

# The impact of spatial heterogeneity on selection: a case study on *Pinus pinaster* breeding seedling orchards

Rafael Zas

**Abstract:** Although failure to account for spatial autocorrelation has been dramatic in some forest progeny trials, little attention has been paid to how this issue may affect selections within the trials. The effects of spatial autocorrelation of height growth on the estimation of genetic gain and on the spatial distribution of the selected trees were studied in four *Pinus pinaster* Ait. progeny trials that were rogued using different selection methods and intensities. When selections are based on unadjusted original values, selected trees tend to be located in the best microsites and are unlikely to be the most genetically superior. This resulted in a loss of genetic gain that varied between 10% and 20% and sometimes exceeded 30%. Differences in the loss of gain among different selection methods and intensities were minor and followed no clear pattern. Selecting on the basis of a conventional model resulted in spatial patterns of the retained trees that were clearly aggregated in all cases. However, selections based on spatially adjusted data resulted in random spatial patterns, except with family selection because of the use of multiple-tree plots. Because clumping of the retained trees may seriously affect the quantity and quality of the seed crop, breeders are strongly encouraged to use appropriate spatial models for roguing breeding seedling orchards.

**Résumé :** Même si le fait de ne pas tenir compte de l'autocorrélation spatiale dans certains tests de descendance en milieu forestier est préoccupant, l'impact potentiel de cette situation sur les sélections effectuées dans ces tests a reçu peu de considération. L'auteur a étudié les effets de l'autocorrélation spatiale de la croissance en hauteur sur les estimations de gain génétique et sur la distribution spatiale des arbres sélectionnés dans quatre tests de descendance de *Pinus pinaster* Ait. qui ont été éclaircis selon différentes méthodes et intensités de sélection. Lorsque les sélections sont effectuées à partir des valeurs originales non ajustées, les arbres sélectionnés ont tendance à être localisés sur les meilleurs microsites. Dans ces conditions, il est donc peu probable que ces arbres soient génétiquement supérieurs. Dans ce cas, des pertes de gain génétique de l'ordre de 10 % à 20 % et parfois en excès de 30 % ont été observées. Les différences entre les pertes de gain dues aux différentes méthodes et intensités de sélection étaient mineures et ne suivaient pas de patron précis. La sélection sur la base d'un modèle traditionnel s'est traduite par une configuration spatiale des arbres sélectionnés qui était clairement agrégée dans tous les cas. Cependant, les sélections basées sur des données qui tiennent compte de l'autocorrélation spatiale ont produit des configurations spatiales aléatoires, à l'exception de la sélection familiale en raison de la présence de parcelles contenant plusieurs arbres. Étant donné que l'agrégation des arbres sélectionnés peut sérieusement affecter la quantité et la qualité des graines, on encourage fortement les améliorateurs à utiliser des modèles spatiaux appropriés lors de l'éclaircie des vergers à graines de semis.

[Traduit par la Rédaction]

## Introduction

Most tree breeding is based on a strategy of recurrent selection for general combining ability in which a broadly based breeding population is maintained through open- or control-pollinated family trials. Future breeding population parents and limited individuals with highest breeding values are selected within these trials for the production population, represented by clonal seed orchards (Burdon and Shelbourne 1971). Genetic parameters, i.e., genetic and environmental variances and covariances, heritabilities, and genetic correla-

tions, are commonly estimated from data collected in progeny trials, as are parental and offspring breeding values. Good estimates of breeding values from the progeny trials are crucial for realizing adequate gains in both breeding populations and seed orchards.

Testing, breeding, and seed production can be combined in a single plantation, called a breeding seedling orchard (BSO) by Barnes (1995). BSOs represent a simple, low-input breeding strategy, where both breeding and the supply of genetically improved seed can be accomplished simultaneously. They are frequently used for secondary tree species and (or) for species where clonal seed orchards are not feasible because of difficulties in vegetative propagation (David et al. 2003; Fedorkov et al. 2005; Hansen et al. 2005; Kang et al. 2005). Typically, BSOs are half-sib plantings that act first as progeny tests and breeding populations for forward selection and, then after genetic thinning, as seed orchards (Barnes 1995). To effectively utilize a BSO for seed production, quite intense roguing is needed to produce seed with reasonable genetic gain. Dif-

Received 12 December 2006. Accepted 20 May 2007. Published on the NRC Research Press Web site at cjfr.nrc.ca on 24 January 2008.

R. Zas,<sup>1</sup> Centro de Investigaciones Ambientales de Lourizán, Apdo. 127, Pontevedra, 36080 Spain (e-mail: rzas.cifal@siam-cma.org).

<sup>1</sup>Present address: Ensis-Genetics, Private bag 3020, Rotorua, New Zealand.

ferent selection methods can be used for roguing BSOs (e.g., individual selection, family selection, combined selection, and group-merit selection), and breeders typically seek the method and selection intensity that optimize the combination of genetic gain and the genetic diversity (GD) of the resulting seed crop after thinning (Olsson et al. 2001; David et al. 2003; Fedorkov et al. 2005).

Recent studies have demonstrated that microsite variation within progeny tests of forest trees may seriously affect the estimation of genetic parameters and breeding values (e.g., Dutkowski et al. 2006; Zas 2006). Forest-tree progeny trials are usually very large because of numerous genetic entries and the large space needed for individual plants. Moreover, they are usually established on irregular forest terrains, resulting in a strong spatial heterogeneity that experimental designs cannot efficiently accommodate (Fu et al. 1999; Gezan et al. 2006). Spatial heterogeneity is commonly reflected in a significant spatial autocorrelation of trait values, which implies a violation of the requirement of data independence in standard parametric statistics (Legendre 1993). The assumption of independence is also needed to fit mixed models using the restricted maximum likelihood (REML) method, although several statistical packages allow incorporation of different structures of correlated errors (Littell et al. 1996). Several authors have shown that conventional analysis of spatially correlated data in inadequately blocked progeny trials may be completely erroneous and concluded that spatial analysis becomes essential in these cases (e.g., Saenz-Romero et al. 2001; Hamann et al. 2002; Dutkowski et al. 2006; Zas 2006). These works have focused mainly on the effects of spatial autocorrelation on the estimates of genetic parameters and breeding values and on the accuracy of the estimates. However, spatial autocorrelation may also alter selections in the progeny trials and, thus, the estimation and capture of genetic gain (Costa e Silva et al. 2001; Dutkowski et al. 2002, 2006). These effects on genetic gain estimates may be critical for deciding the optimal selection method for roguing the BSO. Considering that spatial autocorrelation is the norm rather than the exception in progeny trials (Fu et al. 1999; Dutkowski et al. 2006), knowledge of the extent of these effects becomes very important.

Microsite variation and resulting spatial autocorrelation will also determine the spatial distribution of the remaining trees after thinning BSOs. Greater spatial autocorrelation leads to stronger spatial clumping of the selected trees, because trees located on favorable microsites would have higher probability of being selected. Seed orchard managers have played little attention to this issue, even though it can have important consequences for the genetic merit, GD, and the seed yield of thinned BSOs. For a given number of genetic entries, GD is maximized when all trees in the BSO intermate randomly and in equal proportions (called panmixia). Because the probability of mating increases with proximity (Burczyk and Prat 1997), the GD of the seed crop may decrease as the spatial pattern of the selected trees in the BSO becomes clumped. Moreover, clumping implies a poor exploitation of the available resources and, thus, a loss of potential seed yield.

