



Prediction of genotypic values of maize hybrids in unbalanced experiments

Roberto Fritsche-Neto¹, Manoel Carlos Gonçalves¹, Roland Vencovsky¹, and Cláudio Lopes de Souza Junior^{1*}

Received 12 June 2009

Accepted 15 October 2009

ABSTRACT- *The objective of this study was to evaluate whether the REML/BLUP can be useful for predicting the genotypic values of maize hybrids in a group of unbalanced experiments. A set of 256 single-crosses were evaluated in 13 environments for grain yield, plant height and plant lodging. Sets of hybrids within environments and sets of environments were withdrawn from the experiments to simulate unbalanced data, and the hybrid predictions of the unbalanced data were computed by the REML/BLUP, simulated using the bootstrap resampling procedure. The coefficients of determination and percentage of selection coincidence were computed for the predicted genotypic values of unbalanced data and their means from the balanced data. The REML/BLUP method accurately predicted the genotypic values of missing hybrids under losses of up to 20% of hybrids within environments or a reduction of 23% of the environments, even in the presence of significant and complex hybrid x environment interaction.*

Key words: BLUP, *Zea mays* L., maize hybrid, selection.

INTRODUCTION

The most difficult and resource-demanding phase in a maize breeding program is the experimental evaluation of the hybrids, because usually a large number of genotypes has to be assessed in several environments. Thus, the number of hybrids as well as of environments is limited by the restricted resources, which can lead to a reduction in the number of environments. Sets of hybrids are therefore usually evaluated in different environments, resulting in unbalanced comparisons of the genotypes (Panter and Allen 1995).

Since the 1930s, several methodologies of genetic evaluation have been proposed, among these, the least squares for unbalanced data (Yates 1934). The application of this method is not free of problems, since the variance of the prediction error is minimal, the

functions of the prediction are not always estimable and, depending on the degree of data unbalancing, the values of some genotypes may be under- or overestimated (Henderson 1974).

The equations of mixed models described by Henderson (1963) introduced changes in the estimation of variance components and breeding values (Searle 1971). The method consists basically in the prediction of genetic values, considered random, adjusting the data simultaneously to the fixed effects, to the unequal number of data in the subclasses and to coefficients of relatedness of genotypes (Bernardo 2002). The method proposed by Henderson (1963) therefore has flexible properties, since it can be applied to unbalanced data of different generations and to estimate the breeding values of unevaluated genotypes (Henderson 1984, Piepho et al. 2008).

¹ Departamento de Genética, ESALQ/USP, CP 83, 13400-970, Piracicaba, SP, Brazil. *E-mail: clsouza@esalq.usp.br

Mixed models based on the statistical method REML/BLUP (Restricted Maximum Likelihood / Best Linear Unbiased Prediction) are widely used in animal breeding. However, in plant breeding programs their use has to date been restricted to perennial crops where unbalanced experiments are common. There is little information in the literature about the reliability of REML/BLUP for the prediction of genotypic values in unbalanced experiments for breeding programs of annual crops (Bernardo 2002). The objective of this study was to evaluate whether the REML/BLUP method can be useful for predicting the genotypic values of maize single-crosses in a group of unbalanced experiments.

MATERIAL AND METHODS

Genetic material and experimental development

For this study 250 maize single-crosses (hybrids) from the crosses of five testers with 50 inbred lines were used. The inbred lines were from different populations and commercial hybrids, and had been developed by the Department of Genetics of ESALQ / USP. The 250 single-crosses and six commercial single-crosses (Z-8420, Z-8550, Z-8460 and Z-8480, BRS-2020, and BRS-1010), were evaluated in experiments using a 16 x 16 lattice design, with two replications. Each plot consisted of a 4-m row; the rows were spaced 0.8 m and plants 0.2 m apart, corresponding to a stand of 62,500 plants ha⁻¹.

