

Characterization of maize populations in different environmental conditions by means of Three-Mode Principal Components Analysis

Marta Marticorena^{1, 2}, Sergio Bramardi^{1, 3}, and Raquel Defacio⁴

¹Facultad de Ciencias Agrarias, Universidad Nacional del Comahue, Ruta 151, CC 85, Cinco Saltos (8303), Río Negro, Argentina.

²Facultad de Economía y Administración, Universidad Nacional del Comahue, Buenos Aires 1400(8300), Neuquén, Argentina.

³Facultad de Ciencias Agrarias y Forestales, Universidad Nacional de La Plata, Av. 60 e/ 118 y 119 (1900), La Plata Buenos Aires, Argentina.

⁴Estación Experimental Agropecuaria. INTA, Pergamino, Ruta 32, Km 4,5 (2700), Pergamino, Buenos Aires, Argentina.

Abstract

M. Marticorena, S. Bramardi, and R. Defacio. 2010. Characterization of maize populations in different environmental conditions by means of Three-Mode Principal Components Analysis. Cien. Inv. Agr. 37(3): 91-103. Characterization of 31 native populations of maize conserved at the germplasm bank of the INTA Pergamino Experimental Station, Argentina, was achieved by evaluating 10 quantitative attributes in two different environmental situations. The experimental design generated three-way or three-mode data, repeated observations of a set of attributes for a set of individuals in different conditions. The information was displayed in a three-dimensional array, and the structure of the data was explored using Three-Mode Principal Component Analysis, the Tucker-2 Model. A group of populations was identified that displayed homogeneous behavior in the two environments with respect to the following traits: ear length, prolificacy, grains per meter, yield and 1000 kernel weight. However, another group of populations displayed opposite behavior for the traits of plant height and ear insertion height in the different environment conditions and is indicative of the existence of genotype-environment interaction. In conclusion, Three-Mode Principal Component Analysis is an important tool for characterizing plant genetic resources when their phenotypic values are likely to be affected by environmental conditions.

Key words: Germplasm Bank, genotype-environment interaction, Tucker-2 Model.

Introduction

A crucial step in the process of cataloguing and characterizing the material conserved in

a germplasm bank is to determine the relationships or associations among the taxa (Bramardi, 2000). Depending on the type of traits studied, there are different types of characterizations (Avisé, 2004). In methods where the characterization is supported by the study of phenotypic, morphologic and agronomic descriptors and where there may be an influence from environmental factors (Pan *et al.*, 2004), the traits must be analyzed in different environments.

This leads to an analysis of the variability of the factor and the possibility of variability due to a genotype-environment interaction in the variables observed (Demey, 2008).

These experimental situations generate three-way or three-mode data, repeated observations of a set of variables for a set of individuals, in different conditions. Although it is possible to apply classical methods of multivariate analysis to two-way data matrices generated from each condition, the results from those methods would necessitate explaining the structure of each table separately and the relations and interactions between any of the studied variables, for each occasion. However, it is not feasible to simultaneously study the individual similarities and differences of the different tables or the interactions among the three modes analyzed with this approach (Kroonenberg, 1983).

The third mode in the design requires the use of representative structures and the application of more appropriate methodologies. A three-dimensional array or three-way table (Hitchcock, 1927; Comon, 2001) and the matricial forms of that array (Kolda, 2001 and 2006; Kolda, *et al.*, 2005) allow for an adequate representation of the three-way data (Bolasco, 1986; Coppi, 1986; Kroonenberg, 1983). One method for the study of data with a three-way structure is the Three-Mode Principal Component Analysis. One of the models for this analysis, proposed by Tucker (1966) and modified by Kroonenberg and De Leeuw (1980) regarding the estimation of their parameters, is the Tucker-2 Model. This method includes the reduction of dimensionality of two of the three modes, which not only enables an estimation of the global variability of the data but also provides information on the behavior of the variables or individuals in the different experimental conditions or situations.

The relevance of the maintenance of the genetic variability conserved in germplasm banks depends on the measurement and characterization of that diversity (Cordeiro *et al.*, 2003). Although classifications based on agromorphologic markers using quantitative traits allow a rapid discrimination of the phenotypes (Lowe *et*

al., 1996) with a high capacity for separating taxa (Alfaro and Segovia, 2000), they are highly susceptible to the influence of the environment (Pan *et al.*, 2004).

The goal of this study is to use the Tucker-2 method as a methodological alternative for the study of native populations of maize, from the province of Buenos Aires, Republic of Argentina, conserved in the germplasm bank of INTA Pergamino Experimental Station. Morphologic characteristics have been observed in characterization tests that have been repeated in different environmental situations. The application of this technique allows for (i) the analysis of the relationships among maize populations and the variables observed in the different environments, as a whole, as well as for each population separately; (ii) a study of the variability caused by environmental conditions and the possible variability due to the genotype-environment interaction on the variables observed; (iii) the identification of the populations interacting with the environment from the graphic representations (interactive and cluster Biplots); and (iv) obtaining a characterization of the 'medium', 'average' or 'consensus' populations from the environments or situations under study.

Materials and methods

The analyzed data corresponded to observations made on 31 native populations of maize (the Orange Flint race) from the Province of Buenos Aires, Argentina. These populations are part of a collection of local populations corresponding to different races that are conserved in the Active Germplasm Bank of INTA Pergamino Experimental Station, Argentina. The populations were coded numerically as follows: 3, 12, 15, 16, 17, 19, 20, 21, 22, 25, 26, 27, 28, 38, 39, 40, 43, 44, 45, 47, 61, 62, 82, 86, 87, 90, 92, 98, 102, 151 and 152.

