Crop Breeding and Applied Biotechnology 4:7-11, 2004
Brazilian Society of Plant Breeding. Printed in Brazil

Proposal of methodologies for environment stratification and analysis of genotype adaptability

Devanir Mitsuyuki Murakami[1] and Cosme Damião Cruz[2*]

Received 18 August 2003
Accepted 30 March 2004

ABSTRACT - *A cultivar which performs well in a given environment will not necessarily perform as well in others. Thus studies on environmental stratification must be carried out to provide indications for cultivars. A procedure commonly used is the stratification of environments through the analysis of genotype-environment interactions, which must be minimal or statistically not significant. We propose an alternative procedure based on factor analysis. Moreover, we propose a methodology for the analysis of genotype adaptability through graphic analysis, generated by plotting scores from factor analysis. These new methodologies were used to assess grain yield of maize hybrids evaluated in diferent environments. Results provided the establishment of environmental subgroups with high correlations within the subgroups, and low or no correlations among them. The scores provided plot that helped understand the adaptability of the hybrids. The methodologies proposed for environment stratification and adaptability analysis were effective and of easy interpretation.*

Key words: maize, environmental stratification, adaptability analysis, factor analysis, plot analysis.

INTRODUCTION

Environment means all the non-genetic variables that may affect the phenotypic expression of a certain genotype. The environmental conditions, which exert influence on the expression of a genotype can be separated in the groups predictable and unpredictable. Predictable are the ones due to permanent environmental factors, such as soil fertility, photoperiod, sowing date, density, agricultural practices etc. The unpredictable occur randomly, such as final stand, rainfall distribution, temperature, and incidence of pests and diseases (Arias 1996).

A cultivar (hybrid or variety) of remarkable performance in one environment does not necessarily perform as well in all others. This is a problem when recommendations for general environment are to be given. To solve the question, environmental stratification studies should be carried out to provide appropriate cultivar indications.

The environment stratification efficacy, within years is greater than the proportion of cultivar-environment interactions due to the site. Little research is yet available on regionalization that takes predictable and unpredictable environmental variations into account; that is, which considers the effects of annual fluctuation on cultivar-site interactions (Silva 1981).

In general, analysis methods of the effect of the genotype-environment interaction can be grouped in two main streams. One is based on the regionalization of sites, and the other on the cultivar stability. The first line comprises two approaches: one is founded mainly on environmental characteristics, trying to adjust crop needs to the existing environmental availabilities, principally of climate and soil; the other proceeds from the reactions of cultivars exposed to different environmental conditions, and characterizes homogeneous regions that consider the relative behavior of cultivars (Silva 1981). Environment stratification enjoys widespread use

---

[1] Departamento de Fitotecnia, Universidade Federal de Viçosa (UFV), 36571-000, Viçosa, MG, Brasil.
[2] Departamento de Biologia Geral, BIOAGRO, UFV, 36571-000. *E-mail: cdcruz@ufv.br

among breeders through analyses of the genotype-environment interaction, which must be minimal or statistically not significant.

Recommendations of genotypes based on average yields, considering all trials, favor the most outstanding in the best environments, but do not differentiate those that adapt to best or worst cultivation conditions. Knowledge on the effects of genotype-environment interactions is very important to outline a breeding strategy for the choice of the experimental environment and genotypes, according to their adaptability to a given environmental condition. There are several methodologies for the analysis of stability and adaptability, but none of them considers stratification to discriminate the genotypes.

This study proposes an alternative methodology of environment stratification based on the multivariate technique of factor analysis. Within the same technique we also proposed a methodology of analyzing genotype adaptability by displays created by plotting scores obtained from the factor analysis.

