the variance components associated can be estimated efficiently through the REML procedure (Patterson and Thompson, 1971; Searle et al., 1992; Thompson, 1973; 1977; 1980; 2002; Thompson and Welham, 2003; Cullis et al. 2004). Except for a constant, the residual likelihood function (in terms of its log) to be maximised is given by:

$$L = -\frac{1}{2}\ (\log|X'V^{-1}X| + \log|V| + v\ \log\ \sigma_\varepsilon^2 + y'Py/\sigma_\varepsilon^2)$$

$$= -\frac{1}{2}\ (\log|C^*| + \log|R| + \log|G| + v\ \log\ \sigma_\varepsilon^2 + y'Py/\sigma_\varepsilon^2)$$

where:

$V = R + ZGZ'$; $\quad P = V^{-1} - V^{-1}X\ (X'V^{-1}X)^{-1}X'V^{-1}$.

$v = N\text{-}r(x)$: degrees of freedom, where N is the total number of data and $r(x)$ is the rank of the matrix X.

C^*: Coefficient matrix of the mixed model equations.

Being general, the model (1) encompasses several models inherent to different situations such as:

Univariate model

$G = A\sigma_\tau^2$; $\quad R = I\sigma_\varepsilon^2$, where:

σ_τ^2: variance of the random effects in τ.

A: known matrix of relationships between the τ elements.

σ_ε^2: residual variance.

Repeated measures model including permanent effects (p)

$$y = X\beta + Z\tau + \varepsilon$$

$$= X\beta + Z_1\tau^* + Z_2\rho + \varepsilon,\ \textit{where}:$$

$Var(\tau^*) = A\sigma_\tau^2$; $\quad Var(\rho) = I\sigma_\rho^2$; $\quad R = I\sigma_\varepsilon^2$

σ_ρ^2: variance of permanent effects.

Multivariate models

In the bivariate case:

$$Z = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}; \quad \tau = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix};$$

$$G = A \otimes G_O; \quad R = I \otimes R_O;$$

$$G_O = \begin{bmatrix} \sigma^2_{\tau_1} & \sigma_{\tau_{12}} \\ \sigma_{\tau_{21}} & \sigma^2_{\tau_2} \end{bmatrix}; \quad R_O = \begin{bmatrix} \sigma^2_{\varepsilon_1} & \sigma_{\varepsilon_{12}} \\ \sigma_{\varepsilon_{21}} & \sigma^2_{\varepsilon_2} \end{bmatrix} \quad or \quad R_O = \begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 \\ 0 & \sigma^2_{\varepsilon_2} \end{bmatrix}, \; where:$$

$\sigma_{\tau_{12}}$: random treatment effects covariance between variables 1 and 2.
$\sigma_{\varepsilon_{12}}$: residual covariance between variables 1 and 2.

Spatial models (time series or geostatistical)

$R = \Sigma$: non-diagonal matrix that considers the correlation between residuals through ARIMA models or covariance based on adjusted semivariance.

1.2.2 Univariate Spatial Models for Individual Annual Measures on each Trial

In the context of the agricultural experiments, the general spatial model developed by Martin (1990) and Cullis and Gleeson (1991) has the following form:

$y = X\beta + Z\tau + \xi + \eta$, where:

y: known vector of data, ordered as columns and rows within columns;
τ: unknown vector of treatment effects;
β: unknown vector representing the spatial variation at large scale or global tendency (block effects, polynomial tendency);
ξ: unknown vector representing the spatial variation at small scale (within blocks) or local tendency, modelled as a random vector with zero mean and spatially dependent variance;
η: unknown vector of independent and identically distributed errors.

Through ARIMA models, the error is modelled as a function of a tendency effect (ξ) plus a non correlated random residual (η). So, the vector of errors is partitioned into $\varepsilon = \xi + \eta$, where ξ and η refer to the spatially correlated and independent errors, respectively. The traditional models of analysis do not include the ξ component.

Considering an experiment with rectangular shape in a grid of c columns and r rows, the residuals can be arranged in a matrix in a way that they can be considered as correlated within columns and rows. Writing this residuals in a vector following the field order (by putting each column beneath another), the variance of residuals is given by $Var(\varepsilon) = Var(\xi + \eta) = R = \Sigma = \sigma_{\xi}^{2}[\sum_{c}(\Phi_{c}) \otimes \sum_{r}(\Phi_{r})] + I\sigma_{\eta}^{2}$, where σ_{ξ}^{2} is the variance due to local tendency and σ_{η}^{2} is the variance of the independent residuals.

The matrices $\sum_{c}(\Phi_{c})$ *and* $\sum_{r}(\Phi_{r})$ refer to first order autoregressive correlation matrices with auto-correlation parameters Φ_{c} and Φ_{r} and order equal to the number of columns and rows, respectively. In this case, ξ is modelled as a separable first order auto-regressive process (AR1 x AR1) with covariance matrix $Var(\xi) = \sigma_{\xi}^{2}[\sum_{c}(\Phi_{c}) \otimes \sum_{r}(\Phi_{r})]$ (Gilmour et al.,1997). This model can preserve the design information and structure, which is a desired feature according to Qiao et al. (2000).

One first order correlation matrix AR1(ρ) is of the form (for 4 columns or rows):

$$\sum(\rho) = \begin{bmatrix} 1 & \rho^{|t_2-t_1|} & \rho^{|t_3-t_1|} & \rho^{|t_4-t_1|} \\ \rho^{|t_2-t_1|} & 1 & \rho^{|t_3-t_2|} & \rho^{|t_4-t_2|} \\ \rho^{|t_3-t_1|} & \rho^{|t_3-t_2|} & 1 & \rho^{|t_4-t_3|} \\ \rho^{|t_4-t_1|} & \rho^{|t_4-t_2|} & a^{|t_4-t_3|} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \rho^{1} & \rho^{2} & \rho^{3} \\ \rho^{1} & 1 & \rho^{1} & \rho^{2} \\ \rho^{2} & \rho^{1} & 1 & \rho^{1} \\ \rho^{3} & \rho^{2} & \rho^{1} & 1 \end{bmatrix}$$

Another formulation can be used such as R = S + R*, where $S = Var(\xi) = \sigma_{\xi}^{2}[\sum_{c}(\Phi_{c}) \otimes \sum_{r}(\Phi_{r})]$ and $R^{*} = I\sigma_{\eta}^{2}$. The matrix S can be included in G as well since ξ is another non-correlated (with other effects) random effect in the model (could be included in τ), besides the error.

Using trials established in complete block designs the following models were fitted.

Model 1: Complete block design, block as fixed effects.

Model 2: Complete block design, block as fixed effects + (AR1 x AR1) without inclusion of the independent term error.

Model 3: Complete block design, block as random effects.

Model 4: Complete block design, block as random effects + (AR1 x AR1) without inclusion of the independent term error.

Model 5: Complete block design, block as fixed effects + (AR2 x AR2) without inclusion of the independent term error.

Model 6: Complete block design, block as fixed effects + (AR2 x AR2) without inclusion of the independent term error + inclusion of rows and columns as random effects.

Model 7: Complete block design, block as fixed effects + (AR1 x AR1) with inclusion of the independent term error.

Model 8: Complete block design, block as fixed effects + (AR2 x AR2) with inclusion of the independent term error.

Model 9: Complete block design, block as fixed effects + row and column as random effects (treatment and block are not orthogonal to row and column).

Model 10: Complete block design, block as fixed effects + row and column as random effects + (AR1 x AR1) with inclusion of the independent term error (treatment and block are not orthogonal to row and column).

Model 11: Row and column as random effects, not considering block and spatial structure.

Model 12: Row and column as random effects (not considering block) + (AR1 x AR1) with inclusion of the independent term error.

Model 13: Row as fixed and column as random effects, not considering block and spatial structure.

Model 14: Row as fixed and column as random effects (not considering block) + (AR1 x AR1) with inclusion of the independent term error.

Model 15: Row and column as fixed effects, not considering block and spatial structure.

Model 16: Linear trend across rows and columns, not considering block and spatial structure.

Model 17: Linear trend across rows and columns, considering block but not the spatial structure.

Model 18: Spatial structure (AR1 x AR1) with inclusion of the independent term error (equivalent to model 7 but without adjusting the block effect).

Model 19: Row and column as fixed effects, not considering block but considering the spatial structure (AR1 x AR1).

Other models including splines were also evaluated.

1.2.3 Longitudinal Non-Spatial Models for Several Measures on each Trial

Repeated measures data analyses were approached by several models, including repeatability, multivariate, character process, ante-dependence, random regression and cubic spline models.

Character Process Models

Pletcher and Geyer (1999) suggested the use of character process models for the analysis of repeated measures. These models are based on the theory of stochastic process and were extended by Jaffrezic and Pletcher (2000) aiming at the relaxing its more restrictive assumption of stationarity of correlations. The simplest character process model uses the covariance function $C(t,s) = \sigma_t \sigma_s \rho^{(t-s)}$, where $C(t,s)$ is the covariance between repeated measures in times t and s, σ_t is the standard deviation of the trait in the time t and $\rho^{(t-s)}$ is the correlation between measures in times t and s. For data collected at regularly spaced times, this character process is equivalent to an autoregressive model with heterogeneous variance (ARH).

Ante-dependence Models

The basic idea of the ante-dependence models is that one observation in time t

can be explained by previous observations. Nunez-Anton and Zimmerman (2000) proposed the structured ante-dependence model in which the number of parameters is smaller than that in the traditional ante-dependence models. These models can deal with highly non-stationary correlation patterns and correspond, in their simple specifications, to a non-stationary generalisation of autoregressive models. They also consider the heterogeneity of variance between measures. The covariance matrix is of the form

$$\sum = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_1 & \sigma_1\sigma_3\rho_1\rho_2 & \sigma_1\sigma_4\rho_1\rho_2\rho_3 \\ & \sigma_2^2 & \sigma_2\sigma_3\rho_2 & \sigma_2\sigma_4\rho_2\rho_3 \\ Sim. & & \sigma_3^2 & \sigma_3\sigma_4\rho_3 \\ & & & \sigma_4^2 \end{bmatrix}.$$

Random regression models

By the random regression model (Meyer and Hill, 1997) the treatment effect is modelled by $\sum_r^{l-1} \beta_{ir}\Phi(a_{ik}^*)r$, where the term β_{ir} denotes the set of l random regressions coefficients for the ith treatment, $\Phi(a_{ik}^*)r$ is the rth polynomial on standardised age (a_{ik}^*) of measurement k. The estimated G matrix for treatment effects is given by G = ΦBΦ', where Φ is a matrix containing the random effects of the polynomials for the ages of measurements and B is the estimated variance-covariance matrix of the polynomial coefficients.

Cubic Spline Models

A cubic spline is a smooth curve over an interval formed by linked segments of cubic polynomials at certain knot points, such that the whole curve and its first and second differentials are continuous over the interval (Green and Silverman, 1994). Natural cubic splines can be incorporated into the standard mixed model framework (White, Thompson and Brotherstone, 1999; Verbyla et al., 1999). By the spline model the treatment effect is modelled by $b_{i0} + b_{i1}t_{ik} + \sum_{l=2}^{q-1} b_{il}z_l(t_{ik})$ where b_{i0} denotes the intercept for treatment i, b_{i1} denotes the slope for treatment i and b_{il} denotes the random regression coefficient for the ith treatment at knot l. The t_{ik} denotes the age of measurement

and $z_l(t_{ik})$ represents the spline coefficient for age t_{ik}. The estimated G matrix for treatment effects is given by G = $\Omega Z \Omega'$, where Ω is a matrix containing the random effects of the spline for the ages of measurements and Z is the estimated variance-covariance matrix of the spline coefficients.

1.2.4 Longitudinal Spatial Models for Repeated Measures on each Trial

In this case, the inverse of the correlated error variance matrix is given by

$$R^{-1} = R_o^{-1} \otimes H^{-1} = \begin{bmatrix} H\sigma^2_{\xi_1} & H\sigma_{\xi_{12}} \\ H\sigma_{\xi_{12}} & H\sigma^2_{\xi_2} \end{bmatrix}^{-1}, \quad \text{where:}$$

$$R_O = \begin{bmatrix} \sigma^2_{\xi_1} & \sigma_{\xi_{12}} \\ \sigma_{\xi_{12}} & \sigma^2_{\xi_2} \end{bmatrix} \qquad H = [\sum_c (\Phi_c) \otimes \sum_r (\Phi_r)]$$

1.2.5 Model Fitting Procedure and Model Comparisons

Likelihood Ratio Test (LRT)

Given two nested models U and V with maximum of the residual likelihood function L(U) and L(V) and correspondent number of parameters n_u and n_v, it can be showed that D = -2 log L(U) −2 log L(V) approaches a chi square distribution with n_v - n_u degrees of freedom (assuming U as nested within V). Testing the significance of D against the appropriate chi square distribution constitutes the LRT test. When V is the saturated model, D is called deviance. So, alternatively, the difference between the deviances of the two models can be used to do the LRT test.

The LRT test can be used to compare fitted models provided they have a nested structure and the same fixed effects. This permits comparison of models with different random factors for a constant structure of fixed effects. For comparing spatial models, the LRT statistic can be used to assess the order of the model to be fitted. Then, it is possible to test if an MA(2) model has a better fit than an MA(1), or whether an ARMA (1,1) is better than an AR(1). However, the use of

the LRT is limited to models fitted under the same regime of differencing. Testing models with different structures of fixed effects was considered by Welham and Thompson (1997).

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)

Other criterion for model selection is the Akaike Information Criterion, which penalise the likelihood by the number of independent parameters fitted. By this criterion, any extra parameter must increase the likelihood at least by one unit for entering in the model. The AIC is given by $AIC = -2 \log L + 2p$, where p is the number of parameters estimated. Smaller values of AIC reflect a better global fit (Akaike, 1974). Other approach is the Bayesian Information Criterion (BIC) of Schwarz (1978), which is given by $BIC = -2 \log L + p \log v$, where $v = N - r(x)$ is the number of residual degrees of freedom. BIC and AIC are calculated for each model and the model with the smallest value is chosen as the preferred model. AIC and BIC can be used for comparing non nested models, but the data should be the same which means the fixed effects should be the same. It can also be seen that both AIC and BIC depend on the basic quantity -2 log L.

Variograms

The variogram uses semivariances and is used in both methods of spatial analysis of field data: time series and geostatistics. The semivariance ranges from 0 (at lag 0) to a value equal to the variance of the observations (at a high lag). As the distance (called lags) between observations (plots or plants) increases, the variogram increases in value. The distance in which the variogram reaches a maximum or plateau (called sill c_1) equal to the variance of the data, is called range. The variograms display the spatial behaviour of the variable and inform about the pattern of variability in several directions. A variogram that reaches a sill or plateau is said to be stationary. The assumption of stationarity is made by the character process models such as the autoregressive. The variograms associated to various models were used as a mean of guiding the fitting procedures.

Software

All models were fitted using the software ASREML (Gilmour and Thompson, 1998; Gilmour, Cullis, Thompson and Welham, 2002) which uses the REML procedure through the algorithm average information and sparse matrix techniques (Gilmour, Thompson and Cullis, 1995; Johnson and Thompson, 1995; Thompson, Wray and Crump, 1994; Thompson et al., 2003). The software GENSTAT (Thompson and Welham, 2003; Welham, Thompson and Gilmour, 1998) was also used.

1.3 Application

The data set concerning to tea plant came from three trials established in complete block designs with six plants per plot and in a spacing of 3 x 2 meters. The trait leaf weight was evaluated at individual level in several consecutive years. Trial 1 was established with 141 treatments (open pollinated progenies) and 10 replications, summing 8,460 plants and 16,920 observations (two consecutive years). Trial 2 provided 5,400 observations (60 treatments x 5 replications x 6 plants per plot x 3 annual measures). From trial 3 4,050 observations were analysed (45 treatments, 5 replications, 6 plants per plot and 3 annual measures). The 45 treatments in trial 3 are also in trial 2. The basic model for all trials included block, treatments, plot and residual effects.

1.3.1 Univariate Spatial Models for Individual Annual Measures on each Trial

Results concerning to trial 1 are presented in Table 1.

Table 1. Summary of results concerning to models 1 to 19 for the trait leaf weight in the first and second years of harvest in trial 1. The estimates are: genetic variance among treatments (progenies) ($\hat{\sigma}_{\tau}^{2}$), non-correlated residual variance ($\hat{\sigma}_{\eta}^{2}$), narrow sense heritability ($\hat{h}^{2}$), adjusted narrow sense heritability ($\hat{h}^{2}_{adj} = (4\hat{\sigma}_{\tau}^{2})/(\hat{\sigma}_{\tau}^{2} + \hat{\sigma}_{\eta}^{2})$) proportional only to the unaccounted error (η), shrinkage factor ($\lambda_{1} = (\hat{\sigma}_{\eta}^{2} - 3\hat{\sigma}_{\tau}^{2})/(4\hat{\sigma}_{\tau}^{2})$) of the genetic effects in the mixed model equations and efficiency (Effic.) of models over the model1, in terms of $\hat{h}^{2}_{adj}$.

Leaf weight in the first measure

Model	Deviance	$\hat{h}^{2}$	$\hat{h}^{2}_{adj}$	λ_{1}	$\hat{\sigma}_{\tau}^{2}$	$\hat{\sigma}_{\eta}^{2}$	Effic.
Non-Spatial Models							
1	-3105.78	0.1413	0.1905	4.250	0.0110	0.2214	1.00
3	-3137.40	0.1378	0.1905	4.250	0.0110	0.2214	1.00
9	-3201.88	0.1416	0.1955	4.115	0.0111	0.2160	1.03
11	-3213.00	0.1391	0.1955	4.115	0.0111	0.2160	1.03
13	-3018.58	0.1439	0.1955	4.115	0.0111	0.2160	1.03
15	-2523.58	0.1600	0.2022	3.946	0.0115	0.2160	1.06
16	-3085.37	0.1371	0.1864	4.366	0.0108	0.2210	0.98
17	-3083.82	0.1422	0.1913	4.227	0.0111	0.2210	1.004
Spatial Models without η							
2	-3862.28	0.1652			0.0126		
4	-3894.34	0.1613			0.0126		
5	-4039.31	0.1667			0.0128		
6	-4045.31	0.1690			0.0128		
Spatial Models with η							
7	-4254.48	0.1737	0.3296	2.034	0.0134	0.1492	1.730
8	-4189.15	0.1788	0.3145	2.180	0.0132	0.1547	1.651*
10	-4257.66	0.1728	0.3278	2.051	0.0133	0.1490	1.721
12	-4283.01	0.1725	0.3276	2.052	0.0134	0.1502	1.720
14	-4069.91	0.1737	0.3278	2.051	0.0133	0.1490	1.721
18	-4278.02	0.1717	0.3264	2.063	0.0134	0.1508	1.713
19	-3498.55	0.1792	0.3352	1.983	0.0136	0.1487	1.760

Leaf weight in the second measure

Model	Deviance	$\hat{h}^2$	$\hat{h}^2_{adj}$	λ_1	$\hat{\sigma}^2_\tau$	$\hat{\sigma}^2_\eta$	Effic.
Non-Spatial Models							
1	11853.72	0.1688	0.2346	3.262	0.083	1.339	1.00
3	11836.52	0.1654	0.2346	3.262	0.083	1.339	1.00
9	1168990	0.1691	0.2438	3.101	0.084	1.294	1.04
11	11683.96	0.1668	0.2438	3.101	0.084	1.294	1.04
13	11767.86	0.1730	0.2438	3.101	0.084	1.294	1.04
15	12027.38	0.1781	0.2493	3.012	0.086	1.294	1.06
16	11926.37	0.1620	0.2282	3.383	0.081	1.339	0.97
17	11872.48	0.1702	0.2361	3.235	0.084	1.339	1.006
Spatial Models without η							
2	11424.52	0.1900	-	-	0.094	-	-
4	11407.16	0.1862	-	-	0.094	-	-
5	11247.98	0.1887	-	-	0.093	-	-
6	11208.30	0.1905	-	-	0.092	-	-
Spatial Models with η							
7	11089.97	0.1927	0.3465	1.886	0.0965	1.0175	1.477
8	11084.85	0.1914	0.3630	1.755	0.0956	0.9669	1.547
10	11070.69	0.1933	0.3489	1.866	0.0965	1.0099	1.487
12	11056.93	0.1920	0.3477	1.876	0.0966	1.0148	1.482
14	11150.56	0.1961	0.3491	1.865	0.0965	1.0093	1.488
18	11080.62	0.1907	0.3438	1.909	0.0965	1.0263	1.465
19	11456.87	0.1974	0.3485	1.870	0.0964	1.0101	1.486

Models with the same structure in terms of the fixed effects:
Block as fixed effects: models 1, 2, 5, 6, 7, 8, 9, 10.
Constant as fixed effect: models 3, 4, 11, 12, 18.
Row as fixed effect: models 13 and 14.
Rows and columns as fixed effects: models 15 and 19.

The deviance criterion is not adequate for comparing models with different fixed effects. The AIC criterion can be used but might not reflect superiority for genetic selection. So, the efficiency in terms of the adjusted heritability (proportional only to the unaccounted error) can be used for inference about the best models. The adjusted narrow sense heritability estimates presented in the previous table are referring to individual models rather then parent models.

The two traits (sequence measurements in consecutive years) presented approximately the same behaviour in terms of results across models. Among the non-spatial models, the row-column analysis (models 11, 13 and 15) performed better than the randomised block analysis (models 1 and 3). This can be explained by the local control in two directions provided by the row-column analysis and by the small block provided by rows since each original block was composed by six rows. Due to this last reason there was no need to fit block additionally to the row and column (models 9, 10 and 17). Among the spatial models, the block effect was insignificant in 10, which was then equivalent to 12.

The spatial models (2, 4, 5, 6, 7, 8, 10, 12, 14, 18 and 19) were always much better than the non-spatial ones (1, 3, 9, 11, 13, 15, 16 and 17) as judged by deviances of the models as well as by selection efficiencies in terms of the adjusted heritabilities or shrinkage factors for treatment effects in the mixed model equations (Table 1). The spatial models with inclusion of η (models 7, 8, 10, 12, 14, 18 and 19) were always better than that without η (models 2, 4, 5 and 6) as judged by deviances of the models as well as selection efficiencies in terms of the adjusted heritabilities or shrinkage factors for treatment effects in the mixed model equations (Table 1).

The need for keeping the design features in the analysis can be seen by comparing models 7, 12 and 18, that led to almost the same efficiency. The rate of recovering of design features by spatial analysis is enhanced when the independent error is fitted. A model without plot and design features was fitted for the two traits and provided almost the same efficiency as model 12, showing that sometimes simple spatial models can be used.

For two dimension spatial models without η (models 2 and 5), the model AR2 was better than the AR1 (change in deviance of 176.54). However, that superiority was not kept (change in deviance of 5.12) when models (7 and 8) with

η were fitted. In this case, the log L failed to converge for one of the traits. So there is no need for AR2 in models with inclusion of η. Besides, the two dimensional AR2 models fail to converge in a number of situations, revealing to be an over-parameterised model.

Little difference (in terms of the adjusted heritability), if any, was noted in fitting local control as fixed or random effects in the non-spatial models (models 1 against 3; 11 against 13 or 15) and spatial models (12 against 14 and 19), with a slight superiority for fitting row and column as fixed effects. However, this superiority probably is not real as the columns are incomplete and do not contributes for the recovering of genetic information when the column effect is fitted as fixed. Besides, the column effect was not significant in models 11 and 12 and the column variance reached zero. This prevents its fitting as fixed effects in models 15 and 19, which will lead to over-fitting. When an effect is treated as fixed, it is considered that its determination coefficient is 1. For this to be true, the effect variance should be, at least, greater than zero. Effects with variance tending to zero should not be fitted as fixed. In spatial analysis, the local control effects tend to be forced to zero and so, probably, such effects should be fitted as random. In a non-spatial context, very often is recommended to treat complete local controls as fixed effects for the sake of unbiased prediction/estimation.

The overall best methods for the two traits were 12 and 14, both corresponding to a row-column analysis + a spatial (AR1 x AR1) + independent term error. For these best models, the efficiency over the traditional randomised complete block analysis ranged from 1.48 to 1.76, i.e., 48% to 76% of superiority. Improved designs can be used to have high efficiency when assuming a spatial model such as model 12 (establishing the experiment according to model 12). In other words, appropriate systematic designs are needed when spatial patterns are present in the field. Spatial analysis has been shown to improve the precision and accuracy of treatment estimates, even with designs not optimised spatially. It is expected that designs with good general spatial properties will further increase the efficiency of treatments estimates. This would permit the fitting of only one spatial model to all trials as advocated by Kempton et al. (1994).

Models ARMA and MA were also tried. The models with error structure (ARMA1 x ARMA1) and (AR2 x AR2) are over-parameterised and, as the (MA1 x MA1), failed to converge. Gleeson and Cullis (1987) found that differencing in

one direction and then fitting a moving average (MA) correlation structure for the residuals in the same direction resulted in great gains in efficiency. However, differencing in two dimensional analysis can be prone to discard treatment information according to Kempton et al. (1994), who found that (MA1 x MA1) after first-differencing was inefficient for many trials. Several authors have questioned the need for differencing (Martin, 1990; Zimmerman and Harville, 1991). Others acknowledged that differencing is unnecessary for many trials (Cullis et al. 1998). Furthermore, differencing can often lead to the need for more complex modelling of the variance structure for the plot errors. In geostatistics, trend is modelled as a mixture of spatial covariances and/or deterministic functions of spatial coordinates. Other alternatives to differencing are the inclusion of polynomial functions of the spatial coordinates or the use of smoothing splines to model global trend. Differencing is often wasteful of degrees of freedom and information on treatments or genetic effects.

