A novel individual-tree mixed model to account for competition and environmental heterogeneity: a Bayesian approach

Eduardo P. Cappa[12*] Facundo Muñoz[3] Leopoldo Sanchez[3] Rodolfo J. C. Cantet[24]

[1] Instituto Nacional de Tecnología Agropecuaria (INTA), Instituto de Recursos Biológicos, De Los Reseros y Dr. Nicolás Repetto s/n, 1686, Hurlingham, Buenos Aires, Argentina.

[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

[3] Institut National de la Recherche Agronomique (INRA) Orléans, Unité Amélioration, Génétique et Physiologie Forestières, 2163 Avenue de la Pomme de Pin, CS 40001 ARDON, 45075 Orleans, Cedex 02, France.

[4] Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, Avenida San Martín 4453, C1417DSQ, Buenos Aires, Argentina.

* **Corresponding author**
Eduardo Pablo Cappa
Bosques Cultivados
Instituto de Recursos Biológicos
Centro de Investigación en Recursos Naturales
Instituto Nacional de Tecnología Agropecuaria
De Los Reseros y Dr. Nicolás Repetto s/n
1686 Hurlingham, Buenos Aires
Argentina.

Phone: +54 11 4621 0433
Fax: +54 11 4621 0433
e-mail: ecappa@cnia.inta.gov.ar

Abstract

Negative correlation caused by competition among individuals, and positive spatial correlation due to environmental heterogeneity, may lead to biases in estimating genetic parameters and predicting breeding values (BV) from forest genetic trials. Former models dealing with competition and environmental heterogeneity did not account for the additive relationships among trees or for the full spatial covariance. This paper extends an individual-tree mixed model with direct additive genetic, genetic and environmental competition effects, by incorporating a two-dimensional smoothing surface to account for complex patterns of environmental heterogeneity (Competition + Spatial Model, CSM). We illustrate the proposed model using simulated and real data from a loblolly pine progeny trial. The CSM was compared with three reduced individual-tree mixed models using a real data set, while simulations comprised only CSM versus true parameters comparisons. Dispersion parameters were estimated using Bayesian techniques via Gibbs sampling. Simulation results showed that the CSM yielded posterior mean estimates of variance components with slight or negligible biases in the studied scenarios, except for the permanent environment variance. The worst performance of the simulated CSM was under a scenario with weak competition effects and small-scale environmental heterogeneity. When analyzing real data, the CSM yielded a lower value of the Deviance Information Criterion than the reduced models. Moreover, although correlations between predicted BV calculated from CSM and from a standard model with block effects and direct genetic effects only were high, the ranking among the top 5% ranked individuals showed differences which indicated that the two models will have quite different genotype selections for the next cycle of breeding.

Introduction

        Advanced forest genetic evaluation involves analyzing data from progeny tests using mixed linear models to calculate "best linear unbiased predictors" (BLUP) of tree breeding values (BV). As BLUP prediction depends on the values of the covariance matrices for the assumed model, the specification of the dispersion parameters should take into account the negative correlation caused by competition among individuals, and the positive spatial correlation due to the environmental heterogeneity. In field trials with perennial plants, both phenomena (i.e., competition and environmental heterogeneity) are dynamic and coexist simultaneously (Magnussen 1994). Therefore, statistical genetic analyses neglecting these factors or considering only one of them can lead to biases in the estimation of genetic parameters and in the prediction of individual additive genetic effects (i.e., BLUP of BV). Simulation studies have shown that positive spatial correlation inflates the additive genetic variance, while moderate levels of negative correlation caused by competition depresses it (Magnussen 1994). Therefore, when both competition and environmental heterogeneity are present in a forest genetic trial, a complete model approach that allows fitting simultaneously both processes is necessary (Resende et al. 2005). However, appropriate choice of the model is likely to influence how well the two processes can be separated analytically (Durban et al. 2001).

        Competition reflects the impairing interplay of closely neighboring trees, often when local resources are limiting. It depends on the genetic composition and the spatial arrangement of neighboring trees (Hinson and Hanson 1962), and it can be decomposed into genetic and environmental sources (Magnussen 1989). Cappa and Cantet (2008) presented an approach to account for competition effects in forest genetic evaluation. The mixed linear model included direct and indirect (i.e., competition) genetic effects, as well as environmental competition effects. Competition effects, either genetic or environmental, are identified in the phenotype of a competitor tree by means of the 'intensity of competition' elements (IC). The ICs are inverse functions 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, so that the model accounts for unequal number of neighbors in locations with mortality and borders. Cappa and Cantet (2008) applied this approach to data on growth at breast height in loblolly pines, resulting in a strong negative correlation between direct and competition breeding values of -0.79. In the model where competition effects were not fitted, the value of the additive genetic variance was smaller and the residual variance was larger than in the model with competition effects. However, Stringer et al. (2011) noted that the approach proposed by Cappa and Cantet (2008) was limited due to not taking into account environmental heterogeneity within trials.

        There are usually two types of spatial environmental variation within a site: global trend (or large-scale variation), and/or local trend (or small-scale variation). Both are empirically known in forestry field trials, and result from variations in soil characteristics and terrain orientation. Many studies applying spatial analysis techniques to single forest genetic trials reported a consistent reduction in residual variance and an increase in estimated heritability. This typically results in a gain of accuracy of BV and in greater genetic gain when compared with different *a priori* experimental designs (e.g. Dutkowski et al. 2006; Zas 2006; Cappa

and Cantet 2007; Ye and Jayawickrama 2008). Several approaches have been developed and applied to single forest trials to reduce the effects of environmental variability (Magnussen 1990; Zas 2006; Thomson and El-Kassaby 1988; Costa e Silva et al. 2001; Ericsson 1997). However, fitting a two-dimensional surface proved useful for accommodating complex patterns of spatial heterogeneity within forest genetic trials (Cappa and Cantet 2007; Finley et al. 2009; Cappa et al. 2011). Cappa and Cantet (2007) proposed using tensor products of cubic B-splines (Eilers and Mark 2003) based on a mixed model, by treating the coefficients of B-spline functions as random variables (and using a covariance structure for the random spline effects based on the spatial arrangement) in a two-dimensional grid. The mixed model with the fit of a two-dimensional surface displayed a consistent reduction in the residual variance, an increase in the estimated additive genetic variance and heritability, and an increase in the accuracy of the predicted breeding values of parent and offspring in both large (Cappa and Cantet 2007) and small (< 1 ha) forest genetic trials (Cappa et al. 2011). Finley et al. (2009) proposed an individual linear mixed model with spatial random effects modeled with two methods: the first used an exponential Ornstein–Uhlenbeck process, whereas the second was based on a hierarchical Gaussian predictive process model with a set of knots for rows and columns. In a dataset from Scots pine (*Pinus sylvestris* L.) progeny trial conducted in northern Sweden, they demonstrated that estimates from the predictive process approach provided a reliable reduction in the residual variance and an increment in the heritability even when the underlying spatial process was misaligned with the grid's rows and columns. However, where both environmental heterogeneity and competition effects are important, a full model approach is needed (Resende et al. 2005).

In a simulation study of an empirical sib-trial data set from *Pinus banksiana* Lamb., Magnussen (1994) suggested modeling both effects by means of an iterative nearest neighbor adjustment and a standard linear model (two-way ANOVA). In a series of 70 sugar beet trials, Durban et al. (2001) simultaneously modeled the fertility trend and the genetic competition by means of a mixed model. They used the "treatment interference model" proposed by Besag and Kempton (1986) to model the interplot competition, and a one-dimensional smoothing spline to account for environmental heterogeneity. However, Durban et al. (2001) considered varieties as fixed effects, and did not account for additive relationships among genotypes, a usual assumption in variety selection. A more realistic joint modeling approach was proposed recently by Stringer et al. (2011) and applied in two early-stage data sets from sugarcane trials. They extended the model proposed by Durban et al. (2001) modeling the variability from the local environmental trend with a Kronecker product of first order autoregressive (AR(1)) covariance structures for rows and columns (Gilmour et al. 1997), and included random variety genetic effects. However, when in a forest genetic trial both competition and environmental effects are important, the AR(1) covariance residual structure reflects a mixture of environmental competition and local trend. To minimize this confounding of effects, Stringer et al. (2011) proposed modeling the environmental competition effects using an autoregressive residual process, but in this case of order 2 (EAR2), or 3 (EAR3). While analyzing sorghum trials from Australia, Hunt et al. (2013) accommodated both spatial heterogeneity and interplot competition while using pedigree information. The spatial heterogeneity was modeled with an AR(1) residual covariance structure for rows and columns, and the

interplot competition using the "treatment interference model" (Besag and Kempton 1986). Following Stringer et al. (2005), Resende et al. (2005) modeled simultaneously competition effects and spatial variability in two field forest trials, they used the approach proposed by Besag and Kempton (1986) to account for competition, and the approach of Gilmour et al. (1997) to model via AR(1) row and column environmental variation. In that tree study, however, Resende et al. (2005) did not account for additive relationships among trees, they applied the joint competition and spatial model at the plot level instead of at the individual-tree level, and did not accommodate unequal number of neighbors due to mortality or edge plot effects.

In the current research, statistical and methodological aspects of joint analysis of competition effects and environmental heterogeneity of forest genetic trials are developed. We extend the individual-tree mixed model with additive direct, genetic and environmental competition effects of Cappa and Cantet (2008) with a two-dimensional smoothing surface (Cappa and Cantet 2007) that captures the complex patterns of environmental heterogeneity. A Bayesian approach via Gibbs sampling was employed to make inference in all dispersion parameters of the model. Developments are illustrated with simulated datasets covering a wide range of competition and environmental parameters, and with real data for diameter at breast height of *Pinus taeda* L. used by Cappa and Cantet (2008). The resulting estimates of all dispersion parameters are compared with the corresponding estimates from three reduced individual-tree mixed models: the standard model with block and direct genetic effects only (no competition and spatial continuous effects), the competition model (Cappa and Cantet 2008), and the spatial model (Cappa and Cantet 2007). The ranking of selection candidates calculated for the joint and standard models are compared to determine the importance of simultaneously accounting for competition and environmental heterogeneity effects.

Methods

Competition individual-tree mixed model

We follow closely Cappa and Cantet (2008) to describe the models presented in this research. The individual-tree mixed model that includes direct and competition genetic effects plus a permanent environmental competition effect, is described by the following model equation

$$ \boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z}_d \boldsymbol{a}_d + \boldsymbol{Z}_c \boldsymbol{a}_c + \boldsymbol{Z}_p \boldsymbol{p}_c + \boldsymbol{e} \qquad [1] $$

where $\boldsymbol{y} = [y_i]$ ($n \times 1$) contains the phenotype of tree i ($i = 1, ..., n$; n is the total number of trees with recorded data); $\boldsymbol{X}$ is the $n \times p$ incidence matrix relating records to the vector of fixed effects $\boldsymbol{\beta}$; $\boldsymbol{p}_c$ is a vector that includes permanent environmental effects such that $\boldsymbol{p}_c = [p_{ci}]$, such that $\boldsymbol{p}_c \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n \sigma_p^2)$; and $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n \sigma_e^2)$ is the random of residuals, and σ_e^2 is the residual variance. Direct and competition breeding values for parents without records plus offspring with data in $\boldsymbol{y}$ are included in the random vector $\boldsymbol{a}_d = [a_{dj}]$ and $\boldsymbol{a}_c = [a_{cj}]$, respectively, with $j = 1, ..., q > n$. The breeding values for an observed individual i are the $a_{j(i)}$ components of both vectors. Direct and competition breeding values are related to $\boldsymbol{y}$ by the $n \times q$ incidence matrices $\boldsymbol{Z}_d$ and $\boldsymbol{Z}_c$, respectively. Every row (i) of $\boldsymbol{Z}_d$ has all elements equal to 0 except for a 1 in the column $j(i)$. Similarly,

each row i of matrix $\mathbf{Z}_c$ has all elements equal to zero except in the position $j = j_1, ..., j_{mi}$ corresponding to the m_i neighbors of the tree i, where they are f_{ij}. These positive coefficients can be interpreted as the *intensity of competition* (IC) that each neighbor exerts over the phenotype of the ith tree. As show by Cappa and Cantet (2008), the standardization of the variance of competition effects within the phenotypic variance of any individual-tree, when accounting for unequal number of neighbors, can be achieved by using intensity of competition factors (f_{ij}) under the following restriction:

$$\sum_{j=1}^{m_j} f_{ij}^2 = n_{i\,R\text{-}C}\ f_{i\,R\text{-}C}^2 + n_{i\,D}\ f_{i\,D}^2 = 1$$

where $n_{i\,R\text{-}C}$ and $n_{i\,D}$ are the respective numbers of jth competitors that lie either in the same row or column of the tree i (R-C; see Cappa and Cantet 2008 Figure 1) and jth competitors that lie in the diagonal (D; see Cappa and Cantet 2008 Figure 1), and $f_{iR\text{-}C}$ and f_{iD} are the IC for R-C and for D competitors to tree i. In trees planted at certain distances, Cappa and Cantet (2008) assumed that the ICs are 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. Thus, $f_{iR\text{-}C}$ and f_{iD} are:

$$f_{iD} = \frac{1}{\sqrt{2\,n_{i\,R\text{-}C} + n_{i\,D}}} \qquad f_{iR\text{-}C} = \sqrt{\frac{2}{2\,n_{i\,R\text{-}C} + n_{i\,D}}} \qquad [2]$$

