

# Spatial statistical analysis and selection of genotypes in plant breeding

João Batista Duarte<sup>(1)</sup> and Roland Vencovsky<sup>(2)</sup>

<sup>(1)</sup>Universidade Federal de Goiás, Escola de Agronomia e Engenharia de Alimentos, Caixa Postal 131, CEP 74001-970 Goiânia, GO, Brazil. E-mail: jbduarte@agro.ufg.br <sup>(2)</sup>Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Dep. de Genética, Caixa Postal 83, CEP 13418-070 Piracicaba, SP, Brazil.

**Abstract** – The objective of this study was to evaluate the efficiency of spatial statistical analysis in the selection of genotypes in a plant breeding program and, particularly, to demonstrate the benefits of the approach when experimental observations are not spatially independent. The basic material of this study was a yield trial of soybean lines, with five check varieties (of fixed effect) and 110 test lines (of random effects), in an augmented block design. The spatial analysis used a random field linear model (RFML), with a covariance function estimated from the residuals of the analysis considering independent errors. Results showed a residual autocorrelation of significant magnitude and extension (range), which allowed a better discrimination among genotypes (increase of the power of statistical tests, reduction in the standard errors of estimates and predictors, and a greater amplitude of predictor values) when the spatial analysis was applied. Furthermore, the spatial analysis led to a different ranking of the genetic materials, in comparison with the non-spatial analysis, and a selection less influenced by local variation effects was obtained.

**Index terms:** augmented design, mixed model, information recovery, autocorrelation, correlated data, geostatistics.

## Seleção de genótipos e análise estatística espacial no melhoramento de plantas

**Resumo** – O objetivo deste trabalho foi avaliar a eficiência da análise estatística espacial na seleção de genótipos de plantas num programa de melhoramento. Buscou-se demonstrar os benefícios potenciais dessa abordagem quando as observações experimentais não são espacialmente independentes. O material consistiu de um ensaio de competição de linhagens de soja, com cinco cultivares testemunhas (de efeitos fixos) e 110 novos genótipos (de efeitos aleatórios), delineado em blocos aumentados. O ajuste espacial foi feito pelo modelo linear de campo aleatório (RFLM), com função de autocovariância estimada a partir dos resíduos da análise sob erros independentes. Os resultados apontaram uma autocorrelação residual de magnitude e alcance significativos, o que garantiu à abordagem espacial uma melhoria considerável na discriminação dos tratamentos genéticos – aumento do poder dos testes estatísticos, redução nos erros padrão de estimativas e de preditores e alargamento na amplitude das previsões genotípicas. A análise espacial levou a um diferente ordenamento das linhagens em relação à análise não espacial e, finalmente, a uma seleção menos influenciada por efeitos da variação local.

**Termos para indexação:** delineamento aumentado, modelo misto, recuperação de informação, autocorrelação, dados correlacionados, geoestatística.

### Introduction

In plant breeding, two features indicate the preliminary phases of selective programs: the large numbers of new genotypes to be evaluated and the small amount of material for their propagation. Both of them limit the use of replications of these genetic treatments, which are frequently evaluated in a single experimental plot, i.e., without replications. Federer (1956) proposed the augmented experimental designs to deal with this type

of limitation, which allow the adjustment of the test line (new treatment) means for environmental effects (blocks, lines, or columns) estimated on the basis of repeated check genotypes. The author also presented the corresponding methods of statistical analysis, based on ordinary least squares (OLS) and, therefore, on the assumption of independence among observations.

The limited availability of propagation material, such as seeds and tubers, on the other hand, forces the breeder to adopt small plots, usually with just one or two rows of

plants. This increases the chance of violating the independence among observations assumed when using the OLS method, due to the likely similarity of observations of neighboring plots (Stroup et al., 1994). This phenomenon, referred to as spatial correlation – also called spatial dependence or autocorrelation – can seriously affect the comparison of treatments. Es & Es (1993) have demonstrated that when this correlation occurs, the statistical tests associated with contrasting treatments in plots nearer together have higher probabilities of type II error, which consists of different treatments appearing to be identical. On the other hand, higher probabilities of type I error, i.e., identical treatments appearing to be different, were observed in the contrasts between treatments in which plots were farther apart.

