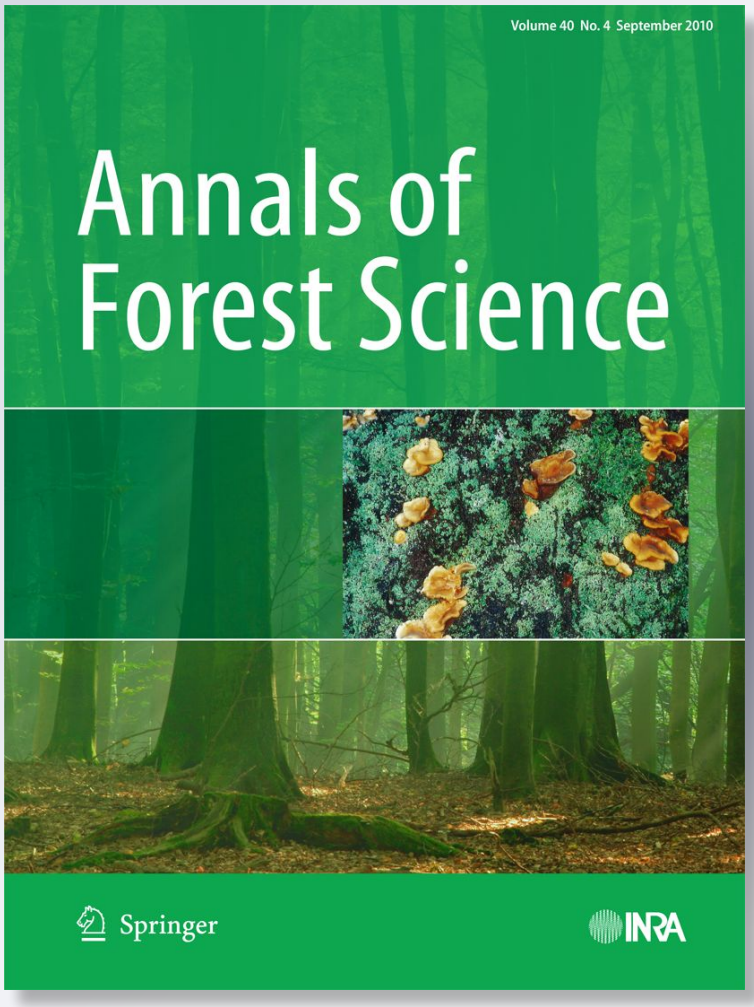
Bayesian inference for multi-environment spatial individual-tree models with additive and full-sib family genetic effects for large forest genetic trials

Eduardo P. Cappa, Alvin D. Yanchuk & Charlie V. Cartwright

Annals of Forest Science
Official journal of the Institut National de la Recherche Agronomique (INRA)

ISSN 1286-4560
Volume 69
Number 5

Annals of Forest Science (2012)
69:627-640
DOI 10.1007/s13595-011-0179-7



Springer

Annals of Forest Science (2012) 69:627–640
DOI 10.1007/s13595-011-0179-7

ORIGINAL PAPER

Bayesian inference for multi-environment spatial individual-tree models with additive and full-sib family genetic effects for large forest genetic trials

Eduardo P. Cappa · Alvin D. Yanchuk · Charlie V. Cartwright

Received: 17 May 2011 / Accepted: 26 December 2011 / Published online: 22 March 2012

Abstract
• ***Context*** The gain in accuracy of breeding values with the use of single trial spatial analysis is well known in forestry. However, spatial analyses methodology for single forest genetic trials must be adapted for use with combined analyses of forest genetic trials across sites.
• ***Aims*** This paper extends a methodology for spatial analysis of single forest genetic trial to a multi-environment trial (MET) setting.
• ***Methods*** A two-stage spatial MET approach using an individual-tree model with additive and full-sib family genetic effects was developed. Dispersion parameters were estimated using Bayesian techniques via Gibbs sampling. The procedure is illustrated using height growth data at age 10 from eight large *Tsuga heterophylla* (Raf.) Sarg. second-generation full-sib progeny trials from two series established across seven sites in British Columbia (Canada) and on one in Washington (USA).
• ***Results*** The proposed multi-environment spatial mixed model displayed a consistent reduction of the posterior mean and an increase in the precision of error variances $\left(\sigma_e^2\right)$ than the model with "sets in replicates" or incomplete block alpha designs. Also, the multi-environment spatial model provided an average increase in the posterior means of the narrow- and broad-sense individual-tree heritabilities h_N^2 and h_B^2, respectively). No consistent changes were observed in the posterior means of additive genetic correlations ($r_{Ajj'}$).
• ***Conclusion*** Although computationally demanding, all dispersion parameters were successfully estimated from the proposed multi-environment spatial individual-tree model using Bayesian techniques via Gibbs sampling. The proposed two-stage spatial MET approach produced better results than the commonly used nonspatial MET analysis.

Keywords Multi-environment spatial model · Model comparison · Western hemlock · Gibbs sampling · Full-sib family genetic effects · Additive genetic correlations

Handling Editor: Bruno Fady

Contribution of the co-authors Conceived and designed the methodology: EPC and ADY. Wrote the program to analyze the data: EPC. Analyzed the data: EPC. Wrote the paper: EPC, ADY and CVC. Organized field work and data collection: CVC. Supervising the work: ADY and CVC. Coordinating the research project: ADY.

E. P. Cappa
Department of Forest Sciences, University of British Columbia,
2424 Main Mall,
Vancouver, British Columbia V6T 1Z4, Canada

E. P. Cappa · A. D. Yanchuk · C. V. Cartwright
British Columbia Forest Service, Tree Improvement Branch,
PO Box 9519, Stn Prov Govt,
Victoria, British Columbia V8W 9C2, Canada

Present Address:
E. P. Cappa (✉)
Instituto Nacional de Tecnología Agropecuaria (INTA),
Instituto de Recursos Biológicos, Consejo Nacional de
Investigaciones Científicas y Técnicas (CONICET),
De Los Reseros y Dr. Nicolás Repetto s/n,
1686, Hurlingham, Buenos Aires, Argentina
e-mail: ecappa@cnia.inta.gov.ar

1 Introduction

Tree improvement programs usually involve the evaluation of similar sets of potential new genotypes in a range of sites and often over several years using a variety of experimental designs. This is often referred to as multi-environment trials (MET), where an environment constitutes a particular site/year combination (Smith et al. 2001). MET allows the study of the differential response of genotypes to different environmental conditions, i.e., quantifying the magnitude of the genotype by

environment interactions if related material is tested. The importance of genotype by environment interactions can be examined by genetic correlations among environments, assuming that a character measured in two environments represents two distinct traits (Falconer and Mackay 1996). Mixed model analyses of multi-environment forest genetics trials data are being increasingly used for estimating between site genetic correlations (e.g., see Costa e Silva et al. 2005; Li et al. 2007).

Spatial environmental variation within a site is largely of two types: global trend or large-scale variation and/or local trend or small-scale variation. Both are well known in forestry field trials as a result of factors such as variations in soil fertility, moisture, depth, or slope. Many studies utilizing spatial analysis of single forest genetic trials display a consistent reduction in the error variance and an increase in the heritability. This typically results in a gain in accuracy of breeding values and 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 for single forest trials to reduce the effects of the environmental variability, such as nearest neighbor techniques (Magnussen 1990), kriging (Zas 2006), nonstochastic functions such as polynomials (Thomson and El-Kassaby 1988) or smoothing splines (Costa e Silva et al. 2001) and post-blocking (Ericsson 1997). Cappa and Cantet (2007) proposed the use of the tensor products of cubic B-splines based on a mixed model framework using Bayesian techniques via Gibbs sampling. This was done by treating the B-spline function parameters as random variables (i.e., using a covariance structure for the random knots effects), in a two-dimensional grid. The two-dimensional surface involving a tensor product of B-splines bases for single forest genetic trials, provided an increase in the heritability and the accuracy of the estimated breeding value of parent and offspring (Cappa and Cantet 2007; Cappa et al. 2011). However, this spatial analysis methodology for single forest genetic trials must be adapted for use with combined analyses across sites, which can be referred to as "spatial MET analysis".

In recent years, spatial MET analyses have become common in agricultural field trials (Cullis et al. 1998; Smith et al. 2001; Casanoves et al. 2005; Oakey et al. 2007). Cullis et al. (1998) presented a one-stage approach for the spatial analysis of plot data from multi-environment variety trials, by fitting both the spatial variability within trials using a separable two-dimensional first-order autoregressive (AR(1)) error covariance structure for rows and columns of each trial, and the heterogeneity in error variance across trials. Smith et al. (2001) extended the analysis to include a factor analytic structure to approximate an unstructured genetic (i.e., variety) covariance matrix between sites. Casanoves et al. (2005) detailed the use of a local spatial correlation between plots within trials through isotropic and anisotropic covariance structures for the error terms and including heterogeneous residual variances across locations for a series of 18 MET of peanut (*Arachis hypogaea* L.). Oakey et al. (2007) extended the model of Smith et al. (2001) to estimate dominance effects with plots of sugarcane with pedigree information for 2,663 individuals on six sites. Although multi-environment mixed models that account for genetic correlations among trials and spatial variability within trials being superior to the classical nonspacial models (i.e., better model fit and more accurate breeding value prediction; Ye and Jayawickrama 2008; Ding et al. 2008), the application of spatial MET analyses in data from forest genetic trials is still rather limited (but see Ye and Jayawickrama 2008; Ding et al. 2008; Hardner et al 2010; de la Mata and Zas 2010).

