Tree Genetics & Genomes (2006) 2: 177–185
DOI 10.1007/s11295-006-0042-4

ORIGINAL PAPER

Rafael Zas

Iterative kriging for removing spatial autocorrelation in analysis of forest genetic trials

Received: 2 September 2005 / Revised: 28 November 2005 / Accepted: 8 March 2006 / Published online: 8 April 2006

Abstract Conventional analysis of spatially correlated data in inadequately blocked field genetic trials may give erroneous results that would seriously affect breeding decisions. Forest genetic trials are commonly very large and strongly heterogeneous, so adjustments for micro-environmental heterogeneity become indispensable. This study explores the use of geostatistics to account for the spatial autocorrelation in four *Pinus pinaster* Ait. progeny trials established on hilly and irregular terrains with a randomized complete block design and large blocks. Data of five different traits assessed at age 8 were adjusted using an iterative method based on semivariograms and kriging, and the effects on estimates of variance components, heritability, and family effects were evaluated in relation to conventional analysis. Almost all studied traits showed nonrandom spatial structures. Therefore, after the adjustments for spatial autocorrelation, the block and family × block variance components, which were extremely high in the conventional analysis, almost disappeared. The reduction of the interaction variance was recovered by the family variance component, resulting in higher heritability estimates. The removal of the spatial autocorrelation also affected the estimation of family effects, resulting in important changes in family ranks after the spatial adjustments. Comparison among families was also greatly improved due to higher accuracy of the family effect estimations. The analysis improvement was larger for growth traits, which showed the strongest spatial heterogeneity, but was also evident for other traits such as straightness or number of whorls. The present paper demonstrates how spatial autocorrelation can drastically affect the analysis of forest genetic trials with large blocks. The iterative kriging procedure presented in this paper is a promising tool to account for this spatial heterogeneity.

Keywords Semivariogram · Geostatistics · Spatial heterogeneity · *Pinus pinaster* · Genetic parameters · Heritability · Tree breeding

R. Zas (✉)
Departamento de Producción Forestal,
Centro de Investigacións Forestais e Ambientais de Lourizán,
Apdo. 127,
Pontevedra, 36080, Spain
e-mail: rzas.cifal@siam-cma.org
Tel.: +34-986805067
Fax: +34-986856420

Introduction

Forest genetic trials (FGT) are usually very large because of the high number of genetic entries and the large space needed for individual plants. Moreover, FGT are often established on hilly and irregular terrains, leading to a spatial heterogeneity within replicated blocks that may limit the efficiency of the tests. In fact, spatial gradients and patchiness have been shown to be the norm in FGT (Costa-Silva et al. 2001; Dutkowski et al. 2002; Fu et al. 1999; Hamann et al. 2002; Magnussen 1990), even in apparently homogeneous farm-field tests with intensive site management (Joyce et al. 2002). When spatial heterogeneity is present, near neighbors are more similar than far neighbors, i.e., data are autocorrelated, and the requirement of data independence in standard parametric statistics is violated (Legendre 1993). Despite having been shown that the conventional analyses of spatially correlated observations in genetic trials may be disastrous (Magnussen 1993a), the incorporation of spatial correlation in data analysis is just slowly emerging.

Several sophisticated incomplete block (IB) designs with small blocks can successfully account for much of this spatial heterogeneity (Fu et al. 1998). Some of these experimental designs, such as alpha designs (Whitaker et al. 1999), have become popular in recent forest genetic trials, but randomized complete block design (RCB) are still being used and were the typical designs in FGT established some decades ago (Loo-Dinkins 1992; Magnussen 1993b). As the evaluation of FGT must be delayed to achieve a reliable prediction of mature performance, there are still many forest tests with RCB

and large blocks that still require evaluation. Nevertheless, the use of IB designs may be unable to account for the complete spatial heterogeneity over short distances (Magnussen 1990), and a mismatch between block boundaries and actual patterns of site variation will occur regardless of the design employed because environmental variation tends to be continuous and smooth (Dutilleul 1993; Dutkowski et al. 2002; Ettema and Wardle 2002).

Interplant competition may be another particularity of older forest trials, which affects correlation between neighbors (Magnussen 1994). To minimize competition between different genetic entries, FGT are commonly established with multi-tree contiguous plots (Magnussen 1993b). In multi-tree contiguous plots, substantial environmental covariance is confounded with genetic covariances in a given plot, resulting in biased genetic parameter estimations and large entry-by-block interaction variance, which may substantially reduce the precision of a field test (Loo-Dinkins 1992).

Besides IB designs, several analytic approaches have been proposed to account for environmental variation and improve the efficiency of the analysis of field genetic trials. Spatial analysis, such as trend analysis, neighbor analysis, or nearest neighbor analysis, have shown to enhance the analysis of agricultural field trials, and are becoming quite popular in agronomy (e.g., Gilmour et al. 1997; Grondona et al. 1996; Qiao et al. 2000, 2004). Although not so commonly used, spatial analytical methods have also given successful results in forest trials (Anekonda and Libby 1996; Costa-Silva et al. 2001; Dutkowski et al. 2002; Hamann et al. 2002; Joyce et al. 2002; Magnussen 1993a, 1994). Among the different methods, geostatistics have been shown to be an excellent tool to model the spatial variation of soil properties (Gallardo 2003; Gallardo and Covelo 2005; López and Arrúe 1995), and also to remove micro-environmental heterogeneity in genetic trials (Hamann et al. 2002).

