

Biplot AMMI graphic representation of specific combining ability

João Batista Duarte^{*1} and Rogério de Melo Costa Pinto²

¹ Escola de Agronomia, Universidade Federal de Goiás (UFG), Caixa Postal 131, CEP 74001-970, Goiânia, GO, Brazil; ² Faculdade de Matemática, Universidade Federal de Uberlândia, Av. João Naves Ávila, 2.121, Campus Santa Mônica, CEP 38400-902, Uberlândia, MG, Brazil. (* Corresponding Author. E-mail: jbduarte@agro.ufg.br)

ABSTRACT

Specific combining ability (SCA) is a genetic property of great interest in breeding programs aiming at hybrid cultivars. It enables assessment of the genotype combinations most promising for hybridization. This study presents the foundations of biplot graphic display associated with AMMI (additive main effects and multiplicative interaction) analysis, and their application to genetic studies of this kind. The analysis allows the partition of the sum of squares of the elements of a diallel table (matrix), the SCA sum of squares, to capture a systematic portion and discard an additional noise to the experimental error. The systematic part or pattern is represented graphically by the biplot. The parental combinations with greatest potential for hybrid production may be identified in this graph, and the SCA divergence among the parents can also be assessed. A numerical example is presented with SAS (Statistical Analysis System) statements for its implementation.

KEY WORDS: Singular decomposition, principal component, eigen analyses, diallel, cross, hybrid.

INTRODUCTION

Breeding for hybrid cultivars is one of the earliest and most efficient methods of genetic improvement. A fundamental step in this process is the genotype assessment in crosses to select the parents which result in the best hybrid combinations. Single cross among all the parents or among those of known diverse origin must be obtained for this purpose. Diallel crosses have been designed to assess the properties of the parents submitted to hybridization. The methods proposed by Griffing (1956) and Gardner and Eberhart (1966) are outstanding because of their simplicity and number of applications.

The genetic properties of greatest interest are the general combining ability (GCA) and the specific combining ability (SCA). The first concept is the mean potential of a genotype for hybrid production when crossed with a series of other genotypes (broad genetic base). The second refers not to individual genotypes, but to properties of a specific cross, that is, it measures the potential of each parent combination for hybrid production.

Once the hybrids have been obtained and experimentally assessed, either with or without their parents, the data are usually treated statistically by diallel or combining ability analyses. The resulting SCA estimates are, finally, presented in two-way tables (females x males) also called diallel tables. Although their interpretation is simple and direct, the

identification of the favorable hybrid combinations could be improved with a graphic representation. The parent clustering in the heterotic groups could also be easily confirmed by visual inspection. It is worth mentioning that in classical diallel analyses the SCA is estimated as residue from fitting of the additive main effects. However, it is risky to assume that this residue is solely due to the genetic interaction between two parents. Part of this component may be the result of other effects (noise) besides the interaction itself (Mandel, 1971). In this sense, the adoption of models that allow the isolation of additional error effects in the SCA estimation is pertinent.

This study presents the fundamentals of biplot representation (Gabriel, 1971), usually used in multivariate analyses, applied to the description of the SCA table or matrix. Considering the conceptual equivalence between SCA and a simple interaction of factors, the theoretical base of this mathematical technique has already been described in the AMMI analysis (Additive Main Effects and Multiplicative Interaction Analysis). In this method, the main effects are additive modeled by analysis of variance, and the interaction effects are multiplicative modeled by singular value decomposition (SVD). This latter procedure is equivalent to a double principal components analysis (PCA), applied jointly to rows and columns of the interaction matrix (Zobel et al., 1988; Gauch and Zobel, 1996; Duarte and Vencovsky, 1999). Although the majority of the applications of

this analysis are related to genotype x environment interaction (GEI) studies, the authors suggest its perfect adaptation to line assessment in test crossings, such as the diallel crosses.

METHODOLOGY

Singular value decomposition

The biplot graphic display is based on a matrix algebra tool called singular value decomposition (SVD). Proposed by Eckart and Young (1936), this procedure allows writing a real matrix \mathbf{B} , of rank p , as a sum of p orthogonal unitary rank matrices:

$$\mathbf{B}_{(l \times c)} = \sum_{k=1}^p \lambda_k \mathbf{u}_k \mathbf{v}_k'$$

where: $k=1, 2, \dots, p$ and $p \leq \min\{l, c\}$

and λ_k is the k^{th} singular value of \mathbf{B} (the square root of the k^{th} non-zero eigenvalue of $\mathbf{B}\mathbf{B}'$ or $\mathbf{B}'\mathbf{B}$), and \mathbf{u}_k and \mathbf{v}_k' are the associated singular vectors, column vector and row vector respectively.

When the sum for k is applied only to the n first terms (with $n < p$), an approximation for \mathbf{B} of rank equal to n is obtained, with interesting mathematical and statistical properties. The resulting matrix represents the best approximation of \mathbf{B} in the sense of having a smaller sum of squares of the deviations of the \mathbf{B} elements compared to corresponding elements of the approximate matrix ($\hat{\mathbf{B}}$). Thus, the denomination n -approximation of least squares of the \mathbf{B} matrix (Good, 1969; Gabriel, 1978).

Arranging the p singular values in decreasing order in a diagonal matrix ($\mathbf{S}_{(p)}$) and the corresponding singular \mathbf{u}_k and \mathbf{v}_k' vectors in two matrices $\mathbf{U}_{(l \times p)}$ and $\mathbf{V}'_{(p \times c)}$, respectively, the previous expression may be written as: $\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{V}'$. Thus \mathbf{U} will have the p vectors \mathbf{u}_k in its columns and \mathbf{V}' , the p vectors \mathbf{v}_k' in its rows, sequentially placed.

