Research Article

Prediction of hybrid means from a partial circulant diallel table using the ordinary least square and the mixed model methods

Américo José dos Santos Reis, Lázaro José Chaves, João Batista Duarte and Edward Madureira Brasil

Escola de Agronomia e Engenharia de Alimentos, Universidade Federal de Goiás, Goiânia, GO, Brazil.

Abstract

By definition, the genetic effects obtained from a circulant diallel table are random. However, because of the methods of analysis, those effects have been considered as fixed. Two different statistical approaches were applied. One assumed the model to be fixed and obtained solutions through the ordinary least square (OLS) method. The other assumed a mixed model and estimated the fixed effects (BLUE) by generalized least squares (GLS) and the best linear unbiased predictor (BLUP) of the random effects. The goal of this study was to evaluate the consequences when considering these effects as fixed or random, using the coefficient of correlation between the responses of observed and non-observed hybrids. Crossings were made between S_1 inbred lines from two maize populations developed at Universidade Federal de Goiás, the UFG-Samambaia "Dent" and UFG-Samambaia "Flint". A circulant inter-group design was applied, and there were five ($s = 5$) crossings for each parent. The predictions were made using a reduced model. Diallels with different sizes of s (from 2 to 5) were simulated, and the coefficients of correlation were obtained using two different approaches for each size of s. In the first approach, the observed hybrids were included in both the estimation of the genetic parameters and the coefficient of correlation, while in the second a cross-validation process was employed. In this process, the set of hybrids was divided in two groups: one group, comprising 75% of the original group, to estimate the genetic parameters, and a second one, consisting of the remaining 25%, to validate the predictions. In all cases, a bootstrap process with 200 resamplings was used to generate the empirical distribution of the correlation coefficient. This coefficient showed a decrease as the value of s decreased. The cross-validation method allowed to estimate the bias magnitude in evaluating the correlation coefficient using the same hybrids, to predict the genetic parameters and the correlation evaluation. The bias was shown to be greater when the OLS method was used. When the correlation coefficients of the observed and estimated hybrid means were obtained through the mixed instead of the fixed model, this decrease was less marked. The selection of hybrids superior to the checks, in terms of grain weight, also differed in the two different approaches. Nineteen percent of the hybrids were shown to be superior to the checks in the fixed models, while only 1.8% of them were superior in the mixed model.

Key words: diallel analysis, BLUP, prediction, cross-validation.

Introduction

Nowadays, one of the major obstacles to corn breeding programs which aim to develop hybrids is the high cost of field evaluation. The strategy originally adopted in breeding programs was to perform all possible crossings in a group of inbred lines and then make an evaluation of the single hybrids obtained, followed by the selection of the most promising ones. However, as the breeding programs became larger, thousands of inbred lines became available. This made the development and evaluation of all possible hybrids extremely difficult, mainly because of the high cost of the assessment phase. So, there was an urgent need to develop procedures to allow the evaluation of a large number of inbred lines from a small sample of hybrids. The prediction of non-observed hybrid performance became possible through the use of genetic parameter estimation. Consequently, estimators or predictors have been sought, in order to maximize the correlation between estimated or predicted genetic values and parametric genetic values. Diallel tables have been one of the main tools for estimating genetic parameters, not only because they provide great amounts of information, but also because of the flexibility in constructing them. For predictive analysis, the scheme proposed by Kempthorne and Curnow (1961), based on a sample of all possible crossings between a group of parents, referred to

Send correspondence to L.J. Chaves. Universidade Federal de Goiás, Escola de Agronomia e Engenharia de Alimentos, Rodovia GO 080 km 0, Campus II, Caixa Postal 131, 74001-970 Goiânia, GO, Brazil. E-mail: lchaves@agro.ufg.br.

as a circulant diallel cross, is noteworthy. Miranda Filho and Vencovsky (1999), using Griffing's model (1956), and Reis (2000), using the model proposed by Gardner and Eberhart (1966), adjusted the circulant design to an interpopulation level. In order to achieve this, it was necessary to obtain the hybrid combinations *ps/2*, where *p* is the number of parents and *s* is the number of combining hybrids in each participating parent. In the second case, evaluation of the parents is also required. When comparing the complete diallel tables, the great reduction in the number of crossings is striking, mainly when there is an increase in the value of *p*.