Galicia (NW Spain) is a hilly and heterogeneous region with acidic and infertile forest soils. Productivity of forest plantations in this region is highly dependent on soil properties (Merino et al. 2003; Sánchez-Rodríguez et al. 2002; Zas and Serrada 2003), which are known to have a small-scale spatial heterogeneity (Gallardo 2003; Gallardo and Covelo 2005). In four RCB *Pinus pinaster* Ait. progeny tests in Galicia, this spatial heterogeneity was reflected by the large proportion of total variance explained by differences among blocks and a large family-by-block interaction (Zas et al. 2004a). Spatial autocorrelation seemed to be very important in these trials, and the selected material displayed an extreme and incongruous environmental sensitivity that should be further investigated. The objectives of the present study were 1) to analyze the spatial structure of different variables in these progeny trials, and 2) to evaluate the effect of spatial heterogeneity on the efficiency of conventional statistical analysis for RCB trials. The spatial variation was modeled using semivariograms and kriging. An iterative method based on these techniques is proposed to remove the spatial heterogeneity from the data.

Materials and methods

Genetic material and test sites

Four *P. pinaster* progeny trials established in Galicia (NW Spain) in 1994–1995 for breeding purposes were used in this study. Each site includes between 79 and 98 open pollinated families obtained from plus-trees selected within the coastal area of Galicia. All sites followed a randomized complete block design with ten blocks, five-tree row plots, and 3×3 m spacing. The trials are situated on hilly and irregular sites representative of the breeding area, where the improved material is to be used. Soils are shallow, acid, coarse-textured, and have relatively low levels of nutrients. Details on the progeny tests, the genetic material, and the genetic parameter estimates for growth, stem form, and branching traits using a standard approach have been published previously (Zas et al. 2004a,b).

The traits considered in this study were total height (H), breast height diameter (D), stem straightness score (STR: 1=straight to 6=very crooked), number of whorls (WH), and branch angle score (ANG: 1=branches at a flat angle to 3=steep branching), i.e., a representation of growth, stem form, polyciclism, and branching habit, traits that are expected to have different spatial pattern structures. All traits were assessed in all living and not badly suppressed plants at age 8 since planting.

Spatial analysis

Because not all saplings were planted following a regular grid, the approximate x–y coordinates of each tree were estimated using orthographic aerial photographs at ~1:1,000 scale. Plantation lines were easily identified and coordinates of the beginning and end of each line were obtained from the photographs. The coordinates of the trees within each line were determined by dividing the plantation lines in $n-1$ equal-length segments, where n is the number of trees in the given line.

Residuals of each variable after subtraction of the family effects were used to explore the spatial structure of the data. A one-way analysis of variance with the family effect considered random was carried out, and the best linear unbiased predictors (BLUPs) of the family effects were obtained using the MIXED procedure in the SAS system (SAS Institute 1999). The spatial structure of the resulting residuals was modeled using a semivariogram, which plots the semi-variance among trees as a function of the distance

between them (Cressie 1993). The semi-variance $\gamma(h)$ was calculated as:

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}[z(s_i) - z(s_{i+h})]^2$$

where n is the number of pairs of neighbor trees separated by distance h (called the lag distance), $z(s_i)$ is the measured trait value for a tree located at s_i, and $z(s_{i+h})$ is the value for a tree located at a distance h from s_i. For randomly distributed data, little change in the semivariance will be obtained when h increases, and the semivariogram will be essentially flat. If spatial dependence is present, semivariance will be small at short distances, will increase at intermediate distances, and will reach an asymptote at large distances (Fig. 1). Experimental semivariograms were constructed using the VARIOGRAM procedure in SAS (Sas Institute, 1999).

An exponential theoretical semivariogram $(\gamma(h) = c_n + c_o(1 - e^{-h/a0}))$ was fitted to the experimental semivariogram using the NLIN procedure in SAS (SAS Institute 1999). This theoretical semivariogram was used to partition the variation of residuals into spatially autocorrelated variation and random error with the kriging method. Kriging computes smooth surfaces of best linear unbiased predictions of values on a spatial grid based on the spatial structure defined by the theoretical semivariogram. At each point, the value of the kriging estimate would be interpreted as the amount of the trait that is due to the spatial position. Thus, the kriging values at each tree location were used to correct the original values of each variable in relation to their spatial variation. The kriging analysis was performed using the KRIG2D SAS procedure (SAS Institute 1999).

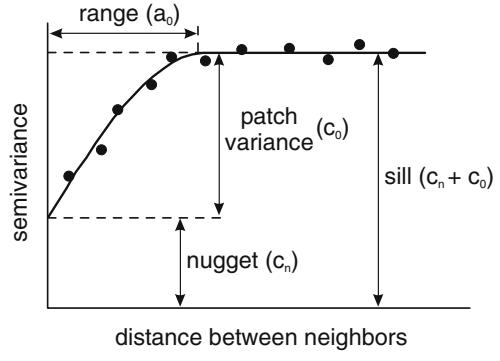

Fig. 1 Hypothetical standard semivariogram showing the observed semivariance for distance classes (*dots*) and the fitted model (*solid line*). Three parameters derived from the semivariogram characterize the spatial structure pattern: (**a**) the distance at which the asymptote begins (a_o), which indicates the range or patch size of heterogeneity below which data are stochastically dependent, (**b**) the asymptote, called *sill* (c_o+c_n), which is an estimate of total population variance, and (**c**) the *nugget* (c_n) or intercept at distance zero, which represents the variance due to sampling error and/or spatial dependence at scales not explicitly sampled. The larger the c_o/c_n ratio, the greater the intensity of the spatial structure

Genetic analysis

Values adjusted for the spatial structure were analyzed using the following random model:

$$(y_{ijk} - k_{ijk}) = \mu + F_i + B_j + FB_{ij} + \delta_{ijk}$$

where y_{ijk} is the value of the original variable of the kth tree of the ith family in the jth block, k_{ijk} is the kriging estimate at the position of that tree, μ is the overall mean, F_i, B_j, and FB_{ij} are the random effects of family i, block j, and the corresponding interaction, and δ_{ijk} is the spatially independent error. The same statistical model was also used to analyze uncorrected original values.