Contrary to agricultural variety trials, analysis in forest genetics tests are typically on individual trees as opposed to using plot data. In addition, forestry trials are often larger than the variety trials because of the larger size of individual plants and the higher replication necessary to achieve satisfactory family estimates (Dutkowski et al. 2002). Therefore, forest genetic trials typically have a large number of trees per site and more test sites. Consequently, fitting a multi-environment individual-tree model with additive and/or full-sib family genetic effects, while accounting for genetic correlations between trials and spatial variability within trials in a one-stage approach, requires a lot of computation time and a large computer memory (Ye and Jayawickrama 2008; de la Mata and Zas 2010). In this study, we developed a two-stage approach for the multi-environment spatial analyses of several large forest genetics trials. A two-stage spatial approach was also used by Ye and Jayawickrama (2008) and de la Mata and Zas (2010) in a relatively small series of forest progeny trials (i.e., no more than four test sites with 2,700 trees per site and no more than five test sites with 2,925 trees per site, respectively). These authors corrected the original data using the spatially dependent residuals from an individual-tree mixed model with a separable two-dimensional AR(1) error covariance structures for rows and columns (Ye and Jayawickrama 2008), or using an iterative spatial analysis method (de la Mata and Zas 2010). However, in these studies, the analysis of the spatially adjusted data was limited by fitting a simpler multi-environment family mixed model (i.e., without account for the additive relationship among trees) or an individual-tree mixed model with a restrictive genetic (co) variance structure across environments. This restrictive genetic (co)variance structure assumes that the genetic variance within all trials is equal and all pairs of covariances (and correlations) are constant across sites; therefore, a precise study of the magnitude and importance of the genotype by environment interactions is not accommodated.

When the number of records and traits (trials in our case) is very large, Misztal (2008) recommended using a Bayesian method via Gibbs sampling algorithm to estimate the variance components. Bayesian theory for statistical analyses has become popular in several areas of quantitative genetics due to the feasibility of doing posterior inference by means of Markov

chain Monte Carlo (MCMC) algorithms. These methods allow marginal inferences on each individual parameter and produce measures of precision of the estimators through variances or standard errors, directly by inspecting the posterior distribution (Sorensen and Gianola 2002). In multiple-trait (or multi-environment) models with additive relationship matrices and several additive, and environmental covariance components, there is no frequentist counterpart to a posterior distribution: i.e., there are no small sampling distributions for (co)variance parameters (or functions of them). Animal breeders have used MCMC techniques such as Gibbs sampling to estimate (co) variance components in multiple-trait animal models (e.g., Van Tassell and Van Vleck 1996). Nevertheless, Gibbs sampling estimates of genetic parameters from multiple-trait individual-tree models are still scarce (e.g., Cappa and Cantet 2007), and estimates from multi-environment individual-tree models even fewer (Gwaze and Woolliams 2001). Gwaze and Woolliams (2001) fitted a multi-environment individual-tree model with an unstructured additive (co)variance matrix between trials and an independent error variance for each site, using Gibbs sampling to make decisions about the optimal selection environment. However, their approach does not considered spatial variability within trials.

Our goal in this paper is to extend the methodology for spatial analysis for a single forest genetic trial proposed by Cappa and Cantet (2007) to a MET setting, using an individual-tree model with additive and full-sib family genetic effects in a two-stage approach. The Bayesian approach via Gibbs sampling was employed to make inferences in all dispersion parameters of the model. Developments are illustrated for tree height at age 10 for 34,143 progeny from 483 families involving 149 parents for eight large second-generation full-sib trials in two series of western hemlock (*Tsuga heterophylla* (Raf.) Sarg.). Additionally, the resulting estimates of all dispersion parameters for the proposed multi-environment spatial individual-tree mixed model are finally compared with corresponding estimates from the classical model including a priori "sets in replicates" or incomplete block alpha designs.

2 Materials and methods

2.1 Genetic material, mating design, and trial description

The western hemlock (*T. heterophylla* (Raf.) Sarg.) data set used in this study originated from the full-sib second-generation Hemlock Tree Improvement Cooperative (HEMTIC) program in the Pacific Northwest. This second-generation population was formed by recombining the best 150 first-generation parents, i.e., by recombining the top 30 parents from each of the five first-generation programs: two each from Oregon (OR) and Washington (WA) and one from BC. Two different mating designs were used: (1) "local diallels" (LD) series, composed of five six-parent disconnected partial (half) diallels representing each of the five programs mentioned above (two each from OR and WA and one from BC), and (2) "elite diallels" (ED) series, composed of the "best" 30 parents in partial diallels (i.e., the six best from parents from each of the five programs as ranked in the first generation tests). Each "best" parent was crossed with two parents from each of the five programs. A detailed description of these first- and second-generation genetic materials used in this study can be found in King and Cress (1991), King et al. (1998), and Jayawickrama (2003).

Eight of the HEMTIC full-sib progeny trials, five from the LD series and three from the ED series were chosen for the two-stage spatial MET analysis (Table 1, Fig. 1). The five LD series trials were planted in 1997 and 1998 and the three ED series sites were planted in 1999. They are distributed along the Pacific coast, from Washington State to the northern end of the Vancouver Island in BC (i.e., from 47° 12′ to 50° 34′ north latitude, from 123° 56′ to 127° 41′ west longitude and from 45 to 380 m of altitude; Fig. 1).

The number of common parents between pairs of sites varied from 134 to 144 within the five LD series trials and was 30 for the three ED series trials. The LD series has from 314 to 342 common families and the ED series from 137 to 140 (Table 2). However, the parents and families forming the two series are only partially connected. Only 30 of the 149 parents in the ED series are also in the LD series, and the ED series has only 15 common families (out of a total of 483) with the LD series. A "sets in replicates" design (Schutz and Cockerham 1966; "S in R"), with families within a local diallel grouped together in a set ("genetic group") and each replicate including all sets, was used for the LD series. The families of the ED series were planted in an incomplete block alpha design (Williams and Matheson 1994; ICB) with 30 replications and 10 incomplete blocks within replications. Single-tree plots were used in all the trials. The five LD series trials were planted at 2×2, 2.5×2.5, and 3×2.5 m spacing and the three ED series trials at 2.5×2.5 m spacing (Table 1). Two check lots occur on almost all test sites; however, these data were not used for the first and second stages of the spatial MET analysis. Total height at age 10 (TH, centimeter) for all surviving trees was measured on all sites. The resulting data set consisted of 34,143 progeny height observations from the eight trials (23,473 from LD series and 10,670 from ED series).

2.2 Statistical models of analysis

2.2.1 The first-stage statistical model for spatial single-environment analysis

In the first-stage of the spatial MET analysis, a two-dimensional smoothed surface involving a tensor product of cubic B-splines bases was estimated for each site of the two series analyzed. In doing so, an individual-tree mixed

Springer INRA

Table 1 Location, sites characteristics, design information, mean with standard deviation, minimum, and maximum values for total height at age of 10 years for each of the eight trials

	Local diallels				Elite diallels			
Site	Jordan 2	Kiyu	Jordan 3	Rupert 1	Humptulips	Stove	Tlupana	Michelsen
Latitude (north)	48°25′ 16″	50°05′ 07″	48°24′ 44″	50°34′ 21″	47°12′ 52″	50°15′ 49″	49°45′ 33″	50°34′ 42″
Longitude (west)	124°00′ 43″	126°30′ 00″	123°59′ 28″	127°23′ 44″	123°56′ 15″	125°40′ 02″	126°23′ 22″	127°41′ 30″
Altitude (m)	110	380	130	110	100	90	45	100
Soil description	Well drained moderately deep, finer textured podzol with limited litter and moder organic layers	Thin silty soils over, very well drained submesic mesotrophic site	Moderately deep coarser soils with limited litter and moder organic layers, well drained	Hummocky silty soils with moderately deep litter and organic soil layers, moderately well drained	Deeper, silty soils with limited organic layers, well drained mesic	Moderately deep silt soils with a moder form orgainic layer, well drained	Moderately deep silty soils with thin litter and organic layers, well drained	Subhygric, eutrophic podzol with a moderately deep organic layer
AMP[a] (mm)	2,428	3,134	2,415	2,212	2,440	1,851	3,987	2,675
AMT[b] (°C)	9.2	7.9	9.3	8.2	10.3	8.4	8.9	8.4
MTCM[c] (°C)/MTWM[d] (°C)	0.4/20.9	−1.7/22.1	0.5/21	0/19.5	1.5/21.9	−1.4/21.1	0.3/21.6	0.4/19.2
Number of initial trees	7,680	7,680	8,320	7,488	8,320	4,800	4,800	4,800
Number Of parents	138	138	149	149	149	30	30	30
Number of families	324	314	342	342	339	141	137	140
Number of check lots	2	2	2	2	2	2	2	2
Number of trees with records	5,305	3,830	4,393	4,384	5,561	3,660	3,428	3,582
Field design[e]	S in R	S in R	S in R	S in R	S in R	ICB	ICB	ICB
Replicates	20	20	20	18	20	30	30	30
Sets	23	23	26	26	26	10	10	10
Rows	104	128	60	144	74	40	97	72
Columns	104	88	199	143	144	186	96	149
Spacing (m)	2×2	2.5×2.5	3×2.5	2.5×2.5	2×2	2.5×2.5	2.5×2.5	2.5×2.5
Mean (cm)(SD)[f]	548.8 (150.1)	488.8 (151.2)	699.6 (148.0)	648.5 (175.8)	606.5 (159.7)	632.9 (141.5)	604.1 (162.1)	697.1 (134.5)