One way of evaluating the predictive capacity of the model, that uses the estimates from a circulant diallel table, is by applying the Pearson correlation coefficient between the responses of the predicted and observed hybrids. Andrade (1995), using $s = 3$, found correlation coefficients varying from 0.82 to 0.96. On the other hand, Araújo (2000), using $s = 4$, found a correlation of 0.86, and Fuzatto (2003) observed correlations between 0.685 and 0.925, using values of *s* from 6 to 2. The correlation increased as the value of *s* decreased. All these authors evaluated ear weight. Gonçalves (1987), using $s = 3$, observed correlations from 0.92 to 0.86 related to the grain weight. In all these experiments, the genetic parameters (general and specific combining ability) were estimated by using the ordinary least squares (OLS) method.

On the other hand, in a circulant diallel table, there is an interest in extrapolating the information obtained about the observed hybrids to a reference population of non-observed hybrids [$(p/2)^2$-*ps/2*]. As emphasized by Searle *et al.* (1992), the main issue is to quantify the performance of a non-realized random variable (non-observed hybrids), given an observation vector (realized observation). Therefore, in this context, according to Henderson (1986), the use of BLUP (Best Linear Unbiased Predictor) would be the most appropriate method to predict the genetic parameters. The use of BLUP in plant breeding has also been advocated

by Bernardo (1994, 1995, 1996a, 1996b). In this particular case, the error variance and the other variance components will influence the genetic parameter estimation, making it possible to obtain the BLUE (Best Linear Unbiased Estimator) for fixed effects and the BLUP for random effects, which is the appropriate approach for mixed linear models (Henderson, 1984). In this method, the known covariances will be considered not only in the statistical tests, but also in the assessment and prediction of effects which directly influence the selection of the inbred lines. In general, the corresponding estimators have lower variances than the ones obtained through the OLS, thus resulting in more reliable estimation (Duarte and Vencovsky, 2001). André (1999) concluded that the BLUP provides better accuracy than the OLS estimators in predicting the general combining ability effects in different conditions of heritability. Besides being possible when the information about co-ancestry between the inbred lines is available, it is also possible to consider the additive effects, the dominant effects and the epistatic interactions. The main restriction found to the use of this approach is its great computational requirement, which no longer represents an obstacle.

The purpose of this paper was to evaluate the efficiency of the mixed linear models methodology in analyzing a partial circulant table, with varying sizes of *s*. This evaluation was performed mainly by correlating the predicted and the observed values of the hybrids.

Material and Methods

Two groups of parents, 34 flint maize inbred lines S_1 and 34 dent maize inbred lines S_1, randomly sampled from two populations, the UFG-Samambaia flint and the UFG-Samambaia dent, were used as the experimental material. These populations were developed at Universidade Federal de Goiás (EA-UFG). The crossings were performed according to a partial circulant diallel design, with five crosses for each parent ($s = 5$) (Table 1), where 165 out of

Table 1 - Scheme of an interpopulational circulant diallel table, with $p = 34$ and $s = 5$.

Flint inbred lines	Dent inbred lines									
	1	2	3	4	5	6	7	...	33	34
1	$y_{1.1}$	$y_{1.2}$	$y_{1.3}$	$y_{1.4}$	$y_{1.5}$					
2		$y_{2.2}$	$y_{2.3}$	$y_{2.4}$	$y_{2.5}$	$y_{2.6}$				
3			$y_{3.3}$	$y_{3.4}$	$y_{3.5}$	$y_{3.6}$	$y_{3.7}$			
4				$y_{4.4}$	$y_{4.5}$	$y_{4.6}$	$y_{4.7}$	...		
5					$y_{5.5}$	$y_{5.6}$	$y_{5.7}$	...		
6						$y_{6.6}$	$y_{6.7}$	...		
7							$y_{7.7}$	...		
⋮										
33	$y_{33.1}$	$y_{33.2}$	$y_{33.3}$						$y_{33.33}$	$y_{33.34}$
34	$y_{34.1}$	$y_{34.2}$	$y_{34.3}$	$y_{34.4}$						$y_{34.34}$

170 originally predicted hybrids were obtained, representing the reference population for the 1156 possible hybrids between these two inbred line groups. These hybrids were evaluated through a randomized complete block design, with four replications. The experimental plots were represented by single rows 5 m long spaced 0.9 m apart, with 25 plants per plot after thinning. The triple hybrid BR-3123 was used as a check, and planting was done on January 6, 1999, in the experimental area at the EA-UFG.