The experiments were assessed in 13 environments, where each environment was represented by a combination of location - year. With a view to evaluate the hybrids in different environments, different planting dates and locations were used. The environments were: Experimental Station (E.St.) Areão, E.St. Caterpillar and E.St. Department of Genetics of the Escola Superior de Agricultura "Luiz de Queiroz"/ Universidade de São Paulo (ESALQ/USP), all in Piracicaba (lat 22° 43' S, long 47° 36' W, alt 547 m asl), state of São Paulo in the 2002/2003, 2003/2004 and 2004/2005 growing seasons; E.St. Anhembi / SP (lat 22° 47' S, long 48° 07' W, alt 480 m asl) in the growing seasons 2003/2004 and 2004/2005, and two experiments were conducted at the E.St. Patos de Minas of the company Biomatrix, in Patos de Minas/MG (lat 18° 35' S; long 46° 31' W; alt 832 m asl) in the 2004/2005 growing season. For soil preparation, planting and cultural practices, the

technical recommendations for each environment were implemented. At the locations E.St. Depto. Genetics, Anhembi and Patos de Minas, the experiments were irrigated as required.

Data of the final stand (ST), plant height (PH), plant lodging (PL), grain moisture content (HUM) and grain yield (GY) were collected. The GY data, measured in kg plot⁻¹, were converted into t ha⁻¹, and corrected to 15% moisture and adjusted by the final plot stand by covariance analysis. The data of PL expressed in percentage (X%) were transformed to $PL = \sqrt{X\% + 0.5}$. The traits GY and PL were evaluated in all environments, while PH was evaluated in all but E.St. Areão and E.St. Caterpillar in the 2004/2005 growing season. These traits were chosen due to their differences in genetic complexity, heritability coefficient (h^2), coefficients of experimental variation and the type of data distribution they normally have. We thereby tried to represent the majority of the traits usually considered in species breeding.

Genetic-Statistical analysis

To predict the genotypic values of the hybrids, considering different percentages of loss of hybrids within environments or of environments, the REML/BLUP method was used. In each of these situations six variations were considered: balanced data (256 hybrids and 13 environments) or loss of 10%, 20%, 30%, 40%, or 50% of the assessed hybrids, corresponding to the loss of 26, 51, 77, 102, or 128 hybrids in each environment, or the elimination of 15%, 23%, 30%, 38%, or 46% of the environments, which corresponds to the elimination of two, three, four, five, or six environments. The simulation of data loss or exclusion was simulated 1,000 times, at random, by bootstrap resampling. On this basis, combined analyses were carried out for the two situations of imbalance and their respective variations, using the mixed model equations, by the method proposed by Bernardo (2002):

I. For the situation of hybrid losses in environments and of balanced data, the following model was used : $y = Xl + Zg + Wb + Tr + e$

y is the $n \times 1$ vector of phenotypic data, where n is the number of observations ranging from 3,328 (50% loss of hybrids) to 6,656 (balanced data), X is the $n \times 13$ incidence matrix of environments; l is the 13 x 1 vector of environmental effects (fixed); Z is the $n \times 256$ incidence matrix of the genotypic effects; g is the 256 x 1 vector of genotypic effects (random), where $G = A\sigma_g^2$,

and where A is the matrix of coefficients of relatedness and σ_g^2 is the genetic variance among hybrids; W is the $n \times 16$ incidence matrix of the block effects within replications; b is the 16×1 vector of the effects of blocks within replications (random); T is the $n \times 2$ incidence matrix of replication effects; r is the 2×1 vector of the replication effects (random); and e is the $n \times 1$ vector of errors, where $e \sim N(O, R)$, $R = I\sigma_e^2$, I is an $n \times n$ identity matrix and σ_e^2 is the variance of the experimental error. II. For the case of environment elimination, the following model was used:

$$y = Xr + Zg + Wb + Tl + e$$

y is the $n \times 1$ vector of phenotypic data, where n is the number of observations, ranging from 3,584 (54% of the environments) 6,656 (all environments), X is the $n \times 2$ incidence matrix of replications; r is the 2×1 vector of the effects of replications (fixed); Z is the $n \times 256$ incidence matrix of the genotypic effects; g is the 256×1 vector of genotypic effects (random), where, $g \sim N(O, G)$; W is the $n \times 16$ incidence matrix of the effects of blocks within replications; b is the 16×1 vector of the block effects within replications (random); T is the $n \times 13$ incidence matrix; l is the 13×1 vector of environmental effects (random); e is the $n \times 1$ vector of errors; $e \sim N(O, R)$ and $R = I\sigma_e^2$.

In the approach of mixed models, G represents the genetic covariance matrix of the hybrids, and is denoted $A\sigma_g^2$. In this study, the coefficients of relatedness (A) was ignored and therefore matrix G was assumed to be equal to $I\sigma_g^2$, that is, $A=I$. Consequently, σ_g^2 represents the genetic variance between the hybrids and vector \hat{g} corresponds to the predictions of the genotypic values of the hybrids (BLUP).