The traditional analysis of variance relies on randomization to neutralize the harmful effects of this type of correlation, but frequently this is not attained adequately (Stroup et al., 1994). For this reason, Kempton et al. (1994) support a greater use of methods that consider some accounting for spatial dependence to improve the precision of variety trials. Recent advances in statistics for spatially distributed data have provided a number of alternative methods. One interesting approach is that of Zimmerman & Harville (1991). In this analysis, the plot effect (trend + error) is modeled in such a way that the observations are collectively taken as a partial outcome of a random field, similar to predictive models used in geostatistical applications (Martínez, 1994). The model aims at estimating the general covariance function, which is used in estimation and prediction, through generalized least squares (GLS). Therefore, it is a mixed linear model with spatially correlated errors, called a random field linear model (RFLM).

Due to the relatively limited use of these techniques among plant breeders, it is necessary to assess their effects on selection of genotypes to finally demonstrate their true potential. This study illustrates the application of the RFLM approach, adapting it to the augmented block design, which is typical in the preliminary phases of the selection process in plant breeding. The attempt does not intend to represent the best spatial approach for the set of data analyzed, but, rather, to demonstrate the benefits of a less restrictive statistical analysis in comparison with the traditional one, based on spatially independent observations.

## Material and Methods

The data used in this study were obtained in a soybean variety trial, with F<sub>6:3</sub> lines of the semi-early maturity group, conducted in the locality of Areão, municipality of Piracicaba, SP, Brazil, in 1999/1995. The trial is part of a selection program conducted to increase soybean yield, carried out by the Department of Genetics of Esalq/USP. Genetic materials were evaluated in augmented block design, with  $t = 5$  check varieties (Bossier, Davis-1, IAC-12, IAS-5 and Viçoja) and  $p = 110$  test lines, distributed in  $b = 4$  blocks with approximately 50 plots each. The plot corresponded to two rows of plants, spaced 0.6 m apart and 5 m long. Only grain yield data (kg ha<sup>-1</sup>) were considered here. For spatial statistical analysis, it was necessary, in addition, to obtain the distances (meters) among plots, which was done from the geographical coordinates of the center of each plot in the experimental field grid – COORDX represents the width coordinate of the plots and COORDY, the length coordinate.

Two mathematical models were used for statistical analysis: i) a model which assumes spatially independent observations; and ii) a model allowing spatial correlations among observations. In both cases, the effects of test lines were taken as random, and here were assumed to be derived from a single base population, that is, varying randomly about a common mean. For this reason, the independent error analysis here does not correspond to the fixed model (OLS). Thus, both are mixed models, despite the adjustment for checks. The only difference between them is the assumption on the experimental error.

In the case of spatially independent observations, this analysis is described as intergenotypic information recovery analysis (Wolfinger et al., 1997; Federer, 1998). One peculiarity of such analysis in the augmented designs is that the mathematical model needs to accommodate two types of treatment effects: fixed effects for the checks ( $t$  populations) and random effects for the test lines. These lines constitute the  $(t+1)^{\text{th}}$  population, which is also assumed to have a fixed effect. Thus, in the first alternative (i), the observations can be individually characterized by the model (an adaptation to the model of Scott & Milliken, 1993, proposed by Duarte, 2000):

$$Y_{ijk} = \mu + b_j + c_k + g_{i(k)} + e_{ijk}$$

in which  $Y_{ijk}$  is the observation in the plot with genotype  $i$ , stemming from population  $k$  in block  $j$ ;  $\mu$  is the constant common to all observations;  $b_j$  is the fixed effect of the  $j^{\text{th}}$  block ( $j=1,2,\dots,b$ );  $c_k$  is the fixed effect of the  $k^{\text{th}}$  population ( $k=1,2,\dots,t,t+1$ ; here the check cultivars plus

the population of lines);  $g_{i(k)}$  is the effect of the  $i^{\text{th}}$  genotype within the  $k^{\text{th}}$  population, assumed to be fixed and with a null mean, if the genotype is a check ( $i_{(k)} = 1$ ), or random with independent distribution  $N(0, \sigma_g^2)$ , if the genotype is a test line ( $i_{(k)} = 1, 2, \dots, p_k$ , with  $\sum p_k = p$ ); and  $e_{ijk}$  is the random experimental error associated with the  $ijk^{\text{th}}$  plot, which is assumed to be independent, that is, null covariance among the errors of different plots, and with distribution  $N(0, \sigma_e^2)$ .