Ten quantitative variables were evaluated in two tests conducted in the 2003/04 season in the localities Pergamino (EEA INTA) and Ferré (Escuela Agrotécnica Salesiana). These variables allowed for morphologic-agronomic characterization and may be influenced, to a certain extent, by environmental conditions.

The variables were as follows: plant height, higher ear insertion height, ear length, ear diameter, grain width, grain length, prolificacy, yield, grains per meter and 1000 kernel weight. The average minimum and maximum values of the traits in each environment are shown in Table 1.

The information corresponding to each locality, Pergamino (Environment one) and Ferré (Environment two), was transferred into Tables T_k ($k = 1$ or 2). A three-way data array $\underline{X}_{(I \times J \times K)}$ was generated from the matrices T_k , where $I = 31$, $J = 10$ and $K = 2$ are the dimensions of the first mode (I, populations), the second mode (J, traits) and the third mode (K, environments), respectively. The matricial forms of the three-dimensional array were obtained using a process known as ‘matricizing’ (Kiers, 2000) or ‘unfolding’ (De Lathauwer, *et al.*, 2000). These matricial forms, the result of concatenating two of the modes (Kroonenberg, 1983) are the extended matrices as follows: $X_{1,2 \times 3}$ of the order (31×20) , in which the lines represent the populations and the columns represent the concatenation of the traits with the environments; $X_{2,3 \times 1}$ of the order (10×62) , in which the lines represent the traits

and the columns represent the concatenation of the environments with the populations; and the matrix $X_{3,1 \times 2}$ of the order (2×310) , in which the lines represent the environments and the columns represent the concatenation of the populations with the traits.

Three-Mode Principal Component Analysis: Tucker-2 Model

The statistical analysis of the information was conducted using the Tucker-2 model. This model for the third unreduced mode was formulated as a factorization of the three-way data array $\underline{X} = \{x_{ijk}\}$,

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q a_{ip} b_{jq} g_{pqk}, \quad i = 1, \dots, 31; \quad j = 1, \dots, 10; \quad k = 1, 2$$

In this equation, P and Q indicate the number of components retained in the first and second modes, respectively; a_{ip} and b_{jq} are the elements of the matrices; and real, orthonormal components $A_{(I \times P)}$ and $A_{(J \times Q)}$ and \underline{g}_{pqk} are the elements of the three-way array, termed in this model as the “matrix of extended links”.

Table 1. Agronomic-morphologic characteristics: mean, minimum and maximum values in each environment.

Character	Environment one (Pergamino)			Environment two (Ferré)		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
AGR1	8.50	6.00	9.80	8.80	7.20	10.60
ALTMZ2	97.20	68.00	121.50	85.10	64.00	111.50
ALTPL3	163.70	122.70	193.00	144.10	118.50	167.50
DMZ4	41.60	33.40	47.00	42.00	34.50	48.10
GRXM5	31.40	21.60	39.80	35.00	29.90	40.00
LGR6	8.10	5.40	12.00	10.10	8.00	12.00
LMZ7	16.10	13.10	19.50	16.60	14.00	19.40
PESO8	286.80	188.00	355.00	287.20	208.00	346.00
PROL9	0.97	0.55	1.36	1.00	0.66	1.33
REND10	5483.00	1341.00	9440.00	6777.00	2614.00	10827.00

¹AGR: grain width (mm). ²ALTMZ: height of ear insertion (cm). ³ALTPL: height of the plant (cm). ⁴DMZ: ear diameter (mm). ⁵GRXM: grains by meter. ⁶LGR: grain length (mm). ⁷LMZ: ear length (cm). ⁸PESO: weight of 1000 grains (g). ⁹PROL: prolificacy (ears/plant). ¹⁰REND: yield (kg/ha).

The axis value to be retained in the first and second mode, P and Q , was determined using the Diffit (Timmerman and Kiers, 2000). For the third mode, $K=2$. Nine possible solutions are presented in the first three columns of Table 2.

Once the number of axis was fixed through the Tucker-2 algorithm (Kroonenberg and De Leeuw, 1980), which allows one to simultaneously approximate the extended matrices, \mathbf{A} , \mathbf{B} and \mathbf{C} were obtained. The algorithm consists of an iterative process whose steps are as follows:

Step 1: Initial solution: $(\mathbf{A}_0, \mathbf{B}_0, \mathbf{C}_0)$.

\mathbf{A}_0 : P first columns of the matrix of eigenvectors of $\mathbf{X}_{1,2 \times 3} \mathbf{X}'_{1,2 \times 3}$

\mathbf{B}_0 : Q first columns of the matrix of eigenvectors of $\mathbf{X}_{2,3 \times 1} \mathbf{X}'_{2,3 \times 1}$

\mathbf{C}_0 : Identity matrix of order 2.

Step 2: \mathbf{A}_1 is obtained from P first eigenvectors of the matrix,

$$[\mathbf{X}_{1,2 \times 3} (\mathbf{B}_0 \otimes \mathbf{C}_0)] [\mathbf{X}_{1,2 \times 3} (\mathbf{B}_0 \otimes \mathbf{C}_0)]'$$

The solution is tested. (See the criteria for convergence*)

Step 3: If the previous triad does not provide a

solution, \mathbf{B}_1 is obtained from the Q first vectors of the matrix

$$[\mathbf{X}_{2,3 \times 1} (\mathbf{C}_0 \otimes \mathbf{A}_1)] [\mathbf{X}_{2,3 \times 1} (\mathbf{C}_0 \otimes \mathbf{A}_1)]'$$

The solution is tested $(\mathbf{A}_1, \mathbf{B}_1, \mathbf{C}_0)$, if not, the process begins the process again calculating \mathbf{A}_2 from \mathbf{B}_1 and \mathbf{C}_0 and so on until the convergence is reached.

