

# Direct and Competition Additive Effects in Tree Breeding: Bayesian Estimation From an Individual Tree Mixed Model

By E. P. CAPPA<sup>1,2,\*</sup> and R. J. C. CANTET<sup>1,3</sup>

(Received 23<sup>th</sup> August 2006)

## Abstract

An individual tree model with additive direct and competition effects is introduced to account for competitive effects in forest genetics evaluation. The mixed linear model includes fixed effects as well as direct and competition breeding values plus permanent environmental effects. Competition effects, either additive or environmental, are identified in the phenotype of a competitor tree by means of 'intensity of competition' elements (IC), which are non-zero elements of the incidence matrix of the additive competition effects. The ICs are inverse function of the distance and the number of competing individuals, either row-column wise or diagonally. The ICs allow standardization of the variance of competition effects in the phenotypic variance of any individual tree, so that the model accounts for unequal number of neighbors. Expressions are obtained for the bias in estimating additive variance using the covariance between half-sibs, when ignoring competition effects for row-plot designs and for single-tree plot designs. A data set of loblolly pines on growth at breast height is used to estimate the additive variances of direct and competition effects, the covariance between both effects, and the variance of permanent environmental effects using a Bayesian method via Gibbs sampling and Restricted Maximum Likelihood procedures (REML) via the Expectation-Maximization (EM) algorithm. No problem of convergence was detected with the model and ICs used when compared to what has been reported in the animal breeding literature for such models. Posterior means (standard error) of the estimated parameters were  $\hat{\sigma}_{Ad}^2 = 12.553$  (1.447),  $\hat{\sigma}_{Ac}^2 = 1.259$  (0.259),  $\hat{\sigma}_{AdAc} = -3.126$  (0.492),  $\hat{\sigma}_p^2 = 1.186$  (0.289), and  $\hat{\sigma}_e^2 = 5.819$  (1.07). Leaving permanent environmental competition effects out of the model may bias the predictions of direct breeding values. Results suggest that selection for increasing direct growth while keeping a low level of competition is feasible.

*Key words:* competition effects, individual tree mixed model, additive and direct competition effects, estimation of additive (co)variances, Gibbs sampling.

## Introduction

The additive genetic variance is a parameter of utmost importance in the genetic improvement of forest

trees as it affects the gain and the precision of selection. In order to avoid bias when estimating additive variance, the statistical model of analysis should include all other sources of genetic variation as well as all identifiable environmental effects. Tree competition for resources may bias breeding value estimation from competing individuals (see for example, MAGNUSSEN, 1993; FOSTER *et al.*, 1998; RADTKE *et al.*, 2003) by inducing a negative correlation between either individual trees or neighbor plots. Competition is defined as the stress suffered by a plant due to the genotype and the spatial arrangement of neighboring trees (HINSON and HANSON, 1962), and is caused by genetic and environmental sources (MAGNUSSEN, 1989). Usually genetics effects of competition are not accounted for in the model of evaluation in spite of evidence of their existence (see the references in the discussion of MAGNUSSEN, 1993 and FOSTER *et al.*, 1998). CANNELL (1978) suggested selecting for non-competitive genotypes in order to increase yield per unit of area in forest trees. This breeding strategy is most effective if direct effects for growth are negatively correlated to competition effects. In this scenario, plants with a large genetic potential for growth tend to induce less competition, so that tree density may be increased and, as a result of both increases, the yield per unit area would be augmented. On the other hand, if the correlation between direct and competition effects is positive, selection for higher growth will result in more competitive individuals. As a consequence, total yield per unit area may be affected as the faster growing individuals would hinder the growth of their neighbors, which in turn would decrease total production. Inclusion of genetic effects of competition results in an increased number of additive dispersion parameters in the model of genetic evaluation compared with those models where competition effects are absent.

In a series of papers, GRIFFING (1967, 1968a, 1968b) described models that include genetic effects of competition among individuals or groups of individuals, and analyzed the consequences of using such models for the response to selection. In these models, the phenotype of an individual is a linear combination of its genetic effects ("direct genetic effects") plus the genetic contributions from other genotypes ("indirect genetic effects"). Whereas direct genetic effects are expressed in the phenotype of an individual, indirect genetic effects are expressed only in the phenotype of another individual. A common example of indirect effects is the maternal effect in mammals (WILLHAM, 1963; LYNCH and WALSH, 1998, chapter 23), which is expressed in the offspring from birth to weaning. When looking at the genetic evaluation of trees, competition is an indirect genetic effect. WRIGHT (1986) obtained expressions for the covariance

<sup>1</sup>) Department of Animal Production, University of Buenos Aires, Avenida San Martín 4453, C1417DSQ Buenos Aires, Argentina.

<sup>2</sup>) Doctoral fellow, "Fondo para la Investigación Científica y Tecnológica" (FONCyT), Argentina.

<sup>3</sup>) "Consejo Nacional de Investigaciones Científicas y Técnicas" (CONICET), Argentina.

<sup>\*</sup>) Corresponding author: EDUARDO PABLO CAPPA, Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, Av. San Martín 4453, C1417DSQ Buenos Aires, Argentina. Phone: +54-11-4524-8000 ext. 8192. Fax: +54-11-4514-8735 or 8737. E-mail: [ecappa@mail.agro.uba.ar](mailto:ecappa@mail.agro.uba.ar)

between relatives in the model of GRIFFING (1967), including additive, dominance and epistasis, for direct and competition effects. For a single trait individual tree model, 22 genetic (co)variance components have to be estimated: 3 for additive effects, 3 for dominance effects, and 16 for epistatic effects. Attempting to fit a model with all 22 parameters to data with the usual family relationships (full and half-sibs) commonly found in trees seems to be hopeless. Instead, it may be feasible to fit additive effects, as it requires estimating only three (co)variance components plus the error term. MUIR and SCHINCKEL (2002) described an animal model with direct and competitive effects. VAN VLECK and CASSADY (2005) used simulation and a fixed number of competitors, to determine whether Restricted Maximum Likelihood (REML, PATTERSON and THOMPSON, 1971) with relationships could untangle the covariance structure of direct and competition genetic variances and the covariance between them, whereas ARANGO *et al.* (2005) attempted to estimate the three (co)variance components to a swine population during a growth test. The results of both studies suggest that the additive relationships among competing individuals present in the data may be essential in disentangling the information to estimate the (co)variance components. MUIR (2005) wrote down the mixed model equations with competition effects to analyze forest tree data, but he ignored the consequences of a variable number of competitors (due for example to mortality or thinning) on estimating the additive genetic variance for competition. All in all, when fitting a model with competition effects it is essential to report a measure of the variability of the estimates of the dispersion parameters.

A shortcoming of REML is that the variability of estimation has to be measured approximately using large sample theory by means of the inverse of the information matrix from the marginal or restricted likelihood, as there is no analytic solution for the variance of the estimating equations: there is no exact sampling distribution for the REML estimators. From a Bayesian perspective REML can be seen as the mode of the joint posterior distribution of all (co)variance components after integrating out the fixed effects using a flat prior for the dispersion parameters (HARVILLE, 1974). Alternatively, a full Bayesian approach by means of the Gibbs sampler can be attempted for estimating the (co)variance components for additive direct and indirect effects, by exploiting the similarity with the model of maternal effects (SORENSEN and GIANOLA, 2002, section 13.3). The basics of the Gibbs sampling is discussed by CASELLA and GEORGE (1992), whereas SORIA *et al.* (1998), GWAZE and WOOLLIAMS (2001), ZENG *et al.* (2004) and CAPPA and CANTET (2006) developed some applications of the sampler to the genetic improvement of forest trees. The goals of this research are: 1) to introduce an additive genetic individual tree model that includes direct and competition effects, accounting for the number and position of competitor trees; 2) to estimate the dispersion parameters of the model (additive variances for direct and competition effects, and the covariance between both effects) using a Bayesian approach by means of the Gibbs sampler. Developments are illustrated with data

on the diameter at breast height in *Pinus taeda* L. at 13 years of age.

## The model

### Breeding values for direct and competition effects

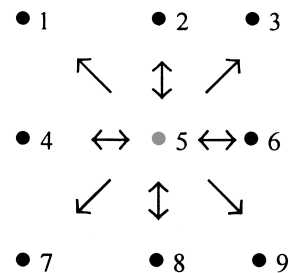
Consider a trait mostly affected by additive genetic effects, with a direct and a competition component (WRIGHT, 1986). The dynamics of the joint genetic variability for additive effects with an indirect component has been developed by WILLHAM (1963) for maternal effects, and more generally discussed by WOLF (2003) within an evolutionary framework. Let the direct component of the breeding value for individual  $i$  be  $\mathbf{a}_{di}$ , and let the competition breeding value be  $\mathbf{a}_{ci}$ . The greater the magnitude of  $\mathbf{a}_{ci}$  the higher the competition effects exerted by tree  $i$ . The phenotypic record of tree  $i$  ( $y_{ij1..jm}$ ) is affected by the  $\mathbf{a}_{cj}$ 's of neighbor trees  $j$  ( $j = j_1, j_2, \dots, j_m$ ), in a similar fashion to maternal effects which are expressed in the progeny's phenotype. Thus,  $y_{ij1..jm}$  depends on the  $\mathbf{a}_{cj}$  of its neighbors but not on its own  $\mathbf{a}_{ci}$ . As an individual tree suffers competition from more than a neighbor, let  $m_i$  be the number of competitors of the  $i^{th}$  tree. For plantations in a regular grid, the maximum value  $m_i$  can take is 8. The assumption is that any tree does not compete with other trees than its nearest neighbors (see *Figure 1*). The notation R-C indicates that the competitor lies, either in the same row or in the same column, and is represented with the symbol ' $\leftrightarrow$ ' in *Figure 1*. In the same way, the letter D refers to competitors that lie diagonal, and these are represented with the diagonal arrows in *Figure 1*. To exemplify, in *Figure 1* the R-C competitors of tree 5 are plants 2, 4, 6 and 8, whereas individuals 1, 3, 7 and 9 are D competitors of 5.