[a] AMP annual mean precipitation

[b] AMT annual mean temperature

[c] MTCT minimum mean temperature of the coldest month (usually January)

[d] MTWM maximum mean temperature of the warmest month (usually August)

[e] S in R "sets in replicates" design, ICB, incomplete block design (alpha design)

[f] SD standard deviation

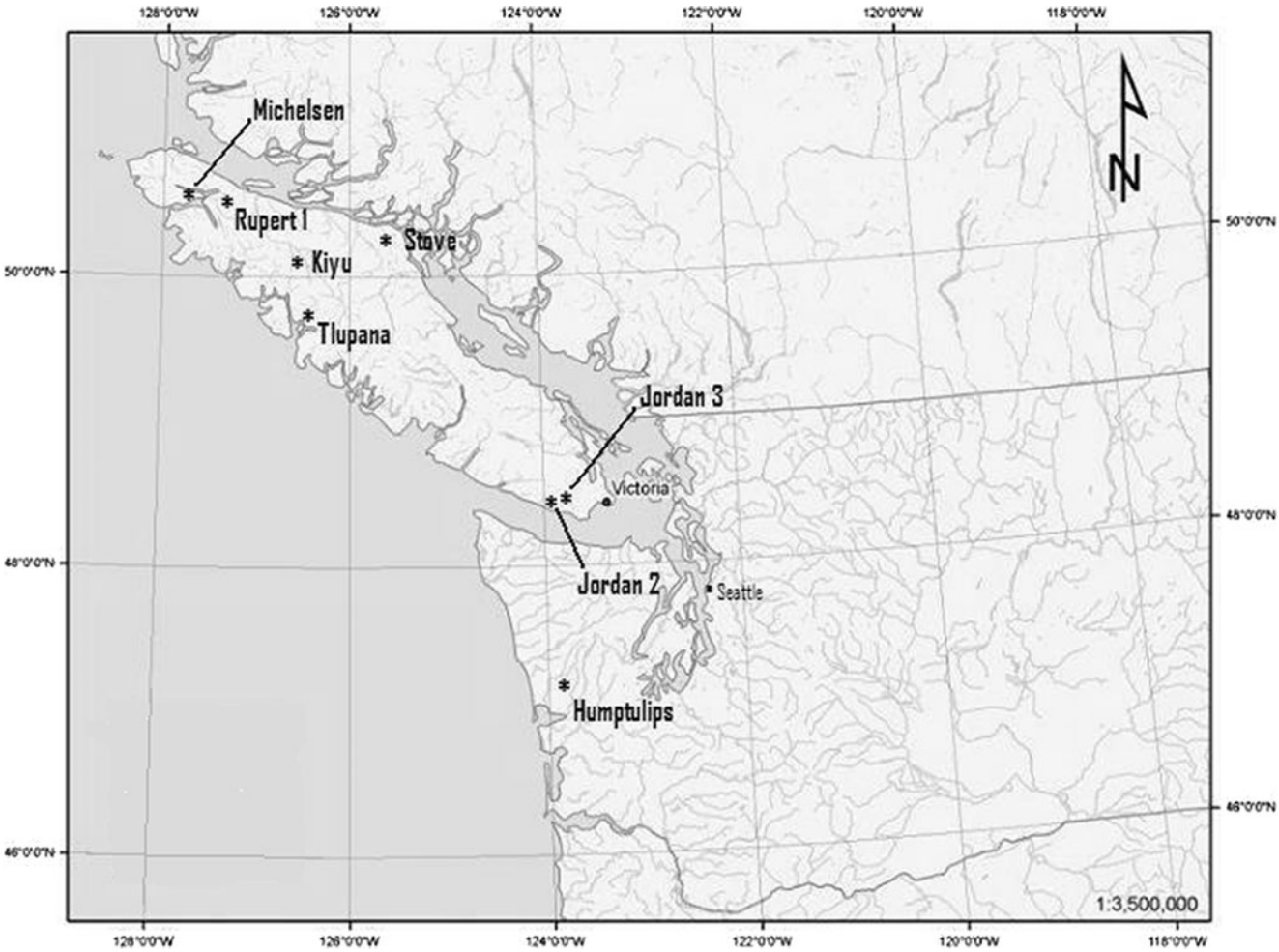


Fig. 1 Location of the eight western hemlock full-sib progeny trials

model with a two-dimensional smoothed surface, as described by Cappa and Cantet (2007) was fitted. The only difference from their individual-tree additive genetic mixed model was the addition of the random full-sib family genetic effects (corresponding to specific combining ability; SCA). Therefore, the individual-tree mixed model with a two-dimensional surface for the original data fitted on each site included: fixed effects of the diallel unit (i.e., genetic groups; LD series), overall site mean (ED series), and random effects of additive (breeding values) and full-sib family

Table 2 Number of parents (above diagonal) and families (below diagonal) in common among the trials analyzed within each series

Local diallels	Jordan 2	Kiyu	Jordan 3	Rupert 1	Humptulips
Jordan 2		134	136	136	136
Kiyu	314		134	134	134
Jordan 3	324	314		144	143
Rupert 1	324	314	342		143
Humptulips	324	314	339	339	
Elite diallels	Stove	Tlupana	Michelsen		
Stove		30	30		
Tuplana	137		30		
Michelsen	140	137			

effects (i.e., SCA). The two-dimensional surface (in matrix notation, $\boldsymbol{Bb}$) was fitted at each site using the tensor products of cubic B-splines bases. The matrix $\boldsymbol{B}$ contains the nonzero B-spline bases needed to express each row and column in terms of cubic B-splines bases. The parametric vector $\boldsymbol{b}$ contains the parameters of the tensor products of B-splines bases. A more detailed explanation of the two-dimensional surface ($\boldsymbol{Bb}$) using the tensor products of B-splines bases, can be found in Cappa and Cantet (2007, pp. 2678–2679). Three models with different numbers of knots for rows and columns were fitted for each site of the two series analyzed (Cappa et al. (unpublished)). The best model for each site (i.e., the suitable number of knots capturing most of the spatial variability) was first used to estimate the two-dimensional smoothed surface, and then to obtain the spatially adjusted data (as described in the following section).

2.2.2 The second-stage statistical model for spatial MET analysis

Given the small number of common parents and families, the different field design for each series, and our overall objective to compare the multi-environment model for the spatially adjusted data on the a priori "S in R" and ICB designs, the MET analyses (i.e., the spatial and nonspatial) were undertaken within each series separately (i.e., for the LD and ED series). Therefore, two multi-environment individual-tree mixed models (i.e., spatial and nonspatial) with additive and full-sib family genetic effects were evaluated within each series. In the second-stage of the spatial MET analysis, the spatially adjusted data are obtained for each tree at each site by subtracting the estimated smoothed surface from the original data. Thus, the spatially adjusted data of tree i at each j site $\boldsymbol{y}_{ij}$, is calculated as

$$\boldsymbol{y}_{ij} = \boldsymbol{y}^{*}_{ij} - \left(\boldsymbol{B}\hat{\boldsymbol{b}}\right)_{ij} \tag{1}$$