Griffing's model (1956) was adopted to describe the observations of the diallel table:

$$y_{ij} = \mu + g_i + g_j + s_{ij} + \varepsilon_{ij} \tag{1}$$

where: y_{ij} is the phenotypic value of the hybrids between the dent line i ($i = 1, 2, ..., I$) and the flint line j ($j = 1, 2, ..., J$); μ is the mean common to the observations; g_i is the general combining ability effect of the ith parent from the dent group, assumed to be random and with the distribution $N \sim (0, \sigma^2_{LD})$; g_j is the general combining ability effect of the jth parent from the flint group, assumed to be random and with the distribution $N \sim (0, \sigma^2_{LF})$; s_{ij} is the specific combining ability effect resulting from the crossing between the parents i and j, assumed to be random and with the distribution $N \sim (0, \sigma^2_{CEC})$; and ε_{ij} is the random error effect with the distribution $N \sim (0, \sigma^2)$.

Fixed model

In the matrix form, the hybrid means can be represented by:

$$\bar{y} = X\beta + \varepsilon \tag{2}$$

where $\bar{y}$ is the mean treatment vector, X is the incidence matrix of the genetic effects, β is the parametric vector, and is the error vector. As X is an incomplete rank column matrix, $X'X$ is singular, not having a single inverse. Therefore, in order to solve the system of normal equations and to obtain single solutions, the following parametric restrictions were adopted: $\sum_{i=1}^{I} g_i = 0, \sum_{j=1}^{J} g_j = 0, \sum_{i=1}^{I} s_{ij} = 0$ and $\sum_{j=1}^{J} s_{ij} = 0$.

Thus, the OLS solutions are given by:

$$\hat{\beta} = (X'X)^{-1}(X'\bar{y}). \tag{3}$$

Mixed linear model

The individual observations can be expressed in the matrix form as follows:

$$y = X\theta + Z_1 a_X + Z_2 a_Y + Z d + \varepsilon \tag{4}$$

where: y is the observation vector; θ is the vector of fixed effects, which here includes the general mean and the block effect; a_X is the vector for the general combining ability of the dent inbred lines; a_Y is the vector for the general combining ability of the flint inbred lines; d is the vector for the specific combing ability; ε is the error vector; and X, Z_1, Z_2 and Z are the incidence matrices for vectors θ, a_X, a_Y, and d, respectively.

In this case, applying generalized least squares (GLS) to calculate the fixed effects and the best linear unbiased prediction for the random effects, as proposed by Henderson (1984), the solutions of the mixed model equations can be obtained by:

$$\begin{bmatrix} \hat{\theta} \\ \hat{a}_X \\ \hat{a}_Y \\ \hat{d} \end{bmatrix} = \begin{bmatrix} X'X & X'Z_1 & X'Z_2 & X'Z \\ Z_1'X & Z_1'Z_1 + A_1^{-1}\gamma_1 & Z_1'Z_2 & Z_1'Z \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + A_2^{-1}\gamma_2 & Z_2'Z \\ Z'X & Z'Z_1 & Z'Z_2 & Z'Z + D^{-1}\gamma_3 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \\ Z'y \end{bmatrix} \tag{5}$$

or:

$$\begin{bmatrix} \hat{\theta} \\ \hat{a}_X \\ \hat{a}_Y \\ \hat{d} \end{bmatrix} = \begin{bmatrix} C_{00} & C_{01} & C_{02} & C_{03} \\ C_{10} & C_{11} & C_{12} & C_{13} \\ C_{20} & C_{21} & C_{22} & C_{23} \\ C_{30} & C_{31} & C_{32} & C_{33} \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \\ Z'y \end{bmatrix}$$

with:

$$\gamma_1 = \frac{\sigma^2}{V_X}, \gamma_2 = \frac{\sigma^2}{V_Y}, \text{ and } \gamma_3 = \frac{\sigma^2}{V_{CEC}}.$$