The total additive genetic competition that is exerted over plant  $i$  from trees  $j_1, j_2, \dots, j_{m_i}$  is equal to

$$f_{i1} \mathbf{a}_{c_1} + f_{i2} \mathbf{a}_{c_2} + \dots + f_{im} \mathbf{a}_{c_m} = \sum_{j=1}^{m_i} f_{ij} \mathbf{a}_{c_j} \quad [1]$$

The element  $f_{ij}$  is interpreted as the *intensity of competition* (IC) that  $\mathbf{a}_{cj}$  ( $j = j_1, j_2, \dots, j_m$ ) exerts over the phenotype of the  $i^{th}$  neighbor tree ( $y_{ij1..jm}$ ). To obtain the IC values, consideration should be given to the fact that, in the absence of inbreeding and of genetic relationships among competitors and regardless of the number of competitors,

$$\text{Var} \left( \sum_{j=1}^{m_i} f_{ij} \mathbf{a}_{c_j} \right) = \sigma_{Ac}^2 \quad [2]$$



*Figure 1.* – A diagram of competition effects in a regular grid.

where  $\sigma_{Ac}^2$  is the additive variance for competition breeding values. This implies that the potential genetic strength for competing with the neighbors, i.e.  $\mathbf{a}_c$ , is split proportionally to  $m_i$ . Let  $f_{ijR-C}$  be the IC in [1] for R-C and  $f_{ijD}$  for D competitors to  $i$ . Also, let  $n_{R-C}$  and  $n_D$  be the respective numbers of R-C and D competitors, and  $n_{R-C} + n_D = m_i$ . Under all these premises, and on using the variance operator in [2], we obtain

$$\begin{aligned} \text{Var}\left(\sum_{j=1}^{m_i} f_{ij} \mathbf{a}_{c_j}\right) &= (n_{R-C} f_{ijR-C}^2 + n_D f_{ijD}^2) \sigma_{Ac}^2 \\ &= \left(\sum_{j=1}^{m_i} f_{ijk}^2\right) \sigma_{Ac}^2 \end{aligned} \quad [3]$$

where  $k = R-C$  or  $D$ . By equating [2] to [3], the ICs are such that

$$\sum_{j=1}^{m_i} f_{ijk}^2 = n_{R-C} f_{ijR-C}^2 + n_D f_{ijD}^2 = 1 \quad [4]$$

Now, it seems reasonable to assume that in trees IC is related to the inverse of the *distance* between  $i$  and  $j$  (i.e. RADTKE *et al.*, 2003). If  $d$  is the regular spacing of the planting design, for R-C competitors the IC is proportional to  $1/d$ . However, the distance of a tree located diagonal to  $i$  is  $2^{1/2} d$  by the Pythagorean theorem, so that competition is proportional to  $1/(2^{1/2} d)$  for D competitors. Now,  $f_{ijR-C} = 1/d$  and  $f_{ijD} = 1/(2^{1/2} d)$ . On solving for  $d$  in both equalities we get  $d = 1/f_{ijR-C}$  and  $d = 1/f_{ijD} 2^{1/2}$ . Next is to equate the resulting expressions so as to obtain

$$\sqrt{2} f_{ijD} = f_{ijR-C} \quad [5]$$

Finally, by replacing with [5] in [4] and solving for  $f_{ijD}$  and  $f_{ijR-C}$  produces

$$n_{R-C} 2 f_{ijD}^2 + n_D f_{ijD}^2 = 1 \quad \text{or} \quad (2n_{R-C} + n_D) f_{ijD}^2 = 1$$

Thus

$$f_{ijD} = \frac{1}{\sqrt{2n_{R-C} + n_D}} \quad [6]$$

A similar argument is used to obtain

$$f_{ijR-C} = \sqrt{\frac{2}{2n_{R-C} + n_D}} \quad [7]$$

Notice that  $f_{ij}$  is not necessarily equal to  $f_{ji}$  as  $i$  and  $j$  may have different number of competitors. An interesting feature of expressions [6] and [7] is that the intensity factors end up being independent on the distance  $d$ . The reason is that in row-column arrays there is a relation between the RC-distance and the D-distance based on the theorem of Pythagoras, and the value of  $d$  cancels out when obtaining [5]. This approach can be simply extended to planting designs when the distance between rows and columns is not the same, i.e. when  $d_{\text{Row}} \neq d_{\text{Column}}$ .

GRIFFIN (1967) and WRIGHT (1986) observed that the covariance between  $\mathbf{a}_{di}$  and  $\mathbf{a}_{ci}$  from the  $i$ -th tree is equal to:

$$\text{cov}(\mathbf{a}_{d_i}, \mathbf{a}_{c_j}) = \mathbf{A}_{ij} \sigma_{AdAc}$$

where  $\mathbf{A}_{ij}$  is the additive genetic relationship between tree  $i$  and its  $j$  neighbor, and  $\sigma_{AdAc}$  is the covariance between direct and competition breeding values. Moreover, as the genetic model is exclusively an additive one, we have that

$$\text{Var}(\mathbf{a}_{d_i}) = (1 + F_i) \sigma_{Ad}^2 \quad \text{Var}(\mathbf{a}_{c_j}) = (1 + F_j) \sigma_{Ac}^2$$

(KEMPTHORNE, 1969; page 349). In this expression,  $F_i$  and  $F_j$  are the inbreeding coefficients of  $i$  and  $j$ , respectively;  $\sigma_{Ad}^2$  is the additive variance for direct effect. In terms of the covariance matrix of breeding values we have

$$\text{Var} \begin{bmatrix} \mathbf{a}_{d_i} \\ \mathbf{a}_{c_j} \end{bmatrix} = \begin{bmatrix} (1 + F_i) \sigma_{Ad}^2 & \mathbf{A}_{ij} \sigma_{AdAc} \\ \mathbf{A}_{ij} \sigma_{AdAc} & (1 + F_j) \sigma_{Ac}^2 \end{bmatrix}$$

*Additive individual tree mixed model with direct and competition breeding values*

On inserting [1] into an additive individual tree model (BORRALHO, 1995) for the record of tree  $i$  competing with neighbors  $j_1, j_2, \dots, j_m$  produces

$$y_{ij1..jm} = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{a}_{d_i} + \sum_{j=1}^{m_i} f_{ij} (\mathbf{a}_{c_j} + \mathbf{p}_{c_j}) + e_{ij1..jm} \quad [8]$$

In [8],  $y_{ij1..jm}$  is the phenotype of  $i$  ( $i = 1, \dots, n$ ;  $n$  is the total number of trees with data recorded). The  $p \times 1$  vector  $\boldsymbol{\beta}$  contains fixed effects such as site or block, and is associated to the data by the  $i^{\text{th}}$  row of the incidence matrix  $\mathbf{X}$  ( $n \times p$ ). The breeding values  $\mathbf{a}_{di}$  and  $\mathbf{a}_{cj}$  and the  $f_{ij}$  are as defined above. Model [8] includes permanent environmental effects through the random variable  $\mathbf{p}_{c_j}$ , so that  $\mathbf{a}_{c_j} + \mathbf{p}_{c_j}$  is the phenotypic effect of competitor  $j$  over the phenotype of  $i$  (MUIR, 2005). Finally,  $e_{ij1..jm}$  is the random error term.

The variance of  $y_{ij1..jm}$  in [8] is equal to

$$\begin{aligned} \text{Var}(y_{ij1..jm}) &= \text{Var}\left(\mathbf{a}_{d_i} + \sum_{j=1}^{m_i} f_{ij} \mathbf{a}_{c_j}\right) + \text{Var}\left(\sum_{j=1}^{m_i} f_{ij} \mathbf{p}_{c_j}\right) + \text{Var}(e_{ij1..jm}) \end{aligned}$$

where the first variance is for the additive effects, the second one for the permanent environmental effects, and the remaining one for the error. It is shown in the Appendix A that total additive variance in model [8] is equal to

$$\begin{aligned} \text{Var}\left(\mathbf{a}_{d_i} + \sum_{j=1}^{m_i} f_{ijk} \mathbf{a}_{c_j}\right) &= (1 + F_i) \sigma_{Ad}^2 \\ &+ \left[(n_{R-C} f_{iR-C}^2 + n_D f_{iD}^2)(1 + F_j) + 2 \sum_{j \neq j'}^{m_i} f_{ij'k} f_{imk} \mathbf{A}_{j'j}\right] \sigma_{Ac}^2 \\ &+ 2 \sum_{j=1}^{m_i} f_{ijk} \mathbf{A}_{ij} \sigma_{AdAc} \end{aligned} \quad [9]$$

When: 1) all individuals are not inbred ( $F_i = F_j = 0$ ), 2) tree  $i$  is unrelated to its competitors ( $\mathbf{A}_{ij} = 0$ , for all  $j$ ), and 3) the competitors are unrelated among themselves ( $\mathbf{A}_{j'j} = 0$ ), expression [9] reduces to  $\sigma_{Ad}^2 + \sigma_{Ac}^2$ .

In matrix notation, the individual tree model [8] is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_d \mathbf{a}_d + \mathbf{Z}_c \mathbf{a}_c + \mathbf{Z}_p \mathbf{p}_c + \mathbf{e} \quad [10]$$

where  $y = [y_{ij1..jm}]$  ( $n \times 1$ ) contains the data;  $\mathbf{X}$  is the  $n \times p$  incidence matrix relating records to the vector of fixed effects  $\boldsymbol{\beta}$ ,  $\mathbf{p}_c$  is a vector such that  $\mathbf{p}_c = [\mathbf{p}_{c_j}]$   $j = 1, \dots, n$ , such that  $\mathbf{p}_c \sim N_n(\mathbf{0}, I_n \sigma_p^2)$ , and  $\mathbf{e}$  ( $n \times 1$ ) is the random vector of i.i.d. errors distributed as  $N_n(\mathbf{0}, I_n \sigma_e^2)$ , being  $\sigma_e^2$  the error variance. Direct breeding values are included in the random vector  $\mathbf{a}_d = [\mathbf{a}_{di}]$  and competition breeding values are in  $\mathbf{a}_c = [\mathbf{a}_{ci}]$ . The same  $q$  individuals having direct breeding values in  $\mathbf{a}_d$  are also in  $\mathbf{a}_c$ , and in the same order. Direct and competition breeding values are related to  $y$  by the  $n \times q$  incidence matrices  $\mathbf{Z}_d$  and  $\mathbf{Z}_c$ , respectively. Every row of  $\mathbf{Z}_d$  has all elements equal to 0 except for a 1 in the column belonging to  $\mathbf{a}_{di}$ . Similarly, matrix  $\mathbf{Z}_c$  has rows with 0 elements but the  $f_{ij}$  's in the columns for the  $\mathbf{a}_{ci}$  of the  $m_i$  competitors of tree  $i$ . To exemplify, suppose that in *Figure 1* there is a missing plant in position 3. The row of  $\mathbf{Z}_c$  relating the record of 5 ( $i = 5$ ) to its competitors is

$$\mathbf{Z}_{c5} = \left[ \frac{1}{\sqrt{11}} \quad \frac{\sqrt{2}}{\sqrt{11}} \quad \frac{\sqrt{2}}{\sqrt{11}} \quad 0 \quad \frac{\sqrt{2}}{\sqrt{11}} \quad \frac{1}{\sqrt{11}} \quad \frac{\sqrt{2}}{\sqrt{11}} \quad \frac{1}{\sqrt{11}} \right],$$

were the columns are related to the competition breeding values of trees 1, 2, 4, 5, 6, 7, 8, and 9. As plant 3 is missing, its competition breeding value is not included in  $\mathbf{a}_c$ . Also the direct breeding values will not be in  $\mathbf{a}_d$ . Using [7] results in  $f_{5jR-C} = [2/(2*4 + 3)]^{1/2} = \sqrt{2}/\sqrt{11}$  for R-C trees 2, 4, 6 and 8, whereas using [6] produces  $f_{5jD} = 1/[2*4 + 3]^{1/2} = 1/\sqrt{11}$  for D trees 1, 7 and 9. The 0 in the fourth column reflects that the competition breeding value of 5 is not related to its own record. Matrix  $\mathbf{Z}_p$  is composed of the non-zero columns of  $\mathbf{Z}_c$  and has order equal to  $n \times n$ .