For a more detailed explanation of the derivation of these formulas, see Cappa and Cantet (2008). 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 $\mathbf{A}\sigma^2_{Ad}$, where the $q \times q$ matrix $\mathbf{A}$ is the additive relationship matrix and σ^2_{Ad} is the variance of the direct additive genetic effects. Also, $\mathbf{a}_c \sim N\,(\mathbf{0}, \mathbf{A}\,\sigma^2_{Ac})$ where σ^2_{Ac} is the variance of the competition breeding values, and $\text{cov}(\mathbf{a}_d, \mathbf{a}_c) = \mathbf{A}\ \sigma_{AdAc}$, where σ_{AdAc} is the covariance between direct and competition breeding values. Thus, the covariance matrix of the stacked vector of breeding values is equal to

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

In the following section, we focus on an extension of model [1] with a two-dimensional smoothing surface to account for large- and/or small-scale environmental heterogeneity.

Joint competition and spatial individual-tree mixed model

The individual-tree mixed model with competition effects [1] can be extended to include a two-dimensional smoothed surface to account for the environmental heterogeneity, as follows

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}\mathbf{b} + \mathbf{Z}_d\mathbf{a}_d + \mathbf{Z}_c\mathbf{a}_c + \mathbf{Z}_p\mathbf{p}_c + \mathbf{e} \qquad [3]$$

where $\boldsymbol{B}\,\boldsymbol{b}$ is the matrix expression approximating the two-dimensional surface with a tensor product of cubic B-spline bases (Eilers and Marx 2003). Matrix $\boldsymbol{B}$ has dimension $n \times (nx_r =$ number of splines for rows $\times\ nx_c =$ number of splines for columns), and is equal to $\boldsymbol{B} = \left[\boldsymbol{B}_r \otimes \boldsymbol{I}_{nx_c}^{'}\ \#\ \boldsymbol{I}_{nx_r}^{'} \otimes \boldsymbol{B}_c\right]$, where the symbols $\otimes$ and $\#$ indicate the Kronecker and Hadamard products of matrices, respectively (Harville 1997). The matrices $\boldsymbol{B}$ contain the B-splines bases evaluated in the corresponding rows and columns for each tree. Calculations of the $\boldsymbol{B}$ coefficients were performed using the recursive algorithm of De Boor (1993). The vector of random effects $\boldsymbol{b}$ contains the $nx_r \times nx_c$ coefficients of the tensor products of B-splines bases. The vector $\boldsymbol{b}$ is assumed normally distributed with mean zero and covariance matrix $\boldsymbol{U}\,\sigma_b^2$. The scalar σ_b^2 is the variance of the B-spline coefficients for rows and columns, and the $\boldsymbol{U}$ matrix is the covariance structure in two dimensions for the B-splines coefficients. In the present study, we select the tridiagonal matrix originally proposed by Green and Silverman (1994; page 13) and then used by Durban et al. (2001) to fit a fertility trend. A more detailed explanation of the two-dimensional surface ($\boldsymbol{B}\,\boldsymbol{b}$) and the $\boldsymbol{U}$ matrix covariance structure can be found in Cappa and Cantet (2007, pp. 2678-2680).

Taking into account the random effects in model [3], the covariance matrix ($\boldsymbol{V}$) of the data vector $\boldsymbol{y}$ is given by:

$$\boldsymbol{V} = \boldsymbol{B}\,\boldsymbol{U}\,\boldsymbol{B}^{'}\sigma_b^2 + \boldsymbol{Z}_d\,\boldsymbol{A}\,\boldsymbol{Z}_d^{'}\sigma_{Ad}^2 + \left[\boldsymbol{Z}_d\,\boldsymbol{A}\,\boldsymbol{Z}_c^{'} + \boldsymbol{Z}_c\,\boldsymbol{A}\,\boldsymbol{Z}_d^{'}\right]\sigma_{AdAc} + \boldsymbol{Z}_c\,\boldsymbol{A}\,\boldsymbol{Z}_c^{'}\sigma_{Ac}^2 + \boldsymbol{Z}_p\,\boldsymbol{Z}_p^{'}\sigma_p^2 + \boldsymbol{I}_n\,\sigma_e^2$$

A set of mixed model equations (Henderson 1984) for model [3] is equal to

$$\begin{bmatrix} \boldsymbol{X}^{'}\boldsymbol{X} & \boldsymbol{X}^{'}\boldsymbol{B} & \boldsymbol{X}^{'}\boldsymbol{Z}_d & \boldsymbol{X}^{'}\boldsymbol{Z}_c & \boldsymbol{X}^{'}\boldsymbol{Z}_p \\ \boldsymbol{B}^{'}\boldsymbol{X} & \boldsymbol{B}^{'}\boldsymbol{B}+\boldsymbol{U}^{-1}\left(\frac{\sigma_e^2}{\sigma_b^2}\right) & \boldsymbol{B}^{'}\boldsymbol{Z}_d & \boldsymbol{B}^{'}\boldsymbol{Z}_c & \boldsymbol{B}^{'}\boldsymbol{Z}_p \\ \boldsymbol{Z}_d^{'}\boldsymbol{X} & \boldsymbol{Z}_d^{'}\boldsymbol{B} & \boldsymbol{Z}_d^{'}\boldsymbol{Z}_d+k_{11}\boldsymbol{A}^{-1} & \boldsymbol{Z}_d^{'}\boldsymbol{Z}_c+k_{12}\boldsymbol{A}^{-1} & \boldsymbol{Z}_d^{'}\boldsymbol{Z}_p \\ \boldsymbol{Z}_c^{'}\boldsymbol{X} & \boldsymbol{Z}_c^{'}\boldsymbol{B} & \boldsymbol{Z}_c^{'}\boldsymbol{Z}_d+k_{21}\boldsymbol{A}^{-1} & \boldsymbol{Z}_c^{'}\boldsymbol{Z}_c+k_{22}\boldsymbol{A}^{-1} & \boldsymbol{Z}_c^{'}\boldsymbol{Z}_p \\ \boldsymbol{Z}_p^{'}\boldsymbol{X} & \boldsymbol{Z}_p^{'}\boldsymbol{B} & \boldsymbol{Z}_p^{'}\boldsymbol{Z}_d & \boldsymbol{Z}_p^{'}\boldsymbol{Z}_c & \boldsymbol{Z}_p^{'}\boldsymbol{Z}_p+\boldsymbol{I}\left(\frac{\sigma_e^2}{\sigma_p^2}\right) \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{b} \\ \boldsymbol{a}_d \\ \boldsymbol{a}_c \\ \boldsymbol{p}_c \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^{'}\boldsymbol{y} \\ \boldsymbol{B}^{'}\boldsymbol{y} \\ \boldsymbol{Z}_d^{'}\boldsymbol{y} \\ \boldsymbol{Z}_c^{'}\boldsymbol{y} \\ \boldsymbol{Z}_p^{'}\boldsymbol{y} \end{bmatrix} \qquad [4]$$

where $\begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} = \boldsymbol{G}_0^{-1}\sigma_e^2$

The conditional likelihood of the observed data is written as being proportional to:

$$p\left(\boldsymbol{y}\mid\boldsymbol{\beta},\boldsymbol{b},\boldsymbol{a}_d,\boldsymbol{a}_c,\boldsymbol{p}_c,\sigma_b^2,\boldsymbol{G}_0,\sigma_p^2,\sigma_e^2\right) \propto \left(\sigma_e^2\right)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma_e^2}\boldsymbol{e}^{'}\boldsymbol{e}\right] \qquad [5]$$

where $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{B}\boldsymbol{b} - \boldsymbol{Z}_d\boldsymbol{a}_d - \boldsymbol{Z}_c\boldsymbol{a}_c - \boldsymbol{Z}_p\boldsymbol{p}_c$.

Bayesian estimation of covariance components

All dispersion parameters of model [3], i.e. σ^2_b , $\sigma^2_{A_d}$, $\sigma_{A_dA_c}$, $\sigma^2_{A_c}$, σ^2_p and σ^2_e , are estimated using a Bayesian approach, by means of Gibbs sampling (Sorensen and Gianola 2002). Below we describe in detail the prior distribution of all parameters, and the joint and conditional posterior densities for the joint individual-tree model with competition and spatial effects, as these are needed to make posterior inference with the Gibbs sampler. In doing so, we follow closely Cappa and Cantet (2007, 2008).

Specification of prior distributions: In a conjugate approach the prior densities for all parameters are chosen to be closed under sampling (Robert and Casella 1999), which means that both prior and posterior belong to the same family of distributions. In order to reflect a prior state of uncertainty for the fixed effects in a mixed linear model, while keeping the posterior distribution proper (Hobert and Casella 1996), $\boldsymbol{\beta}$ is taken to be N_p (**0**, **K**). The matrix **K** is diagonal with large elements ($k_{ii} > 10^8$), so that the prior density of $\boldsymbol{\beta}$ is then proportional to

$$p\left(\boldsymbol{\beta} \mid \boldsymbol{K}\right) \propto \left|\prod_{i=1}^{p} k_{ii}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{p}\frac{\beta_i^2}{k_{ii}}\right\} \tag{6}$$

The vector **b** is distributed *a priori* as $\boldsymbol{b} \sim N_b(\boldsymbol{0}, \boldsymbol{U}\,\sigma^2_b)$, so that the prior distribution of **b** is

$$p\left(\boldsymbol{b} \mid \sigma^2_b\right) \propto \sigma_b^{2\,-\frac{nx*nx}{2}} \exp\left\{-\frac{\boldsymbol{b}'\boldsymbol{U}^{-1}\boldsymbol{b}}{2\sigma^2_b}\right\} \tag{7}$$

The joint prior distribution of the direct and competition breeding values $\boldsymbol{a} = [\boldsymbol{a}_d{}',\, \boldsymbol{a}_c{}']'$ can be written as

$$p\left(\boldsymbol{a} \mid \boldsymbol{A}, \boldsymbol{G}_0\right) \propto \left|\boldsymbol{G}_0\right|^{-\frac{q}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{G}_0^{-1}\boldsymbol{S}_g\right)\right) \tag{8}$$

where

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

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

$$p\left(\boldsymbol{p}_c \mid \sigma^2_p\right) \propto \sigma_p^{2\,-\frac{n}{2}} \exp\left(-\frac{\boldsymbol{p}_c'\boldsymbol{p}_c}{2\sigma^2_p}\right) \tag{9}$$

The covariance matrix of the direct and competition breeding values ($\boldsymbol{G}_0$) follows *a priori* an Inverted Wishart (IW) density: $\boldsymbol{G}_0 \sim \text{IW}\,(\boldsymbol{G}_0^*,\, \upsilon_g)$, where $\boldsymbol{G}_0^*$ is the prior covariance matrix and υ_g is the degree of belief parameter. Thus:

$$p\left(\boldsymbol{G}_0 \mid \boldsymbol{S}_g, \upsilon_g\right) \propto \left|\boldsymbol{G}_0\right|^{-\frac{(\upsilon_g+3)}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{G}_0^* \boldsymbol{G}_0^{-1}\right)\right) \tag{10}$$

Following Sorensen and Gianola (2002), we choose to use independent scaled inverted chi-square densities as prior distributions for the variance components σ_b^2, σ_p^2 and σ_e^2:

$$p\left(\sigma_b^2 \mid \upsilon_b, \delta_b^2\right) \propto \left(\sigma_b^2\right)^{-\left(\frac{\upsilon_b}{2}+1\right)} \exp\left\{-\frac{\upsilon_b \delta_b^2}{2\sigma_b^2}\right\} \tag{11}$$

$$p\left(\sigma_p^2 \mid \upsilon_p, \delta_p^2\right) \propto \left(\sigma_p^2\right)^{-\left(\frac{\upsilon_p}{2}+1\right)} \exp\left\{-\frac{\upsilon_p \delta_p^2}{2\sigma_p^2}\right\} \tag{12}$$

$$p\left(\sigma_e^2 \mid \upsilon_e, \delta_e^2\right) \propto \left(\sigma_e^2\right)^{-\left(\frac{\upsilon_e}{2}+1\right)} \exp\left\{-\frac{\upsilon_e \delta_e^2}{2\sigma_e^2}\right\} \tag{13}$$

Parameters of the densities [11], [12], and [13] are the hypervariances δ_b^2, δ_p^2 and δ_e^2, and the degrees of freedom υ_b, υ_p and υ_e, respectively.