Different models were applied in I and II because the nature of the model effects changed according to the situation considered. In the first, all environments were included, to recover the information of the hybrids lost in the environments, and in the second, a sample of the environments was used, in which the experiments within them were balanced. To obtain the solutions, the components of genetic and non-genetic variance were assumed as unknown, as observed in practice, and estimated by the restricted maximum likelihood (REML). This method is an iterative process, for which the numerical algorithm EM was used, with alternating steps of expectation and maximization, characterized as EM-REML. Thus, based on initial arbitrary values of $\hat{\sigma}_g^2$, $\hat{\sigma}_b^2$ and of $\hat{\sigma}_r^2$, in situation I, and of $\hat{\sigma}_g^2$, $\hat{\sigma}_b^2$ and of $\hat{\sigma}_l^2$ in

situation II, the solutions \hat{g} , \hat{b} and \hat{r} , were obtained in I, and \hat{g} , \hat{b} and \hat{l} in II. These results are used to obtain new estimates of variance components and so on, until convergence is achieved.

To identify possible changes in the hybrid ranking, according to the environments and the magnitude of their effects, the mean square of the hybrid \times environment interaction was decomposed in its simple and complex components for GY, according to the method described by Robertson (1959).

Assessment of reliability of REML/BLUP

To verify the reliability of REML/BLUP for the prediction of genotypic values of maize hybrids, the correlation coefficients of the genotypic values predicted by REML/BLUP in the situation of balanced data and of those found in the other situations of data loss were calculated. For this purpose, Pearson's correlation coefficient was used in the analysis of the variables GY and PH, and for PL, which does not have a normal distribution, Spearman's correlation coefficient. This resulted in 1,000 r values for each unbalanced data set considered. Based on these results, the mean r values and their confidence intervals were estimated and subsequently squared to compute the coefficient of determination (R^2); their significances ($H_0: \rho=0$) was evaluated by the F test (Steel and Torrie 1980).

The hybrid ranking in the different imbalance situations was also compared, performing truncation selection of 5%, 10%, 15% and 20% of the superior hybrids, corresponding, respectively, to the selection of 13, 26, 39, and 52 hybrids. The selection aimed at higher grain yield, less lodging and reduced plant height. The percentage of selection coincidence was calculated based on the number of hybrids in common, selected in the situation of balanced data and in each of the variations of imbalance, resulting in 1000 percentage values of coincidence. These were used to compute the mean percentages and their confidence intervals, estimated at 95% confidence. The upper and lower limits of these intervals, be it for the correlation coefficients or for the percentage of selection coincidence, corresponded to the 25th and 975th positions, respectively, in the distribution of bootstrap values. All statistical analyses were performed using the software package Statistical Analysis System (SAS) version 9.1 (SAS Institute 2003).

RESULTS AND DISCUSSION

For the traits grain yield, plant lodging and plant height, the ratios between the highest and lowest variances of the experimental errors were 4.44, 3.21 and 1.88, respectively. The variance of experimental error was therefore considered homogeneous (Pimentel Gomes 2000). Significant differences ($p \leq 0.01$) were detected among the hybrids by the F test for all traits as well as for the environmental effects. The significant difference for environments was mainly due to climatic variations between locations. With regard to the hybrids, the differences were a result of the existing genetic variability. Besides, the hybrid x environment interaction was significant for all traits, indicating that the hybrid performance differed according to the environment of evaluation. For grain yield, 96.27% of the mean square of hybrid x environment interaction was accounted for by the complex part of its decomposition. This indicates that there were major differences in hybrid ranking from one environment to another, so that evaluation and subsequent selection must not be based on data of one location only. In this situation the use of experimental techniques for the accurate prediction of genotypic values is particularly appropriate, mainly when the genotypes are not evaluated in all environments.

The overall mean grain yield was 8.56 t ha^{-1} and coefficient of variation (CV%) 13.22%. For plant lodging, the mean rate was 2.19%, with a coefficient of variation of 51.71%. For plant height, the overall mean was 226.70 cm, with a coefficient of variation of 3.43%. Thus, all observed CV% values were within the limits reported

as acceptable for these traits, including for plant lodging (Alves et al. 2002, Aguiar et al. 2003, Lima et al. 2006).