The covariance matrix of  $\mathbf{a}_d$  is  $\mathbf{A}\boldsymbol{\sigma}_{Ad}^2$ . The  $q \times q$  matrix  $\mathbf{A} = [\mathbf{A}_{ij}]$  has diagonal elements equal to  $1 + F_i$ , and off-diagonals equal to the additive relationships  $\mathbf{A}_{ij}$ . Also,  $\mathbf{a}_c \sim (\mathbf{0}, \mathbf{A}\boldsymbol{\sigma}_{Ac}^2)$  and  $\text{cov}(\mathbf{a}_d, \mathbf{a}_c) = \mathbf{A}\boldsymbol{\sigma}_{AdAc}$ . Now, we are able to write the total additive covariance matrix in a more compact manner as follows

$$\text{Var} \begin{bmatrix} \mathbf{a}_d \\ \mathbf{a}_c \end{bmatrix} = \begin{bmatrix} \boldsymbol{\sigma}_{Ad}^2 & \boldsymbol{\sigma}_{AdAc} \\ \boldsymbol{\sigma}_{AdAc} & \boldsymbol{\sigma}_{Ac}^2 \end{bmatrix} \otimes \mathbf{A} = \mathbf{G}_0 \otimes \mathbf{A}$$

Taking into account the random effects in model [10], the (co)variance matrix ( $\mathbf{V}$ ) of  $y$  is given by:

$$\mathbf{V} = \mathbf{Z}_d \mathbf{A} \mathbf{Z}_d' \boldsymbol{\sigma}_{Ad}^2 + (\mathbf{Z}_d \mathbf{A} \mathbf{Z}_c' + \mathbf{Z}_c \mathbf{A} \mathbf{Z}_d') \boldsymbol{\sigma}_{AdAc} + \mathbf{Z}_c \mathbf{A} \mathbf{Z}_c' \boldsymbol{\sigma}_{Ac}^2 + \mathbf{Z}_p \mathbf{Z}_p' \boldsymbol{\sigma}_p^2 + \mathbf{I}_n \boldsymbol{\sigma}_e^2 \quad [11]$$

#### Bayesian estimation of (co)variance components

As in SORIA *et al.* (1998), GWAZE and WOOLLIAMS (2001), ZENG *et al.* (2004) and CAPPALÀ and CANTET (2006), we will estimate the dispersion parameters  $\boldsymbol{\sigma}_{Ad}^2$ ,  $\boldsymbol{\sigma}_{AdAc}$ ,  $\boldsymbol{\sigma}_{Ac}^2$ ,  $\boldsymbol{\sigma}_p^2$  and  $\boldsymbol{\sigma}_e^2$  using a Bayesian approach by means of Gibbs sampling (SORENSEN and GIANOLA, 2002). Under normality of breeding values and errors, the conditional likelihood of the observed data can be written as being proportional to:

$$p(y | \boldsymbol{\beta}, \mathbf{a}_c, \mathbf{a}_d, \mathbf{p}_c, \mathbf{G}_0, \boldsymbol{\sigma}_p^2, \boldsymbol{\sigma}_e^2) \propto (\boldsymbol{\sigma}_e^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\boldsymbol{\sigma}_e^2} \mathbf{e}' \mathbf{e} \right] \quad [12]$$

where  $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_d \mathbf{a}_d - \mathbf{Z}_c \mathbf{a}_c - \mathbf{Z}_p \mathbf{p}_c$ . Conjugate prior densities are chosen for all parameters. In order to reflect a prior state of uncertainty for the fixed effects

and to obtain a proper posterior distribution (HOBERT and CASELLA, 1996), we take  $\boldsymbol{\beta} \sim N_p(\mathbf{0}, \mathbf{K})$ . Matrix  $\mathbf{K}$  is diagonal with large elements ( $k_{ii} > 10^8$ ). Also, the joint prior distribution of the direct and competition breeding values ( $\mathbf{a}_d, \mathbf{a}_c$ ) is

$$\begin{bmatrix} \mathbf{a}_d \\ \mathbf{a}_c \end{bmatrix} | \mathbf{A}, \mathbf{G}_0, \sim N \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \mathbf{G}_0 \otimes \mathbf{A} \right)$$

On defining  $\mathbf{a} = [\mathbf{a}_d', \mathbf{a}_c']'$ , the corresponding density can be written as

$$p(\mathbf{a} | \mathbf{A}, \mathbf{G}_0) \propto |\mathbf{G}_0|^{-\frac{q}{2}} \exp \left\{ -\frac{1}{2} \mathbf{a}' (\mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1}) \mathbf{a} \right\} \quad [13]$$

Let

$$\mathbf{S}_g = \begin{bmatrix} \mathbf{a}_d' \mathbf{A}^{-1} \mathbf{a}_d & \mathbf{a}_d' \mathbf{A}^{-1} \mathbf{a}_c \\ \mathbf{a}_d' \mathbf{A}^{-1} \mathbf{a}_c & \mathbf{a}_c' \mathbf{A}^{-1} \mathbf{a}_c \end{bmatrix}$$

Then

$$\begin{aligned} \mathbf{a}' (\mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1}) \mathbf{a} &= [\mathbf{a}_d \quad \mathbf{a}_c]' \begin{bmatrix} \mathbf{a}_d' \mathbf{A}^{-1} \mathbf{a}_d & \mathbf{a}_d' \mathbf{A}^{-1} \mathbf{a}_c \\ \mathbf{a}_d' \mathbf{A}^{-1} \mathbf{a}_c & \mathbf{a}_c' \mathbf{A}^{-1} \mathbf{a}_c \end{bmatrix} \begin{bmatrix} \mathbf{a}_d \\ \mathbf{a}_c \end{bmatrix} \\ &= \text{tr} \left( \mathbf{G}_0^{-1} \mathbf{S}_g \right) \end{aligned}$$

Therefore, [13] can be expressed as

$$p(\mathbf{a} | \mathbf{A}, \mathbf{G}_0) \propto |\mathbf{G}_0|^{-\frac{q}{2}} \exp \left\{ -\frac{1}{2} \text{tr} (\mathbf{G}_0^{-1} \mathbf{S}_g) \right\} \quad [14]$$

*A priori* the permanent environmental effects are distributed as  $\mathbf{p}_c \sim N_n(\mathbf{0}, \boldsymbol{\sigma}_p^2 \mathbf{I}_n)$ :

$$p(\mathbf{p}_c | \boldsymbol{\sigma}_p^2) \propto (\boldsymbol{\sigma}_p^2)^{-\frac{n}{2}} \exp \left\{ -\frac{\mathbf{p}_c' \mathbf{p}_c}{2\boldsymbol{\sigma}_p^2} \right\} \quad [15]$$

The matrix of the additive (co)variance components  $\mathbf{G}_0$  follows *a priori* an Inverted Wishart (IW) density:  $\mathbf{G}_0 \sim \text{IW}(\mathbf{G}_0^*, \nu_g)$  where  $\mathbf{G}_0^*$  is the prior covariance matrix and  $\nu_g$  are the degrees of freedom. Thus:

$$p(\mathbf{G}_0 | \mathbf{S}_g, \nu_g) \propto |\mathbf{G}_0|^{-\frac{(\nu_g+3)}{2}} \exp \left\{ -\frac{1}{2} \text{tr} (\mathbf{G}_0^* \mathbf{G}_0^{-1}) \right\} \quad [16]$$

*A priori* the permanent environmental variance has a scaled inverted  $\chi^2$  density so that:

$$p(\boldsymbol{\sigma}_p^2 | \nu_p, S_p^2) \propto (\boldsymbol{\sigma}_p^2)^{-\left(\frac{\nu_p}{2}+1\right)} \exp \left\{ -\frac{\nu_p S_p^2}{2\boldsymbol{\sigma}_p^2} \right\} \quad [17]$$

with 'hyperparameters'  $\nu_p$ , the 'degree of belief' and  $S_p^2$  the hypervariance. Finally, and following the approach of JENSEN *et al.* (1994), the residual variance is assumed to follow *a priori* a scaled inverted  $\chi^2$  with density proportional to:

$$p(\boldsymbol{\sigma}_e^2 | \nu_e, S_e^2) \propto (\boldsymbol{\sigma}_e^2)^{-\left(\frac{\nu_e}{2}+1\right)} \exp \left\{ -\frac{\nu_e S_e^2}{2\boldsymbol{\sigma}_e^2} \right\} \quad [18]$$

where  $\nu_e$  and  $S_e^2$  are the 'hyperparameters'.

Multiplying [12] with [14], [15], [16], [17], and [18], produces the joint posterior density for all parameters, and this is proportional to

$$\begin{aligned}
& p(\boldsymbol{\beta}, \mathbf{a}_d, \mathbf{a}_c, \mathbf{p}_c, \mathbf{G}_0, \sigma_p^2, \sigma_e^2 | \mathbf{y}) \propto \\
& p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{a}_c, \mathbf{a}_d, \mathbf{p}_c, \mathbf{G}_0, \sigma_p^2, \sigma_e^2) p(\mathbf{a}_c, \mathbf{a}_d | \mathbf{A}, \mathbf{G}_0) \\
& p(\mathbf{p}_c | \sigma_p^2) p(\mathbf{G}_0 | \mathbf{S}_g, \nu_g) p(\sigma_p^2 | \mathbf{S}_p^2, \nu_p) \\
& p(\sigma_e^2 | \mathbf{S}_e^2, \nu_e)
\end{aligned} \quad [19]$$