*Criteria for convergence: The solution must involve the values of \mathbf{A} and \mathbf{B} stabilized for a real positive value in which ϵ is arbitrarily small:

$$\|\mathbf{A}_{i+1} - \mathbf{A}_i\| \leq \epsilon; \|\mathbf{B}_{i+1} - \mathbf{B}_i\| \leq \epsilon;$$

Once \mathbf{A} and \mathbf{B} are found, \mathbf{C} is calculated in their extended matricial form for the first mode, $\mathbf{C}_{1,2 \times 3} = \mathbf{A} \mathbf{X}_{1,2 \times 3} (\mathbf{B} \otimes \mathbf{C})'$. $\mathbf{C}_{2,3 \times 1}$ and $\mathbf{C}_{3,1 \times 2}$ are calculated for symmetry.

Biplot representations

A Biplot (Gabriel, 1971) is a graphical representation of multivariate data. The essential characteristic that differentiates it from the graphical representations associated with classic methods of dimension reduction is that, in this

Table 2. Adjustment difference and increase for the preselected solutions.

n	P_n^1	Q_n^2	K_n^3	S_n^4	Fit $_n^5$ (%)	DiffFit $_n^6$	Incr $_n^7$
1	2	1	2	5	27.6803	27.6803	1.9237
2	2	2	2	6	37.5193	9.8390	-
3	3	3	2	8	51.9079	14.3886	1.2608
4	3	4	2	9	55.1103	3.1024	-
5	4	4	2	10	63.3191	11.4117	1.6710
6	5	4	2	11	67.9710	4.6519	-
7	5	5	2	12	74.8001	6.8291	1.2426
8	7	4	2	13	77.4962	2.6961	-
9	7	5	2	14	82.9918	5.4956	

¹ P_n : number of axes to retain in the first way. ² Q_n : number of axes to retain in the second way. ³ K_n : amount of axes to retain in the third way. ⁴ $S_n = P_n + Q_n + K_n$. ⁵Fit $_n$ (%): percentage of variation explained by the solution ($P \times Q \times K$). ⁶DiffFit $_n$: difference between each adjustment with the previous one (DiffFit $_n - \text{DiffFit}_{n-1}$). ⁷Incr: increase in variation (DiffFit $_n / \text{DiffFit}_{n-1}$).

case, a joint representation of lines and columns of a data matrix is possible. The Biplot interpretation is based on very simple geometric concepts. In general terms, the similarity between ‘individuals’ (lines) is an inverse function of the distance between the points representing them. The lengths and angles of the vectors representing the ‘variables’ are interpreted in terms of variability and covariability, respectively. The relations between the lines and columns are interpreted in terms of dot products, that is, in terms of the orthogonal projections of the point’s “line” on the vector’s “column”.

Carlier and Kroonenberg (1996) showed that the graphic representation of the Tucker models could be carried out using the Biplot methods proposed by Gabriel (1971). They formulated two types of Biplot representations for data with three-way structure: the Interactive Biplot (or the Biplot with multiplicative structure) and the joint Biplot. To construct these Biplots, it is necessary to utilize the three-way decomposition associated with the Tucker models.

The interactive Biplot (Bradru and Gabriel, 1978; Cox and Gabriel, 1982; Van Eeuwijk and Kroonenberg, 1998) consists of combining two of the modes (J and K), obtaining markers (a_j) for the individuals (lines) and markers (d_{jk}) for the variables that were concatenated with the conditions (columns). Therefore, this graphic reflects different behaviors among these elements that correspond to the results in the analysis of the link matrix. The interactive Biplot allows for the visualization of the inter-structure of the different data tables.

The joint Biplot (Kroonenberg, 1983) is a Biplot conditional to one of the modes; through the plane \mathbf{G}_r the axes of the joint Biplot are rescaled according to their relative importance. Therefore, it provides a faithful visual interpretation of the intra-structure of each Table.

The algorithm to obtain the Tucker-2 model program and the results are elaborated on in the software MABLAB (Version 5, U.S.A.). The results are modifications of the results

presented by Varela (2002) and Baccalá (2004).

Results and discussion

Selection of the number of axes

The algorithm requires that the value of the axis to be retained in each mode is determined *a priori*, the sum of axes to be retained in each mode (S_n), the difference between a previous fit and a further (DiffFit_n) and the increase in the variation are presented in Table 2 along the nine possible solutions resulting from the application of the DiffFit method. The solutions ($2 \times 2 \times 2$), ($3 \times 4 \times 2$), ($5 \times 4 \times 2$) and ($7 \times 4 \times 2$) were discarded, as they do not verify $\text{DiffFit}_{n+1} < \text{DiffFit}_n$ for $S_{n+1} > S_n$. Among the remaining solutions, the solution ($3 \times 3 \times 2$) was chosen because the relationship between the percentage of variation explained (51.9%) and the low- dimension of the modes was considered acceptable.

Extended Link Matrix

The frontal planes of the extended link matrix contain the information corresponding to each environment. The same is shown in matrices \mathbf{G}_1 and \mathbf{G}_2 (Table 3). It is feasible to infer that in both environments, the most relevant relationships are positive and result along the first components of the populations and the traits (\mathbf{g}_{1IK}), unlike the moderate and opposite relations among the third components of the two modes (\mathbf{g}_{23K}), as well as between the second component of the populations and the third of the variables (\mathbf{g}_{23K}). This may be deduced from the absolute values and the signs of the homologous elements of the matrices.