Using the expectation maximization-restricted maximum likelihood (EM-REML) algorithm (Dempster *et al.*, 1977) to obtain the solution of this system, the variance component estimators are given by:

$$\sigma^2 = \frac{y'y - \hat{\theta}'X'y - \hat{a}'_X Z_1'y - \hat{a}'_Y Z_2'y - \hat{d}'Z'y}{n - r(X)},$$

$$\hat{V}_X = \frac{\hat{a}'_X A_1^{-1}\hat{a}_X + Tr(A_1^{-1}C_{11})\hat{\sigma}^2}{p_1},$$

$$\hat{V}_Y = \frac{\hat{a}'_Y A_2^{-1}\hat{a}_Y + Tr(A_2^{-1}C_{22})\hat{\sigma}^2}{p_2}, \text{ and}$$

$$\hat{V}_{CEC} = \frac{\hat{d}'D^{-1}\hat{d} + Tr(D^{-1}C_{33})\hat{\sigma}^2}{(p \times s)/2} \tag{6}$$

where: p_1, p_2, p, and s are the numbers of flint inbred lines, dent inbred lines, total number of inbred lines and the number of crosses for each inbred line, respectively. In (6), $r(X)$ is the rank of X, and Tr is the trace operation. As the inbred lines were considered unrelated and since the two groups

are not related to each other, the matrices A_1, A_2 and D are identity matrices. When assuming the existence of co-ancestry between the parents, matrices A_1 and A_2 will present values equal to 1.0 in the diagonal and the co-ancestry coefficient between parents off the diagonal. Thus the diagonal of matrix D is also composed of values equal to 1.0, and the off-diagonal values are the products of the co-ancestry coefficients between the parents.

An interactive process was conducted, in accordance with (6) and (5), until a convergence was obtained, attributing an initial randomized value to the variance components. As only the estimates of the variance components, and not their parametric values, were known, the EBLUP (Empirical Best Linear Unbiased Predictor) was obtained from the random effects (Littel *et al.*, 1996). However, for the selection based on isolated traits, the rank of the candidates for selection is not as influenced by errors in the estimation of variance components (Resende, 2002), when there are balanced data and when only one population is considered (Duarte and Vencovsky, 2001).

The diallels were simulated to evaluate the models' goodness of fit and the way that correlation is obtained, in order to make the predictions of non-observed hybrids. The *s* sizes ranged from 5 to 2, and the correlation was made in two different ways. First, the correlation was obtained by using the sample of observed hybrids to calculate both the parameters and the coefficient of correlation. Second, a cross-validation procedure was applied to the original data set. This set was randomly divided into two groups: one, constituting 75% of the original set, was used to estimate the genetic parameters, and the other, composed of the remaining observed hybrids (25%), was used to validate the predictions. In all cases, a process of 200 resamplings with replacement (bootstrap) was employed to generate empirical distributions of the correlation coefficient estimates.