SORENSEN and GIANOLA (2002, page 575) gave the posterior distribution for all parameters of maternal models. On expanding their model to include permanent competition effects, the posterior conditional density for the Gibbs sampling of  $\boldsymbol{\beta}$ ,  $\mathbf{a}_d$ ,  $\mathbf{a}_c$  and  $\mathbf{p}_c$  is equal to

$$\begin{aligned}
& \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{a}_d \\ \mathbf{a}_c \\ \mathbf{p}_c \end{bmatrix} | \mathbf{y}, \mathbf{G}_0, \sigma_p^2, \sigma_e^2 \sim N \\
& \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}}_d \\ \hat{\mathbf{a}}_c \\ \hat{\mathbf{p}}_c \end{pmatrix} \left[ \begin{array}{cccc} X'X + K^{-1} & X'Z_d' & X'Z_c' & X'Z_p' \\ Z_d'X & Z_d'Z_d + k_{11}A^{-1} & Z_d'Z_c + k_{12}A^{-1} & Z_d'Z_p \\ Z_c'X & Z_c'Z_d + k_{21}A^{-1} & Z_c'Z_c + k_{22}A^{-1} & Z_c'Z_p \\ Z_p'X & Z_p'Z_d & Z_p'Z_c & Z_p'Z_p + I_{\left(\frac{\nu_p}{\sigma_p^2}\right)} \end{array} \right]^{-1}
\end{aligned} \quad [20]$$

where  $\begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} = \mathbf{G}_0^{-1} \sigma_e^2$ , and  $\hat{\boldsymbol{\beta}}$ ,  $\hat{\mathbf{a}}_d$ ,  $\hat{\mathbf{a}}_c$  and  $\hat{\mathbf{p}}_c$

are the solutions of the following system of equations

$$\begin{bmatrix} X'R^{-1}X + K^{-1} & X'R^{-1}Z_d' & X'R^{-1}Z_c' & X'R^{-1}Z_p' \\ Z_d'R^{-1}X & Z_d'R^{-1}Z_d + k_{11}A^{-1} & Z_d'R^{-1}Z_c + k_{12}A^{-1} & Z_d'R^{-1}Z_p \\ Z_c'R^{-1}X & Z_c'R^{-1}Z_d + k_{21}A^{-1} & Z_c'R^{-1}Z_c + k_{22}A^{-1} & Z_c'R^{-1}Z_p \\ Z_p'R^{-1}X & Z_p'R^{-1}Z_d & Z_p'R^{-1}Z_c & Z_p'R^{-1}Z_p + I_{\left(\frac{\nu_p}{\sigma_p^2}\right)} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}}_d \\ \hat{\mathbf{a}}_c \\ \hat{\mathbf{p}}_c \end{bmatrix} = \begin{bmatrix} X'R^{-1}\mathbf{y} \\ Z_d'R^{-1}\mathbf{y} \\ Z_c'R^{-1}\mathbf{y} \\ Z_p'R^{-1}\mathbf{y} \end{bmatrix}$$

Expression [20] may suggest that sampling of  $\boldsymbol{\beta}$ ,  $\mathbf{a}_d$ ,  $\mathbf{a}_c$  and  $\mathbf{p}_c$  is in block. However, it is simpler to sample the elements of those vectors individually, as discussed by SORENSEN and GIANOLA (2002, page 566, expressions (13.11) and (13.12)), which was the way it was done in the current research.

Collecting the second and third terms in the right of [19], the full conditional posterior distribution of  $\mathbf{G}_0$  is equal to

$$\begin{aligned}
& p(\mathbf{G}_0 | \boldsymbol{\beta}, \mathbf{a}_c, \mathbf{a}_d, \mathbf{p}_c, \sigma_p^2, \sigma_e^2, \mathbf{y}) \propto \\
& |\mathbf{G}_0|^{-\frac{(\nu_g + q + 3)}{2}} \exp\left\{-\frac{1}{2} \text{tr} \mathbf{G}_0^{-1} (\mathbf{S}_g + \mathbf{G}_0^*)\right\}
\end{aligned} \quad [21]$$

Expression [21] is the kernel of a  $2 \times 2$  scaled inverted Wishart distribution, with degrees of freedom equal to  $(\nu_g + q + 3)$  and scale matrix  $\mathbf{S}_g + \mathbf{G}_0^*$ .

For the permanent error variance, the full conditional posterior distribution is

$$\begin{aligned}
& p(\sigma_p^2 | \boldsymbol{\beta}, \mathbf{a}_d, \mathbf{a}_c, \mathbf{p}_c, \mathbf{G}_0, \sigma_e^2, \mathbf{y}) \propto \\
& (\sigma_p^2)^{-\left(\frac{n + \nu_p + 2}{2} + 1\right)} \exp\left\{-\frac{\tilde{\nu}_p \tilde{\delta}_p^2}{2\sigma_p^2}\right\}
\end{aligned} \quad [22]$$

which is a scaled inverted  $\chi^2$  density with  $\tilde{\nu}_p = n + \nu_p$  degrees of freedom and scale parameter

$$\tilde{\delta}_p^2 = \frac{(\mathbf{p}_c' \mathbf{p}_c + \nu_p \delta_p^2)}{(\nu_p + n)}.$$

Finally, the full conditional posterior density of the residual variance is proportional to

$$\begin{aligned}
& p(\sigma_e^2 | \boldsymbol{\beta}, \mathbf{a}_c, \mathbf{a}_d, \mathbf{p}_c, \mathbf{G}_0, \sigma_p^2, \mathbf{y}) \propto \\
& (\sigma_e^2)^{-\left(\frac{n + \nu_e + 2}{2} + 1\right)} \exp\left\{-\frac{\tilde{\nu}_e \tilde{\delta}_e^2}{2\sigma_e^2}\right\}
\end{aligned} \quad [23]$$

which is a scaled inverted  $\chi^2$  density with  $\tilde{\nu}_e = n + \nu_e$  degrees of freedom and scale parameter

$$\tilde{\delta}_e^2 = \frac{(\mathbf{e}' \mathbf{e} + \nu_e \delta_e^2)}{(\nu_e + n)}.$$

At each iteration, the Gibbs algorithm proceeds by first sampling  $\boldsymbol{\beta}$ ,  $\mathbf{a}_d$ ,  $\mathbf{a}_c$  and  $\mathbf{p}_c$  from [20], then  $\sigma_e^2$  from [23],  $\sigma_p^2$  from [22], and finally  $\sigma_{Ad}^2$ ,  $\sigma_{AdAc}$  and  $\sigma_{Ac}^2$  from [21]. A program was written in FORTRAN to perform all these calculations with the data and model described below.

## An application to Loblolly pines

### Data

An additive individual tree mixed model with direct and competition breeding values was applied to a progeny data set derived from 20 open-pollinated families of Loblolly pine (*Pinus taeda* L.), originated from Marion (Florida, USA) and belonging to CIEF (Forestry Research and Experimentation Centre). Five lots of commercial seeds were used as control populations. The trait analyzed was diameter at breast height (1.3 m, DBH) measured at age 13 from 932 trees. The trial site was located in Villa Olivari (lat. 27° 36' S long. 56° 55' W), northern Corrientes province Argentina, where soils are deep, sandy, yellowish and quartzous. Families were arranged in randomized complete blocks, with eight replicates of 5 trees in line per plot, and the spacing was 3.5 m x 3.5 m. The data available are summarized in *Table 1*.

*Table 1.* – Means and number of records in the Loblolly pine data set.

Trees with records ( <i>n</i> )	932
Families	20
Individuals in the pedigree ( <i>q</i> )	957
Mean diameter (DBH, cm) (SD)	27.21 (4.56)

SD = Standard deviation.

### Model and posterior inference

Besides  $\mathbf{a}_d$ ,  $\mathbf{a}_c$  and  $\mathbf{p}_c$ , the individual tree model included a vector  $\boldsymbol{\beta}$  with 8 block effects. The order of  $\mathbf{Z}_d$  and  $\mathbf{Z}_c$  was 932 x 957, and of  $\mathbf{Z}_p$  was 932 x 932. Whenever a tree had all 8 neighbor competitors,  $n_{R-C} = n_D = 4$ . On replacing with these values into [7] and [6] produces the non-zero elements of  $\mathbf{Z}_c$ :  $f_{ijR-C} = [2/(2*4 + 4)]^{1/2} = 1/\sqrt{6}$  and  $f_{ijD} = 1/[2*4 + 4]^{1/2} = 1/\sqrt{12}$ .

As there no other estimates of  $\sigma_{AdAc}$  and  $\sigma_{Ac}^2$ , several Markov chains with different prior values of  $\sigma_{AdAc}$  (+, 0, and -) and of  $\sigma_{Ac}^2$  (high and low relative to  $\sigma_{AdAc}^2$ ) were run. Results were quite similar so the chain with the best convergence properties was used to estimate the (co)variance components. The prior variance for  $\sigma_p^2$  ( $\delta_p^2$ ) was chosen to be equal to the prior value for  $\sigma_{Ac}^2$ . A single Gibbs chain of 1010 000 samples each were drawn as discussed above, and the first 10000 iterates were discarded due to *burn-in*. The autocorrelations were calculated with “*Bayesian Output Analysis*” (BOA version 1.0.1, Smith 2003) for all lags from 1 to 100. To account for the impact of autocorrelations in the chain on measures of variability, posterior standard errors of each parameter were corrected for an ‘effective sample size’ (ESS, Neal in KASS *et al.*, 1998), which was calculated as:

$$ESS = \frac{1000000}{1 + 2 \sum_{i=1}^{100} \rho(i)}$$

where  $\rho(i)$  is the autocorrelation measured at lag  $i$ . The marginal posterior densities for all parameters were estimated using the Gaussian kernel method (SILVERMAN, 1986; chapter 2):

$$f(\theta) = \frac{1}{10000 h} \sum_{i=1}^{10000} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{z - \theta_i}{h} \right)^2 \right] \quad [24]$$

where  $f(\theta)$  is the estimated posterior density,  $\theta_i$  ( $i = 1, \dots, 10000$ ) is a sampled value and  $h$  is the window width estimated by cross-validation. The basic idea of this procedure is to withdraw one observation at a time and estimate the density. After repeating the procedure  $n$  times the average of the logs of the estimated densities is maximized with respect to  $h$ . Further details can be consulted in SILVERMAN (1986, section 3.4.4). The procedure is implemented in the function *density* within the free-software *R* (<http://www.r-project.org/>). Mean, mode,

median, standard deviation (SD), and 95% high posterior density interval (95% HPD), were obtained with BOA for all parameters from the individual marginal posteriors, under *R*.

### Results

Posterior statistics for  $\sigma_{Ad}^2$ ,  $\sigma_{Ac}^2$ ,  $r_{AdAc}$ ,  $\sigma_p^2$  and  $\sigma_e^2$  are shown in *Table 2*. For all parameters posterior means and medians were quite similar, whereas the modes were somewhat smaller. Therefore, the marginal posterior distributions were slightly right skewed (*Figure 2*). The marginal posterior means and SDs of  $\sigma_{Ad}^2$  and  $\sigma_{Ac}^2$  were respectively equal to 12.553 and 1.259, and 1.447 and 0.259. The marginal posterior mean of  $r_{AdAc}$  was moderate to large and negative -0.788 and the SD was 0.056. The posterior means of  $\sigma_p^2$  and  $\sigma_e^2$  were equal to 1.186 and 5.819, respectively, and their SDs were 0.289 and 1.070. None of the 95% HPD for  $\sigma_{Ad}^2$ ,  $\sigma_{Ac}^2$ ,  $r_{AdAc}$ ,  $\sigma_p^2$  or  $\sigma_e^2$  included 0, which suggests that these parameters are different from zero.