Within the framework of the maritime pine (*Pinus pinaster* Ait.) breeding program in Galicia (northwestern Spain), the supply of genetically improved seeds has re-

**Table 1.** Details of the progeny trials.

|                      | Site  |       |       |       |
|----------------------|-------|-------|-------|-------|
|                      | A     | B     | C     | D     |
| <b>Situation</b>     |       |       |       |       |
| Longitude (°W)       | 8.41  | 8.74  | 8.11  | 8.00  |
| Latitude (°N)        | 42.14 | 42.63 | 42.21 | 42.65 |
| Altitude (m)         | 525   | 300   | 530   | 700   |
| <b>Design</b>        |       |       |       |       |
| No. of families      | 87    | 78    | 98    | 94    |
| No. of trees         | 3708  | 3050  | 3848  | 4076  |
| Area (ha)            | 4.07  | 3.67  | 4.59  | 4.63  |
| <b>Height growth</b> |       |       |       |       |
| Height mean (cm)     | 513.3 | 490.9 | 557.8 | 440.5 |
| CV height (%)        | 19.8  | 23.2  | 20.8  | 17.9  |

cently decreased dramatically because of the infection of the clonal seed orchards by several fungal pathogens. Therefore, there is an urgent need to increase the seed-production population, and BSOs appear to be an attractive alternative because of their immediate availability, low cost, and simplicity. *Pinus pinaster* progeny trials in Galicia were established following randomized complete block (RCB) designs with large blocks (Zas et al. 2004). Several traits evaluated in these trials at age 8 years showed strong spatial autocorrelation that the experimental layout could not account for efficiently (Zas 2006). Conventional analysis of these traits resulted in underestimated heritabilities and erroneous parental breeding-value estimates. After appropriate spatial adjustments, dramatic family rank changes were observed, and accuracy of family effect estimations was greatly improved (Zas 2006). The present study focuses on how spatial autocorrelation may affect the decisions about roguing these progeny trials to transform them in BSOs. Specifically, the objectives of the present paper are (i) to analyze the effect of spatial autocorrelation on the genetic gain estimates for height growth in four *P. pinaster* progeny trials under different selection scenarios and (ii) to analyze how the spatial structure of the data and the selection method influences the spatial distribution of the selected trees, i.e., the spatial pattern of the trees in the thinned BSO.

## Materials and methods

### Experimental data

The study was conducted in four open-pollinated *P. pinaster* progeny trials located in Galicia (northwestern Spain) (Table 1). The trials, established in 1994–1995, consisted of between 79 and 98 open-pollinated families obtained from a total of 116 plus trees selected within the Atlantic area of Galicia. All sites followed a RCB design with 10 blocks, five-tree row plots, and 3 m × 3 m spacing. Further details of the progeny trials, the genetic material, and the genetic parameter estimates for growth, stem form, and branching traits are given in Zas et al. (2004). Height growth at age 8 years from planting, which showed strong spatial autocorrelation in all sites (Zas 2006), was the trait considered for this study. Positions of all trees were spatially referenced using orthographic aerial photographs at ~1:1000 scale.

### Statistical analysis

Conventional analyses for the RCB design within each site were carried out using the following random model

$$[1] \quad y_{ijk} = \mu + F_i + B_j + FB_{ij} + \varepsilon_{ijk}$$

where  $y_{ijk}$  is the measured height of the  $k$ th tree of the family  $i$  in the block  $j$ ;  $\mu$  is the overall mean;  $F_i$ ,  $B_j$ , and  $FB_{ij}$  are the random effects of family  $i$ , block  $j$ , and the corresponding interaction (in practice, probably a plot  $\times$  environment effect); and  $\varepsilon_{ijk}$  is the experimental random error.

Variance components and best linear unbiased predictors (BLUPs) of family effects were estimated using the REML method of the MIXED procedure in SAS (SAS Institute Inc. 1999). Unless the covariance structure of residuals is specifically declared, residuals of these mixed models must comply with three main assumptions, namely, normal distribution, constant variance across groups, and mutual independence. The first two assumptions are typically fulfilled when analyzing a continuous variable, such as tree height, in progeny trials; in the case of the *P. pinaster* tests analyzed here, a graphical inspection of the residuals revealed nonsignificant departures from these assumptions. However, the independence assumption would be violated if the data are spatially autocorrelated.

Therefore, the spatial autocorrelation of the residuals was checked by constructing a semivariogram, which plots the semivariance among trees as a function of the distance separating them (Rossi et al. 1992). A flat semivariogram will indicate random spatial structures, whereas the semivariance will be small at short distances in the presence of spatial dependence and will increase as distance between trees increases. Semivariograms were computed using the VARIOGRAM procedure of SAS (SAS Institute Inc. 1999).

### Spatial analysis

Data were adjusted for spatial autocorrelation using the iterative spatial analysis (ISA) procedure proposed by Zas (2006). Briefly, the method consists in subtracting from each observation the amount of the effect that is attributable to the spatial covariance of the data. After subtracting preliminary estimates of family effects, the spatial structure of the residuals was modeled using a semivariogram. An exponential theoretical semivariogram was fitted to the observed semivariogram and the resulting sill, nugget, and range parameter estimates were used to divide the variation of residuals into spatially autocorrelated variation and random error with the kriging method (Rossi et al. 1992). At each point, the kriging estimate can be interpreted as the amount of the trait effect that is due to the spatial position. Thus, the kriging values at each tree location were used to correct the original observations. The adjusted values were reanalyzed and a new estimation of the family effects was obtained and used to generate new residuals from the original data. The process was repeated iteratively until convergence of the BLUP estimates of family effects. All analysis were carried out in SAS (SAS Institute Inc. 1999). Variance components and BLUPs of family effects for adjusted values were estimated using the REML method of the MIXED procedure, whereas semivariograms and kriging were computed using the VARIOGRAM and KRIGE2D SAS procedures (SAS In-

stitute Inc. 1999). Further details of the method are given in Zas (2006).

### Genetic parameters

Individual ( $\hat{h}_i^2$ ) and family ( $\hat{h}_f^2$ ) heritabilities for both original and adjusted values were estimated as

$$[2] \quad \hat{h}_i^2 = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_f^2 + \hat{\sigma}_{fb}^2 + \hat{\sigma}_e^2}$$

$$[3] \quad \hat{h}_f^2 = \frac{\hat{\sigma}_f^2}{\hat{\sigma}_f^2 + (\hat{\sigma}_{fb}^2/B) + (\hat{\sigma}_e^2/BN)}$$

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

Within-family heritability ( $\hat{h}_w^2$ ) was estimated as (Falconer and Mackay 2001)

$$[4] \quad \hat{h}_w^2 = \frac{1-r}{1-t} \hat{h}_i^2$$

where  $r$  is the genetic correlation among half-sibs ( $r = 0.25$ ),  $t$  is the intraclass phenotypic correlation for half-sibs ( $t = \sigma_f^2/(\sigma_f^2 + \sigma_{fb}^2 + \sigma_e^2) = 0.25\hat{h}_i^2$ ), and  $\hat{h}_i^2$  is the individual heritability estimate.