The use of SVD to obtain the approximation of a matrix by another of a lower rank represents exactly the application adopted by the AMMI analysis to interpret GEI (the interactions matrix assumes the role of \mathbf{B}). In the present case, $\mathbf{B}_{(f \times m)}$ is the SCA matrix, where f is the number of rows (female) and m is the number of columns (males) in the diallel table. In a complete diallel (all the genotypes crossed among each other without reciprocal effect) \mathbf{B} is a square and symmetric matrix. However, in a partial diallel

(crosses only among genotypes of different groups) \mathbf{B} is no longer symmetrical and can even be rectangular, but these facts do not preclude the application of the SVD in view of its generality.

Another important property of the SVD is that it determines a partition of the sum of the squares of the elements of the original matrix ($SS_{\mathbf{B}}$). Considering that \mathbf{B} is the matrix of the estimates of the SCA deviations (\hat{s}_{ij} , with zero mean), this sum of squares relates directly with the SS_{SCA} of the diallel analysis of variance. Thus, $SS_{\mathbf{B}} = SS_{SCA}$ and $SS_{\mathbf{B}} = 2SS_{SCA}$, in the cases of partial diallel and complete diallel (symmetric \mathbf{B}), respectively. In the more general situations (rectangular \mathbf{B}), the partitioning of the $SS_{\mathbf{B}}$ may be illustrated in algebraic terms as:

$$\begin{aligned} \mathbf{B}_{(f \times m)} &= \lambda_1 \mathbf{u}_1 \mathbf{v}_1' + \lambda_2 \mathbf{u}_2 \mathbf{v}_2' + \dots + \lambda_p \mathbf{u}_p \mathbf{v}_p' = \mathbf{U}\mathbf{S}\mathbf{V}' \\ \left[\begin{array}{c} \mathbf{B}_{(f \times m)} \\ \mathbf{B} = [b_{ij}] \end{array} \right] &= \left[\begin{array}{c} \mathbf{u}_1 = [u_{1j}] \\ \mathbf{u}_2 = [u_{2j}] \\ \dots \\ \mathbf{u}_p = [u_{pj}] \end{array} \right] \left[\begin{array}{c} \mathbf{S} \\ \mathbf{S} = [s_{ij}] \end{array} \right] \left[\begin{array}{c} \mathbf{v}_1' = [v_{1i}] \\ \mathbf{v}_2' = [v_{2i}] \\ \dots \\ \mathbf{v}_p' = [v_{pi}] \end{array} \right] \\ \sum_{ij} b_{ij}^2 &= \sum_{ij} s_{ij}^2 = \sum_{ij} b_{1ij}^2 + \sum_{ij} b_{2ij}^2 + \dots + \sum_{ij} b_{pij}^2 = SS_{SCA} \\ \sum_{ij} s_{ij}^2 &= \sum_{k=1}^p \lambda_k^2 = \lambda_1^2 + \lambda_2^2 + \dots + \lambda_p^2 = \sum_{k=1}^p SS_{IPCA_k} = SS_{SCA} \end{aligned}$$

Therefore, SS_{SCA} is partitioned in components relative to each term of the SVD or, in multivariate language, at each interaction principal components axis (IPCA k , as in the notation of the AMMI analysis). Emphasize that the interaction assessed here is SCA. Note that it is not necessary to find the p partial matrixes to obtain such partition, it is enough to obtain the singular values of \mathbf{B} . The sum of squares for the k^{th} multiplicative term (singular or principal interaction axis) is equivalent to the square of the corresponding singular value: $SS_{SCA(IPCA_k)} = \lambda_k^2$. Given the properties of SVD, the p components of SS_{SCA} are also orthogonal and, therefore, independent. Furthermore, the number of degrees of freedom associated to the k -th component is asymptotically given by: $DF_{IPCA_k} = f + m - l - 2k$. This allow the application of the F test as in a conventional analysis of variance (Gollob, 1968). For complete diallel, a partition of the degrees of freedom is not yet available. However, the absence of a statistical test criterion does not mathematically invalidate the biplot graphic representation of the a matrix.

By analogy with PCA, cumulatively taking the successive partition terms, sequentially better approximations of the original SS_{SCA} can be obtained. However, as $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_p^2$, it is possible that

only the few first terms (one, two or three) already describe a high proportion of this sum of squares. This allows the description of the SCA using a parsimonious model (in the sense of using few degrees of freedom) and with good capacity to explain the original variability present in the matrix. Therefore, the analysis allows a interpretation of the SS_{SCA} by an approximation of rank n of the SCA matrix. As n is possibly smaller than p (e.g. $n=1, 2$ or 3) the analysis results in an informative model which can be represented graphically, since a complex p -dimensional space is represented in a bi or tridimensional graph.

In the context of the AMMI method, the objective of the analysis is to describe the interaction by a reduced number of axes. It is expected that, successively, each axis will capture a smaller portion of the pattern underlying to the interaction (general law which governs the phenomena) and more noise present in the data. Thus, a simpler model would discard a high proportion of noise which has no agronomic interest and can hamper good predictions, retaining mainly the portion rich in deterministic information (Gauch and Zobel, 1988; Gauch, 1990). This separation in pattern and noise effects may also be expressed in terms of sum of squares:

$$SS_{SCA} = \sum_{k=1}^p \lambda_k^2 = \left(\sum_{k=1}^n \lambda_k^2 \right) + \left(\sum_{k=n+1}^p \lambda_k^2 \right)$$

$$= (SS_{SCA_pattern}) + (SS_{SCA_noise}).$$

The biplot graphic display

An important property of the SVD approximation is its allowance for the representation of effects of each row and of each column in a single graph called *biplot*. The name does not come from the space dimension of representation, but from the fact that it displays simultaneously the rows and the columns of the matrix.