Joint and conditional posterior densities: multiplying expression [5] with densities [6] to [13] produces the joint posterior density for all parameters, and this is proportional to

$$p\left(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{a}_d, \boldsymbol{a}_c, \boldsymbol{p}_c, \sigma_b^2, \boldsymbol{G}_0, \sigma_p^2, \sigma_e^2 \mid \boldsymbol{y}\right) \propto$$
$$p\left(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{a}_c, \boldsymbol{a}_d, \boldsymbol{p}_c, \sigma_b^2, \boldsymbol{G}_0, \sigma_p^2, \sigma_e^2\right) p\left(\boldsymbol{\beta} \mid \boldsymbol{K}\right) p\left(\boldsymbol{b} \mid \sigma_b^2\right) p\left(\boldsymbol{a}_c, \boldsymbol{a}_d \mid \boldsymbol{A}, \boldsymbol{G}_0\right) \tag{14}$$
$$p\left(\boldsymbol{p}_c \mid \sigma_p^2\right) p\left(\boldsymbol{G}_0 \mid \boldsymbol{S}_g, \upsilon_g\right) p\left(\sigma_b^2 \mid \delta_b^2, \upsilon_b\right) p\left(\sigma_p^2 \mid \delta_p^2, \upsilon_p\right) p\left(\sigma_e^2 \mid \delta_e^2, \upsilon_e\right)$$

The posterior conditional density for the Gibbs sampling of $\boldsymbol{\beta}$, $\boldsymbol{b}$, $\boldsymbol{a}_d$, $\boldsymbol{a}_c$, and $\boldsymbol{p}_c$ is equal to

$$
\begin{bmatrix} \beta \\ b \\ a_d \\ a_c \\ p_c \end{bmatrix} \Big| y, \sigma_b^2, G_0, \sigma_p^2, \sigma_e^2 \sim
$$

$$
N \left(\begin{bmatrix} \beta \\ b \\ a_d \\ a_c \\ p_c \end{bmatrix}, \begin{bmatrix} X'X & X'B & X'Z_d & X'Z_c & X'Z_p \\ B'X & B'B + U^{-1}\left(\frac{\sigma_e^2}{\sigma_b^2}\right) & B'Z_d & B'Z_c & B'Z_p \\ Z_d'X & Z_d'B + k_{11}A^{-1} & Z_d'Z_d + k_{11}A^{-1} & Z_d'Z_c + k_{12}A^{-1} & Z_d'Z_p \\ Z_c'X & Z_c'B + k_{21}A^{-1} & Z_c'Z_d + k_{21}A^{-1} & Z_c'Z_c + k_{22}A^{-1} & Z_c'Z_p \\ Z_p'X & Z_p'B & Z_p'Z_d & Z_p'Z_c & Z_p'Z_p + I\left(\frac{\sigma_e^2}{\sigma_p^2}\right) \end{bmatrix}^{-1} \right)
$$

[15]

where $\hat{\beta}, \hat{b}, a_d, a_c$, and p_c are the solutions to equation [4].

Expression [15] may suggest that sampling of β, b, a_d, a_c, and 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 third and seventh terms in the right of [14], the full conditional posterior distribution of σ_b^2 is the scaled inverted chi-square

$$
p\left(\sigma_b^2 | \beta, b, a_d, a_c, p_c, G_0, \sigma_p^2, \sigma_e^2, y \right) \propto \left(\sigma_b^2\right)^{-\left(\frac{n + \upsilon_b + 2}{2} + 1\right)} \exp\left\{ -\frac{\tilde{\upsilon}_b \tilde{\delta}_b^2}{2\sigma_b^2} \right\}
$$

[16]

with $\tilde{\upsilon}_b = nx * nx + \upsilon_b$ and $\tilde{\delta}_b^2 = \left(b'U^{-1}b + \upsilon_b \delta_b^2 \right) / \tilde{\upsilon}_b$.

Collecting the fourth and sixth terms on the right of [14], the full conditional posterior distribution of G_0 is equal to

$$
p\left(G_0 | \beta, b, a_c, a_d, p_c, \sigma_p^2, \sigma_e^2, y \right) \propto |G_0|^{-\frac{(\upsilon_g + q + 3)}{2}} \exp\left\{ -\frac{1}{2} \mathrm{tr}\left[G_0^{-1}\left(S_g + G_0^* \right) \right] \right\}
$$

[17]

Expression [17] is the kernel of a 2×2 scaled inverted Wishart distribution, with degrees of freedom equal to $(\upsilon_g + q + 3)$ and scale matrix $S_g + G_0^*$.

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

$$p\left(\sigma_p^2 \mid \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{a}_c, \boldsymbol{a}_d, \boldsymbol{p}_c, \sigma_b^2, \boldsymbol{G}_0, \sigma_e^2, \boldsymbol{y}\right) \propto \left(\sigma_p^2\right)^{-\left(\frac{n+\upsilon_p+2}{2}+1\right)} \exp\left\{-\frac{\tilde{\upsilon}_p \tilde{\delta}_p^2}{2\sigma_p^2}\right\} \qquad [18]$$

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

$$\tilde{\delta}_p^2 = \left.\left(\boldsymbol{p}_c'\boldsymbol{p}_c + \upsilon_p \delta_p^2\right)\middle/\left(\upsilon_p + n\right)\right. .$$

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

$$p\left(\sigma_e^2 \mid \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{a}_c, \boldsymbol{a}_d, \boldsymbol{p}_c, \sigma_b^2, \boldsymbol{G}_0, \sigma_p^2, \boldsymbol{y}\right) \propto \left(\sigma_e^2\right)^{-\left(\frac{n+\upsilon_e+2}{2}+1\right)} \exp\left\{-\frac{\tilde{\upsilon}_e \tilde{\delta}_e^2}{2\sigma_e^2}\right\} \qquad [19]$$

i.e. a scaled inverted χ^2 density with $\tilde{\upsilon}_e = n + \upsilon_e$ degrees of freedom and scale parameter

$$\tilde{\delta}_e^2 = \left.\left(\boldsymbol{e}'\boldsymbol{e} + \upsilon_e \delta_e^2\right)\middle/\left(\upsilon_e + n\right)\right. .$$

At any iteration of the Gibbs algorithm, we first sampled from distribution [15], then from [19], next from [17], next from [18], and finally from [16], to start the process back again. A program was written in FORTRAN to perform all calculations. The FORTRAN program is available from the first author on request.

An application to loblolly pines
Data

Data came from a field trial of 20 open-pollinated families of loblolly pine (*Pinus taeda* L.), originating from Marion County, Florida, USA, and belonging to CIEF (Forestry Research and Experimentation Centre). Five lots of commercial seeds were used as control populations (about 7% of the total of phenotypic data). These trees were included in the analysis due to their contribution to the estimation of environmental variation (Dutkowski et al. 2006). The trait analyzed was diameter at breast height (1.3 m, DBH) measured at age 13 on the 932 surviving trees (93.2 % of survival). The trial site was located in Villa Olivari (lat. 27º 36' S long. 56º 55' W), northern Corrientes province, Argentina. Families were arranged in randomized complete blocks, with eight replicates of five-trees row-plots, and the spacing was 3.5 m × 3.5 m. Further details about the experimental data are given in Cappa and Cantet (2008).

Models of analysis

Four individual-tree mixed models were evaluated in the loblolly pine data set. All models included a random direct additive genetic effect and a random effect of commercial seed lots. The latter was to avoid

biasing in the estimates of the additive genetic variances caused by the inclusion of these trees with unknown parents in the additive relationships matrix *A* (Dutkowski et al. 2006). The standard individual-tree mixed model ("standard model", TM) also included a fixed block effects. The second individual-tree mixed model, the "competition model" (CM) included fixed block effects, direct additive genetic and competition effects, and permanent competition effects (see expression [1]). The third model, the "spatial model" (SM) included the same effects as in TM plus a tensor product of B-splines (to account for environmental heterogeneity). Four spatial structures with different numbers of knots for rows and columns were fitted: 10×10, 12×12, 15×15, and 18×18. The model including 15×15 knots displayed the better fit (based on the smallest Deviance Information Criterion, see below), and captured most of the spatial variability (i.e., visual comparison between the spatial patterns of the residuals and the resulting estimated surface shows minor differences). Therefore, this model was used for further comparison. The fourth individual-tree mixed model, i.e. the "competition + spatial model" (CSM) included a two-dimensional smoothed surface, a direct and competition genetic effects, plus permanent competition effects (model [3]). The Deviance Information Criterion (DIC, Spiegelhalter et al. 2002) was employed to compare the models. The DIC combines a measure of model fit (the posterior mean deviance), with a measure of model complexity (the "effective number of parameters"). Models with more parameters display better fit, but at the expense of adding complexity. Similar to Akaike Information Criterion (AIC), DIC penalizes the additional parameters that improve the fit while in the search for a more parsimonious model. Therefore, models with the smallest value of DIC should be favored, as this indicates a better fit and a lower degree of model complexity. Total tree breeding value (TBV$_i$) from CSM [3] was calculated following Costa e Silva and Kerr (2013; equation 14); i.e., TBV$_i$ = a_{di} + 2.32 a_{ci}, where the quantity 2.32 is the sum of the products of the means across all focal individuals in the loblolly pine trial for the number of their neighbors and IC elements in the row, column and diagonal directions $\bar{n}_R \bar{f}_{ijR} + \bar{n}_C \bar{f}_{ijC} + \bar{n}_D \bar{f}_{ijD}$. It is important to examine the impact of competition effects and environmental heterogeneity on selection decisions (in addition to their impact on the variance component), when both types of effects exist. In that sense, Spearman rank correlations were calculated to assess the extent by which the ranking of predicted tree breeding values for all individuals (offspring) differed among the standard and the competition + spatial models. Additionally, the proportion of common individuals (offspring) in the top 5% (47 trees) from the two models was also compared.

Identification of spatial and competition effects

As in Gilmour et al. (1997), we examined the spatial distribution of residual to identify spatial patterns in the data, using an exploratory model with fixed overall mean and random commercial seed lot effect and direct breeding values. The distribution of the residuals from this model of DBH is displayed in Figure 1. The grayscale intensity represents the magnitude of the residuals - the darker and bright the square, the larger the residual values - and shows the small-scale environmental variation observed within the loblolly pine trial studied.
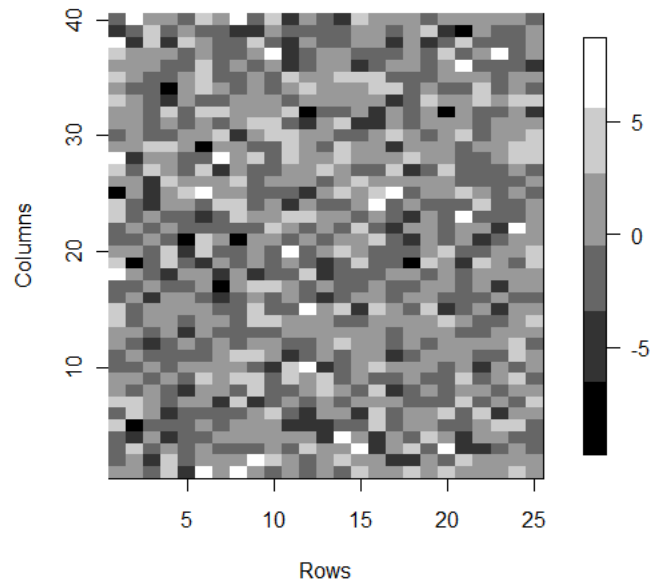
Figure 1: Spatial patterns of the residuals from an exploratory model of tree diameter at breast height.