### Discussion

It has been observed that competition among trees may bias the estimated breeding value of a plant from those of its competitors (MAGNUSSEN, 1993; FOSTER *et al.*, 1998; RADTKE *et al.*, 2003, among others). In the current research, we presented an individual tree mixed model that allows disentangling breeding values for direct and competition effects and estimating their variances plus the covariance between both effects, as well as the variance of permanent environmental competition effects. Estimation of the dispersion parameters was accomplished using a Bayesian method with the Gibbs sampler originally proposed by JENSEN *et al.* (1994) for maternal effects in animals. For direct and competition effects, VAN VLECK and CASSADY (2005), ARANGO *et al.* (2005) and MUIR (2005) estimated the (co)variance components by REML. We did not find problems of convergence and sensitivity to starting values, as reported by VAN VLECK and CASSADY (2005) and ARANGO *et al.* (2005) when estimating the competitive (co)variance components. It is unlikely that the difference in performance is due to the use of a different method of estimation, but to different amount of information on competitive effects for forest trees compared with animals. In trees, each individual may be competing with 8 others at different

*Table 2.* – Posterior statistics for direct additive variance  $\sigma_{Ad}^2$ , competition additive variance  $\sigma_{Ac}^2$ , direct and competition additive correlation  $r_{AdAc}$ , permanent environmental variance  $\sigma_p^2$  and error variance  $\sigma_e^2$ .

	Mean	Median	Mode	SD	95% HPD	ESS
$\sigma_{Ad}^2$	12.553	12.531	10.075	1.447	10.179 – 14.971	21743
$\sigma_{Ac}^2$	1.259	1.237	1.012	0.259	0.876 – 1.726	13314
$r_{AdAc}$	-0.788	-0.795	-0.843	0.056	-0.868 – -0.688	16088
$\sigma_p^2$	1.186	1.152	0.869	0.289	0.779 – 1.718	18979
$\sigma_e^2$	5.819	5.780	4.816	1.07	4.163 – 7.685	18553

SD = standard deviation, HPD = high posterior density interval, ESS = effective sample size.

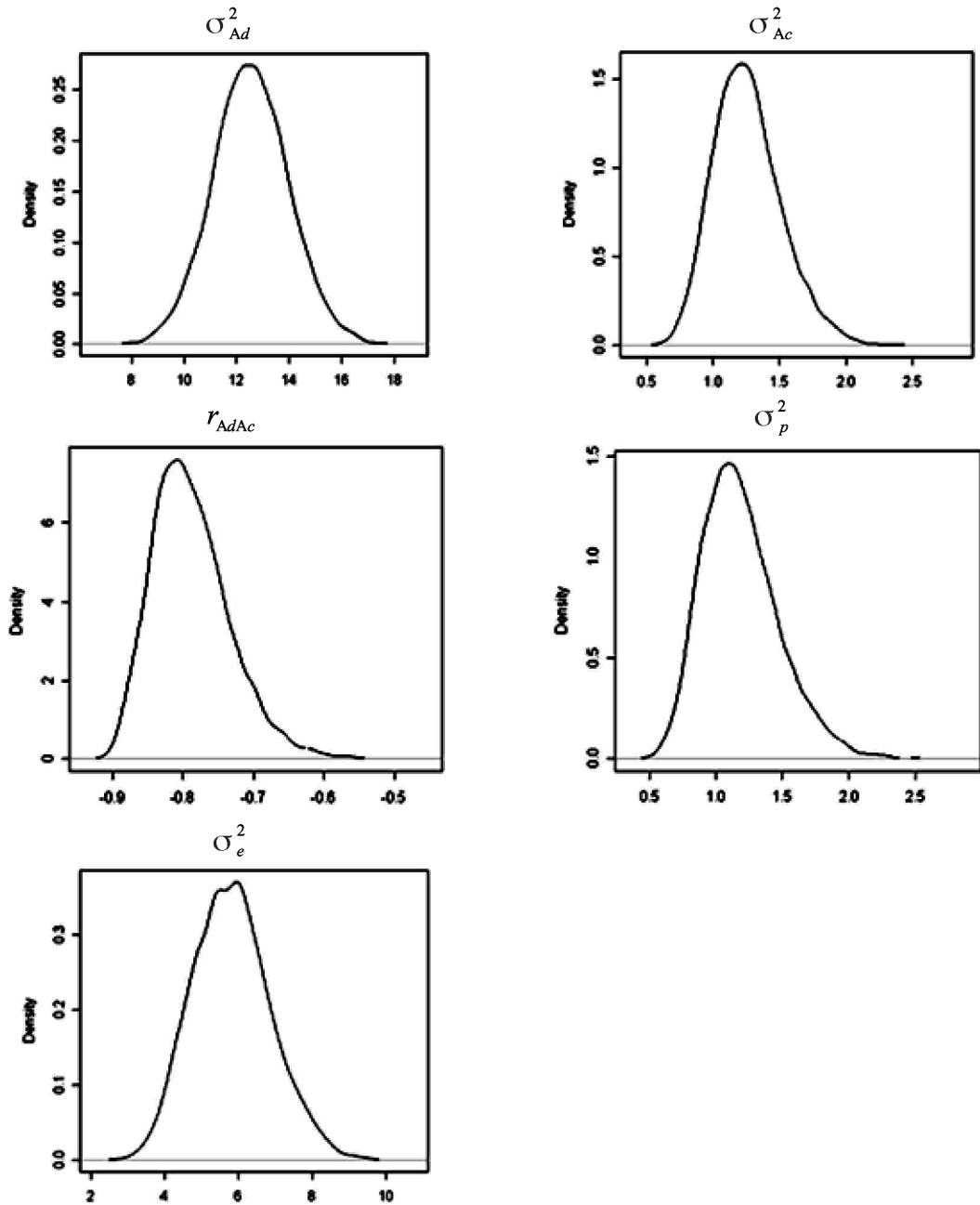


Figure 2. – Marginal posterior distributions of direct additive variance  $\sigma_{Ad}^2$ , competition additive variance  $\sigma_{Ac}^2$ , direct and competition additive correlation  $r_{AdAc}$ , permanent environmental variance  $\sigma_p^2$  and error variance  $\sigma_e^2$ .

intensities (see [6] and [7]), as compared with animal breeding data for individuals managed in common pens where all ICs are equal. This informativeness of the IC through matrix  $Z_c$  allowed estimating additive effects of competition and permanent environmental competitive effects.

Posterior means of the (co)variance components for a model without permanent environmental competitive effects were  $\hat{\sigma}_{Ad}^2 = 13.527$ ,  $\hat{\sigma}_{Ac}^2 = 1.488$ ,  $\hat{r}_{AdAc} = -0.659$ , and  $\hat{\sigma}_e^2 = 5.417$ . Notice that the estimate of  $\sigma_{Ad}^2$  was larger (13.527 vs. 12.553) and the estimate of  $\sigma_e^2$  was smaller (5.417 vs. 5.819) than when permanent environmental competitive effects were fitted. This may suggest that leaving those environmental effects out of the model may bias the predictions of direct breeding val-

ues. A possible evidence of this bias is that the predictions of permanent environmental competition effects of trees having less than 8 competitors (for example, those plants nearby a site without a tree) were almost a quarter of a standard deviation larger than the same effects but for trees having all 8 competitors. Thus, the presence of dead trees promotes a less stressful environment for the surrounding plants that may bias the prediction of direct breeding values, if permanent environmental effects are not accounted for in the model.

As expression [9] shows, ignoring the number of competitors, the additive relationships between the individual with the competitors, and the relationships among the competitors themselves results in biased estimates of the dispersion parameters. ARANGO *et al.* (2005)

accounted for different number of competitors by including the covariable  $1/n$  or  $1/\sqrt{n}$ , whereas VAN VLECK and CASSADY (2005) did not account for a variable number of competitors. As progeny tests in forest trees last several years, mortality or tree stand failure are quite common, and this results in variable number of competitors. We accounted for unequal number of neighbors ( $m_i < 8$ ) for either mortality or border location by adjusting  $f_{ijk}$  so that always

$$\sum_{j=1}^m f_{ijk}^2 = n_{R-C} f_{ijR-C}^2 + n_D f_{ijD}^2 = 1.$$

We also accounted for the additive relationships between any tree and its competitors and among the competitors themselves. VAN VLECK and CASSADY (2005) did not account for additive relationships between individuals in the management unit, which bias the estimation of  $\sigma_{Ac}^2$ .

An individual tree model that includes direct and competition breeding values, taking into account a variable number of competitors and the relationships among all trees, allows different selection goals and schemes which capitalizes on (or attenuates) the impact of competition effects. Compared to a model with direct breeding values, the application of the model discussed here to forest breeding data requires only the positions (row and column) of all the trees in the trial. The estimated value of  $\sigma_{Ad}^2$  was almost ten times higher than the magnitude of  $\sigma_{Ac}^2$ , whereas the magnitude of the genetic correlation between direct and competitive effects was sizable ( $-0.788$ ). Therefore, selecting for high direct effects and low competitive effects to increase the yield per unit area is facilitated for a relatively high negative correlation between both types of effects.