The variograms for the best models were stationary and exhibited approximately the same pattern. The autocorrelation coefficients for models without independent errors were approximately 0.21 and 0.29 for AR Column and 0.13 and 0.14 for AR Row, for the two traits, respectively. For models with independent errors, the autocorrelation coefficients were approximately 0.79 and 0.75 for AR Column and 0.50 and 0.52 for AR Row, for the two traits, respectively. These high autocorrelation coefficients obtained show that the AR process is modelling fertility gradient rather than competition. This is coherent with the spacing used (3 by 2 meters) and with crop management in which each year all the leaves are harvested. These features tend to avoid competition between plants.

Although the variograms have shown a reasonable behaviour, models with splines were also tried, some extending the previous model 12 and others using only splines to account the spatial variation. Results are presented in Table 2.

Table 2. Results concerning to some models for the trait leaf weight in the first two years of harvest in trial 1. The estimates are: genetic variance among treatments (progenies) ($\hat{\sigma}^2_\tau$), non-correlated residual variance ($\hat{\sigma}^2_\eta$), adjusted narrow sense heritability ($\hat{h}^2_{adj} = (4\hat{\sigma}^2_g)/(\hat{\sigma}^2_g + \hat{\sigma}^2_\eta)$) proportional only to the unaccounted error (η) and efficiency (Effic.) of models over the model 1, in terms of $\hat{h}^2_{adj}$. Spl(rc) means cubic splines applied on row and columns.

Model	Deviance	$\hat{h}^2_{adj}$	$\hat{\sigma}^2_\tau$	$\hat{\sigma}^2_\eta$	Eff.
			Leaf weight 1 – Trial 1		
1	-3105.78	0.1905	0.0110	0.2214	1.00
11	-2523.58	0.1955	0.0111	0.2160	1.03
Spl(rc)	-3202.72	0.1961	0.0112	0.2173	1.03
12	-4283.01	0.3276	0.0134	0.1502	1.72
12 + Spl(rc)	-4270.12	0.3302	0.0134	0.1489	1.73
			Leaf weight 2 – Trial 1		
1	11853.72	0.2346	0.0830	1.3390	1.00
11	11683.96	0.2438	0.0840	1.2940	1.04
Spl(rc)	11731.32	0.2464	0.0852	1.2979	1.05
12	11056.93	0.3477	0.0966	1.0148	1.48
12 + Spl(rc)	11070.08	0.3492	0.0965	1.0088	1.49

It can be seen that the extended model 12 through the inclusion of splines did not improve the fit. The deviances of the extended models were higher as the spline variance component is constrained to be positive, but the efficiencies in terms of the adjusted heritability were practically the same (Table 2).

The approach of using splines in place of AR(1) x AR(1) process for modelling spatial variation was suggested by Kempton (1999) and used by Durban, Currie and Kempton (2001). In our data set, such approach showed to be very inefficient being comparable only with the random row and column analysis (model 11).

Results concerning to individual analysis of trial 2 of tea plant are presented in Table 3.

Table 3. Results concerning to some models for the trait leaf weight in the first three years of harvest in trial 2. The estimates are: genetic variance among treatments (progenies) ($\hat{\sigma}^2_\tau$), non-correlated residual variance ($\hat{\sigma}^2_\eta$) and adjusted narrow sense heritability ($\hat{h}^2_{adj} = (4\hat{\sigma}^2_g)/(\hat{\sigma}^2_g + \hat{\sigma}^2_\eta)$) proportional only to the unaccounted error (η) and efficiency (Effic.) of models over the model 1, in terms of $\hat{h}^2_{adj}$.

Model	Deviance	$\hat{h}^2_{adj}$	$\hat{\sigma}^2_\tau$	$\hat{\sigma}^2_\eta$	Eff	Local Control (signif.)
Leaf weight 1						
1	-1964.99	0.4778	0.0137 ± 0.004	0.1010 ± 0.004	1.00	Not sig.
7	-2038.17	0.5248	0.0140 ± 0.004	0.0927 ± 0.004	1.10	
8	-2026.02nc	-	-	-	-	
11	-1995.02	0.4633	0.0131 ± 0.004	0.1000 ± 0.004	0.97	
12	-2059.70	0.5198	0.0138 ± 0.004	0.0924 ± 0.004	1.09	Not sig.
13	-1855.95	0.4602	0.0130 ± 0.004	0.1000 ± 0.004	0.96	Row *
15	-1661.66	0.3899	0.0108 ± 0.004	0.1000 ± 0.004	0.82	C**;r*
Leaf weight 2						
1	830.15	0.7076	0.1038 ± 0.03	0.483 ± 0.02	1.00	* 6%
7	703.19	0.7931	0.1061 ± 0.02	0.429 ± 0.02	1.12	
8	747.34 nc	-	-	-	-	-
11	795.78	0.7128	0.1017 ± 0.03	0.469 ± 0.02	1.01	
12	686.88	0.7934	0.1059 ± 0.02	0.428 ± 0.02	1.12	Not sig.
13	871.88	0.7134	0.1018 ± 0.03	0.469 ± 0.02	1.01	R**
15	984.87	0.6416	0.0896 ± 0.03	0.469 ± 0.02	0.91	C**;r*
Leaf weight 3						
1	3415.14	0.5887	0.345 ± 0.10	1.999 ± 0.07	1.00	**
7	3215.12	0.6852	0.351 ± 0.08	1.698 ± 0.07	1.16	
8	3358.81n	-	-	-	-	-
11	3379.20	0.6041	0.335 ± 0.10	1.883 ± 0.07	1.03	
12	3212.90	0.6849	0.350 ± 0.08	1.694 ± 0.07	1.16	Not sig.
13	3378.11	0.6069	0.337 ± 0.10	1.884 ± 0.07	1.03	R**
15	3401.23	0.5176	0.280 ± 0.10	1.884 ± 0.07	0.88	R**;c ns
18	3213.71	0.6858	0.352 ± 0.08	1.701 ± 0.07	1.16	

It can be seen that the best models for all three traits were 7 (complete block design + (AR1 x AR1) + η) and 12 (row-column design + (AR1 x AR1) + η) in terms of efficiency over the base model 1 (block analysis) and deviance values. The efficiencies (between 1.09 and 1.16) were in general much lower than the previous (of the order of 1.48 to 1.76) reported for trial 1. This is because there is much less environmental variability in this trial as revealed by the low significance of block effects for two of the three traits. Due to the same reason the efficiencies of row-column over block designs were small or did not exist in this case. For the trait leaf weight 1, block and row effects should not be fitted as fixed because they were non-significant. So, the results concerning to models 1, 11, 13 and 15 are not comparable for the trait 1.

For this trial, column effects should not be fitted as fixed (model 15) as it is so small (size 30) and genetic information would be lost. With spatial analysis and inclusion of the independent error in the model there was no need to include the design features in the model, even when the block effects were significant (trait 3). It can be seen from the deviance values that the model 7 and 18 were equivalent (Table 3). The model 8 with (AR2 x AR2) structure did not converge for all traits. The auto-correlation coefficients were of the order of 0.80 and 0.90 between rows and columns, respectively, for the three traits (0.79 and 0.87; 0.79 and 0.87; 0.81 and 0.90, for traits 1, 2 and 3, respectively, according to the model 12).

Results concerning to individual analysis of trial 3 of tea plant are presented in Table 4.

Table 4. Results concerning to some models for the trait leaf weight in the first three years of harvest in trial 3. The estimates are: genetic variance among treatments (progenies) ($\hat{\sigma}^2_\tau$), non-correlated residual variance ($\hat{\sigma}^2_\eta$) and adjusted narrow sense heritability ($\hat{h}^2_{adj} = (4\hat{\sigma}^2_g)/(\hat{\sigma}^2_g + \hat{\sigma}^2_\eta)$) proportional only to the unaccounted error (η) and efficiency (Effic.) of models over the model1, in terms of $\hat{h}^2_{adj}$.

Model	Deviance	$\hat{h}^2_{adj}$	$\hat{\sigma}^2_\tau$	$\hat{\sigma}^2_\eta$	Eff.	Local Control (signif.)
Leaf weight 1						
1	-1815.12	0.8736	0.0223 ± 0.005	0.0798 ± 0.003	1.00	Block ns
11	-1880.34	0.8735	0.0221 ± 0.005	0.0791 ± 0.005	1.00	C*,r ns
12	-1910.51	0.9360	0.0212 ± 0.005	0.06940 ± 0.004	1.07	C*,r ns
12AR2	-1909.93ns	-	-	-	-	-
Leaf weight 2						
1	365.48	0.8527	0.0997 ± 0.03	0.368 ± 0.02	1.00	
11	138.86	0.9159	0.1078 ± 0.03	0.3630 ± 0.02	1.07	C*,r ns
12	156.77	0.8830	0.1000 ± 0.02	0.3530 ± 0.05	1.04	C ns;r ns
12AR2	216.08 n.c	-	-	-	-	-
Leaf weight 3						
1	3437.61	1.05	1.268 ± 0.40	3.584 ± 0.15	1.00	
11	3323.76	1.06	1.252 ± 0.31	3.460 ± 0.15	1.00	c*; r*
13	3319.46	1.06	1.256 ± 0.31	3.468 ± 0.15	1.00	r*
12	3191.46	1.09	1.211 ± 0.29	3.240 ± 0.14	1.03	c ns;r ns
12AR	3284.14 nc	-	-	-	-	-

The auto-correlation coefficients between rows and columns were 0.45 and 0.83; 0.95 and 0.89; 0.91 and 0.96, for traits 1, 2 and 3, respectively, according to the model 12. This was the best model in terms of deviance. However, it does not provide significantly better efficiencies than the non-spatial models, except for trait 1. This is because there is small environmental variability in this trial as revealed by the non-significance of row effects for two of the three traits and by the high values of the adjusted heritability. The heritabilities greater than 1 can be due to an unrealistic assumption of half sib parentage between individuals in a family. The models with (AR2 x AR2) error structure were non-significant over the (AR1 x AR1) or did not converge.

1.3.2 Longitudinal Non-Spatial Models for Several Measures on each Trial

Results concerning to repeatability and multivariate models for the repeated measures in trial 1 are presented in Tables 5 and 6. Block, measure and block x measure interaction effects were fitted as fixed.

Table 5. Estimates of the variance parameters: genetic among treatments (progenies) ($\hat{\sigma}^2_\tau$), among plots ($\hat{\sigma}^2_\kappa$), permanent ($\hat{\sigma}^2_\rho$) and residual ($\hat{\sigma}^2_\eta$). Repeatability and multivariate models with original data in trial 1.

	Repeatability Model	Multivariate Model*	
Parameters estimates	Both weight	Leaf weight 1	Leaf weight 2
$\hat{\sigma}^2_\tau$	0.1913	0.0462	0.4397
$\hat{\sigma}^2_\kappa$	0.0739	0.0365	0.13854
$\hat{\sigma}^2_\rho$	0.3038	-	-
$\hat{\sigma}^2_\eta$	0.5413	0.2214	1.3390
Deviance	14173.70	3357.02	

* The genetic and residual covariances involving the pair of ages 1-2 were 0.1393 and 0.3687, respectively.

The associated repeatability coefficient was 0.51, which can be classified as intermediate. The genetic correlation coefficient between the two measures in the multivariate analysis was 0.98. These results show that probably the trait is not changing so much genetically from one another measure or age. However, it can be seen that there is heterogeneity of variance between the measures. The deviance values show that the multivariate model is much better than the repeatability. This justifies the preference by the multivariate model. Results with standardised (divided by the phenotypic standard deviation) data are presented in Table 6.

Table 6. Estimates of the variance parameters: genetic among treatments (progenies) ($\hat{\sigma}^2_\tau$), among plots ($\hat{\sigma}^2_\kappa$), permanent ($\hat{\sigma}^2_\rho$) and residual ($\hat{\sigma}^2_\eta$). Repeatability and multivariate models for standardised data in trial 1.

	Repeatability Model	Multivariate Model*	
Parameters estimates	Both weight	Leaf weight 1	Leaf weight 2
$\hat{\sigma}^2_\tau$	0.1799	0.1488	0.2184
$\hat{\sigma}^2_\kappa$	0.0848	0.1176	0.0764
$\hat{\sigma}^2_\rho$	0.4541	-	-
$\hat{\sigma}^2_\eta$	0.2465	0.7124±	0.6649
Deviance	7637.22	7316.52	

* The genetic and residual covariances involving the pair of ages 1-2 were 0.1763 and 0.4661, respectively.

With standardised data, the associated repeatability coefficient was 0.75, which is higher than the previous one. The standardisation led to an increased permanent variance estimate, while the others (except by the independent error) variance components were kept approximately constant (in comparison to the data in original scale) by the repeatability model. The genetic correlation coefficient between the two measures in the multivariate analysis was 0.98, which is the same as in the previous analysis. However, it can be seen that the heterogeneity of variance was reduced after standardisation. The deviance values show that the repeatability and multivariate models became closer after standardisation.

Nevertheless, the AIC values were 7334.52 and 7645.22 for the multivariate and repeatability models. This shows that multivariate model, although less parsimonious, is still better than the repeatability model. So, in practice, the multivariate model should be used for selection. In case of choice in favour of the repeatability model, the data should at least be standardised. The use of multivariate model for selection implies giving weight to genetic values predicted for the two measures. These weights should be 0.5 if the two ages have equal importance. If the last measure provides a better representation of a mature trait, higher weight should be given for this measure. Nonetheless, the high genetic correlation may suggest that the weights should be 0.5 for each measure.

Estimates for the multivariate model with original data in trial 2 are presented in Table 7. Block, measure and block x measure interaction effects were fitted as fixed.

Table 7. Estimates of the variance and covariance parameters for the multivariate model with original data in trial 2, concerning to three repeated measures.

Treatment (genetic)			Plot			Residual		
Covar.\Variance\Correl.			Covar.\Variance\Correl.			Covar.\Variance\Correl.		
0.0134	0.9239	0.9984	0.0248	0.8638	0.7095	0.1011	0.6766	0.6128
0.0342	0.1020	0.9211	0.0422	0.0964	0.9123	0.1495	0.4827	0.7686
0.0673	0.1711	0.3380	0.0817	0.2072	0.5352	0.2755	0.7551	1.9990
Deviance = -787.172								

The deviance value (Table 7) reveals that the multivariate model is far more suitable for the original data than the repeatability model (deviance 5070.64). Such model gave high values for the genetic correlations between pairs of measures. The correlations were all within the parameter space but the model had to be constrained to achieve this. Without constraining the G matrix to be positive definite, correlations higher than 1 and negative variance components were obtained. In the constrained model, the G matrix is bent and this process involves shrinking the variances towards their mean. The unconstrained analysis is less biased because bias is introduced when constraining the solution to the parameter space. Convergence turned difficult as the number of measure increased. So, more suitable models needed to be searched.

Results concerning to the character process model called first order autoregressive with heterogeneous variance (ARH) for the treatments effects are presented in Table 8.

Table 8. Estimates of the variance and covariance parameters for the character process model called first order autoregressive with heterogeneous variance (ARH) applied to original data in trial 2, concerning to three repeated measures.

Treatment (genetic)			Plot			Residual		
Covar.\Variance\Correl.			Covar.\Variance\Correl.			Covar.\Variance\Correl.		
0.0129	0.9761	0.9528	0.0254	0.8667	0.7532	0.1011	0.6766	0.6128
0.0357	0.1033	0.9761	0.0430	0.0968	0.9109	0.1495	0.4827	0.7686
0.0619	0.1792	0.3261	0.0891	0.2104	0.5510	0.2755	0.7551	1.9990
Deviance = -782.94								

The ARH and multivariate models presented almost the same deviance and the AIC values were -750.94 and -751.17, respectively, which are basically the same -751. So, the two models are equivalents by the parsimony criterion. However, the ARH presented easy convergence without constraining the G matrix to be positive definite, fitted a small (two less than the multivariate model) number parameters and gave correlations within the parameter space. Besides, it gave a more realistic correlation between the most distant measures 1 and 3. The ARH model is then much preferred. Such model assumes stationarity and same correlation in all intervals of same lag.

Other model evaluated was the structured ante-dependence model (SAD) which has also parsimony and does not assume stationarity. Results are presented in Table 9.

Table 9. Estimates of the variance and covariance parameters for the structured ante-dependence model (SAD) with original data in trial 2, concerning to three repeated measures.

Treatment (genetic)			Plot			Residual		
Covar.\Variance\Correl.			Covar.\Variance\Correl.			Covar.\Variance\Correl.		
0.0128	0.9840	0.9580	0.0254	0.8667	0.7532	0.1011	0.6766	0.6128
0.0358	0.1032	0.9730	0.0430	0.0968	0.9109	0.1495	0.4827	0.7686
0.0618	0.1784	0.3250	0.0891	0.2104	0.5510	0.2755	0.7551	1.9990
Deviance = -783.04								

The SAD and ARH models presented basically the same deviance (-783) and then are equivalent by this criterion. Nonetheless, the SAD model fitted one more parameter than the ARH model and is not preferred in terms of parsimony by the AIC rule. The results for plot and residual effects were exactly the same by the two models. The genetic components were slightly different but are both coherent in terms of the magnitude of the correlation coefficients, i.e., smaller for the lag 1-3. This was not achieved by the multivariate model. Both models could be used efficiently in practice. The SAD model allows for different correlation for lags of same size.

These two classes of models were also used for modelling the other random terms of the model. Results concerning to correlations for treatment and plot terms modelled by ARH and SAD are presented in Table 10.

Table 10. Estimates of the correlation parameters for the structured ante-dependence model (SAD) and character process (ARH) for modelling both the treatment and the plot effects. Original data set in trial 2 (three repeated measures) was used.

Treatment (genetic)			Plot			Residual		
ARH\SAD			ARH\SAD			ARH\SAD		
-	0.990	0.968	-	0.851	0.769	-	0.6766	0.6128
0.982	-	0.977	0.882	-	0.903	0.6766	-	0.7686
0.964	0.982	-	0.778	0.882	-	0.6128	0.7686	-
Deviance ARH\SAD = -780.76\-782.20								

The results show that the plot effect can be perfectly modelled by the ARH or SAD process. The deviance values were close to the previous where the plot effect was modelled in a multivariate fashion. The AIC values here were -752.76 and -750.20 for ARH and SAD, respectively, which are close to the values -751 and -749 for ARH and SAD, respectively, obtained with the two models but with multivariate plot effect. Comparing these four AIC values, the choice is for the ARH model for both treatment and plot effects (AIC -752.76).

The modelling of the residual term by the ARH was also evaluated. The resultant deviance for modelling the three effects simultaneously as an ARH process gave a deviance of only -677.46. Also the residual correlations obtained were very different than the previous ones. Then the residual should be modelled in a multivariate way.

Other approaches were also evaluated. The banded correlation or Toeplitz model converged with a deviance of -794.28. Nevertheless, gave a genetic correlation higher than one, just the correlation supposed to be the small one. When constraining the model to give a positive definite G matrix, the convergence failed. This model assumes equal correlation for lags of same size as does the ARH model, but the elements of the several diagonals are different and not a function of the correlation for lag 1.

Random regression models were also tried and results for the full constrained model are presented in Table 11.

Table 11. Estimates of the variance and covariance parameters for the full (quadratic fit) random regression model with original data in trial 2, concerning to three repeated measures.

Treatment (genetic)			Plot			Residual		
Covar.\Variance\Correl.			Covar.\Variance\Correl.			Covar.\Variance\Correl.		
0.0134	0.9239	0.9984	0.0248	0.8638	0.7095	0.1011	0.6766	0.6128
0.0342	0.1020	0.9211	0.0422	0.0964	0.9123	0.1495	0.4827	0.7686
0.0673	0.1711	0.3380	0.0817	0.2072	0.5352	0.2755	0.7551	1.9990
Deviance = -787.172								

Results were identical to those (which were not suitable) from the multivariate analysis as expected for the full fitting of the random regression model, i.e., for fitting a quadratic polynomial. In a search for parsimony a reduced fit was tried and the results are presented in Table 12.

Table 12. Estimates of the variance and covariance parameters for the reduced (linear fit) random regression model with original data in trial 2, concerning to three repeated measures.

Treatment (genetic)			Plot			Residual		
Covar.\Variance\Correl.			Covar.\Variance\Correl.			Covar.\Variance\Correl.		
0.0098	1.0552	1.0765	0.0248	0.8638	0.7095	0.1011	0.6766	0.6128
0.0331	0.1004	1.0040	0.0422	0.0964	0.9123	0.1495	0.4827	0.7686
0.0563	0.1677	0.2791	0.0817	0.2072	0.5352	0.2755	0.7551	1.9990
Deviance = -777.08								

The deviance (-777) of the model is higher than that (-783) obtained from the ARH and SAD models for treatment effects (Tables 8 and 9). The AIC value is −747 which is higher than that obtained for ARH (-751) and SAD (-749) models. So, the reduced random regression model is not a choice. Also, this model showed a poor reconstruction of the G matrix for treatment effects leading all correlations to be higher than 1 (Table 12). These results are in accordance with Apiolaza, Gilmour and Garrick (2000) who found that random regression models were often inappropriate.

The fit of smoothing cubic splines was also tried. The deviance obtained was only −748.33, which was the worst between the parsimonious models tried. This result was expected as function of the small number of ages available for fitting.

In conclusion, the best approaches for trial 2 were the ARH and SAD models for treatment and plot effects. These models should be extended and used in conjunction with the spatial models for the residuals.

Results concerning to multivariate models for the repeated measures (original data) in trial 3 are presented in the Table 13. Block, measure and block x measure interaction effects were fitted as fixed.

Table 13. Estimates of the variance and covariance parameters for multivariate model with original data in trial 3, concerning to three repeated measures.

Treatment (genetic)			Plot			Residual		
Covar.\Variance\Correl.			Covar.\Variance\Correl.			Covar.\Variance\Correl.		
0.0273	0.9864	0.9246	0.0117	0.8757	0.7782	0.0798	0.5714	0.5383
0.0468	0.0825	0.9569	0.0400	0.1780	0.9556	0.0979	0.3680	0.7611
0.1714	0.3086	1.2600	0.1107	0.5305	1.7320	0.2879	0.8742	3.5840
Deviance = -60.3412								

The genetic correlation coefficient between the two first measures in the multivariate analysis was 0.986, which is close to the values obtained for ages 1 and 2 in trials 1 and 2. The genetic correlation between ages 1 and 3 was 0.92 and between 2 and 3 was 0.97. These results show that probably the trait is approximately the same in all three ages. These results are coherent showing that the small correlation occurred between ages 1 and 3.

The deviance values reveal that the multivariate model is far more suitable for the original data than the repeatability model (deviance 6754.42). Such model gave high values for the genetic correlations between pairs of measures. The correlations were all within the parameter space and the model had not to be constrained to achieve this. Even so, more parsimonious suitable models were searched.

The ARH and SAD models were used for modelling both the treatment and the plot effects. Results concerning to correlations and deviances are presented in Table 14.

Table 14. Estimates of the correlation parameters for the structured ante-dependence model (SAD) and character process (ARH) for modelling both the treatment and the plot effects in trial 3 (three repeated measures).

Treatment (genetic)			Plot			Residual		
ARH\SAD			ARH\SAD			ARH\SAD		
1.0000	0.9472	0.9214	1.0000	0.8575	0.8118	1.0000	0.5742	0.5345
0.9640	1.0000	0.9728	0.9417	1.0000	0.9467	0.5705	1.0000	0.7618
0.9290	0.9640	1.0000	0.8869	0.9417	1.0000	0.5317	0.7629	1.0000
Deviance ARH\SAD = -60.50\-64.16								

The structured ante-dependence model (SAD) and character process (ARH) for modelling both the treatment and the plot effects gave smaller deviance than the multivariate model. So they are better. Besides, the ARH fitted four less parameters and the SAD fitted two less parameters than the multivariate model. The AIC values for AHR and SAD were -32.59 and -32.16, respectively, that are basically the same. So, either of these two models could be used.

The modelling of the residual term by the ARH was also evaluated. The resultant deviance for modelling the three effects simultaneously as an ARH process gave a deviance of only 73.26. Also the residual correlations obtained were very different than the previous ones. Then the residual should be modelled in a multivariate way.

Again, the best approaches were the ARH and SAD models for treatment and plot effects. These models should be extended and used in conjunction with the spatial models for the residuals.

1.3.3 Longitudinal Spatial Models for Repeated Measures on each Trial

Results concerning to multivariate spatial model for trial 1 are presented in Table 15.

Table 15. Estimates of the variance parameters: genetic among treatments (progenies) ($\hat{\sigma}^2_\tau$), among plots ($\hat{\sigma}^2_\kappa$), residual ($\hat{\sigma}^2_\eta$) and respective covariance and correlation by the multivariate spatial model for leaf weight in trial 1.