To understand the mathematical origin of a biplot consider the SVD n -approximation for the \mathbf{B} matrix of rank $p > n$: $\hat{\mathbf{B}}_{(f \times m)} = \sum_{k=1}^n \lambda_k \mathbf{u}_k \mathbf{v}'_k = \tilde{\mathbf{U}} \tilde{\mathbf{S}} \tilde{\mathbf{V}}'$; where $\tilde{\mathbf{U}}_{(f \times n)}$ has in its columns only the n first $\mathbf{u}_{(f \times 1)}$ vectors; $\tilde{\mathbf{V}}'_{(n \times m)}$ has in its rows the n first $\mathbf{v}'_{(1 \times m)}$ vectors; and $\tilde{\mathbf{S}}_{(n)}$ is the diagonal matrix with the first singular values, $\lambda_1, \lambda_2, \dots, \lambda_n$ (Duarte and Vencovsky, 1999). The expression may further be written as: $\hat{\mathbf{B}} = (\tilde{\mathbf{U}} \tilde{\mathbf{S}}^{1/2})(\tilde{\mathbf{S}}^{1/2} \tilde{\mathbf{V}}')$; or simply $\hat{\mathbf{B}} = (\tilde{\mathbf{G}})(\tilde{\mathbf{H}}')$, where $\tilde{\mathbf{G}}_{(f \times n)} = (\tilde{\mathbf{U}} \tilde{\mathbf{S}}^{1/2})$ and $\tilde{\mathbf{H}}'_{(n \times m)} = (\tilde{\mathbf{S}}^{1/2} \tilde{\mathbf{V}}')$. Thus, $\tilde{\mathbf{G}}$ matrix will have f vectors called row markers, each one with n elements, the coordinates of the females (F) for

each singular axis selected. Similarly, $\tilde{\mathbf{H}}'$ will have m column markers, each also with n elements, corresponding to the male (M) coordinates for the same n axes. Given that these axes are orthogonal, the f females and m males can be plotted in a single cartesian system of n perpendicular axes. In practice, similar to the scatter plots in PCA, the rank of the approximate matrix (n) to be displayed should be one, two or three, which would allow a graphic representation in one, two or three dimensions, respectively.

A property of biplot representation is that any element of the approximate matrix is exactly the inner product of the vectors (markers) corresponding to its row and to its column (Gabriel, 1971). In this case, the product between the coordinates of the i^{th} female and the j^{th} male yields the expected SCA value for that hybrid combination ($\hat{s}_{ij}^{\text{AMMI}}$). Clearly, this value does not correspond to that in the diallel table because the biplot represents the SVD n -approximation of the original \mathbf{B} matrix. However, it is expected that this prediction, rich in pattern, is closer to the real SCA (s_{ij}) than the values \hat{s}_{ij} of the diallel table, which are potentially rich in noise (Gauch, 1990). In this sense, a male/female pair with coordinates of high magnitude and the same signal has high SCA (positive interaction), characterizing a favorable combination for hybridization. On the other hand, a combination whose coordinates have opposite signs and negative SCA is unfavorable. Thus the intergroup genotype distances (F/M) have inverse interpretation in relation to the distance visualized in the graph, that is, points close together do not indicate genotypic similarity but rather affinity for hybrid combination.

On the other hand, inspecting only the female (or male) markers, it is possible to assess the intragroup divergences. In this case, close points indicate greater genotypic similarity. In complete diallels, the females and males are the same genotypes and, therefore, such inspection would enable the assessment of the divergence among all the parents. Thus for genotype clustering, it would be enough to plot only the female markers (or the males) as in PCA display. However, in this case the opportunity to graphically predict the SCA magnitude and to identify the most promising hybrid combinations would be lost.

APPLICATION

To illustrate the use of the biplot graphic display in SCA description, Griffing (1956) method IV results

(diallel with F1 hybrids, without parents and reciprocals) were taken. The \hat{s}_{ij} estimates for grain yield in maize are expressed by the symmetric $\mathbf{B}_{(9 \times 9)}$ matrix (F indicates females, and M, males):

$$\mathbf{B} = \begin{matrix} & \text{M1} & \text{M2} & \text{M3} & \text{M4} & \text{M5} & \text{M6} & \text{M7} & \text{M8} & \text{M9} & \\ \left[\begin{array}{cccccccccc} 0.00 & \mathbf{4.94} & \mathbf{14.46} & \mathbf{-2.76} & \mathbf{2.80} & \mathbf{-12.31} & \mathbf{-16.93} & \mathbf{-8.67} & \mathbf{18.47} & & \text{F1} \\ 4.94 & 0.00 & \mathbf{-6.54} & \mathbf{14.14} & \mathbf{8.90} & \mathbf{5.68} & \mathbf{13.07} & \mathbf{-2.47} & \mathbf{-37.73} & & \text{F2} \\ 14.46 & -6.54 & 0.00 & \mathbf{-29.94} & \mathbf{19.01} & \mathbf{15.80} & \mathbf{4.48} & \mathbf{-6.96} & \mathbf{-10.31} & & \text{F3} \\ -2.76 & 14.14 & -29.94 & 0.00 & \mathbf{11.50} & \mathbf{-8.21} & \mathbf{3.77} & \mathbf{-19.47} & \mathbf{30.97} & & \text{F4} \\ 2.80 & 8.90 & 19.01 & 11.50 & 0.00 & \mathbf{-36.16} & \mathbf{-0.77} & \mathbf{4.99} & \mathbf{-10.27} & & \text{F5} \\ -12.31 & 5.68 & 15.80 & -8.21 & -36.16 & 0.00 & \mathbf{-10.78} & \mathbf{32.77} & \mathbf{13.22} & & \text{F6} \\ -16.93 & 13.07 & 4.48 & 3.77 & -0.77 & -10.78 & 0.00 & \mathbf{5.66} & \mathbf{1.50} & & \text{F7} \\ -8.67 & -2.47 & -6.96 & -19.47 & 4.99 & 32.77 & 5.66 & 0.00 & \mathbf{-5.85} & & \text{F8} \\ 18.47 & -37.73 & -10.31 & 30.97 & -10.27 & 13.22 & 1.50 & -5.85 & 0.00 & & \text{F9} \end{array} \right. \end{matrix}$$