Variance components and best linear unbiased predictors (BLUPs) of family effects were estimated using the restricted maximum likelihood method of the MIXED procedure in SAS (SAS Institute 1999). Individual h_i^2 and family h_f^2 heritabilities were estimated as:

$$h_i^2 = \frac{\sigma_A^2}{\sigma_f^2 + \sigma_{fb}^2 + \sigma_e^2}$$

$$h_f^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_{fb}^2/B + \sigma_e^2/BN}$$

where σ_A^2 is the additive variance which was assumed to be $\sigma_A^2 = 4\sigma_f^2$, σ_f^2 is the family variance, σ_{fb}^2 is the variance of the family-by-block interaction, σ_e^2 is the residual variance, and B and N are the number of blocks and the harmonic mean of the number of trees per plot. Approximate standard errors of heritabilities were estimated according to Wright (1976).

The accuracy of the BLUPs of family effects ($r_{g\hat{g}}$), i.e. the correlation between the true and the predicted genetic value, was calculated as (Dutkowski et al. 2002):

$$r_{g\hat{g}} = \sqrt{1 - \frac{PEV}{\sigma_f^2}}$$

where PEV is the prediction error variance of family BLUPs (the square of the standard errors) and σ_f^2 is the family variance.

Iterative procedure

The spatial analysis presented above used residuals adjusted for family effects. However, if spatial heterogeneity is quantitatively important, the estimates of the family effects from the original values could be strongly biased. The family BLUPs estimates after adjustment for spatial heterogeneity are supposed to be better predictors of true

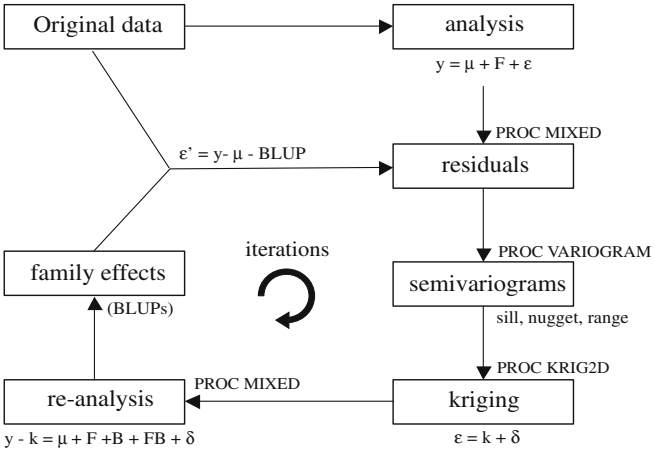

Fig. 2 Scheme of the proposed iterative spatial adjustment procedure. The spatial structure of the residuals adjusted for family effects (ε) was determined using semivariograms. An exponential theoretical semivariogram was fitted to the observed semivariogram and the resulting *sill*, *nugget*, and *range* were used to divide the variation of residuals into spatially autocorrelated variation (k) and random error (δ) with the kriging method. The original values adjusted for the spatial structure (y-k) were re-analyzed and a new estimation of the family effects (BLUPs) was obtained and used to generate new residuals (ε') from the original data. The process was repeated iteratively until convergence of the BLUPs estimations of the family effects

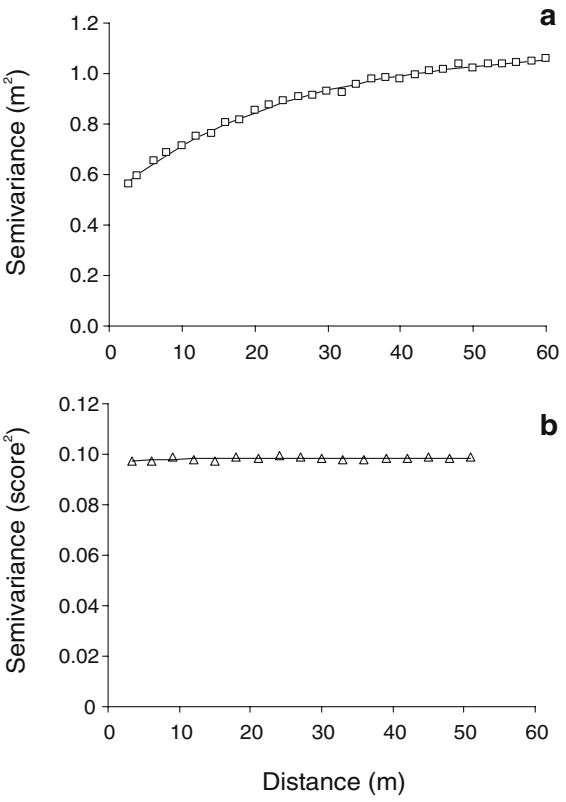

Fig. 3 Examples of the semivariograms of residuals after subtracting family effects (first iteration) for height in site B (**a**), and for the branch insertion angle in site A (**b**). The reduction of the semivariance at short distances in (**a**) indicated a patchy structure, whereas, the flat semivariogram in (**b**) indicated random spatial variation

family effects and can be used to obtain new residuals from the original data. A new semivariogram and kriging estimates were obtained from these new residuals. These kriging estimates were then used to correct original values, and a new estimation of family effects was obtained. This process was repeated iteratively until convergence (stability of family ranks) of the BLUPs estimates of family effects (Fig. 2).