	Correlation	Covariance	Variance	
Parameters estimates	Both weight	Both weight	Leaf weight 1	Leaf weight 2
$\hat{\sigma}^2_\tau$ (Treatment)	0.9736	0.1390	0.0465	0.4386
$\hat{\sigma}^2_\kappa$ (Plot)	0.9771	0.0251	0.0068	0.0975
$\hat{\sigma}^2_\eta$ (Independent Error)	0.6697	0.3378	0.1952	1.3030
Deviance	2764.12			

It can be seen that the multivariate spatial model is the best option for the analysis and selection concerning this experiment. The deviance of this model (2764.12) is much lower than that of the multivariate non-spatial model (3357.02). The models fitted 11 and 9 random effects and the AIC values were 2786.12 and 3375.02 for the multivariate spatial and non-spatial models, respectively. So, the choice is for multivariate spatial model. The variogram showed the same behaviour as in the two univariate spatial analyses.

The genetic variance components stayed almost the same as in the multivariate non-spatial model. However, the plot variances were greatly reduced. The residual variances were also reduced by spatial analysis as expected. The genetic correlation showed about the same magnitude as in the non-spatial analysis. In conclusion, for trial 1, the selection should be practised according to the multivariate spatial analysis with weights to be given to genetic values in each measure.

For trial 2, the superior approaches for analysing the repeated measures were extended by incorporating spatially correlated residuals. Three models were tried: ARH for treatments, ARH for treatments and plots and SAD for treatments and plots. The deviance values obtained were -769.92, -767.70 and -769.82, respectively. These values are higher than that obtained with the best non-spatial models and the autocorrelation parameters were fixed at boundary of 1, revealing that there is no need for spatial analysis for this multivariate data. This was expected as the efficiency of spatial analysis for the univariate case in this experiment was low as a function of the low environmental variability in the trial. In the multivariate case for the repeated measures, the amount of information about one individual increase and the model is automatically improved becoming more difficult to add important information from the spatial analysis. Besides, the autocorrelation estimates approached 1, revealing that the estimated correlated error was of small magnitude.

For trial 3, the ARH model for both treatment and plot effects was extended by incorporating spatially correlated residuals. The autocorrelation parameters were fixed at boundary 1 and the analysis did not converge. As mentioned for the trial 2, this result was expected.

1.4 Conclusions

- For individual analysis the best model out of 19 was the row-column analysis + a spatial autoregressive (AR1 x AR1) correlated error + independent term error (efficiency between 1.09 and 1.76 over block analysis, i.e., between 9% and 76% of improvement).
- The traits (sequence measurements in consecutive years) gave approximately the same behaviour in terms of results and variograms across models.
- In general, the best approaches involved the modelling of treatment effects by ante-dependence or autoregressive models with heterogeneous variance and the modelling of error as a spatial autoregressive (AR1 x AR1) correlated error + independent term error.

2. Factor Analytic Multiplicative Mixed Models in the Analysis of Multiple Experiments

2.1 Introduction

Analysis of experiments repeated on several sites or environments are very common and important in agriculture. Such trials aim at providing inferences concerning to responses on both broad (in the average of all sites) and specific environments. To attain this, all the information should be analysed simultaneously. Traditional analysis of these multi-environment trials (MET) has been done through joint analysis of variance (ANOVA) and linear regression techniques. In general, stability and adaptability approaches (Finlay and Wilkinson, 1963; Eberhart and Russell, 1966) have been used to study treatment x environment interaction, mainly referred to as genotype x environment interaction or g x e. In spite of their generalised use, these regression based methods present limitations that have been reported in literature, such as inefficiency in the presence of non-linearity generating simplified response models (Crossa, 1990; Duarte and Vencovsky, 1999). Some proposed models (Cruz et al., 1989) correct this inefficiency but the g x e component has been estimated but not decomposed into the pattern (tendency) and noise components.

A first attempt to circumvent these limitations was the proposed technique called AMMI (Additive Main Effects and Multiplicative Interaction Analysis). This technique has been well described by Gauch (1988; 1992) and attributed to Fisher and Mackenzie (1923) and Gollob (1968). Another denomination of the method is PCA (Doubled Centred Principal Components Analysis). AMMI may be viewed as a procedure to separate pattern (the g x e interaction) from noise (mean error of treatment mean within trial). This is achieved by PCA, were the first axes (i.e. the axes with the largest eigenvalues), recover most of the pattern, whilst most of the noise ends up in later axes. The pattern can be viewed as the whole g x e effect weighted by an estimate of the pattern-to-noise ratio associated with the respective effect. This pattern-to-noise ratio is a variance component ratio analogue to a repeatability or heritability coefficient (Piepho, 1994). The multiplicative models AMMI have been popularised in a fixed model context and found a number of applications (Gauch, 1988; 1992; Crossa et al., 1990). AMMI analysis combines in a model, additive components for main effects (treatments and environments) and multiplicative components

for g x e effects. It combines a univariate technique (ANOVA) for the main effects and a multivariate technique (PCA-principal component analysis) for g x e effects. Crossa (1990) suggests that the use of multivariate techniques permits a better use of information than the traditional regression methods.

Although useful, the AMMI models present at least five great limitations: consider the genotype and g x e effects as fixed; is suitable only for balanced data sets; does not consider the spatial variation within trials; does not consider the heterogeneity of variance between trials; does not consider the different number of replications across sites. These features are not realistic in analysing field data, where the data are generally unbalanced and treatments (genotypes) are in a great number not supporting the assumption of fixed genotype effects (implicit heritability at mean level equal 1). The AMMI model estimates phenotypic and not genotypic values. If genotypes are considered as random, effects can be predicted by best linear unbiased prediction (BLUP). Hill and Rosenberger (1985) and Stroup and Mulitze (1991) showed that assuming random genotypes may be preferable in terms of predictive accuracy even when genotypes would be considered fixed by conventional standards. Assuming genotype as random effects it is possible to obtain shrinkage predictions of the random interaction g x e terms and so to separate pattern and noise as do AMMI models. In this sense, BLUP and AMMI may be seen as two approaches to achieve the same goal, namely to separate pattern from noise. The BLUP procedure estimates the GLS of interaction effects and then weights them by an estimate of the correspondent pattern-to-noise ratios. However, the BLUP procedure has a number of advantages that circumvent all the limitations of AMMI. It has also been shown that BLUP can be predictively more accurate than AMMI models (Piepho, 1994).

The full multivariate BLUP model is the best approach for analysing data on multiple experiments. This model provides response on each environment through the use of all information. However, with great number of experiments the mixed model analysis is unlikely to converge. The variance-covariance matrix in this case is completely unstructured, which means a great number of parameters to be estimated. So, the parsimonious model behind AMMI is an interesting feature. Van Eeuwijk et al. (1995) suggested to obtain a genotype by environment BLUP and then subject this table to AMMI analysis, using an single value decomposition procedure. A better approach was found by Piepho (1998). In a mixed model setting, he presented a multiplicative factor analytic model with

random genotype and g x e effects, which is conceptually and functionally better than AMMI. In the same context, Smith, Cullis and Thompson (2001) presented a general class of factor analytic multiplicative mixed models that encompass the approach of Piepho (1998) and include separate spatial errors for each environment. Such general class of models provides a full realistic approach for analysing MET data (Thompson et al., 2003).

The multivariate technique of factor analysis (Lawley and Maxwell, 1971; Mardia, Kent and Bibby, 1988; Comrey and Lee, 1992) provides simplification of correlated multivariate data as do other multivariate methods such as principal components analysis and canonical transformation. These techniques consider the correlation between variables and generate a new set of independent (non-correlated) variables. The technique of factor analysis can be considered as an extension of the principal component analysis. The factor analytic variance-covariance structure may be regarded as an approximation to the completely unstructured variance-covariance matrix and can provide parsimonious models.

Analysis of multi-environment trials (MET) has also been traditionally based on simple models assuming error variance homogeneity between trials, independent error within trial, genotype x environment (g x e) effects as a set of independent random effects. The combined analysis of MET data through realistic models is a complex statistical problem, which requires extensions to the standard linear mixed model. Such extensions have been done recently. Cullis, Gogel, Verbyla and Thompson (1998) presented a spatial mixed model analysis for MET data, which fits a separate error structure for each site, circumventing the assumptions of error variance homogeneity among trials and independent error within trial. The relaxation of the assumption concerning to independence of g x e effects can be achieved with the use of multiplicative models.

In a mixed model setting, multiplicative models for random g x e interaction terms induce correlations between the interactions. Mixed models with multiplicative terms are closely related to the so-called factor analytic variance-covariance structure advocated by Jennrich and Schluchter (1986). Piepho (1997) proposed multiplicative mixed models for multi-environment analysis but assumed random environment rather than random genotype effects. The same author proposed the use of factor analytic multiplicative mixed (FAMM) models with random genotype effects (Piepho, 1998). Smith, Cullis and Thompson (2001) presented a general class of FAMM models that encompass the approach

of Piepho (1998) and provides: accounting of heterogeneity of g x e variance; accounting of correlation among g x e interactions; appropriate spatial error variance structures for individual trials. This factor analytic multiplicative mixed spatial (FAMMS) model provides parsimonious models for large multivariate data sets and a better conceptual approach for interaction effects based on multiplicative model. The model can be regarded as a random effects analogue of AMMI. Smith, Cullis and Thompson (2001) reported that the advantages of FAMMS models are numerous and include: (i) within trial spatial variation can be accommodated; (ii) between trial error variance heterogeneity can be accommodated; (iii) unbalanced data are easily handled; (iv) genotype effects and g x e interactions can be regarded as random, leading to better predictions; (v) the goodness of fit of the model, i.e., number of multiplicative terms needed, can be formally tested through REMLLRT. Through a unified mixed model approach the stability parameters are integrated into broad (selection for an average environment), specific (selection for specific environments) and new-environment (selection for a non-tested environment) inferences. Also, traditional methods such as that of Wricke (1965) can be applied over the predicted genotypic values, eliminating the original disadvantage of the method, concerning the consideration of phenotypic rather than genotypic values of interaction effects.

The present paper deals with the application of FAMM and FAMMS models in two large unbalanced data sets aiming at the emphasising their advantages over AMMI models in terms of the assumptions of error variance homogeneity between trials and independent error within trials. Also, the ability of FAMM models in providing parsimonious models is also stressed.

2.2 Factor Analytic Models

A model concerning to evaluation of several treatments or genotypes in several environments is given by:

$Y_{ij} = \mu + g_i + e_j + ge_{ij} + \varepsilon_{ij}$, where:

μ, g, e, ge and ε are the fixed constant, genotype, environment, genotype x environment interaction and within environment error effects, respectively. The μ and e effects can be regarded as fixed and the others as random. A model referring to random genotype effects in each environment can be written as:

$Y_{ij} = \mu + g_{ij} + e_j + \varepsilon_{ij}$.

In the context of MET data, the factor analysis approach can be used to provide a class of structures for the variance-covariance matrix G_0. The model is postulated in terms of the unobservable genotype effects in different environments:

$g_{ij} = \sum_{r=1}^{k} \lambda_{jr} f_{ir} + \delta_{ij}$, where:

g_{ij} : efféct of the genotype i in environment j;

λ_{jr} : loading for factor r in environment j;

f_{ir} : score for genotype i in factor r;

δ_{ij} : error representing the lack of fit of the model.

The FAMM model is presented according to Smith, Cullis and Thompson (2001). Applied to the g genotype effects on s environments, the factor analytic model postulates dependence on a set of random hypothetical factors $f_r^{(g\,x\,1)}, (r = 1 .... k < s)$. In vector notation, the factor analytic model for these effects is

$g_s = (\lambda_1 \otimes I_g) f_1 + ... + (\lambda_k \otimes I_g) f_k + \delta$, where:

$\lambda_r^{(s\,x\,1)}$: loadings or weights of the factors in environments;

$\delta^{(gs\,x\,1)}$: vector of residuals or lack of fit for the model (also called vector of specific factors).

In a compact way, the model is:

$g_s = (\Lambda \otimes I_g) f + \delta$, where:

$\Lambda^{(s\,x\,k)} = [\lambda_1 ..... \lambda_k]$;

$f^{(gk\,x\,1)} = (f_1', f_2' .... f_k')'$.

The joint distribution of f *and* δ is given by

$$\begin{pmatrix} f \\ \delta \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_k \otimes I_g & 0 \\ 0 & \Psi \otimes I_g \end{pmatrix}\right]$$, where:

$\Psi = diag\ (\psi_1 .....\psi_p)$;

ψ_i : specific variance for the *i*th trial.

The variance matrix for the genotype effects on environments is given by

$$\begin{aligned} \text{var}(g_s) &= (\Lambda \otimes I_g)\,\text{var}(f)(\Lambda' \otimes I_g) + \text{var}(\delta) \\ &= (\Lambda\Lambda' + \Psi) \otimes I_g . \end{aligned}$$

The model for genotype effects in each environment leads to a model for G in which:

$\sigma_{g_{jj'}} = \sum_{r=1}^{k} \lambda_{jr}^2 + \psi_j$: genotype variance in environment j;

$\sigma_{g_{jj'}} = \sum_{r=1}^{k} \lambda_{jr}\lambda_{j'r}$: genotype covariance between environments j and j';

$\rho_{g_{jj'}} = \sum_{r=1}^{k} \lambda_{jr}\lambda_{j'r} / [(\sum_{r=1}^{k} \lambda_{jr}^2 + \psi_j)(\sum_{r=1}^{k} \lambda_{j'r}^2 + \psi_{j'})]^{1/2}$: genotype correlation between environments j and j'

The equation for g_s has the form of a (random) regression on k environmental covariates $\lambda_1 ....\lambda_k$ in which all regressions pass through the origin. It may be more appropriate to allow a separate (non-zero) intercept for each genotype. This is equivalent to the model with genotype main effects, g, and a k-factor analytic model for g x e interaction. Then, the expression for g_s turns to

$$\begin{aligned} g_s &= (1_s \otimes I_g)g + ge \\ &= (1_s \otimes I_g)g + (\Lambda \otimes I_g)f + \delta . \end{aligned}$$

The vector g has mean zero and variance $\sigma_g^2 I$ or $\sigma_g^2 A$ where A is a genetic relationship matrix. The model can be written as

$$g_s = (\sigma_g 1_s \otimes I_g) f_0 + (\Lambda \otimes I_g) f + \delta$$
$$= (\Lambda_g \otimes I_g) f_g + \delta,$$

where:

$$\Lambda_g^{s(k+1)} = [\sigma_g 1_s \ \Lambda];$$
$$f_0 = g / \sigma_g;$$
$$f_g' = (f_0' f').$$

Thus the model with genotype main effects and a k-factor analytic model for g x e interactions is a special case of a (k + 1)-factor analytic genotype effects in each environment, in which the first set of loadings are constrained to be equal.

The feature that distinguishes equations for g_s from standard random multivariate regression problems is that both the covariates and the regression coefficients are unknown and therefore must be estimated from the data. The model is then a multiplicative model of environment and genotypes coefficients (known as loadings and factorial scores, respectively). Here lies the analogy with AMMI models. However, a key difference is that the multiplicative model in equation for g_s accommodates random effects, whereas AMMI is a fixed-effects model. The FAMM models are also called random AMMI.

2.3 General Linear Mixed Model and REML Estimation of Factor Analytic, Multivariate and Spatial Models

A general linear mixed model has the form (Henderson, 1984; Searle et al. 1992; Thompson et al., 2003):

$$y = X\beta + Z\tau + \varepsilon \quad (1),$$

with the following distributions and structures of means and variances:

$\tau \sim N(0,\ G)$ $\qquad E(y) = X\beta$

$\varepsilon \sim N(0,\ R)$ $\qquad Var(y) = V = ZGZ'+R$

where:
y: known vector of observations.
β: parametric vector of fixed effects, with incidence matrix X.
τ: parametric vector of random effects, with incidence matrix Z.
ε: unknown vector of errors.
G: variance-covariance matrix of random effects.
R: variance-covariance matrix of errors.
0: null vector.

Assuming G and R as known the simultaneous estimation of fixed effects and the prediction of the random effects can be obtained through the mixed model equations given by:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \tilde{\tau} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

The solution to this system of equations for $\hat{\beta}$ and $\tilde{\tau}$ leads to identical results as that obtained by:

$\hat{\beta} = (X'V^{-1}X)^{-}X'V^{-1}y$: generalised least estimator (GLS) or best linear unbiased estimator (BLUE) of β;

$\tilde{\tau} = GZ'V^{-1}(y - X\hat{\beta}) = C'V^{-1}(y - X\hat{\beta})$: best linear unbiased predictor (BLUP) of τ; where $C' = GZ'$: covariance matrix between τ and y.

When G and R are not known, the variance components associated can be estimated efficiently through the REML procedure (Patterson and Thompson, 1971; Searle et al., 1992; Thompson, 1973; 1977; 1980; 2002; Thompson and Welham, 2003). Except for a constant, the residual likelihood function (in terms of its log) to be maximised is given by:

$$L = -\frac{1}{2}\ (\log|X'V^{-1}X| + \log|V| + v\ \log\ \sigma_\varepsilon^2 + y'Py/\sigma_\varepsilon^2)$$

$$= -\frac{1}{2}\ (\log|C^*| + \log|R| + \log|G| + v\ \log\ \sigma_\varepsilon^2 + y'Py/\sigma_\varepsilon^2)$$

where:

$V = R + ZGZ'$; $\quad P = V^{-1} - V^{-1}X\ (X'V^{-1}X)^{-1}X'V^{-1}$

$v = N\text{-}r(x)$: degrees of freedom, where N is the total number of data and $r(x)$ is the rank of the matrix X.

C^*: Coefficient matrix of the mixed model equations.

Being general, the model (1) encompass several models inherent to different situations such as:

Univariate model

$G = A\sigma_\tau^2$; $\quad R = I\sigma_\varepsilon^2$, where:

σ_τ^2 : variance of the random effects in τ.

A: known matrix of relationships between the τ elements.

σ_ε^2 : residual variance.

Multivariate models

In the bivariate case:

$$Z = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}; \quad \tau = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix};$$

$G = A \otimes G_O$; $\quad R = I \otimes R_O$;

$$G_O = \begin{bmatrix} \sigma_{\tau_1}^2 & \sigma_{\tau_{12}} \\ \sigma_{\tau_{21}} & \sigma_{\tau_2}^2 \end{bmatrix}; \quad R_O = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & \sigma_{\varepsilon_{12}} \\ \sigma_{\varepsilon_{21}} & \sigma_{\varepsilon_2}^2 \end{bmatrix} \quad or \quad R_O = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & 0 \\ 0 & \sigma_{\varepsilon_2}^2 \end{bmatrix}, \ where:$$

$\sigma_{\tau_{12}}$: random treatment effects covariance between variables 1 and 2.

$\sigma_{\varepsilon_{12}}$: residual covariance between variables 1 and 2.

Spatial models (time series or geostatistical)

$R = \Sigma$: non-diagonal matrix that considers the correlation between residuals through ARIMA models or covariance based on adjusted semivariance.

In the context of the agricultural experiments, the general spatial model developed by Martin (1990) and Cullis and Gleeson (1991) has the following form:

$y = X\beta + Z\tau + \xi + \eta$, where:

y: known vector of data, ordered as columns and rows within columns;

τ: unknown vector of treatment effects;

β: unknown vector representing the spatial variation at large scale or global tendency (block effects, polynomial tendency);

ξ: unknown vector representing the spatial variation at small scale (within blocks) or local tendency, modelled as a random vector with zero mean and spatially dependent variance;

η: unknown vector of independent and identically distributed errors.

Through ARIMA models, the error is modelled as a function of a tendency effect (ξ) plus a non correlated random residual (η). So, the vector of errors is partitioned into $\varepsilon = \xi + \eta$, where ξ and η refer to the spatially correlated and independent errors, respectively. The traditional models of analysis do not include the ξ component.

Considering an experiment with rectangular shape in a grid of c columns and r rows, the residuals can be arranged in a matrix in a way that they can be considered as correlated within columns and rows. Writing this residuals in a vector following the field order (by putting each column beneath another), the variance of residuals is given by $Var(\varepsilon) = Var(\xi + \eta) = R = \Sigma$
$= \sigma_{\xi}^{2}[\sum_{c}(\Phi_{c}) \otimes \sum_{r}(\Phi_{r})] + I\sigma_{\eta}^{2}$, where σ_{ξ}^{2} is the variance due local tendency and σ_{η}^{2} is the variance of the independent residuals.

The matrices $\sum_{c}(\Phi_{c})$ *and* $\sum_{r}(\Phi_{r})$ refer to first order autoregressive correlation matrices with auto-correlation parameters Φ_{c} and Φ_{r} and order

equal to the number of columns and rows, respectively. In this case, ξ is modelled as a separable first order auto-regressive process (AR1 x AR1) with covariance matrix $Var(\xi) = \sigma^2_{\xi}[\sum_{c}(\Phi_c) \otimes \sum_{r}(\Phi_r)]$ (Gilmour, Cullis and Verbyla, 1997). The auto-regressive parameters are efficiently estimated by REML (Cooper and Thompson, 1973; Gilmour, Thompson and Cullis, 1995).

The mixed model equations and variance structure for spatial factor analytic models can be given by

$$\begin{bmatrix} \hat{\beta} \\ \tilde{g}_s \\ \tilde{\kappa} \end{bmatrix} = \begin{bmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} & Z'R^{-1}W \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + C^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \end{bmatrix} \text{ where:}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_s \end{bmatrix}; \qquad \tilde{g}_s = \begin{bmatrix} \tilde{g}_1 \\ \vdots \\ \tilde{g}_s \end{bmatrix}; \qquad \tilde{\kappa} = \begin{bmatrix} \tilde{\kappa}_1 \\ \vdots \\ \tilde{\kappa}_s \end{bmatrix}$$

$$R^{-1} = R_O^{-1} \otimes H^{-1}; \qquad G^{-1} = G_O^{-1} \otimes A^{-1}; \qquad C^{-1} = C_O^{-1} \otimes I$$

$$R_O = \begin{bmatrix} \sigma^2_{\xi_1} & 0 \\ & \\ 0 & \sigma^2_{\xi_s} \end{bmatrix}; \qquad G_O = \begin{bmatrix} \sigma_{g_{11}} & \sigma_{g_{1s}} \\ \sigma_{g_{1s}} & \sigma_{g_{ss}} \end{bmatrix}; \qquad C_O = \begin{bmatrix} \sigma^2_{\kappa_1} & 0 \\ 0 & \sigma^2_{\kappa_s} \end{bmatrix}, \text{ where:}$$

$$R^{-1} = \begin{bmatrix} H_1\sigma^2_{\xi_1} & 0 \\ 0 & H_s\sigma^2_{\xi_s} \end{bmatrix}^{-1}$$

β and κ: vectors of fixed effects and random plot effects, respectively.

$H_1 = [\sum_{c_1}(\Phi_{c_1}) \otimes \sum_{r_1}(\Phi_{r_1})]$: spatial correlation matrix for the environment 1;

$H_s = [\sum_{c_s}(\Phi_{c_s}) \otimes \sum_{r_s}(\Phi_{r_s})]$: spatial correlation matrix for the environment s;

$$H = \begin{bmatrix} H_1 & 0 \\ 0 & H_s \end{bmatrix}$$

In this case, the genotype main effects are fitted implicitly in $\tilde{g}_s = [\tilde{g}_1 \ldots \tilde{g}_s]'$. The explicit fitting of genotype main effects term is achieved by including another

random vector for these main effects in the mixed model equations. After that, the $\tilde{g}_s$ effects in the mixed model equations will represent g x e interactions.

Solving the mixed model equations above provides BLUPs of genotype effects in individual environments. The BLUPs of the genotype's factorial scores f can then be obtained from $\tilde{g}_s$ as

$$\tilde{f}_s = \hat{\text{var}}(f)[Z\ (\hat{\Lambda} \otimes I_g)]'\hat{P}y$$
$$= [\hat{\Lambda}'(\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi})^{-1} \otimes I_g]\tilde{g}_s.$$

The estimates are:

$\hat{\Lambda}$: matrix of estimated loadings;

$\hat{\Psi}$: matrix of estimated specific variances.

The BLUPs of the residuals of the g x e interactions can be obtained by

$$\tilde{\delta} = [\hat{\Psi}(\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi})^{-1} \otimes I_g]\tilde{g}_s.$$

It can be seen that the factor analytic model requires calculations of the parameters Λ and Ψ which compound the variance-covariance matrix G_0, and can be estimated by REML (Patterson and Thompson, 1971) through the algorithm average information (Gilmour, Thompson and Cullis, 1985; Johnson and Thompson, 1995). A specific REML algorithm for factor analytic models was developed by Thompson et al. (2003).