Of further note is the fact that ignoring competitive effects would result in biased estimates of the additive variance ( $\sigma_{Ad}^2$ ). To specify the covariance between relatives in the additive model two individuals ( $\mathbf{x}$  and  $\mathbf{y}$ ) are required. If indirect effects are involved more individuals are needed. For example, the covariance between relatives with maternal effects requires four animals ( $\mathbf{x}$ ,  $\mathbf{y}$ , and respective dams  $w$  and  $z$ ), and up to 18 individuals have to be considered in the additive model with competition effects in [8]:  $\mathbf{x}$  and its competitors  $w_1, \dots, w_8$ , and  $\mathbf{y}$  and its competitors  $z_1, \dots, z_8$ . On applying the covariance operator on the additive effects of model [8] results in the following expression for the genetic covariance between the records of  $\mathbf{x}$  and  $\mathbf{y}$

$$\begin{aligned} & \text{cov} \left( \mathbf{a}_{dx} + \sum_{i=1}^8 f_{xw_i} \mathbf{a}_{cw_i}, \mathbf{a}_{dy} + \sum_{j=1}^8 f_{yz_j} \mathbf{a}_{cz_j} \right) \\ &= \text{cov}(\mathbf{a}_{dx}, \mathbf{a}_{dy}) + \text{cov} \left( \sum_{i=1}^8 f_{xw_i} \mathbf{a}_{cw_i}, \mathbf{a}_{dy} \right) \\ &+ \text{cov} \left( \mathbf{a}_{dx}, \sum_{j=1}^8 f_{yz_j} \mathbf{a}_{cz_j} \right) + \text{cov} \left( \sum_{i=1}^8 f_{xw_i} \mathbf{a}_{cw_i}, \sum_{j=1}^8 f_{yz_j} \mathbf{a}_{cz_j} \right) \\ &= A_{xy} \sigma_{Ad}^2 + \left( \sum_{i=1}^8 f_{xw_i} A_{yw_i} + \sum_{j=1}^8 f_{yz_j} A_{xz_j} \right) \sigma_{AdAc} \\ &+ \left( \sum_{j=1}^8 f_{xw_i} f_{yz_j} A_{w_i z_j} \right) \sigma_{Ac}^2 \end{aligned}$$

Expression [25] is of general form and parameters are associated to additive relationships between individuals in the following way:  $\sigma_{Ad}^2$  with  $\mathbf{x}$  and  $\mathbf{y}$ ; the covariance  $\sigma_{AdAc}$  with the relationships between the individual  $\mathbf{x}$  ( $\mathbf{y}$ ) and competitors  $w_i$  ( $z_j$ ), and  $\sigma_{Ac}^2$  is associated with the relationships among the members of both groups of competitors. To get further insight into the model with competition consider the situation where  $\mathbf{x}$  and  $\mathbf{y}$  are direct competitors. Schematically

$$\begin{array}{cccc} \bullet w_1 & \bullet w_2 = z_1 & \bullet w_3 = z_2 & \bullet z_3 \\ \bullet w_4 & \bullet \mathbf{x} = z_4 & \bullet \mathbf{y} = w_5 & \bullet z_5 \\ \bullet w_6 & \bullet w_7 = z_6 & \bullet w_8 = z_7 & \bullet z_8 \end{array}$$

Trees  $\mathbf{x}$  and  $\mathbf{y}$  are in the center, surrounded by their respective competitors  $w_1, \dots, w_8, z_1, \dots, z_7$ , and  $z_8$ . Notice that 6 individuals are common competitors, implying that  $w_2 = z_1$ ,  $x = z_4$ , and so on. Expression [25] is now equal to

$$\begin{aligned} & \text{cov} \left( \mathbf{a}_{dx} + \sum_{i=1}^8 f_{xw_i} \mathbf{a}_{cw_i}, \mathbf{a}_{dy} + \sum_{j=1}^8 f_{yz_j} \mathbf{a}_{cz_j} \right) = A_{xy} \sigma_{Ad}^2 \\ &+ \left[ f_{xy} (1 + F_y) + f_{yx} (1 + F_x) + \sum_{i=1}^7 f_{yw_i} A_{yw_i} + \sum_{j=1}^7 f_{xz_j} A_{xz_j} \right] \sigma_{AdAc} \\ &+ \left( \sum_{w_i=z_j} f_{xw_i} f_{yz_j} (1 + F_{w_i}) + \sum_{w_i \neq z_j} f_{xw_i} f_{yz_j} A_{w_i z_j} \right) \sigma_{Ac}^2 \end{aligned}$$

Alternatively, if  $\mathbf{x}$  and  $\mathbf{y}$  are distant from each other the scheme is as follows

$$\begin{array}{cccccc} \bullet w_1 & \bullet w_2 & \bullet w_3 & \cdots & \bullet z_1 & \bullet z_2 & \bullet z_3 \\ \bullet w_4 & \bullet \mathbf{x} & \bullet w_5 & \cdots & \bullet z_4 & \bullet \mathbf{y} & \bullet z_5 \\ \bullet w_6 & \bullet w_7 & \bullet w_8 & \cdots & \bullet z_6 & \bullet z_7 & \bullet z_8 \end{array}$$

The dots (...) in the scheme stress the fact that trees are separated by at least a column, or they may be even positioned in totally different rows, or in any other position within the trial. The covariance in [25] reduces to

$$\begin{aligned} & \text{cov} \left( \mathbf{a}_{dx} + \sum_{i=1}^8 f_{xw_i} \mathbf{a}_{cw_i}, \mathbf{a}_{dy} + \sum_{j=1}^8 f_{yz_j} \mathbf{a}_{cz_j} \right) \\ &= A_{xy} \sigma_{Ad}^2 + \left[ \sum_{i=1}^8 f_{xw_i} A_{yw_i} + \sum_{j=1}^8 f_{yz_j} A_{xz_j} \right] \sigma_{AdAc} \\ &+ \left( \sum_{i=1}^8 \sum_{j=1}^8 f_{xw_i} f_{yz_j} A_{w_i z_j} \right) \sigma_{Ac}^2 \end{aligned}$$

Formula [26] may be applied to row-plot designs where related individuals are planted nearby, an example of which is the structure in the data set analyzed in the current research. On the other hand, expression [27] is associated with single-tree plot designs where individuals of the same family are positioned distant to each other. Consider the estimation of heritability in a single-tree plot design using half-sib families. The additive variance will be estimated as 4 times the covariance between half-sibs, i.e. the variance between mothers. Trees  $\mathbf{x}$  and  $\mathbf{y}$  have a relationship equal to  $A_{xy} = 1/4$  and are distant to each other, and this defines the first term in [27]. In order for the covariance between half-sibs to be an unbiased estimator of the additive variance, all terms related with  $\sigma_{AdAc}$  and  $\sigma_{Ac}^2$  should be equal to



zero. The second term will be null as long as competitors of any individual ( $\mathbf{x}$ , say) are unrelated with a distant half-sib ( $\mathbf{y}$ ). However, it is unlikely that all competitors of any plant  $\mathbf{x}$  are unrelated to the competitors of all possible  $\mathbf{y}$ , and this will introduce bias in the estimation of additive variance. All sources of variation and covariation that are not accounted for the covariance between half-sibs will fall into the error term. Of particular interest is the covariance between any two unrelated and distant individuals  $\mathbf{x}$  and  $\mathbf{y}$ . It is most likely that the second and third terms in [27] are not zero as some  $\mathbf{x}$ 's are related to any of the  $z$ -competitors, or some  $\mathbf{y}$ 's are related to any of the  $w$ -competitors, or some  $w$ -competitors are related to any of the  $z$ -competitors. Any of these covariances will go into the error and will bias the error variance upwards. Therefore, even if the bias on the additive variance is small, the estimate of heritability will be affected as the error variance is most likely biased. However, the direction and magnitude of the bias will depend on the sign and the magnitude of  $\sigma_{AdAc}$  as compared with the magnitude of  $\sigma_{Ac}^2$ . In comparison, in data structures where related individuals are next to each other and competing such as in the row-plot field design, the covariance between half-sibs will be more affected than in the case of the single-tree plot. This is due to the fact that the third term in [26] will not be null as the competitors of  $\mathbf{x}$  are related to the competitors of  $\mathbf{y}$ , being many times the same individuals ( $\mathbf{x} = z_4$ ,  $\mathbf{y} = w_5$ ). Also the second term in [26] will not be zero as when looking at the covariance between  $\mathbf{x}$  and the competitors of  $\mathbf{y}$  (or  $\mathbf{y}$  and the competitors of  $\mathbf{x}$ ),  $\mathbf{x}$  is also a competitor of  $\mathbf{y}$  and  $\mathbf{y}$  a competitor of  $\mathbf{x}$ . Hence,  $f_{xy}$  and  $f_{yx}$  are not zero. For the error variance we will look at the covariance between unrelated individuals ( $\mathbf{A}_{xy} = 0$ ) that are either competing or distant. As in the case of single-tree plot, in the row-plot design the error variance will also be affected, as the second and third terms in [26] will not be zero: any pair of unrelated  $\mathbf{x}$  and  $\mathbf{y}$  will have relatives competing to the other individual in the pair, i.e. there will be  $z$ -competitors related to  $\mathbf{x}$ ,  $w$ -competitors related to  $\mathbf{y}$ , and  $z$ -competitors related to  $w$ -competitors. The size of the bias when estimating the additive and error variances will depend on the sign and the magnitude of  $\sigma_{AdAc}$  as compared with the magnitude of  $\sigma_{Ac}^2$ . When looking into the data analysis, estimates of the dispersion parameters in the model with competition effects were  $\hat{\sigma}_{Ad}^2 = 12.553$ ,  $\hat{\sigma}_{AdAc} = -3.126$ ,  $\hat{\sigma}_{Ac}^2 = 1.259$ ,  $\hat{\sigma}_p^2 = 1.186$ , and  $\hat{\sigma}_e^2 = 5.819$ . On the other hand, the estimated variances in the model with direct effects only (no competition) were  $\hat{\sigma}_{Ad}^2 = 10.644$ , and  $\hat{\sigma}_e^2 = 9.257$ . The difference between the estimates of  $\sigma_{Ad}^2$  can be explained by the negative sign and the absolute value (3.126) of  $\sigma_{AdAc}$  relative to the small value of  $\hat{\sigma}_{Ac}^2$ , which gives more weight to the 2<sup>nd</sup> than the 3<sup>rd</sup> term in [26]. As a consequence,  $\hat{\sigma}_{Ad}^2$  in the model with competition was higher than in the model excluding competitive effects. The value of  $\hat{\sigma}_e^2$  in the model were competition effects were absent was larger than in the model with competition. This is probably due to the larger number of covariances related to 3<sup>rd</sup> as compared to 2<sup>nd</sup> term in [26]. A quick look at this formula shows that, whereas a maximum of 16 elements are related to  $\sigma_{AdAc}$ , up to 64 elements are associated with  $\sigma_{Ac}^2$ . Therefore, even though

$\hat{\sigma}_{AdAc}$  was negative and larger in absolute value than  $\hat{\sigma}_{Ac}^2$ , the higher number of elements in the third term gave more weight to  $\sigma_{Ac}^2$  than to  $\sigma_{AdAc}$ .