where, $\boldsymbol{y}^{*}_{ij}$ is the original data of tree i for site j and $\left(\boldsymbol{B}\hat{\boldsymbol{b}}\right)_{ij}$ is the estimate smoothed surface of of tree i for site j from the first stage. Then, the full multi-environment model is analyzed by considering the set of observations on each environment a different trait and combining the analysis over the five and three sites within of each LD and ED series, respectively. Since diallel mating units are confounded with geographic origins of parents in the LD series, genetic groups or diallel units are included in the analyses to account for genetic differences among programs and genetic sampling effects among diallels. A previous analysis found significant differences among genetic group in the five trials of the LD series. Therefore, the individual tree spatially adjusted data, were analyzed by fitting the multi-environment individual-tree mixed models for the LD and ED series with partial diallel units (or genetic groups) within site, and overall site mean as fixed effects, respectively. Let the subscript j index the sites ($j=1, \ldots, s$), i index the trees ($i=1, \ldots, q_s$) for each site, and l index the groups ($l=1, \ldots, g_s$) for site s. Let y_{ijl} be the spatially adjusted data of tree i from site j scored in group l. The a_{ijl} and f_{ijl} are the additive breeding value and the SCA of individual (or family) i for site j, scored in group l. Then, we can write the following multi-environment individual-tree mixed model with additive and full-sib family genetic effects for spatially adjusted data (hereafter multi-environment spatial model) as:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ . \\ \boldsymbol{y}_s \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & . & \boldsymbol{0} \\ . & . & . \\ \boldsymbol{0} & . & \boldsymbol{X}_s \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ . \\ \boldsymbol{\beta}_s \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_{a1} & . & \boldsymbol{0} \\ . & . & . \\ \boldsymbol{0} & . & \boldsymbol{Z}_{as} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ . \\ \boldsymbol{a}_s \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_{f1} & . & \boldsymbol{0} \\ . & . & . \\ \boldsymbol{0} & . & \boldsymbol{Z}_{fs} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_1 \\ . \\ \boldsymbol{f}_s \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ . \\ \boldsymbol{e}_s \end{bmatrix} \tag{2}$$

or, more compactly as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z}_a\boldsymbol{a} + \boldsymbol{Z}_f\boldsymbol{f} + \boldsymbol{e} \tag{3}$$

The breeding values of all individuals for all sites are included in $\boldsymbol{a}_1'|\ldots|\boldsymbol{a}_s'] = \boldsymbol{a}$. This vector has zero expectation and a covariance matrix equal to $\boldsymbol{G}_{\mathrm{A}} \otimes \boldsymbol{A}$, where $\boldsymbol{G}_{\mathrm{A}}$ is the $s \times s$ additive genetic (co)variance matrix between the five and three trials of the LD and ED series, with diagonal elements $\sigma^2_{\mathbf{A}_{j,j}}$: the additive variance of site j and off-diagonal elements $\sigma_{\mathrm{A}_{j,j'}}$: the additive genetic covariance between sites j and j'. The square matrix $\boldsymbol{A}$, of order $q \times q$, contains the additive relationships (Henderson 1984) among all trees of all sites within each series, either parents without data or offspring with records in $\boldsymbol{y}$. The total number of trees at all sites within each series, including parents without phenotypic data, is q.

The full-sib family genetic effects for all sites within each series are included in $[\boldsymbol{f}_1'|\ldots|\boldsymbol{f}_s'] = \boldsymbol{f}$. This vector has zero expectation and the diagonal $f \times f$ (co)variance matrix $\boldsymbol{G}_{\mathrm{F}}$, where $\boldsymbol{G}_{\mathrm{F}} = \bigoplus_{j=1}^{s} \boldsymbol{I}_{n_j}\sigma^2_{f_j}$ and $\sigma^2_{f_j}$ are scalars representing the family variance for each site j, with $j=1, \ldots, s$. The symbol $\bigoplus_{j=1}^{s}$ indicate the "direct sum" of matrices notation. Therefore, the family genetic covariance between sites j and j' are assumed to be

zero within both series, this allows a more parsimonious model.

To obtain a matrix formulation of the distribution of e, the error terms are ordered by tree within site. Then, the expected value of e is zero and covariance matrix $\boldsymbol{R}$, where $\boldsymbol{R} = \bigoplus_{j=1}^{s} \boldsymbol{I}_{\boldsymbol{n_j}} \sigma^2_{e_j}$, and $\sigma^2_{e_j}$ are scalars representing the error variance for each site j, with j=1, …, s. Notice that the family and residual covariances are assumed to be zero, even though trees across sites are additive genetically related, i.e., additive genetic covariance between any two sites within each series exists, but no family and environmental covariance is assumed. Vectors $\boldsymbol{a}$, $\boldsymbol{f}$, and $\boldsymbol{e}$ are assumed to be independent and normally distributed.

For the nonspatial MET analysis, the multi-environment individual-tree model with additive and full-sib family genetic effects (3), were fitted using the original data $\left(\boldsymbol{y}^*_{ij}\right)$ instead of the spatially adjusted data $\left(\boldsymbol{y}_{ij}\right)$ of tree i at each j site for each series. To account for the environmental variability, this multi-environment nonspatial model also included from 18 to 20 replicate effects for the five LD series trials, and 30 replicate effects for the ED series trials, as fixed design effects.

2.3 Bayesian estimation

All dispersion parameters of the model (3) are estimated by the Bayesian approach, Gibbs sampling (Sorensen and Gianola 2002). We describe in detail the prior distribution of all parameters, the likelihood of the data, and the joint and conditional posterior densities for the multi-environment individual-tree model of spatially adjusted data, as these are necessary to make posterior inference by means of Gibbs sampling.

2.3.1 Specification of prior distributions and likelihood for combined sites

Following Cappa and Cantet (2006) for a multi-trait individual-tree model, here we chose to use conjugate prior densities. In a conjugate approach, the prior densities for all parameters are selected to be closed under sampling, 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, $\boldsymbol{\beta}$ is taken to be $\boldsymbol{\beta} \sim N_p\ (0,\ \boldsymbol{K})$, where p is the number of fixed effects. The matrix $\boldsymbol{K}$ is diagonal with large elements ($k_{ii} > 10^8$, Cantet et al. 2004), and the prior density of $\boldsymbol{\beta}$ (the fixed effects) is then proportional to:

$$p(\boldsymbol{\beta}|\boldsymbol{K}) \propto \left|\prod_{i=1}^{p} k_{ii}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{p}\frac{\boldsymbol{\beta}_i^2}{k_{ii}}\right\} \tag{4}$$

The vector of breeding values is distributed a priori as $\boldsymbol{a} \sim N_q\ (\boldsymbol{0},\ \boldsymbol{G}_{\mathrm{A}} \otimes \boldsymbol{A})$; see (7) in Cappa and Cantet (2006), so that:

$$p(\boldsymbol{a}|\boldsymbol{G}_{\mathrm{A}},\ \boldsymbol{A}) \propto |\boldsymbol{G}_{\mathrm{A}}|^{-\frac{q}{2}}.|\boldsymbol{A}|^{-\frac{s}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{a}'\left(\boldsymbol{G}_{\mathrm{A}}^{-1} \otimes \boldsymbol{A}^{-1}\right)\boldsymbol{a}\right\} \tag{5}$$

A priori, the additive covariance matrix $\boldsymbol{G}_{\mathrm{A}}$ follows an inverted Wishart density as $\boldsymbol{G}_{\mathrm{A}} \sim \mathrm{IW}\ (\boldsymbol{G}_{\mathrm{A}}^*, n_A)$ with parameters $\boldsymbol{G}_{\mathrm{A}}^*$ and n_A. In a Bayesian setting, $\boldsymbol{G}_{\mathrm{A}}^*$ is the hypercovariance and n_A the *degrees of belief* (Sorensen and Gianola 2002; page 57) so that:

$$p\left(\boldsymbol{G}_{\mathrm{A}}\left|\boldsymbol{G}_{\mathrm{A}}^*,\ n_A\right.\right) \propto \left|\boldsymbol{G}_{\mathrm{A}}^*\right|^{\frac{n_A}{2}} \left|\boldsymbol{A}\right|^{-\frac{(n_A+s+1)}{2}} \exp\left\{-\frac{1}{2}\ tr\left(\boldsymbol{G}_{\mathrm{A}}^*\boldsymbol{G}_{\mathrm{A}}^{-1}\right)\right\} \tag{6}$$

The vector of full-sib family genetic effects is distributed a priori as $\boldsymbol{f} \sim N_f\ (0,\ \boldsymbol{G}_{\mathrm{F}})$, so that:

$$p(\boldsymbol{f}|\boldsymbol{G}_{\mathrm{F}}) \propto \left|\boldsymbol{G}_{\mathrm{F}}\right|^{-\frac{f}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{f}'\boldsymbol{G}_{\mathrm{F}}^{-1}\boldsymbol{f}\right\} \tag{7}$$

A priori, the diagonal family covariance matrix $\boldsymbol{G}_{\mathrm{F}}$ (i.e., the $\sigma^2_{f_j}$ at each j sites), follows an scaled inverted x^2 density as Inv-x^2 $\left(\delta^2_{f_j}, \upsilon_f\right)$, with hypervariance $\delta^2_{f_j}$ and degrees of belief υ_f, so that:

$$p\left(\sigma^2_{f_j}\left|\upsilon_f, \sigma^2_{f_j}\right.\right) \propto \left(\sigma^2_{f_j}\right)^{-\left(\frac{\upsilon_f}{2}+1\right)} \exp\left\{-\frac{\upsilon_f\delta^2_{f_j}}{2\sigma^2_{f_j}}\right\} \tag{8}$$

Finally, a priori, the residual variance at each site $\sigma^2_{e_j}$ has a scaled inverted x^2 density as Inv-x^2 $\left(\delta^2_{e_j}, \upsilon_e\right)$, with hypervariance $\delta^2_{e_j}$ and degrees of belief υ_e, so that:

$$p\left(\sigma^2_{e_j}\left|\upsilon_e, \sigma^2_{e_j}\right.\right) \propto \left(\sigma^2_{e_j}\right)^{-\left(\frac{\upsilon_e}{2}+1\right)} \exp\left\{-\frac{\upsilon_e\delta^2_{e_j}}{2\sigma^2_{e_j}}\right\} \tag{9}$$

In the Bayesian view of the mixed linear model (Sorensen and Gianola 2002), the likelihood of the data is proportional to

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{f}, \boldsymbol{R}) \propto |\boldsymbol{R}|^{-\frac{1}{2}}.\exp\left[-\frac{1}{2}\boldsymbol{e}'\boldsymbol{R}^{-1}\boldsymbol{e}\right] \tag{10}$$

where $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Z}_{\boldsymbol{a}}\boldsymbol{a} - \boldsymbol{Z}_{\boldsymbol{f}}\boldsymbol{f}$.