The OLS was used to estimate the genetic parameters from (3), and to obtain the best linear unbiased prediction, as done in (5). All analyses were performed using the SAS® PROC IML system (Little *et al.*, 1996). The program used to calculate the EBLUP was adapted from André (1999):

```
options no date no number ps = 1000 ls = 90;
data;
input efix pi pj cruz y;
cards;
;
proc iml;
START REML;
n1 = nrow(x1); /*counts the number of observations*/
n = n1;
ncru = unique(p3);
ncru1 = union(ncru, ncru)`;
ncrut = nrow(ncru1); /*counts the number of crosses */
z1 = design(p1); /* incidence matrix of p1*/
z2 = design(p2); /* incidence matrix of p2*/
z3 = design(p3); /* incidence matrix of p3*/

ng1 = ncol(z1);/*counts the number of genitors in group I*/
ng2 = ncol(z2);/* counts the number of genitors in group II*/
x = design(x1); /* incidence matrix of the fixed effects*/

nf = ncol(x); /*counts the number of fixed effects*/
z = z1||z2||z3; /*incidence matrix of the random effects*/
k = x||z1||z2||z3; /*overall incidence matrix*/

vgI = 558; /* initial value of the CGC variance for the dent group */
vgII = 647; /* initial value of the CGC variance for the flint group*/
vs = 83; /* initial value of the CEC variance*/
r = 118; /* initial value of the error variance*/

iter = 0; /*initializes the interaction counting */
maxiter = 2000; /*defines the maximum interactions*/
conv = 1E-8; /*defines the criterion of convergence */
do until (dif );

g1 = I(ng1); /* parentage matrix A associated to the flint group */
g2 = I(ng2); /* parentage matrix A associated to the dent group */
g3 = I(ncrut); /* parentage matrix A associated to sij*/
g = block((g1)/vgI,(g2)/vgII,(g3)/vs); /*defines the inverse matrix G */

t = inv(((x`*x)||(x`*z))//((z`*x)||(z`*z+g*r)));
/* concatenates and inverts the t matrix*/
solution = t*(x`*y//z`*y); /*solution of the system*/

u = solution[nf:nf]; /*solution of the fixed effect*/
si = solution[nf+1:ng1+1]; /*solution of the CGC GI*/
sj = solution[ng1+2:ng1+ng2+1]; /*solution of the CGC GII*/
sij = solution[ng1+ng2+2:ng1+ng2+ncrut+1];/*solution for sij*/

r1 = (y`*y-solution`*k`*y)/(n-nf);/*component of the residual variance 1*/
vgI1 = (si`*g1*si+trace(g1*(t[1:ng1,1:ng1]))*r)/(ng1);/*component of the CGC variance of the flint group*/
vgII1 = (sj`*g2*sj+trace(g2*(t[ng1+1:ng1+ng2,ng1+1:ng1+ng2]))*r)/(ng2);
/*component of the CGC variance of the dent group */
vs1 = (sij`*g3*sij+trace(g3*(t[ng1+ng2+1:ng1+ng2+ncrut,ng1+ng2+1:ng1+ng2+ncrut]))*r)/(ncrut);/*component of the sij variance*/

dif = max(abs(r1-r), abs(vgI1-vgI), abs(vgII1-vgII));
/*calculation of the differences in estimation s*/
if iter = maxiter then dif = 0;
/*condition for the maximum of interactions */
if iter then do;
var_CGCI = vgI;
var_CGCII = vgII;
var_sij = vs;
erro = r;
difer = dif;
n_iter = iter;
end;
else do;
var_CGCI = var_CGCI//vgI;
var_CGCII = var_CGCII//vgII;
var_sij = var_sij//vs;
erro = erro//r;
difer = difer//dif;
n_iter = n_iter//iter;
end;
iter = iter+1;
vgI = vgI1;
vgII = vgII1;
 vs = vs1;
r = r1;
  end;
print , 'VARIANCE COMPONENT ESTIMATIONS (REML)',
n_iter [format = 3.]
var_CGCI [format = 12.6]
var_CGCII [format = 12.6]
```

```
var_sij [format = 12.6]
error [format = 12.6]
difer [format = 12.8];

vectoru = solution[1:nf];
vectorCGCI = solution[nf+1:ng1+1];
vectorCGCII = solution[ng1+2:ng1+ng2+1];
vectorsij = solution[ng1+ng2+2:ng1+ng2+ncrut+1];
crossings = (1:ncrut)`;
genitorsg1 = (1:ng1)`;
genitorsg2 = (1:ng2)`;

print ,'EBLUP ESTIMATIONS OBTAINED BY THE MIXED
MODELS ',
vectoru [format = 12.2] vectorCGCI [format = 12.3] vectorCGCII [for-
mat = 12.3] vectorsij [format = 12.3];

zu = x*vector u;
zvi = z1c*vector CGCI;
zvj = z2c*vector CGCII;
zsij = z3*vector sij;

predictions MM = zu+zvi+zvj+zsij;
print ,'PREDICTIONS OF HYBRIDS',
predictions MM [format = 12.3];
FINISH;
use data;
read all into data;
x1 = data(|,1|); /*defines the mean vector */
p1 = data(|,2|); /*defines the vector of the CGC effects for the flint
group*/
p2 = data(|,3|); /*defines the vector of the CGC effects for the dent
group*/
p3 = data(|,4|); /*defines the vector of the CEC effects*/
y = data(|,5|); /*defines the vector of observations*/
RUN REML;
quit;
```

Results and Discussion

An increase was found in the correlation coefficient as the value of s was reduced, when the observed hybrids were used both in the estimation of the genetic parameters and of the correlation coefficient. However, the standard deviations associated with those estimates increased as s decreased. The value of the correlation coefficient decreased from 0.916 ± 0.0727 ($s = 2$) to 0.742 ± 0.0090 ($s = 5$), using the OLS, and from 0.851 ± 0.0217 ($s = 2$) to 0.733 ± 0.0049 ($s = 5$), when the mixed model was applied (Table 2).