In the current research, a Bayesian procedure coupled with a Markov Chain Monte Carlo technique (Gibbs sampling), has been used to estimate the (co)variance components. An alternative approach for estimating dispersion parameters is the use of REML. Comparison of frequentist and Bayesian estimators is difficult due to the fact that central issues related to the comparison of frequentist estimators (such as repeated sampling or bias) do not have the same meaning in the Bayesian school (GELMAN *et al.*, 1995, page 108). When comparing REML *vs* Bayes Gibbs sampling for estimating (co)variance components in mixed models by stochastic simulation, both methods were seemingly unbiased (VAN TASSELL *et al.*, 1995; DUANGJINDA *et al.*, 2001). Models compared included different genetic or environmental effects, and different data based selection policies were performed. For the sake of completeness, we obtained REML estimates of the dispersion parameters using the EM algorithm (DEMPSTER *et al.*, 1977), using formulae described in Appendix B. The estimated (co)variance components were  $\hat{\sigma}_{Ad}^2 = 13.889$ ,  $\hat{\sigma}_{AdAc} = -3.335$ ,  $\hat{\sigma}_{Ac}^2 = 1.521$ ,  $\hat{\sigma}_p^2 = 1.150$ , and  $\hat{\sigma}_e^2 = 3.997$ . On the other hand, the estimated REML-EM variances in the model with direct effects only (no competition effects) were  $\hat{\sigma}_{Ad}^2 = 7.572$ , and  $\hat{\sigma}_e^2 = 12.496$ . Although there were some differences, both sets of estimates of the (co)variance components for the model with competition effects obtained were similar. For the model without competition effects, the REML-EM estimate of the additive variance was smaller and the estimated error variance was larger than corresponding Bayesian estimates. An analytical comparison can be established for the prediction of breeding values using either REML + BLUP or Bayesian posterior means. ARORA and LAHIRI (1997, theorem 1, page 1056) showed that BLUP prediction of random effects from a general mixed model with estimated variance components (for example, those resulting from the use of REML) have the same expected value as the Bayesian posterior means from the same model. However, the mean square error of the Bayesian posterior mean is always smaller than the one obtained from the REML + BLUP predictions. Thus, one may expect the predictions to be similar on average, but the mean square of the Bayes posterior means will be smaller than their BLUP counterparts. The exclusion of competition effects introduces bias in the prediction of breeding values for direct effects, either in an individual tree model or in a parental or family model. If the individual tree model with competition effects is difficult to fit, the same predictions of breeding values can be calculated by means of an equivalent model (HENDERSON, 1977) that has a reduced number of equations: the number of fixed effects plus the number of parent trees. This is the topic of a future publication.

Another subject for research in the future is the search for optimal experimental designs to estimate direct and competition (co)variance components, as for example the use of single-tree plot *vs* lineal or squared plots. In all cases the additive relationships between competing individuals should be of major concern.

## Acknowledgments

This research was supported by grants of Agencia Nacional de Ciencia y Tecnología (FONCyT PICT 08-09502) and Universidad de Buenos Aires (UBACyT G018, 2004-2007), of Argentina. The authors would like to thank to Forestry Research and Experimentation Centre (CIEF, Buenos Aires, Argentina) for kindly providing the *Pinus taeda* L. data set used for the study, and to two anonymous reviewers for constructive comments on an earlier version of this manuscript.

## References

- ARANGO, J., I. MISZTAL, S. TSURUTA, M. CULBERTSON and W. HERRING (2005): Estimation of variance components including competitive effects of Large White growing gilts. *J. Anim. Sci.* **83**: 1241–1246.
- ARORA, V. and P. LAHIRI (1997): On the superiority of the Bayesian method over the BLUP in small area estimation problems. *Statistica Sinica* **7**: 1053–1063.
- BORRALHO, N. M. G. (1995): The impact of individual tree mixed models (BLUP) in tree breeding strategies. In *Proceedings of the CRC-IUFRO Conference: Eucalypts plantations: Improving fibre yield and quality*. 12–24 February. 1995, Hobart, Tasmania, Australia. Edited by B. M. POTTS, N. M. G. BORRALHO, J. B. REID, R. N.
- CANTET, R. J. C., D. GIANOLA, I. MISZTAL, R. L. FERNANDO (1993): Estimates of dispersion parameters and of genetic and environmental trends for weaning weight in Angus cattle using a maternal animal model with genetic grouping. *Livest. Prod. Sci.* **34**: 203–212.
- CAPPA, E. P. and R. J. C. CANTET (2006): Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs. *Can. J. For. Res.* **36**: 1276–1285.
- CASELLA, G. and E. I. GEORGE (1992): Explaining the Gibbs Sampler. *Am. Stat.* **46**: 167–174.
- DEMPSTER, A. P., N. M. LAIRD and D. B. RUBIN (1977): Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society* **39**: 1–38.
- DUANGJINDA, M., I. MISZTAL, J. K. BERTRAND, S. TSURUTA (2001): The empirical bias of estimates by restricted maximum likelihood, Bayesian method, and method R under selection for additive, maternal, and dominance models. *J. Anim. Sci.* **79**: 2991–2996.
- FOSTER, G. S., R. J. ROUSSEAU and W. L. NANCE (1998): Eastern cottonwood clonal mixing study: intergenotypic competition effects. *Forest Ecology and Management*. **112**: 9–22.
- GELMAN, A., J. B. CARLIN, H. S. STERN and D. B. RUBIN (1995): *Bayesian data analysis*. Chapman and Hall, New York, USA.
- GRIFFING, B. (1967): Selection in reference to biological groups. I. Individual and group selection applied to populations of unordered groups. *Aust. J. Biol. Sci.* **20**: 127–139.
- GRIFFING, B. (1968a): Selection in reference to biological groups. II. Consequences of selection in groups of one size when evaluated in groups of a different size. *Aust. J. Biol. Sci.* **21**: 1163–1170.
- GRIFFING, B. (1968b): Selection in reference to biological groups. III. Generalized results of individual and group selection in terms of parent-offspring covariances. *Aust. J. Biol. Sci.* **21**: 1171–1178.
- GWAZE, D. P. and J. A. WOOLLIAMS (2001): Making decisions about the optimal selection environment using Gibbs sampling. *Theor. Appl. Genet.* **103**: 63–69.
- HARVILLE, D. A. (1974): Bayesian inference for variance components using only error contrasts. *Biometrika* **61**: 383–384.
- HENDERSON, C. R. (1977): Best linear unbiased prediction of breeding values not in the model for records. *J. Dairy Sci.* **60**: 783–787.
- HINSON, K. and W. D. HANSON (1962): Competition studies in soybeans. *Crop. Sci.* **2**: 117–123.
- HOBERT, J. P. and G. CASELLA (1996): The effects of improper priors on Gibbs sampling in hierarchical linear models. *J. Amer. Statist.* **91**: 1461–1473.
- JENSEN, J., C. S. WANG, D. A. SORENSEN and D. GIANOLA (1994): Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. *Acta Agric. Scand.* **44**: 193–201.
- KASS, R. E., B. P. CARLIN, A. GELMAN and R. M. NEAL (1998): Markov chain Monte Carlo in practice: a roundtable discussion. *Amer. Stat.* **52**: 93–100.
- KEMPTHORNE, O. (1969): *An Introduction to Genetic Statistics*. Iowa State Univ. Press, Ames Iowa.
- LYNCH, M. and B. WALSH (1998): *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland MA.
- MAGNUSSEN, S. (1989): Effects and adjustments of competition bias in progeny trials with single-tree plots. *Forest Science* **35**(2): 532–547.
- MAGNUSSEN, S. (1993): Bias in genetic variance estimates due to spatial autocorrelation. *Theoretical and Applied Genetics* **86**: 349–355.
- MUIR, W. M. and A. SCHINCKEL (2002): Incorporation of competitive effects in breeding programs to improve productivity and animal well being. CD-ROM Communication No. 14-07 in *Proc. 7<sup>th</sup> World Cong. Genet. Appl. Livest. Prod.*, Montpellier France.
- MUIR, W. M. (2005): Incorporation of competitive effects in forest tree or animal breeding programs. *Genetics* **170**: 1247–1259.
- PATTERSON, H. D. and R. THOMPSON (1971): Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**: 545–554.
- RADTKE, P. J., J. A. WESTFALL and H. E. BURKHART (2003): Conditioning a distance-dependent competition index to indicate the onset of intertree competition. *Forest Ecology and Management* **175**: 17–30.
- SILVERMAN, B. W. (1986): *Density estimation for statistics and data analysis*. Chapman and Hall, London.
- SMITH, B. J. (2003): *Bayesian Output Analysis Program (BOA) version 1.0 user's manual*. Available from <http://www.public-health.uiowa.edu/boa/Home.html>.
- SOERENSEN, D. and D. GIANOLA (2002): *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer-Verlag, New York.
- SORIA, F., F. BASURCO, G. TOVAL, L. SILIÓ, M. C. RODRIGUEZ and M. TORO (1998): An application of Bayesian techniques to the genetic evaluation of growth traits in *Eucalyptus globulus*. *Can. J. For. Res.* **28**: 1286–1294.
- VAN TASSELL, C. P., G. CASELLA and E. J. POLLAK (1995): Effects of Selection on Estimates of Variance Components Using Gibbs Sampling and Restricted Maximum Likelihood. *J. Dairy Sci.* **78**: 678–692.
- VAN VLECK, L. D. and J. P. CASSADY (2005): Unexpected estimates of variance components with a true model containing genetic competition effects. *J. Anim. Sci.* **83**: 68–74.
- WILLHAM, R. L. (1963): The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* **19**: 18–27.

WOLF, J. B. (2003): Genetic architecture and evolutionary constraint when the environment contains genes. *Proc. Nat. Academy of Sciences* **100**: 4655–4660.

WRIGHT, A. J. (1986): Individual and group selection with competition. *Theoretical and Applied Genetics* **72**: 256–263.

ZENG, W., S. K. GHOSH and B. LI (2004): A Blocking Gibbs Sampling Method to Detect Major Genes Affecting a Quantitative Trait for Diallel Mating Design. *Genetical Research* **84**: 1–12.

## Appendix A

*Derivation of the additive genetic variance with additive competition effects*

The additive genetic variance for direct and competition breeding values in (6) is:

$$\text{Var}(a_{d_i}) + \text{Var}\left(\sum_{j=1}^{m_i} f_{ijk} a_{c_j}\right) + 2 \text{Cov}\left(a_{d_i}, \sum_{j=1}^{m_i} f_{ijk} a_{c_j}\right) \quad [\text{A.1}]$$

Using the variance operator in the first term produces

$$\text{Var}(a_{d_i}) = (1 + F_i) \sigma_{Ad}^2 \quad [\text{A.2}]$$

For the second term in [A.1] we have

$$\begin{aligned} \text{Var}\left(\sum_{j=1}^{m_i} f_{ijk} a_{c_j}\right) &= \text{Var}(f_{i1D} a_{c_1} + f_{i2R-C} a_{c_2} + \dots + f_{imk} a_{c_m}) \\ &= \text{Var}(f_{i1D} a_{c_1}) + \text{Var}(f_{i2R-C} a_{c_2}) + \dots + \text{Var}(f_{imk} a_{c_m}) \\ &+ 2 \left[ \text{Cov}(f_{i1D} a_{c_1}, f_{i2R-C} a_{c_2}) + \dots + \text{Cov}(f_{ij'k} a_{c_{j'}}, f_{imk} a_{c_m}) \right] \\ &= f_{i1D}^2 \text{Var}(a_{c_1}) + \dots + f_{imk}^2 \text{Var}(a_{c_m}) \\ &+ 2 \left[ f_{i1D} f_{i2R-C} \text{Cov}(a_{c_1}, a_{c_2}) + \dots + f_{ij'k} f_{imk} \text{Cov}(a_{c_{j'}}, a_{c_m}) \right] \\ &= \sum_{j=1}^{m_i} f_{ijk}^2 \text{Var}(a_{c_j}) + 2 \sum_{j \neq j'}^{m_i} f_{ij'k} f_{imk} \text{Cov}(a_{c_{j'}}, a_{c_m}) \\ &= \sum_{j=1}^{m_i} f_{ijk}^2 (1 + F_j) \sigma_{Ac}^2 + 2 \sum_{j \neq j'}^{m_i} f_{ij'k} f_{imk} A_{jj'} \sigma_{Ac}^2 \end{aligned}$$

where  $A_{jj'}$  indicates the relationship between competitors  $j$  and  $j'$ . Therefore

$$\begin{aligned} \text{Var}\left(\sum_{j=1}^{m_i} f_{ijk} a_{c_j}\right) &= \left[ (n_{R-C} f_{iR-C}^2 + n_D f_{iD}^2) (1 + F_j) + 2 \sum_{j \neq j'}^{m_i} f_{ij'k} f_{imk} A_{jj'} \right] \sigma_{Ac}^2 \quad [\text{A.3}] \end{aligned}$$