The matrix symmetry comes from the assumption of no reciprocal effects, although the analysis is sufficiently generic to accommodate non-symmetric matrices. The SCA mean square is 339.44, with 27 degrees of freedom, which corresponds to: $SS_{SCA} = 9164.28$ (Griffing, 1956). The SS of the matrix elements ($\sum b_{ij}^2 = \sum \hat{s}_{ij}^2 = 18328.55$) is then equivalent to twice the SS_{SCA} (\hat{s}_{ij} values repeated above and below the main diagonal). Therefore, the partition of the sum of square of the matrix elements corresponds to

partitioning the SS_{SCA} .

As mentioned earlier, the SVD decomposition of the \mathbf{B} matrix is: $\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{V}'$. The \mathbf{U} , \mathbf{S} and \mathbf{V} matrices, that determine this partition, can be obtained using a computer program for matrix algebra. The *Proc IML* (interactive matrix language procedure) of the SAS (Statistical Analysis System) can be used, for example, with the syntax "call svd (U,S,V,B);". The results of the application of this command are the matrices:

i) $\mathbf{U}_{(9 \times 8)}$ matrix of singular vectors associated with rows (females):

$$\mathbf{U} = \begin{bmatrix} 0.2640 & 0.1759 & -0.1622 & 0.0515 & 0.5091 & 0.5569 & 0.4353 & -0.0210 \\ -0.3450 & -0.3612 & -0.0589 & -0.4124 & -0.3872 & 0.0198 & 0.5486 & -0.1207 \\ -0.3671 & 0.2619 & 0.2526 & -0.2661 & 0.5169 & -0.3951 & -0.1281 & -0.3335 \\ -0.0160 & 0.5778 & -0.5138 & 0.2488 & -0.3196 & -0.3324 & 0.1029 & 0.0760 \\ 0.3470 & -0.3116 & -0.2969 & -0.4184 & 0.1818 & -0.2327 & -0.2924 & 0.4853 \\ 0.5286 & -0.1209 & 0.5549 & 0.2996 & -0.1352 & -0.3438 & 0.2396 & -0.0566 \\ 0.2452 & 0.0337 & -0.0381 & -0.1533 & -0.3288 & 0.3567 & -0.5120 & -0.5524 \\ -0.3097 & 0.2535 & 0.4556 & 0.0130 & -0.1937 & 0.3505 & -0.1962 & 0.5676 \\ -0.3470 & -0.5090 & -0.1933 & 0.6373 & 0.1566 & 0.0201 & -0.1975 & -0.0446 \end{bmatrix}$$

ii) $\mathbf{S}_{(8)}$ matrix of singular values:

$$\mathbf{S} = \text{diag} \{77.1865 \ 63.3445 \ 60.1728 \ 54.7776 \ 36.8391 \ 17.6392 \ 8.1048 \ 1.7241\}$$

iii) $\mathbf{V}'_{(8 \times 9)}$ matrix of singular vectors associated with columns (males):

$$\mathbf{V}' = \begin{bmatrix} -0.2640 & 0.3450 & 0.3671 & 0.0160 & -0.3470 & -0.5286 & -0.2452 & 0.3097 & 0.3470 \\ -0.1759 & 0.3612 & -0.2619 & -0.5778 & 0.3116 & 0.1209 & -0.0337 & -0.2535 & 0.5090 \\ -0.1622 & -0.0589 & 0.2526 & -0.5138 & -0.2969 & 0.5549 & -0.0381 & 0.4556 & -0.1933 \\ 0.0515 & -0.4124 & -0.2661 & 0.2488 & -0.4184 & 0.2996 & -0.1533 & 0.0130 & 0.6373 \\ 0.5091 & -0.3872 & 0.5169 & -0.3196 & 0.1818 & -0.1352 & -0.3288 & -0.1937 & 0.1566 \\ -0.5569 & -0.0198 & 0.3951 & 0.3324 & 0.2327 & 0.3438 & -0.3567 & -0.3505 & -0.0201 \\ 0.4353 & 0.5486 & -0.1281 & 0.1029 & -0.2924 & 0.2396 & -0.5120 & -0.1962 & -0.1975 \\ 0.0210 & 0.1207 & 0.3335 & -0.0760 & -0.4853 & 0.0566 & 0.5524 & -0.5676 & 0.0446 \end{bmatrix}$$

Note that the \mathbf{S} matrix of non-null singular values is of 8×8 dimension indicating that the rank of \mathbf{B} is also $p=8$, which corresponds to the number of parents minus one. Thus the SS_{SCA} may be partitioned in up to eight components (IPCA axes) as shown in Table 1. In the present case (complete diallel), each component corresponds to half the square of each singular value, that is: $SS_{SCA(IPCA_k)} = (\lambda_k^2)/2$, with

$k=1,2,\dots,8$. In the partial diallels (non-symmetric matrices and possibly rectangular matrices) the partition is direct, without division by two. In this example, the biplot representation of the SCA is illustrated taking the first three singular axes.