With assumption of the model $y = X\beta + Z[(\Lambda \otimes I_g)f + \delta] + \varepsilon$, the predicted effects of genotypes in an average environment ($\tilde{g}_{\bar{s}}$) can be given by the formula:

$$\tilde{g}_{\bar{s}} = \bar{\hat{\beta}} + [(\bar{\hat{\lambda}}_1\ \bar{\hat{\lambda}}_2 \ldots\ldots \bar{\hat{\lambda}}_k) \otimes I_g]\tilde{f}.$$

The quantities $\bar{\hat{\lambda}}_r$ and $\tilde{f}$ are the mean across environments of the estimated loadings for the rth factor, and the estimated factorial scores for genotypes, respectively. This is a prediction at the average values of the loadings. By definition of the loadings these are predictions of genotype means for an

environment that is average in the sense of having average covariance with all other environments. The prediction of overall genotype performance is the same irrespective of the inclusion of genotype main effects in the model. The issue of interpretation of the genotype main effects included is important. These are not main effects in the usual sense, namely a measure of overall genotype performance, but are merely intercepts in the regression. They therefore reflect genotype performance in an environment that has zero values of the loadings. That inclusion would provide results of genotype main effects identical to the predicted values for an average environment ($\widetilde{g}_{\bar{s}}$) (Smith, Cullis and Thompson, 2001).

One form of obtaining the overall performance of genotypes is by forming the two-way table of predicted genotype means for each environment and then averaging across environments to obtain the overall genotype means. These predicted means are also given by the formula:

$$\widetilde{g}_{\bar{s}m} = \overline{\hat{\beta}} + [(\overline{\hat{\lambda}}_1 \overline{\hat{\lambda}}_2 .....\overline{\hat{\lambda}}_k) \otimes I_g]\widetilde{f} + \overline{\hat{\delta}}$$

This formula differs from $\widetilde{g}_{\bar{s}}$ only by the adding of the unexplained g x e effects, which refers to the lack of fit from the factor analysis. This overall performance is only likely to be a good predictor if the correlation of genotype in different environments is high.

2.4 Constraints and Rotation on Loadings and Interpretation of Environmental Loadings and Factorial Scores

When the number k of factors is greater than 1, constraints must be imposed on the factor analytic parameters in order to ensure identifiability. This arises because the distribution of $(\Lambda \otimes I_g)f$ is singular. It can be shown that k(k-1)/2 independent constraints must be imposed on the elements of Λ. According to Mardia, Kent and Bibby (1988), the factor analytic model is not unique under rotation so the constraints must be chosen to ensure uniqueness. One set of constraints that fulfils this requirement is to set all k(k-1)/2 elements in the upper triangle of Λ to be zero, i.e., $\lambda_{jr} = 0 \; for \; j < r = 2...k$ (Jennrich and Schluchter, 1986). The implication of the constraints is that the number of variance

parameters in the factor analytic model with k terms is given by pk + p-k(k-1)/2 (Smith, Cullis and Thompson, 2001).

The nonuniqueness of Λ when k > 1 introduces ambiguity in the interpretation o the environmental loadings and genotype scores. The constrained form of Λ is merely for computational ease and has no biological basis. So, rotation of loadings is advocated for generating meaningful results. Lawley and Maxwell (1971) describe a number of useful rotations. In MET data the required rotation is $\Lambda^* = \Lambda T$, where T is an orthogonal matrix. According to Johnson and Wichern (1988), the axes can then be rotated in a certain angle ϕ and the rotated loadings can be given by $\Lambda^* = \Lambda T$, with $T = \begin{bmatrix} \cos\phi & \text{sen}\phi \\ -\text{sen}\phi & \cos\phi \end{bmatrix}$.

The loadings from factor analytic models are useful for clustering environments in terms of genetic correlations. The graphical display of loadings from a model with k > 1 can be very informative in this respect.

In factor analysis, the main interest is centred on the parameters of the factor model. Nevertheless, the predicted values of the common factors, named factor scores, are particularly useful in cluster analysis. Besides their utility in predicting genotype averages, the genotype's factorial scores can also be plotted for the factors 1 and 2 for example, permitting inference about the grouping of genotypes based on their similarity.

2.5 Goodness of Fit, Model Comparison and Fitting Procedure

Selection of FAMM Models

In a search for parsimonious models the adequacy of the FAMM models of several orders k can be formally tested, as it is fitted within a mixed model framework. The model with k factors, denoted FAk, is nested within the model with k + 1 factors. Models including the main genotype effect (g) are intermediate between the factor analytic models of order k (FAk) and of order FAk + 1. The model FA1 + g is intermediate to models FA1 and FA2. Residual maximum likelihood ratio tests (REMLLRT) can be used to compare such models. Other approaches for testing the goodness-of-fit of factor analytic models

involve comparisons with the unstructured covariance matrix (Mardia, Kent and Bibby,1988), which is very hard to obtain with a great number of environments.

Likelihood Ratio Test (LRT)

Given two nested models U and V with maximum of the residual likelihood function L(U) and L(V) and correspondent number of parameters n_u and n_v, it can be showed that D = -2 log L(U) −2 log L(V) approaches a chi square distribution with n_v - n_u degrees of freedom (assuming U as nested within V). Testing the significance of D against the appropriate chi square distribution constitutes the LRT test. When V is the saturated model, D is called deviance. So, alternatively, the difference between the deviance of the two models can be used to do the LRT test.

The LRT test can be used to compare fitted models provided they have a nested structure and the same fixed effects. This permits comparison of models with different random factors for a constant structure of fixed effects. For comparing spatial models, the LRT statistic can be used to assess the order of the model to be fitted. However, the use of the LRT is limited to models fitted under the same regime of differencing.

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)

Other criterion for model selection is the Akaike Information Criterion, which penalise the likelihood by the number of independent parameters fitted. By this criterion, any extra parameter must increase the likelihood at least by one unit for entering in the model. The AIC is given by AIC = -2 log L + 2 p, where p is the number of parameters estimated. Smaller values of AIC reflect a better global fit (Akaike, 1974). Other approach is the Bayesian Information Criterion (BIC) of Schwarz (1978), which is given by BIC = -2 log L + p log v, where v = N – r(x) is the number of residual degrees of freedom. BIC and AIC are calculated for each model and the model with the smallest value is chosen as the preferred model. AIC and BIC can be used for comparing non nested models, but the data should be the same which means the fixed effects should be the same.

Software

All models were fitted using the software ASREML (Gilmour and Thompson, 1998, 2002; Gilmour, Cullis, Thompson and Welham, 2002) which uses the REML

procedure through the average information algorithm (Gilmour, Thompson and Cullis, 1995; Johnson and Thompson, 1985; Thompson et al., 2003). The software GENSTAT (Thompson and Welham, 2003) was also used.

2.6 Applications

Two large unbalanced data sets were used. The first one concerned to 200 eucalypt treatments (progenies) evaluated for the trait trunk circumference on six sites in lattice designs with different replication numbers in each trial. The total number of plants evaluated was 65000. The second data set concerned to 60 tea plant treatments (progenies) evaluated in complete block designs for the trait leaf weight in three consecutive years and in two trials. Trial 1 provided 5400 observations (60 treatments x 5 replications x 6 plants per plot x 3 annual measures) and trial 2 provided 4050 observations (45 treatments, 5 replications, 6 plants per plot and 3 annual measures). The 45 treatments in trial 2 are also in trial 1.

2.6.1 Eucalypt Data Set

Results concerning to several models applied to eucalypt data set on six environments are presented in Table 1.

Table 1. Residual log-likelihoods (Log L) and likelihood ratio statistic (LRT) for the sequence models fitted to the eucalypt data.

Model for G	Log L	LRT	Var. param. ln G	Var. param. total
1.Uniform for g x e	-151100	-	1	3
2.Uniform for g	-149228	-	1	3
3.Uniform for g + g x e	-147892	2672	2	4
4.FA1, var. homog.	-147619	546	12	14
5.FA2, var. homog.	-147562	114	17	19
6.Multiv.var. homog.	-147556	12	21	23
7.FA1, var. heterog.	-146381	-	12	19
8.FA1 +g, var.heterog.	-146381	0	13	20
9.FA2, var. heterog.	-146325	112	17	24
10.Multiv. var. heterog.	-146318	14	21	28

The first part of Table 1 contains only models (1 to 6) fitted with assumption of homogeneous error variance. Model 1 fitted treatment effects on each environment and considered a common error variance for all environments. Model 2 fitted treatment effects on an average environment and considered a

common error variance for all environments. Model 3 fitted treatment effects on an average environment plus g x e interaction and considered a common error variance for all environments. Model 4 fitted a factor analytic structure of order 1 for treatment effects and considered a common error variance for all environments. Model 5 fitted a factor analytic structure of order 2 for treatment effects and considered a common error variance for all environments. Model 6 fitted a full multivariate unstructured for treatment effects and considered a common error variance for all environments. The second part of the same table contains only models (7 to 10) with assumption of heterogeneous error variance. Models 7 and 9 fitted a factor analytic structure of order 1 and 2, respectively, for treatment effects. Model 8 fitted a factor analytic structure of order 1 for treatment effects plus treatment main effects. Model 10 fitted a full multivariate unstructured for treatment effects.

Contrasting the two parts in terms of the Log L it can be seen that the models allowing error variance heterogeneity are far better than the models assuming variance homogeneity. This shows the superiority of FAMM models over AMMI models, which do not consider the error variance heterogeneity. Common error variance for all trials is implicit in the AMMI approach. Even the full multivariate model (6) for G_0 (21 parameters) with homogeneous variance is worst than the FA1 model (7) for G_0 (12 parameters) with heterogeneous variance. This confirms the great importance of considering error variance heterogeneity in MET analysis. And this can only be done in the mixed modelling framework. So, it is a great advantage the factor analytic models being embedded in this framework.

Other important feature of the FAMM models is the providing of parsimonious models in relation to the full unconstrained multivariate approach. The multivariate approach is prohibitive with a great (usually > 5) number of environments, generating over-parameterised and hard-to-converge models. Results from Table 1 reveal that the model FAMM with two factors (FA2) is practically equivalent (REMLLRT of 14 and 12 on 4 degrees of freedom, p value $> .01$) to the full multivariate model in both situations, with and without allowing for variance heterogeneity. So, in practice a model with four less parameters can be used. It is worthy mention that all the FAMM models converged without a need for constraining the G_0 matrix.

A Model including the main genotype effect (g) is intermediate between the factor analytic models of order k (FAk) and of order FAk + 1, as it is FAk + 1 with

constraints. The model FA1 + g is intermediate to models FA1 and FA2. In the present data set the models FA1 and FA1 + g were equivalent, giving the same Log L. In fact, the estimate of the variance component for genotype effects was on the boundary; that is it was estimated as zero. The role of genotype main effects in an FA model is purely in terms of the search for a parsimonious variance structure between a given FAk model and a FAk + 1 model. The approach for prediction of overall genotype means across environments is the same irrespective the inclusion of genotype main effects (Smith, Cullis and Thompson, 2001). In a factor analytic context, the model without genotype main effects is equivalent to a model for genotype effects in each environment.

Overall, the best parsimonious model was the FA2 with heterogeneous variance for errors (model 9 in Table 1). Results concerning to loadings, common, specific and error variances provided by this model are presented in Table 2.

Table 2. Estimated loadings (on the correlation scale), common (communality), specific and error variances for the model FA2 fitted to the eucalypt data.

	Original Loadings and (Rotated)		Common Variance (%)	Specific Variance (%)	Error Variance
Location	Factor 1	Factor 2			
1. L1	0.845 (0.433)	0.498 (0.880)	0.962	0.038	20.0422
2. L2	0.791 (0.443)	0.398 (0.767)	0.784	0.216	20.5270
3. L3	0.837 (0.450)	0.454 (0.839)	0.907	0.093	22.6041
4. L4	0.907 (0.596)	0.295 (0.745)	0.910	0.090	44.5751
5. L5	0.979 (0.761)	0.104 (0.624)	0.969	0.031	38.0380
6. L6	0.904 (0.837)	-0.149 (0.372)	0.839	0.161	28.9856
Eigenvalues	4.639	0.710			
Accu. Var. Explained	0.773	0.892			

It can be seen that the FA2 model explained a large amount (almost 90%) of the total genotypic variance. The first factor explained 77.3% of the variation and the second factor added 11.9%. The specific variances (in percentage of the total) were low, except for the environments 2 and 6, which were 22% and 16%, respectively. The high values of the common variance (or communality) show that the two factors explained a great percentage of the variance of each

environment and that the FA2 model fitted well to the data set (Table 2).

The genotypic variance-covariance matrix and the correlations (obtained by $\Lambda\Lambda'+\Psi$ from model FA2 on the correlation scale) involving the several environments are presented in Table 3.

Table 3. Estimated genotypic covariance\variance\correlation matrix associated to model FA2 applied to eucalypt data set.

	L1	L2	L3	L4	L5	L6
L1	**6.312**	0.867	0.933	0.914	0.879	0.689
L2	6.964	**10.225**	0.843	0.835	0.812	0.655
L3	7.375	8.481	**9.905**	0.893	0.867	0.689
L4	8.132	9.463	9.959	**12.555**	0.919	0.776
L5	6.566	7.754	8.108	9.682	**8.837**	0.869
L6	5.135	6.207	6.425	8.148	7.659	**8.784**

It can be observed that there is heterogeneity among the specific variances concerning to several environments (diagonal of Table 3). This justifies the use of models with heterogeneous specific variances. Piepho (1997, 1998) proposed the use of a factor analytic model with common specific variance for all sites. However, Smith, Cullis and Thompson (2001) noted that models with heterogeneous specific variances were significantly better. It can be seen that there is also heterogeneity of covariance between the several combinations of environments. These covariances represent the genotypic variance free from interaction effects between each two sites. This heterogeneity explains the better fit of FAk and multivariate models over the model 3, which includes g + g x e. When there are only two environments, the bivariate and model 3 tend to give the same fitting (see results from tea plant data set).

Results about correlations reveal that the first four environments have smaller correlations with the environment 6, which has higher correlations with environment 5 (Table 3). It can be observed that factor analysis put greater emphasis on environments 5 and 6 in the factor 1 (rotated loadings higher than 0.76) and higher emphasis on sites 1, 2, 3 and 4 in factor 2 (rotated loadings higher than 0.74) (Table 2). This is the logic of factor analysis: to separate groups of traits with high correlations between them in each group and then put higher weights in traits of a group in one factor (factor 1) and higher weights in

traits of another group in the other factor (factor 2). Plotting the first set of loadings against the second will show the clustering of environments: L1, L2, L3 and 4 close together in one group and L5 and L6 in a second group. Other advantage of FAMM models over AMMI is that they provide an estimate of the full correlation structure, facilitating practical decisions to be made.

The FAMM and AMMI models are also useful for the clustering of environments based on their similarity in terms of genetic correlations. This can be done through biplots (AMMI) or plot of loadings from the first factor against the loadings from the second factor (FAMM). The full structure of correlation provided by the FAMM models can be also subjected to methods of cluster analysis or other multivariate methods. Such methods traditionally operate on correlations estimated by pairs of environments through balanced ANOVA. The FAMM models use the information on all environments simultaneously to give the correlation for pairs of environments, so providing more precise estimates.

2.6.2 Tea Plant Data Set

Multi-environment Spatial Analysis for each Trait

The two trials contain 45 treatments in common, so it was possible to analyse all data simultaneously. Although not all progenies were represented in the two environments, the FAMM models were applied. An important remark is that the factor analysis under the mixed model can be done with incomplete data sets.

Firstly, multi-environment spatial analyses were done for each trait in a combination of the two trials. Three objectives pursued by breeders were considered: selection for specific environments (multivariate multi-environment spatial model), selection for an average environment (univariate multi-environment spatial model), selection for a non-tested environment (univariate multi-environment spatial models, including the genotype x environment interaction effects). The main features concerning variance structures of the models are presented in the sequence.

Selection for Specific Environments

$R^{-1} = R_O^{-1} \otimes H^{-1}$; $G^{-1} = G_O^{-1} \otimes A^{-1}$, where:

$$G_O = \begin{bmatrix} \sigma^2_{\tau_1} & \sigma^2_{\tau} \\ \sigma^2_{\tau} & \sigma^2_{\tau_2} \end{bmatrix}$$

$\tilde{\tau}_i$: vector of genetic effects in the environment i;

σ^2_{τ} : genetic covariance between two environments.

Selection for an Average Environment

$R^{-1} = R_O^{-1} \otimes H^{-1}$; $G^{-1} = (1/\sigma^2_{\tau_m}) A^{-1}$, where:

$$R_O = \begin{bmatrix} \sigma^2_{\xi_1} & 0 \\ 0 & \sigma^2_{\xi_2} \end{bmatrix} \quad H = \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} \quad R^{-1} = \begin{bmatrix} H_1\sigma^2_{\xi_1} & 0 \\ 0 & H_2\sigma^2_{\xi_2} \end{bmatrix}^{-1}$$

$H_1 = [\sum_{r_1} (\Phi_{r_1}) \otimes \sum_{r_1} (\Phi_{r_1})]$: spatial correlation matrix for the environment 1;

$H_2 = [\sum_{r_2} (\Phi_{r_2}) \otimes \sum_{r_2} (\Phi_{r_2})]$: spatial correlation matrix for the environment 2;

$\tilde{\tau}_m$: vector of genetic effects in an average environment;

$\sigma^2_{\tau_m}$: genetic variance for an average environment.

Selection for a New Environment

$R^{-1} = R_O^{-1} \otimes H^{-1}$; $G^{-1} = (1/\sigma^2_{\tau}) A^{-1}$; $Q^{-1} = (1/\sigma^2_{ge}) I$

σ^2_{ge} : variance of the g x e interaction effects;

ge: vector of g x e interaction effects;

Q: variance-covariance matrix of g x e interaction effects.

Results concerning to the first objective are presented in Table 4. The plot effect was not fitted because it was non-significant with spatial analysis.

Table 4. Estimates of the variance parameters: genetic among treatments (progenies) in environment 1 ($\hat{\sigma}^2_{\tau_1}$), genetic among treatments (progenies) in environment 2 ($\hat{\sigma}^2_{\tau_2}$), genetic covariance among treatments across sites ($\hat{\sigma}_{\tau_{12}}$), correlated residual in site 1 ($\hat{\sigma}^2_{\xi_1}$), correlated residual in site 2 ($\hat{\sigma}^2_{\xi_2}$), non-correlated residual in site 1 ($\hat{\sigma}^2_{\eta_1}$), non-correlated residual in site 2 ($\hat{\sigma}^2_{\eta_2}$), narrow sense heritability in site 1 ($\hat{h}^2_1$), narrow sense heritability in site 2 ($\hat{h}^2_2$), respective adjusted heritabilities ($\hat{h}^2_{adj_1}$ and $\hat{h}^2_{adj_2}$) and residual auto-correlation coefficients between columns (AR Column i) and rows (AR Row i), in the specific trial or site i.

Parameters estimates	First year	Second year	Third year
$\hat{\sigma}^2_{\tau_1}$	0.0157 ± 0.004	0.1074 ±0.02	0.3573 ±0.08
$\hat{\sigma}^2_{\tau_2}$	0.0214 ± 0.005	0.0978 ±0.02	1.1526±0.27
$\hat{\sigma}_{\tau_{12}}$	0.0087± 0.003	0.0585±0.02	0.3669 ±0.12
$\hat{\sigma}^2_{\xi_1}$	0.0296± 0.006	0.1439±0.03	0.9032±0.18
$\hat{\sigma}^2_{\xi_2}$	0.0183 ± 0.018	0.1286±0.04	1.9108 ±0.62
$\hat{\sigma}^2_{\eta_1}$	0.0948±0.004	0.4326±0.02	1.7135±0.07
$\hat{\sigma}^2_{\eta_2}$	0.0797±0.003	0.3531±0.02	3.2352±0.14
$\hat{h}^2_1$	0.4492	0.6283	0.4806
$\hat{h}^2_2$	0.7163	0.7017	0.7261
AR Column 1	0.8073±0.05	0.8463±0.04	0.8875±0.03
AR Row 1	0.8000±0.05	0.7967±0.05	0.8137±0.05
AR Column 2	0.9816±0.03	0.9192±0.04	0.9603±0.02
AR Row 2	0.9960±0.01	0.9482±0.02	0.9100±0.03
Deviance	-3966.24	829.13	6393.20
$\hat{h}^2_{adj_1} = (4\hat{\sigma}^2_{g_1})/(\hat{\sigma}^2_{g_1} + \hat{\sigma}^2_{\eta_1})$	0.5683	0.7956	0.6902
$\hat{h}^2_{adj_2} = (4\hat{\sigma}^2_{g_2})/(\hat{\sigma}^2_{g_2} + \hat{\sigma}^2_{\eta_2})$	0.8466	0.8676	1.04

The genetic correlations between environments were about 0.48, 0.57 and 0.57 for leaf yield in years 1, 2 and 3, respectively. The magnitudes of these correlations reveal a need for specific selection for each site. The bivariate model involving the two sites was fitted also assuming variance homogeneity across sites and independent errors. The deviance values obtained were -3756.5, 1127.04 and 6883.92, for the three traits, respectively. These are much higher than the -3966.24, 833.70 and 6393.74 obtained with the model allowing heterogeneity of variance and spatial errors. Such results reinforce that FAMM models could be more adequate than AMMI models, which do not allow for heterogeneity of variance and spatial errors. The residual auto-correlation coefficients were very high for the site 2 and spatial analysis could be abdicated for this site without loss of efficiency.

Results concerning to the second objective are presented in Table 5.

Table 5. Estimates of the variance parameters: genetic among treatments (progenies) in an average environment ($\hat{\sigma}^2_{\tau}$), correlated residual in site 1 ($\hat{\sigma}^2_{\xi_1}$), correlated residual in site 2 ($\hat{\sigma}^2_{\xi_2}$), non-correlated residual in site 1 ($\hat{\sigma}^2_{\eta_1}$), non-correlated residual in site 2 and respective residual auto-correlation coefficients between columns (AR Column i) and rows (AR Row i), in the specific trial or site i.

Parameters estimates	First year	Second year	Third year
$\hat{\sigma}^2_{\tau}$	0.01397±0.003	0.0797±0.02	0.4310 ±0.09
$\hat{\sigma}^2_{\xi_1}$	0.0338± 0.007	0.1606±0.03	0.9470±0.18
$\hat{\sigma}^2_{\xi_2}$	0.0182 ± 0.005	0.1350±0.04	1.9259±0.62
$\hat{\sigma}^2_{\eta_1}$	0.09757±0.004	0.4423±0.02	1.7335±0.07
$\hat{\sigma}^2_{\eta_2}$	0.07531±0.004	0.3580±0.02	3.4773±0.15
AR Column 1	0.8049±0.05	0.8154±0.05	0.8766±0.03
AR Row 1	0.8365±0.05	0.8094±0.05	0.8169±0.05
AR Column 2	0.8893±0.05	0.9000±0.04	0.9487±0.02
AR Row 2	0.7336±0.08	0.9290±0.03	0.9103±0.03
Deviance	-3917.40	883.21	6483.50

This model (Table 5), albeit more parsimonious than the full multivariate (Table 4), gave a significant higher deviance and higher AIC value. So, the multivariate is preferred and selection for an average environment can be done by taking means of predicted genetic values in each environment. The superiority of the multivariate model can be explained by the heterogeneity of genetic variance across sites (Table 4). Data standardisation should correct this and make the univariate (for an average environment) model suitable.

Results concerning to the third objective are presented in Table 6.

Table 6. Estimates of the variance parameters: genetic among treatments (progenies) free of g x e interaction effects ($\hat{\sigma}^2_{\tau}$), g x e interaction effects ($\hat{\sigma}^2_{ge}$), correlated residual in site 1 ($\hat{\sigma}^2_{\xi_1}$), correlated residual in site 2 ($\hat{\sigma}^2_{\xi_2}$), non-correlated residual in site 1 ($\hat{\sigma}^2_{\eta_1}$), non-correlated residual in site 2 and respective residual auto-correlation coefficients between columns (AR Column i) and rows (AR Row i), in the specific trial or site i.

Parameters estimates	First year	Second year	Third year
$\hat{\sigma}^2_{\tau}$	0.00865±0.003	0.0588±0.02	0.3305±0.12
$\hat{\sigma}^2_{ge}$	0.00976±0.003	0.0442±0.01	0.3412±0.09
$\hat{\sigma}^2_{\xi_1}$	0.0298± 0.007	0.1437±0.03	0.9047±0.18
$\hat{\sigma}^2_{\xi_2}$	0.0183 ± 0.02	0.1286±0.04	1.8799±0.64
$\hat{\sigma}^2_{\eta_1}$	0.09469±0.004	0.4327±0.02	1.7111±0.07
$\hat{\sigma}^2_{\eta_2}$	0.07979±0.003	0.3505±0.01	3.2502±0.14
AR Column 1	0.8078±0.05	0.8466±0.04	0.8888±0.03
AR Row 1	0.8004±0.06	0.7968±0.05	0.8144±0.05
AR Column 2	0.9817±0.03	0.9187±0.04	0.9591±0.02
AR Row 2	0.9959±0.09	0.9485±0.02	0.9161±0.04
Deviance	-3965.28	829.22	6409.64

Comparing results from Tables 5 and 6 it can be seen by the deviance values that the model with interaction (Table 6) fits better to data, revealing the significance of the g x e interaction effects.

This model gave approximately the same deviance and smaller AIC values in relation to the full multivariate (Table 4). Then it should be preferred. The g x e component encompassed all the heterogeneity of genetic variance. From this model, predicted genetic values can be derived for each treatment (parent or individual) in each environment by summing the correspondent g and g x e predicted effects. After, the mean of predicted genetic values of each treatment over several environments can be taken aiming at the selection for an average environment.