### Selection scenarios and genetic gain

Five different selection methods were considered: individual (IS), family (FS), within-family (WFS), combined (CS), and family + within-family selection (F + WFS). Estimates of genetic gains for each method and for each selection ratio (from 1% to 40% of the total trees selected) were calculated as follows: (i)  $\Delta\hat{G}(\%) = \hat{S}_i \hat{h}_i^2 / \mu$  for IS; (ii)  $\Delta\hat{G}(\%) = \hat{S}_f \hat{h}_f^2 / \mu$  for FS; (iii)  $\Delta\hat{G}(\%) = \hat{S}_w \hat{h}_w^2 / \mu$  for WFS; and (iv)  $\Delta\hat{G}(\%) = (\hat{S}_f \hat{h}_f^2 + \hat{S}_w \hat{h}_w^2) / \mu$  for CS and F + WFS, where  $\mu$  is the overall mean;  $\hat{S}_i$ ,  $\hat{S}_f$ , and  $\hat{S}_w$  are the estimated selection differentials for individual values, family effects, and within-family deviations for the selected trees, respectively; and  $\hat{h}_i^2$ ,  $\hat{h}_f^2$ , and  $\hat{h}_w^2$  are the individual-, family-, and within-family heritability estimates, respectively. Values were previously adjusted for differences in block effects by subtracting the BLUPs of block effects (David et al. 2003). For IS, those trees with the lowest phenotypic values are thinned out. For FS, trees of those families with the best family-effect estimates are selected, whereas WFS involves selecting those trees with the highest within-family deviations. For CS, trees were selected based on the index  $H_w \hat{h}_w^2 + H_f \hat{h}_f^2$ , where  $H_w$  is the within-family deviation (observed individual phenotypic value – family mean) and  $H_f$  is the family effect (overall mean + family BLUP). For F + WFS, the family size was arbitrarily set at 15 individuals, which represents a WFS rate of around 37% (15 of 40 individuals per family on average).

For each selection scenario, three estimates of genetic gains were calculated according to which data were consid-

**Table 2.** Variance component estimates using the conventional model (N) and the iterative spatial analysis (ISA) in four progeny trials, and spatial structure of corresponding residuals.

| Site | Model | Variance components <sup>a</sup> |              |                 |              | Residual spatial structure <sup>b</sup> |           |                   |
|------|-------|----------------------------------|--------------|-----------------|--------------|---|-----------|-------------------|
|      |       | $\sigma_f^2$                     | $\sigma_b^2$ | $\sigma_{fb}^2$ | $\sigma_e^2$ | $R^2$                                   | $a_o$ (m) | $c_o/(c_o + c_n)$ |
| A    | N     | 380.5***                         | 2140.2*      | 1963.4***       | 6136.7***    | 0.994                                   | 20.5      | 0.34              |
|      | ISA   | 271.0***                         | 0.0          | 0.0             | 4668.7***    | Flat semivariogram                      |           |                   |
| B    | N     | 305.9**                          | 3960.6*      | 2859.0***       | 6297.9***    | 0.997                                   | 23.8      | 0.54              |
|      | ISA   | 363.1***                         | 0.0          | 0.0             | 4046.2***    | Flat semivariogram                      |           |                   |
| C    | N     | 247.6**                          | 6247.9*      | 1661.9***       | 5960.6***    | 0.996                                   | 80.7      | 0.50              |
|      | ISA   | 276.9***                         | 0.0          | 15.4            | 5314.5***    | Flat semivariogram                      |           |                   |
| D    | N     | 172.3**                          | 1310.4*      | 1646.7***       | 3243.1***    | 0.992                                   | 29.9      | 0.49              |
|      | ISA   | 104.7***                         | 0.0          | 0.0             | 2299.4***    | Flat semivariogram                      |           |                   |

<sup>a</sup> $\sigma_f^2$ ,  $\sigma_b^2$ ,  $\sigma_{fb}^2$ , and  $\sigma_e^2$  are the family, block, family  $\times$  block interaction, and error variance estimates, respectively.

<sup>b</sup> $R^2$ ,  $a_o$ ,  $c_o$ , and  $c_n$  are the coefficient of determination, the range, the patch variance, and the nugget resulting of fitting the exponential model to the observed semivariogram, respectively. The ratio  $c_o/(c_o + c_n)$  is a measure of the intensity of the spatial structure.

ered: original values (two possibilities) or values adjusted for spatial autocorrelation. When selections are based on original values, selected trees are not necessarily the genetically most superior individuals but tend to be those trees located in the best microsite environments. Although data were adjusted for differences in block effects, blocks were clearly heterogeneous internally (Zas 2006), so trees located in favored microsites within blocks had higher probability of being selected. Two estimates of genetic gain can be computed when selections are based on original unadjusted values: (i) the expected genetic gain without considering the spatial heterogeneity,  $\Delta\hat{G}_{EXP} = \hat{S}^2 \hat{h}^2$ , where  $\hat{S}$  is the selection differential for original values (adjusted for block effects) and  $\hat{h}^2$  is the heritability estimate based on conventional analysis, and (ii) the adjusted genetic gain,  $\Delta\hat{G}_{ADJ} = \hat{S}' \hat{h}'^2$ , where the same trees selected from the unadjusted heights are again selected, but the selection differential,  $\hat{S}'$ , and heritability,  $\hat{h}'^2$ , applied are based on spatially adjusted data. Thus,  $\Delta\hat{G}_{EXP}$  is the predicted genetic gain ignoring the spatial autocorrelation of the data, whereas  $\Delta\hat{G}_{ADJ}$  is the adjusted genetic gain reached (after corrections of heritability and selection differential for spatial autocorrelation) when selections are still based on unadjusted original values. A third estimate of genetic gain ( $\Delta\hat{G}_{SPA}$ ) is obtained when both selection and prediction of genetic gain are based on spatially adjusted values.

### Genetic diversity

The GD of the retained trees was calculated as

$$[5] \quad GD = 1 - \Theta$$

where  $\Theta$  is the group coancestry of the population, which describes the probability that two genes are identical by descent

$$[6] \quad \Theta = \sum_{i=1}^k \frac{n_i}{n} \left[ \frac{0.5 + 0.5r(n_i - 1)}{n} \right]$$

where  $k$  is the total number of families selected,  $n_i$  is the number of individuals in family  $i$ , and  $n$  is the total number of trees retained (David et al. 2003).

### Spatial distribution of selected trees

The spatial pattern of the selected trees under the different selection scenarios was analyzed using spatial analysis by distance indices (SADIE) software for mapped data (Perry 1995; the program is available from <http://www.rothamsted.ac.uk/pie/sadie>). This method uses an interesting algorithm based on a biological model to evaluate the distance to regularity ( $D$ ), which is the minimum distance needed to move the analyzed points to a regular arrangement similar to a hexagonal lattice. Details of this algorithm are given in Perry (1995). Usually, the larger the  $D$ , the more spatially aggregated are the data. To quantify the magnitude of  $D$ , SADIE compares it with the  $D$  estimated from spatially random distributions of the same number of points within the sample area. The ratio between the observed value of  $D$  and the mean  $D$  from a high number of randomizations (in our case 200) provides the index of aggregation  $I_p$ . Values of  $I_p$  greater than unity indicate aggregation, values close to unity indicate random spatial distribution, and values below unity indicate spatial regularity. Furthermore, the method provides a formal test for departures from randomness by calculating the proportion,  $P_p$ , of randomizations that generates values of  $D$  greater than the observed  $D$  for the actual data. The null hypothesis of spatial randomness may be rejected at the 5% level in favor of the alternative hypothesis of aggregation if  $P_p < 0.05$  or in favor of spatial regularity if  $P_p > 0.95$ .

Because this method is sensitive to irregular shapes of the study area (Rozas and Camarero 2005), as with the progeny trials studied here, we applied the method to the greatest possible rectangle circumscribed into each progeny test.