Table 4 shows \mathbf{g}_{pqk}^2 and the proportion of the variation explained by each combination of components. The element \mathbf{g}_{pqk}^2 of \mathbf{G} represents the intensity of the relationship between the p-th component of the first mode and the q-th component of the second mode in the k-th condition, and it indicates the extent to which the total variation is explained by the combination of the components p and q for the condition k. It is clear that the most important relationships result among the first

Table 3. Frontal planes of the extended core matrix.

$\tilde{G}_1 \{ \tilde{g}_{pq1} \}^1$			$\tilde{G}_2 \{ \tilde{g}_{pq2} \}^2$		
10.2607	1.8486	3.5531	8.3808	3.0217	-0.6006
-2.0151	3.9903	3.3061	-1.3223	3.3322	-4.2687
-1.2908	3.3546	-3.4303	1.3253	3.3413	3.2864

^{1,2} \tilde{g}_{pqk} : for k=1,2; relationships between the pth components of the populations and the qth components of the characters in environment k.

Table 4. Proportion of variation explained by each combination of components.

Environment one (Pergamino)					
\tilde{g}_{pq1}^2			Proportion of variation explained		
105.2820	3.4172	12.6245	0.1754	0.0057	0.0210
4.0606	15.9224	10.9303	0.0068	0.0265	0.0182
1.6661	11.2533	11.7669	0.0028	0.0188	0.0196
Environment two (Ferré)					
\tilde{g}_{pq2}^2			Proportion of variation explained		
70.2378	9.1307	0.3607	0.1171	0.0152	0.0006
1.7485	11.1035	8.2218	0.0029	0.0185	0.0304
1.7564	11.1643	10.8004	0.0029	0.0186	0.0180

^{1,2} \tilde{g}_{pqk}^2 : for k=1,2; variation explained by the combination of component p and q for the environment k.

components of the populations and variables in both environments. It is therefore evident that the first direction is common for both environments.

The percentages of variation indicate the most important combinations of components, but not the tendency of the relationships. This information results from the signs of the elements of the extended link matrix; the signs of the heavier elements in the components of the matrices **A** and **B**. For the elements \tilde{g}_{pqk} with a positive sign, the combinations (a_{ip}, b_{jq}), whose result is also positive, will show higher values

of positive relations when they are compared with the combinations of negative categories in the same combination of components (which will show the highest negative relations) as it may be visualized in the Interactive Biplot (Figure 1).

The graphic representations

The Interactive Biplot, Figure 1, provides the following information: with the exception of the plant height and the ear insertion height

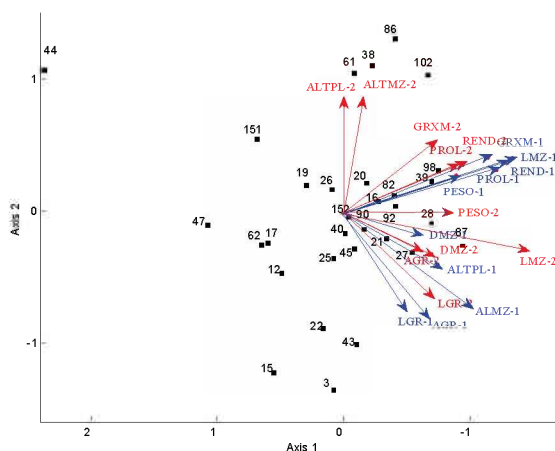


Figure 1. Interactive Biplot. Plane 1-2. Markers for the populations: represented by points. Markers for characters concatenated with environment: represented by vectors. In the environment k, for $k=1,2$: AGR-k: grain width (mm); ALTMZ-k: height of ears insertion superior (cm); ALTPL-k: height of the plant (cm); DMZ-k: ears diameter (cm); GRXM-k: grains by meter; LGR-k: grain length (mm); LMZ-k: ears length (cm); PESO-k: weight of 1000 grains (g); PROL-k: prolificacy (ears/plant); REND-k: yield (g).

of environment two, the rest of the traits for both localities contribute to the formation of the first axis of the inter-structure; this indicates that there is no interaction between the environment and these variables. In the first two localities, populations 28, 39, 87, 39, 98 and 102 show very high values for the traits evaluated, which are clearly different from other populations, such as 17, 44, 47, 62 and 151. The traits for grains per meter, yield and prolificacy show a higher contribution from environment one, and ear length is a trait common to both environments. Regarding the last variable, although the angle forming the vector in environment one corresponds with that of environment two, it is comparatively larger than the other variables, but insufficiently large to be considered a significant difference. However, the angle formed between the vectors of the variables plant height and ear insertion height, corresponding to the different environments, is important. In both cases the vectors are approximately straight; this indicates an unstable behavior regarding the environmental conditions. It is therefore clear that the populations do not have a common structure regarding these variables in the two localities. In environment two, populations 38, 61, 86 and 102 present a higher plant height and ear insertion height than the rest of the popula-

tions, while the values in environment one are medium to low. Likewise, populations 3, 15, 22 and 43, where the highest values are recorded in environment one, presented the lowest values into environment two for both variables. This clearly shows a genotype-environment interaction for plant height and ear insertion. On the other hand, the populations with higher variability are farther from the center of the coordinates in regard to populations.

The entries of accessions 28, 39, 87 and 98 present the highest values in both environments for all of the variables, except for the traits plant height and ear insertion. Populations 39 and 98 show high values for these traits in both environments, while accessions 28 and 87 show the highest values of all of the populations studied with regard to plant height and ear insertion in environment one, but in environment two the values for these two variables were medium.

Population 44 presents the lowest values among all of the entries for all of the traits studied in both environments, with the exception of plant height and ear insertion in environment two. A similar situation is observed in population 151, although the values are not as extreme in comparison to population 44.