Another alternative is the obtaining of treatment effects in each environment directly by fitting only the g x e component, i.e., overlooking the g main effects. Applying this approach for the measure in the first year, the variance component for g x e obtained was 0.01858 which is approximately equivalent to the sum of variance component for g and g x e presented in Table 6, as expected. The deviance obtained was -3957.20 which is significantly (by LRT) higher then the -3965.28 reported in Table 6. This shows that the model with g is better.

Factor Analytic Models (Spatial and Non-Spatial) for Multivariate and Multi-Environment Data

Although the univariate model with g and g x e for treatment effects is sufficient for the multi-site analysis of individual traits, the univariate approach is not appropriate for all six measures together due to the great variance heterogeneity between measures in each site. So, a multivariate approach for the six traits together with fit of individual permanent effects in each site was adopted. The fit of permanent effects aimed at the elimination the residual covariance between measures in each site. The model is an extension (increasing the number of traits to six and including permanent effects) of that concerning to selection for specific environments.

However, the fit of this model not converged with spatial errors and a non-spatial model was fitted. Results are presented in the sequence together with the factor analytic models, which were fitted as alternative parsimonious models.

Results concerning to factor analytic models for the six repeated measures in two environments in comparison with the multivariate model are presented in Table 7. In all models the individual permanent effects were fitted as a mean of eliminating the residual correlation between repeated measures in each site.

Table 7. REML log-likelihoods (LogL) and REMLLRT (LRT) for comparing models of fitting covariances structures involving six traits. Models fitted were multivariate for treatments and non-spatial for residuals (MNS), factor analytic of order 1 for treatments and non spatial for residuals (FA1NS), factor analytic of order 1 for treatments and spatial (including both the correlated and independent term) for residuals (FA1S).

Number of Variance parameters					
Model for G	G	Total	LogL	LRT(P value)	%Variance
MNS	21	28	-2335.67	-	
FA1NS	12	19	-1848.10	975.14(0.001)	
FA1S	12	37	-585.31	2525.58(0.001)	71.5

It can be seen that the best model was the factor analytic with spatial error (FA1S). This model was superior to that one with non-spatial error (FA1NS). This fact is sufficient to show the superiority of factor analytic multiplicative mixed models (FAMM) over the additive main and multiplicative interaction effects (AMMI), which assumes fixed treatment effects and do not permit to model separate spatial errors. The proportion of genetic variance explained by the FA1S was 71.5%. This value is sufficient for the purpose of the analysis, i.e., genetic selection.

The non-spatial factor analytic model showed to be superior to the non-spatial multivariate model (MNS), revealing the advantages of the factor analytic models in terms of parsimony and ability of fitting. The MNS model, although with more parameters, showed a smaller LogL and was hard to converge, demanding restriction on G to be positive definite. Even so, the convergence was not so reliable, as ASREML fixed some variance components on the boundaries. In fact, it might not converged to a maximum likelihood solution. Other models like the full multivariate with spatial error and factor analytic of order 2 did not converge.

Results concerning to genetic correlation for the best model (FA1S) are presented in Table 8.

Table 8. Estimated genetic correlations obtained from the FA1S modelling.

Genetic correlations						
Trait	1	2	3	4	5	6
1	1	0.982	0.999	0.665	0.852	0.745
2		1	0.982	0.585	0.794	0.653
3			1	0.664	0.851	0.744
4				1	0.870	0.862
5					1	0.935
6						1

The estimated correlations are relatively coherent with previous estimates and expectation: higher correlation between repeated measures within site and lower correlations across sites. This, together with the suitable proportion of genetic variance explained by the FA1S model reveals the adequacy of the factor analytic model for analysis of this sort of data. Otherwise, the whole data set could not be analysed simultaneously. The variograms showed adequate behaviour.

Gilmour and Thompson (2002) reported the computational aspects of analysing six traits in an animal breeding context, when some traits are highly correlated. They conclude that the Factor Analytic and Cholesky models appear best in this situation. We confirm the adequacy of FA models. The Cholesky appear to be inadequate for our data set with errors non-correlated across traits, as we fit the permanent effect to account the correlation across traits within sites and the errors are non-correlated across sites.

Practical experiments with several perennial plants generate annually throughout the world a large amount of data on repeated measures. These measures are usually taken only three or four times before selection, since more than that, leads to less genetic gain per unit of time. Suitable models should be found for application in such kind of data in one or several experiments simultaneously. For analysing multi-environment data sets with longitudinal data, the factor analytic multiplicative mixed model proved to be a very useful tool, mainly when applied together with spatial analysis. The software ASReml showed to be

essential for modelling the complex data structure involving repeated measures, spatial dependency and multi-environment data sets in perennial plants. FAMM and FAMMS models can also be used for studies concerning QTL (quantitative trait loci) x environment interaction. This approach can be better than that advocated by Romagosa et al. (1996), based on AMMI analysis.

2.7 Conclusions

- Parsimonious FAMM models were found for the two data sets: FA2 for eucalyptus data set and FA1 for tea plant data set.
- There were great advantages of heterogeneous variance FAMM models over homogeneous variance FAMM models. This reveals the superiority of FAMM models over AMMI models.
- It was noted heterogeneity among the specific variances in individual environments so factor analytic models with common specific variances for all sites were not suitable.
- FAMM models provided estimates of the full correlation structure, facilitating practical decisions to be made.
- FAMM models with heterogeneous variance among traits and spatial errors within traits were advantageous over FAMM models with variance homogeneity and non-spatial error. This also shows the superiority of FAMM models over AMMI models, which do not allow for dependent or spatial errors.
- For analysing multi-environment data sets with longitudinal data, the FAMM models proved to be a very useful tool, mainly when applied together with spatial analysis.

3. Analysis of Interference and Environmental Trend in Field Trials by Joint Modelling of Competition and Spatial Variability

3.1 Introduction

Analysis of plant field experiments should be based on realistic approaches taking into account the biological process associated to the trait evaluated as well as the environmental influences. Experimental designs play a key role in providing reliable data sets for analysis. However, the local control schemes relying on block can be inefficient in accounting of all environmental gradients and trends and even the incomplete blocks do not provide a complete evaluation of the environmental effects. The spatial dependency or environmental trend within blocks, due to fertility and other environmental effects, should be considered through appropriate models of spatial analysis. Additionally, competition effects of neighbouring plants can also cause bias in treatment comparisons, due to interference of one genotype on phenotypic response of a neighbour plant or plot. So competition models should be also employed aiming at evaluation of interference effects.

There are two underlying assumptions in the classical block model. Firstly, that the fertility associated with plots in a block is constant (or nearly so). Secondly, that the response on a plot due to a particular treatment does not directly affect the response on a neighbouring plot. The first assumption is concerned with an environmental or residual effect called spatial trend, whilst the second assumption is concerned to treatment effect and is referred to as interference (Durban, Hackett and Currie, 1999). The trend effect is of common occurrence and correction for it is likely to increase heritability and precision estimates, as it is an environmental effect. The interference can occur only in some plant species and in determined phase of growth. So it depends on the biology of the species and its adjustment is likely to reduce the heritability estimates, as it is concerned to treatment effects. Adjustments for both effects are likely to reduce bias.

Experimenters should know about the competition effects in the species subjected to research aiming to choose between models with or without competition effects. Such effects has been reported in several important crops such as wheat, barley, oat, triticale, field beans, rice, cassava, sugar beet,

potatoes, swedes, kale (Talbot et al., 1995). In perennial plants, competition has also been found in forest trees (Correll and Anderson, 1983), cocoa (Glendinning and Vernon, 1965; Lotode and Lachenaud, 1988), oil palm (Nouy et al., 1990) and robusta coffee (Montagnon et al., 2001). Interference depends also on the size and form of the plot and has been reported to be very common in sugarcane trials designed in single-furrow plots (Stringer and Cullis, 2002a and b). In forest trees, competition effects depend mainly on age of measurement. Under competition effects, the best genotypes tend to exhibit overestimates of their superiority due to greater aggressiveness over the worst genotypes which exhibit sensitivity to competition. Models for evaluating the aggressiveness and sensitivity of genotypes were presented by Kempton (1982).

An important feature of the plant interference and spatial trend effects is their influence on the fitted models. Spatial trend generates positive auto-correlation between neighbouring plants or plots and plant interference due to competition generates negative auto-correlation between them. Firstly fitting of spatial models can reveal the need for competition models. High (say > 0.3) positive auto-correlation coefficients estimates obtained in spatial analysis reveal that spatial trend is predominant over competition and negative or near zero auto-correlation coefficients estimates reveals strong competition effects probably together with spatial trend. Also, firstly fitting a competition model can reveal the significance of such effects. In some circumstances, modelling only one of the effects, can be inappropriate. So the two effects should be modelled together. Durban, Currie and Kempton (2001) reported stronger fertility trend and stronger competition effects estimates in sugar beet when adjusting for these two effects simultaneously than when the two effects were modelled separated.

Fertility trend has been well accommodated through the residual autoregressive models of Gleeson and Cullis (1987), Cullis and Gleeson (1991) and Gilmour, Cullis and Verbyla (1997). Models for competition in plants have been proposed. Mead (1967) presented a theory of the original pure-stand competition. Other relevant papers are Pierce (1957), Draper and Guttman (1980), Kempton (1982), Besag and Kempton (1986), Pithuncharurnlap, Basford and Federer (1993), Talbot et al. (1995), Durban, Hackett and Currie (1999), Durban, Currie and Kempton (2001), Stringer and Cullis (2002b). Such models will be considered in details in the next section. Pierce (1957), Draper and Guttman (1980), Kempton (1982), Besag and Kempton (1986) took into account only competition through

autoregressive models at phenotypic and/or also at treatment or genotypic levels. Pithuncharurnlap, Basford and Federer (1993) considered simultaneously environmental trend through the residual autoregressive model of Gleeson and Cullis (1987) in one direction and competition through the genotypic approach of Besag and Kempton (1986). Durban, Hackett and Currie (1999) and Durban, Currie and Kempton (2001) considered simultaneously the two effects by modelling trend through cubic smoothing splines within blocks and interference by the phenotypic model of Kempton (1982). Stringer and Cullis (2002b) attempted to the joint modelling of spatial and competition effects through the methods of Gilmour, Cullis and Verbyla (1997) and genotypic model of Besag and Kempton (1986), respectivamente.

The present paper aims at accounting simultaneously for trend and interference in field trials of perennial plants such as forest trees and sugarcane. The objectives are the comparison and extension of alternative models, the quantification of competition levels in these species and the inference about the need for more complex models in routine of data analysis in these crops.

3.2 Competition Models

A simple tool to diagnostic the presence of competition effects in a field trial consists in performing the scatter plot between the residual of a central plot (adjusted for genotype and block effects) and the mean of the adjacent plots (adjusted for blocks aiming at the elimination of the positive correlation due to fertility). Also, a correlation coefficient between residuals and the performance (corrected for blocks) of the neighbours can inform about the presence of competition. Spatial analysis through autoregressive models and sample variogram can inform about the competition as well. Low positive and negative auto-correlation coefficients in spatial analysis show the presence of competition. Sample variograms exhibiting spikes and high and low points (alternating ridges) reveal negative correlation between residuals and so competition.

3.2.1 Phenotypic Interference

Kempton (1982) presented the following model for competition.

$$Y_{ij} = \tau_i + \beta X_j + \varepsilon_{ij}, \qquad (1)$$

where:

y_{ij} : observed value of the genotype i in plot j;

τ_i : fixed effect of treatment or genotype i;

β : competition coefficient, common to all genotypes;

X_j : mean of the neighbouring plots of the genotype i in plot j;

ε_{ij} : error independently and normally distributed with zero mean and variance σ^2.

The model assumes observations adjusted for the general mean and ignores the block effect. The covariate X is given by $X = \sum y / p$ where p is the number of neighbouring plots considered. Normally p can be 2 (evaluation at plot level, several plants per plot), 4 (evaluation at plant level with one or several plants per plot) or 8 (evaluation at plant level with one or several plants per plot).

The τ_i effect represents the genotype effect expected when the variety is grown under the competitive stress of the trial. Its performance in monoculture is estimated by $\tau_{ic} = \tau_i /(1 - \beta)$. Since β is negative, it can be seen that the performances of the best treatments are reduced after the correction for competition. This is because under competition the more aggressive varieties tend to have their performances overestimated in detriment of the more sensitivity varieties. If the experimenter is interested in assessing comparative varietal performance in monocultures, this correction should be made. The differences observed between performances of genotypes in the trials and in commercial plantings arise partially because the allocation of varieties in trials are not balanced for neighbouring varieties, but largely because a selected variety is likely to be highly competitive in the trial and therefore plants are liable to show natural depression in yield when grown as a monoculture. This has been observed in sugarcane in Brazil, confirming a need for corrections to be made.

The parameters can be estimated simultaneously by least square through the following set of equations.

$$\hat{\tau}_i = (g/n)\sum_j (Y_{ij} - \hat{\beta}X_j)$$

$$\hat{\beta} = \sum_j^n (Y_{ij} - \hat{\tau}_i)X_j / (\sum_{j=1}^n X_j^2).$$

The summation in equation for τ_i extends only over the set of plots j containing the genotype i (n/g plots, where n is the total number of plots in the trial and g is the number of genotypes or treatments). In the equation for β all plots are used, as the competition coefficient is common for all genotypes. β is a regression coefficient relating the residuals with the mean value (as a covariate) of the neighbouring plants or plots.

This least square approach is valid when the covariate is another variate different from the main trait of interest, for example, the main trait being the yield and the covariate being the plant height. However, when the covariate is defined to be the same as the main trait (for example, both being yield), the least square approach produces an invalid estimate of β (as the competition coefficient appears in both the mean and variance of y). An efficient estimation can be performed using maximum likelihood. The significance of β in the model can be tested through the likelihood ratio test. The omission of the competition effect can increase the deviance, ℓ, of the model. To test for the significance of competition, having adjusted for varietal effects, the statistic $\ell(y \mid \hat{\sigma}, \hat{\tau}) - \ell(y \mid \hat{\sigma}, \hat{\tau}, \hat{\beta})$ should be used, which under the null hypothesis should approximate to a χ^2 distribution with 1 degree of freedom.

The same model can be re-written by considering only two plants or plots as neighbours:

$$Y_{ij} = \tau_i + (1/2)\beta (Y_{j+1,s} + Y_{j-1,t}) + \varepsilon_{ij} \qquad (2)$$

where $Y_{j+1,s}$ and $Y_{j-1,t}$ are the performances (for the same trait) of genotypes s and t in plots neighbouring the genotype i. Situations can exist where the competition coefficients depend on the particular genotypes grown in the plots. In such cases, specific competition coefficients $\beta_{is} = \delta_i \gamma_s$ may be demanded, where δ_i represents the sensitivity of the genotype i to competition and γ_s represents the aggressiveness of genotype s and may be standardised so

that $\sum_s \gamma_s = g$. So the model (2) can be re-written as

$$Y_{ij} = \tau_i + (1/2)(\beta_{is} Y_s + \beta_{it} Y_t) + \varepsilon_{ij}$$

In matrix notation the model (2) can be re-written as (Besag and Kempton, 1986):

$$y = Xb + Z\tau + \beta W y + \varepsilon \qquad (3)$$

where:

W: is a n x n weight or regressor matrix which has the off-diagonal elements (j, j ± 1) or the principal off-diagonals equal to (1/2), otherwise zero;

b: is a vector of design features such as blocks, with incidence matrix X.

The vector τ can be interpreted as centred genotype effects in the absence of competition or under the average competitive stress in the trial. But when grown in a monoculture, the best varieties would produce a more competitive environment than that of the trial average and so will not perform as well as in the trial. Then τ should be divided by a factor $(1-\beta)$ to represent the pure stand effects. The competition effects increase the range and variability of genotype effects as they amplify the values of the more aggressive genotypes. The correction using the factor $(1-\beta)$ causes shrinkage in genotype effects, leading to more realistic results.

According to Kempton (1985), an alternative form for the model (2) is $Y_{ij} = \tau_{ic} + \beta(Y_{j+1} + Y_{j-1} - 2Y_j) + \varepsilon_{ij}$. In such a case the treatment effect is fitted already corrected for the neighbour effects, i.e., represents the pure stand productivity.

3.2.2 Genotypic Interference

Draper and Guttman (1980) have ignored the errors in Y_s and Y_t and used the model (2) as

$$Y_{ij} = \tau_i + (1/2)\beta(\tau_s + \tau_t) + \varepsilon_{ij} \qquad (4)$$

This model considers that the competition have more to do with the genotype rather than with the phenotype of the plants. This makes sense, since the aggressiveness and sensitivity of the genotypes are likely to be due to genetic causes and also to depend on another traits like height, canopy size and tillering

ability. In such model, the regression coefficient relates the genetic effect of the neighbours to the residual value of each central plant.

Pierce (1957) considered a model of plot interference in which each treatment i has a direct effect τ_i on the plot to which it is applied and a neighbour effect ϕ_i on each neighbouring plot. Genotype competition can be considered in this way, as the causes of competition are often unknown. Following Besag and Kempton (1986), the model is of the form:

$$y = Xb + Z\tau + NZ\phi + \varepsilon \qquad (5)$$

where:

ϕ : is a vector of centred on neighbour treatment effects (indirect effect produced on neighbours), which are genotypic and not phenotypic;

N: is the neighbour incidence matrix of dimension n x n, composed by 0 and 1.

It can be seen explicitly from model (5) that competition effects are concerned with treatment effects (depend on Z matrix) and not residual ones. Due to this reason the auto-regressive approach for the residuals only, can be inappropriate to account for interplant or interplot competition.

Draper and Guttman (1980) included a special case of (5) in which $\phi_i = \lambda\tau_i$, where λ is a coefficient of interference, common to all genotypes. The model is:

$$\begin{aligned} y &= Xb + H\tau + \varepsilon \\ &= Xb + Z\tau + NZ\lambda\tau + \varepsilon \end{aligned} \qquad (6)$$

where $H = (I + \lambda N)Z$, so that the model is non linear in τ and λ. The treatment effect for pure stand planting is given by $\tau_i^* = (1 + v\lambda)\tau_i$.

The component ϕ_i in (5) can be positive or negative depending on aggressiveness of the treatment. If negative (for aggressive varieties), the absolute value of ϕ_i should be subtracted from τ_i through $\tau_i^* = \tau_i + v\phi_i$ giving the treatment effect for pure stand planting, where v is the number of neighbours considered. If positive (sensitive variety), ϕ_i will be summed in the expression for τ_i^*. The neighbouring effect is not always correlated (negatively) to the trait being evaluated as it can depend on other traits such height and vigour of the

plants. In cases in which ϕ_i is unrelated to τ_i , the models (1), (2), (3), (4) and (6) are inadequate as they consider an unique competition coefficient for all genotypes. So, the model (5) tends to be better as it permits the neighbour genotypic effects to be individually specified. Also, a ranking based on the component ϕ_i can be performed aiming at the selection of low-competition and high-production varieties for high-density planting.

3.3 Joint Modelling of Competition Effects and Fertility Trends

Pithuncharurnlap, Basford and Federer (1993) attempted to include both trend and competition in a spatial model. They put together the uni-dimensional autoregressive model of Gleeson and Cullis (1987) for modelling fertility trend in one dimension and the genotypic competition model (5) of Besag and Kempton (1986) for modelling interference. The model is of the form:

$$y = Xb + Z\tau + NZ\phi + \xi + \eta \qquad (7)$$

where:
ξ : random vector of correlated errors;
η : random vector of non-correlated errors.

The competition was modelled as part of the treatment structure and the trend in only one dimension was modelled as part of the structure of errors.

Durban, Currie and Kempton (2001) commented about the problem of simultaneously modelling of two types of local correlation. According to them, some difficult might be anticipated in the joint modelling of trend and competition by the phenotypic model, since both are correlation effects. They suggested different mechanisms to specify the two effects, allowing them to be separately estimated.

Durban, Hackett and Currie (1999) and Durban, Currie and Kempton (2001) considered simultaneously the two effects by modelling trend through cubic smoothing splines within blocks and interference by the phenotypic model of Kempton (1982). However, splines might not be the best option for modelling spatial trend. Very often, two-dimension separable auto-regressive models provide a better fit (Gilmour, Cullis and Verbyla, 1997).

Stringer and Cullis (2002b) used the same model as (7), but assumed τ_i and ϕ_i as random effects (model 8). In this case, there is a covariance between τ_i and ϕ_i. The covariance matrix between them is:

$G = \begin{pmatrix} g_{\tau\tau} & g_{\tau\phi} \\ g_{\tau\phi} & g_{\phi\phi} \end{pmatrix}$, where $g_{\tau\tau}$ is the variance component for the direct genotypic effects, $g_{\phi\phi}$ is the variance component for the on neighbour genotypic effects and $g_{\tau\phi}$ is the covariance component between the direct and on neighbour genotypic effects.

According to model (6) of Draper and Guttman (1980), the variance-covariance matrix G is given by:

$$G = \begin{pmatrix} g_{\tau\tau} & \lambda_1 g_{\tau\tau} \\ \lambda_1 g_{\tau\tau} & \lambda_1^2 g_{\tau\tau} \end{pmatrix}.$$

This variance-covariance matrix is of reduced rank (rank = 1). Thompson et al. (2003) describe how to deal with models of this sort.

Stringer and Cullis (2002b) advocated a sequential approach, commencing by modelling trend, then checking the variograms and auto-correlations, and finally undertaking the modelling of the competition.

3.4. Competition Models in Perennial Crops and Forest Trees

The competition models applied in perennial plants and forest trees have been the same (with small modifications) as applied in annual crops, which were described in the previous topics. Correll and Anderson (1983) applied the competition model of Draper and Guttman (1980) together (but not simultaneously) with the spatial analysis of Papadakis to account for interference and trend, respectively. Magnussen and Yeatman (1987) used two approaches: the competition index of Hegyi (1974) as covariate and a modification of the competition model of Kempton (1982). The competition index of Hegyi (1974) was proposed in the context of competitive pressure on single individuals in natural stands and is given by:

$C_i = (\sum_{j=1}^{8} Y_j / Y_i) / Dist_{ij}$, where:

C_i: Competition index of the subject tree or plant;
Y_i: observed value of the subject tree i;
Y_j: observed value of the competitor tree j;
$Dist_{ij}$: distance between tree i and j.

The use of this index as a covariate produce results similar to that obtained with the method of Kempton (1982) when applied to the average of the 8 neighbours assumed equally spaced in relation to the subject tree. So, the advantage of the Hegyi's index refers only to the consideration of the different distances between the subject tree and the neighbours. Leonardecz-Neto (2002) also applied this index in forest trees.

The modification on Kempton (1982), introduced by Magnussen and Yeatman (1987) was the consideration of two competition coefficients β, one for individuals of different treatments and other for individuals of the same treatment in a plot, i.e., one competition coefficient for related individuals and other for unrelated individuals. This sort of model is a first-order auto-normal scheme of a two-dimensional Markov process (Besag, 1974) if the experimental design is regarded as a regular lattice of point sites with continuous variables having a multivariate normal distribution, and assuming stability in both time and space.

Magnussen (1994) considered the simultaneous adjustment for spatial and competition effects by using modifications of the approach used by Correll and Anderson (1983), based on the Papadakis method. Kusnandar (2001) extended the model of Kempton (1982) to two dimensions, considering competition in the row and column directions, under a mixed effects model. Montagnon et al. (2001) reported the first paper dealing with competition in coffee. They used specific competition coefficients for each treatment but only at residual level. They used the multiple linear regression technique to estimate the competition or partner effects (Gallais, 1975).

The competition models used in forest trees did not consider specific competition coefficients for each treatment and the partitioning of treatment effect into direct and on neighbour effect. Also, for joint modelling of competition and trend, the spatial approach used was the Papadakis method, which can not be the best one. Besides, when using the phenotypic model of interference and the Hegyi index, in which the covariate is defined to be the same as the main trait (for example, both being height of the plants), the least square approach produces an invalid estimate of β . An efficient estimation can be performed using profile maximum likelihood, which were not used by the authors mentioned before. So, in the next topic we propose new modelling for trend and competition effects in perennial plants.

3.5 Proposed Competition and Spatial Models for Perennial Plants

3.5.1 Competition and Spatial Model for Single Tree Plot Design (Four Neighbours)

The model (8) of Stringer and Cullis (2002b) can be used to account for any number of neighbours. In this case, the neighbour plants or plots belong to different treatment (variety), and the spacing between the subject (S) plant and the neighbours (N) are the same.

Layout

N

N S N

N

The competition effect on four neighbours can be specified individually (when the neighbour effect depends on shading) as east, west, north and south neighbours or just in one coefficient encompassing all the horizontally and vertically neighbours in ϕ_{HV} . Such model is detailed below.

Model

$$y = Xb + Z\tau + N_{HV} Z\phi_{HV} + \xi + \eta \qquad (8)$$

τ : random vector of genotype effects in the absence of competition or under the average competitive stress in the trial;

N_{HV} : incidence matrix for horizontally and vertically neighbours;

ϕ_{HV} : random vector of genotype effects on horizontally and vertically neighbours.