2.3.2 Joint and conditional posterior densities for combined sites

The joint posterior density is written as the product of the expressions (10) and (4–9) which then results in

$$
\begin{aligned}
&p(\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{f},\boldsymbol{G}_{\mathrm{A}},\boldsymbol{G}_{\mathrm{F}},\boldsymbol{R}|\boldsymbol{y}) \propto \\
&|\boldsymbol{R}|^{-\frac{1}{2}}\exp\left[-\tfrac{1}{2}\boldsymbol{e}'\boldsymbol{R}^{-1}\boldsymbol{e}\right]\exp\left\{-\tfrac{1}{2}\sum_{i=1}^{p}\frac{\beta_i^{*2}}{k_{ii}}\right\} \\
&\exp\left\{-\tfrac{1}{2}\boldsymbol{a}'\left(\boldsymbol{G}_{\mathrm{A}}^{-1}\otimes\boldsymbol{A}^{-1}\right)\boldsymbol{a}\right\}|\boldsymbol{G}_{\mathrm{A}}|^{-\frac{(n_{\mathrm{A}}+s+q+1)}{2}}\exp\left\{-\tfrac{1}{2}\,tr\left(\boldsymbol{G}_{\mathrm{A}}^{*}\boldsymbol{G}_{\mathrm{A}}^{-1}\right)\right\} \\
&\exp\left\{-\tfrac{1}{2}\boldsymbol{f}'\boldsymbol{G}_{\mathrm{F}}^{-1}\boldsymbol{f}\right\}\left(\sigma_{f_j}^2\right)^{-\left(\frac{\upsilon_f}{2}+1\right)}\exp\left\{-\frac{\upsilon_f\delta_{f_j}^2}{2\sigma_{f_j}^2}\right\}\quad\left(\sigma_{e_j}^2\right)^{-\left(\frac{\upsilon_e}{2}+1\right)}\exp\left\{-\frac{\upsilon_e\delta_{e_j}^2}{2\sigma_{e_j}^2}\right\}
\end{aligned}
\tag{11}
$$

To take advantage of all information in the data about any parameter, inferences about $\boldsymbol{\beta}$, $\boldsymbol{a}$, $\boldsymbol{f}$, $\boldsymbol{G}_{\mathrm{A}}$, $\boldsymbol{G}_{\mathrm{F}}$, and $\boldsymbol{R}$ are based on their respective marginal posterior densities. A useful property of these marginal distributions is their lack of dependence upon any particular value of the other parameters. Thus, each marginal density is obtained by integrating out the joint distribution (11) with respect to the parameters other than the one of interest. This is accomplished by using the MCMC procedure known as Gibbs sampling, while taking advantage that the marginal conditional densities resulting from (10) are feasible for sampling. For a multitrait model with additive effects, the conditional densities were obtained as shown in Van Tassell and Van Vleck (1996), Sorensen and Gianola (2002), and Cappa and Cantet (2006). Thus, for the linear parameters in $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{a}}$ and $\widehat{\boldsymbol{f}}$ of model (3) the posterior conditional density is equal to

$$
\begin{aligned}
&\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{a}}\\ \widehat{\boldsymbol{f}}\end{bmatrix}\Bigg|\boldsymbol{y},\ \boldsymbol{G}_{\mathrm{A}},\ \boldsymbol{G}_{\mathrm{F}},\sigma_{e_j}^2,\sim N_{p+f+sq} \\
&\left(\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{a}}\\ \widehat{\boldsymbol{f}}\end{bmatrix},\begin{bmatrix}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}+\boldsymbol{K}^{-1} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z}_a & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z}_f\\ \boldsymbol{Z}_a'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}_a'\boldsymbol{R}^{-1}\boldsymbol{Z}_a+\boldsymbol{G}_{\mathrm{A}}^{-1}\otimes\boldsymbol{A}^{-1} & \boldsymbol{Z}_a'\boldsymbol{R}^{-1}\boldsymbol{Z}_f\\ \boldsymbol{Z}_f'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}_f'\boldsymbol{R}^{-1}\boldsymbol{Z}_a & \boldsymbol{Z}_f'\boldsymbol{R}^{-1}\boldsymbol{Z}_f+\boldsymbol{G}_{\mathrm{F}}^{-1}\end{bmatrix}^{-1}\right)
\end{aligned}
\tag{12}
$$

The vectors $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{a}}$ and $\widehat{\boldsymbol{f}}$ in (16) are the solutions to the following system of equations

$$
\begin{bmatrix}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X}+\boldsymbol{K}^{-1} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z}_a & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z}_f\\ \boldsymbol{Z}_a'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}_a'\boldsymbol{R}^{-1}\boldsymbol{Z}_a+\boldsymbol{G}_{\mathrm{A}}^{-1}\otimes\boldsymbol{A}^{-1} & \boldsymbol{Z}_a'\boldsymbol{R}^{-1}\boldsymbol{Z}_f\\ \boldsymbol{Z}_f'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}_f'\boldsymbol{R}^{-1}\boldsymbol{Z}_a & \boldsymbol{Z}_f'\boldsymbol{R}^{-1}\boldsymbol{Z}_f+\boldsymbol{G}_{\mathrm{F}}^{-1}\end{bmatrix}\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{a}}\\ \widehat{\boldsymbol{f}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y}\\ \boldsymbol{Z}_a'\boldsymbol{R}^{-1}\boldsymbol{y}\\ \boldsymbol{Z}_f'\boldsymbol{R}^{-1}\boldsymbol{y}\end{bmatrix}
\tag{13}
$$

The posterior conditional density of the covariance matrix $\boldsymbol{G}_{\mathrm{A}}$ is an inverted Wishart with scaling matrix $\boldsymbol{G}_{\mathrm{A}}{}^{*}+\boldsymbol{S}$ and degrees of belief equal to $n_A+q+s+1$

$$
p\left(\boldsymbol{G}_{\mathrm{A}}\middle|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{f},\sigma_{f_1}^2,\ldots,\sigma_{f_s}^2,\sigma_{e_1}^2,\ldots,\sigma_{e_s}^2\right)\propto\left|\boldsymbol{G}_{\mathrm{A}}\right|^{-\frac{(n_A+q+s+1)}{2}}\exp\left\{-\frac{1}{2}tr\left[\left(\boldsymbol{G}_{\mathrm{A}}^{*}+\boldsymbol{S}\right)\boldsymbol{G}_{\mathrm{A}}^{-1}\right]\right\}
\tag{14}
$$

being $\boldsymbol{S}$ equal to

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{a}_1'\boldsymbol{A}^{-1}\boldsymbol{a}_1 & \boldsymbol{a}_1'\boldsymbol{A}^{-1}\boldsymbol{a}_2 & . & . & \boldsymbol{a}_1'\boldsymbol{A}^{-1}\boldsymbol{a}_s \\ \boldsymbol{a}_2'\boldsymbol{A}^{-1}\boldsymbol{a}_1 & \boldsymbol{a}_2'\boldsymbol{A}^{-1}\boldsymbol{a}_2 & . & . & \boldsymbol{a}_2'\boldsymbol{A}^{-1}\boldsymbol{a}_s \\ . & . & \boldsymbol{a}_i'\boldsymbol{A}^{-1}\boldsymbol{a}_j & . & . \\ . & . & . & . & . \\ \boldsymbol{a}_s'\boldsymbol{A}^{-1}\boldsymbol{a}_1 & \boldsymbol{a}_s'\boldsymbol{A}^{-1}\boldsymbol{a}_2 & . & . & \boldsymbol{a}_s'\boldsymbol{A}^{-1}\boldsymbol{a}_s \end{bmatrix} \quad (15)$$