In model (ii), the term  $e_{ijk}$  is assumed to have the distribution  $e_{ijk} \sim N[0, C(h)]$ , in which  $C(h)$  is the covariance between two errors of plots which are  $h$  units of distance apart ( $h \geq 0$ ). If such errors are denoted by  $e_{(s)}$  and  $e_{(s+h)}$ , in which  $s$  represents the spatial position of the  $ijk^{\text{th}}$  plot, in the RFLM approach,  $C(h)$  is defined as (Littell et al., 1996):

$$C(h) = \begin{cases} \sigma^2, & \text{if } h = 0; \text{ and} \\ \sigma_{e_{(s)}, e_{(s+h)}} = \sigma^2 [f(h)], & \text{if } h > 0 \end{cases}$$

Thus, it is assumed that the covariance of the errors is a function of the distance that separates the corresponding plots ( $f(h)$ ). However, this is not predetermined, but is estimated from the “uniformity experiment” suggested by the residuals ( $\hat{e}_{ijk}$ ) of the independent error model adjustment.

Representing the observations by a vector  $y$ , both models can be expressed in matrix notation by the general mixed linear model (Henderson, 1984):  $y = X\beta + Z\gamma + \varepsilon$ ; with  $\gamma \sim N(\phi, G)$ ,  $\varepsilon \sim N(0, R)$ ,  $E(y) = X\beta$  and  $\text{Var}(y) = V = ZGZ' + R$ .

The fixed effects are in parametric vector  $\beta$ ; the random effects, in parametric vector  $\gamma$ , except the errors that are in vector  $\varepsilon$ ;  $X$  and  $Z$  are incidence matrices of the effects contained in  $\beta$  and  $\gamma$ , respectively. The random genotypic effects ( $\gamma$ ) are assumed, without loss of generality, to have a normal distribution with a null mean ( $\phi$ ) and matrix of covariance  $G = I\sigma_g^2$  (where  $I$  is an identity matrix). The experimental errors are presumed to have a normal distribution with a null mean and a generic matrix of covariance  $R$ . Thus, in the first model (i),  $R = I\sigma_e^2$ , while in the other (ii),  $R = \Sigma$ , i.e., a non-diagonal matrix with structure defined by the general covariance function and by the autocorrelation range.

The first step in this spatial analysis is the adjustment of the model which postulate spatial independence among observations. The components of variance  $\sigma_g^2$  and  $\sigma_e^2$  were estimated by restricted maximum likelihood

(REML). The estimated residual vector of this adjustment is:  $\hat{\varepsilon} = y - \hat{y}$ , with  $\hat{y} = X\beta^0 + Z\tilde{\gamma}$ , in which  $\beta^0 = (X'V^{-1}X)^{-1}X'V^{-1}y$  and  $\tilde{\gamma} = GZ'V^{-1}(y - X\beta^0)$ , representing the solution vectors of the mixed model equations (Henderson, 1984). The residuals were then used to estimate the spatial correlation structure. This was done graphically by means of a so-called semivariogram or simply variogram (Stroup et al., 1994).

In this graphic representation, estimated values of semivariance,  $\hat{S}(h)$ , are plotted against their respective distances  $h$ , resulting in a scatter plot (sample variogram). The semivariance is defined as:  $S(h) = 1/2 \text{Var}[e_{(s+h)} - e_{(s)}]$ ; which is estimated by  $\hat{S}(h) = \frac{1}{2N(h)} \sum_{N(h)} [\hat{e}_{(s+h)} - \hat{e}_{(s)}]^2$ , with

$N(h)$  being the number of differences at the distance  $h$ . In this graph, values  $\hat{S}(h)$  that are distributed randomly as a function of  $h$  reflect independent observations (residuals). The typical configuration of spatial dependence among observations occurs if values  $\hat{S}(h)$  tend to increase as  $h$  increases up to a certain distance (range), after which the semivariance stabilizes reaching a plateau (or sill). Less variability is associated with smaller distances. The spatial correlation range ( $a$ ) is the mean distance influence of an observation (plot), asserted here to be uniform in all directions (isotropy). The sill ( $\sigma^2$ ) corresponds to the intrinsic variance of the variable under study ( $\text{Var}[e_{(s)}] = \text{Cov}[e_{(s)}, e_{(s)}]$ ), which is also equivalent to the covariance between residuals of plots separated by a distance equal to or greater than the range ( $\text{Cov}[e_{(s)}, e_{(s+h)}]$ , with  $h \geq a$ ).