The presence of competition was first detected by a diagnostic plot of residuals of DBH from the exploratory model, against the mean DBH of nearest 8 neighbors, following Durban et al. (2001) (Figure 2), where a negative correlation (r = -0.22) suggested the presence of competition. Second, the correlation between direct and competition genetic effects from CSM [3] was also used to identify competition effects at the genetic level. High and negative correlation between direct and competition BV (higher than -0.3, Kusnandar 2001) revealed strong genetic competition, meaning that a tree with a positive breeding value for its own growth has - on average - a negative genetic influence on the growth of its neighbors (Costa e Silva et al. 2013). The magnitude of this correlation was moderate to large and negative (–0.85, see below). Finally, a two-dimensional autoregressive (AR(1) × AR(1)) residual structure was fitted to the standard model using the ASREML program (Gilmour et al. 2006). The sign and magnitude of the autocorrelation parameters for row and column suggest that competition between neighbors may be present, and it is dominant over the environmental heterogeneity at the small scale (ρ_{row} = -0.34 and ρ_{column} = -0.10).
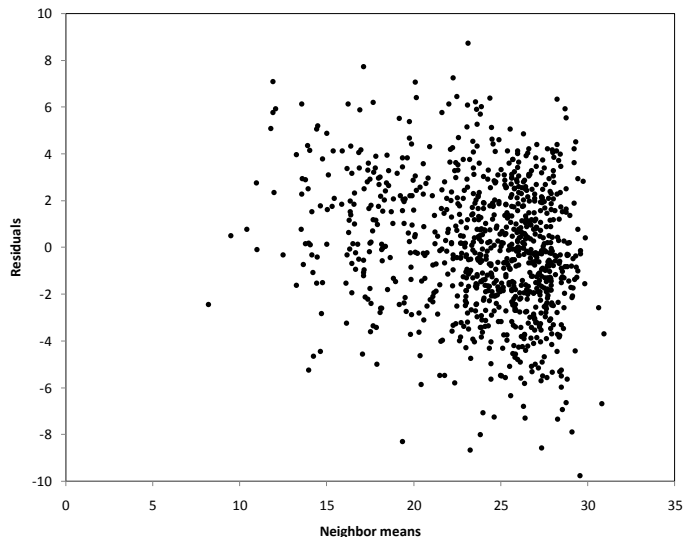
Figure 2: Residuals from the exploratory model vs. mean of tree diameter at breast height of the nearest neighbor trees.

Computational details and posterior inference

The values of the hypervariances in the priors of σ^2_{Ad} and σ^2_e were estimated from the same data set using an empirical Bayes approach via Gibbs sampling, with an individual-tree model including fixed effects of blocks, and random additive genetic effects (standard model). As there was no prior information on the hypervariance δ^2_b, we tried different values in the interval $[0, \delta^2_e)$ and found that the algorithm converged always to the same posterior mean of σ^2_b in the spatial model. The same strategy was used for σ_{AcAd} and σ^2_{Ac}, by trying different prior values of σ_{AcAd} (+, 0, and -) and of σ^2_{Ac} (high and low relative to σ^2_{Ad}) under CM. The hypervariance δ^2_p was chosen to be equal to the prior value for σ^2_{Ac}. The deviance information criterion (DIC) was computed for each model using the output from the Gibbs sampling.

A single Gibbs chain of 1,010,000 samples were drawn as discussed above. The first 10,000 iterations were discarded as *burn-in*. Convergence was monitored by plotting the iterations against the parameter value (trace plots) and against the mean of the draws up to each iteration (running mean plots) for each parameter, and using the *Z* criterion of Geweke (1992) for each parameter. To evaluate the impact of autocorrelations in the variability of the samples, the 'effective sample size' (ESS) proposed by R. Neal (Kass et al. 1998) was calculated for each parameter as:

$$\mathrm{ESS} = \frac{1\,000\,000}{1 + 2\sum_{i=1}^{50} \rho_i}$$

where ρ_i is the autocorrelation measured at lag i. Means, modes, medians, standard deviations, and 95% high posterior density intervals (95% HPD) were then calculated with '*Bayesian Output Analysis*' (BOA version 1.1.7, Smith 2003) for all parameters from the individual marginal posteriors, under the free-software R (R Development Core Team 2011).

Simulation example

To further illustrate the performance of the proposed model, a stochastic simulation study was carried out. In order to simplify the simulation procedure without loss of generality, data were generated according to a structure mimicking loblolly pine data, using the same design (e.g., the same number of families and trees per family) and similar variance components. Simulated data were investigated under 27 different scenarios varying the type of environmental heterogeneity (i.e., Small = small-scale variation, Large = large-scale variation, and Mixed = large-scale together with small-scale variation; see Supplementary Table S1). These spatial variations were simulated with a B-spline model, in a rectangular region of 25 rows and 40 columns with 10 knots in each dimension for the small-scale scenario, and 4 knots in each dimension for the large-scale. The Mixed scenario was obtained from the sum of a small-scale surface and a large-scale surface. Some of the parameters used in the simulation were fixed: $\sigma^2_{Ad} = 10$; $\sigma^2_b = 31.52$, $\sigma^2_p = 1$, and $\sigma^2_e = 5$. However, in view of the importance of the additive genetic variance for competition breeding values (σ^2_{Ac}) and the correlation between direct and competition breeding values (ρ_{AdAc}), we examine three values of these parameters. Specifically, we used the values of 1, 2.5, and 5 for σ^2_{Ac} which represent a 10, 25 and 50 percent, respectively, of the additive genetic variance for direct effects (σ^2_{Ad}). Since negatives correlations values are expected when strong competition for resources exist (e.g. Muir 2005), we used the values -0.3, -0.6, and -0.9 for ρ_{AdAc} representing low, moderate, and high levels of competition, respectively. Altogether, the three environmental scales of variations, three levels of genetic competition variance, and three values of correlation between direct and competition genetic effects resulted in $3^3 = 27$ scenarios. We produced 6 replicates for each scenario, giving rise to 162 simulated data sets. All data sets were simulated through the function breedR.sample.phenotype() of the R-package breedR (Muñoz and Sanchez, 2014). Finally, the simulated data were analyzed using the CSM (model [3]), and the simulated values were compared with the true values using the mean and the average relative deviation from the true value (bias) averaged over the 6 replicates. The average root mean squared relative error was used to measure the performance of the proposed joint CSM [3] in the different scenarios studied, according to the following formula:

$$\left[\frac{1}{6} \sum_{l=1}^{6} \left(\frac{S_l - T_l}{T_l} \right)^2 \right]^{1/2}$$

where S_l are the simulated values and T_l the true values for each dispersion parameters of model [3], i.e., $l = 1$, …, 6.

The hypervariances of all parameters were set equal to the true values used in the simulated data, and degrees of belief were set to 5 to reflect a relatively high degree of uncertainty. A single Gibbs chain of 510,000 samples were drawn, with first 10,000 iterations were discarded as *burn-in*.

Results

Results from the simulation

Table 1 shows the posterior means of each parameter under each of the 27 simulation scenarios. The deviations of the estimates relative to true values are shown in the Supplementary Table S2. In general, the CSM (Model 3) yielded posterior means of variance components with slight or negligible biases, except for the permanent environment variance that showed important systematic overestimations across scenarios. For the rest of variables, estimates of parameters of interest agree well with the true values (Table 1 and Supplementary Table S2). Variances of estimates (Table 1) over 6 replicates, were generally small with the exception of the permanent environment variance, particularly under small-scale and mixed-scale variation and lower competition additive genetic variance. The average root mean squared relative error varies from 0.33 to 2.18 for the different scenarios studied. Additionally, the marginal posterior (and prior) distributions of each parameter in each scenario are shown in the Supplementary Figure S1. A quick view of the figure clearly suggests that the posterior distributions have distinguishable modes for all of the parameters. The prior distributions follow the shape of the posteriors for all parameters; but, in general, the priors are at a lower value, reflecting the vague prior information characteristics induced by the degrees of freedom used (i.e., 5 in all parameters). In summary, the proposed CSM gives reliable inferential answers about unknown parameters in the model under the different scenarios studied.

Table 1. Posterior means (and standard deviations) of estimates of the direct additive genetic variance σ^2_{Ad}, the competition additive genetic variance σ^2_{Ac}, the correlation between direct and competition genetic effects ρ_{AdAc}, the variance of the B-spline coefficients σ^2_b, the permanent environmental variance σ^2_p and the residual variance σ^2_e by scenario, averaged over the 6 replicated simulations.

Scenario[1]	σ^2_{Ad}	σ^2_{Ac}	ρ_{AdAc}	σ^2_b	σ^2_p	σ^2_e
Small / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.3$	9.05 (1.42)	1.33 (0.55)	-0.29 (0.12)	26.82 (6.05)	2.28 (0.78)	5.14 (1.15)
Small / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.6$	9.10 (1.41)	1.02 (0.37)	-0.62 (0.12)	26.20 (5.79)	3.04 (0.73)	5.16 (1.17)
Small / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.9$	9.25 (1.18)	1.21 (0.29)	-0.87 (0.07)	25.44 (5.38)	2.19 (0.55)	5.18 (1.00)
Small / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.3$	9.84 (1.51)	2.86 (0.86)	-0.33 (0.10)	28.86 (6.93)	2.49 (0.90)	4.93 (1.21)
Small / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.6$	8.92 (1.39)	3.01 (0.79)	-0.55 (0.10)	32.39 (7.04)	2.26 (0.82)	5.12 (1.15)
Small / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.9$	9.58 (1.25)	2.77 (0.59)	-0.80 (0.06)	27.67 (5.82)	2.02 (0.62)	5.26 (1.06)
Small / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.3$	9.13 (1.64)	4.29 (1.05)	-0.19 (0.09)	25.77 (6.31)	1.97 (0.89)	5.84 (1.32)
Small / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.6$	8.88 (1.36)	5.37 (1.14)	-0.61 (0.08)	32.33 (7.30)	2.34 (0.97)	5.83 (1.15)
Small / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.9$	10.57 (1.16)	6.22 (0.98)	-0.77 (0.05)	25.04 (5.44)	1.92 (0.72)	4.04 (0.90)
Large / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.3$	9.51 (1.49)	0.84 (0.33)	-0.23 (0.14)	51.24 (23.14)	1.93 (0.60)	5.02 (1.17)
Large / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.6$	9.00 (1.39)	1.02 (0.34)	-0.62 (0.11)	42.82 (19.77)	2.19 (0.54)	5.37 (1.14)
Large / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.9$	9.45 (1.21)	0.91 (0.19)	-0.89 (0.06)	55.18 (24.72)	1.75 (0.40)	5.22 (0.99)
Large / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.3$	8.98 (1.58)	2.47 (0.71)	-0.28 (0.11)	38.06 (17.44)	1.60 (0.65)	5.91 (1.30)
Large / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.6$	8.90 (1.46)	3.13 (0.71)	-0.56 (0.09)	40.99 (18.94)	1.51 (0.58)	5.58 (1.21)
Large / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.9$	10.09 (1.26)	2.97 (0.54)	-0.83 (0.05)	34.07 (15.64)	1.25 (0.42)	5.08 (1.01)
Large / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.3$	9.03 (1.66)	4.79 (1.01)	-0.26 (0.09)	40.52 (19.09)	1.69 (0.78)	5.98 (1.35)
Large / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.6$	9.95 (1.32)	5.85 (1.01)	-0.62 (0.07)	36.05 (17.17)	1.95 (0.73)	4.52 (1.06)
Large / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.9$	9.87 (1.23)	5.42 (0.74)	-0.80 (0.05)	53.50 (24.15)	1.20 (0.47)	4.90 (0.97)
Mixed / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.3$	9.32 (1.46)	0.86 (0.36)	-0.19 (0.15)	42.25 (8.19)	2.59 (0.75)	4.95 (1.14)
Mixed / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.6$	8.76 (1.40)	1.06 (0.41)	-0.57 (0.13)	33.49 (7.02)	3.17 (0.77)	5.37 (1.17)
Mixed / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.9$	9.12 (1.24)	0.84 (0.20)	-0.86 (0.07)	38.54 (7.38)	2.58 (0.60)	5.29 (1.04)
Mixed / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.3$	8.79 (1.55)	2.65 (0.79)	-0.24 (0.11)	38.81 (7.83)	1.88 (0.79)	6.03 (1.26)
Mixed / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.6$	8.72 (1.48)	3.35 (0.90)	-0.51 (0.10)	40.56 (7.69)	2.14 (0.84)	5.58 (1.24)
Mixed / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.9$	9.47 (1.3)	3.16 (0.65)	-0.82 (0.06)	26.60 (5.52)	1.73 (0.60)	5.54 (1.09)
Mixed / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.3$	8.99 (1.63)	5.49 (1.32)	-0.21 (0.09)	31.62 (7.35)	2.30 (1.13)	5.77 (1.31)
Mixed / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.6$	8.98 (1.33)	6.61 (1.26)	-0.62 (0.07)	33.56 (7.11)	2.21 (0.92)	5.57 (1.11)
Mixed / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.9$	9.47 (1.21)	6.03 (0.95)	-0.77 (0.05)	38.12 (6.98)	1.70 (0.67)	4.85 (0.98)

Note:

[1] Small corresponds to small-scale environmental variation; Large correspond to large-scale environmental variation; Mixed correspond to small-scale together with large-scale variation.