Before constructing the biplot graphic display it is necessary to obtain the \mathbf{G} and \mathbf{H}' matrices, such that $\mathbf{G}=\mathbf{U}\mathbf{S}^{1/2}$ and $\mathbf{H}'=\mathbf{S}^{1/2}\mathbf{V}'$ (so that $\mathbf{B}=\mathbf{G}\mathbf{H}'$):

i) $\mathbf{G}_{(9 \times 8)}$ matrix of row markers (females F1, F2, ..., F9):

$$\mathbf{G} = \begin{bmatrix} 2.3198 & 1.4001 & -1.2579 & 0.3811 & 3.0903 & 2.3389 & 1.2392 & -0.0276 & \text{F1} \\ -3.0311 & -2.8750 & -0.4565 & -3.0524 & -2.3502 & 0.0834 & 1.5617 & -0.1585 & \text{F2} \\ -3.2249 & 2.0848 & 1.9591 & -1.9692 & 3.1374 & -1.6595 & -0.3646 & -0.4379 & \text{F3} \\ -0.1409 & 4.5983 & -3.9856 & 1.8411 & -1.9398 & -1.3960 & 0.2928 & 0.0998 & \text{F4} \\ 3.0482 & -2.4803 & -2.3030 & -3.0966 & 1.1035 & -0.9773 & -0.8325 & 0.6372 & \text{F5} \\ 4.6443 & -0.9626 & 4.3047 & 2.2172 & -0.8207 & -1.4441 & 0.6821 & -0.0743 & \text{F6} \\ 2.1540 & 0.2680 & -0.2958 & -1.1342 & -1.9958 & 1.4980 & -1.4577 & -0.7254 & \text{F7} \\ -2.7205 & 2.0173 & 3.5341 & 0.0964 & -1.1755 & 1.4722 & -0.5587 & 0.7452 & \text{F8} \\ -3.0488 & -4.0508 & -1.4991 & 4.7167 & 0.9508 & 0.0844 & -0.5623 & -0.0585 & \text{F9} \end{bmatrix}$$

ii) $\mathbf{H}'_{(8 \times 9)}$ matrix of column markers (males: M1, M2,...M9):

$$\mathbf{H}' = \begin{array}{c} \begin{array}{cccccccccc} \text{M1} & \text{M2} & \text{M3} & \text{M4} & \text{M5} & \text{M6} & \text{M7} & \text{M8} & \text{M9} \end{array} \\ \begin{bmatrix} -2.3198 & 3.0311 & 3.2249 & 0.1409 & -3.0482 & -4.6443 & -2.1540 & 2.7205 & 3.0488 \\ -1.4001 & 2.8750 & -2.0848 & -4.5983 & 2.4803 & 0.9626 & -0.2680 & -2.0173 & 4.0508 \\ -1.2579 & -0.4565 & 1.9591 & -3.9856 & -2.3030 & 4.3047 & -0.2958 & 3.5341 & -1.4991 \\ \hline 0.3811 & -3.0524 & -1.9692 & 1.8411 & -3.0966 & 2.2172 & -1.1342 & 0.0964 & 4.7167 \\ 3.0903 & -2.3502 & 3.1374 & -1.9398 & 1.1035 & -0.8207 & -1.9958 & -1.1755 & 0.9508 \\ -2.3389 & -0.0834 & 1.6595 & 1.3960 & 0.9773 & 1.4441 & -1.4980 & -1.4722 & -0.0844 \\ 1.2392 & 1.5617 & -0.3646 & 0.2928 & -0.8325 & 0.6821 & -1.4577 & -0.5587 & -0.5623 \\ 0.0276 & 0.1585 & 0.4379 & -0.0998 & -0.6372 & 0.0743 & 0.7254 & -0.7452 & 0.0585 \end{bmatrix} \end{array}$$

The dotted rows in \mathbf{G} and \mathbf{H}' delimit at the left and above the $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{H}}'$ matrices, respectively, and the coordinates corresponding to females and to males for the construction of a three dimensional biplot (*biplot-AMMI3*). This representation captures 74.15% of the original SS_{SCA} (Table 1) and the predominant pattern associated with the SCA. Thus, it is expected that the SCA predictions obtained by

the SVD 3-approximation ($\hat{\mathbf{B}} = \tilde{\mathbf{G}} \tilde{\mathbf{H}}'$) should be more realistic than those presented in the original diallel table (matrix \mathbf{B}). If only the first two columns of \mathbf{G} and the first two rows of \mathbf{H}' were taken, the biplot would be represented in a plan (*biplot-AMMI2*); however, it would explain only 54.40% of the SS_{SCA} . Finally, the \hat{S}_{ij}^{AMMI} values predicted by the SVD 3-approximation of the \mathbf{B} matrix are expressed by:

$$\hat{\mathbf{B}} = \begin{array}{c} \begin{array}{cccccccccc} \text{M1} & \text{M2} & \text{M3} & \text{M4} & \text{M5} & \text{M6} & \text{M7} & \text{M8} & \text{M9} \end{array} \\ \begin{bmatrix} -5.76 & 11.63 & 2.10 & -1.10 & -0.70 & -14.84 & -5.00 & -0.96 & 14.63 & \text{F1} \\ 11.63 & -17.24 & -4.68 & 14.61 & 3.16 & 9.35 & 7.43 & -4.06 & -20.20 & \text{F2} \\ 2.10 & -4.68 & -10.91 & -17.85 & 10.49 & 25.42 & 5.81 & -6.06 & -4.32 & \text{F3} \\ -1.10 & 14.61 & -17.85 & -5.28 & 21.01 & -12.08 & 0.25 & -23.74 & 24.17 & \text{F4} \\ -0.70 & 3.16 & 10.49 & 21.01 & -10.14 & -26.46 & -5.22 & 5.16 & 2.70 & \text{F5} \\ -14.84 & 9.35 & 25.42 & -12.08 & -26.46 & -3.97 & -11.02 & 29.79 & 3.81 & \text{F6} \\ -5.00 & 7.43 & 5.81 & 0.25 & -5.22 & -11.02 & -4.62 & 4.27 & 8.10 & \text{F7} \\ -0.96 & -4.06 & -6.06 & -23.74 & 5.16 & 29.79 & 4.27 & 1.02 & -5.42 & \text{F8} \\ 14.63 & -20.20 & -4.32 & 24.17 & 2.70 & 3.81 & 8.10 & -5.42 & -23.46 & \text{F9} \end{bmatrix} \end{array}$$

Considering the observed high correlation between $\hat{\mathbf{B}}$ and \mathbf{B} ($r = \sqrt{0.74} = 0.86$) it can be said that in this example the error associated with a decision based on \mathbf{B} instead of on $\hat{\mathbf{B}}$ would not be great. It is justified because, under an adequate choice of the number of multiplicative terms (or singular axes), a low magnitude correlation between $\hat{\mathbf{B}}$ and \mathbf{B} is related to the predominance of noise in the \mathbf{B} matrix. Thus, in this situation, it may be wrong to interpret the residues \hat{s}_{ij} as reliable estimates of the true SCA (s_{ij}).

Based on the $\hat{\mathbf{B}}$ matrix the following crosses were outstanding among the more favorable combinations: 6x8, 3x6, 4x9 and 4x5 (female x male or vice-versa). Also, note that in the case of complete diallels and under the assumption of no reciprocal effects, the SCA prediction matrix is also symmetric. Further, the SCA of each parent with itself (s_{jj}) is estimated, allowing inferences on varietal heterosis (Cruz and Vencovsky, 1989). The majority of these values were negative, indicating that this type of cross is not adequate for hybrid extraction. The parents identified by the numbers 8, 6 and 7, in this order, should show greater varietal heterosis than those of the numbers 9 and 2. Further note that, in spite of the arbitrariness of the assignment of zero values for the main diagonal elements of \mathbf{B} , the analysis was able to predict the distinct reaction of the parents when crossed with themselves. This opens the perspective of predicting the SCA from unavailable crosses, avoiding the need to assess all possible single crosses.

The possibility of predicting the hybrid performance of non-realized crosses comes from the fact that the analysis deals with the interaction (SCA) not only as a property of each parental combination, but also in an integrated manner, as a phenomenon determined together by rows and columns of the \mathbf{B} matrix. However, if the major interest is for this type of

prediction, it is recommended that one proceeds interactively according to Gauch (1992). Furthermore, to predict the genotypic responses it is necessary to take into account the main effects associated with each parent. Charcosset et al. (1993) used models with multiplicative interaction terms (AMMI analysis) with a similar purpose of predicting the performance of line hybrids in maize, when the assessment of all the single crosses were not possible. The authors concluded that the models offer the possibility of improving prediction efficiency, without supplementary measurements. The problem could also be treated by the mixed models approach, obtaining the BLUP (*best linear unbiased predictor*) of the s_{ij} of missing hybrids (Bernardo, 1996; André, 1999), followed by the SVD adjustment of these predictions.

The coordinates of each parent, used to construct the graphics, are summarized in Table 2. A SAS program to execute the analysis is freely available from the web-address: www.agro.ufg.br.

The relationships commented upon may be seen by looking at the corresponding biplot. Figure 1, although bringing all the biplot-AMMI3 information, does not allow a perfect assessment of the relative distances among the parents because it is a three dimensional graph projected on the paper plane. Thus, for a more rigorous inspection, it is convenient to project the points of the 3-D space in at least two complementary planes, for example, *IPCA1 vs. IPCA2* and *IPCA1 vs. IPCA3* (Figure 2). Therefore, the closeness between two points (genotypes) must always be confirmed by inspection in these principal planes.

Figures 1 and 2 summarize the potential of the different parental combinations to produce hybrids. These graphics do not explain the whole SCA value. However, according to the AMMI analysis concept (as in PCA), the true pattern of interaction (SCA) is

Table 1. The partition of SS_{SCA} (specific combining ability sum of the squares) by singular decomposition, proportion retained in each singular axis and accumulated percentage up to k^{th} axis, in a diallel AMMI analysis.

$IPCA_k$	SS_{IPCA_k}	SS_{IPCA_k}/SS_{SCA}	% Accumulated
1	2978.8753	0.3251	32.5053
2	2006.2649	0.2189	54.3975
3	1810.3817	0.1975	74.1523
4	1500.2918	0.1637	90.5234
5	678.5611	0.0740	97.9278
6	55.5711	0.0170	99.6254
7	32.8437	0.0036	99.9838
8	1.4862	0.0002	100.0000
Total	9164.2760	1.0000	-

described mainly by the first singular axes. Each axis is a linear combination of the rows and columns (vectors) which form the B matrix. As all these vectors contain \hat{s}_{ij} values, the axes represent SCA linear functions constructed with all the matrix data. Finally, as the successive axes capture ever less pattern and more noise, discarding the latter becomes necessary to avoid the incorporation of substantial errors in the description of the interaction.