METHODOLOGY

Factor analysis is a multivariate technique that allows a reduction of the great number of observed original variables to a small number of abstract variables, called *factors*. These may be independent or correlated and each factor assembles original variables strongly linked among each other, but weakly correlated to the other factors (Johnson and Wichern 1992). This way, when one same variable is evaluated in different environments, the obtained values in every environment can be looked upon as a different variable, allowing the use of a multivariate technique for the analysis. Results of the factor analysis then provide information that enables the formation of environment subgroups (stratification) so that there are high correlations within the subgroups and low or no correlations among them.

In this study, the grain yield (kg ha^{-1}) of several commercial maize hybrids was evaluated in various trials at three different periods: a) two in the autumn (alternative) growing season of 1996 and 1997; and, b) one in the summer (conventional) growing season of 1996/97. A randomized complete block design with three replications was used. Each plot consisted of four lines with a length of 5.20 m spaced 0.80 to 0.90 m between the rows and 0.20 m between plants, depending on the environment. Only the two central rows were used for data collection.

Mean grain yields (kg ha^{-1}) were considered for the factor analysis; ***n*** represented the tested hybrids and ***v*** the respective variables of each studied environment. Data were processed by software GENES (Cruz 1997).

Factor analysis, as described in (Johnson and Wichern 1992) follow the model

$$Xj = I_{j1}F_1 + I_{j2}F_2 + \dots + I_{jm}F_m + \varepsilon_j$$

where

X_j is the jth variable, with j = 1,2,....v,

I_{jk} is the factor loading for the jth variable associated to the kth factor, where k = 1,2,... m (number of common factors),

F_k is the jth common factor,

ε_j is the specific factor.

One of the main objectives of factor analysis is the characterization of common factors. This characterization is a function of the variables with highest loading within these factors. Summing up, a factor can only be clearly characterized as such if there are some variables with a high factor loading, and moreover, if these have some functional relationship among each other. However, cases assuming similar values, the factor loading are not distant enough from each other that a perfect characterization of the factors is not possible.

A way to overcome this trouble in factor analysis is by rotation, where the matrix of the initial factor loading $[L]_{pxm}$ (p = number of variables and, m = number of common factors) is multiplied afterwards by an orthogonal matrix $[T]_{mxm}$. Thereby, [L][T] brings forth a new matrix $[L]^*_{pxm}$ where $[T]'[T] = I_{mxm}$ and [L]* is a new matrix pxm of rotated factor loadings.

For the model of orthogonal factors, the most commonly used method for rotation is called varimax normal, whose matrix [T]* is formed by sines and cosines of the angle θ established between the axes of the factors. The rotation is always carried out two by two, independent of the number of common factors, with totally C_m^2 rotations, where m is the number of common factors.

The multiplication [L][T] spawns a new matrix of factor loadings [L]*, where some variables within each factor have a loading near to 1 and the others near to zero, which allows an interpretation of the factors. In other words, the variance of squares of the factor loadings can be maximized in each factor, and consequently, the sum of theses variances.

One aspect of the factor analysis that has to be determined is the number of final factors. Generally, the number of final factors has been defined as equal to the number of eigenvalues superior to the units found in the matrix of phenotypic correlations of standardized variables (Castoldi 1997). In this study, the final number of factors was established as equal to the number of eigenvalues above 1. Nevertheless, in cases where the percentage of the variability explained by eigenvalues above 1 was low, other factors were considered until an adequate proportion. In general, over 80% of the total variation was obtained.

Environment stratification (grouping environments) was based on final factor loadings (obtained after rotation), as described by Johnson and Wichern (1992). The factor loadings above or equal to 0.70 with positive signal indicate environments with high correlations within each factor which allow the establishment of environment subgroups (environment stratification); low factor loadings (≤ 0.5) make the identification of environmental groups impossible; and intermediate factor loading do not warrant a definition of groups.