Results from the real data analysis of loblolly pines

Lack of convergence of the Gibbs sampler was not detected by inspection of trace plots of all unknowns parameters from the four models evaluated in the real data set (Supplementary Figure S2). When comparing the four models, it is clear that adding competition and/or environmental heterogeneity effects improved the fit compared to TM. This can be seen in Table 2 *via* the DIC, for which CSM resulted in the lowest value, followed by CM and SM. The CM showed a greater reduction in DIC than SM, most likely due to the fact that competition had a higher effect than the small-scale environmental heterogeneity in this loblolly pine data set.

Table 2: Deviance Information Criterion (DIC) and posterior statistics for the direct additive genetic variance σ^2_{Ad}, the competition additive genetic variance σ^2_{Ac}, the estimated correlation between direct and competition genetic effects ρ_{AdAc}, the variance of commercial seed lots σ^2_t, the variance of the B-spline coefficients σ^2_b, the permanent environmental variance σ^2_p, and the residual variance σ^2_e for diameter at breast height in the loblolly pine data set and from the four models evaluated.

Model[a]	DIC	Parm.[b]	Mean	Median	Mode	SD[c]	95% HPD[d]	ESS[e]
TM	2686.24	σ^2_{Ad}	5.76	5.57	4.74	1.52	3.62 – 8.54	17,512
		σ^2_t	20.72	18.80	15.39	8.74	10.66 – 37.33	294,926
		σ^2_e	12.44	12.52	12.55	1.41	10.00 – 14.63	22,268
CM	2515.20	σ^2_{Ad}	8.04	7.98	7.95	1.43	5.78 – 10.48	15,864
		σ^2_{Ac}	1.56	1.54	1.39	0.33	1.08 – 2.14	15,877
		ρ_{AdAc}	-0.85	-0.86	-0.87	0.05	-0.91 - -0.76	16,694
		σ^2_p	1.21	1.17	1.20	0.30	0.79 – 1.76	21,008
		σ^2_t	19.39	17.59	14.85	8.19	9.98 – 34.96	268,454
		σ^2_e	8.25	8.22	7.63	1.26	6.23 – 10.36	17,351
SP	2593.88	σ^2_{Ad}	8.52	8.34	8.32	2.22	5.18 – 12.46	15,279
		σ^2_b	3.00	2.87	2.92	0.88	1.83 – 4.63	46,867
		σ^2_t	16.30	14.80	11.32	6.80	8.46 – 29.29	314,012
		σ^2_e	10.19	10.26	10.84	1.85	7.02 – 13.11	16,394
CSM	2481.25	σ^2_{Ad}	8.75	8.70	10.09	1.52	6.35 – 11.33	15,681
		σ^2_{Ac}	1.62	1.59	1.77	0.33	1.12 – 2.19	16,796
		ρ_{AdAc}	-0.85	-0.86	-0.86	0.05	-0.92 - -0.77	16,694
		σ^2_b	2.88	2.75	2.25	0.83	1.78 – 4.41	47,504
		σ^2_p	0.98	0.95	1.05	0.24	0.65 – 1.42	21,260
		σ^2_t	19.45	17.65	13.13	8.20	10.01 – 35.00	267,294
		σ^2_e	7.54	7.51	6.17	1.27	5.51 – 9.67	16,461

Note:
[a] **TM (Standard model)**: Individual-tree mixed model does not account for competition effects and continuous environmental heterogeneity (only fit blocks fitted as fixed effects).
CM (Competition model): Individual-tree mixed model that accounts competition effects.
SP (Spatial model): Individual-tree mixed model that accounts environmental heterogeneity using P-splines with 15 knots for rows and 15 knots for columns.
CSM (Competition + Spatial model): Individual-tree mixed model that accounts competition effects and environmental heterogeneity using P-splines with 15 knots for rows and 15 knots for columns.
[b] **Parm.** = Parameter.
[c] **SD** = standard deviation.
[d] **HPD =** high posterior density interval.
[e] **ESS =** effective sample size.

Posterior summaries for σ^2_{Ad}, σ^2_{Ac}, ρ_{AdAc}, σ^2_b, σ^2_p, σ^2_t and σ^2_e are shown in Table 2. Posterior means, medians, and modes were quite similar, except for σ^2_t, and σ^2_e from CSM. The 95% high posterior density intervals for these variance components were shifted away from zero, which reflects the constraint of the priors but also the fact that all the effects in these models had detectable variation. TM resulted in the

lowest component of σ^2_{Ad} among all 4 models, with a corresponding maximum value under CSM. Conversely, σ^2_e was highest for TM and lowest for CSM, and differences between both can be easily labeled as competition and environmental heterogeneity components. Relative to TM, the reduction in σ^2_e was much larger for the CM (34%) than for the SM (18%). This further confirms the observation that competition effects predominate over the small-scale environmental heterogeneity in this loblolly pine trial. Although smaller when compared with the estimates of σ^2_{Ad}, the estimates of σ^2_{Ac} were similar in CM (1.56) and CSM (1.62). The marginal posterior mean of ρ_{AdAc} from CM was large and negative -0.85, and the SD was 0.05, revealing a strong competition between additive genetic effects. While CSM showed similar marginal posterior mean of ρ_{AdAc} to CM, the estimate of σ^2_p was smaller in the former (1.21 and 0.98, respectively). The estimate of σ^2_b from SM (3.00) was slightly higher than the corresponding values of CSM (2.88).

The Spearman rank correlations between offspring predicted breeding values from TM, and the direct and total breeding values from CSM (i.e., the one with the smallest DIC) were high and positive (0.96 and 0.70, respectively).

Discussion

This research presents an extension of the method of Cappa and Cantet (2008) to account for environmental heterogeneity within genetic trials using a smoothing surface. This extension produces a more realistic model, making possible to fit simultaneously competition effects and spatial environmental variability. To do this, we included in the individual-tree model with competition effects proposed by Cappa and Cantet (2008) a surface that is smoothed in the direction of both the columns and the rows, to account for large- and/or small-scale environmental variation.

There are other approaches that simultaneously dealt with competition and spatial modelling. Stringer et al. (2011) developed a joint model to account for interplot competition in one dimension and spatial variability. They fitted the "treatment interference model" presented by Besag and Kempton (1986) and originally proposed by Pearce (1957), to model the genetic competition as a random effect. In this approach, each treatment is assumed to have a direct effect and a neighbor effect on adjacent plots. In other words, the average genotypic value is associated with the nearest neighboring plots of many plants, rather than with the observed individual-tree data. Moreover, it does not account for the genetic relationship between units (plots or trees). If selection is designed to select genotypes for the next generation of the breeding population, the most convenient way is to add a random genetic effect that accounts for the additive relationship between individuals, as we have proposed here. Additionally, our approach accounts for unequal numbers of neighbors due to mortality or border location, and for different competition intensity between the trees in the different spatial directions (i.e., row, column and diagonal), by incorporating the IC elements (Cappa and Cantet 2008). Recently, Costa e Silva et al. (2013) applied the competition individual-tree mixed model proposed by Cappa and Cantet (2008) and the covariance residual structure EAR3 after Stringer et al.

(2011), to jointly model competition effects and local environmental heterogeneity of a large progeny trial of *Eucalytus globulus*. Different from the original formulation, Costa e Silva et al. (2013) modeled the large-scale environmental variation using the mixed model proposed by Verbyla et al. (1999) to fit a smoothing spline in row and/or column directions only.

In the present study, a two-dimensional surface using the tensor product of B-splines bases (Marx and Eilers 2005) was used to model both global- and small-scale spatial variation. The literature is extremely limited on the application of a tensor product of B-splines bases to capture both kinds of spatial heterogeneity in forest genetic evaluation (Cappa and Cantet 2007; Cappa et al. 2011). In research that is not yet published, we observed the utility of the approach in accommodating complex patterns of spatial heterogeneity in several large forest genetic trials of western hemlock (*Tsuga heterophylla* (Raf.) Sarg.). Stringer et al. (2011) proposed to account for global trend by including design factors based on row and column coordinates, or by fitting in one dimension either low order polynomials or cubic smoothing splines (Verbyla et al. 1999). Durban et al. (2001) also proposed to model the large-scale fertility variation using cubic smoothing splines in one dimension (Green and Silverman 1994). Nevertheless, in forest genetic trials where trees are planted in squares or rectangles, a large portion of the global trend is usually present in two dimensions, and one-dimensional polynomials or cubic smoothing splines may not completely account for spatial covariance (Cappa and Cantet 2007). Moreover, it is extremely rare to find large-scale continuous spatial variability either in the direction of the rows or of the columns, and some sort of interaction between rows and columns has to be considered to account for such variability (Federer 1998). Additionally, polynomials do a poor job when fitting observations in the extremes, and small changes in the data produce a dramatic effect in the estimated values of the parameters, and this is specially so for polynomials of higher degree.

As interacting trees share the same environment when competition and environmental heterogeneity are present, there is a risk of confounding environmental competition and local trend effects (Bijma 2013). Following Cappa and Cantet (2008) (model [1]), the proposed model [3] included permanent competition effects through a random variable (p_c) to model a competition effect at the environmental or residual level. To model the permanent environmental competition effects and local environmental trend, Stringer et al. (2011) proposed including an autoregressive residual process EAR2 or EAR3. The EAR3 has two parameters, one nominally representing the local trend, the other representing the environmental competition. They recommended modeling the permanent environmental competition using an EAR2 in trials where competition was dominant, while an EAR3 allowed for competition and local trend variability. However, when the permanent environmental competition seemed to be dominant over small-scale trend, Costa e Silva and Kerr (2013) suggested that the EAR2 did not appear to produce a better fit than the AR(1) in preliminary analysis of diameter growth from forest genetic trials. Instead, in an empirical study of the trait DBH, Costa e Silva et al. (2013) used a separable EAR3 residual structure in the column direction. However, when the environmental trend was stronger than the permanent environmental competition and the full model was fitted (i.e., a mixed model including genetic and environmental competition effects and large - and small - environmental trend), the autocorrelation coefficient for columns was estimated with a large standard error.