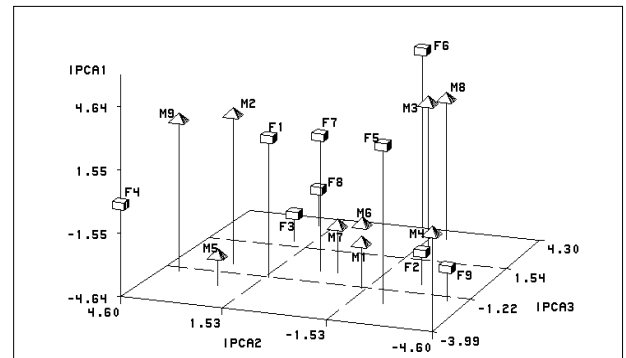
Conscious of these concepts and observing the pairs of points corresponding to each female and male combination, the proximity between F6/M3, F6/M8, F4/M9 and F4/M5 can be depicted. This indicates that such combinations are promising for single cross production in this set of parents. A strong characteristic in all of them is the high magnitude and same sign coordinates for the genotypes of each pair. On the other hand, it can be deduced that the genotypes with good GCA should have low magnitude coordinates (small interaction) as is the case of parents 1 and 7 (Figure 1). These lines are shown to have high GCA in Griffing analysis (1956).

Among the combinations with lower potential for hybrid production are: F2/M9, F6/M5 and F4/M8, all showing great distance among the respective F/M points in the Figures 1 and 2. Further, the low potential is also observed for crosses among the same genotype (F6/M6, F9/M9, etc.), which always show points in opposite position. It is interesting to note that the F3xM5 cross although shown as fairly favorable in

Table 2. Coordinates of the parents for the three first principal axes (IPCA1, IPCA2 and IPCA3) of a biplot graphic display, in a diallel AMMI analysis.

Point	Female/Male	IPCA ₁	IPCA ₂	IPCA ₃
1	F1	2.3197	1.4001	-1.2579
2	F2	-3.0311	-2.8749	-0.4565
3	F3	-3.2249	2.0848	1.9591
4	F4	-0.1408	4.5982	-3.9855
5	F5	3.0481	-2.4802	-2.3030
6	F6	4.6443	-0.9625	4.3046
7	F7	2.1540	0.2680	-0.2958
8	F8	-2.7205	2.0172	3.5341
9	F9	-3.0488	-4.0507	-1.4990
10	M1	-2.3197	-1.4001	-1.2579
11	M2	3.0311	2.8749	-0.4565
12	M3	3.2249	-2.0848	1.9591
13	M4	0.1408	-4.5982	-3.9855
14	M5	-3.0481	2.4802	-2.3030
15	M6	-4.6443	0.9625	4.3046
16	M7	-2.1540	-0.2680	-0.2958
17	M8	2.7205	-2.0172	3.5341
18	M9	3.0488	4.0507	-1.4990

Figure 1. Biplot AMMI3 for specific combining ability (SCA) in maize grain yield of a diallel cross; the cubes identify female genotypes (F), and the pyramids, the male genotypes (M).



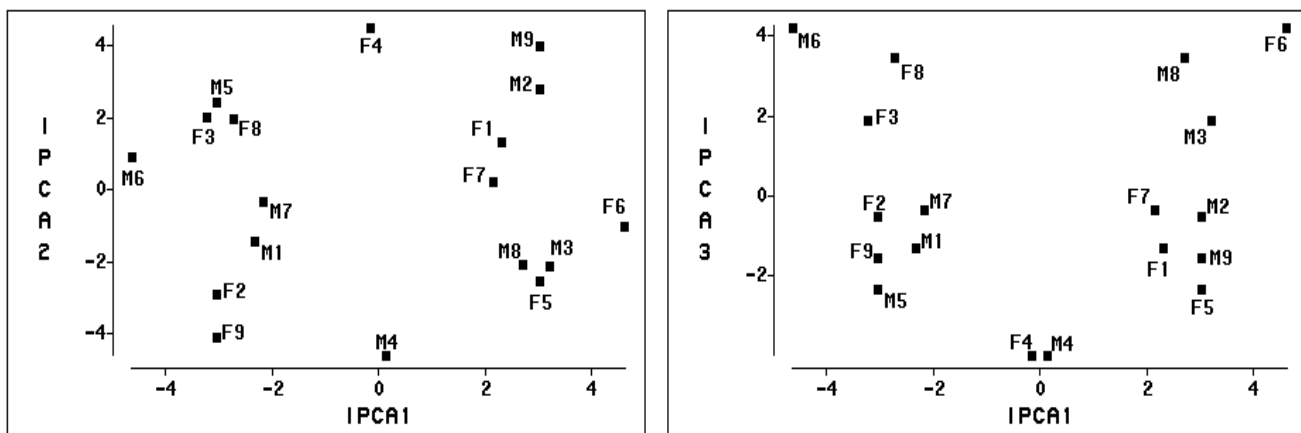
the original diallel table, was shown to be only reasonable by the present analysis. This happened because the AMMI method discards part of the information contained in the original matrix as of predominantly erratic in nature. The option for the biplot-AMMI3 implied in disregarding 26% of the SS_{SCA} . These results can be confirmed in the \hat{B} matrix for \hat{s}_{ij}^{AMMI} prediction.