There is an advantage in evaluating spatial dependence by means of the variogram. Under stationarity – spatial law unaffected by translation – the variogram has a direct and simple relation to the function of autocovariance  $C(h)$ , that is:  $S(h) = \sigma^2 - C(h)$ ; in which  $\sigma^2 = C(h = 0)$  (Es & Es, 1993; Stroup et al., 1994; Pannatier, 1996). Thus, fitting a continuous model to the sample variogram, the corresponding spatial covariance function for this relation is obtained. The most commonly utilized variogram functions are the so-called spherical, exponential and Gaussian models (Grondona & Cressie, 1991; Zimmerman & Harville, 1991; Vieira, 2000). Due to the wide application of the variogram in geostatistics,

software that facilitates this adjustment is available (ex: Variowin; Geo-Eas). In such programs, the search for the function that best fits the observational points is carried out by changing slightly the values of  $\sigma^2$  and  $a$ . In the present case, the exponential model provided the best fit, corresponding to the following covariance function (for isotropic random fields):  $C(h) = \sigma^2 \exp(\frac{-3h}{a})$ .

After defining parameters ( $\sigma^2$  and  $a$ ) and the general covariance function, the next step is fitting the model to account for spatial dependence ( $R = \Sigma$ ). This involves obtaining estimates, predictors, and statistical tests related to treatment effects, which must be free from estimated autocorrelation effects. To evaluate only the effects of the spatial adjustment on the statistical analysis, the same estimate of  $\sigma_g^2$  obtained in the former analysis (under  $R = I\sigma_e^2$ ) was used. The following procedure consisted of resolving the mixed model equations (Henderson, 1984):

$$\begin{bmatrix} X'\Sigma^{-1}X & X'\Sigma^{-1}Z \\ Z'\Sigma^{-1}X & Z'\Sigma^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta^0 \\ \tilde{\gamma} \end{bmatrix} = \begin{bmatrix} X'\Sigma^{-1}y \\ Z'\Sigma^{-1}y \end{bmatrix},$$

which solutions have already been reported.

## Results and Discussion

### Characterization of spatial covariance

The experimental data showed a positive spatial correlation of first-order to sixth-order, in the series of residuals (Table 1). This fact is indicative of the violation of spatial independence among observations postulated by the first model (under  $R = I\sigma_e^2$ ). Residuals of this

**Table 1.** First to tenth-order autocorrelations ( $\hat{\rho}$ ) and the respective Durbin-Watson statistics (d) for the residual ( $\hat{e}_{ijk}$ ) series obtained from the adjustment of an augmented block design model with independent errors.

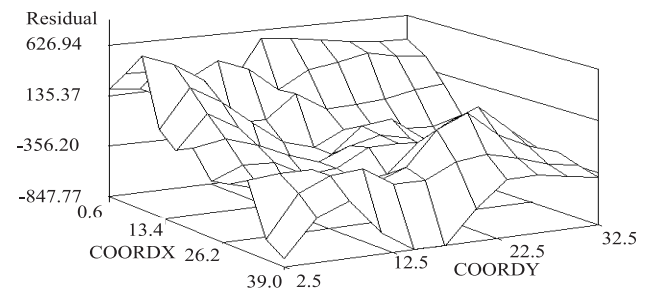
Order	$\hat{\rho}$	d	Pr<d <sup>(1)</sup>
1	0.64125	0.7003	0.0001
2	0.50359	0.9599	0.0001
3	0.42519	1.0895	0.0001
4	0.38857	1.1313	0.0001
5	0.29623	1.2942	0.0001
6	0.16385	1.5386	0.0143
7	0.09528	1.6624	0.0839
8	-0.00189	1.8176	0.3440
9	-0.06130	1.9241	0.6153
10	-0.10327	2.0027	0.7964

<sup>(1)</sup>Marginal probability of the statistical test ( $H_0: \rho = 0$ ) (one-sided to the left, in this case).

analysis were not randomly distributed in the experimental field (Figure 1). Rather, there is a clear tendency for larger residual values ( $\hat{e}_{ijk}$ ) to be concentrated in the top of the field map graph, that is, to be associated with plots having smaller COORDX values.

This fact also determined a predominant gradient in the direction of plot widths (COORDX). Considering this is how the blocks were constructed, it is possible that such an orientation may not have been ideal. Given the features of the residual surface, which provides an estimate of the uniformity trial underlying the experiment, it is reasonable to suppose that a lengthwise blocking of the plots would have been more effective in controlling local variation. The possibility of making this diagnosis represents an advantage of the spatial approach, which creates perspectives for the application of alternative forms of *a posteriori* local control or post-blocking (Federer, 1998).