Results

Spatial structure

Residuals after subtracting family effects revealed nonrandom spatial structures for almost all variables and sites. The exponential theoretical semivariogram fitted well to the observed semivariogram for all traits and sites, except for the branch angle. The theoretical semivariograms explained 98 to 99% of the observed variation for growth traits (r^2=0.98 to 0.99) and 80–97% for stem form and the number of whorls. These semivariograms indicated that data from near neighbors were more similar than those from far neighbors, revealing an spatial autocorrelation, but the pattern and intensity of this spatial dependence varied greatly among the different traits and sites. Spatial heterogeneity was much more evident for growth traits, as revealed by higher path-nugget variance ratios (c_o/c_n), but it was also present for stem form and polyciclism traits (Table 1). Height growth was the trait for which the intensity of the spatial structure was greatest, whereas, the branch angle showed weak or nearly null spatial dependence. A comparison between the semivariograms for these two traits can be observed in Fig. 3. According to the

Table 1 Patch size (range in meters, a_o) and intensity of the spatial pattern structure (patch variance to nugget variance ratio, c_o/c_n) derived from the theoretical semivariograms fitted to different variables[1] adjusted for family effects (first iteration) in the four test sites

	H		*D*		STR		WH		ANG	
	a_o (m)	c_o/c_n	a_o (m)	c_o/c_n	a_o (m)	c_o/c_n	a_o (m)	c_o/c_n	a_o (m)	c_o/c_n
Site A	20.5	0.51	20.9	0.34	45.7	0.14	21.2	0.12	5.1	0.10
Site B	23.8	1.16	13.1	0.85	21.3	0.21	14.3	0.16	8.9	0.24
Site C	80.7	1.01	100.9	0.72	30.7	0.22	47.3	0.42	26.6	0.10
Site L	29.9	0.98	27.9	0.62	6.8	0.29	20.0	0.08	15.4	0.04

[1] *H* Height, *D* diameter, *STR* straightness, *WH* number of whorls, *ANG* branch angle

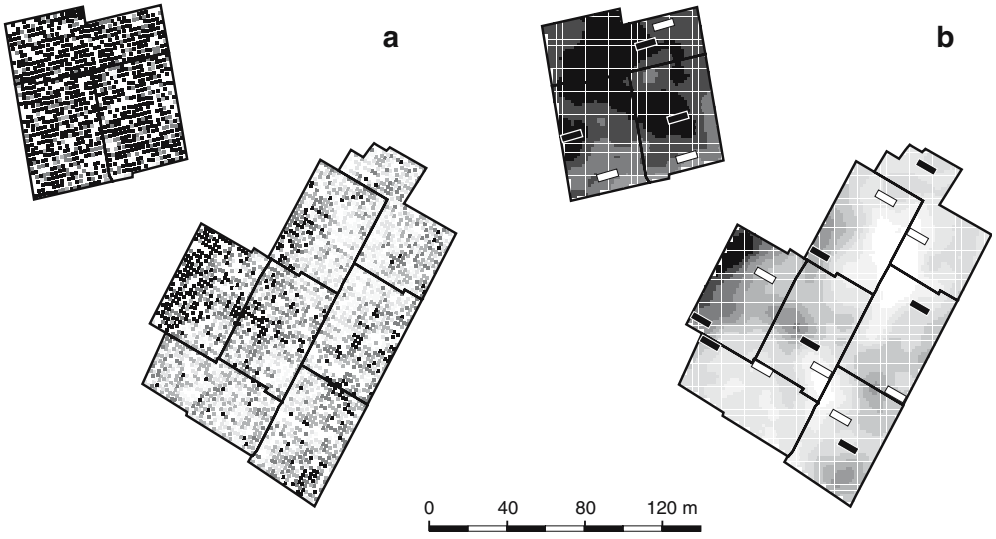

Fig. 4 Plot of height residuals after subtracting family effects in site C showing nonrandom spatial variation (**a**) and modeling of this variation through iterative kriging (**b**). Values range from <−1 STD (*white*) to >+1 STD (*black*) by 0.5 STD steps, where STD is the standard deviation of the original variable. *Black lines* are the block boundaries. The locations of two contrasting families that showed drastic rank changes after spatial adjustment (see Fig. 7) are indicated in (**b**) by *black* (family 1023) and *white* (family 2078) *rectangles*. Note that the black family was randomly placed in better microsites than the white one in almost all blocks

differences between the c_o/c_n ratios, the flat semivariogram for the branch angle contrasts with the clear reduction of the semi-variance at short distances for height growth. The spatial distribution of residuals for height growth in site C is presented, as an example, in Fig. 4a.

The range (a_o) or patch size of the exponential theoretical semivariograms was also greatly variable but, in general, clearly lower than the block size (~0.4 ha), indicating a spatial heterogeneity within blocks that implies a violation of the block design assumptions. For growth traits, the path size varied between approximately 15 to 30 m, with the exception of site C, where the patchy structure was confounded with a gradient trend extending for over several blocks (Fig. 4), leading to longer range estimates (Table 1).

The values presented in Table 1 and Fig. 3 corresponded to the first iteration of the spatial analysis, which is typical of what one will initially observe when exploring the spatial structure of a trait in a genetic trial. After successive iterations, i.e., successive estimations of the family effects to be subtracted to the original values, the spatial structure remained similar, with a slight increase (5–15%) in the c_o/c_n ratio in all cases.

The example of the spatial variation of height in site C, modeled through kriging after successive iterations, is presented in Fig. 4b.