### Results

The results of the analysis of unadjusted and spatially adjusted data and the results of the corresponding spatial structure of the residuals (Table 2) show clearly both the need for spatial adjustment and its effects on the variance component estimates. The semivariogram of the residuals of the conventional model fitted the exponential model well in all cases ( $R^2 > 0.99$  and  $p < 0.001$ , Table 2) indicating that residuals were spatially autocorrelated. Thus, the independence assumption was clearly violated. The spatial structure

**Table 3.** Individual ( $h_i^2$ ), family ( $h_f^2$ ) and within family ( $h_w^2$ ) heritability estimates ( $\pm$ SE) for height growth at age 8 years in four progeny trials using standard analysis and iterative spatial analysis (ISA).

| Site | Standard analysis |                 |                 | Iterative spatial analysis |                 |                 |
|------|-------------------|-----------------|-----------------|----------------------------|-----------------|-----------------|
|      | $h_i^2$           | $h_f^2$         | $h_w^2$         | $h_i^2$                    | $h_f^2$         | $h_w^2$         |
| A    | 0.18 $\pm$ 0.04   | 0.52 $\pm$ 0.06 | 0.14 $\pm$ 0.03 | 0.22 $\pm$ 0.04            | 0.70 $\pm$ 0.07 | 0.17 $\pm$ 0.04 |
| B    | 0.13 $\pm$ 0.03   | 0.41 $\pm$ 0.06 | 0.10 $\pm$ 0.03 | 0.33 $\pm$ 0.06            | 0.78 $\pm$ 0.10 | 0.27 $\pm$ 0.05 |
| C    | 0.13 $\pm$ 0.03   | 0.44 $\pm$ 0.05 | 0.10 $\pm$ 0.03 | 0.20 $\pm$ 0.04            | 0.68 $\pm$ 0.06 | 0.16 $\pm$ 0.03 |
| D    | 0.14 $\pm$ 0.03   | 0.41 $\pm$ 0.05 | 0.11 $\pm$ 0.03 | 0.17 $\pm$ 0.04            | 0.65 $\pm$ 0.06 | 0.14 $\pm$ 0.03 |

explained between 34% and 54% of total variation, and the spatial dependence was evident at distances of 20–80 m. However, after spatial adjustments, the residuals of the model showed a flat semivariogram in all cases, indicating random spatial variation and, thus, residual independence. Although among-block variance were significantly greater than zero in all cases, blocks were not able to absorb all the spatial variation in the conventional model. After spatial adjustments, block and plot variances disappeared, and residual variation was strongly reduced (Table 2). As a consequence, heritability estimates were always higher for spatially adjusted data than for original data, but the increase in the estimates varied greatly among the test sites (Table 3). Site B was the one with the greatest increase in heritability estimates, whereas heritabilities in site A just slightly increased after spatial adjustment. The effect of spatial autocorrelation on estimates of variance components, heritability, and family effects has been extensively treated elsewhere (Zas 2006) and will be not further considered here.

Figure 1 represents the expected ( $\Delta\hat{G}_{EXP}$ ) and adjusted ( $\Delta\hat{G}_{ADJ}$ ) genetic gain estimates for selections based on original data in relation to the genetic gain that would be expected, at the same selection rate, for selections based on spatially adjusted data ( $\Delta\hat{G}_{SPA}$ ). When the spatial autocorrelation of the data is ignored,  $\Delta\hat{G}_{EXP}$  varied between 40% and 120% of the potential genetic gain  $\Delta\hat{G}_{SPA}$ . For all sites and selection methods, except for IS and WFS in sites A and D,  $\Delta\hat{G}_{EXP}$  was lower than  $\Delta\hat{G}_{SPA}$ . Site B showed the lowest values of  $\Delta\hat{G}_{EXP}$  in relation to  $\Delta\hat{G}_{SPA}$ , whereas both values were much more similar in site A. On the other hand, adjusted genetic gain by means of standard analysis ( $\Delta\hat{G}_{ADJ}$ ) was always lower than  $\Delta\hat{G}_{SPA}$  and usually was between 80% and 90% of  $\Delta\hat{G}_{SPA}$ , which represents a loss of genetic gain of 10%–20% in relation to the potential genetic gain that could be expected if selections and calculations were based on spatially adjusted data. Maximum loss of genetic gain was 33% for F + WFS in site C at a selection rate of 1%. Differences in the loss of genetic gain among the different selection methods did not follow a clear pattern, although it was slightly lower for FS and F + WFS selections. The loss of genetic gain was relatively constant at different selection rates, especially for IS, WFS, and CS.

On the other hand, the spatial adjustments had little influence on the number of families retained and on the GD estimates for the selected trees (Fig. 2). For all selection methods, except for CS, the effect on GD was null or unappreciable. In CS and at the highest selection rates, GD was slightly lower for selections based on the spatial models.

Moreover, differences in the loss of GD among the different selection methods were only relevant at high selection rates, where GD was substantially reduced in the case of FS and F + WFS selection. For stand densities appropriate for seed production in BSOs, differences in GD among the different selection methods were very low.

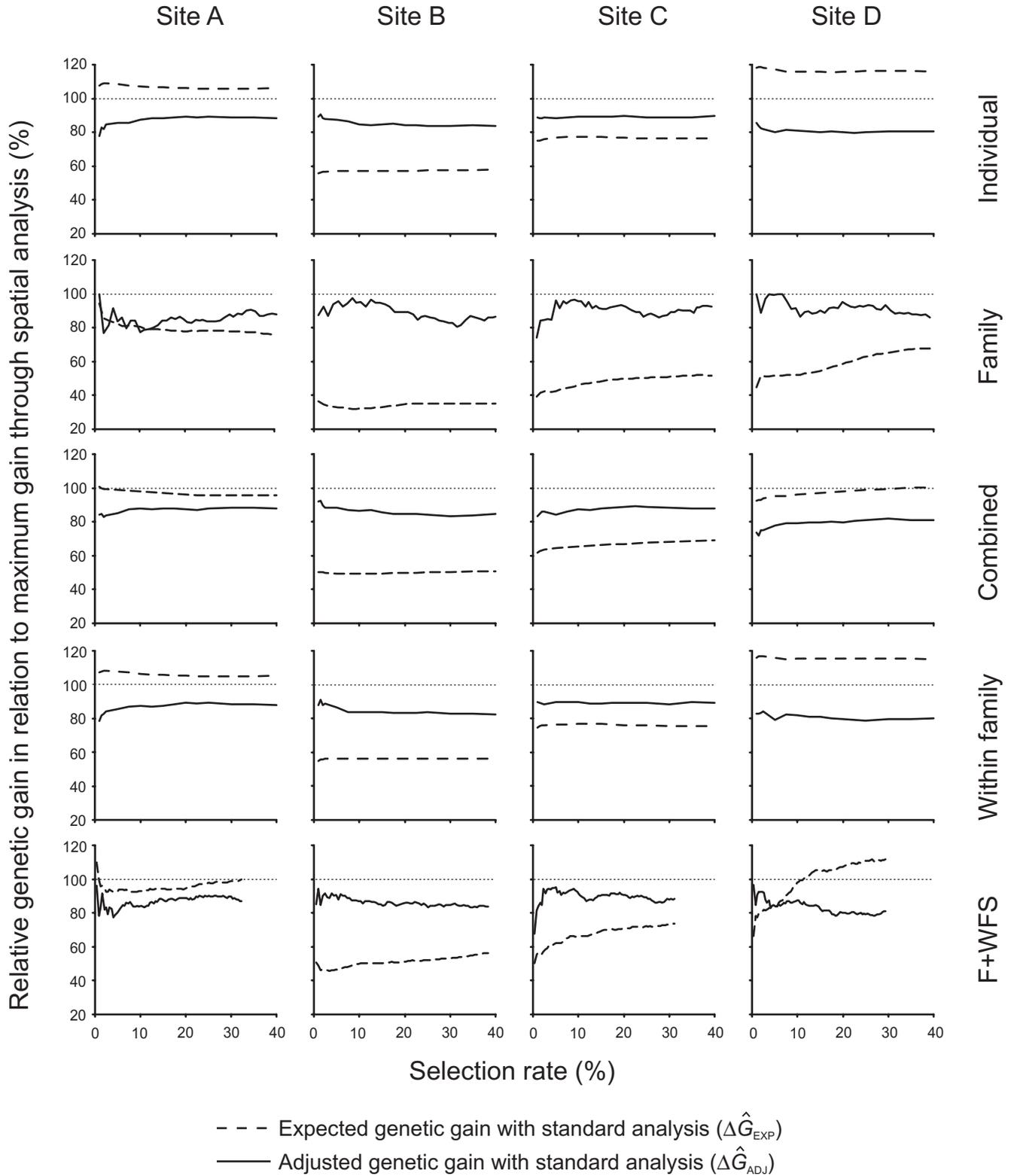
The spatial pattern of the trees selected from original unadjusted values was clearly aggregated for all sites, selection methods, and selection intensities. For example, Fig. 3 shows the estimates of the SADIE clumping indices  $I_p$  for IS, FS, and CS in site C. Despite the moderately low values of  $I_p$  at the very high culling rates for IS and CS, the values of  $I_p$  for selections based on original values were always greater than unity, and the associated probabilities  $P_p$  were always lower than 0.05 indicating that the corresponding spatial patterns were significantly ( $p < 0.05$ ) aggregated. This can also be observed in Table 4 for a selection intensity of 30%. When selections are carried out based on original unadjusted data, the spatial distribution of the selected trees was always significantly aggregated ( $I_p > 1$  and  $P_p < 0.05$ ) for all sites and selection methods. However, when selection is based on spatially adjusted data, the spatial distribution of the selected trees always followed an essentially random (or regular) pattern ( $I_p$  close to 1 and  $P_p > 0.05$ ), except with FS for which the spatial pattern of the selected trees was always significantly aggregated (Table 4, Fig. 3). In FS, all trees of the selected families are included in the selection. Because the experimental units consisted in five-tree row plots, selected trees will be an ensemble of five-tree rows, which necessarily incurs clumping. Apart from FS, CS consistently showed the largest values of  $I_p$  for selections based on spatially adjusted values, although the corresponding spatial pattern was always significantly random.