Populations 12, 17, 47 and 62 present the lowest values (except for population 44) in grains per meter, yield, prolificacy, 1000 kernel weight and ear length in both environments. Population 47 presents the lowest values within this group. Additionally, low values are observed for plant height and ear insertion in both environments, in the four entries, where the values for population 12 are lower in environment two than in environment one. Populations 17 and 62, on the contrary, have lower values in environment one than in environment two, and entry 47 presents in environment one the lowest values of the total of all of the populations studied. Populations 17, 47 and 62 have low values in ear diameter, grain length and grain width in the two environments. Entry 12 has low values in ear diameter and medium values in grain length in both environments.

Entries 38, 61, 86 and 102 present high values in grains per meter, yield, prolificacy, ear length and 1000 kernel weight in both environments, and the values of population 102 are the highest of all of the populations studied. This group of populations has low values for ear diameter, grain length and grain width in both environments and in plant height and ear insertion in environment one. These two traits present the highest values of all of the populations in environment two.

Populations 3, 15, 22 and 43 have low values in grains per meter, yield, prolificacy and 1000

kernel weight in both environments and ear length in environment one. The values for population 15 are among the lowest values, presenting a low value for ear length in environment two. The four entries have high values for ear diameter, grain length and grain width in both environments, as well as for plant height and ear insertion height in environment one. On the other hand, they show the lowest values among all of the populations in plant height and ear height in environment two. Accessions 3 and 43 have high values for ear length in environment two, and entry 22 has a medium value.

The categories of populations and traits are represented in the Joint Biplots, Figures 2 and 3, projected on the components of each of the environments. In environment one, all of the variables, except grain width and grain length, contribute to the formation of the first axis of variability. The traits grains per meter, prolificacy, ear length and yield are highly correlated and demonstrate the highest contribution. To a lesser extent, these four variables are positively associated with ear insertion height (another variable with a high contribution to axis one), and they are nearly independent from grain width and grain length (which show the highest contribution to the formation of the second axis). Therefore, they are essential for differentiating populations regarding this axis.

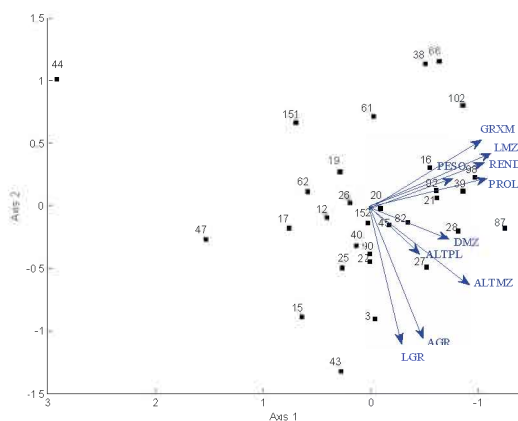


Figure 2. Joint Biplot – Environment one. Plane 1-2. Markers for the populations: represented by points. Markers for characters: represented by vectors. AGR: grain width (mm); ALTMZ: height of ears insertion superior (cm); ALTPL: height of the plant (cm); DMZ: ears diameter (cm); GRXM-k: grains by meter; LGR: grain length (mm); LMZ: ears length (cm); PESO: weight of 1000 grains (g); PROL: prolificacy (ears/plant); REND-k: yield (g).

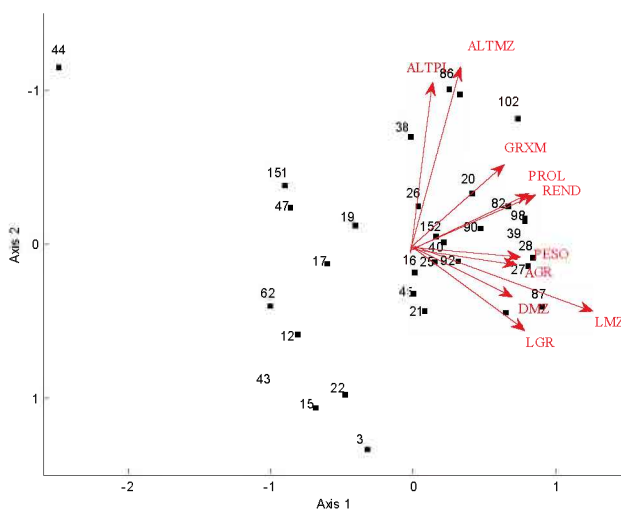


Figure 3. Joint Biplot – Environment two. Plane 1-2. Markers for the populations: represented by points. Markers for characters: represented by vectors. AGR: grain width (mm); ALTMZ: height of ears insertion superior (cm); ALTPL: height of the plant (cm); DMZ: ears diameter (cm); GRXM-k: grains by meter; LGR: grain length (mm); LMZ: ears length (cm); PESO: weight of 1000 grains (g); PROL: prolificacy (ears/plant); REND-k: yield (g).

In environment two, except for plant height and ear insertion height, the rest the variables influence the formation of the first axis. Additionally, ear length has a larger contribution, and it is practically independent of plant height and ear insertion height, both variables that contribute to forming the second axis. Therefore, these two variables present the highest influence in forming population clusters in the two axes.

When the three graphics presented are compared, taking into account the correlations between the populations distances on the correlation corresponding to the interactive Biplot and the distances observed in the Joint Biplots (representing the configurations from each environment, as shown in Table 5) the Interactive Biplot may be considered as a ‘consensus’ or ‘average’ configuration.

Table 5. Correlations between configurations of Joint Biplots: A1 and A2, and Interactive Biplot: A1-A2.