## Appendix B

*REML-EM equations for the (co)variance components in a model with additive direct, additive competition breeding values and permanent environmental competition effects.*

In order to obtain REML-EM of dispersion parameters in model [10], we take a similar approach to CANTET *et al.* (1993). Let the mixed model equations for the mixed model [10] be:

For the third term in [A.1] we use the covariance operator so that

$$\begin{aligned} \text{Cov}\left(a_{d_i}, \sum_{j=1}^{m_i} f_{ijk} a_{c_j}\right) &= \text{Cov}(a_{d_i}, f_{i1D} a_{c_1} + f_{i2R-C} a_{c_2} + \dots + f_{imk} a_{c_m}) \\ &= \text{Cov}(a_{d_i}, f_{i1D} a_{c_1}) + \text{Cov}(a_{d_i}, f_{i2R-C} a_{c_2}) \\ &+ \dots + \text{Cov}(a_{d_i}, f_{imk} a_{c_m}) \quad [\text{A.4}] \\ &= f_{i1D} \text{Cov}(a_{d_i}, a_{c_1}) + f_{i2R-C} \text{Cov}(a_{d_i}, a_{c_2}) \\ &+ \dots + f_{imk} \text{Cov}(a_{d_i}, a_{c_m}) \\ &= \sum_{j=1}^{m_i} f_{ijk} A_{ij} \sigma_{AdAc} \end{aligned}$$

Now, by replacing in [A.1] with [A.2], [A.3] and [A.4] gives the additive variance terms for the variance of  $y_{ij1..jm}$  in [9] as

$$\begin{aligned} \text{Var}\left(a_{d_i} + \sum_{j=1}^{m_i} f_{ijk} a_{c_j}\right) &= (1 + F_i) \sigma_{Ad}^2 \\ &+ \left[ (n_{R-C} f_{iR-C}^2 + n_D f_{iD}^2) (1 + F_j) + 2 \sum_{j \neq j'}^{m_i} f_{ij'k} f_{imk} A_{jj'} \right] \sigma_{Ac}^2 \\ &+ 2 \sum_{j=1}^{m_i} f_{ijk} A_{ij} \sigma_{AdAc} \end{aligned}$$

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_d' & X'R^{-1}Z_c' & X'R^{-1}Z_p' \\ Z_d'R^{-1}X & Z_d'R^{-1}Z_d + k_{11}A^{-1} & Z_d'R^{-1}Z_c + k_{12}A^{-1} & Z_d'R^{-1}Z_p \\ Z_c'R^{-1}X & Z_c'R^{-1}Z_d + k_{21}A^{-1} & Z_c'R^{-1}Z_c + k_{22}A^{-1} & Z_c'R^{-1}Z_p \\ Z_p'R^{-1}X & Z_p'R^{-1}Z_d & Z_p'R^{-1}Z_c & Z_p'R^{-1}Z_p + I \begin{pmatrix} \frac{1}{\sigma_d^2} \\ \frac{1}{\sigma_c^2} \end{pmatrix} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a}_d \\ \hat{a}_c \\ \hat{\rho}_c \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z_d'R^{-1}y \\ Z_c'R^{-1}y \\ Z_p'R^{-1}y \end{bmatrix} \quad [\text{B.1}]$$

The inverse of the coefficient matrix in [B.1] is

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z_d' & X'R^{-1}Z_c' & X'R^{-1}Z_p' \\ Z_d'R^{-1}X & Z_d'R^{-1}Z_d'+k_{11}A^{-1} & Z_d'R^{-1}Z_c'+k_{12}A^{-1} & Z_d'R^{-1}Z_p' \\ Z_c'R^{-1}X & Z_c'R^{-1}Z_d'+k_{21}A^{-1} & Z_c'R^{-1}Z_c'+k_{22}A^{-1} & Z_c'R^{-1}Z_p' \\ Z_p'R^{-1}X & Z_p'R^{-1}Z_d' & Z_p'R^{-1}Z_c' & Z_p'R^{-1}Z_p'+I\begin{pmatrix} z \\ \sigma_z^2 \end{pmatrix} \end{bmatrix}^{-1} = \begin{bmatrix} C^{\beta\beta} & C^{\beta\alpha_d} & C^{\beta\alpha_c} & C^{\beta p} \\ C^{\alpha_d\beta} & C^{\alpha_d\alpha_d} & C^{\alpha_d\alpha_c} & C^{\alpha_d p} \\ C^{\alpha_c\beta} & C^{\alpha_c\alpha_d} & C^{\alpha_c\alpha_c} & C^{\alpha_c p} \\ C^{p\beta} & C^{p\alpha_d} & C^{p\alpha_c} & C^{p p} \end{bmatrix}$$

Then, the REML-EM estimating equations of the dispersion parameters in the  $k^{\text{th}}$  iteration are

$$\hat{\sigma}_{Ad}^{2[k]} = \frac{[\hat{a}_d'A^{-1}\hat{a}_d]^{[k]} + \text{tr}(A^{-1}C^{\alpha_d\alpha_d})\hat{\sigma}_e^{2[k-1]}}{q}$$

$$\hat{\sigma}_{Ac}^{2[k]} = \frac{[\hat{a}_c'A^{-1}\hat{a}_c]^{[k]} + \text{tr}(A^{-1}C^{\alpha_c\alpha_c})\hat{\sigma}_e^{2[k-1]}}{q}$$

$$\hat{\sigma}_{AcAd}^{2[k]} = \frac{[\hat{a}_d'A^{-1}\hat{a}_c]^{[k]} + \text{tr}(A^{-1}C^{\alpha_d\alpha_c})\hat{\sigma}_e^{2[k-1]}}{q}$$

$$\hat{\sigma}_p^{2[k]} = \frac{[\hat{p}_c'\hat{p}_c]^{[k]} + \text{tr}(C^{p p})\hat{\sigma}_e^{2[k-1]}}{n}$$

$$\hat{\sigma}_e^{2[k]} = \frac{[\hat{e}'\hat{e}]^{[k]} + (p+2q-f^{[k]})\hat{\sigma}_e^{2[k-1]}}{n}$$

where  $\hat{e}^{[k]} = y - X\hat{\beta}^{[k]} - Z_d\hat{a}_d^{[k]} - Z_c\hat{a}_c^{[k]} - Z_p\hat{p}_c^{[k]}$ , and

$$f^{[k]} = \left[ \text{tr}(A^{-1}C^{dd})g^{dd} + 2\text{tr}(A^{-1}C^{dc})g^{dc} + \text{tr}(A^{-1}C^{cc})g^{cc} + \text{tr}(C^{p p})\hat{\sigma}_p^{2[k-1]} \right]$$

$$\text{with } G_0^{-1} = \begin{bmatrix} g^{dd} & g^{dc} \\ g^{cd} & g^{cc} \end{bmatrix}.$$

## Inheritance and Genetic Gain in Wood Stiffness in Radiata Pine Assessed Acoustically in Young Standing Trees

By A. C. MATHESON<sup>1</sup>, W. J. GAPARE<sup>1,\*</sup>, J. ILIC<sup>2</sup> and H. X. WU<sup>1</sup>

(Received 5<sup>th</sup> October 2006)

### Abstract

Wood stiffness, measured in terms of its modulus of elasticity (MoE) is an important characteristic of radiata pine for structural products. To select high stiffness radiata pine for breeding purpose, rapid, inexpensive methods for measuring wood stiffness are desirable. In this study, we explored acoustic instruments to measure stiffness of young standing trees in radiata pine and examined inheritance and genetic gain for stiffness in an Australian national breeding program. Time of flight of sound waves was recorded in standing trees in two progeny trials, one in eastern Victoria (Flynn) aged 8 years and the other in South Australia (Kromelite) aged 7 years. Average time of flight at Kromelite was higher than at Flynn, (519  $\mu\text{s}/\text{metre}$  compared to 463  $\mu\text{s}/\text{metre}$ ) which corresponds to 3.7 GPa and 4.7 GPa for MoE, respectively. Heritability for time of flight was higher at Flynn ( $h^2 = 0.67 \pm 0.10$ ) than at Kromelite ( $h^2 = 0.30 \pm 0.14$ ). Selection of the best 10% for time of flight based

on pooled data would result in 21% genetic gain in wood stiffness.

*Key words:* wood stiffness, acoustics, heritability, genetic gain, radiata pine.

### Introduction

Australia has an advanced breeding program for *Pinus radiata* D. Don (POWELL *et al.*, 2005), which has over the last 50 years significantly improved many characteristics of this widely planted fast growing conifer (MATHESON *et al.*, 1986; COTTERILL and DEAN, 1990; WU *et al.*, 2004). Traditionally trees have been selected based on visual qualities including diameter, height, branching, straightness, observable defects, age and site characteristics (WU and MATHESON, 2002). However, neither site nor these visual characteristics are good predictors of the mechanical properties of the wood products. In addition, as the main uses of the products are structural applications either as solid wood or as engineered wood products, there is a demand to plant trees with high stiffness. In the framework of a genetic improvement program, visually unobservable characteristics such as wood stiffness could be considered as selection criteria in the same way as growth or form, if they can be measured inexpensively, to maintain or improve the mechanical properties of wood produced.

<sup>1</sup>) Ensis<sup>3</sup>, CSIRO Forestry and Forest Products, P. O. Box E4008, Canberra, ACT 2604, Australia.

<sup>2</sup>) Ensis, CSIRO Forestry and Forest Products, Private Bag 10, Clayton South, Victoria 3168, Australia.

<sup>3</sup>) Ensis is a joint venture between CSIRO FFP P/L and Scion Australasia P/L.

\*) Corresponding author: Telephone: +61-2-6281-8322, Fax: +61-2-6281-8312. E-Mail: [Colin.Matheson@csiro.au](mailto:Colin.Matheson@csiro.au)