The analysis of genotypic adaptability as proposed in this study, based on factor analysis, is realized graphically using the scores, which were generated through this methodology. For this purpose, parallel axes are sketched based on the score means, establishing four quadrants. In quadrants II and IV we find genotypes with specific adaptability to a certain sub-region, determined by the factor. Quadrant I holds genotypes of broad adaptability, while the genotypes in quadrant III are useless and of low performance; they should be discarded or not indicated for cultivation. In situations where only 2 factors sufficiently explain more than 80% of the total variance, defining only two sub-regions, the analysis of adaptability is based on the analysis of single plot. When more than two sub-regions are formed, the number of displays corresponds to the combination of m environments taken in pairs so that the genotype indications become more efficient (specific) than traditional methods. One reason is that this method does not take only favorable or unfavorable environments into consideration, but environment groups established according to their similarities (factor analysis), in the present case, the similarity of grain yield.

RESULTS AND DISCUSSION

Twelve commercial maize hybrids and four environments [1 = Lagoa Funda farm, Campo Verde county (previous crop: maize), 2 = Lagoa Funda farm (previous crop: soybean), 3 = Juriti farm, Primavera do Leste county (previous crop: maize) and, 4 = Girassol farm located in the Petrolina High Lands (previous crop: maize)] were studied in the trials of the autumn (alternative) growing season of 1996.

In the factor analysis, only one eigenvalue presented a value above 1, making up 57.40% of the total variation. Since one single factor contributed very little, another factor was considered to obtain an adequate variability. Thus, 81.03% of the total variation was obtained when taking the first two factors into consideration. This way, two final factors were fixed to obtain the final factor loadings (after rotations) (Table 1). With the proposed methodology based on factor analysis, it was possible to group the environments 1, 2, and 4 by the final factor loadings of the first factor (factor 1). Environment 3 appeared isolated in the second factor (Table 1). To understand and confirm the groups created by factor analysis it is necessary to analyze the results of correlations (r) among environments. Highest correlations were observed between the environments 1 and 2 (r = 0.63), 1 and 4 (r = 0.73), and 2 and 4 (r = 0.47), indicating a high correlation between the environments 1, 2, and 4; thus, these environments are grouped together. Environment 3 presented low correlations with all the other environments ($r_{1,3} = 0.13$, $r_{2,3} = 0.22$ and $r_{3,4} = 0.17$), and appears isolated in the second factor for this reason. However, when there is a high number of environments the analysis of the numerous correlation combinations is somewhat time-consuming. The proposed methodology offers significant simplification of theses analyses since it requires observation of the final factor loadings only.

In the trials of the summer (conventional) growing season of 1996/97, 16 commercial maize hybrids evaluated in six environments [1 = Lagoa Funda farm, Campo Verde county (previous crop: maize); 2 = Juriti farm, Primavera do Leste county (previous crop: soybean); 3 = Juriti farm (previous crop: maize); 4 = São Roque farm, Primavera do Leste county (previous crop: soybean); 5 = São Carlos farm, Rondonópolis county (previous crop: maize) and 6 = UFMT/FAMEV farm, Santo Antonio do Leverger county (previous crop: vegetables)] presented two eigenvalues with values above 1, making up only 67.30% of the total variation. To achieve at least 80%, four eigenvalues were considered and contributed with 87.86% of the total variation. The final factor loadings (after rotations) with the fixation of four factors are shown in Table 2. The first factor (factor 1) grouped the environments 3, 4, and 5 together. Environments 1, 2, and 6 appear isolated in the factors 2, 3, and 4, respectively, and do not form any subgroup (Table 2).