When only one kind of marker is taken, for example the F points (or the M points), the divergences and similarities among the genotypes which make up the parent group can be assessed. Thus the divergence between genotypes 3 and 6 is notorious, as well as those between 4 and 9. However, it is also evident the similarity between the genotypes 3 and 8, and between 2 and 9 (Figures 1 and 2). Observe also that the first principal axis (IPCA1), capturing 32% of the SCA, is strongly determined by the divergence among two groups of genotypes (2, 3, 8, 9) and (1, 5, 6, 7). The two best hybrid combinations found resulted exactly from the cross between genotypes belonging to distinct groups (3x6 and 6x8). The two other axes, each explaining about 20% of the referred variability, relate more with intragroup genotypic divergences. Therefore, the biplot display was shown to be an additional tool to guide the formation of heterotic groups and the planning of future crosses.

CONCLUSIONS

The adoption of biplot-AMMI to represent the SCA allows a quick visualization of the favorable and unfavorable genotypic combinations for hybrid production. The use of this graphic display also allows a quick and easy assessment of the genotypic divergences and similarities among the parents

Figure 2. Two first principal planes for specific combining ability (SCA) in maize grain yield of a diallel cross; F (females) identify the genotypes in the rows and M (males) the genotypes in the columns of the diallel table.



involved in the diallel crosses. Further, the possibility of discarding an additional noise may guarantee improvements in estimates of SCA.

Considering the integrated treatment which the AMMI analysis offers to the data, it is possible to predict the SCA, even for hybrid combinations not assessed experimentally. This opens the perspective of incorporating this method of analysis, associated with biplot representation, in the routine procedures of prediction of hybrid means in breeding programs.

RESUMO

Representação gráfica *BILOT-AMMI* para a capacidade específica de combinação

No melhoramento para a obtenção de híbridos, uma propriedade genética de grande interesse é a capacidade específica de combinação (CEC). Através dela é possível avaliar as combinações de genótipos mais promissoras à hibridação. O presente trabalho tem o objetivo de apresentar os fundamentos da representação *biplot*, associada à análise *AMMI* (*additive main effects and multiplicative interaction*), bem como sua aplicação aos estudos genéticos dessa natureza. A análise permite uma decomposição da soma de quadrados dos elementos da tabela (matriz) dialélica, a soma de quadrados para CEC, de maneira a captar uma porção sistemática e descartar um resíduo adicional ao erro do experimento. A parte sistemática ou *padrão* é, então, representada graficamente por meio do *biplot*. Neste gráfico pode-se identificar as combinações de genitores com maior potencialidade para a produção de híbridos, bem como avaliar a divergência entre os parentais, em termos de CEC. Um exemplo numérico

é apresentado com instruções computacionais para a sua implementação.

REFERENCES

- André, C.M.G. 1999. Avaliação da melhor predição linear não tendenciosa BLUP associada ao uso de marcadores moleculares na análise dialélica. M.S. Thesis. Universidade Federal de Lavras, Lavras.
- Bernardo, R. 1996. Best linear unbiased prediction of maize single-cross performance. *Crop Science*. 36:50-56.
- Charcosset, A.; Denis, J.B.; Lefort-Buson, M. and Gallais, A. 1993. Modelling interaction from top-cross design data and prediction of F1 hybrid value. *Agronomie*. 13:597-608.
- Cruz, C.D. and Vencovsky, R. 1989. Comparison of some methods of diallel analysis. *Revista Brasileira de Genética*. 12:425-438.
- Duarte, J.B. and Vencovsky, R. 1999. Interação genótipos x ambientes, uma introdução à análise AMMI. Série Monografias, n.9. Sociedade Brasileira de Genética, Ribeirão Preto.
- Eckart, C. and Young, G. 1936. The approximation of one matrix by another of lower rank. *Psychometrika*. 1:211-218.
- Gabriel, K.R. 1978. Least squares approximation of matrices by additive and multiplicative models. *J. R. Statist. Soc.* 40:186-196.
- Gabriel, K.R. 1971. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*. 58:453-467.
- Gardner, C.O. and Eberhart, S.A. 1966. Analysis and

interpretation of the variety cross diallel and related populations. *Biometrics*. 22:439-452.

Gauch, H.G. 1990. Full and reduced models for yield trials. *Theoretical Applied Genetics*. 80:153-160.

Gauch, H.G. and Zobel, R.W. 1996. AMMI analysis of yield trials. p.85-122. In: Kang, M.S. and Gauch, H.G.(Eds.). *Genotype by environment interaction*. CRC Press, Boca Raton.

Gauch, H.G. and Zobel, R.W. 1988. Predictive and postdictive success of statistical analysis of yield trials. *Theoretical Applied Genetics*. 76:1-10.

Gauch, H.G. 1992. *Statistical analysis of regional yield trials: AMMI analysis of factorial designs*. Elsevier, New York.

Gollob, H.F. 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika*. 33:73-115.

Good, I.J. 1969. Some applications of the singular decomposition of a matrix. *Technometrics*. 11:823-831.

Griffing, B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. *Austr. J. Biol. Sci.* 9:463-493.

Mandel, J. 1971. A new analysis of variance model for non-additive data. *Technometrics*. 13:1-18.

Zobel, R.W.; Wright, M.J. and Gauch, H.G. 1988. Statistical analysis of a yield trial. *Agronomy Journal*. 80:388-393.

Received: August 20, 2001;

Accepted: May 23, 2002.