The spatial distribution of the selected trees for a selection rate of 30% is shown in Fig. 4 for IS and FS in site D. With IS, the spatial pattern of the selected trees is clearly different when selections are based on original versus spatially adjusted data. With original values, the selected trees were clearly aggregated around the best within-block microsites, whereas the spatial pattern became random when selections are based on spatially adjusted data. However, with FS, the spatial pattern of the selected trees was clearly aggregated in both cases.

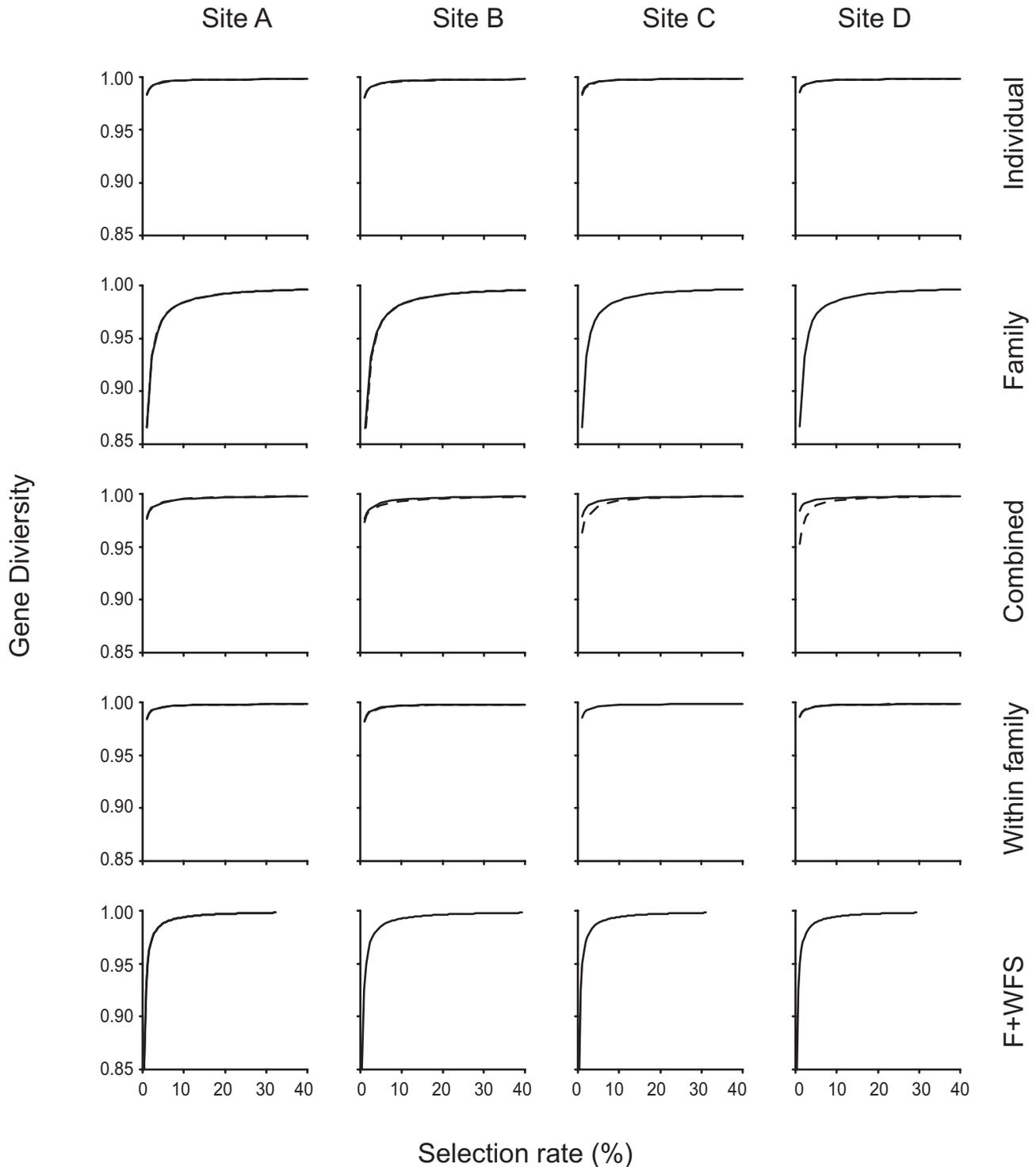
## Discussion

When the spatial autocorrelation of the data is ignored, selection differential is inflated because part of the superiority of the selected trees actually originates from their posi-

**Fig. 1.** Relative expected genetic gain ( $\Delta\hat{G}_{EXP}$ , broken line, %) and adjusted genetic gain ( $\Delta\hat{G}_{ADJ}$ , solid line, %) for selections based on original unadjusted data, as a ratio of gain from selection on spatially adjusted data ( $\Delta\hat{G}_{SPA}$ ). Results are presented for different sites, selection methods, and intensities. Expected genetic gain is calculated using the selection differential and heritability estimate derived from a conventional statistical model. Adjusted genetic gain for selections based on unadjusted data is calculated using the selection differential adjusted for spatial variation and the heritability estimate calculated using the spatial model.