The posterior conditional density of the family variance at each j site, $\sigma^2_{f_j}$, has a scaled inverted x^2 distribution for each jth site with parameters $\widetilde{\upsilon}_{f_j} = f_j + \upsilon_f$ and $\widetilde{\delta}^2_{f_j} = \left(\boldsymbol{f}_j'\boldsymbol{f}_j + \upsilon_f\delta^2_{f_j}\right)/\widetilde{\upsilon}_{f_j}$, such that it is written a

$$p\left(\sigma^2_{f_j}\middle|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{f},\boldsymbol{G}_A,\sigma^2_{e_1},...,\sigma^2_{e_s}\right) \propto \left(\sigma^2_{f_j}\right)^{-\left(\frac{\upsilon_f+f_j+2}{2}+1\right)} \exp\left\{-\frac{\widetilde{\upsilon}_{f_j}\widetilde{\delta}^2_{f_j}}{2\sigma^2_{f_j}}\right\} \quad (16)$$

Finally, the marginal posterior conditional density of the error variances $\sigma^2_{e_j}$ has a scaled inverted x^2 distribution with parameters $\widetilde{\upsilon}_{e_j} = n_j + \upsilon_e$ and $\widetilde{\delta}^2_{e_j} = \left(\boldsymbol{e}_j'\boldsymbol{e}_j + \upsilon_e\delta^2_{e_j}\right)/\widetilde{\upsilon}_{e_j}$, such that it is written as

$$p\left(\sigma^2_{e_j}\middle|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{a},\boldsymbol{f},\boldsymbol{G}_A,\sigma^2_{f_1},\ldots,\sigma^2_{f_s}\right) \propto \left(\sigma^2_{e_j}\right)^{-\left(\frac{\upsilon_e+n_j+2}{2}+1\right)} \exp\left\{-\frac{\widetilde{\upsilon}_{e_j}\widetilde{\delta}^2_{e_j}}{2\sigma^2_{e_j}}\right\} \quad (17)$$

At any iteration of the Gibbs algorithm, we first sampled from distribution (12), then from (17), (14), and finally from (16), for the process to start back again. A program was written in FORTRAN to perform all calculations. The FORTRAN program is available from the first author upon request.

2.3.3 Computational details, posterior inference, and model comparison

The values of the hypervariances for the diagonal elements of $\boldsymbol{G}_A^*$, $\delta^2_{f_j}$, and $\delta^2_{e_j}$ were estimated from the same spatially adjusted data set using an empirical Bayes approach via a single trial Gibbs sampling with an individual-tree model including fixed effects of genetic groups (for the LD series) and an overall site mean (for the ED series) and random additive and full-sib family genetic effects. Priors for the additive covariances $\sigma_{A_{j,j'}}$ were obtained using the average information algorithm implemented in the software package ASReml (Gilmour et al. 2006) with a multi-environment individual-tree model (3) from the same spatially adjusted data. The degrees of belief were then set to 10 (i.e., n_A=υ_f=υ_e=10) to reflect a relatively high degree of uncertainty (Sorensen and Gianola 2002; page 57).

At the end of each iteration of the Gibbs sampling $\widetilde{\sigma}_{A_{j,j'}}$, $\widetilde{\sigma}^2_{A_{j,j}}$, $\widetilde{\sigma}^2_{f_j}$, and $\widetilde{\sigma}^2_{e_j}$ were reparametrized to additive genetic correlations ($r_{Ajj'}$), dominance variance $\left(\sigma^2_{Dj}\right)$, and individual narrow- and broad-sense heritability (h^2_N and h^2_B, respectively) as follows:

$$r_{Ajj'} = \frac{\widetilde{\sigma}_{A_{j,j'}}}{\sqrt{\widetilde{\sigma}^2_{A_{j,j}}\,\widetilde{\sigma}^2_{A_{j',j'}}}}; \quad \sigma^2_{Dj} = 4 \times \widetilde{\sigma}^2_{f_j};$$

$$h^2_{Nj} = \frac{\widetilde{\sigma}^2_{A_{j,j}}}{\widetilde{\sigma}^2_{A_{j,j}} + \widetilde{\sigma}^2_{f_j} + \widetilde{\sigma}^2_{e_j}}; \quad h^2_{Bj} = \frac{\widetilde{\sigma}^2_{A_{j,j}} + \widetilde{\sigma}^2_{Dj}}{\widetilde{\sigma}^2_{A_{j,j}} + \widetilde{\sigma}^2_{f_j} + \widetilde{\sigma}^2_{e_j}}$$

A single Gibbs chain of 202,000 (LD series) and 302,000 (ED series) iterations was sampled. These iterations took approximately 6 (LD series) and 13 (ED series) days, respectively, to deliver its entire inferential output, on a single 2.20 GHz Intel(R) Core(TM)2 DUO processor with 3.0 Gbytes of random access memory, using a 32-bits Microsoft Window XP. The first 2,000 iterates were discarded due to burn-in. Convergence was monitored by plotting the iterations against the mean of the draws up to each iteration (running mean plots) for each parameter. Means, modes, medians, standard deviations, and 95% high posterior density intervals, were then calculated for all parameters from the individual marginal posteriors using the R package "Bayesian Output Analysis" (BOA version 1.0.1; Smith 2003).

The estimates of all dispersion parameters were used to compare the fit of the multi-environment spatial and non-spatial models within each series. Further model comparison were provided by the accuracy of prediction of breeding values (i.e., the correlation between the true and predicted breeding values) for each parent, offspring, and family, which were computed using the following expression (Mrode 2006; page 90):

$$r_{ij} = \sqrt{\frac{\left(\widetilde{\sigma}^2_{A_{j,j}} - \mathrm{PEV}_{ij}\right)}{\widetilde{\sigma}^2_{A_{j,j}}}}$$

The acronym PEV_{ij} stands for "prediction error variance" (Henderson 1984) of predicted breeding values to tree i and site j using the "best linear unbiased predictors" of parents, offspring, and families. The PEV is calculated as the diagonal elements of the inverse of the coefficient matrix from the mixed model equations (Henderson 1984) in (13). To make the accuracies comparable across the multi-environment spatial and nonspatial models, the required variance components to set up the mixed model equations were those estimated from the multi-environment spatial model. Spearman correlations were also calculated to compare whether the ranking of predicted breeding values differed among models.

3 Results

The proposed multi-environment individual-tree mixed model for spatially adjusted data produced a reduction of the posterior means of the estimates error variances $\left(\sigma_e^2\right)$ from 7.8% to 38.3% in the LD series and from 3.9% to 14.6% in the ED series (Table 3). The reduction in the σ_e^2 of the multi-environment spatial model compared to multi-environment nonspatial model was on average 20.7% for the "S in R" (LD series), but only 8.9% for the ICB (ED series) designs. This suggests that the a priori ICB design (ED series) was more effective in removing the spatial variability than the a priori "S in R" design (LD series).

The changes in the absolute value of posterior means of the additive variance $\left(\sigma_A^2\right)$ across the five LD series trials showed small but generally inconsistent changes between the multi-environment nonspatial and spatial models, with σ_A^2 both decreasing and increasing in value from −3.9% to 5.0%. However, in the three trials of the ED series, the estimated posterior means of σ_A^2 from the multi-environment spatial model were greater than the observed multi-environment nonspatial model estimates (from 15.8% to 24.1%). A different tendency was observed for the estimate of dominance variances $\left(\sigma_D^2\right)$, where some higher values were obtained from the multi-environment spatial model in the three LD series trials with nonzero estimates. Similar posterior means of σ_D^2 were estimated for the ED series trials between both models (Table 3). The estimates h_N^2 and h_B^2 showed inconsistent changes between the series, increasing (to a maximum of 66.7% in the LD series) or decreasing (less than 11.1% in the ED series) from the multi-environment nonspatial to spatial models (Table 3).

The posterior means of additive genetic correlations ($r_{Ajj'}$) showed nonsignificant and inconsistent changes between the multi-environment spatial and nonspatial models across sites within each series, with $r_{Ajj'}$ both decreasing (by −0.06 to −0.17 correlation units in the LD series) and increasing (by 0.01 to 0.02 correlation units in the LD series and be 0.01 to 0.04 correlation units in the ED series; Table 4).

The average accuracy of predicted breeding values showed no or very small differences (i.e., usually in the third decimal place), but largely in favor of the proposed spatial MET analysis, in both series (not shown). The Spearman correlation between predicted breeding values from the two multi-environment models compared was high for both the LD and ED series and equal to 0.95 and 0.99 for parents, 0.99 and 0.99 for offspring, and 0.93 and 0.97 for families, respectively; therefore, a small amount of rank change took place for either parents or families between models.