The opposite results were found when cross-validation was employed, that is, the value of the correlation coefficient increased as the value of s increased. Likewise, the related standard deviation decreased with the increase of s. When the OLS method was applied, the correlation coefficient varied from 0.260 ± 0.1217 ($s = 5$) to 0.100 ± 0.2441 ($s = 2$), while with the mixed model the variation ranged from 0.370 ± 0.1063 ($s = 5$) to 0.120 ± 0.2278 ($s = 2$) (Table 3). It is interesting to note that the greatest mean values of the correlation coefficient ($r = 0.41$) were obtained when using $s = 4$, in the analysis made through the mixed model. The empirical distributions of correlation co-

Table 2 - Mean values ($\hat{R}$), associated standard deviations (DP), and maximum and minimum values for the correlation coefficient between the predicted and observed hybrid means, under different sizes of s[1].

Parameters	$s = 5$	$s = 4$	$s = 3$	$s = 2$
	OLS			
$\hat{R}$	0.742^{**}	0.795^{**}	0.801^{**}	0.916^{**}
DP	0.009	0.014	0.034	0.073
Maximum	0.757	0.820	0.852	0.990
Minimum	0.708	0.759	0.670	0.658
	Mixed model			
$\hat{R}$	0.733^{**}	0.783^{**}	0.798^{**}	0.851^{**}
DP	0.005	0.007	0.012	0.022
Maximum	0.743	0.792	0.820	0.871
Minimum	0.717	0.755	0.750	0.779

[1]Observed hybrid means for both the estimation of parameters and the calculation of the correlation coefficient, using 200 resamplings with reposition.

Table 3 - Mean values ($\hat{R}$), associated standard deviations (DP), and maximum and minimum values for the correlation coefficient between the predicted and observed hybrid means using different sizes of s, obtained by a cross-validation and 200 resamplings with reposition.

Parameters	$s = 5$	$s = 4$	$s = 3$	$s = 2$
	OLS			
$\hat{R}$	0.260^{**}	0.240^{**}	0.110^{ns}	0.100^{ns}
DP	0.122	0.169	0.183	0.244
Maximum	0.550	0.620	0.620	0.620
Minimum	-0.200	-0.180	-0.440	-0.590
	Mixed model			
$\hat{R}$	0.370^{**}	0.410^{**}	0.300^{**}	0.120^{ns}
DP	0.106	0.114	0.162	0.228
Maximum	0.650	0.710	0.670	0.720
Minimum	-0.040	0.090	-0.270	-0.690

efficient estimates for each case are shown in Figure 1. It is important to highlight that the maximum theoretical limit of this correlation is not 1.0, but the square root of the heritability coefficient (Vencovsky and Barriga, 1992). In the present work, this limit was equal to 0.734, which does not seem so unrealistic when compared with the correlations found by Bernardo (1996a). This author evaluated 4099 hybrids among several heterotic groups, using the mixed model method associated with the co-ancestry data between the parents. In his experiment, the correlations ranged from 0.136 to 0.762, with theoretical maximum limits of 0.554 and 0.864, respectively.