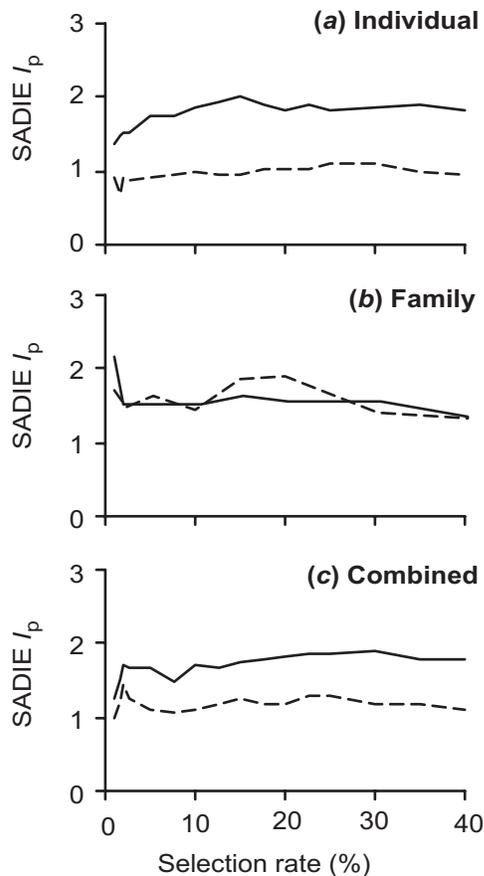
**Fig. 2.** Gene diversity of the retained trees in four progeny trials thinned using different selection methods and intensities for the conventional statistical model (solid line) and iterative spatial analysis (broken line).



tion in growth-favoring microsites. Thus, the more precise selection differential, calculated on the basis of spatially adjusted data, will be always lower than that calculated for original unadjusted values. On the other hand, heritability is usually underestimated when based on unadjusted data (Costa e Silva et al. 2001; Hamann et al. 2002; Zas 2006). Thus, the effect on the estimation of genetic gain would de-

pend on the relative magnitude of these over- and underestimates. This would explain the big differences between  $\Delta\hat{G}_{EXP}$  and  $\Delta\hat{G}_{SPA}$  for different sites and selection methods (Fig. 1). The  $\Delta\hat{G}_{EXP}$  values for IS and WFS in sites A and D were higher than  $\Delta\hat{G}_{SPA}$ , because the overestimation of the selection differential was higher than the underestimation of individual and within-family heritability (just

**Fig. 3.** Spatial analysis by distance indices (SADIE) of aggregation ( $I_p$ ) for the spatial distribution of the remaining trees at site C after selecting the best trees using different selection methods and intensities. Results are presented for (a) individual, (b) family, and (c) combined selections based on a conventional statistical model (solid line) or on iterative spatial analysis (ISA) (broken line).



~20% in contrast with ~40%–60% for sites B and C; Table 3). By contrast, the large underestimation of all heritabilities in site B using the conventional analysis (Table 3) explains the high differences there between  $\Delta\hat{G}_{EXP}$  and  $\Delta\hat{G}_{SPA}$  for all selection methods (Fig. 1).

The real loss of genetic gain when selecting on unadjusted data, compared with the potential gain that would be achieved using spatially adjusted data, is because the selected trees are not actually the genetically best but are, to some extent, those that were favored in growth by their locations. Blocks of forest genetic trials in RCB designs are commonly very large and heterogeneous, so adjustments for block effects are clearly not enough to account for all the spatial variation (Saenz-Romero et al. 2001; Dutkowski et al. 2002, 2006; Zas 2006). The results presented here indicated that the loss of genetic gain through selection on unadjusted data in the presence of spatial autocorrelation varied typically between 10% and 20% and it can be over 30%, which means the loss of genetic gain of up to 4% of mean height. This loss of gain is quantitatively important, but more importantly, the tree breeder has no reliable prediction of genetic gain if spatial autocorrelation is ignored.

An increase in predicted genetic gain following spatial analysis was reported before (Costa e Silva et al. 2001;

Dutkowski et al. 2002, 2006). Selecting the best 20% of the parents or the best 5% of the offspring on the basis of a spatial model, Costa e Silva et al. (2001) found gain improvements in height growth of up to 23% and 13%, respectively, in relation to selection with a standard model in several progeny trials of different forest tree species. Among the different trials analyzed by these authors, a *P. pinaster* progeny trial in Portugal, not far from the trials of the present study, showed the greatest improvement. This trial, which followed a RCB design with plots of 8 trees ( $2 \times 4$ ) and a markedly patchy spatial variation, also stood out in further studies by the authors. Improvements in gain in this trial for diameter at age 5 years were 2.8% and 5.1% of the trial mean for parental and offspring selection, respectively (Dutkowski et al. 2002), and reached 28% and 59% for diameter at age 12 years (Dutkowski et al. 2006). However, this trial was an extreme case, and Dutkowski et al. (2006), analyzing a large data set of around four traits in each of 55 genetic trials, found generally lower improvements in genetic gains. Among all these traits and trials, height growth and RCB designs with large blocks showed the greatest benefits of spatial analysis. Height is also expected to be the trait with the highest response in our case, because it was the trait with the strongest spatial structure (Zas 2006).

For IS and WFS, the impact of spatial autocorrelation in selections was relatively independent on the selection intensity. However, for FS and F + WFS selection, the impacts were more variable, depending on the proportion of selected families in common between the standard and the spatial approach, and the mean value of these selections. For some selection intensities, the improvement in genetic gain was low and even null (Fig. 1), as observed in some cases by Dutkowski et al. (2006) for parent selection.

In the case of the conversion of the progeny trials to BSOs for seed production (Barnes 1995), a selection ratio of 30%, which would lead to densities of around 250 trees·ha<sup>-1</sup> (~6 m spacing), would be appropriate. Gains under this selection rate varied between 1.4% (FS in site D) and 5.2% (CS in site B). Although not high, these gains are considerable for a low-input breeding strategy such as the BSO and are within the range of those obtained by other authors after thinning BSOs (Fedorkov et al. 2005; Hansen et al. 2005; Kang et al. 2005). Nevertheless, this gain should be added to that achieved by the original selection of the plus tree parents, which is not known in our case; however, judging from results of other authors, it should be around 5%–10% (Butcher and Hopkins 1993). The interest of BSOs is high in the maritime pine breeding program in northwestern Spain, because the existing first-generation clonal seed orchards have been infected by several fungal pathogens that impede commercial seed production. There is a strong demand in Galicia for *P. pinaster* seed for reforestation that could be partially satisfied without much effort.

When roguing BSOs, tree breeders must also consider the GD of the retained trees. However, results presented here indicated that spatial adjustment had little effect on the GD estimates. Only in the case of using high selection rates with CS, was there a slightly reduced GD for selections based on spatially adjusted values, probably because FS became rela-