In the 1997 autumn experiments, 15 maize cultivars and eight environments [1 = Lagoa Funda farm, Campo Verde county (previous crop: soybean); 2 = Juriti farm, Primavera do Leste county (previous crop: maize); 3 = Juriti farm, Primavera do Leste county (previous crop: soybean); 4 = São Roque farm, Primavera do Leste county (previous crop: maize); 5 = São Roque farm (previous crop: soybean); 6 = Cuiabana farm, Primavera do Leste county (previous crop: soybean); 7 = Santa Maria farm, Rondonópolis county (previous crop = soybean) and 8 = Ponte de Pedra farm, Rondonópolis county (previous crop: soybean)] were investigated in the environment stratification studies; which results are shown in Table 3. Three eigenvalues lay above 1, making up 75.15% of the total variance. It was necessary to include one more eigenvalue to reach a minimum of 80%; thus, four eigenvalues contributed

Table 1. Establishment of environment subgroups according to the factor analysis method, considering 12 commercial maize hybrids and four environments for the trials of the autumn (alternative) growing season of 1996

Eigenvalues estimated			Factor loadings after rotation		
Eigenvalues	Eigenvalues (%)	% Acumulated	Environments	Factor 1	Factor 2
2.296178	57.40	57.40	1	0.928586	0.013651
0.945238	23.63	81.03	2	0.778291	0.199406
0.533372	13.33	94.37	3	0.097937	0.990425
0.225212	5.63	100.00	4	0.859728	0.061598

with 83.92% to the total variation. The final factor loadings obtained after the rotation, fixing four final factors, provide the following information: the environments 4, 5, and 7 grouped together in factor 1, and the environments 1, 3, and 8 in factor 2. The environments 2 and 6 appear isolated in the factors 3 and 4, respectively (Table 3).

For a presentation of the new adaptability analysis proposals, only the grain yield data obtained in the autumn (alternative) growing season of 1996 were considered, evaluated in four environments mentioned above, whose scores and plotting can be seen in Figure 1. The hybrids C-701, C-435, and XL-604 fit into in quadrant I which expresses broad adaptability since they presented a good performance in the sub-region 1, determined by factor 1, as well as in the sub-region 2, determined by factor 2. Hybrids XL-660 and EXCELER presented specific adaptability to sub-region 2 with environment 3; while C-505, XL-380, XB-8030, and XB-995074 presented adaptability to sub-region 1, with the environments 1, 2, and 4. The hybrids DENSUS, G-600 and XB-8028 fit into quadrant III, and represent the genotypes with the lowest performance which should consequently be disposed of, or not indicated.

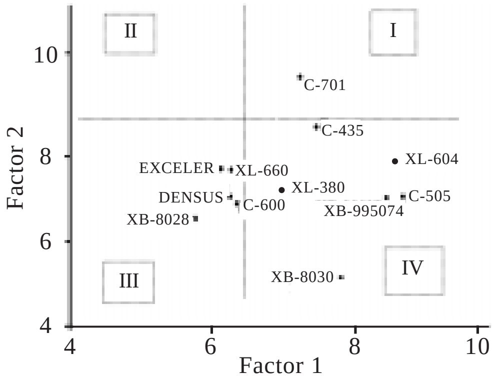

Figure 1. Scores plot from factor analysis considering 12 commercial maize hybrids evaluated in four environments, in the trials of the autumn (alternative) growing season of 1996.

All these hybrids, analyzed by the methodology of Eberhart and Russell (1966) presented broad (or general) adaptability (data not shown). Thus, the proposed methodology had the highest capacity for genotypic discrimination for adaptability. Besides its better differentiating capacity, the proposed methodology also presents more coherent results for the analyzed environment strata types.

Table 2. Establishment of environment subgroups according to results of the factor analysis, considering 16 commercial maize hybrids and six environments for the summer (conventional) growing season of 1996/97

Eigenvalues estimated			Factor loadings after rotation				
Eigenvalues	% Eigenvalues	% Acumulated	Environments	Factor 1	Factor 2	Factor 3	Factor 4
2.909479	48.49	48.49	1	0.037070	0.972076	0.106073	0.121295
1.128784	18.81	67.30	2	0.211584	0.187897	0.881064	0.227767
0.618765	10.31	77.62	3	0.693213	-0.043580	0.455264	0.163412
0.614567	10.24	87.86	4	0.877912	0.218251	-0.010726	0.257832
0.402525	6.71	94.57	5	0.725576	-0.149522	0.505983	-0.002983
0.32588	5.43	100.00	6	0.204888	0.133140	0.200737	0.940682