	A1	A2	A1-A2
A1	1.00	0.77	0.91
A2	0.77	1.00	0.90
A1-A2	0.91	0.91	1.00

From the results presented, it may be deduced that the environments present a similar structure in the first axis, where the highest variability is projected. On the contrary, the behavior is opposite in the combination of the second axis of the populations and the third axis of the variables. This indicates that the populations and the heavier traits in these components are the most strongly affected by the environment. This situation contrasts with the populations of homogeneous behavior in the two environments and indicates the existence of genotype-environment interaction.

Unlike most works related to the study of genetic diversity, the method proposed provides an integral analysis of three-way structure data and represents an important tool for the research of triple interactions. The Ballot representations induced by the model allow for the visualization of the analytic results and for the identification of individuals and variables interacting with the different environmental conditions.

The techniques more commonly used for characterizations using phenotypical markers based on quantitative variables are the analysis of Three-Mode Principal Component Analysis

and Cluster Analysis (Alika *et al.*, 1993; Badea *et al.*, 2008; Brandolini and Brandolini, 2001; Gouesnard *et al.*, 1997; Lucchin *et al.*, 2003; Nghia *et al.*, 2008; Ruiz de Galarreta and Álvarez, 2001). The limitation of these methodologies is that they do not allow for the study of the relations among the configurations obtained in the different environments, the information referred to as the genotype-environment interaction.

A wide majority of the procedures analyzing the relationships among characterizations create correlations using matrices of distances and/or similarities (Franco *et al.*, 2001; García *et al.*, 2007; Kalita *et al.*, 2007; Lanza *et al.*, 1997; Tar'an *et al.*, 2005; Syamkumar and Sasikumar, 2007), but these procedures do so without providing information on the variables responsible for the discrepancies or concordances between environments.

The methods using additive and multiplicative models for the simultaneous evaluation the populations studied in different tests, according to quantitative variables and estimating the genotype-environment interaction (Cornelius *et al.*,

1992; Crossa and Cornelius, 1993; Reeb *et al.*, 2007; Taba *et al.*, 1998), do not enable the identification of the populations that are affected by the environment.

Techniques exist for the study of simultaneously different configurations based on the search of a consensus configuration such as Generalized Procrustes Analysis (GPA), STATIS (Structuration des Tableaux A Trois Indices de la Statistique) and Main Meta-Components (Bramardi *et al.*, 2005; Defacio *et al.*, 2006; Defacio *et al.*, 2007; Faccioli *et al.*, 1995; Lorea *et al.*, 2006; Milbourne *et al.*, 1997; Lopes *et al.*, 2006; Esposito *et al.*, 2007) among others. However, none of these techniques includes the projection of the variables responsible for the consensus configuration in the analysis.

In conclusion, the Three-Mode Principal Component Analysis provides both useful analytic and graphic tools to study and characterize phylogenetic resources, especially when this characterization is based on the study of phenotypical, morphologic and agronomic descriptors where the influence of environmental factors is possible.

Resumen

M. Marticorena, S. Bramardi y R. Defacio. 2010. Caracterización de poblaciones de maíz en distintas condiciones ambientales mediante Análisis de Componentes Principales de tres modos. Cien. Inv. Agr. 37(3): 91-103. Se analizaron datos correspondientes a la caracterización de 31 poblaciones nativas de maíz, conservadas en el banco de germoplasma de la Estación Experimental Agropecuaria INTA Pergamino (EEA INTA Pergamino), Argentina, sobre las que se evaluaron 10 variables cuantitativas en dos situaciones ambientales distintas. Esta situación experimental generó datos de tres vías o modos: observaciones de un conjunto de variables sobre un conjunto de individuos repetidas en distintas condiciones. Se concentró la información en un arreglo tridimensional, para luego explorar la estructura del cuerpo de datos empleando "análisis de componentes principales de tres modos", Modelo Tucker-2. Los resultados obtenidos permitieron determinar un grupo de poblaciones de comportamiento homogéneo en los dos ambientes, en cuanto a los caracteres longitud de mazorca, prolificidad, granos por metro, rendimiento y peso de 1.000 granos. En cambio, otro grupo de poblaciones manifestaron la incidencia de las condiciones ambientales, en el comportamiento opuesto reflejado en las variables altura de planta y altura de inserción de mazorca. Esta situación, contrastada con las poblaciones de comportamiento homogéneo en los dos ambientes, es indicativa de la existencia de interacción genotipo-ambiente. Se concluye que el análisis de componentes principales de tres modos es una herramienta importante para la caracterización de recursos fitogenéticos, cuyos valores fenotípicos y agronómicos son susceptibles de ser afectados por condiciones ambientales.

Palabras clave: Banco de Germoplasma, Interacción genotipo-ambiente, Modelo Tucker-2.