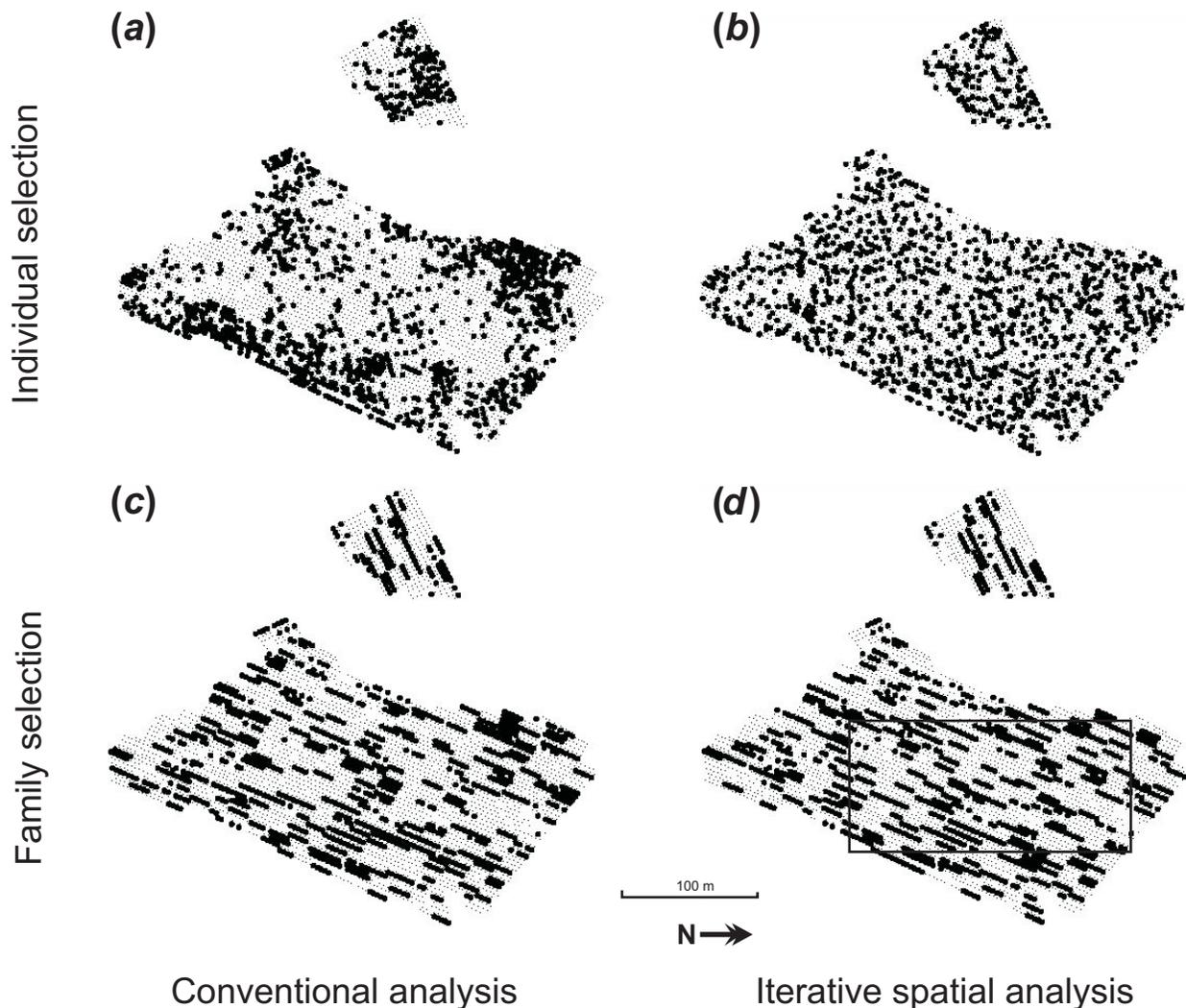
**Table 4.** Spatial analysis by distance indices (SADIE) of aggregation ( $I_p$ ) and associated probability levels ( $P_p$ ) for the spatial distribution of the remaining trees in the four progeny trials after selecting the best 30% of the trees using different selection methods.

| Selection method    | Analysis | Site A      |       | Site B      |       | Site C      |       | Site D      |       |
|---------------------|----------|-------------|-------|-------------|-------|-------------|-------|-------------|-------|
|                     |          | $I_p$       | $P_p$ | $I_p$       | $P_p$ | $I_p$       | $P_p$ | $I_p$       | $P_p$ |
| Individual          | N        | <b>1.77</b> | 0.003 | <b>1.40</b> | 0.005 | <b>1.87</b> | 0.003 | <b>2.92</b> | 0.003 |
|                     | ISA      | 0.93        | 0.713 | 1.01        | 0.405 | 1.12        | 0.107 | 0.88        | 0.993 |
| Within family       | N        | <b>1.76</b> | 0.003 | <b>1.50</b> | 0.005 | <b>1.83</b> | 0.005 | <b>2.70</b> | 0.005 |
|                     | ISA      | 0.76        | 0.997 | 1.11        | 0.155 | 0.91        | 0.850 | 0.88        | 0.995 |
| Family <sup>a</sup> | N        | <b>1.73</b> | 0.003 | <b>1.86</b> | 0.005 | <b>1.54</b> | 0.005 | <b>1.63</b> | 0.003 |
|                     | ISA      | <b>1.63</b> | 0.003 | <b>1.84</b> | 0.005 | <b>1.40</b> | 0.005 | <b>1.33</b> | 0.003 |
| Combined            | N        | <b>1.88</b> | 0.003 | <b>1.43</b> | 0.005 | <b>1.89</b> | 0.005 | <b>3.00</b> | 0.005 |
|                     | ISA      | 1.12        | 0.120 | 1.14        | 0.120 | 1.18        | 0.050 | 1.00        | 0.465 |
| F+WFS*              | N        | <b>1.74</b> | 0.003 | <b>1.41</b> | 0.005 | <b>1.81</b> | 0.003 | <b>2.56</b> | 0.005 |
|                     | ISA      | 0.98        | 0.520 | 0.80        | 0.965 | 0.93        | 0.733 | 0.81        | 0.995 |

**Note:** Results are presented for selection using the conventional statistical model (N) or the iterative spatial analysis (ISA) procedure. Spatial distributions with indices  $I_p$  shown in boldface are significantly aggregated ( $p < 0.05$ ).

\*In FS and F + WFS selection the final number of selected trees was, given the constraints of family size, as close as possible to the corresponding 30%.

**Fig. 4.** Spatial distribution of the remaining trees after roguing site D, leaving the 30% best trees through individual selection (a and b) or family selection (c and d) and based on a conventional statistical model (a and c) or on the iterative spatial analysis (b and d). The rectangle considered for the calculation of the spatial analysis by distance indices (SADIE) is shown in Fig. 4d.



tively more important in this case. However, for densities appropriate for seed production in BSOs, differences in GD were minor, and spatial adjustments did not alter the general picture. With low selection rates, differences in GD among the different selection methods were also small, and only FS showed slightly lower GD estimates.

Thinning the progeny trials from original unadjusted data led to aggregate distributions of the selected trees whatever the method and intensity of selection (Figs. 3 and 4 and Table 4). Clumping will seriously hamper seed production in the resulting thinned BSO. The optimal method and intensity of selection for thinning BSOs is usually determined based on the estimated genetic gain and gene diversity after selections (David et al. 2003; Fedorkov et al. 2005; Kang et al. 2005). The GD of the seed crop is usually estimated under panmixia. However, variation in strobili production, fecundity, and flowering synchronization are known to reduce the effective gene diversity (e.g., Kang et al. 2001). Although not so intensively studied, the distance between mates also influences the diversity of the seed crop and the effective population size. Closely adjacent trees would have higher probability of intermating than distant trees, and hence, aggregate distributions will be less panmictic than random distributions (Burczyk and Prat 1997). Furthermore, aggregate distributions imply a waste of the available space, an increase in competition between trees, and higher local density, which causes narrower crowns and lower seed production. Apart from genetic gain and gene diversity considerations, the spatial distribution of the remaining trees should always be taken into account when planning thinning in BSOs. In our case, with the selection rate of 30% that gave an appropriate density for seed production, FS should be clearly rejected because of spatial aggregation (Table 4). Furthermore, although CS gave the highest gain and acceptable GD, this method generated the largest  $I_p$  values, which suggested slight clumping in some cases, such as in site C (Table 4). In these cases, an alternative selection method such as IS or F + WFS should be considered to achieve random or regular spatial distribution at the cost of losing some genetic gain.