The variogram obtained for distances less than 30 m is showed in Figure 2. The configuration of the dots is typical of stochastic processes with spatial dependence, that is, with decreasing variability as distance decreases. After 20 m (range) the variability tends to stabilize. The value of this plateau represents the residual variance among independent plots, and the existence of the increasing variogram with a plateau is an indication that the intrinsic hypothesis of stationarity was satisfied (Vieira, 2000). Furthermore, on the assumption of isotropy, the continuous function that best fits the dots is the exponential semivariance model:  $S(h) = \sigma^2 [1 - \exp(\frac{-3h}{a})]$ , with:  $\sigma^2 = 126450 \text{ (kg ha}^{-1}\text{)}^2$  and  $a = 20.4 \text{ m}$ . Consequently, the respective autocovariance function is expressed by  $C(h) = 126450 \exp(\frac{-3h}{6,8})$ . This defines the residual

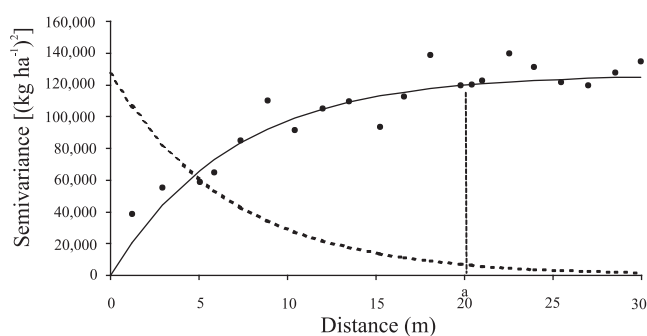


**Figure 1.** Residuals ( $\text{kg ha}^{-1}$ ) of the augmented block design model adjustment, with recovery of test lines information, under independent errors as a function of the plot center coordinates, in meters (COORDX and COORDY).

covariance matrix  $R = S$ , whose main diagonal elements were all equal to 126450 and the off-diagonal elements were equal to  $126450 \exp(-\frac{h}{6.8})$ , in which  $h$  is the distance that separates each two plots identified by a row and a column in the matrix. Thus, the spatial covariance inherent to the experiment was characterized. The implications of the use, or not, of this information in statistical analysis are evaluated in the following section.

### Comparison of the spatial and non-spatial statistical analysis models

Models with a larger number of covariance parameters always exhibit better fit than those with a simpler structure. For this reason, comparative criteria that penalize the more parametrized models, such as the Akaike's Information Criterion (AIC) and Schwarz's Bayesian Criterion (BIC) should be adopted. Both are based on the value which maximizes the restricted likelihood logarithm,  $L_{REML}(G,R)$ , reduced from a function of the number of parameters. Thus, the model with the greatest AIC or BIC values should be preferred (Littell



**Figure 2.** Sample variogram (dots) of the residuals of an augmented block design analysis that assumed independent errors, and adjustment (continuous line) by the exponential semivariance model [ $a = 20.4$  m;  $s^2 = 126450$  (kg ha<sup>-1</sup>)<sup>2</sup>]. The dotted line illustrates the corresponding autocovariance function.

et al., 1996). The results in Table 2 show that the covariance structure  $R = \Sigma$  provides a better fit to the respective model in comparison with the independent error model ( $R = I\sigma_e^2$ ).

With regard to statistical tests related to the genotypic effects, it was observed that variation among the six fixed populations was not significant in the first analysis (at the 5% level of probability), but reached high statistical significance ( $p$ -value < 0.01) in the spatial analysis (Table 3). With further partitioning of the population effects, the F values were also higher under the spatial analysis, both in the detection of differences among checks (four degree of freedom) and in the contrast between checks and test lines (one degree of freedom). Considering the three contrasts chosen to illustrate the comparison among some of these lines, the superiority of the spatial analysis was again evident and even greater. While the analysis under  $R = I\sigma_e^2$  did not detect any difference among these genotypes ( $p$ -value > 0.90), the spatial analysis showed that two of the three contrasts were significant ( $p$ -value < 0.025). These results reflect greater genotypic discrimination ability under the spatial analysis, compared with the non-spatial procedure.