Neighbour Incidence Matrix (N_{HV})

Field Array

1 5 9
2 6 10
3 7 11
4 8 12

1 2 3 4 5 6 7 8 9 10 11 12

0 1 0 0 1 0 0 0 0 0 0 0
0 1 0 0 1 0 0 0 0 0 0
0 1 0 0 1 0 0 0 0 0
0 0 0 0 1 0 0 0 0
0 1 0 0 1 0 0 0
0 1 0 0 1 0 0
0 1 0 0 1 0
0 0 0 0 1
0 1 0 0
0 1 0
0 1
0

Covariance Matrix of Direct and on Neighbour Treatment Effects

$$G = \begin{pmatrix} g_{\tau\tau} & g_{\tau\phi_{HV}} \\ g_{\tau\phi_{HV}} & g_{\phi_{HV}\phi_{HV}} \end{pmatrix}$$

Corrected Treatment Effect

$$\tau_i^* = \tau_i + v_{HVi}\phi_{HV_i}$$

v_{HVi} : number of horizontally plus vertically neighbours of the genotype i in the trial, i.e., 4.

3.5.2 Competition Model for Single Tree Plot Design (Eight Neighbours)

Aiming to take into account the different distances between the neighbours and the subject tree, the model (8) should be extended to (9).

Layout

N	N	N
N	S	N
N	N	N

Model

$$y = Xb + Z\tau + N_{HV}Z\phi_{HV} + N_D Z\phi_D + \xi + \eta \qquad (9)$$

τ : random vector of genotype effects in the absence of competition or under the average competitive stress in the trial;
N_{HV} : incidence matrix for horizontally and vertically neighbours;
ϕ_{HV} : random vector of genotype effects on horizontally and vertically neighbours;
N_D : incidence matrix for diagonally neighbours;
ϕ_D : random vector of genotype effects on diagonally neighbours.

Neighbour Incidence Matrix (N_D)

Field Array

1	5	9
2	6	10
3	7	11
4	8	12

1 2 3 4 5 6 7 8 9 10 11 12

0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 1 0 1 0 0 0 0 0
0 0 0 1 0 1 0 0 0 0
0 0 0 1 0 0 0 0 0
0 0 0 0 0 1 0 0
0 0 0 1 0 1 0
0 0 0 1 0 1
0 0 0 1 0
0 0 0 0
0 0 0
0 0
0

Covariance Matrix of Direct and on Neighbour Treatment Effects

$$G = \begin{pmatrix} g_{\tau\tau} & g_{\tau\phi_{HV}} & g_{\tau\phi_D} \\ & g_{\phi_{HV}\phi_{HV}} & g_{\phi_{HV}\phi_D} \\ & & g_{\phi_D\phi_D} \end{pmatrix}$$

Corrected Treatment Effect

$$\tau_i^* = \tau_i + \nu_{HVi}\phi_{HV_i} + \nu_{Di}\phi_{D_i}$$

ν_{HVi} : number of horizontally plus vertically neighbours of the genotype i in the trial, i.e., 4.

ν_{Di} : number of diagonally neighbours of the genotype i in the trial, i.e., 4.

3.5.3 Competition and Spatial Model for Multiple Tree Plot Design (Four Neighbours)

In this case, the neighbour plants belong to different treatment (variety) in one dimension (horizontally, in general) and belong to the same variety in the other direction (vertically, usually). Also, the spaces between the subject (S) plant and the neighbours (N) are the same. We should change model (8) to (10).

Layout

N_S

N_O S N_O

N_S

Model

$$y = Xb + Z\tau + N_H Z\phi_H + N_V Z\phi_V + \xi + \eta \qquad (10)$$

τ : random vector of genotype effects in the absence of competition or under the average competitive stress in the trial;
N_H : incidence matrix for horizontally neighbours;
ϕ_H : random vector of genotype effects on horizontally neighbours;
N_V : incidence matrix for vertically neighbours;
ϕ_V : random vector of genotype effects on vertically neighbours.

Neighbour Incidence Matrix (N_H)

Field Array

1 5 9
2 6 10
3 7 11
4 8 12

1 2 3 4 5 6 7 8 9 10 11 12

0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 1 0 0 0 0 0
0 0 0 0 1 0 0 0 0
0 0 0 0 1 0 0 0
0 0 0 0 1 0 0
0 0 0 0 1 0
0 0 0 0 1
0 0 0 0
0 0 0
0 0
0

Covariance Matrix of Direct and on Neighbour Treatment Effects

$$G = \begin{pmatrix} g_{\tau\tau} & g_{\tau\phi_H} & g_{\tau\phi_V} \\ & g_{\phi_H\phi_H} & g_{\phi_H\phi_V} \\ & & g_{\phi_V\phi_V} \end{pmatrix}$$

Corrected Treatment Effect

$$\tau_i^* = \tau_i + v_{Hi}\phi_{H_i} + v_{Vi}\phi_{V_i}$$

v_{Hi} : number of horizontally neighbours of the genotype i in the trial, i.e., 2.
v_{Vi} : number of vertically neighbours of the genotype i in the trial, i.e., 2.

In this case, the neighbour effect in the own variety is estimated by ϕ_{V_i} .

3.5.4 Competition and Spatial Model for Multiple Tree Plot Design (Eight Neighbours)

Aiming to take into account the different distances between the neighbours and the subject tree, the model (10) should be extended to (11).

Layout

N	N_S	N
N_O	S	N_O
N	N_S	N

Model

$$y = Xb + Z\tau + N_H Z\phi_H + N_V Z\phi_V + N_D Z\phi_D + \xi + \eta \qquad (11)$$

τ : random vector of genotype effects in the absence of competition or under the average competitive stress in the trial;
N_H : incidence matrix for horizontally neighbours;
ϕ_H : random vector of genotype effects on horizontally neighbours;
N_V : incidence matrix for vertically neighbours;
ϕ_V : random vector of genotype effects on vertically neighbours.

N_D : incidence matrix for diagonally neighbours;
ϕ_D : random vector of genotype effects on diagonally neighbours.

Covariance Matrix of Direct and on Neighbour Treatment Effects

$$G = \begin{pmatrix} g_{\tau\tau} & g_{\tau\phi_H} & g_{\tau\phi_V} & g_{\tau\phi_D} \\ & g_{\phi_H\phi_H} & g_{\phi_H\phi_V} & g_{\phi_H\phi_D} \\ & & g_{\phi_V\phi_V} & g_{\phi_V\phi_D} \\ & & & g_{\phi_D\phi_D} \end{pmatrix}$$

Corrected Treatment Effect

$$\tau_i^* = \tau_i + v_{Hi}\phi_{H_i} + v_{Vi}\phi_{V_i} + v_{Di}\phi_{D_i}$$

v_{Hi} : number of horizontally neighbours of the genotype i in the trial, i.e., 2..
v_{Vi} : number of vertically neighbours of the genotype i in the trial, i.e., 2.
v_{Di} : number of diagonally neighbours of the genotype i in the trial, i.e., 4.

In this case, the neighbour effect in the own variety is estimated by ϕ_{V_i} .

3.5.5 Generalised Competition and Spatial Model

A more generalised model suitable for any experimental layout (one or several plants per plot) and number of neighbours is given by (12):

Layout

N	N	N
N	S	N
N	N	N

Model

$$y = Xb + Z\tau + N_E Z\phi_E + N_W Z\phi_W + N_N Z\phi_N + N_S Z\phi_S + N_{NE} Z\phi_{NE} + N_{SE} Z\phi_{SE} + N_{NW} Z\phi_{NW} + N_{SW} Z\phi_{SW} + \xi + \eta \qquad (12)$$

τ : random vector of genotype effects in the absence of competition or under the average competitive stress in the trial;
N_E : incidence matrix for eastern neighbours;
ϕ_E : random vector of genotype effects on eastern neighbours;
N_W : incidence matrix for western neighbours;
ϕ_W : random vector of genotype effects on western neighbours.
N_N : incidence matrix for northern neighbours;
ϕ_N : random vector of genotype effects on northern neighbours.
N_S : incidence matrix for southern neighbours;
ϕ_S : random vector of genotype effects on southern neighbours;
N_{NE} : incidence matrix for north-eastern neighbours;
ϕ_{NE} : random vector of genotype effects on north-eastern neighbours.
N_{SE} : incidence matrix for south-eastern neighbours;
ϕ_{SE} : random vector of genotype effects on south-eastern neighbours.
N_{NW} : incidence matrix for north-western neighbours;
ϕ_{NW} : random vector of genotype effects on north-western neighbours.
N_{SW} : incidence matrix for south-western neighbours;
ϕ_{SW} : random vector of genotype effects on south-western neighbours.

This full model and nested models within it can be used to infer about the significance of specific neighbour positions. The final model kept must allow for the covariance between the random effects remained.

Corrected Treatment Effect for the Full Model

$$\tau_i^* = \tau_i + \phi_E + \phi_W + \phi_N + \phi_S + \phi_{NE} + \phi_{SE} + \phi_{NW} + \phi_{SW}$$

This generalised model demands large data sets to be fitted, as many degrees of freedom are necessary to fit all the effects. Trials with great number of treatment and limited number of replications are not suitable for the application of this model and perhaps neither the model (8) to (11). In such case, the alternative phenotypic approach of Kempton (1982) together with the spatial analysis of Cullis and Gleeson (1991) and Gilmour, Cullis and Verbyla (1997) should be used.

3.5.6 Phenotypic Competition and Spatial Model

Apparently, the phenotypic competition approach of Kempton (1982) together with the spatial analysis of Cullis and Gleeson (1991) and Gilmour, Cullis and Verbyla (1997) was not used simultaneously. Such simultaneous modelling can be specified according to the following model.

$$y = Xb + Z\tau + \beta Wy + \xi + \eta \qquad (13)$$

where:

W: is a n x n weight or regressor matrix which, in conjunction with y, provides the average value of the neighbours as a covariate. In general, the mean of the two or of the four neighbours can be used. Diagonally neighbours are expected to have non significant effects because of the greater distance from the subject tree and the positive effects produced on growth of the other closer neighbours of the subject tree.

Estimation and prediction concerning this model demands the use of the profile likelihood which is detailed in the item 6.

3.5.7 Missing Plant Effects

For inference about treatments or varieties, the effects of missing plants is considered by omitting (or fitting as fixed effects) the zeros correspondent to missing plots for the purpose of predicting the direct effects and through the consideration of the zeros for the purpose of predicting the on neighbours effects, according to the models (8) to (12). This can be achieved by coding all the zeros neighbour values as belonging to a variety not yet coded in the treatment column, i.e. coding them as a new variety. The effect of missing plants will be reflected on ϕ and then on τ_i^*. However, for individual selection of trees, a further correction of the observed value of a tree can be necessary, when an individual model is not used, i.e., when using a reduced animal model.

For individual selection it is common to use the reduced individual model for predicting the individual genetic value (a). By this approach, the prediction equation is $\hat{a} = Z\hat{\tau} + Z\hat{\phi} + h_d^2(y - X\hat{b} - Z\hat{\tau} - Z\hat{\phi})$, where h_d^2 is the within

variety heritability. The observed values in y can be corrected through the use of growth traits of the neighbours as covariates. The same matrices N_{HV}, N_H, N_V and N_D can be used to obtain the numerical values of the covariates. One (model 9), two (models 10 and 11) or three regression coefficients may be necessary according to the model. Generically, the covariate values can be obtained by $X = (N'N)^{-1}N'y$ and the corrected y values to enter in the expression for $\hat{a}$ are given by $y_c = y - \beta(X - \overline{X})$, where β is a competition coefficient at phenotypic level.

In model (13), the zeros as neighbour values are considered in the computation process of the average of neighbours as a covariate.

3.6 Profile Likelihood and Generalisation of REML (GREML)

It is not possible to use ordinary REML for the phenotypic competition model of Kempton (1982), as the competition coefficient appears in both the mean and variance of y. However, a generalisation of REML can be applied for estimating the parameters of the model. The generalisation (GREML) involves adjusting profile likelihood (through the adjusted profile score) for the parameter of interest in a general class of models. Such adjustment can be done by using the method of McCullagh and Tibshirani (1990), which remove bias from maximum likelihood estimates.

The inference in the presence of nuisance parameters is a difficult problem in statistics. From the likelihood perspective, the simplest approach is to maximise out the nuisance parameters for fixed values of the parameters of interest and to construct the so-called profile likelihood. In other words, such solution refers to replace the nuisance parameters in the likelihood function with their maximum likelihood estimates for fixed values of the parameters of interest. This gives the profile likelihood. The profile likelihood is then treated as an ordinary likelihood function for estimation and inference about the parameters of interest. Unfortunately, with large numbers of nuisance parameters, this procedure can produce inefficient or even inconsistent estimates. The inherent problems in the use of profile likelihoods are biased parameters estimates and optimistic estimates of standard errors.

Modifications to the profile likelihood with an aim to alleviate these problems were proposed. Barndorff-Nielsen (1983, 1986) proposed the modified profile likelihood, which is closely related to conditional profile likelihood proposed by Cox and Reid (1987) in which they suggested a likelihood ratio test constructed from the conditional distribution of the observations given maximum likelihood estimates for the nuisance parameters. McCullagh and Tibshirani (1990) proposed a simpler alternative approach named adjusted profile likelihood. Their method depends on the observation that the score function computed from the full log-likelihood function has (i) zero expectation and (ii) variance equal to the negative of the expected derivative matrix. A score function that has property (i) is said to be unbiased, while if has property (ii) is said to be information unbiased. By association, it can be said that a likelihood function is unbiased/information unbiased if its score function is unbiased/information unbiased. In contrast to the score function computed from the full log-likelihood, the score function computed from the profile log-likelihood is, in general, neither unbiased nor information unbiased. McCullagh and Tibishirani's idea is that the profile log-likelihood score function be centred and scaled so that it too is unbiased and information unbiased (Durban and Currie, 2000).

McCullagh and Tibshirani (1990) concentrated on giving asymptotic formulae for their corrections in a very general setting. Durban and Currie (2000) gave exact expressions for the adjustments for a general non-linear normal regression model. In its more general form, the model allows both the mean and the variance of y to depend on the parameter of interest. An example of this general form is a regression model with autoregressive terms such as the phenotypic model of competition. The exact adjustment for the profile likelihood for such model improves the estimation of the variance and competition parameters.

According to the phenotypic competition model, $y = Xb + Z\tau + \beta Wy + \varepsilon$, we can write:

$Dy = Xb + Z\tau + \varepsilon$, where $D = I - \beta W$.

$y = D^{-1}Xb + D^{-1}Z\tau + D^{-1}\varepsilon$, where:

$y \sim N(D^{-1}Xb, \sigma^2 D^{-1}VD^{-1\prime})$.
$\tau \sim N(0, \sigma^2 G)$.
$\varepsilon \sim N(0, \sigma^2 R)$.
$V = ZGZ + R$.

Considering *b* as a nuisance parameter and $\theta = (\beta, \tau, \sigma^2)$ as parameters of interest, the log-likelihood of $Dy \sim N(Xb, \sigma^2 V)$ is given by

$$\ell = \ell(\theta, b; y) = -(n/2)\log 2\pi - (n/2)\log \sigma^2 - (1/2)\log \left| D^{-1} V D^{-1'} \right| - (1/2\sigma^2)(Dy - Xb)' V^{-1} (Dy - Xb)$$

Taking the derivative of this log-likelihood with respect to b and equating to zero gives the maximum likelihood estimate of *b* which is given by

$$\hat{b} = (X' V^{-1} X)^{-1} X' V^{-1} Dy$$

According to McCullagh and Tibshirani (1990), the profile log-likelihood is obtained by replacing the nuisance parameters by their maximum likelihood estimates. Substituting $\hat{b} = (X'V^{-1}X)^{-1}X'V^{-1}Dy$ into $\ell = \ell(\theta, b; y)$ gives the profile log-likelihood ℓ_p, which ignoring constants is equivalent to

$$\ell_p(\theta; y) = -(n/2)\log \sigma^2 - (1/2)\log \left| D^{-1} V D^{-1'} \right| - (1/2\sigma^2) y' D' PVPDy$$
$$= \log |D| - (n/2)\log \sigma^2 - (1/2)\log |V| - (1/2\sigma^2) y' D' PDy$$

where $P = V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1}$.

From this profile log-likelihood, adjusted profile score equations can be obtained. The adjusted profile score equations for the variance parameters are equivalent to the REML score equations based on *Dy*.

The residual log-likelihood based on *Dy* is given by

$$\ell_{Re} = -[(n-p)/2]\log \sigma^2 + \log |D| - (1/2)\log |V| - (1/2)\log \left| X' V^{-1} X \right| - (1/2\sigma^2) y' D' PDy$$

A key difference between this and the residual log-likelihood on y is the additional term $\log|D|$. So, REML on Dy can be obtained by using the standard algorithms used in ASREML and GENSTAT, but $\log|D|$ should be also obtained and added in the log L. This can be done in an easier way using GENSTAT.

The presence of the competition coefficient (parameter of interest) in both the mean and variance of y leads to difficulties. However, the McCullagh and Tibishirani's adjustments apply well in this situation and the resulting adjusted profile likelihood equals the residual maximum likelihood (REML) of Patterson and Thompson (1971). The adjusted profile score equations are equivalent to the REML score equations based on the adjusted response Dy . In practice, D is replaced by its estimate. The adjusted score produces both a REML type adjustment to the estimates of variance components and an adjustment to the estimate of β, removing its bias. Profile log likelihoods and adjusted profile scores for the parameters of interest are presented by Durban and Currie (2000) for the fixed model case.

In the context of the model (13), the competition parameter and variance components were estimated as follows: (i) obtaining of REML on Dy for several given values of β ; (ii) obtaining of (Log|D|) for given values of β; (iii) obtaining of the profile likelihood (LogL + Log|D|) for a range of β.

3.7 Estimation/Prediction Procedures and Softwares

Variance components associated to several models were estimated through the REML procedure (Patterson and Thompson, 1971; Searle et al., 1992; Thompson, 1973, 1977, 1980, 2002; Thompson and Welham, 2003; Cullis et al. 2004). Random effects were predicted by the BLUP procedure (Henderson, 1973; Thompson, 1979).

All models were fitted using the software ASREML (Gilmour and Thompson, 1998, 2002; Gilmour, Cullis, Thompson and Welham, 2002; Gilmour et al. 2002) which uses the REML procedure through the average information algorithm and sparse matrix techniques (Gilmour, Thompson and Cullis, 1995; Johnson and Thompson, 1995; Thompson, Wray and Crump, 1994; Thompson et al., 2003). The software GENSTAT (Thompson and Welham, 2003) was also used.

3.8 Applications to Experimental Data

3.8.1 General Results from Five Different Species

Five data sets concerning to different crops were used: circumference of the trunk in two years old *Eucalyptus spp* trees, evaluated in a lattice design with single tree plots, 240 treatments (clones) and 40 replications; leaf weight in a harvest of tea plants evaluated in a complete block design with 141 treatments (half sib families), ten replications and six plants per plot; number of stems in sugarcane evaluated in a complete block design with 128 treatments (clones) and two replications; diameter of the trunk evaluated in 13 years old *Pinus caribaea* var. *bahamensis* trees evaluated in a lattice design with 121 treatments (half sib families), six replications and six plants per plot; circumference of the trunk in 18 years old *Eucalyptus maculata* trees evaluated in a complete block design with 25 treatments (half sib families) and 36 replications in single tree plots.

Results concerning to the basic traditional and spatial analysis (AR1 x AR1) are presented in Table 1.

Table 1. Residual log-likelihoods (Log L) and estimates of the genetic variance among treatments ($\hat{\sigma}^2_\tau$), residual variance ($\hat{\sigma}^2$), adjusted heritability ($\hat{h}^2_{adj}$) proportional only to the unaccounted error, proportional error variances associated to spatial and non-spatial analysis ($\hat{\sigma}^2_s / \hat{\sigma}^2_{ns}$) and auto-correlation coefficients associated to columns (ARc) and rows (ARr). Data sets referring to five different species.

Data Set	Log L	$\hat{\sigma}^2_\tau$	$\hat{\sigma}^2$	$\hat{h}^2_{adj}$	$\hat{\sigma}^2_s / \hat{\sigma}^2_{ns}$	ARc	ARr
E. spp - Traditional	-19487.4	15.414	24.827	0.383	-	-	-
E. spp - Spatial	-19407.2	15.514	24.607	0.387	0.991	0.81*	0.99*
Tea Plant-Traditional	1552.89	0.0110	0.2214	0.1905	-	-	-
Tea Plant-Spatial	2127.24	0.0134	0.1492	0.3296	0.674	0.79*	0.75*
Sugarcane-Traditional	-1023.77	685.98	228.73	0.7499	-	-	-
Sugarcane-Spatial	-1023.05	684.17	229.26	0.7490	1.000	-0.12^{ns}	0.035^{ns}
Pinus-Traditional	-6584.67	1.0403	18.255	0.2156	-	-	-
Pinus-Spatial	-6559.19	0.9621	17.891	0.2040	0.980	-0.10*	-0.13*
E. maculata-Traditional	-2940.18	37.25	601.44	0.2333	-	-	-
E. maculata-Spatial	-2935.12	35.80	596.61	0.2264	0.991	0.10*	0.10*

These five data sets provided the three practical situations that can occur in field experiments: (i) absence of spatial trends within the levels of local control and absence of competition effects (Eucalyptus spp data set); (ii) presence of spatial trends within the levels of local control and absence of competition effects (tea plant data set); (iii) presence of competition effects (sugarcane, Pinus and *E. maculata* data sets).

The spatial trend is likely to occur in any field trial. However, sometimes it can be taken into account by the control local associated with more elaborated experimental designs such as lattices and row/column. This happened here for the *Eucalyptus spp* trial experiment established in lattice design, which did not show benefits from the spatial analysis. This can be seen from the high relation (0.99) between the two error variances concerning to spatial (including the independent error term) and non-spatial analysis and the high auto-correlation coefficients. These very high auto-correlation coefficients reveal that the auto-regressive process is modelling global trend and that there is no competition effects acting. The global trend is being taken into account by the blocks in the traditional lattice analysis and this is confirmed by the zero value for the block variance in spatial analysis and by its significant value in the lattice analysis. These results can be seen from Table 2, which compares three models of analysis for the *Eucalyptus spp* data set. The absence of competition was expected as the trees were only two years old which is an age not suitable for competition in forest trees species.

Table 2. Residual log-likelihoods (Log L) and estimates of the genetic variance among treatments ($\hat{\sigma}^2_\tau$), residual variance ($\hat{\sigma}^2$), correlated error variance ($\hat{\sigma}^2_\xi$), block variance ($\hat{\sigma}^2_b$), proportional error variances associated to spatial and non-spatial analysis ($\hat{\sigma}^2_s/\hat{\sigma}^2_{ns}$) and auto-correlation coefficients associated to columns (ARc) and rows (ARr). Data set concerning to Eucalyptus.

Model of Analysis	Log L	$\hat{\sigma}^2_\tau$	$\hat{\sigma}^2$	$\hat{\sigma}^2_\xi$	$\hat{\sigma}^2_b$	$\hat{\sigma}^2_s/\hat{\sigma}^2_{ns}$	ARc	ARr
Eucalyptus-Traditional	-19487.4	15.414	24.827	-	1.576	1.000	-	-
Eucalyptus-Spatial + η	-19407.2	15.514	24.607	6.686	0.023	0.991	0.81*	0.99*
Eucalyptus-Spatial	-19484.7	15.316	-	24.732	1.687	0.996	0.00^{ns}	-0.03^{ns}

It can be seen that when the correlated (spatial) plus the independent errors were fitted, the spatial term modelled only the global tendency of blocks, which turned zero in $\hat{\sigma}^2_b$. When the model without the independent error was fitted,

the auto-correlation coefficients revealed no correlated error and the global tendency was modelled by the blocks of the lattice design (Table 2).

The tea plant trial experiment established in complete block design showed significant spatial trend and benefits from the spatial analysis. This can be seen from the low relation (0.67) between the two error variances concerning to spatial and non-spatial analysis and the high auto-correlation coefficients. These high auto-correlation coefficients reveal that there are no competition effects acting, which is also expected as all leaves are collected every year in each plant justifying the absence of aboveground competition. The reduced error variance and higher adjusted heritability reveal the presence of spatial trend within blocks and the benefits of the spatial analysis. In general, it can be seen from Table 1 that when the auto-correlation coefficients tends to 1 or 0, spatial analysis tends to give no practical ($\hat{\sigma}^2_s / \hat{\sigma}^2_{ns}$) improvement in the fit, despite significant changes in Log L.

The results concerning sugarcane, Pinus and *E. maculata* showed the presence of competition, which is expected in sugarcane (Stringer and Cullis, 2002a and b) and in older trees. So such cases demanded extended models of analysis. For these data set the independent error in spatial models were non significant. This is in accordance with Gilmour et al. (1997) who reported that when the autoregressive parameters are near 0, it is often impossible or very difficult to estimate $\hat{\sigma}^2_\eta$.

3.8.2 Phenotypic Competition Models via Profile Likelihood in Sugarcane

Competition models for sugarcane (only two replications) could only be applied through the phenotypic model of interference as the degrees of freedom were not sufficient to fit all the effects needed in the genotypic models. Results concerning to several models in sugarcane are presented in Tables 3, 4 and 5.