Influence on variance components

When the original variables were analyzed without adjustment for spatial autocorrelation, the block variance for growth traits was extremely high (13–44%), indicating important spatial variation at large scale (Fig. 5). Family-by-block interaction was also surprisingly and incongruently high (10–26%), while family variance was proportionally insignificant (0.7–3.6%). In homogeneous blocks, these results would indicate an extreme genotypic sensibility to within-site variation, but, in fact, the excessive family-by-block interaction variance is a consequence of the patchy within-block heterogeneity and the experimental design with five contiguous tree plots. In the case of straightness, number of whorls, and branch angle, the block and family-by-block interaction variances were not so large, but were comparatively important in relation to the family variance.

After correction for spatial heterogeneity through iterative kriging, the scene became completely different. A great proportion of the original variation was described by the spatial structure of the data, as reflected by the

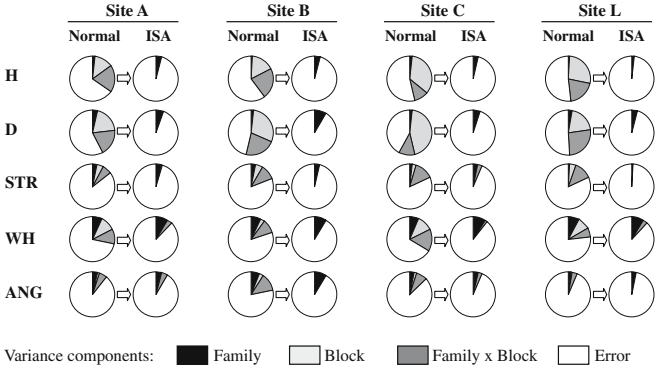

Fig. 5 Effect of the iterative spatial analysis (ISA) procedure on the variance component estimates for different variables in the four test sites in relation to the standard analysis (normal). After the spatial adjustment, the block variance disappeared and the block by family variance was greatly reduced. The reduction of the interaction variance was partially recovered by the family variance. See Table 1 for variable codification

important proportion of covariance removed after the spatial adjustments (42–67% for growth traits and 5–40% for the others). When this covariance was removed, differences among blocks disappeared, and the family-by-block interaction approached zero or was substantially reduced in all cases (Fig. 5). This reduction in variance was accounted by an increase in the relative value of both the residual and family variances. The effects on variance components were reflected in an increase of heritability, especially for growth traits (Table 2). After the iterative process, individual heritability increased from 0.04–0.11 to 0.10–0.17 for diameter, and from 0.13–0.18 to 0.17–0.33 for height growth (Table 2), whereas, family heritability increased from 0.17–0.40 to 0.50–0.64 for diameter, and from 0.41–0.52 to 0.65–0.78 for height growth (data not shown). The asymptotic approximation of individual heritability estimates through successive steps of the iterative procedure is presented for height growth in Fig. 6a. Note that the estimations after the first step can be fairly under- or overestimated.

Some effective local environmental control via blocking was evident through the high block variance for growth traits in all sites (13–44%). However, the covariance removed by means of spatial modeling was clearly higher (42–67%) indicating that the block structure was not enough to account for the complete spatial heterogeneity. The relative efficiency of blocking, measured as the ratio between the block variance and the removed spatial covariance, varied between 0.28 to 0.73 for growth traits, being higher for those sites and traits with large range (a_0) estimations, i.e., with large-scale gradients.

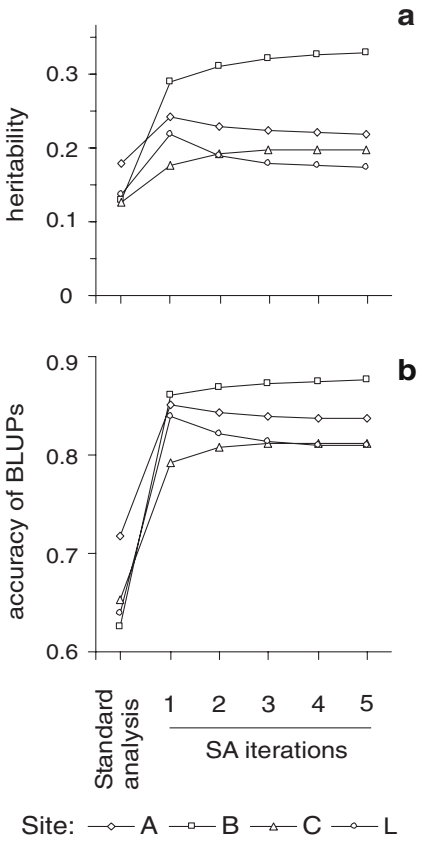

Fig. 6 Asymptotic approximation of **a** the heritability estimate and **b** the accuracy of family BLUPs for height growth through successive iterations of the spatial analysis (*SA*) procedure. Note that just one step of the spatial adjustment can give unbiased heritability estimations

Influence on family effect estimations

The coefficients of determination (r^2) between the family effects (BLUPs) using the standard analysis and the iterative spatial analysis are shown in Table 2. A high correspondence between both estimations can be observed in those variables that displayed a relatively heterogeneous spatial structure, i.e., STR, WH, and ANG. On the contrary, the estimation of the family effects for growth traits, which showed to be noticeably heterogeneous, substantially varied after correction for spatial autocorrelation. It should be noted that these relationships in growth traits correspond with important family rank changes (up to 40–50 positions, Fig. 7) that would implicate a serious impact on selection breeding decisions. To better illustrate the impact of spatial

Table 2 Effects of the spatial analysis procedure on the heritability estimates and coefficient of determination (r^2) between family effects (BLUPs) estimated using the standard (N) and the iterative spatial analysis (ISA) approach