The coefficients of determination (R^2) of the predicted genotypic values in the situation of balanced data and the two situations of imbalance, with the respective variations were high and highly significant ($p \leq 0.01$) for all traits (Tables 1 and 2). The higher the R^2 , the higher the coincidence in hybrid ranking is expected to be, that is, hybrids that are superior in one will also be superior in another situation. Besides, in all imbalance situations, the mean R^2 coefficients were very similar. Considering the mean R^2 values and their bootstrap confidence intervals for GY and PH, it was observed that REML/BLUP prediction of the genotypic values with up to 50% imbalance was reliable, since these predictions did not differ significantly from those obtained with balanced data (considering an accuracy of over 90%, i.e., $R^2 \geq 0.90$). However, the situation was different for PL, where the desired accuracy was maintained only up to 40% imbalance within environments, or with a reduction of up to 30% of environments. This was probably due to the high values of CV% observed for this trait (Tables 1 and 2). According to Bernardo (1996), there is a tendency to obtain higher R^2 estimates between the observed and predicted values with an increase in the number of hybrids. Reis et al. (2005) investigated the causes of this correlation and identified a bias in these estimates when the sample size is small and in the absence of a cross-validation procedure. This is relevant not only to assess the ability of the model to describe the set of observed data, but, above all, its predictive reliability.

Table 1. Coefficients of determination (R^2) and respective confidence intervals of the genotypic values predicted in the case of balanced data or for the different variations of hybrid loss within environments, for the traits grain yield (GY, t ha^{-1}), plant lodging (PL) and plant height (PH, cm) in maize

Trait	Percentage of hybrid loss ¹				
	10%	20%	30%	40%	50%
GY	0.96** ² [0.94 - 0.98] ³	0.96** [0.94 - 0.98]	0.96** [0.94 - 0.98]	0.96** [0.94 - 0.98]	0.92** [0.88 - 0.96]
PL	0.94** [0.90 - 0.96]	0.92** [0.90 - 0.96]	0.86** [0.81 - 0.92]	0.79** [0.72 - 0.86]	0.76** [0.67 - 0.85]
PH	0.96** [0.94 - 0.98]	0.96** [0.94 - 0.98]	0.96** [0.94 - 0.98]	0.96** [0.94 - 0.98]	0.90** [0.86 - 0.98]

** Significant at 1% probability by the F test

¹ Rounded values

² Mean coefficient of determination

³ Upper and lower limits by bootstrapping ($\alpha=0.05$)

Table 2. Coefficients of determination (R^2) and respective confidence intervals of the predicted genotypic values in the case of balanced data and the predicted values in the different variations of elimination of environments for the traits grain yield (GY, t ha⁻¹), plant lodging (PL) and plant height (PH, cm) in maize

Trait	Percentage of lost environments ¹				
	15%	23%	30%	38%	46%
GY	0.97** ² [0.96 - 0.98] ³	0.96** [0.94 - 0.98]	0.94** [0.92 - 0.98]	0.94** [0.90 - 0.96]	0.92** [0.86 - 0.94]
PL	0.95** [0.94 - 0.96]	0.92** [0.88 - 0.94]	0.89** [0.85 - 0.92]	0.89** [0.77 - 0.90]	0.79** [0.71 - 0.86]
PH	0.97** [0.96 - 0.98]	96.04** [0.94 - 0.98]	94.09** [0.92 - 0.98]	94.09** [0.90 - 0.96]	92.16** [0.88 - 0.94]

¹** Significant at 1% probability by the F test; ² Mean coefficient of determination; ³ Upper and lower limits by bootstrapping ($\alpha=0.05$)

It is therefore clear that a reduction in the number of hybrids also decreases the prediction potential of a model. Therefore, this study also showed the percentage of hybrids that would be selected in common by predictions using balanced data and in situations of imbalance.