It is relevant to emphasize that the increase observed in the correlation coefficients, when the size of s is decreased, does not mean that the lower values of s allow better predictions. As stated by Gauch and Zobel (1988), it
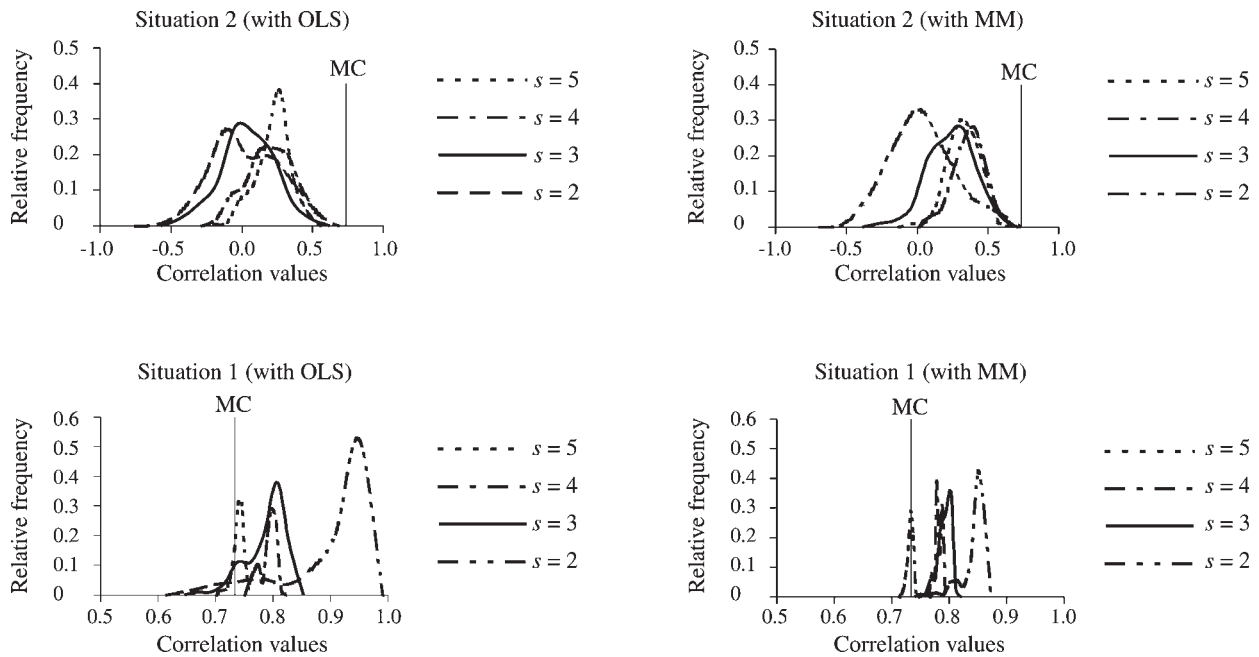
Figure 1 - Empiric distributions of the correlation coefficients between the predicted and observed hybrid means, considering the two following situations: Situation 1: the observed hybrid means are included in the estimation of the genetic parameters and in the calculation of the correlation coefficient. Situation 2: the observed hybrid means were assigned to two different groups, where 75% of them were used to calculate the parameters, and the remaining 25% to estimate the correlation coefficient (MC is the maximum limit of the correlation coefficient, OLS is the ordinary least square estimation, and MM is the mixed model approach).

means that the correlation is measuring the postdictive ability of the model, that is, with the decrease in s, the model can better explain the observed data. Moreover, when the correlation coefficients were evaluated through a cross-validation process, an increase of the correlation coefficients was observed whenever the values of s were increased. In this situation, not only the ability of the model to predict non-realized observations is evaluated, but also its ability to describe the set of observed data. Thus, it is possible to assess the predictive ability of the model by approaching its predictions to the data not included in the analysis, simulating future responses that have not been measured yet.

It is thus clear that a reduction in the value of s also decreases the predictive potential of the model. Furthermore, the correlation coefficient calculated through the observed hybrids, during both the assessment of $\hat{g}_i$ and of $\hat{g}_j$ and the model's validation, yields bias. This bias can be calculated assuming that the average correlation coefficient obtained by using the cross-validation is the parametric value for each value of s. In this case, an increase in the bias of the correlation coefficient estimate can be observed when the value of s decreases (Table 4). The bias is of greater magnitude when the OLS method is employed in the analysis, ranging from 0.482 with $s = 5$ to 0.816 with $s = 2$. Using the mixed models, the value found was 0.363 with $s = 5$ and 0.731 with $s = 2$. Another indicator of this bias can be observed in Figure 1, where the distributions

Table 4 - Bias of the correlation coefficient for different sizes of s under two statistical analysis models (ordinary least square - OLS and mixed models).

Values of	Mixed model	OLS
$s = 5$	0.363	0.482
$s = 4$	0.373	0.555
$s = 3$	0.498	0.691
$s = 2$	0.731	0.816

obtained using cross-validation do not exceed the maximum theoretical limit of the correlation coefficient (MC). However, this is not true for the first situation.

If hybrids with yield mean superior to the check mean were to be selected, considering all possible hybrids in the diallel table (1156 hybrids), and if the prediction was made through OLS analysis with $s = 5$, 19% of the hybrids would be selected. By using a mixed model, only 1.8% would be selected. However, Spearman's correlation coefficient between the ranks of hybrid means by the two analyses was equal to 0.95.