Variable[1]	Model	Site A		Site B		Site C		Site L	
		h_i^2	r^2	h_i^2	r^2	h_i^2	r^2	h_i^2	r^2
D	*N*	0.11±0.03	0.75	0.04±0.02	0.69	0.09±0.03	0.81	0.04±0.02	0.59
	ISA	0.15±0.04		0.17±0.04		0.16±0.04		0.10±0.03	
H	*N*	0.18±0.04	0.74	0.13±0.03	0.75	0.13±0.03	0.78	0.14±0.03	0.67
	ISA	0.22±0.04		0.33±0.06		0.20±0.04		0.17±0.04	
STR	*N*	0.13±0.03	0.95	0.13±0.03	0.92	0.09±0.03	0.91	0.04±0.02	0.84
	ISA	0.14±0.03		0.15±0.04		0.13±0.03		0.05±0.02	
WH	*N*	0.31±0.06	0.96	0.30±0.06	0.92	0.29±0.05	0.91	0.35±0.06	0.96
	ISA	0.36±0.06		0.36±0.06		0.36±0.06		0.36±0.06	
ANG	N	0.15±0.04	0.99	0.25±0.05	0.94	0.13±0.03	0.97	0.10±0.03	0.98
	ISA	0.16±0.04		0.35±0.06		0.14±0.04		0.11±0.03	

[1]See Table 1 for variable codification

heterogeneity on family effect estimations, two families with contrasting rank changes are outlined in Fig. 7, and their field locations indicated in Fig. 4b. Family 1023 (black rectangles) was arranged in better microsites than family 2,078 (white rectangles) within almost all blocks. As a result, the family effect estimate for family 1,023 using the standard approach was clearly higher than that for family 2,078, i.e., the BLUPs of families 1,023 and 2,078 were overestimated and underestimated, respectively. After removal of spatial covariance, the BLUP of family 2,078 greatly increased, whereas, that for family 1,023 decreased, resulting in a drastic switch of family ranks.

Comparisons among families were not only enhanced through larger family variances and wider ranges of family effects, but also because the accuracy of the BLUPs of family effects was greatly improved after correction for spatially autocorrelated variation, increasing from 0.40–0.63 to 0.72–0.79 for diameter and from 0.63–0.72 to 0.81–0.88 for height growth. The estimation of the accuracy of BLUPs for height growth through successive steps of the iterative procedure is presented in Fig. 6b.

The noticeable shift of family effects after the first iteration of the iterative procedure justifies the application of successive steps of the proposed method. Important family rank changes were still present after the second step but they became irrelevant after further iterations (Fig. 7). Similar results were observed for growth traits in the four sites, but the convergence of the family effect estimates was achieved earlier for the other traits.

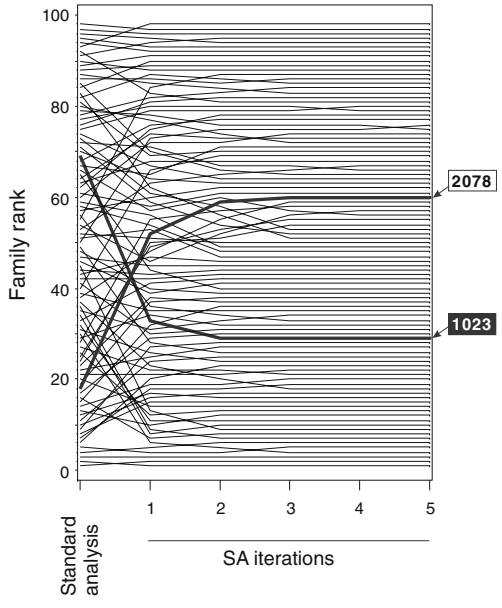

Fig. 7 Family rank changes for height in site C through successive iterations of the spatial analysis (*SA*) procedure. Serious rank changes are evident after the first step of the *SA*, they remained significant in the second step, but they became unimportant in successive iterations, until convergence. Note the drastic rank changes of some families between the standard approach and the last iteration of the SA. Two of these families (2078 and 1023), whose field locations are indicated in Fig. 4b, are outlined

Discussion

The results presented in this paper indicated a strong spatial autocorrelation that may seriously affect the genetic analysis of the data. As forewarned before, the block structure of a RCB design may be not enough to account for this spatial heterogeneity, and the conventional statistical analysis may result in erroneous variance and family effect estimates (Costa-Silva et al. 2001; Dutkowski et al. 2002; Hamann et al. 2002; Joyce et al. 2002; Magnussen 1993a). When blocks are large and spatial dependence is present, some spatial analysis technique is needed. The iterative kriging procedure used in this paper effectively removed the spatial variation, and the block and family-by-block interaction variances disappeared or almost disappeared. Other spatial analysis procedures, such as the nearest neighbor (NN) adjustment, have been used to account for spatial heterogeneity, but several concerns have been noted. The number of neighbors (Joyce et al. 2002) and the weights (Magnussen 1993a) to be used in the NN-covariance adjustments have an important impact on the analysis and are usually arbitrarily defined. The kriging method performs an unbiased estimation of the spatial structure using the complete information within the ratio (range) that corresponds to actual spatial dependence, and uses interpolation weights derived from complex calculations based on the distances between points and the theoretical semivariogram function (Cressie 1993). The kriging method can then be considered an extrapolation of the NN methodology, where the number of neighbors and the weights to be used in the adjustments are optimally derived form the actual spatial structure of the data. Kriging estimates are, thus, more reliable indicators of the amount of the trait that is specifically due to the tree position.

As observed in the present paper, Hamann et al. (2002), analyzing three red alder genetic trials found that a single kriging step, removed block effects, reduced the family-by-block interaction, and increased the family and provenance variance components, heritability increasing accordingly. However, the rank order of family or provenance means did not changed as much as in the *P. pinaster* tests in the present study. A lower intensity of the spatial heterogeneity, as indicated by lower c_o/c_n ratios and relatively lower block variances, can explain these differences. Costa-Silva et al. (2001), including a spatially correlated residual in the statistical analysis, also found an increase in the genetic variance in several FGT and relatively large changes in rank of the genetic entries in relation to the standard analysis. Similar results using the NN method were reported by Joyce et al. (2002) in a black spruce progeny test, and by Anekonda and Libby (1996) in a clonal test of redwood.