This superiority was confirmed when the predicted genotypic values (EBLUP) were considered. While in the first analysis these varied between -98.2 and 100.5 (complete data in Duarte, 2000), with a range of about

**Table 2.** Characteristics of the fitting of non-spatial and spatial models for grain yield data (kg ha<sup>-1</sup>) in a soybean variety trial in an augmented block design.

Description	Value	
	Non-spatial analysis	Spatial analysis
$L_{REML}(G, R)$	-881.125	-838.975
Akaike's Information Criterion (AIC)	-883.125	-841.975
Schwarz's Bayesian Criterion (BIC)	-885.896	-846.131

**Table 3.** Statistical tests of some genotypic effects obtained on the basis of spatial and non-spatial analysis models<sup>(1)</sup>.

Source	NDF <sup>(2)</sup>	Non-spatial analysis			Spatial analysis		
		DDF <sup>(2)</sup>	F	Pr > F	DDF <sup>(2)</sup>	F	Pr > F
Populations	5	12.0	2.71	0.0727	31.9	3.95	0.0067
Checks	4	11.4	0.93	0.4793	31.5	1.79	0.1547
Checks vs. Test lines	1	15.5	9.89	0.0065	33.3	10.28	0.0030
Line G1 vs. Line G3	1	0.20	0.00	0.9859	23.7	0.46	0.5064
Line G1 vs. Line G24	1	0.20	0.01	0.9712	25.8	6.00	0.0214
Line G3 vs. Line G24	1	0.20	0.01	0.9575	25.0	9.82	0.0044

<sup>(1)</sup>Grain yield data, in kg ha<sup>-1</sup>; soybean variety trial in augmented block design. <sup>(2)</sup>NDF and DDF are the numbers of degrees of freedom of numerator and denominator of F (Snedecor) statistic, respectively; the last obtained by the Satterthwaite approximation.

200 kg ha<sup>-1</sup>, in the spatial analysis the detected range was greater than 500 kg ha<sup>-1</sup> (values between -337 and 200.5). This represents an increase of more than 150% in the differentiation among the test lines, in favor of the spatial analysis. The smaller standard errors associated with EBLUP also confirm the better genotypic discrimination of this analytic model. Pontes (2002) has demonstrated a gain of 7% in the efficiency of these predictors when an iterative process to estimate the variogram and its parameters ( $a$  and  $\sigma^2$ ) was used.

When a selection intensity of 25% of the most productive lines was assumed (28 in 110 genotypes), a coincidence of only 46% between the two statistical analysis models was observed (Table 4). In addition, among the genotypes selected by the more traditional analysis (non-spatial), at least 30% would occupy poor ranking positions in the spatial analysis (up to fiftieth position). Examples include the following lines:

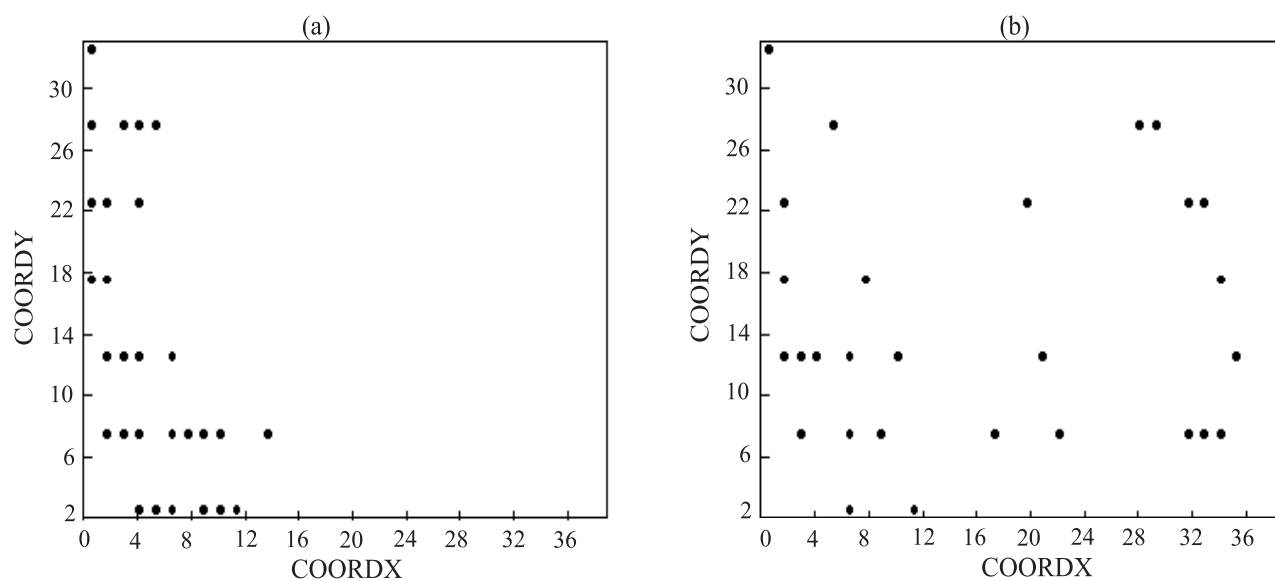
USP 93-2048, USP 93-2393, USP 93-2153 and USP 93-2198. On the other hand, four lines classified in the spatial analysis as among the ten most productive would be discarded using the other analysis (under  $R = I\sigma_e^2$ ).