Table 3. Residual log-likelihoods (Log L) on Dy, determinant component (Log|D|), sum of the two components (LogL + Log|D|) giving the profile likelihood, and auto-correlation coefficients associated to columns (ARc) and rows (ARr), for different values of the competition coefficient (β) in the sugarcane data set. A model with both correlated error (spatial) and phenotypic competition effect was used.

β Values	Log L	Log|D|	LogL + Log|D|	ARc	ARr
-0.80	-1017.61	-31.07	-1048.68	0.43**	0.51**
-0.75	-1017.06	-26.21	-1043.27	0.42**	0.49**
-0.70	-1016.71	-22.05	-1038.76	0.40**	0.48**
-0.65	-1016.56	-18.46	-1035.02	0.38**	0.46**
-0.60	-1016.59	-15.33	-1031.92	0.36**	0.43**
-0.55	-1016.80	-12.60	-1029.40	0.34**	0.41**
-0.50	-1017.15	-10.22	-1027.37	0.31**	0.38**
-0.45	-1017.64	-8.14	-1025.78	0.29**	0.35**
-0.40	-1018.24	-6.34	-1024.58	0.25**	0.32**
-0.35	-1018.91	-4.79	-1023.70	0.22**	0.29**
-0.30	-1019.62	-3.49	-1023.11	0.18**	0.26**
-0.25	-1020.35	-2.40	-1022.75	0.14 ns	0.23**
-0.20	-1021.06	-1.53	-1022.59	0.10 ns	0.19*
-0.15	**-1021.72**	**-0.85**	**-1022.57**	**0.05 ns**	**0.15 ns**
-0.10	-1022.29	-0.38	-1022.67	0.00 ns	0.12 ns
-0.05	-1022.74	-0.09	-1022.83	-0.06 ns	0.07 ns
0.00	-1023.05	0.00	-1023.05	-0.12 ns	0.03 ns
0.05	-1023.21	-0.09	-1023.30	-0.18**	0.00 ns
0.10	-1023.21	-0.38	-1023.59	-0.23**	0.05 ns
0.20	-1022.83	-1.53	-1024.36	-0.33**	-0.14 ns
0.40	-1021.98	-6.34	-1028.32	-0.48**	-0.29**

Table 3 presents the profile likelihood for a range of competition coefficients (β) in a model with both correlated error (spatial) and phenotypic competition effects. It can be seen that the maximisation of the likelihood function occurred for β = -0.15, with a LogL + Log|D| = - 1022.57. The associated residual autocorrelation coefficients were not significant showing that the phenotypic competition coefficient encompassed the whole correlation pattern, including the genetic competition effect and a balance between residual competition effects and environmental trend within blocks.

Table 4 presents the profile likelihood for a range of competition coefficients (β) in a model with only phenotypic competition effects. It can be seen that the maximisation of the likelihood function occurred for β = -0.05, with a LogL + Log|D| = - 1023.28. The two LogL + Log|D| values mentioned are close to each other and the results confirm that a model with only phenotypic competition effects is enough. It can be asserted also that competition effects of small magnitude are present in the trial.

Table 4. Residual log-likelihoods (Log L) on Dy, determinant component (Log|D|), sum of the two components (LogL + Log|D|) giving the profile likelihood, and auto-correlation coefficients associated to columns (ARc) and rows (ARr), for different values of the competition coefficient (β) in the sugarcane data set. A model with only competition effect was used.

β Values	Log L	Log|D|	LogL + Log|D|
-0.40	-1030.34	-6.34	-1036.68
-0.30	-1026.46	-3.49	-1029.95
-0.20	-1023.95	-1.53	-1025.48
-0.15	**-1023.28**	**-0.85**	**-1024.13**
-0.10	-1023.03	-0.38	-1023.41
-0.05	**-1023.19**	**-0.09**	**-1023.28**
0.00	-1023.77	0.00	-1023.77

Table 5 presents comparative results from a series of models applied to the sugarcane data set.

Table 5. Residual log-likelihoods (Log L) and estimates of the genetic variance among treatments ($\hat{\sigma}^2_\tau$), residual variance ($\hat{\sigma}^2$), heritability ($\hat{h}^2_{mh}$), competition coefficient ($\hat{\beta}$) and auto-correlation coefficients associated to columns (ARc) and rows (ARr). Sugarcane data set.

Model	Log L	$\hat{\sigma}^2_\tau$	$\hat{\sigma}^2$	$\hat{h}^2_{mh}$	$\hat{\beta}$	ARc	ARr
Traditional	-1023.77	685.98	228.73	0.7499	-	-	-
Spatial	-1023.05	684.17	229.26	0.7490	-	-0.12^{ns}	0.035^{ns}
Competition (Profile)	-1023.28	667.48	231.43	0.7425	-0.05	-	-
Competition + Spatial (Profile)	-1022.57	660.02	232.96	0.7391	-0.15	0.05^{ns}	0.15^{ns}
Competition (Covariate)	-1025.55	674.03	230.56	0.7451	-0.10*	-	-
Competition + Spatial (Covariate)	-1018.97	466.27	347.69	0.5728	-0.64*	0.38*	0.45*

The traditional, spatial (autoregressive in two dimensions), competition (using profile likelihood) and competition + spatial (using profile likelihood) models gave basically the same results in terms of the residual log-likelihoods, residual variance and heritability. This is due to the small magnitudes of the competition effects. The competition model (3) taking the average of the four neighbours (horizontally and vertically) as a covariate (fixed effect) and treatment effects as random, confirmed the presence of the competition effects ($\hat{\beta}$ = -0.10). Also it gave the same heritability as the traditional and the spatial models. However, the ordinary REML procedure applied here is not adequate because the competition coefficient appears in both the mean and variance of y and so can not be fitted as a covariate. The exact procedure of profile likelihood provides an exact adjustment and precise fitting for such model which improves the estimation of the variance and competition parameters. The competition coefficient estimate changed from -0.10 with ordinary REML to -0.05 with the profile REML.

For the competition + spatial model, the ordinary REML procedure using a covariate gave very different results concerning to Log L, residual variance and heritability. The competition coefficient and auto-correlation parameters estimates were considerable higher than that obtained with the profile REML. This difference reveals the importance of using the more accurate profile REML procedure. The competition + spatial model using a covariate (model (13)) gave a much higher competition coefficient (-0.64 against -0.15 of the profile REML) and the auto-correlation parameters were positive and high (0.38 and 0.45), i.e., they are modelling spatial trend. These results were obtained using positive starting values for the auto-correlation parameters. However, using negative starting values for such parameters, convergence with different results was obtained. The values at convergence were 0.40 for $\hat{\beta}$ and -0.47 and -0.29 for the auto-correlation parameters. Such estimates are non sense because opposite

signs are expected. These results revealed the inconsistence and inadequateness of the covariate approach for modelling competition and trend simultaneously. The profile likelihood approach should be always used instead.

The variety effects should be corrected by using the expression $\tau_c = \tau / (1 - \hat{\beta})$. In this case, the competition coefficient was –0.05 and so the variety effects should be divided by 1.05 or multiplied by 0.95. This is equivalent to multiply the heritability at clone mean level by 0.95.

3.8.3 Genotypic and Phenotypic Competition Models in Eucalyptus maculata

Phenotypic and genetic competition models for *E. maculata* data set are presented in Tables 6 to 8.

Table 6. Residual log-likelihoods (Log L) on Dy, determinant component (Log|D|), sum of the two components (LogL + Log|D|) giving the profile likelihood, and autocorrelation coefficients associated to columns (ARc) and rows (ARr), for different values of the competition coefficient (β) in the **Eucalyptus maculata** data set. Spatial and competition model was used.

β Values	Log L	Log|D|	LogL + Log|D|	ARc	ARr
-0.60	-2924.59	-47.53	-2972.120	0.25^{**}	0.25^{**}
-0.50	-2924.19	-31.71	-2955.900	0.19^{**}	0.20^{**}
-0.40	-2925.21	-19.69	-2944.900	0.13^{**}	0.14^{**}
-0.30	-2927.10	-10.83	-2937.930	0.07^{*}	0.08^{*}
-0.20	-2929.51	-4.742	-2934.250	0.01^{ns}	0.02^{ns}
-0.15	-2930.83	-2.654	-2933.484	-0.02^{ns}	-0.01^{ns}
-0.10	**-2932.20**	**-1.175**	**-2933.375**	$\mathbf{-0.05^{ns}}$	$\mathbf{-0.04^{ns}}$
-0.05	-2933.64	-0.293	-2933.933	-0.07^{*}	-0.07^{*}
0.00	-2935.12	0.000	-2935.120	-0.10^{**}	-0.10^{**}
0.10	-2938.30	-1.175	-2939.480	-0.15^{**}	-0.15^{**}

Table 6 presents the profile likelihood for a range of competition coefficients (β) in a model with both correlated error (spatial) and phenotypic competition effects. It can be seen that the maximisation of the likelihood function occurred for β = -0.10, with a LogL + Log|D| = -2933.38. The associated residual autocorrelation coefficients were not significant showing that the phenotypic competition coefficient encompassed the whole correlation pattern, including the

genetic competition effect and a balance between residual competition effects and environmental trend. These results are coincident or analogous with those obtained for the sugarcane data set.

Table 7 presents the profile likelihood for a range of competition coefficients (β) in a model with only phenotypic competition effects. It can be seen that the maximisation of the likelihood function occurred for β = -0.10, with a LogL + Log|D| = - 2934.30. The two LogL + Log|D| values mentioned are close to each other and the results confirm that a model with only phenotypic competition effects is enough. It can be asserted also that competition effects are present in the trial.

Table 7. Residual log-likelihoods (Log L) on Dy, determinant component (Log|D|), sum of the two components (LogL + Log|D|) giving the profile likelihood, and auto-correlation coefficients associated to columns (ARc) and rows (ARr), for different values of the competition coefficient (β) in the **Eucalyptus maculata** data set. Only competition model was used.

β Values	Log L	Log|D|	LogL + Log|D|
-0.40	-2933.55	-19.69	-2953.240
-0.30	-2929.77	-10.83	-2940.600
-0.20	-2929.63	-4.742	-2934.370
-0.10	**-2933.13**	**-1.175**	**-2934.305**
-0.05	-2936.22	-0.293	-2936.513
0.00	-2940.18	0.000	-2940.180
0.10	-2950.59	-1.175	-2951.770

Table 8 presents comparative results from a series of models applied to the *E. maculata* data set.

Table 8. Residual log-likelihoods (Log L) and estimates of the genetic variance among treatments ($\hat{\sigma}^2_\tau$), residual variance ($\hat{\sigma}^2$), heritability ($\hat{h}^2_{adj}$), competition coefficient ($\hat{\beta}$) and auto-correlation coefficients associated to columns (ARc) and rows (ARr). **Eucalyptus maculata** data set with several models.

Model	Log L	$\hat{\sigma}^2_\tau$	$\hat{\sigma}^2$	$\hat{h}^2_{adj}$	$\hat{\beta}$	ARc	ARr
(a) Traditional	-2940.18	37.25	601.44	0.233	-	-	-
(b) Spatial	-2935.12	35.80	596.61	0.226	-	-0.10**	-0.10**
(c) Competition (Profile)	-2934.30	36.28	590.72	0.231	-0.10	-	-
(d) Competition + Spatial (Profile)	-2933.38	35.83	588.00	0.229	-0.10	-0.05ns	-0.04 ns
(e) Competition (Covariate)	-2932.19	34.82	585.97	0.224	0.25**	-	-
(f) Spatial + Competition (Cov.)	-2927.19	30.34	644.52	0.180	0.52**	0.21**	0.21**
(g) Spatial + G Competition	-2935.12	35.80	596.61	0.226		-0.10**	-0.10**

The spatial model showed to be better than the traditional in terms of the residual log-likelihoods. Traditional, spatial, competition (using profile likelihood) and competition + spatial (using profile likelihood) models gave basically the same results in terms of the residual variance and heritability. Adjustment for competition did not reduce the heritability estimate. This is because competition is only at the residual level (discussed later) in this data set, i.e., is an environmental effect. The competition model (3) taking the average of the four neighbours (horizontally and vertically) as a covariate (fixed effect) and treatment effects as random, confirmed the significance of the competition effects ($\hat{\beta}$ = -0.25). Also it gave the same heritability as the traditional and the spatial models. However, the ordinary REML procedure applied here is not adequate because the competition coefficient appears in both the mean and variance of y and so can not be fitted as a covariate. The exact procedure of profile likelihood provides an exact adjustment and precise fitting for such model which improves the estimation of the variance and competition parameters. The competition coefficient estimate changed from -0.25 with ordinary REML to -0.10 with the profile REML. As observed for the sugarcane data set, the ordinary REML procedure overestimated the competition effects.

For the competition + spatial model, the ordinary REML procedure using a covariate gave very different results concerning to Log L, residual variance and heritability. The competition coefficient and auto-correlation parameters estimates were considerable higher than that obtained with the profile REML. This difference reveals the importance of using the more accurate profile REML procedure. The competition + spatial model using a covariate (model (13)) gave a much higher competition coefficient (-0.52 against -0.10 of the profile REML)

and the auto-correlation parameters were positive and high (0.21 and 0.21), i.e., they are modelling spatial trend.

The genetic competition + spatial model gave the same results as the spatial model, revealing no significance of genetic effects for competition (Table 8). So, the plausible competition coefficient is -0.10 and, alternatively, competition effects can be accounted for by the spatial model. It can be seen that the auto-correlation parameters estimates with the spatial model were also -0.10. When applied on neighbours in rows and columns separately, both estimated competition coefficients were about -0.10, i.e., identical to the values obtained for the auto-correlation parameters. This shows that, with no genetic competition, the spatial model and the phenotypic competition model are modelling the same effects, named a balance between residual competition and residual environmental trend. Residual competition and environmental trends are confounded effects and can not be separated. However, there is no practical need for such separation. The spatial model and the phenotypic competition model differ only in the presence of competition at genetic level (case of the Pinus data set, discussed later).

A comparison involving the traditional, spatial and competition models in terms of variety ranking is presented in Table 9. It can be seen that the three models produced very similar ranking and predicted treatment or variety effects. The same varieties can be selected by the three models with selection intensities of 20% (best 5 selected) or 50% (best 13 selected). This result is expected with low competition coefficients as that (-0.10) obtained in the present work. Using simulations, Kusnandar (2001) reported that competition models did not perform any better when the magnitude of competition parameters was small (between 0.0 and -0.10). According to the author, competition models turned more efficient with competition parameters higher than -0.3.

The variety effects should be corrected by using the expression $\tau_c = \tau \ /(1 - \hat{\beta})$. In this case, the competition coefficient was –0.10 and so the variety effects for the competition model (and also for the spatial model) in Table 9 should be divided by 1.10 or multiplied by 0.91. This is equivalent to multiply the heritability at treatment mean level by 0.91. For the traditional analysis such heritability is 0.69 and for the competition model is 0.69 as well. Multiplying this last value by 0.91 gives 0.63, which is smaller and more realistic than the 0.69 obtained through the traditional analysis. So, the use of competition and spatial

Table 9. Comparison involving the traditional, spatial and competition models in terms of variety ranking and predicted variety effects. **Eucalyptus maculata** data set.

	Variety Ranking			Variety Predicted Effects		
Varieties	Competition	Spatial	Traditional	Competition	Spatial	Traditional
579	1	1	1	10.26	10.12	10.78
565	2	2	2	8.104	8.281	8.153
572	3	3	3	6.854	6.993	6.933
580	4	5	4	5.042	4.522	5.263
577	5	4	5	4.786	4.954	4.653
573	6	7	6	2.384	1.509	2.558
576	7	12	7	2.136	2.453	2.194
584	8	8	10	1.556	1.858	1.438
561	9	9	9	1.552	1.620	1.490
562	10	6	8	1.334	1.276	1.553
581	11	10	12	1.184	0.975	1.110
563	12	11	11	1.166	2.407	1.122
574	13	13	13	0.629	-0.122	0.680

models in this case will therefore be of more importance in the estimation of genetic gains rather than of varietal selection.

The competition models using profile likelihood in both data sets, sugarcane and Eucalyptus, gave coherent results in terms of the non-significance of autoregressive terms in the joint model spatial + phenotypic competition. This is as expected, as the adjustment for competition effects addresses largely the same source of variation as the autoregressive parameters, when there is no competition at the genetic level. In this situation, the phenotypic competition model and the spatial model are likely to give the same results. In absence of genetic competition, the phenotypic competition method turns into the Papadakis method and is expected the produce the same results as the approaches of Papadakis (1937), Bartelett (1978) and Kempton and Howes (1981) for fertility trends. As the two dimensional separable autoregressive model encompasses the Papadakis method (Gilmour et al., 1997), the phenotypic competition model and the spatial model are expected to produce the same results in absence of genetic competition. Such results were not achieved by using the ordinary REML procedure. It is also important to mention that the use of the profile likelihood is an improved procedure over the Papadakis method. When fitting the Papadakis or the two dimensional separable autoregressive methods, a mixture of residual competition and local environmental trend is being modelled. Correll and Anderson (1983) found that the Papadakis term and the intervarietal competition were effectively uncorrelated. This is expected as the residual and genetic components of competition are likely to be independent effects.

In parametric terms, the competition effect of a plant i is given by $c_i = \phi_i + \gamma_i$, where ϕ_i is the genotypic competition effect and γ_i is the residual competition effect. The parametric model for the total residual effect is given by $e_i = \gamma_i + \xi_i + \eta_i$ and so the parametric model for the phenotypes (in terms of a vector) can be decomposed into $y = Xb + Z\tau + NZ\phi + \gamma + \xi + \eta$. The phenotypic competition model treats the elements ϕ_i, γ_i, ξ_i and η_i altogether in $\phi_i + \gamma_i + \xi_i + \eta_i$. The autoregressive spatial model considers $e_i = \gamma_i + \xi_i + \eta_i$. From these formulas it can be seen that the phenotypic competition and autoregressive spatial models are identical in absence of genetic competition. In general, the following models are optimal (in terms of considering all the specified effects in the model for phenotype) in the following situations:

(i) Autoregressive Spatial Model: optimal in absence of competition at the genetic level;

(ii) Phenotypic Competition Model via Profile Likelihood: optimal in any situation;

(iii) Phenotypic Competition + Autoregressive Spatial Model via Profile Likelihood: optimal in any situation, as it tends to be equivalent to (ii);

(iv) Genotypic Competition + Autoregressive Spatial Model: optimal in any situation;

(v) Genotypic Competition Model: optimal in absence of residual competition and local environmental trend.

It can also be pointed out that competition models are only needed when such competition has a genetic base. Without genetic competition, the traditional and/or autoregressive spatial models are sufficient. So, it is recommended to verify the significance of genetic competition effects as a first step in the analysis. This result will guide the statistician to better model choices for further analysis.

In the presence of genetic competition, there are two options: (a) use of a simultaneous model for genetic competition and for fertility trends (via autoregressive spatial model); (b) use of a phenotypic competition model using profile likelihood. The phenotypic model in (b) considers implicitly three effects: genetic competition, residual competition and environmental trend. The model in (a) consider explicitly the genetic competition and also allow for the covariance between treatment and competition effects. So, such model tends to be more precise and should be the choice for practical applications.

3.8.4 Genotypic and Phenotypic Competition Models in Pinus

For Pinus, genotypic and phenotypic competition models were applied. Results are presented in Table 10.

Table 10. Residual log-likelihoods (Log L) and estimates of the genetic variance among treatments ($\hat{\sigma}^2_\tau$), residual variance ($\hat{\sigma}^2$), heritability ($\hat{h}^2_{adj}$), competition coefficient ($\hat{\beta}$) and auto-correlation coefficients associated to columns (ARc) and rows (ARr). Pinus data set.

Model	Log L	$\hat{\sigma}^2_\tau$	$\hat{\sigma}^2$	$\hat{h}^2_{adj}$	$\hat{\beta}$	ARc	ARr
(a) Traditional	-6584.67	1.0403	18.255	0.2156	-	-	-
(b) Spatial	-6559.19	0.9621	17.891	0.2040	-	-0.10*	-0.13*
(c)Competition + Spatial (Profile)	-6512.25	1.1174	16.975	0.2470	-0.18*	-0.03 ns	-0.05 ns
(d) Competition (Covariate)	-6498.27	1.1795	16.960	0.2600	-0.23*	-	-
(e) Spatial + Competition(Cov.)	-6496.79	1.1497	16.945	0.2541	-0.22*	-0.01 ns	-0.04 ns
(f) Genotypic Compet.-North	-6574.99	1.0515	18.192	0.2186	-	-	-
(g) Spatial + G North	-6547.81	0.9837	17.831	0.2091	-	-0.10*	-0.13*
(h) Spatial + G North + Cov	-6543.87	0.9476	17.779	0.2024	-	-0.10*	-0.13*

The spatial model gave better fit than the traditional and revealed the presence of competition according to the significant negative auto-correlation coefficients for columns and rows. The presence of competition was confirmed by the significance of the phenotypic competition coefficient in (c) from Table 10, in a model which includes also spatial errors. This model, fitted via profile likelihood, gave no significance for the spatial autocorrelation parameters. The phenotypic competition models were also fitted using the covariate approach (models d and e in Table 10). As expected the competition coefficients were overestimated by the covariate approach. It can be seen that the phenotypic competition model differ from spatial model only in the presence of competition at genetic level. This occurred in the present data set (autoregressive competition parameter higher than the autoregressive spatial parameters) but not in the previous ones.

Considering competition at both levels, genotypic and residual, can be a better approach. This was done according to the models (g) and (h). Firstly, a model without the spatial term but including all the eight competitors was evaluated. This model revealed significance only for the northern neighbours at the genotypic level. So a genotypic competition model including only the northern neighbours was fitted in (f) from the Table 10. This model proved to be intermediate between the traditional and the spatial models (a) and (b), respectively, as can be seen from the residual log-likelihoods. So, the competition at the residual level proved to be higher than that at genotypic level.

Modelling competition simultaneously at the genotypic (northern neighbours) and residual levels according to the model (g) gave a better fit and showed the same values for the auto-correlation coefficients for columns and rows as in the spatial model in (b). This confirms that the spatial analysis was modelling competition only at residual level and that this is not sufficient in this case, as competition is also due to genetic causes.

A more complete model allowing for the covariance between direct and on neighbours effects was fitted as (h) in Table 10. This model gave a better fit than the model (g) without such covariance. Also gave a smaller heritability estimate as expected under competition adjustment and proved to be modelling competition adequately at both genotypic and residual levels. The same model revealed a negative genetic correlation between direct and on neighbour effects, of magnitude −0.68. This reveals the same tendency as observed by the phenotypic competition coefficient. The model also showed an adjusted heritability of 7.4% for the indirect effect on northern neighbours, i.e., heritability of the competition effects. The significant effects of only northern neighbours are likely to be due to shading according to the sun position in the region.

An explicit comparison between the phenotypic spatial (c) and genotypic spatial (h) models can not be done as they contain different fixed effects. Theoretically and conceptually the genotypic model is more complete. The models in (c) and (h) were compared in terms of variety ranking and genetic gain. Taking the model (h) as the best or correct one, it was verified the following coincidence (with selection by model h) rates with selection of the best 10% varieties: 91.7% for model (c), 83.3% for model (b) and 75.0% for model (a). So, the selection efficiency of the phenotypic model of competition was close to that of the genotypic model. However, the estimated genetic gains were 5.68% for the phenotypic model and 4.37% for the genotypic, which means an overestimation of 30% according to model (c), as expected due to the higher heritability estimate provided by such model.

Table 11 shows the negative genetic correlation between direct and on neighbour effects obtained with model g. The high and negative on neighbour effects of the best three varieties show that they are very aggressive and had their real value overestimated in the models without genetic competition. This shows the inefficiency of simple spatial and non-spatial models when there is genetic competition.

Table 11. Predicted random effects for the best varieties by the genotypic competition + spatial model. Pinus data set.

Varieties	Direct (τ_i)	Variety Ranking	
		Indirect (ϕ_i)	Total ($\tau_i + \phi_i$)
98	3.251	-1.238	2.013
96	2.034	-0.646	1.388
70	2.018	-0.932	1.086
20	1.061	-0.203	0.858
25	1.447	-0.745	0.702
66	0.839	-0.158	0.681
106	0.842	-0.186	0.656
99	1.046	-0.406	0.640
69	0.735	-0.176	0.558
21	0.504	0.044	0.547
45	0.612	-0.060	0.546
107	1.044	-0.499	0.544

3.9 Conclusions

- Results showed that the phenotypic competition coefficient encompassed the whole correlation pattern, including the genetic competition effect and a balance between residual competition effects and environmental trend.
- The exact procedure of REML profile likelihood provides an exact adjustment and precise fitting of phenotypic competition models and improves the estimation of the variance and competition parameters.
- Results revealed the inconsistence and inadequateness of the covariate approach for modelling competition and trend simultaneously. The profile likelihood approach should be always used instead.
- The spatial model and the phenotypic competition model differ only in the presence of competition at genetic level.
- In general, the following models are optimal according to the situations: (i) Autoregressive Spatial Model: optimal in absence of competition at the genetic level; (ii) Phenotypic Competition Model via Profile Likelihood: optimal in any situation; (iii) Phenotypic Competition + Autoregressive Spatial Model via Profile Likelihood: optimal in any situation, as it tends to be equivalent to (ii) because the adjustment for competition effects, besides considering genetic competition, addresses largely the same source of

variation as the autoregressive parameters; (iv) Genotypic Competition + Autoregressive Spatial Model: optimal in any situation; (v) Genotypic Competition Model: optimal in absence of residual competition and local environmental trend.