However, Magnussen (1993a), using simulated data, found that spatial autocorrelation artificially inflates genetic variance and heritability in multi-unit plot designs. In these cases, spatial autocorrelation may inflate the absolute family variance, as trees in a single plot would be more similar, but this should be clearly compensated by an increase in the family-by-block interaction variance

(generated by the within-block spatial heterogeneity), resulting in a downward bias in the heritability estimation. The small-scale patchy structure in the simulated data of Magnussen (1993a) may have affected family variance in a greater extent than the family-by-block interaction variance, and it would explain why he did not find a downward bias in the heritability estimation. In the present study, the absolute value of the family variance did not consistently change after the spatial analysis, but there was a consistent increase in its relative value (Fig. 5) and, as a consequence, in the heritability estimation (Table 2).

Pre-subtraction of family means could bias the estimation of genetic parameters as environmental covariance can be confounded with covariance among family members, as would be the case if a family occurred in better-than-average microsites in each block (Joyce et al. 2002; Loo-Dinkins 1992). However, this will probably not occur in all the ten blocks of the *P. pinaster* progeny trials. This bias has been considered a main constraint of NN adjustments in analysis of genetic field data but is overcome with the iterative process proposed in this paper. A family that occurs in better-than-average microsites in all blocks will initially generate a downward bias in the kriging estimates around the positions where the family occurs. However, the following estimation of the family effect will be lower and this bias will tend to disappear in further iterations. In fact, the small-scale spatial variation of residuals after the iterative process did not differ for within- and across-row orientations, as indicated by similar anisotropic semivariograms (data not shown).

The need for an iterative procedure was evident given the changes in variance components, accuracy of family BLUPs, and family ranks that remained after the first step. Other convergence criteria, such as the changes in covariance removed or the changes in accuracy of BLUPs, could also be considered to finalize the iterative process. However, the stability of family ranks appeared to be the most restrictive criteria, and the changes in covariance removed or in accuracy of BLUPs were insignificant after the family ranks were stabilized. Similarly, Magnussen (1993a) also found that an iteration of the NN adjustment procedure proved to be more efficient than the non-iterated adjustment.

Other authors have modeled the spatial variation in two steps: first, general gradients in rows and columns are identified by linear regression, and then, detrended residuals are explored for small-scale spatial variation (Fu et al. 1999; Joyce et al. 2002). Although such two-step analyses may be of advantage in the characterization of spatial variation (Fu et al. 1999), the more important aspect in the analysis of RCB field genetic trials is the within-block spatial heterogeneity, i.e., the small-scale spatial variation. The efficiency of block designs is higher in the case of general gradients. Semivariograms should be then constructed up to distances similar to the block sizes. Fu et al. (1999) found that the spherical covariance model fit well to detrended residuals of height in 50 out of 66 Douglas-fir progeny trials, and indicated that the large-scale gradients in other directions, rather than rows and columns, may explain the lack of fit in the remaining sites. If raw residuals (not detrended) were used for semivariogram fitting (as in this case), these large-scale spatial variations could be accounted for by other theoretical functions, such as the exponential function that reaches the asymptote (*sill*) more gradually. Nevertheless, one should try several spatial models and decide which model is the best suited in each case. Several spatial covariance models (Spherical, Exponential, Gaussian, Lineal, Power, nested models, etc.) are available in SAS (SAS Institute, 1999).

Another question is how we should proceed when a trial is divided in several discrete zones; such is the case in Fig. 4. The reliability of the kriging method is smaller near the plot boundaries because there are fewer neighbors to be used for kriging estimations, resulting in higher standard errors of the kriging estimates. Thus, the kriging method will perform worse in divided sites or in sites with irregular shapes (higher proportion of trees near boundaries). In the case of several discrete zones, separate semivariograms could be constructed for each zone. In the *P. pinaster* sites with discrete zones of this paper (sites B and C), the iterative kriging method seemed to be robust, and the results were virtually identical using either separate semivariograms or a unique semivariogram for the whole site (data not shown). Nevertheless, divided sites or sites with irregular shapes should be avoided in future FGT to optimize the effectiveness of spatial analysis techniques.

Growth traits were most affected by spatial heterogeneity, but other traits less influenced by the variation in soil properties, such as straightness and number of whorls, were not completely free of this phenomenon. Besides the soil spatial dependence, which nevertheless may exist (del-Río et al. 2004), subjectively assessed traits may be affected by an autocorrelation caused by the relativity of the assessor (Sierra de Grado et al. 1999) that can be, again, effectively removed by the proposed spatial analysis procedure. As indicated by Hamann et al. (2002) for phenology traits, the reduced changes in variance components and family effect estimation when there is low spatial autocorrelation suggest that there would be little harm in applying the proposed procedure even in these cases.