References

- Alfaro, Y. and V. Segovia. 2000. Maíces del sur de Venezuela clasificados por taxonomía numérica. I. Caracteres de la planta. *Agronomía Tropical* 50 (3): 413-433.
- Alika J.E., M.E. Aken'Ova, and C.A. Fatokun. 1993. Variation among maize (*Zea mays* L.) accessions of Bendel State, Nigeria. Multivariate analysis of agronomic data. *Euphytica* 66:65-71.
- Avise, J.C. 2004. *Molecular Markers, Natural History and Evolution*. Sinauer Associates. 2nd edition. Sunderland, USA. 684 pp.
- Baccalá, N. 2004. Contribuciones al análisis de matrices de datos multivía: Tipología de las variables. Tesis Doctoral. Universidad de Salamanca. Salamanca, España. 323 pp.
- Badea, A., F. Eudes, R.J. Graf, A. Laroche, D.A. Gaudet, and R.S. Sadasivaiah. 2008. Phenotypic and marker-assisted evaluation of spring and winter wheat germplasm for resistance to fusarium head blight. *Euphytica* 164:803-819.
- Bolasco, S. 1986. Per una teoria sulla costruzione e l'analisi delle matrici a tre modi (Towards a theory on the construction and the analysis of three-mode matrices). In *Proceedings of the 33rd Meeting of the Italian Statistical Society*. Bari, Italy. p. 183-195.
- Bradu, D., and K.R. Gabriel. 1978. The Biplot as a diagnostic tool for models of two-way tables. *Technometrics* 20(1): 47-68.
- Bramardi, S. J. 2000. Estrategias para el análisis de datos en la caracterización de recursos fitogenéticos. Tesis Doctoral, Departamento de estadística, Universidad Politécnica de Valencia, España. 390 pp.
- Bramardi, S.J., G.P. Bernet, M.J. Asíns, and E.A. Carbonell. 2005. Simultaneous Agronomic and Molecular Characterization of Genotypes via the Generalised Procrustes Analysis. An Application to Cucumber. *Crop Science* 45:1603-1609.
- Brandolini A, and A. Brandolini. 2001. Classification of Italian maize (*Zea mays* L.) germplasm. *Plant Genetic Resource Newsletter* 126: 1-11.
- Carrier, A., and P.M. Kroonenberg. 1996. Decompositions and Biplots in Three-way Correspondence Analysis. *Psychometrika* 61(2): 355-373.
- Comon, P. 2001. Tensor decompositions. In: *Mathematics in Signal Processing V*. McWhirter J.G., and I. K. Proudler (eds.). Oxford University Press. Oxford, UK. p. 1-24.
- Coppi, R. 1986. Analysis of Three-Way Data Matrices Based on Pairwise Relation Measures. In: F. de Antoni, N. Lauro and A. Rizzi (eds.). *Physica-Verlag*. Heidelberg, Alemania. p. 129-139.
- Cordeiro, G.M., Y.B. Pan, and R.J. Henry. 2003. Sugarcane microsatellites for the assessment of genetic diversity in sugarcane germplasm. *Plant Science* 165(1):181-189.
- Cornelius, P.L., M. Seyedsadr, and J. Crossa. 1992. Using the shifted multiplicative model to search for "separability" in crop cultivar trials. *Theor. Appl. Genet.* 84:161-172.
- Cox, C., and K.R. Gabriel. 1982. Some comparisons of Biplot display and pencil-and-paper E.D.A. methods. In: Launer R.L., and A.F. Siegel (eds.). *Modern data analysis*. London: Academic Press. p. 45-82.
- Crossa, J., and Cornelius, P. L. 1993. Recent developments in multiplicative models for cultivar trials. In: D.R. Buxton, R. Shibles, R.A. Forsberg, B.L. Blad, K.H. Asay, G.M. Paulsen, and R.F. Wilson (eds.). *International Crop Science I*. Crop Science Society of America, Inc. Madison, USA. p. 571-577.
- Defacio, R.A., N.M. Paz, S.J. Bramardi, M.E. Ferrer, and A.R. Schlatter. 2006. Relación entre datos agro-morfológicos y moleculares en 10 poblaciones locales de maíz. VII Congreso Latinoamericano de Sociedades de Estadística. Rosario, Santa Fé, Argentina. (Resúmenes) p 222. (Actas en CD) 1-6.
- Defacio, R.A., N.M. Paz, F.M. Scropanich, S.J. Bramardi, A.R. Schlatter, and M.E. Ferrer. 2007. Análisis de 40 poblaciones locales de maíz (*Zea mays*) de la raza cristalino Colorado en base a marcadores agro-morfológicos y moleculares. XXXVI Congreso Argentino de Genética. Pergamino. Argentina. (Resúmenes) p S166.
- De Lathauwer, L., B. De Moor, and J. Vandewalle. 2000. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* 21: 1253-1278.
- Demey, J.R. 2008. Diversidad genética en bancos de germoplasma: Un enfoque Biplot. Tesis Doctoral, Departamento de estadística, Universidad de Salamanca, España. 244 pp.
- Esposito, M.A., E.A. Martin, V.P. Cravero, and E. COUNTRY. 2007. Characterization of pea accessions by SRAP's markers. *Scientia Horticulturae* 113(4): 329-335.