There seems to be a problem of residual confounding in this analysis, a ubiquitous process in mixed model inference with errors following autoregressive processes (Gustafson and Greenland 2006; Paciorek 2010). The issue can be compounded by the fact that models including autoregressive processes are prone to bias, especially if the series is short (Shaman and Stine 1988). In these conditions, fitting permanent environmental effects as random variables, apart from the error term, and with an informative covariance structure between direct and competition breeding values seems to be advantageous. The procedure mimics the fitting of a prior distribution to avoid residual confounding as suggested by Gustafson and Greenland (2006). In their words, "the benefit from modeling residual confounding is maintained when the prior distributions employed only roughly correspond to reality". Additionally, while the competition additive genetic effects extend over the whole area of the trial, the environmental permanent competition effects act locally, i.e. in the nearest eight neighbors. In our approach, every row of the incidence matrix Z_p has all elements equal to zero except for a 1 in the column belonging to a_{pj} of the nearest j neighbor competitor tree. However, the EAR3 residual structure could be unnecessary in most cases, given that it is unlikely that the environmental competition effects be spatially extended in the direction of columns or/and rows beyond of first- or second-order competitors.

There are few empirical studies that combine competition and spatial variability in agronomic crops (Durban et al. 2001; Stringer et al. 2011; Hunt et al. 2013) and forest genetic trials (Magnussen 1994; Resende et al. 2005 and Costa e Silva et al. 2013). When the root mean squared relative error was calculated to evaluate the performance of the proposed joint model in the different scenarios studied, lower values were found for those scenarios with large-scale environmental heterogeneity and stronger competition genetic effects (i.e., highest and negative correlation between direct and competitive additive genetic effects). On the contrary, the highest values were observed for those scenarios with small-scale environmental heterogeneity and weak competition effects. These results suggest that the performance of the proposed model under situations with weak competition effects and environmental heterogeneity operating at the same spatial scales (i.e., small-scale) may be somewhat limited. In the empirical analysis reported here, conditions *a priori* were more favorable than the worst-case scenario from simulations, with the former showing strong competition and small-scale environmental heterogeneity. In these circumstances, the proposed CSM showed better fit (i.e., smallest DIC) than simpler models with no competition and spatial continuous effects, or either of these effects alone (i.e., TM, CM, and SM, Table 2). Similar results have been found in crop (Stringer et al. 2011; Hunt et al. 2013) and forest species (Resende et al. 2005 and Costa e Silva et al. 2013), when competition and environmental spatial effects are present regardless of whether the spatial trend predominated over competition or *vice-versa*. However, different traits responded differently to competition. The substantial improvement in the fit from CSM was followed by CM, and this may be attributed to the fact that DBH is more affected by competition than, for example, the height trait (Hunnrup et al. 1998; Dutkowski et al. 2006; Ye and Jayawickrama 2008).

Our results from the simulation data using the proposed CSM show that the permanent environment variance estimates were biased upwards in most scenarios. However, it is important to note that this component accounts for a very small part of the variation. We further investigated the possible causes of this

bias, and one possible explanation is the fact that there were slight differences between the incidence matrices used to simulate the data and to analyze the data. Another possible source of this bias is the simple form of the covariance matrix used for these permanent environmental effects, i.e., an identity matrix. This simple structure specifies that these environment effects are independent. A more complex and informative covariance structure for this environmental parameter will be considered in a future work.

Our empirical results show that not fitting the detected competition effects (i.e., TM or SM) and/or small-scale trend (i.e., TM or CM) resulted in a consistent increase in the posterior mean of σ_e^2 and a decrease in the posterior mean of σ_{Ad}^2 (Table 2). Costa e Silva and Kerr (2013) have observed a similar phenomenon when comparing TM with the CM proposed by Cappa and Cantet (2008). This was observed from simulated data where competition was the most important source of residual autocorrelation, and the environmental heterogeneity was negligible so that it was not modeled. These results are also in agreement with those of Brotherstone et al. (2011) who studied the diameter trait in a 19-year old Sitka spruce clonal trial growing in Scotland. They fitted the simple competition model proposed by Bijma (2007) jointly with an AR(1) × AR(1) covariance structure for the residual. There was evidence of a global trend in the latter model, but this effect was not formally accounted for.

In the current research, where the empirical data shows strong competition effects, ignoring the genetic and environmental competition effects leads to overestimation of environmental heterogeneity; i.e., the spatial model yielded estimates of σ_b^2 that were higher than those of CSM (3.00 vs. 2.88, respectively). While CM yielded estimates of ρ_{AdAc} that were equal to those of CSM (-0.85), ignoring the environmental heterogeneity led to overestimation of the environmental competition effects; i.e., CM yielded higher σ_p^2 than the estimate from CSM (1.21 vs. 0.98, respectively). This is the first study applying the competition model proposed by Cappa and Cantet (2008) plus a two-dimensional smoothing surface to account for environmental heterogeneity in forest genetic trials. Therefore, it is not possible to compare the variance estimates σ_b^2 and σ_p^2 with estimates from other studies.

Trees may compete at an early stage for water and nutrients, and after canopy closure they compete mainly for light (Brotherstone et al. 2011). In small row-plot designs (as with the five-tree row-plot) the sharing of these resources could be different, because the interacting plants are genetically different (Stanger et al. 2011). In this work, a maximum of eight first-order neighbor trees were used in both CM and CSM models. However, the larger negative autocorrelation coefficient for row (-0.34) than for column (-0.10) identified in the first step of the analysis, suggests a slight asymmetric competition. When an asymmetric genetic competition effect was fitted in the CM and CSM models, i.e., when considering only the maximum of two row-neighbor trees to calculate the ICs, the values of DIC were higher (2571.97 and 2563.04, respectively) than the respective values considering all the eight first-order neighbors trees (Table 2). Moreover, posterior means of the covariance components for the CM model fitted only with the row-neighbor trees were σ_{Ad}^2 = 5.87, σ_{Ac}^2 = 1.52, ρ_{AdAc} = -0.80, σ_p^2 = 1.64, σ_t^2 = 19.23 and σ_e^2 = 9.71. Notice that the

estimate of σ^2_{Ad} was smaller (5.87 *vs*. 8.04) and the estimate of σ^2_e was larger (9.71 *vs*. 8.25) than when the eight first-order neighbors were fitted. Similar results were obtained for the CSM. Therefore, this result suggests that, if left out of the analysis, those competitors may bias the predictions of the direct breeding values.

From the tree breeder's viewpoint, a relevant question is whether or not the predicted tree BV from TM and the predicted direct and total tree BV from CSM lead to a different ranking. In our study, rank correlations of predicted BV were high, but not perfect. Assuming 100 % survival and using simulated data with a similar additive covariance matrix of breeding values (i.e., $\sigma^2_{Ad} = 20$; $\sigma^2_{Ac} = 2$; $\rho_{AdAc} = -0.9$), Costa e Silva and Kerr (2013), also found that the Spearman correlations were high and positive between predicted breeding values from SM and direct (0.99) and total breeding values (0.79) from a model including both competition and spatial effects. Although our empirical study showed high correlations between predicted tree BVs, the ranking among the top 5% (47) individuals from TM and CSM displayed some differences. The proportion of common individuals within the top 47 trees (5%) was 0.83 between the BV of TM and direct BV from CSM, and 0.34 between the BVs of TM and total BV from CSM. These differences demonstrate that using overly simplified models may substantially compromise selection decisions and genetic progress.

Summary and Conclusions

A novel Bayesian approach to effectively model simultaneously and effectively two of the sources of potential bias in forest genetic trials - the genetic and environmental competition and environmental heterogeneity - was developed in the current research. Models that were previously suggested for dealing with competition and spatial environmental variation neglect some aspect, like the additive relationships among trees, or the full spatial covariance. The fitting of a more complex model to accommodate both competition and environmental heterogeneity effects was illustrated using simulated data and also real data of diameter at age 13, from an open-pollinated progeny trial of loblolly pine. The simulation and the real data example showed the importance of simultaneously accounting for competition and environmental heterogeneity effects, and make it possible to understand the dynamics of both phenomena and their effects on the estimation of genetic parameters and the prediction of breeding values in single forest genetics trials. In particular, the real data analysis carried out with the proposed joint model showed that: a) by not fitting the detected competition and the small-scale trend resulted in a consistent increase in the residual variance and in a decrease in the estimated direct genetic variance; b) correlations between predicted tree BV from TM and CSM were high; however, the ranking among the top 5% ranked individuals showed differences which indicated that the two models will have quite different genotype selections for the next cycle of breeding; and c) total tree breeding values led to more realistic results.

Data Archiving Statement

We followed standard Tree Genetics and Genomes policy. Simulated data set used in this manuscript will be made available in the Zenodo repository, http://dx.doi.org/10.5281/zenodo.32036. Supplementary information of the Pinus taeda L. trial and family numbers will be also made available in the Zenodo repository, http://dx.doi.org/10.5281/zenodo.32040. In addition, diameter at breast height of the *Pinus taeda* L. data set will be available upon request.

References

Besag J, and Kempton R (1986) Statistical analysis of field experiments using neighbouring plots. Biometrics **42**: 231-251.

Bijma P, Muir WM, Van Arendonk JAM (2007) Multilevel selection 1: quantitative genetics of inheritance and response to selection. Genetics **175**: 277–288.

Bijma P (2013) The quantitative genetics of indirect genetic effects: a selective review of modelling issues. Heredity doi: 10.1038/hdy.2013.15.

Brotherstone S, White IMS, Sykes R, Thompson R, Connolly T, Lee S, Woolliams J (2011) Competition effects in a young sitka spruce (*Picea sitchensis*, Bong. Carr) Clonal Trial. Silvae Genet. **60**: 149–155.

Cappa EP, Cantet RJC (2007) Bayesian estimation of a surface to account for a spatial trend using penalized splines in an individual-tree mixed model. Can. J. For. Res. **37**: 2677-2688.

Cappa EP, Cantet RJC (2008) Direct and competition additive effects in tree breeding: Bayesian estimation from an individual tree mixed model. Silvae Genet. **57**: 45-56.

Cappa EP, Lstiburek M, Yanchuk AD, El-Kassaby YA (2011) Two-dimensional penalized splines via Gibbs sampling to account for spatial variability in forest genetic trials with small amount of information available. Silvae Genet. **60**: 25-35.

Costa e Silva J, Dutkowski GW, Gilmour AR (2001) Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. Can. J. For. Res. **31**: 1887-1893.

Costa e Silva J, Kerr RJ (2013) Accounting for competition in genetic analysis, with particular emphasis on forest genetic trials. Tree Genetics & Genomes **9**: 1-17.

Costa e Silva J, Potts BM, Bijma P, Kerr RJ, Pilbeam DJ (2013) Genetic control of interactions among individuals: contrasting outcomes of indirect genetic effects arising from neighbour disease infection and competition in a forest tree. New Phytol. **197**: 631-41.

De Boor C (1993) B(asic)-spline basics. Fundamental Developments of Computer-Aided Geometric Modeling. L. Piegl, ed. Academic Press, San Diego, CA.

Durban M, Currie I, Kempton R (2001) Adjusting for fertility and competition in variety trials. J. Agric. Sci. (Camb.) **136**: 129–140.

Dutkowski GW, Costa e Silva J, Gilmour AR, Wellendorf H, Aguiar A (2006) Spatial analysis enhances modeling of a wide variety of traits in forest genetic trials. Can. J. For. Res. **36**: 1851-1870.

Eilers PHC, Marx BD (2003) Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. Chemometr. Intell. Lab. Syst. **66**: 159-174.

Ericsson T (1997) Enhanced heritabilities and best linear unbiased predictors through appropriate blocking of progeny trials. Can. J. For. Res. **27**: 2097–2101.

Federer WT (1998) Recovery of interblock, intergradient, and intervarietal information in incomplete block and lattice rectangle designed experiments. Biometrics **54**: 471-481.

Finley AO, Banerjee S, Waldmann P, Ericsson T (2009) Hierarchical spatial modeling of additive and dominance genetic variance for large spatial trial datasets. Biometrics:65, 441–451.

Geweke J (1992) Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM(eds) Bayesian statistics 4. Oxford University Press, Oxford.

Gilmour AR, Cullis BR, Verbyla AP (1997) Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments. J. Agric. Biol. Environ. Stat. **2**: 269-293.

Gilmour, A.R., Gogel, BJ, Cullis, B.R., and Thompson, R. 2006. ASReml User Guide Release 2.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. p 267.

Green P J, Silverman BW (1994) Nonparametric Regression and Generalized Linear Model. Chapman & Hall, London, UK.

Gustafson P, Greenland S (2006) The Performance of Random Coefficient Regression in Accounting for Residual Confounding. Biometrics **62**:760–768.