The results presented in this paper demonstrate how spatial heterogeneity can drastically affect the genetic analysis of FGT with large blocks on hilly and irregular terrains. To leave data unadjusted in the presence of spatial autocorrelation is clearly unacceptable, as already pointed out (Costa-Silva et al. 2001; Dutkowski et al. 2002; Hamann et al. 2002; Magnussen 1993b, 1994). Without removal of spatial effects, erroneous variance estimations can lead to unfortunate decisions when planning the advanced breeding strategy and to incorrect estimations of genetic gains. Family selection can also be strongly affected, as severe changes were observed in family ranks after the spatial adjustment. Individual selections within the progeny trials for further breeding generations would also be mistaken, and potential genetic gains would be lost. The proposed iterative kriging procedure is a promising tool to account for the spatial heterogeneity. Although this method is quite arduous to apply and

requires a long time to complete the analysis, the limitations in computer resources are no longer an obstacle, as was argued some years ago (Joyce et al. 2002; Magnussen 1993a), and all the complete spatial analyses can be carried out using the common SAS-STAT software package (SAS Institute 1999). There are still many RCB trials with large blocks in which this kind of spatial adjustments would be vital. Semivariograms should be routinely inspected to test if residuals of the models are spatially autocorrelated, and, in that case, the iterative kriging method is strongly recommended to remove this spatial heterogeneity.

Acknowledgements This study was financed by the Spanish 'Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria' (Grant INIA-RTA2-109) and the Galician government (Grant PGIDIT02-PXIC50201PN). The author thanks Dr. Luis Sampedro, Dr. Vicente Rozas, and three anonymous referees for valuable comments on earlier versions of this paper.

References

Anekonda TS, Libby WJ (1996) Effectiveness of nearest-neighbor data adjustment in a clonal test of redwood. Silvae Genet 45:46–51

Costa-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

Cressie NAC (1993) Statistics for spatial data. Wiley, New York

Dutilleul P (1993) Spatial heterogeneity and the design of ecological field experiments. Ecology 74:1646–1658

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

Ettema CH, Wardle DA (2002) Spatial soil ecology. Trends Ecol Evol 17:177–182

Fu YB, Clarke GPY, Namkoong G, Yanchuk AD (1998) Incomplete block designs for genetic testing: statistical efficiencies of estimating family means. Can J For Res 28:977–986

Fu YB, Yanchuk AD, Namkoong G (1999) Spatial patterns of tree height variations in a series of Douglas-fir progeny trials: implications for genetic testing. Can J For Res 29:714–723

Gallardo A (2003) Spatial variability of soil properties in a floodplain forest in Northwest Spain. Ecosystems 6:564–576

Gallardo A, Covelo F (2005) Spatial pattern and scale of leaf N and P concentration in a *Quercus robur* population. Plant Soil 273:269–277

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

Grondona MO, Crossa J, Fox PN, Pfeiffer WH (1996) Analysis of variety trials using two-dimensional separable ARIMA processes. Biometrics 52:763–770

Hamann A, Namkoong G, Koshy MP (2002) Improving precision of breeding values by removing spatially autocorrelated variation in forestry field experiments. Silvae Genet 51:210–215

Joyce DG, Ford R, Fu YB (2002) Spatial patterns of tree height variations in a black spruce farm-field progeny test and neighbors-adjusted estimations of genetic parameters. Silvae Genet 51:13–18

Legendre P (1993) Spatial autocorrelation: a trouble or new paradigm? Ecology 74:1659–1673

Loo-Dinkins J (1992) Field test design. In: Fins L, Friedman ST, Brotschol JV (eds) Handbook of quantitative genetics. Kluwer, Dordrecht, The Netherlands, pp 96–139

López MV, Arrúe JL (1995) Efficiency of an incomplete block design based on geostatistics for tillage experiments. Soil Sci Soc Am J 59:1104–1111

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

Magnussen S (1993a) Bias in genetic variance estimates due to spatial autocorrelation. Theor Appl Genet 86:349–355

Magnussen S (1993b) Design and analysis of tree genetic trials. Can J For Res 23:1144–1149

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

Merino A, Rodríguez-López A, Brañas J, Rodríguez-Soalleiro R (2003) Nutrition and growth in newly established plantations of *Eucalyptus globulus* in northwestern Spain. Ann For Sci 60:509–517

Qiao CG, Basford KE, DeLacy IH, Cooper M (2000) Evaluation of experimental designs and spatial analyses in wheat breeding trials. Theor Appl Genet 100:9–16

Qiao CG, Basford KE, DeLacy IH, Cooper M (2004) Advantage of single-trial models for response to selection in wheat breeding multi-environment trials. Theor Appl Genet 108:1256–1264

del-Río M, Bravo F, Pando V, Sanz G, Sierra de Grado R (2004) Influence of individual tree and stand attributes in stem straightness in *Pinus pinaster* Ait. stands. Ann For Sci 61:141–148

Sánchez-Rodríguez F, Rodríguez-Soalleiro R, Español E, López CA, Merino A (2002) Influence of edaphic factors and tree nutritive status on the productivity of *Pinus radiata* D. Don plantations in northwest Spain. For Ecol Manag 171:181–189

SAS Institute (1999) SAS/STAT User's guide, Version 8. SAS Institute, Cary, North Carolina

Sierra de Grado R, Diez-Barra R, Alia Miranda R (1999) Evaluación de la rectitud del fuste en seis procedencias de *Pinus pinaster* Ait. Investig Agrar Sist Recur For 8:264–278

Whitaker D, Williams ER, John JA (1999) CycDesign. A package for the computer generation of experimental designs. CSIRO forestry and forest products, Camberra, Australia

Wright JW (1976) Introduction to forest genetics. Academic, New York

Zas R, Merlo E, Fernández-López J (2004a) Genetic parameter estimates for Maritime pine in the Atlantic coast of North-west Spain. For Genet 11:45–53

Zas R, Merlo E, Fernández-López J (2004b) Genotype × environment interaction in maritime pine families in Galicia, Northwest Spain. Silvae Genet 53:175–182

Zas R, Serrada R (2003) Foliar nutrient status and nutritional relationships of young *Pinus radiata* D. Don plantations in north-west Spain. For Ecol Manag 174:167–176