The results for the percentage of coincidence of the selected hybrids and their bootstrap confidence intervals indicated that REML/BLUP allowed reliable results in predicting genotypic values with of up to 20% imbalance or with a reduction of 23% in the number of environments (Tables 3 and 4). This corroborates

Table 3. Percentage of hybrids selected in common and respective confidence intervals of the predicted genotypic values in the situation of balanced data and the predicted values in the different variations of losses of maize hybrids in the environments for grain yield (GY, t ha⁻¹), plant lodging (PL, %) and plant height (PH, cm) in maize

Trait	IS ²	Percentage of lost hybrids ¹				
		10%	20%	30%	40%	50%
GY	5	84.62 ³ [77.61 - 89.56] ⁴	76.92 [67.05 - 84.1]	76.92 [67.05 - 84.12]	76.92 [67.05 - 84.12]	69.23 [56.87 - 78.54]
	10	92.31 [88.59 - 94.85]	88.56 [83.05 - 92.22]	84.62 [77.61 - 89.56]	84.62 [77.61 - 89.56]	76.92 [67.05 - 84.12]
	15	87.18 [81.23 - 91.33]	87.18 [81.23 - 91.33]	84.62 [77.61 - 89.56]	82.05 [74.05 - 87.76]	74.36 [63.61 - 82.28]
	20	92.31 [88.59 - 94.85]	92.31 [88.59 - 94.85]	88.46 [83.05 - 92.22]	84.62 [77.61 - 89.56]	80.77 [72.28 - 86.86]
	5	84.62 [77.61 - 89.56]	84.62 [77.61 - 89.56]	61.54 [47.05 - 72.80]	61.54 [47.05 - 72.8]	38.46 [19.60 - 54.6]
PL	10	73.08 [61.91 - 81.35]	73.08 [61.91 - 81.35]	65.38 [51.91 - 75.69]	61.54 [47.05 - 72.80]	57.69 [42.27 - 69.87]
	15	84.62 [77.61 - 89.56]	82.05 [74.05 - 87.76]	76.92 [67.05 - 84.12]	64.10 [50.28 - 74.73]	58.97 [43.85 - 70.85]
	20	86.54 [80.32 - 90.89]	82.69 [74.94 - 88.21]	76.92 [67.05 - 84.12]	71.15 [59.38 - 79.95]	71.15 [59.38 - 79.95]
	5	92.31 [88.59 - 94.85]	92.31 [88.59 - 94.85]	92.31 [88.59 - 94.85]	84.62 [77.61 - 89.56]	84.62 [77.61 - 89.56]
	10	88.46 [83.05 - 92.22]	88.46 [83.05 - 92.22]	84.62 [77.61 - 89.56]	84.62 [77.61 - 89.56]	76.92 [67.05 - 84.12]
PH	15	89.74 [84.89 - 93.10]	87.18 [81.23 - 91.33]	84.62 [77.61 - 89.56]	79.49 [70.53 - 85.95]	76.92 [67.05 - 84.12]
	20	90.38 [85.81 - 93.54]	90.38 [85.81 - 93.54]	84.62 [77.61 - 89.56]	84.62 [77.61 - 89.56]	75.00 [64.47 - 82.74]

¹ Rounded values; ² selection intensity (%) ³ Mean percentage; ⁴ Upper and lower limits by bootstrapping ($\alpha=0.05$)

Table 4. Percentage of hybrids selected in common and respective confidence intervals of the predicted genotypic values in the situation of balanced data and the predicted values in the different variations of of elimination of environments for grain yield (GY, t ha⁻¹), plant lodging (PL, %) and plant height (PH, cm) in maize