Table 3. Establishment of environment subgroups according to the factor analysis considering 15 commercial maize hybrids and eight environments for the trials of the 1997 autumn growing season

Eigenvalues estimated			Factor loadings after rotation				
Eigenvalues	% Eigenvalues	% Acumulated	Environments	Factor 1	Factor 2	Factor 3	Factor 4
3.364409	42.05	42.05	1	0.213691	0.754132	0.165753	0.182136
1.460224	18.25	60.31	2	0.245516	0.102475	-0.005168	0.951071
1.187655	14.85	75.15	3	-0.001567	0.859374	0.069393	0.058369
0.701374	8.77	83.92	4	0.903602	0.179132	0.091672	0.167588
0.551977	6.90	90.82	5	0.794121	0.356650	-0.254148	0.086813
0.474373	5.93	96.75	6	0.012099	0.152255	0.970875	-0.008909
0.189368	2.37	99.12	7	0.724894	-0.004350	0.453949	0.400370
0.070622	0.88	100.00	8	0.291044	0.809062	-0.012165	-0.064914

CONCLUSIONS

The obtained results allow the conclusion that the suggested methodology of environment stratification, based on the multivariate technique of factor analysis was efficient to combine environments according to the similarities in grain yield of the evaluated genotypes. Moreover, the proposed methodology for the analysis of the genotypic adaptability, also based on the factor analysis, revealed a great potential.

Proposição de metodologias para estratificação ambiental e análise de adaptabilidade de genótipos

RESUMO - *O bom desempenho de um cultivar em um ambiente pode não se repetir em outros ambientes. Assim, estudos da estratificação ambiental devem ser realizados para recomendação de cultivares. Comumente utiliza-se para esse fim a análise da interação genótipos-ambientes, a qual deve ser mínima ou estatisticamente não significativa. É proposta uma metodologia alternativa de estratificação baseada na análise de fatores. Também propõe-se uma metodologia de análise da adaptabilidade de genótipos pela análise gráfica gerada pela plotagem de escores da análise de fatores. As novas metodologias foram aplicadas no estudo da produtividade de grãos de híbridos de milho ensaiados em diferentes ambientes. Os resultados permitiram o estabelecimento de subgrupos de ambientes, tendo altas correlações dentro e baixa ou nenhuma entre subgrupos. Com os escores, obteve-se um gráfico para a visualização da adaptabilidade dos híbridos. As metodologias propostas para estratificação de ambientes e análise da adaptabilidade foram eficientes e de fácil interpretação.*

Palavras-chave: milho, estratificação ambiental, análise de estabilidade, análise de fator, análise gráfica.

REFERENCES

Arias ERA (1996) **Adaptabilidade e estabilidade das cultivares de milho avaliadas no Estado de Mato Grosso do Sul e avanço genético obtido no período de 1986/87 a 1993/94**. PhD Thesis, Universidade Federal de Lavras, Lavras, 132p.

Castoldi FL (1997) **Comparação de métodos multivariados aplicados na seleção em milho**. PhD Thesis, Universidade Federal de Viçosa, Viçosa, 118p.

Cruz CD (1997) **Programa GENES: Aplicativo Computacional em Genética e Estatística**. Editora UFV, Viçosa, 442p.

Eberhart SA and Russell WA (1966) Stability parameters for comparing varieties. **Crop Sci. 6**:36-40.

Johnson RA and Wichern DW (1992) **Applied Multivariate Statistical Analysis**. Englewood Cliffs, New Jersey-USA, 607p.

Silva EC (1981) **Classificações ambientais para controlar a interação genótipo x ambiente com aplicação à cultura do trigo (*Triticum aestivum* L.) no Rio Grande do Sul**. PhD Thesis. Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 76p.