- Competition models are only needed when such competition has a genetic base. Without genetic competition, the traditional and/or autoregressive spatial models are sufficient.
- In the presence of genetic competition, there are two options: (a) use of a simultaneous model for genetic competition and for fertility trends via the autoregressive spatial model; (b) use of a phenotypic competition model using profile likelihood. The phenotypic model in (b) considers implicitly three effects: genetic competition, residual competition and environmental trend. The model in (a) considers explicitly the genetic competition and also allow for the covariance between treatment and competition effects. So, such model tends to be more precise and should be the choice for practical applications.

Acknowledgements

We would like to thank Arthur Gilmour (NSW-Agriculture, Australia), Brian Cullis (NSW-Agriculture, Australia), Ari Verbyla (Department of Statistics, University of Adelaide, Australia), Sue Welham (Biomathematics Unit, Rothamsted Research, UK), Joanne Stringer (Bureau of Sugarcane Experiment Stations, Australia) for helpful discussions. We also would like to thank Jose Alfredo Sturion (Embrapa, Brasil), Estefano Filho (Embrapa, Brasil), Gabriel Rezende (Aracruz Celulose), Aurelio Mendes (Aracruz Celulose), Marcio Barbosa (UFV), Mario Moraes (UNESP) and Robson Missio (UNESP) for providing data sets used in this work.

4. REFERENCES

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, p. 716-723, 1974.

APIOLAZA, L. A.; GILMOUR, A. R.; GARRICK, D. J. Variance modelling of longitudinal height data from a *Pinus radiata* progeny test. **Canadian Journal of Forest Research,** v. 30, p. 645-654, 2000.

ATKINSON, A. C. The use of residuals as a concomitant variable. **Biometrika**, v. 56, p. 33-41, 1969.

BARNDORFF-NIELSEN, O. E. Inference on full or partial parameters, based on the standardised log likelihood ratio. **Biometrika**, v. 73, p. 307-322, 1986.

BARNDORFF-NIELSEN, O. E. On a formula for the distribution of the maximum likelihood estimator. **Biometrika**, v. 70, p. 343-365, 1983.

BARTLETT, M. S. Nearest neighbour models in the analysis of field experiments. **Journal of the Royal Statistical Society.** Series B, v. 40, p. 147-174, 1978.

BARTLETT, M. S. The approximate recovery of information from replicated field experiments with large blocks. **Journal of Agricultural Science**, v. 28, p. 418-427, 1938.

BESAG, J. Spatial interaction and the statistical analysis of lattice systems. **Journal of the Royal Statistical Society**. Series B, v. 36, p. 192-236, 1974.

BESAG, J.; KEMPTON, R. A. Statistical analysis of field experiments using neighbouring plots. **Biometrics**, v. 42, p. 231–251, 1986.

COMREY, A. L.; HOWARD, B. L. **A first course in factor analysis**. Mahwah: Lawrence Erlbaum Assoc., 1992. 448 p.

COOPER, D. M.; THOMPSON, R. A note on the estimation of the parameters of the autoregressive moving average process. **Biometrika**, v. 64, p. 625-628, 1977.

CORREL, R. L. ; ANDERSON, R. B. Removal of intervarietal competition effects in forestry varietal trials. **Silvae Genetica**, v. 32, n. 5-6, p. 162-165, 1983.

COSTA E SILVA, J.; DUTKOWSKI, G. W.; GILMOUR, A. R. Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. **Silvae Genetica**, v. 31, p. 1887-1893, 2001.

COX, D. R.; REID, N. Parameter orthogonality and approximate conditional inference. **Journal of the Royal Statistical Society**. Series B, v. 49, p. 1-39, 1987.

CRESSIE, N. A. C. **Statistics for spatial data analysis**. New York: John Wiley & Sons, 1993. 900 p.

CROSSA, J. Statistical analysis of multi-location trials. **Advances in Agronomy**, v. 44, p. 55-85, 1990.

CROSSA, J.; GAUCH, H. G.; ZOBEL, R. W. Additive main effects and multiplicative interaction analysis of two international maize cultivars trials. **Crop Science**, v. 30, n. 3, p. 493-500, 1990.

CRUZ, C. D.; TORRES, R. A. A.; VENCOVSKY, R. An alternative to the stability analysis proposed by Silva and Barreto. **Brazilian Journal of Genetics**, v. 12, p. 567-580, 1989.

CULLIS, B. R.; SMITH, A. B.; VERBYLA, A. P.; WELHAM, S. J; THOMPSON, R. **Mixed models for data analysts**. New York: CRC Press, 2004. 288 p.

CULLIS, B. R.; GOGELL, B.; VERBYLA, A.; THOMPSON, R. Spatial analysis of multi-environment early generation variety trials. **Biometrics**, v. 54, p. 1-18, 1998.

CULLIS, B. R.; GLEESON, A. C. Spatial analysis of field experiments-an extension at two dimensions. **Biometrics**, v. 47, p. 1449-1460, 1991.

DRAPER, N. R.; GUTTMAN, I. Incorporating overlap effects from neighbouring units into response surface models. **Applied Statistics**, v. 29, p. 128-134, 1980.

DUARTE, J. B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal**. 2000. 293 f. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba.

DUARTE, J. B.; VENCOVSKY, R. **Interação genótipos x ambientes**: uma introdução a análise AMMI. Ribeirão Preto: Sociedade Brasileira de Genética, 1999. 60 p. (Série Monografias, 9).

DURBAN, M.; CURRIE, I. D. Adjustment of the profile likelihood for a class of normal regression models. **Scandinavian Journal of Statistics**, v. 27, p. 535-542, 2000.

DURBAN, M.; CURRIE, I.; KEMPTON, R. Adjusting for fertility and competition in variety trials. **Journal of Agricultural Science**, v. 136, p. 129-149, 2001.

DURBAN, M.; HACKET, C.; CURRIE, I. Blocks, trends and interference in field trials. In: INTERNATIONAL WORKSHOP ON STATISTICAL MODELLING, 14., 1999, Graz. **Proceedings**. p. 492-495. Disponível em: <http://www.ma.hw.ac.uk/~iain/research/graz.pdf> Acesso em: 10 jan. 2004.

EBERHART, S. A.; RUSSELL, W. A. Stability parameters for comparing varieties. **Crop Science**, v. 6, p. 36-40, 1966.

EEUWIJK, F. A. van; KEIZER, L. C. P.; BAKKER, J. J. Linear and bilinear models for the analysis of multi-environment trials. II. An application to data from Dutch Maize Variety Trials. **Euphytica**, v. 84, p. 9-22, 1995.

FINLAY, K. W.; WILKINSON, G. N. The analysis of adaptation in a plant breeding programme. **Australian Journal of Agricultural Research**, v. 14, p. 742-754, 1963.

FISHER, R. A. **Statistical methods for research workers**. London: Oliver and Boyd, 1925. 314 p.

FISHER, R. A.; MACKENZIE, W. A. Studies in crop variation. II. The manurial response of different potato varieties. **Journal of Agricultural Science**, v. 13, p. 311-320, 1923.

GALLAIS, A. Sur quelques aspects de la competition en amelioration des plantes. **Annales des Ameliorations des Plantes**, v. 25, p. 51-64, 1975.

GAUCH, H. G. Model selection and validation for yield trials with interaction. **Biometrics**, v. 44, p. 705-715, 1988.

GAUCH, H. G. **Statistical analysis of regional yield trials**: AMMI analysis of factorial designs. Amsterdam: Elsevier, 1992. 172 p.

GILMOUR, A. R.; CULLIS, B. R.; WELHAM, S. J.; GOGEL, B. J.; THOMPSON, R. An efficient computing strategy for prediction in mixed linear models. **Computational Statistics and Data Analysis**, v. 44, n. 4, p. 571-586, 2004.

GILMOUR, A. R.; CULLIS, B. R.; WELHAM, S. J.; THOMPSON, R. **ASREML reference manual**: release 1.0. 2nd. ed. Harpenden: Rothamsted Research, Biomathematics and Statistics Department, 2002. 187 p.

GILMOUR, A. R.; THOMPSON, R. Estimating parameters of a singular variance matrix in ASREML. In: AUSTRALASIAN GENSTAT CONFERENCE, 2002, Busselton. **Proceedings.** Busselton: Atlas Conferences Inc. Disponível em: <http://atlas-conferences.com/c/a/j/n/41.htm>. Acesso em: 05 fev. 2004.

GILMOUR, A. R.; THOMPSON, R. Modelling variance parameters in ASREML for repeated measures. In: WORLD CONGRESS ON GENETIC APPLIED TO LIVESTOCK PRODUCTION, 6., 1998, Armidale. **Proceedings**... Armidale: AGBU: University of New England, 1998. v. 27, p. 453-454.

GILMOUR, A. R.; THOMPSON, R.; CULLIS, B. R. Average information REML: an efficient algorithm for parameter estimation in linear mixed models. **Biometrics**, v. 51, p. 1440-1450, 1995.

GILMOUR, A. R.; THOMPSON, R.; CULLIS, B. R.; WELHAM, S. J. ASREML estimates variance matrices from multivariate data using the animal model. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier. **Proceedings.** Montpellier: INRA, 2002. 8 p. (Communication, 28-05).

GILMOUR, A. R. Post blocking gone too far! Recovery of information and spatial analysis in field experiments. **Biometrics**, v. 56, p. 944–946, 2000.

GILMOUR, A. R.; CULLIS, B. R.; VERBYLA, A. P. Accounting for natural and extraneous variation in the analysis of field experiments. **Journal of Agricultural Biological and Environmental Statistics**, v. 2, p. 269-293, 1997.

GILMOUR, A. R.; CULLIS, B. R.; FRENSHAM, A. B.; THOMPSON, R. (Co)variance structures for linear models in the analysis of plant improvement data. In: COMPSTAT98: Computational Statistics, 1998, Heidelberg: **Proceedings**. Heidelberg: Physica-Verlag, 1998. p. 53-64.

GLEESON, A. C. Spatial analysis. In: KEMPTON, R.A; FOX, P. N. **Statistical methods for plant variety evaluation**. London: Chapman & Hall, 1997. p. 68-85.

GLEESON, A. C.; CULLIS, B. R. Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. **Biometrics**, v. 43, p. 277-288, 1987.

GLENDINNING, D. R.; VERNON, A. J. Inter-varietal competition in cocoa trials. **Journal of Horticultural Science**, v. 40, p. 317-319, 1965.

GODDARD, M. E. A mixed model for analysis of data on multiple genetic markers. **Theoretical and Applied Genetics**, v. 83, p. 878-886, 1992.

GOLLOB, H. F. A statistical model which combines features of factor analytic and analysis of variance technique. **Psychometrika**, v. 33, n. 1, p. 73-115, 1968.

GREEN, P. J.; SILVERMAN, B. W. **Nonparametric regression and generalized linear models**. London: Chapman & Hall. 1994. 182 p.

GREEN, P. J.; JENNISON, C.; SEHEULT, A. Analysis of field experiments by least square smoothing. **Journal of the Royal Statistical Society**. Series B, v. 47, n. 2, p. 299-315, 1985.

GRONDONA, M. O.; CRESSIE, N. Using spatial considerations in the analysis of experiments. **Technometrics**, v. 33, p. 381-392, 1991.

GRONDONA, M. O.; CROSSA, J.; FOX, P. N.; PFEIFFER, W. H. Analysis of variety yield trials using two-dimensional separable ARIMA processes. **Biometrics**, v. 52, p. 763-770, 1996.

HEGYI, F. A simulation model for managing jack pine stands. In: FRIES, J. (Ed). **Growth models for tree and stand simulation**. Stockholm: Department of Forest Resources, 1974. p. 74-85.

HENDERSON, C. R. Sire evaluation and genetic trends. In: ANIMAL BREEDING AND GENETICS SYMPOSIUM IN HONOR OF J. LUSH, 1973, Champaign. **Proceedings**. Champaign: American Society of Animal Science, 1973. p. 10-41.

HENDERSON, C. R. **Aplications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462 p.

HILL, R. R.; ROSENBERGER, J. L. Methods for combining data from germplasm evaluation trials. **Crop Science**, v. 25, p. 467-470, 1985.

JAFFREZIC, F.; PLETCHER, S. D. Statistical models for estimating the genetic basis of repeated measures and other function valued traits. **Genetics**, v. 156, p. 913-922, 2000.

JAFFREZIC, F.; WHITE, I. M. S.; THOMPSON, R. Use of the score test as a goodness-of-fit measure of the covariance structure in genetic analysis of longitudinal data. **Genetics Selection Evolution**, v. 35, n. 2, p. 185-198, 2003.

JAFFREZIC, F.; WHITE, I. M. S.; THOMPSON, R.; VISSCHER, P. M; HILL, W. G. Statistical models for the genetic analysis of longitudinal data. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier. **Proceedings**. Montpellier: INRA, 2002. 4 p. (Communication, 16-08).

JENNRICH, R. L.; SCHLUCHTER, M. D. Unbalanced repeated measures models with structured covariance matrices. **Biometrics**, v. 42, p. 805-820, 1986.

JOHNSON, D. L.; THOMPSON, R. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. **Journal of Dairy Science**, v. 78, p. 449-456, 1995.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. Englewood: Prentice Hall Inc., 1988. 594 p.

JOYCE, D.; FORD, R.; FU, Y. B. Spatial patterns of tree height variations in a black spruce farm-field progeny test and neighbors-adjusted estimations of genetic parameters. **Silvae Genetica**, v. 51, n. 1, p. 13-18, 2002.

KEMPTON, R. A. Adjustment for competition between varieties in plant breeding trials. **Journal of Agricultural Science**, Cambridge, v. 98, p. 599-611, 1982.

KEMPTON, R. A. Discussion on the analysis of designed experiments and longitudinal data using smootthing splines. **Journal of the Royal Statistical Society**. Series C, v. 48, p. 300-301, 1999. Original text by Verbyla, A. P; Cullis, B. R.; Kenward, M. G.; Welham, S. J.

KEMPTON, R. A. Statistical models for interplot competition. **Aspects of Applied Biology**, v. 10, p. 11-120, 1985.

KEMPTON, R. A.; HOWES, C. W. The use of neighbouring plot values in the analysis of variety trials. **Applied Statistics**, v. 30, p. 59-70, 1981.

KEMPTON, R. A.; SERAPHIN, J. C.; SWORD, A. M. Statistical analysis of two dimensional variation in variety yield trials. **Journal of Agricultural Science**, v. 122, p. 335-342, 1994.

KIRKPATRICK, M.; HILL, W. G.; THOMPSON, R. Estimating the covariance structure of traits during growth and ageing, illustrated with lactations in dairy cattle. **Genetical Research**, v. 64, p. 57-69, 1994.

KUSNANDAR, D. **The identification and interpretation of genetic variation in forestry plantation**. Perth: The University of Western Australia, 2001. 3 p. PhD Thesis-Abstract.

LAWLEY, D. N.; MAXWELL, A. E. **Factor analysis as a statistical tool.** 2nd. ed. London: Lawrence Erlbaum Assoc, 1992. 448 p.

LEONARDECZ-NETO, E. **Competição intergenotípica na análise de testes de progênies em essencias florestais.**2002. 60 f. Tese (Doutorado) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba.

LOTODÉ, R.; LACHENAUD, P. Méthodologie destinée aux essais de sélection du cacaoyer. **Café Cacao Thé,** v. 32, n. 4, p. 275-292, 1988.

MAGNUSSEN, S. A method to adjust simultaneously for spatial microsite and competition effects. **Canadian Journal of Forestry Research**, v. 24, p. 985-995, 1994.

MAGNUSSEN, S.; YEATMAN, C. W. Adjusting for inter-row competition in a jack pine provenance trial. **Silvae Genetica**, v. 36, n. 5-6, p. 206-214, 1987.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. London: Academic Press, 1988.

MARTIN, R. J. The use of time-series models and methods in the analysis of agricultural field trials. **Communications in Statistics**. Part A: Theory and Methods, v. 19, n. 1, p. 55-81, 1990.

McCULLAGH, P.; TIBSHIRANI, R. A simple method for the adjustment of profile likelihoods. **Journal of the Royal Statistics Society**. Series B, v. 52, p. 325-344, 1990.

MEAD, R. A mathematical model for the estimation of interplant competition. **Biometrics**, v. 23, p. 189-205, 1967.

MEYER, K.; HILL, W. G. Estimation of genetic and phenotypic covariance functions for longitudinal or repeated records by restricted maximum likelihood. **Livestock Production Science**, v. 47, p. 185-200, 1997.

MONTAGNON, C.; FLORI, A.; CILAS, C. A new method to assess competition in coffee clonal trials with single-tree plots in Cote d'Ivoire. **Agronomy Journal**, v. 93, p. 227-231, 2001.

MOREAU, L.; MONOD, H.; CHARCOSSET, A.; GALLAIS, A. Marker assisted selection with spatial analysis of unreplicated field trials. **Theoretical and Applied Genetics**, v. 98, p. 234-242, 1999.

NOUY, B.; ASMADY, A.; LUBIS, R. Effets de competition a Nord-Sumatra dans le essais genetiques sur palmier a huile. Consequences sur l'evaluation du materiel vegetal. **Oleagineux**, v. 45, p. 245-255, 1990.

NUNEZ-ANTON, V.; ZIMMERMANN, D. L. Modelling non-stationary longitudinal data. **Biometrics**, v. 56, p. 699-705, 2000.

PAPADAKIS, J. Advances in the analysis of field experiments. **Proceedings of the Academy of Athens**, v. 59, p. 326-342, 1984.

PAPADAKIS, J. **Agricultural Research**: principles, methodology, suggestions. Buenos Aires: Ed. Argentina, 1970.

PAPADAKIS, J. S. **Méthode statistique pour des expériences sur champ**. Thessalonike: Institut d'Amélioration des Plantes à Salonique, 1937. 30 p. (Bulletin, 23).

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, v. 58, p. 545-554, 1971.

PEARCE, S. C. Experimenting with organisms as blocks. **Biometrika**, v. 44, p. 141-149, 1957.

PIEPHO, H. P. Analysing genotype-environment interaction data by mixed models with multiplicative terms. **Biometrics**, v. 53, p. 761-767, 1997.

PIEPHO, H. P. Best linear unbiased prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. **Theoretical and Applied Genetics**, v. 89, p. 647-654, 1994.

PIEPHO, H. P. Empirical best linear unbiased prediction in cultivar trials using factor analytic variance-covariance structures. **Theoretical and Applied Genetics**, v. 97, p. 195-201, 1998.

PITHUNCHARURNLAP, M.; BASFORD, K. E.; FEDERER, W. T. Neighbour analysis with adjustment for interplot competition. **Australian Journal of Statistics**, v. 35, n. 3, p. 263-270, 1993.

PLETCHER, S. D.; GEYER, C. J. The genetic analysis of age-dependent traits: modelling a character process. **Genetics**, v. 153, p. 825-833, 1999.

QIAO, C. G.; BASFORD, K. E.;DELACY, I. H.;COOPER, M. Evaluation of experimental designs and spatial analysis in wheat breeding trials. **Theoretical and Applied Genetics**, v. 100, p. 9-16, 2000.

RESENDE, M. D. V. de; STURION, J. A. **Análise genética de dados com dependência espacial e temporal no melhoramento de plantas perenes via modelos geoestatísticos e de séries temporais empregando REML/BLUP ao nível individual.** Colombo: Embrapa Florestas, 2001. 80 p. (Embrapa Florestas. Documentos, 65).

RESENDE, M. D. V. de; THOMPSON, R.; WELHAM, S. J. Multivariate spatial statistical analysis in perennial crops. In: INTERNATIONAL BIOMETRIC SOCIETY CONFERENCE, 2003, British Region, 2003. **Proceedings**. Reading: University of Reading, School of Applied Statistics, 2003. p. 70-71.

ROMAGOSA, I.; ULLRICH, S. E.; HAN, F.; HAYES, P. M. Use of additive main effects and multiplicative interaction model in QTL mapping for adaptation in barley. **Theoretical and Applied Genetics**, v. 93, p. 30-37, 1996.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, v. 6, n. 2, p. 461-464, 1978.

SEARLE, S. R.; CASELLA, G.; Mc CULLOCH, C. E. **Variance components**. New York. J. Wiley. 1992. 528 p.

SMITH, A.; CULLIS, B. R.; THOMPSON, R. Analysing variety by environment data using multiplicative mixed models and adjustment for spatial field trend. **Biometrics**, v. 57, p. 1138-1147, 2001.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics**. 2th. ed. New York: Mac Graw-Hill, 1980. 633 p.

STRINGER, J. K.; CULLIS, B. R. Application of spatial analysis techniques to adjust for fertility trends and identify interplot competition in early stage sugarcane selection trials. **Australian Journal of Agricultural Research**, v. 53, n. 8, p. 911-918, 2002a.

STRINGER, J. K.; CULLIS, B. R. Joint modelling of spatial variability and interplot competition. In: AUSTRALASIAN PLANT BREEDING CONFERENCE, 12., 2002, Perth. **Proceedings**. Perth: University of Western Australia, 2002b. p. 614-619. McComb, J. A. (Ed).

STROUP, W. W.; MULITZE, D. K. Nearest neighbour adjusted best linear unbiased prediction. **American Statistician**, v. 45, p. 194-200, 1991.

TALBOT, M.; MILNER, A. D.; NUTKINS, M. A. E.; LAW, J. R. Effect of interference between plots plots on yield performance in crop variety trials. **Journal of Agricultural Science**, Cambridge, v. 124, p. 335-342, 1995.

THOMPSON, R. Maximum likelihood estimation of variance components. **Math. Operationsforsh. Statist.**, v. 11, p. 545-561, 1980.

THOMPSON, R. A review of genetic parameters estimation. In: WORLD CONGRESS OF GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7., 2002, Montpellier. **Proceedings**. Montpellier: INRA, 2002. 1 CD-ROM. Communication n. 17-01.

THOMPSON, R. Sire evaluation. **Biometrics**, v. 35, p. 339-353, 1979.

THOMPSON, R. The estimation of heritability with unbalanced data. **Biometrics**, v. 33, p. 485-504, 1977.

THOMPSON, R. The estimation of variance and covariance components when records are subject to culling. **Biometrics**, v. 29, p. 527-550, 1973.

THOMPSON, R.; CULLIS, B. R.; SMITH, A. B.; GILMOUR, A. R. A sparse implementation of the average information algorithm for factor analytic and reduced rank variance models. **Australian and New Zealand Journal of Statistics**, v. 45, n. 4, p. 445-459, 2003.

THOMPSON, R.; WELHAM, S. J. REML analysis of mixed models. In: PAYNE et al. (Ed.). **GenStat 6 Release 6.1**: the guide to GenStat, v. 2 – statistics. Harpenden: Rothamsted Research, 2003. p. 469-560.

THOMPSON, R.; WRAY, N. R; CRUMP, R. E. Calculation of prediction error variances using sparse matrix methods. **Journal of Animal Breeding and Genetics**, v. 111, p. 102-109, 1994.

VERBYLA, A. P.; CULLIS, B. R.; KENWARD, M. G.; WELHAM, S. J. The analysis of designed experiments and longitudinal data using smoothing splines. **Journal of the Royal Statistical Society.** Series C, v. 48, p. 269-311, 1999.

WELHAM, S. J.; THOMPSON, R. Likelihood ratio tests for fixed models terms using residual maximum likelihood. **Journal of the Royal Statistical Society**. Series B, v. 59, p. 701-719, 1997.

WELHAM, S. J.; THOMPSON, R.; GILMOUR, A. R. A general form for specification of correlated error models with allowance for heterogeneity. In: COMPSTAT98: Computational Statistics, 1998, Heidelberg: **Proceedings**. Heidelberg: Physica-Verlag, 1998. p. 479-484.

WHITE, I. M. S.; THOMPSON, R.; BROTHERSTONE, S. Genetic and environmental smoothing of lactation curves with cubic splines. **Journal of Dairy Science**, v. 82, p. 632-638, 1999.

WILKINSON, G. N.; ECKERT, S. R.; HANCOCK, T. W.; MAYO, O. Nearest neighbor (NN) analysis of field experiments. **Journal of the Royal Statistical Society.** Series B, v. 45, p. 151-211, 1983.

WILKS, S. S. The large sample distribution of the likelihood ratio for testing composite hypothesis. **Annals of Mathematical Statistics**, v. 9, p. 60-62, 1938.

WILLIAMS, E. R. A neighbor model for field experiments. **Biometrika**, v. 73, p. 279-287, 1986.

WRICKE, G. Zur berechning der okovalens bei sommerweizen und hafer. **Pflanzen-zuchtung**, v. 52, p. 127-138, 1965.

ZIMMERMAN, D. I.; HARVILLE, D. A. A random field approach to the analysis of field-plot experiments and other spatial experiments. **Biometrics**, v. 47, p. 223-239, 1991.