The disagreement between these selections can be better understood if the spatial positions of plots with the selected lines in the experimental field are considered. The evidence of the effect of spatial adjustment on selection can be seen in Figure 3. When the non-spatial model was used, the selected genotypes were located exclusively in the left side strip of the experimental field, probably its most fertile area. However, when the spatial adjustment was taken into account, the selected genotypes were detected in plots scattered throughout the whole experimental area. The predominance of genotypes from the left side-stripe can be explained as a result of possible remaining fertility effects or of the

**Table 4.** Predictors (EBLUP) of individual genotypic effects of 28 soybean lines, respective standard errors and ranking of test lines under two statistical analysis models<sup>(1)</sup>.

Lines	Non-spatial analysis			Spatial analysis		
	EBLUP	Standard error	Rank	EBLUP	Standard error	Rank
USP 93-2802	100.48	123.96	1	200.50	103.71	1
USP 93-2850	96.97	123.98	2	172.97	103.68	2
USP 93-2547	76.77	123.98	3	66.91	103.27	20
USP 93-2075	76.37	123.95	4	64.51	107.80	22
USP 93-2302	72.64	123.96	5	107.14	103.34	11
USP 93-2114	70.08	123.95	6	42.59	107.77	31
USP 93-2623	69.27	123.96	7	67.11	103.33	19
USP 93-2642	68.46	123.98	8	13.76	111.13	50
USP 93-2722	66.80	123.96	9	83.48	103.35	14
USP 93-2753	65.68	123.96	10	156.46	103.34	4
USP 93-2171	62.45	123.95	11	126.62	103.54	6
USP 93-2159	58.72	123.96	12	150.22	103.69	5
USP 93-2479	56.78	123.98	13	-21.24	111.08	65
USP 93-2881	53.56	123.96	14	172.84	103.34	3
USP 93-2187	53.41	124.01	15	47.13	111.11	29
USP 93-2037	52.79	123.95	16	5.82	111.74	53
USP 93-2148	52.43	123.96	17	38.83	103.63	35
USP 93-2475	50.86	123.96	18	33.96	103.35	37
USP 93-2474	48.70	124.01	19	64.17	111.68	23
USP 93-2048	48.53	123.95	20	-45.25	104.17	83
USP 93-2198	47.49	123.96	21	-32.91	111.20	75
USP 93-2266	46.90	124.01	22	25.62	103.35	42
USP 93-2153	46.51	123.95	23	-40.82	103.73	81
USP 93-2916	44.66	124.01	24	17.93	103.33	48
USP 93-2418	43.98	124.01	25	76.09	102.76	18
USP 93-2393	40.76	123.96	26	-43.15	103.34	82
USP 93-2077	38.15	123.98	27	-4.27	107.06	55
USP 93-2985	37.30	123.95	28	25.77	108.87	41

<sup>(1)</sup>Grain yield data in kg ha<sup>-1</sup>; variety trial with 110 lines and five checks, in augmented block design.



**Figure 3.** Location of plots with the most productive test lines (25%) among the 110 evaluated lines (unreplicated), using two models of statistical analysis: with  $R = I \sigma^2$  (non-spatial) (a) and with  $R = \Sigma$  (spatial) (b) of a soybean variety trial, in augmented block design.

breeder's preference in allocating genotypes of the same parent side by side. In any event, what is expected from experiments of this nature is an outcome as shown in part (b) of Figure 3 rather than one displayed in its part (a). Similar results are also reported by Besag & Kempton (1986), Cullis et al. (1989), and Kempton & Gleeson (1997).