The use of the mixed model approach was more efficient than the OLS in the operation and management of this data set, resulting in more accurate estimates of correlation coefficients between observed and non-observed hybrids. Values of $s < 4$ have yielded poorer predictions for both the mixed model and the OLS analysis. The use of the same data set to estimate the parameters and to evaluate the

model does not permit inferences about future responses of the hybrids.

The circulant crossing method, with $s = 5$ and $s = 4$, associated with the methodology of mixed models, allowed to predict non-observed hybrid means, and showed good reliability, which is very important in the initial stages of the evaluation of inbred lines.

References

Andrade JAC (1995) Dialelo parcial circulante interpopulacional em milho (*Zea mays* L.) com dois níveis de endogamia dos parentais. Tese de Doutoramento, Universidade de São Paulo, Piracicaba, SP.

André CMG (1999) Avaliação da melhor predição linear não tendenciosa (BLUP) associada ao uso de marcadores moleculares na análise dialélica. Dissertação de Mestrado, Universidade Federal de Lavras, Lavras.

Araújo PM (2000) Dialelo parcial circulante interpopulacional e cruzamentos "top-crosses" na avaliação de linhagens parcialmente endogâmicas de milho (*Zea mays* L.). Tese de Doutoramento, Universidade de São Paulo, Piracicaba, SP.

Bernardo R (1994) Prediction of maize single-cross performance using RLFPs and information from related hybrids. Crop Sci 1:20-25.

Bernardo R (1995) Genetic model for predicting maize performance in unbalanced yield trial data. Crop Sci 1:141-147.

Bernardo R (1996a) Best linear unbiased prediction of maize single-cross performance. Crop Sci 1:50-56.

Bernardo R (1996b) Best linear unbiased prediction of the performance of crosses between untested maize inbreds. Crop Sci 4:872-876.

Dempster AP, Laird NM and Rubim DB (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Ser. B 1:1-38.

Duarte JB and Vencosky R (2001) Estimação e predição por modelo linear misto com ênfase no ordenamento de médias de tratamentos genéticos. Sci Agri 1:109-117.

Fuzatto SR (2003) Dialelo parcial circulante interpopulacional em milho (*Zea mays* L.): Efeito do número (s) de cruzamentos.

Tese de Doutoramento, Universidade de São Paulo, Piracicaba, SP.

Gardner CO and Eberhart SA (1966) Analysis and interpretation of the variety cross diallel and related populations. Biometrics 22:439-452.

Gauch HG and Zobel RW (1988) Predictive and postdictive success of statistical analysis of field trial. TAG 76:1-10.

Gonçalves PS (1987) Esquema circulante de cruzamentos para avaliação de linhagens de milho (*Zea mays* L.) ao nível interpopulacional. Tese de Doutoramento, Universidade de São Paulo, Piracicaba, SP.

Griffing B (1956) Concept of general and specific combining ability in relation to diallel crossing systems. Australian Journal of Biological Science 9:463-493.

Henderson CR (1984) Applications of Linear Models in Animal Breeding. University of Guelph, Ontario, 462 pp.

Henderson CR (1986) Recent developments in variance and covariance estimation. J Anim Sci 63:208-216.

Kempthorne O and Curnow RN (1961) The partial diallel crosses. Biometrics 17:229-250.

Littel RC, Mccuthan BG, Stroup WW and Wolfinger RD (1996) SAS® System for Mixed Models. Cary, SAS Institute Inc., 633 pp.

Miranda Filho JB and Vencovsky R (1999) The partial circulant diallel crosses at the interpopulation level. Genetics and Molecular Biology 4:249-255.

Reis AJS (2000) Estudos de componentes de heterose ao nível interpopulacional a partir de cruzamentos dialélicos parciais circulantes. Dissertação de Mestrado, Universidade Federal de Goiás, Goiânia, GO.

Resende MDV (2002) Procedimentos ótimos de seleção com dados balanceados e desbalanceados. In: Resende, MDV (ed) Genética Biométrica e Estatística no Melhoramento de Plantas Perenes. Embrapa Informação Tecnológica, Brasília, pp 209-345.

Searle SR, Casella G and Mcculloch CE (1992) Variance components. John Wiley & Sons, New York, 501 pp.

Vencovsky R and Barriga P (1992) Genética Biométrica no Fitomelhoramento. Sociedade Brasileira de Genética, Ribeirão Preto, 496 pp.

Associate Editor: José Francisco Ferraz de Toledo