Hannrup B, Wilhelmsson L, Danell Ö (1998) Time Trends for Genetic Parameters of Wood Density and Growth Traits in *Pinus sylvestris* L. Silvae Genetic. **47**: 214-219.

Harville DA (1997) Matrix algebra from a statistician's perspective. Springer-Verlag. New York.

Henderson CR (1984) Applications of Linear Models in Animal Breeding. Canada, University of Guelph, Guelph, Ont.

Hinson K, Hanson WD (1962) Competition studies in soybeans. Crop. Sci. **2**: 117-123.

Hobert JP, Casella G (1996) The effects of improper priors on Gibbs sampling in hierarchical linear models. J. Amer. Statist. **91**: 1461-1473.

Hunt CH, Smith AB, Jordan DR, Cullis BR (2013) Predicting additive and non-additive genetic effects from trials where traits are affected by interplot competition. J. Agric. Biol. and Envir. S. **18**: 53-63.

Kass RE, Carlin BP, Gelman A, Neal RM (1998) Markov chain Monte Carlo in practice: a roundtable discussion. Amer. Stat. **52**: 93-100.

Kusnandar D (2001) The Identification and Interpretation of Genetic Variation in Forestry Plantation. PhD Thesis, University of Western Australia, Crawley, Australia.

Magnussen S (1989) Effects and adjustments of competition bias in progeny trials with single-tree plots. Forest Science **35**: 532-547.

Magnussen S (1990) Application and comparison of spatial models in analysing tree-genetics field trials. Can. J. For. Res. **20**: 536-546.

Magnussen S (1994) A method to adjust simultaneously for statial microsite and competition effects. Can. J. For. Res. **24**: 985-995.

Marx BD, Eilers PH (2005) Multidimensional penalized signal regression. Technometrics **47**: 13-22.

Muñoz F, Sanchez L (2014) breedR: Statistical methods for forest genetic resources analysts. R package version 0.7-16. https://github.com/famuvie/breedR

Paciorek CJ (2010) The Importance of Scale for Spatial-Confounding Bias and Precision of Spatial Regression Estimators. Statistical Science **25**: 107-125.

Pearce SC (1957) Experimenting with organisms as blocks. Biometrika, **44**: 141-149.

Radtke PJ, Westfall JA, Burkhart HE (2003) Conditioning a distance-dependent competition index to indicate the onset of intertree competition. Forest Ecology and Management **175**: 17-30.

Resende MDV, Stringer J, Cullis B, Thompson R (2005) Joint modelling of competition and spatial variability in forest field trials. Rev. Mat. Estat. **23**: 7-22.

Robert CP, Casella GC (1999) Monte Carlo statistical methods. Springer-Verlag, New York, USA.

Shaman P, Stine RA (1988) The Bias of Autoregressive Coefficient Estimators. J. of the Amer. Statist. Assoc. **83**: 842-848.

Smith BJ (2003) Bayesian Output Analysis Program (BOA) version 1.0 user's manual. Available from http://www.public-health.uiowa.edu/boa/.

Sorensen D, Gianola D (2002) Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. Springer-Verlag, New York.

Stanger TK, Galloway GM, Retief ECL (2011) Final results from a trial to test the effect of plot size on *Eucalyptus* hybrid clonal ranking in coastal Zululand, South Africa. Southern Forests, 73: 131–135.

Stringer JK, Cullis BR, Thompson R (2005) Joint modelling of spatial variability and interplot competition to improve the efficiency of plant improvement. In Proceedings of the: Thredbo Statistical Meeting. Canberra, 6 – 11 February 2005. Australian National University/Australasian Region of the International Biometric Society v. 1, p.41.

Stringer JK, Cullis BR, Thompson R (2011) Joint modeling of spatial variability and within-row interplot competition to increase the efficiency of plant improvement. J. Agric. Biol. Environ. Stat. **16**: 269–281.

Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). Journal of the Royal Statistical Society Series B **64**: 583-639.

Thomson AJ, El-Kassaby YA (1988) Trend surface analysis of a Douglas-fir provenace-progeny transfer test. Can. J. For. Res. **18**: 515-520.

Verbyla AP, Cullis BR, Kenward MG, Welham SJ (1999) The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). Applied Statistics **48**: 69-311.

Ye TZ, Jayawickrama KJS (2008) Efficiency of using spatial analysis in firest-generation coastal Douglas-fir progeny tests in the US Pacific Northwest. Tree Genetics & Genomes **4**: 677-692.

Zas R (2006) Iterative kriging for removing spatial autocorrelation in analysis of forest genetic trials. Tree Genetics & Genomes **2**: 177-185.

Supplementary material

Table S1. Combination of environmental heterogeneity and competition true parameters used in the simulated example (see caption of Table 1 for dispersion parameters' abbreviations).

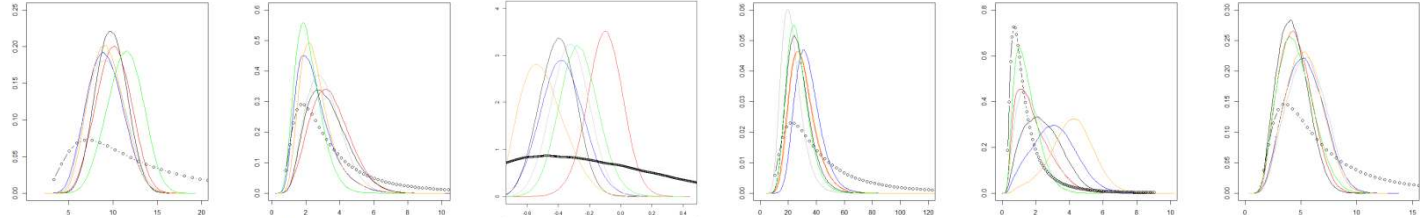| Scenarios	Environmental heterogeneity[1]	Competition[2]	σ^2_{Ad}	σ^2_{Ac}	ρ_{AdAc}	σ^2_b	σ^2_p	σ^2_e
1	Small	+	10	1	-0.3	31.52	1	5
2	Small	++	10	1	-0.6	31.52	1	5
3	Small	+++	10	1	-0.9	31.52	1	5
4	Small	+	10	2.5	-0.3	31.52	1	5
5	Small	++	10	2.5	-0.6	31.52	1	5
6	Small	+++	10	2.5	-0.9	31.52	1	5
7	Small	+	10	5	-0.3	31.52	1	5
8	Small	++	10	5	-0.6	31.52	1	5
9	Small	+++	10	5	-0.9	31.52	1	5
10	Large	+	10	1	-0.3	31.52	1	5
11	Large	++	10	1	-0.6	31.52	1	5
12	Large	+++	10	1	-0.9	31.52	1	5
13	Large	+	10	2.5	-0.3	31.52	1	5
14	Large	++	10	2.5	-0.6	31.52	1	5
15	Large	+++	10	2.5	-0.9	31.52	1	5
16	Large	+	10	5	-0.3	31.52	1	5
17	Large	++	10	5	-0.6	31.52	1	5
18	Large	+++	10	5	-0.9	31.52	1	5
19	Mixed	+	10	1	-0.3	31.52	1	5
20	Mixed	++	10	1	-0.6	31.52	1	5
21	Mixed	+++	10	1	-0.9	31.52	1	5
22	Mixed	+	10	2.5	-0.3	31.52	1	5
23	Mixed	++	10	2.5	-0.6	31.52	1	5
24	Mixed	+++	10	2.5	-0.9	31.52	1	5
25	Mixed	+	10	5	-0.3	31.52	1	5
26	Mixed	++	10	5	-0.6	31.52	1	5
27	Mixed	+++	10	5	-0.9	31.52	1	5

Note:
[1] Small corresponds to small-scale environmental variation; Large correspond to large-scale environmental variation; Mixed correspond to small-scale together with large-scale variation.
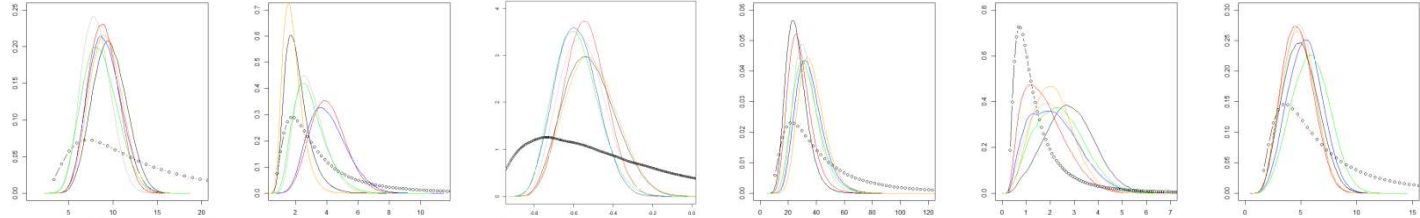[2] Intensity of competition based on the information from ρ_{AdAc} : + low; ++ medium; +++ strong.

Table S2. Posterior means of the relative deviations for the direct additive genetic variance σ^2_{Ad}, the competition additive genetic variance σ^2_{Ac}, the correlation between direct and competition genetic effects ρ_{AdAc}, the variance of the B-spline coefficients σ^2_b, the permanent environmental variance σ^2_p and the residual variance σ^2_e by scenario, averaged over the 6 replicated simulations.

Scenario[1]	σ^2_{Ad}	σ^2_{Ac}	ρ_{AdAc}	σ^2_b	σ^2_p	σ^2_e
Small / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.3$	-0.095	0.327	-0.048	-0.149	1.281	0.028
Small / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.6$	-0.090	0.022	0.037	-0.169	2.045	0.031
Small / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.9$	-0.075	0.209	-0.029	-0.193	1.191	0.036
Small / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.3$	-0.016	0.142	0.093	-0.085	1.493	-0.014
Small / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.6$	-0.108	0.204	-0.086	0.028	1.257	0.024
Small / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.9$	-0.042	0.107	-0.107	-0.122	1.020	0.052
Small / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.3$	-0.087	-0.142	-0.366	-0.182	0.973	0.168
Small / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.6$	-0.112	0.075	0.012	0.026	1.343	0.166
Small / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.9$	0.053	0.280	-0.175	-0.168	0.555	-0.162
Large / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.3$	-0.049	-0.164	-0.219	0.625	0.926	0.003
Large / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.6$	-0.100	0.017	0.041	0.359	1.187	0.074
Large / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.9$	-0.055	-0.086	-0.015	0.751	0.746	0.045
Large / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.3$	-0.102	-0.011	-0.076	0.207	0.601	0.183
Large / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.6$	-0.110	0.252	-0.069	0.300	0.512	0.117
Large / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.9$	0.009	0.190	-0.075	0.081	0.248	0.016
Large / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.3$	-0.097	-0.041	-0.147	0.286	0.686	0.196
Large / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.6$	-0.005	0.171	0.034	0.144	0.953	-0.095
Large / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.9$	-0.013	0.084	-0.116	0.697	0.204	-0.020
Mixed / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.3$	-0.068	-0.145	-0.360	0.340	1.593	-0.010
Mixed / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.6$	-0.124	0.060	-0.054	0.063	2.172	0.075
Mixed / $\sigma^2_{Ac}=1$ / $\rho_{AdAc}=-0.9$	-0.088	-0.155	-0.043	0.223	1.576	0.058
Mixed / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.3$	-0.121	0.062	-0.213	0.231	0.884	0.206
Mixed / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.6$	-0.128	0.338	-0.158	0.287	1.140	0.115
Mixed / $\sigma^2_{Ac}=2.5$ / $\rho_{AdAc}=-0.9$	-0.053	0.266	-0.089	-0.156	0.733	0.108
Mixed / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.3$	-0.101	0.098	-0.311	0.003	1.297	0.155
Mixed / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.6$	-0.102	0.321	0.031	0.065	1.207	0.113
Mixed / $\sigma^2_{Ac}=5$ / $\rho_{AdAc}=-0.9$	-0.053	0.207	-0.147	0.209	0.697	-0.030

Note:
[1] Small corresponds to small-scale environmental variation; Large correspond to large-scale environmental variation; Mixed correspond to small-scale together with large-scale variation.

Figure S1. Marginal posterior densities by scenario for direct additive variance σ^2_{Ad}, the competition additive genetic variance σ^2_{Ac}, the correlation between direct and competition genetic effects ρ_{AdAc}, the variance of the B-spline coefficients σ^2_b, the permanent environmental variance σ^2_p and the residual variance σ^2_e. The line represents each of the six replicates generated for each scenario and parameter. Prior distributions are shown as dotted lines.