Trait	IS ²	Percentage of lost environments ¹				
		15%	23%	30%	38%	46%
GY	5	84.62 ³	84.62	84.62	76.92	69.23
		[77.61 - 89.56] ⁴	[77.61 - 89.56]	[77.61 - 89.56]	[67.05 - 84.12]	[56.87 - 78.54]
	10	88.46	88.46	80.77	80.77	76.92
		[83.05 - 92.22]	[83.05 - 92.22]	[72.28 - 86.86]	[72.28 - 86.89]	[67.05 - 84.12]
	15	92.31	90.38	88.46	88.46	86.54
PL	5	[88.59 - 94.85]	[85.81 - 93.54]	[83.05 - 92.22]	[83.05 - 92.22]	[80.32 - 90.89]
		89.74	87.18	87.18	82.05	79.49
	20	[84.89 - 93.10]	[81.23 - 91.33]	[81.23 - 91.33]	[74.05 - 87.76]	[70.53 - 85.95]
	5	84.62	76.92	76.92	61.54	46.15
		[77.61 - 89.56]	[67.05 - 84.12]	[67.05 - 84.12]	[47.05 - 72.80]	[28.44 - 60.82]
10	88.46	76.92	76.92	61.54	57.69	
	[83.05 - 92.22]	[67.05 - 84.12]	[67.05 - 84.12]	[47.05 - 72.80]	[42.27 - 69.87]	
15	84.62	82.05	79.49	71.79	71.79	
	[77.61 - 89.56]	[74.05 - 87.76]	[70.53 - 85.95]	[60.22 - 80.41]	[60.22 - 80.41]	
PH	5	88.46	80.77	78.85	78.85	69.23
		[83.05 - 92.22]	[72.28 - 86.86]	[69.65 - 85.49]	[69.65 - 85.49]	[56.87 - 78.54]
	20	92.31	92.31	92.31	84.62	84.62
	5	[88.59 - 94.85]	[88.59 - 94.85]	[88.59 - 94.85]	[77.61 - 89.56]	[77.61 - 89.56]
		92.31	92.31	92.31	80.77	76.92
10	[88.59 - 94.85]	[88.59 - 94.85]	[88.59 - 94.85]	[72.28 - 86.86]	[67.05 - 84.12]	
	89.74	87.18	84.62	82.05	76.92	
15	[84.89 - 93.10]	[81.23 - 91.33]	[77.61 - 89.56]	[74.05 - 87.76]	[67.05 - 84.12]	
	90.38	90.38	88.46	88.46	84.62	
20	[85.81 - 93.54]	[85.81 - 93.54]	[83.05 - 92.22]	[83.05 - 92.22]	[77.61 - 89.56]	

¹Rounded values; ² selection intensity (%) ³ Mean percentage; ⁴ Upper and lower limits, respectively, by bootstrapping ($\alpha=0.05$)

the flexibility of REML/BLUP, cited by Henderson (1984), which allows a certain degree of imbalance and the presence of complex genotype-environment interaction, even for traits such as PL, with non-normal distribution and high CV% values. The reason is that the method uses data from all experiments to estimate the fixed effects of each one, provided there is a relation between them, in other words, plant material common to the different experiments (Henderson 1963).

Another important factor, aside from the percentage of selected hybrids in common, is the coincidence in hybrid ranking. The ranking can be modified according to the choice of the statistical model, in the situations where information of relatedness between the selection units is available, as well as in situations of data imbalance (Henderson 1984, Phiepo et al. 2008). According to Duarte and Vencovsky (2001),

the shrinking of predicted means, i.e., the shrinkage effect, above all for populations with low genetic variability, may also lead to changes in the relative ranking of progenies or hybrids of different populations, even in the case of orthogonality and balance. Moreover, in the usual situations of incomplete blocks subject to planned or unplanned imbalance, as in the case of this study, where the lattice design was used, changes in ranking are expected even when no observations are withdrawn.

In this study considerable differences were stated in the hybrid ranking, which increased significantly both with the imbalance degree (loss of hybrids in environments and elimination of environments) and the genetic complexity of the trait and the coefficient of experimental variation. For grain yield and particularly for plant lodging, the reduced selection intensity and

the degree of data imbalance resulted in a marked increase of selection coincidence (data not shown). However, for plant height, better results were obtained at higher selection intensities, even with a higher degree of data imbalance. The ranking of the best hybrid remained unchanged, regardless of the statistical method, trait or degree of data imbalance. Reis et al. (2005), in a study with similar objectives, but using diallel analysis, also found small differences in the ranking of the best hybrids, causing no difficulties for their identification.

Although the results indicate that with REML/BLUP the number of evaluation environments can be reduced by 23%, it is recommended that instead, the number of hybrids be increased by about 20%. The experiments would be planned so that around 80% of all test hybrids would be randomly evaluated in each environment. This would make a combined analysis possible without losing experimental quality, but with a cost reduction in the evaluation of plant material. The arguments cited by Bernardo (1996) were therefore corroborated, showing that the use of predictive techniques is not only relevant, but also a viable

alternative to reduce the cost of hybrid evaluations in maize breeding programs.

CONCLUSIONS

The REML/BLUP is reliable to predict the genotypic values of missing maize hybrids under conditions of up to 20% loss of hybrids in environments or 23% reduction of environments, even in the presence of significant and complex hybrid x environment interaction. This methodology may therefore be useful to increase the efficiency of breeding programs.