Table 3 Posterior means (95% high posterior density interval) for the additive genetic variance $\left(\sigma_A^2\right)$, dominance genetic variance $\left(\sigma_D^2\right)$, individual narrow-sense heritability (h_N^2), individual broad-sense heritability (h_B^2), and error variance $\left(\sigma_e^2\right)$ of total height from the multi-environment nonspatial (NonSP) and spatial (SP) model

Site	MET model[a]	σ_A^2	σ_D^2	h_N^2	h_B^2	σ_e^2
Local diallel						
Jordan 2	NonSP	8.3 (5.2–10.8)	5.7 (3.1–9.3)	0.04 (0.02–0.05)	0.07 (0.05–0.09)	203.0 (196.2–210)
	SP	8.1 (4.4–11.2)	7.4 (3.9–12.4)	0.05 (0.03–0.07)	0.10 (0.07–0.13)	149.9 (144.6–155.4)
Kiyu	NonSP	6.5 (3.8–9.1)	5.5 (2.9–9.2)	0.03 (0.02–0.04)	0.06 (0.04–0.08)	209.6 (201.4–218)
	SP	6.5 (3.6–8.8)	7.0 (3.5–11.9)	0.05 (0.03–0.06)	0.10 (0.07–0.14)	129.3 (124.1–134.6)
Jordan 3	NonSP	9.0 (5.4–11.8)	5.4 (2.8–9.4)	0.05 (0.03–0.06)	0.07 (0.05–0.10)	190.1 (183.1–197.4)
	SP	9.8 (6.3–12.1)	6.0 (3–10.3)	0.06 (0.04–0.08)	0.10 (0.07–0.13)	148.4 (142.8–154.1)
Rupert 1	NonSP	17.5 (11–24)	0.0 (0.0–0.0)	0.06 (0.04–0.08)	0.06 (0.04–0.08)	269.1 (258.7–279.7)
	SP	18.2 (12.7–24.4)	0.0 (0.0–0.0)	0.07 (0.05–0.09)	0.07 (0.05–0.09)	243.1 (233.6–252.7)
Humptulips	NonSP	12.2 (7.9–18)	0.0 (0.0–0.0)	0.06 (0.04–0.09)	0.06 (0.04–0.09)	182.7 (176.2–189.3)
	SP	11.6 (8.6–14.3)	0.0 (0.0–0.0)	0.06 (0.05–0.08)	0.06 (0.05–0.08)	168.4 (162.8–174.2)
Elite diallel						
Stove	NonSP	21.9 (13.1–35.6)	10.0 (5.5–0)	0.11 (0.07–0.18)	0.17 (0.12–0.24)	166.3 (157–175)
	SP	18.5 (12.5–26.7)	9.9 (5.6–0)	0.11 (0.08–0.16)	0.17 (0.13–0.23)	142.1 (135.2–149)
Tuplana	NonSP	19.0 (11.5–30.6)	0.9 (0.4–0)	0.09 (0.05–0.13)	0.09 (0.06–0.14)	204.6 (194.8–214.5)
	SP	14.5 (9.8–20.2)	0.9 (0.4–0)	0.07 (0.05–0.10)	0.08 (0.05–0.10)	187.5 (179.4–195.7)
Michelsen	NonSP	10.1 (5.8–16.2)	7.6 (4–0)	0.06 (0.04–0.09)	0.10 (0.07–0.15)	158.4 (151.5–165.4)
	SP	7.6 (4.9–10.7)	7.4 (3.9–0)	0.05 (0.03–0.07)	0.09 (0.07–0.13)	152.3 (146.1–158.7)

Note that additive genetic, dominance genetic and error variances were divided by 100 for optimal display

[a] MET model NonSP: multi-environment nonspatial model with replicates and incomplete block effects, MET model SP: multi-environment spatial model

Table 4 Posterior means (95% high posterior density interval) of the additive genetic correlations ($r_{A_{jj'}}$) between sites within each series for the multi-environment spatial (above diagonal) and nonspatial (below diagonal) models

Local diallels	Jordan 2	Kiyu	Jordan 3	Rupert 1	Humptulips
Jordan 2		0.87 (0.74–0.93)	0.73 (0.59–0.86)	0.64 (0.34–0.81)	0.88 (0.83–0.94)
Kiyu	0.87 (0.77–0.92)		0.77 (0.66–0.86)	0.80 (0.67–0.90)	0.73 (0.62–0.82)
Jordan 3	0.67 (0.53–0.83)	0.60 (0.49–0.74)		0.61 (0.43–0.76)	0.89 (0.82–0.94)
Rupert 1	0.66 (0.41–0.81)	0.81 (0.71–0.91)	0.49 (0.28–0.65)		0.61 (0.39–0.75)
Humptulips	0.82 (0.71–0.90)	0.58 (0.47–0.69)	0.90 (0.85–0.96)	0.50 (0.27–0.67)	
Elite diallels	Stove	Tlupana	Michelsen		
Stove		0.89 (0.80–0.94)	0.79 (0.63–0.89)		
Tuplana	0.91 (0.83–0.96)		0.97 (0.95–0.99)		
Michelsen	0.83 (0.70–0.92)	0.98 (0.96–0.99)			

4 Discussion

This paper presents an extension of the method of Cappa and Cantet (2007) to account for the spatial variation in multi-environment large forest genetic trials using a multi-environment individual-tree model with additive and full-sib family genetic effects in a two-stage spatial MET analysis. Our approach differs from previous works (Ye and Jayawickrama 2008; de la Mata and Zas 2010), in that we analyzed the spatially adjusted data by fitting a multi-environment individual-tree mixed model with an unstructured additive genetic (co)variance matrix between trials, allowing both, different additive genetic variances for each trial and covariances (and correlations) for all different pairs of trials. Additionally, the proposed multi-environment spatial individual-tree model also accounted for the family (i.e., SCA) variance for each trial. Ye and Jayawickrama (2008) fitted one- and two-stage family and individual-tree mixed model with a common additive variance and covariance across trials. However, the assumption of a constant genetic variance and covariance across trial for the genetic (co)variance matrix may lead to inaccurate biased predictions of genetic values, and thus may affect selection decision and estimates of genetic gain (Costa e Silva et al. 2005). Using a two-stage approach, de la Mata and Zas (2010) analyzed spatially adjusted data using a family mixed model with different constraints to the genetic (co)variance matrix across trial, from a simple diagonal matrix with the same genetic variance for all the trials to an unstructured family (co)variance structure. The multi-environment spatial family model reduces computational requirements by a reduction in the size of the system of equations to be solved, compared with the multi-environment spatial individual-tree model; however, only half-sib parents are being evaluated.

In the analyses here reported, additive genetic correlation across trials within each series, and additive, family, and error variances, and functions of these estimates for each trial, were successfully estimated for eight large second-generation trials of western hemlock using a Bayesian procedure coupled with a Markov chain Monte Carlo technique (Gibbs sampling). An alternative approach for estimating dispersion parameters is the use of restricted maximum likelihood procedures (REML; Patterson and Thompson 1971). However, when the number of trees per site and the number of sites is large, the estimation of parameters for a large unstructured additive genetic (co) variance matrix while accounting for the spatial variability within trial using the average information-REML algorithm, becomes computationally difficult (i.e., failing to converge; Ye and Jayawickrama 2008; Ding et al. 2008). Ye and Jayawickrama (2008) highlighted the difficulties in convergence when fitting an individual-tree mixed model in one-stage assuming variance homogeneity across trials, an independent error variance between sites and a separable two-dimensional AR(1) error covariance structures for rows and columns within trials, using average information-REML. Ding et al. (2008) also fitted an individual-tree model with additive and family genetic effects using the average information algorithm with unstructured genetic covariances between trials, an independent error variance between sites, and a separable two-dimensional AR(1) error covariance structure within trial from data on 12,460 individuals originating from 169 to 216 full-sib families of *Pinus radiata* D. Don at five sites. They also found trouble with convergence when fitting the spatial model in a one-stage process. Using the average information-REML algorithm, Hardner et al. (2010) proposed reducing the number of parameters needed to describe the family (co)variance matrix across trials by using a factor-analytic parameterisation in a one-stage spatial analysis. Their factor analysis was successfully applied to the estimation of genetic parameters of diameter at breast height for a total of 841 genotypes of eucalypt hybrid families assessed across 21 trials. However, Smith et al. (2001) noted difficulties with convergence of these factor-analytic models using the average information algorithm. Misztal (2008) suggested that for complex models (i.e., numerous traits and multiple random effects) and poor starting values (i.e., initial estimates of (co)

variance components), the average information algorithm can "overshoot", resulting in either very slow convergence or parameter estimates that lie out of the parameter space. In spite of Bayesian approaches having better properties for parameter estimation in many situations, it is still difficult to compare frequentist and Bayesian estimators, 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).