Considering that the cause of the divergence in the two selections was the genotypic adjustment for position effects, which are of purely environmental nature, it can be concluded that, in similar conditions, the use of spatial analysis can assure greater efficiency to the breeding programs.

### Conclusions

1. In variety trials with large numbers of treatments and limited availability of propagation material, experimental observations can not be spatially independent; in such conditions, spatial analysis allows better discrimination among genotypes, because it provides increased power in statistical tests, reduced standard errors of genotypic estimates, and greater amplitudes among predicted values.

2. The spatial analysis can be led to a different ranking of the genetic materials, in comparison with the non-spatial analysis, and a selection less influenced by local

variation; such differences may have important consequences for the final outcome of plant breeding programs.

### Acknowledgements

To Dr. Natal Antônio Vello (Esalq, USP), who kindly provided the experimental data used in this study; to the Brazilian agencies Capes and CNPq, for the fellowships to first and second authors, respectively.

### References

- BESAG, J.; KEMPTON, R. Statistical analysis of field experiments using neighbouring plots. *Biometrics*, v.42, p.231-251, 1986.
- CULLIS, B.R.; LILL, W.J.; FISHER, J.A.; READ, B.J. A new procedure for the analysis of early generation variety trials. *Applied Statistics*, v.38, p.361-375, 1989.
- DUARTE, J.B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento vegetal**. 2000. 293p. Tese (Doutorado) - Universidade de São Paulo, Piracicaba. <available on <http://www.teses.usp.br>>
- ES, H.M. van; ES, C.L. van. Spatial nature of randomization and its effect on the outcome of field experiments. *Agronomy Journal*, v.85, p.420-428, 1993.
- FEDERER, W.T. Augmented (or hoonuiaku) designs. *Hawaiian Planter's Records*, v.55, p.191-208, 1956.

- FEDERER, W.T. Recovery of interblock, intergradient, and intervarietal information in incomplete block and lattice rectangle designed experiments. **Biometrics**, v.54, p.471-481, 1998.
- GRONDONA, M.O.; CRESSIE, N. Using spatial considerations in the analysis of experiments. **Technometrics**, v.33, p.381-392, 1991.
- HENDERSON, C.R. **Applications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462p.
- KEMPTON, R.A.; GLEESON, A.C. Unreplicated trials. In: KEMPTON, R.A.; FOX, P.N. (Ed.). **Statistical methods for plant variety evaluation**. London: Chapman & Hall, 1997. cap.6, p.86-100.
- KEMPTON, R.A.; SERAPHIN, J.C.; SWORD, A.M. Statistical analysis of two-dimensional variation in variety yield trials. **Journal of Agricultural Science**, v.122, p.335-342, 1994.
- LITTELL, R.C.; MILLIKEN, G.A.; STROUP, W.W.; WOLFINGER, R.D. **SAS system for mixed models**. Cary, NC: SAS Institute, 1996. 633p.
- MARTÍNEZ, R. Control de la correlacion espacial en experimentos de campo en el sector agricola. **Agronomia Colombiana**, v.11, p.83-89, 1994.
- PANNATIER, Y. **Variowin**: Software for spatial data analysis in 2D. Lausanne: Springer, 1996. 91p.
- PONTES, J.M. **A geoestatística**: aplicações em experimentos de campo. 2002. 82p. Dissertação (Mestrado) - Universidade Federal de Lavras, Lavras.
- SCOTT, R.A.; MILLIKEN, G.A. A SAS program for analyzing augmented randomized complete-block designs. **Crop Science**, v.33, p.865-867, 1993.
- STROUP, W.W.; BAENZIGER, P.S.; MULITZE, D.K. Removing spatial variation from wheat yield trials: a comparison of methods. **Crop Science**, v.34, p.62-66, 1994.
- VIEIRA, S.R. Uso de geoestatística em estudos de variabilidade espacial de propriedades do solo. In: NOVAIS, R.F. (Ed.). **Tópicos em Ciência do Solo**. Viçosa: SBCS, 2000. p.1-54.
- WOLFINGER, R.D.; FEDERER, W.T.; CORDERO-BRANA, O. Recovering information in augmented designs, using SAS PROC GLM and PROC MIXED. **Agronomy Journal**, v.89, p.856-859, 1997.
- ZIMMERMAN, D.I.; HARVILLE, D.A. A random field approach to the analysis of field-plot experiments and other spatial experiments. **Biometrics**, v.47, p.223-239, 1991.

---

Received on March 25, 2004 and accepted on September 14, 2004