- Faccioli, P., V. Terzi, A. Monetti, J. Nicola, N. Pechioni. 1995. B-hordein STS markers for barley genotype identification: comparison with RFLPs, hordein APAGE and morpho-physiological traits. *Seed Science Technology* 23:415-427.
- Franco, J., J. Crossa, J.M. Ribaut, J. Betran, M.L. Warburton, M. Khairallah. 2001. A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. *Theoretical Applied of Genetics* 103:944-952.
- Gabriel, K.R. 1971. The biplot graphic display of matrices with applications to principal components analysis. *Biometrika*. 58, 453-467.
- Garcia, M.V., P.A. Balatti, M.J. Arturi. 2007. Genetic variability in natural populations of *Paspalum dilatatum* Poir. Analyzed by means of morphological traits and molecular markers. *Genetic Resources and Crop Evolution* 54(5):935-946.
- Gouesnard B., J. Dallard, A. Panouillé, and A. Boyat. 1997. Classification of French maize populations based on morphological traits. *Agronomie* 17:491-498.
- Hitchcock, F. L. 1927. Multiple invariants and generalized rank of a p-way matrix or tensor. *Journal of Mathematics and Physics* 7: 39-79.
- Kalita, M.C., T. Mohapatra, X. Dhandapani, D.K. Yadava, K. Srinivasan, A.K. Mukherjee, R.P. Sharma. 2007. Comparative evaluation of RAPD, ISSR and anchored-SSR markers in the assessment of genetic diversity and fingerprinting of oilseed *Brassica* genotypes. *Journal of Plant Biochemistry and Biotechnology* 16(1):41-48.
- Kiers, H. A. L. 2000. Towards a standardized notation and terminology in multiway analysis. *J. Chemometr.* 14: 105-122.
- Kolda, T. G. 2001. Orthogonal tensor decompositions, *SIAM Journal on Matrix Analysis and Applications* 23: 243-255.
- Kolda, T. G., B.W. Bader, and J.P. Kenny. 2005. Higher-order web link analysis using multi-linear algebra, in *ICDM 2005: Proceedings of the 5th IEEE International Conference on Data Mining*, IEEE Computer Society. Houston, Texas, USA. p. 242-249.
- Kolda, T. G. 2006. Multilinear operators for higher-order decompositions. Technical Report Number SAND2006-2081, Sandia National Laboratories. Albuquerque, New Mexico and Livermore, California. Available online at: <http://www.prod.sandia.gov/cgi-bin/techlib/access-control.pl/2006/062081.pdf> (Website accessed: 12 April, 2010).
- Kroonenberg, P.M. and J. De Leeuw. 1980. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* 45: 69-97.
- Kroonenberg, P.M. 1983. Three-Mode principal components analysis. Theory and Applications. Leiden, The Netherlands: DSWO-Press. 349 pp.
- Lanza, L.L.B., C.L. Jr. de Souza, L.M.M. Ottononi, M.L.C. Vieira, and A.P. de Souza. 1997. Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. *Theoretical and Applied Genetics* 94(8):1023-1030.
- Lopes, C.A., M.E. Rodríguez, A. Querol, S. Brarmardi, and A.C. Caballero. 2006. Relationship between molecular and enological features of Patagonian wine yeasts: relevance in selection protocols *World Journal of Microbiology and Biotechnology* 22:827-833.
- Lorea R.D., C. Delucchi, R. Defacio, G.H. Eyhérbide, A.R. Schlatter, and C.G. López. 2006. Correspondencia entre la caracterización de ocho poblaciones nativas de maíz a través de variables fenotípicas de mazorca y marcadores microsateélites (ssr). VII Congreso Latinoamericano de Sociedades de Estadística. Rosario, Santa Fe, Argentina (Resumen). p. 253.
- Lowe, A.J., O. Hanotte, and L. Garino. 1996. Standardization of molecular genetic techniques for the characterization of germplasm collection: the case of random amplified polymorphic DNA (RAPD). *Plant Genetic Resources Newsletter* 107:50-54.
- Lucchin M., G. Barcaccia, and P. Parrini. 2003. Characterization of a flint maize (*Zea mays* L. convar. Mays) Italian landrace: I. Morpho-phenological and agronomic traits. *Genetic Resource and Crop Evolution* 50:315-327.
- Milbourne, D., R. Meyer, J.E. Bradshaw, E. Baird, N. Bonar, J. Provan, W. Powell, and R. Waugh. 1997. Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. *Molecular Breeding* 3:127-136.
- Nghia, N.A., J. Kadir, E. Sunderasan, M.P. Abdulhah, A. Malik, and S. Napis. 2008. Morphological and Inter Simple Sequence Repeat (ISSR) Markers Analyses of *Corynespora cassicola* Isolates from Rubber Plantations in Malaysia. *Mycopathologia* 166:189-201.

- Pan, Y.B., D.M. Burner, B.L. Legendre, M.P. Grisham, and W.H. White. 2004. An assessment of the genetic diversity within a collection of *Saccharum spontaneum* L. with RAPD-PCR. *Genetic Resources and Crop Evolution* 51(1):895-903.
- Reeb P.D., S.J. Bramardi, and R.A. Defacio. 2007. Estimación y tratamiento del efecto ambiente y sus interacciones en la caracterización de Recursos Genéticos de un Banco de Germoplasma. En: XI Conferencia Española y I Encuentro Iberoamericano de Biometría. 20 -22 de junio. Salamanca, España. p. 187-188.
- Ruiz de Galarreta J.L., and A. Álvarez. 2001. Morphological classification of maize landraces from Northern Spain. *Genetic Resources and Crop Evolution* 48:391-400.
- Syamkumar, S., and B. Sasikumar. 2007. Molecular marker based genetic diversity analysis of *Curcuma* species from India. *Scientia Horticulturae* 112(2):235-241.
- Taba, S., J. Franco, and J. Crossa. 1998. Evaluation of Caribbean maize accessions to develop a core subset. *Crop Sci.* 38: 1378-1386.
- Tar'an, B., C. Zhang, T. Warkentin, A. Tullu, and A. Vandenberg. 2005. Genetic diversity among varieties and wild species accessions of pea (*Pisum sativum* L.) based on molecular markers, and morphological and physiological characters. *Genome* 48(2):257-272.
- Timmerman, M.E., and H. Kiers. 2000. Three-mode principal components analysis. Choosing numbers of components and sensitivity to local optima. *The British Journal of the Mathematical and Statistical Psychology* 53: 1-16.
- Tucker, L.R. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika* 31: 279-311.
- Van Eeuwijk, F.A., and P.M. Kroonenberg. 1998. Multiplicative Models for Interaction in Three-Way ANOVA, with Applications to Plant Breeding. *Biometrics* 54: 1315-1333.
- Varela, M. 2002. Los métodos Biplots como herramienta de Análisis de Interacción de orden superior en un Modelo Lineal/Bilineal. Tesis Doctoral. Universidad de Salamanca. España. 203 pp.