Scenario[1]	σ^2_{Ad}	σ^2_{Ac}	ρ_{AdAc}	σ^2_b	σ^2_p	σ^2_e
Small / σ^2_{Ac} = 1 / ρ_{AdAc} =-0.3						
Small / σ^2_{Ac} = 1 / ρ_{AdAc} =-0.6						
Small / σ^2_{Ac} = 1 / ρ_{AdAc} =-0.9						

Small /σ^2_{Ac} = 2.5 /ρ_{AdAc} =-0.3

Small /σ^2_{Ac} = 2.5 /ρ_{AdAc} =-0.6

Small /σ^2_{Ac} = 2.5 /ρ_{AdAc} =-0.9

Small /σ^2_{Ac} = 5 /ρ_{AdAc} =-0.3

Small /σ^2_{Ac} = 5 /ρ_{AdAc} =-0.6

Small / σ^2_{Ac} **= 5 /** ρ_{AdAc} **=-0.9**

Large / σ^2_{Ac} **= 1 /** ρ_{AdAc} **=-0.3**

Large / σ^2_{Ac} **= 1 /** ρ_{AdAc} **=-0.6**

Large / σ^2_{Ac} **= 1 /** ρ_{AdAc} **=-0.9**

Large / σ^2_{Ac} **= 2.5 /** ρ_{AdAc} **=-0.3**

Large / σ^2_{Ac} = 2.5 / ρ_{AdAc} =-0.6

Large / σ^2_{Ac} = 2.5 / ρ_{AdAc} =-0.9

Large / σ^2_{Ac} = 5 / ρ_{AdAc} =-0.3

Large / σ^2_{Ac} = 5 / ρ_{AdAc} =-0.6

Large / σ^2_{Ac} = 5 / ρ_{AdAc} =-0.9

Mixed / $\sigma^2_{Ac} =$ **1 /** ρ_{AdAc} **=-0.3**

Mixed / $\sigma^2_{Ac} =$ **1 /** ρ_{AdAc} **=-0.6**

Mixed / $\sigma^2_{Ac} =$ **1 /** ρ_{AdAc} **=-0.9**

Mixed / $\sigma^2_{Ac} =$ **2.5 /** ρ_{AdAc} **=-0.3**

Mixed / $\sigma^2_{Ac} =$ **2.5 /** ρ_{AdAc} **=-0.6**

Mixed /σ^2_{Ac} = 2.5 /ρ_{AdAc} =-0.9

Mixed /σ^2_{Ac} = 5 /ρ_{AdAc} =-0.3

Mixed /σ^2_{Ac} = 5 /ρ_{AdAc} =-0.6

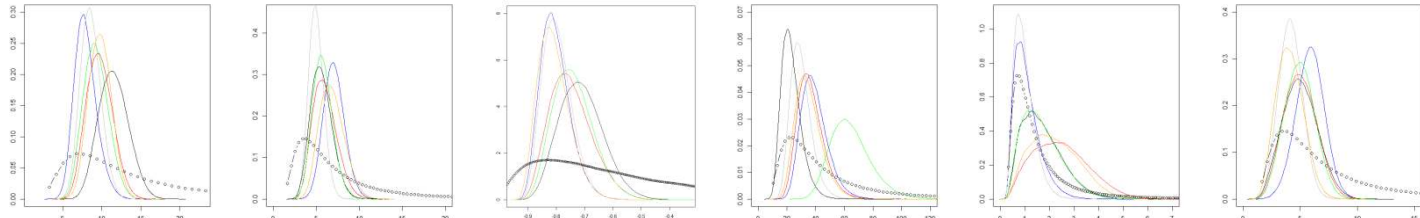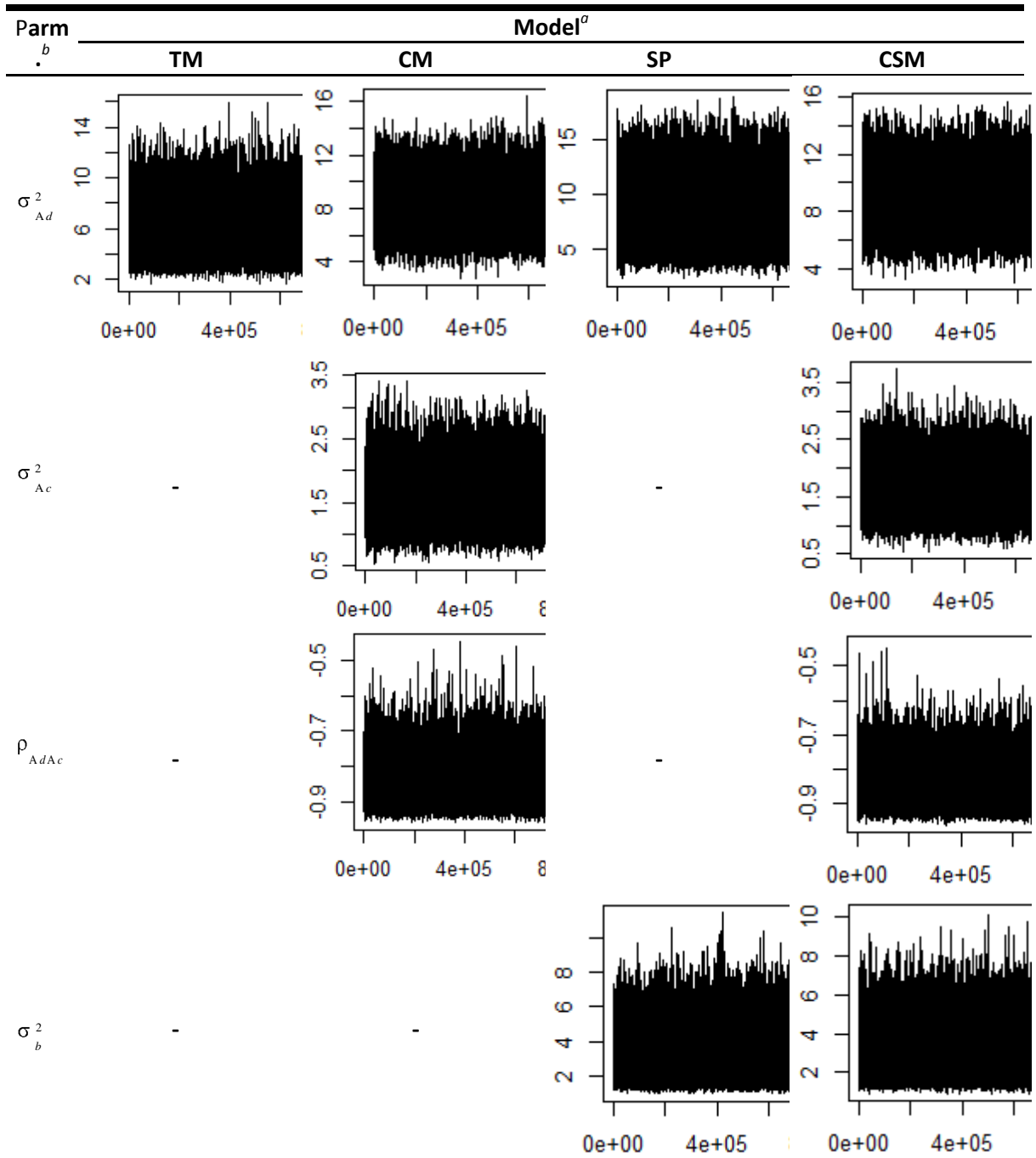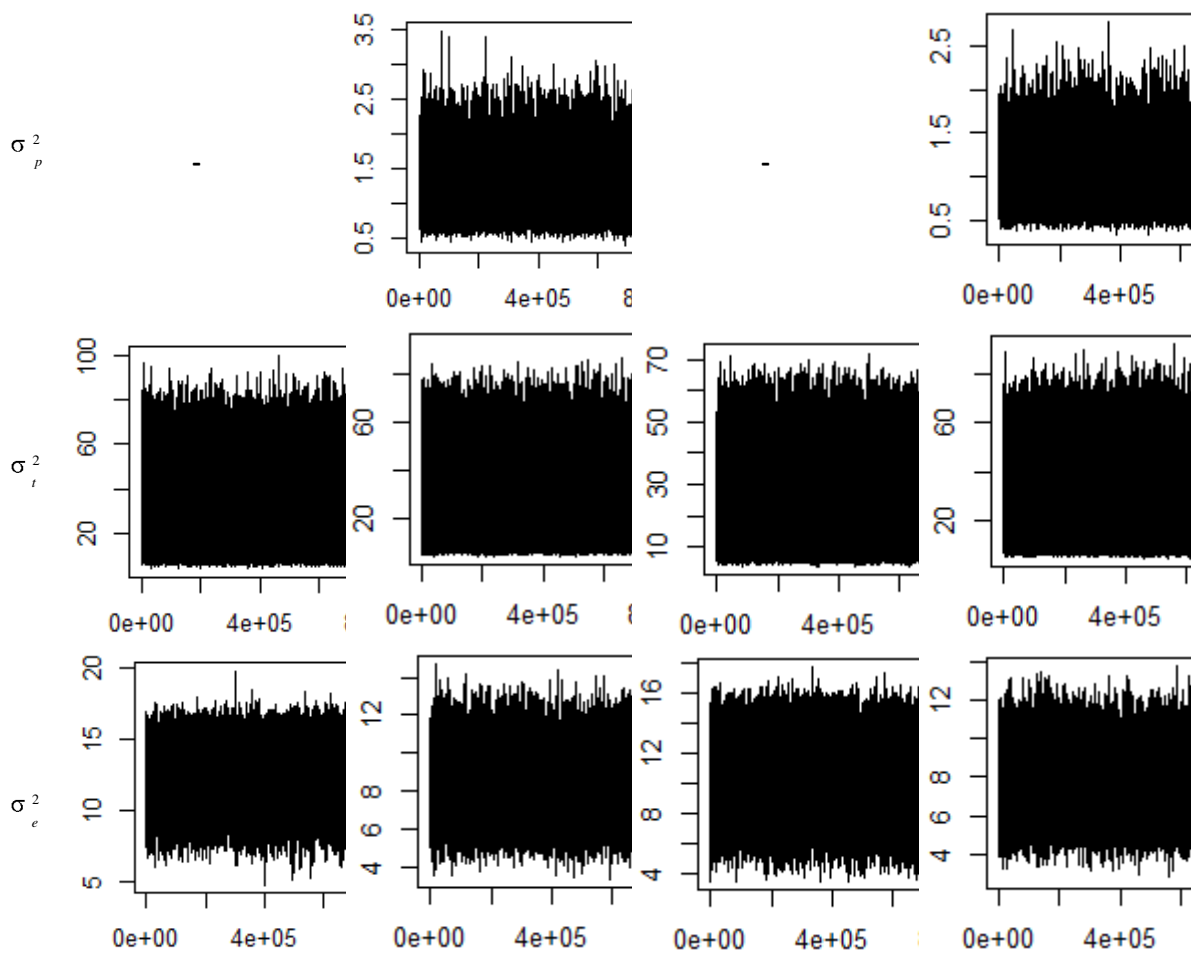**Mixed /σ^2_{Ac} = 5 /ρ_{AdAc} =-0.9**

Figure S2. Trace plots of the direct additive genetic variance σ^2_{Ad}, the competition additive genetic variance σ^2_{Ac}, the correlation between direct and competition genetic effects ρ_{AdAc}, the variance of the B-spline coefficients σ^2_b, the permanent environmental variance σ^2_p and the residual variance.

Note:

[a] **TM (Standard model)**: Individual-tree mixed model does not account for competition effects and continuous environmental heterogeneity (only fit blocks fitted as fixed effects).

CM (Competition model): Individual-tree mixed model that accounts competition effects.

SP (Spatial model): Individual-tree mixed model that accounts environmental heterogeneity using P-splines with 15 knots for rows and 15 knots for columns.

CSM (Competition + Spatial model): Individual-tree mixed model that accounts competition effects and environmental heterogeneity using P-splines with 15 knots for rows and 15 knots for columns.

[b] **Parm.** = Parameter.