ACKNOWLEDGEMENTS

The authors thank the employees of the Department of Genetics of ESALQ / USP, for their contribution in carrying out the experiments, as well as Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), for financial support.

Predição de valores genotípicos de híbridos de milho em experimentos desbalanceados

RESUMO - *O objetivo deste trabalho foi avaliar se o método REML/BLUP pode ser útil para prever valores genotípicos de híbridos de milho em grupo de experimentos desbalanceados. Um conjunto de 256 híbridos simples foi avaliado em 13 ambientes para os caracteres produção de grãos, altura da planta e acamamento de plantas. As predições dos híbridos foram computadas via REML/BLUP considerando as situações de dados balanceados, retiradas de híbridos dentro de ambientes e retiradas completas dos dados de ambientes, as quais foram simuladas utilizando o método de amostragem bootstrap. Foram computados valores dos coeficientes de determinação e de porcentagens de coincidência de seleção entre os valores genotípicos preditos sob dados balanceados e sob dados desbalanceados. O método REML/BLUP prediz com acurácia os valores genotípicos de híbridos faltantes sob condições de até 20% de perda de híbridos dentro de ambientes ou a redução de 23% dos ambientes, mesmo na presença da interação híbridos x ambiente significativa e do tipo complexa.*

Palavras chave: REML/BLUP, *Zea mays* L., híbridos de milho, seleção.

REFERENCES

- Aguiar AM, Carlini-Garcia LA, Silva AR, Santos MF, Garcia AAF and Souza Junior CL (2003) Combining ability of inbred lines of maize and stability of their respective single-crosses. **Scientia Agrícola** **60**: 83-89.
- Alves GF, Ramalho MAP and Souza JC (2002) Alterações nas propriedades genéticas da população CMS-39 submetida à seleção massal para prolificidade. **Revista Brasileira de Milho e Sorgo** **1**: 89-101.
- Bernardo R (1996) Best linear unbiased prediction of maize single-cross performance. **Crop Science** **36**: 50-56.
- Bernardo R (2002) **Breeding for quantitative traits in plants**. Stemma Press, Woodbury, 369p.
- Duarte J and Vencovsky R (2001) Estimativa e predição por modelo linear misto com ênfase na ordenação de médias de tratamentos genéticos. **Scientia Agrícola** **58**: 109-117.
- Henderson CR (1963) Selection index and expected genetic advance. In: Hanson WD and Robinson HF (Eds.)

- Statistical genetics and plant breeding**, NAS/NCR, Washington, p. 141-163.
- Henderson CR (1974) General flexibility of linear model for sire evaluation. **Journal of Dairy Science** **57**: 973-972.
- Henderson CR (1984) **Applications of linear models in animal breeding**. University of Guelph Press, Guelph, 462p.
- Lima MLA, Souza Junior CL, Bento DAV, Souza AP and Carlini-Garcia LA (2006) Mapping QTL for grain yield and plant traits in a tropical maize population. **Molecular Breeding** **17**: 227-239.
- Panter DM and Allen FL (1995) Using best linear unbiased predictions to enhance breeding for yield in soybean: II Selection of superior crosses from a limited number of yield trials. **Crop Science** **35**: 405-410.
- Piepho HP, Möhring J, Melchinger AE and Büchse A (2008) BLUP for phenotypic selection and variety test. **Euphytica** **161**: 209-228.
- Pimentel Gomes F (2000) **Curso de Estatística Experimental**. Editora FEALQ, Piracicaba, 478p.
- Reis AJS, Chaves LJ, Duarte JB and Brasil EM (2005) Prediction of hybrid means from a partial circulant diallel table using the ordinary least square and the mixed model methods. **Genetics and Molecular Biology** **28**: 314-320.
- Robertson A (1959) **Experimental design on the measurement of heritabilities and genetic correlations: biometrical genetics**. Pergamon Press, New York, 186p.
- Sas Institute (2003) **SAS/STAT software 9.1**. SAS, Cary, 1028p.
- Searle SR (1971) **Linear models**. John Wiley Press, New York, 532p.
- Steel RGD and Torrie JH (1980) **Principles and procedures of statistics: a biometrical approach**. McGraw-Hill Press, New York, 631p.
- Yates F (1934) The analysis of multiple classifications with unequal numbers in the different subclasses. **Journal of American Statistical Association** **29**: 51-66.