The SADIE methodology (Perry 1995) appeared as a simple, effective, and versatile tool for analyzing the spatial distribution of the selected trees. Among other available outputs, SADIE computes an overall aggregation index irrespective of distance and gives a simple and intuitively satisfying test for departures from randomness (Rozas and Camarero 2005; Perry et al. 2006). The method has been successfully used in numerous animal and plant ecology studies (e.g., Maestre et al. 2005; Birkhofer et al. 2006). The results presented here demonstrated further usefulness of the method in a field as different as tree breeding.

This study has demonstrated that spatial analysis is essential for efficient selection in progeny trials, particularly those with multiple-tree plots with large and spatially heterogeneous blocks. Although the loss of genetic gain when spatial autocorrelation is ignored was generally just moderate, the resulting spatial distribution of the selected trees may be strongly aggregated, which may seriously affect the quantity and quality of the seed crop in thinned breeding seedling orchards. Furthermore, the spatial distribution of the selected trees should be considered even when spatial models are

used, as different selection methods may result in different degrees of spatial randomness of the selected trees.

## Acknowledgments

This study was funded by the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (grant Nos. RTA02-109 and RTA05-173) and the Galician government (grant No. PGIDIT02-PXIC50201PN). I thank Pepe Zas for his help in developing appropriate routines for the automatic execution of the SADIE program for a large collection of input files. Thanks are extended to Dr. Luis Sampedro, Dr. Esther Merlo, Dr. Raquel Díaz, Gloria Bustingorri, and one anonymous referee for valuable comments on earlier versions of the manuscript. The thorough revision of the manuscript and the very helpful comments by Dr. Rowland Burdon and Dr. Tony Shelbourne (Ensis-Genetics, Rotorua, New Zealand) are specially acknowledged.

## References

- Barnes, R.D. 1995. The breeding seedling orchard in multiple population breeding strategy. *Silvae Genet.* **44**: 81–88.
- Birkhofer, K., Henschel, J., and Scheu, S. 2006. Spatial-pattern analysis in a territorial spider: evidence for multi-scale effects. *Ecography*, **29**: 641–648. doi:10.1111/j.0906-7590.2006.04661.x.
- Burczyk, J., and Prat, D. 1997. Male reproductive success in *Pseudotsuga menziesii* (Mirb.) Franco: the effects of spatial structure and flowering characteristics. *Heredity*, **79**: 638–647. doi:10.1038/sj.hdy.6882600.
- Burdon, R.D., and Shelbourne, C.J.A. 1971. Breeding populations for recurrent selection: conflicts and possible solutions. *N.Z. J. For. Sci.* **1**: 174–193.
- Butcher, T.B., and Hopkins, E.R. 1993. Realised gains from breeding *Pinus pinaster*. *For. Ecol. Manage.* **58**: 211–231. doi:10.1016/0378-1127(93)90146-E.
- Costa e Silva, J., Dutkowski, G.W., and Gilmour, A.R. 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. doi:10.1139/x01-123.
- David, A., Pike, C., and Stine, R. 2003. Comparison of selection methods for optimizing genetic gain and gene diversity in a red pine (*Pinus resinosa* Ait.) seedling seed orchard. *Theor. Appl. Genet.* **107**: 843–849. doi:10.1007/s00122-003-1330-0. PMID:12838390.
- Dutkowski, G.W., Costa e Silva, J., Gilmour, A.R., and Lopez, G.A. 2002. Spatial analysis methods for forest genetic trials. *Can. J. For. Res.* **32**: 2201–2214. doi:10.1139/x02-111.
- Dutkowski, G.W., Costa e Silva, J., Gilmour, A.R., Wallendorf, H., and Aguiar, A. 2006. Spatial analysis enhances modelling of a wide variety of traits in forest genetic trials. *Can. J. For. Res.* **36**: 1851–1870. doi:10.1139/X06-059.
- Falconer, D.S., and Mackay, T.F.C. 2001. *Introducción a la genética cuantitativa*. 4th ed. Editorial Acribia S.A., Zaragoza, Spain.
- Fedorkov, A., Lindgren, D., and David, A. 2005. Genetic gain and gene diversity following thinning in a half-sib plantation. *Silvae Genet.* **54**: 185–189.
- Fu, Y., Yanchuk, A.D., and 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. doi:10.1139/x99-046.
- Gezan, S.A., White, T.L., and Huber, D.A. 2006. Comparison of experimental designs for clonal forestry using simulated data. *For. Sci.* **52**: 108–116.
- Hamann, A., Namkoong, G., and Koshy, M.P. 2002. Improving

- precision of breeding values by removing spatially autocorrelated variation in forestry field experiments. *Silvae Genet.* **51**: 210–215.
- Hansen, J.K., Wellendorf, H., and Kjaer, E.D. 2005. Low cost improvement of coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii* (Mirb.) Franco) by application of the breeding seed orchard approach in Denmark. *Silvae Genet.* **54**: 218–227.
- Kang, K.S., Lindgren, D., and Mullin, T.J. 2001. Prediction of genetic gain and gene diversity in seed orchard crops under alternative management strategies. *Theor. Appl. Genet.* **103**: 1099–1107. doi:10.1007/s001220100700.
- Kang, K.S., El-Kassaby, Y.A., Han, S.U., and Kim, C.S. 2005. Genetic gain and diversity under different thinning scenarios in a breeding seed orchard of *Quercus accutissima*. *For. Ecol. Manage.* **212**: 405–410. doi:10.1016/j.foreco.2005.03.063.
- Legendre, P. 1993. Spatial autocorrelation: a trouble or new paradigm? *Ecology*, **74**: 1659–1673. doi:10.2307/1939924.
- Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. 1996. SAS System for mixed models. SAS Institute Inc. Cary, N.C.
- Maestre, F., Escudero, A., Martinez, I., Guerreros, C., and Rubio, A. 2005. Does spatial pattern matter to ecosystem functioning? Insights from biological soil crusts. *Funct. Ecol.* **19**: 566–573. doi:10.1111/j.1365-2435.2005.01000.x.
- Olsson, T., Lindgren, D., and Li, B. 2001. Balancing genetic gain and relatedness in seed orchards. *Silvae Genet.* **50**: 222–227.
- Perry, G.L.W., Miller, B.P., and Enright, N.J. 2006. A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. *Plant Ecol.* **187**: 59–82. doi:10.1007/s11258-006-9133-4.
- Perry, J.N. 1995. Spatial analysis by distance indices. *J. Anim. Ecol.* **64**: 303–314. doi:10.2307/5892.
- Rossi, R.E., Mulla, D.J., Journel, A.G., and Franz, E.H. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecol. Monogr.* **62**: 277–314. doi:10.2307/2937096.
- Rozas, V., and Camarero, J.J. 2005. Técnicas de análisis espacial de patrones de puntos aplicadas en ecología forestal. *Invest. Agrar. Sist. Recur. For.* **14**: 79–97.
- Saenz-Romero, C., Nordheim, E.V., Guries, R.P., and Crump, P.M. 2001. A case study of a provenance/progeny test using trend analysis with correlated errors and SAS PROC MIXED. *Silvae Genet.* **50**: 127–135.
- SAS Institute Inc. 1999. SAS/STAT user's guide. Version 8. SAS Institute Inc., Cary, N.C.
- Wright, J.W. 1976. Introduction to forest genetics. Academic Press, New York.
- Zas, R. 2006. Iterative kriging for removing spatial autocorrelation in analysis of forest genetic trials. *Tree Genet. Genomes*, **2**: 177–186. doi:10.1007/s11295-006-0042-4.
- Zas, R., Merlo, E., and Fernández-López, J. 2004. Genetic parameter estimates for maritime pine in the Atlantic coast of North-west Spain. *For. Genet.* **11**: 45–53.