Our results showed that the fit of a multi-environment spatial model to eight large second-generation full-sib progeny trials of *T. heterophylla* (Raf.) Sarg. with a priori "S in R" and ICB designs, displayed a consistent reduction in the estimated posterior mean of σ_e^2 within each trial (average reduction of 16.0%) relative to a conventional multi-environment nonspatial model. As expected, these reductions in σ_e^2 confirm that there are spatial variations that are not adequately accounted for the a priori design effects. The reduction of the estimates σ_e^2 agree with results of Ye and Jayawickrama (2008) who showed an average reduction of 32.5% across the five Douglas fir MET tests, and Ding et al. (2008) across five radiata pine trials, with an average reduction of 1.3%. Moreover, in our data, the multi-environment spatial model produced an increase in the precision (i.e., narrower values for the 95% high posterior density interval, with an average increase of 18.8%) for the estimation of σ_e^2. The estimated posterior means of σ_A^2 within each trial, showed no consistent patterns (i.e., both decreasing and increasing) when compared between the multi-environment spatial model and the conventional MET analysis. Similar results have been found by Ye and Jayawickrama (2008) and Ding et al. (2008) who reported a variation from −10.1% to 34.3% and from −58.8% to 14.8%, respectively, using a one-stage spatial analysis. The multi-environment spatial model increased the average estimates of the posterior means of h_N^2 and h_B^2 (on average, across all sites of both series by 11.2% and 18.5%, respectively), though these estimates showed inconsistent changes between series. The higher individual-tree heritabilities estimated for the proposed spatial MET analysis in each of the five trials of the LD series, are likely associated with the higher reduction in the estimation of σ_e^2 obtained for the LD (20.7%), relative to the ED (8.9%) series. Consistent increases (from 3.7% to 65.4%) for individual-tree heritabilities were also found by Ye and Jayawickrama (2008). Interestingly, our results indicated that the average accuracy of predicted breeding values showed indistinguishable differences, and little re-ranking took place for parents, offspring, and families in both series. This differs from the results reported by Ye and Jayawickrama (2008) where the spatial analysis of MET data provided more accurate breeding values (ranged from 1% to 20% for parent and from 1% to 25% for offspring) than the classical nonspatial model.

The effect of the spatial analysis on the genotype by environment interaction from data of multi-environment forest genetic trials has not been extensively studied. The analysis of the western hemlock data revealed that the multi-environment spatial individual-tree model, showed no consistent changes on the $r_{Ajj'}$ between pairs of sites within both LD and ED series. These results agree with those obtained by Ding et al. (2008) who reported an inconsistent effect on estimate genetic correlations (changes ranged from −15.4% to 20.0%) using a one-step spatial approach, and Ye and Jayawickrama (2008) who showed a decrease from 4.3% to 14.0% in the genotype by environment interaction from the conventional (nonspatial) multi-environment to the multi-environment spatial model.

Although applying a one-stage spatial MET analysis to compare the results with those obtained from the proposed procedure (i.e., two-stage spatial MET analysis) was not computationally possible, Ye and Jayawickrama (2008) compared the effect of one- vs. two-stage spatial MET. They found that both methodologies had similar effects on the derived parameters such as heritability, accuracy of breeding values, rank correlations of breeding values (≥0.99) and predicted genetic gains (a difference<0.2%), and concluded that the two-stage analysis is a suitable method for fitting large regular spatial MET data. However, they reported that the family×trial interactions in the two-stage spatial analysis were slightly higher than those from the one-stage spatial analysis. Cullis et al. (1998) compared one- vs. two-stage spatial analysis in barley variety field trials. In this case, the two-stage approach implies that the genotype effects were conditionally fixed for each experiment during the first stage to obtain the variety means. However, using data from all sites, in the second stage of the analysis genotype effects were considered random in the mixed model used. They reported similar results from the one- versus two stage-spatial analyses, with the exception of the genotype×environment variance for the two-stage being slightly higher than for the one-stage analysis.

As a final comment, regardless of being able to complete the two-stage spatial MET analysis with an unstructured additive genetic (co)variance matrix between trials and an SCA variance by trial, computations were still demanding. However, as computing capabilities of desktop computers increases, analyses such as the one presented here, with very large forest genetics trial data will become more common.

5 Summary and conclusions

Given the necessity of accounting for both genetic correlations between trials and spatial variability within trials, we presented an extension of the single-environment spatial model of Cappa and Cantet (2007). We developed a two-

stage spatial MET approach for the analysis of large forest genetic trials, using a multi-environment individual-tree model with additive and full-sib family genetic effects. Previous attempts at a spatial MET analysis with a two-stage approach have been limited by simpler models (i.e., without accounting for the additive relationship among trees) or restrictive genetic (co)variance structures across environment. The procedure of fitting a more complex model for multi-environment large forest trials was empirically illustrated using the height data at age 10, from eight large second-generation trials of western hemlock formed by recombining the top 150 first-generation parents. Although computationally demanding, the (co)variance parameters (and functions of them) were successfully estimated from the proposed multi-environment spatial individual-tree model using a Bayesian method by means of the Gibbs sampling. The proposed two-stage spatial MET approach produced better results than the commonly used non-spatial MET analysis. The multi-environment individual-tree model of spatially adjusted data displayed a consistent and important reduction of the posterior mean as well as an increase in the precision of σ_e^2, and an average increase in the posterior means of h_N^2 and h_B^2.

Acknowledgments The authors would like to thank to the associate editor and two anonymous referees for their insightful comments on an earlier version of this manuscript.

Funding This research was supported by funds from British Columbia Forests Service and Range, and support to the senior author as a post-doctoral fellow at the University of British Columbia.

References

Cantet RJC, Birchmeier AN, Steibel JP (2004) Full conjugate analysis of normal multiple traits with missing records using a generalized inverted Wishart distribution. Genet Sel Evol 36:49–64

Cappa EP, Cantet RJC (2006) Bayesian inference for normal multiple-trait individual-tree models with missing records via full conjugate Gibbs. Can J For Res 36:1276–1285

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, 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 Genetica 60:25–35

Casanoves F, Macchiavelli R, Balzarini M (2005) Error variation in multienvironment peanut trials: within-trial spatial correlation and between-trial heterogeneity. Crop Sci 45:1927–1933

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, Dutkowski GW, Borralho NMG (2005) Across-site heterogeneity of genetic and environmental variances in the genetic evaluation of *Eucalyptus globulus* trials for height growth. Ann For Sci 62:183–191

Cullis BR, Gogel B, Verbyla A, Thompson R (1998) Spatial analysis of multi-environment early generation variety trials. Biometrics 54: 1–18

de la Mata R, Zas R (2010) Transferring Atlantic maritime pine improved material to a region with marked Mediterranean influence in inland NW Spain: a likelihood-based approach on spatially adjusted field data. Eur J For Res 129:645–658

Ding M, Tier B, Dutkowski G (2008) Multi-environment trial analysis for *Pinus radiata*. N Z J For Sci 38:143–159

Dutkowski GW, Costa e Silva J, Gilmour AR, Lopez GA (2002) Spatial analysis methods for forest genetic trials. Can J For Res 32:2201–2214

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

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

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman Group Ltd., New York

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis. Chapman and Hall, New York, USA

Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2006) ASReml User Guide Release 2.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Gwaze DP, Woolliams JA (2001) Making decisions about the optimal selection environment using Gibbs sampling. Theor Appl Genet 103:63–69

Hardner CM, Dieters M, Dale G, DeLacy I, Basford KE (2010) Patterns of genotype-by-environment interaction in diameter at breast height at age 3 for eucalypt hybrid clones grown for reafforestation of lands affected by salinity. Tree Genet Genomes 6:833–851

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Guelph, ON, Canada

King JN, Cress DW (1991) Breeding plan proposal for western hemlock cooperative tree improvement. The Hemlock Tree Improvement Cooperative. Internal report, 14 p.

King JN, Cartwright C, Cress DW (1998) Western hemlock tree improvement: selection of P-1 parents. The Hemlock Tree Improvement Cooperative. Internal report.

Jayawickrama KJS (2003) Genetic improvement and deployment of western hemlock in Oregon and Washington: review and future prospects. Silvae Genetica 52:26–36

Li Y, Dutkowski GW, Apiolaza LA, Pilbeam D, Potts BM (2007) The genetic architecture of a *Eucalyptus globulus* full-sib breeding population in Australia. For Genet 12:167–179

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

Misztal I (2008) Reliable computing in estimation of variance components. J Anim Breed Genet 125:363–370

Mrode R (2006) Linear models for the prediction of animal breeding values. CABI, Wallingford, UK

Oakey H, Verbyla A, Cullis BR, Wei X, Pitchford WS (2007) Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. Theor Appl Genet 114:1319–1332

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

Schutz WM, Cockerham CC (1966) The effect of field blocking on gain from selection. Biometrics 22:843–863

Smith AB, Cullis BR, Thompson R (2001) Analysing variety by environment data using multiplicative mixed models and adjustments for spatial field trends. Biometrics 57:1138–1147

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

Sorensen D, Gianola D (2002) Likelihood, Bayesian, and MCMC methods in quantitative genetics. Springer, New York, USA

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

Van Tassell CP, Van Vleck LD (1996) Multiple-trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. J Anim Sci 76:2586–2597

Williams ER, Matheson AC (1994) Experimental design and analysis for use in tree improvement. CSIRO, Melbourne, Australia

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

Zas R (2006) Iterative kriging for removing spatial autocorrelation in analysis of forest genetic trials. Tree Genet Genomics 2: 177–185