Using artificial neural networks to select upright cowpea (*Vigna unguiculata*) genotypes with high productivity and phenotypic stability

L.M.A. Barroso¹, P.E. Teodoro², M. Nascimento¹, F.E. Torres³, A.C.C. Nascimento¹, C.F. Azevedo¹ and F.R.F. Teixeira¹

¹Department of Statistics, Federal University of Viçosa, Viçosa, MG, Brasil
²Department of General Biology, Federal University of Viçosa, Viçosa, MG, Brasil
³Department of Crop Science, State University of Mato Grosso do Sul, Aquidauana, MS, Brasil

Corresponding author: P.E. Teodoro
E-mail: eduteodoro@hotmail.com

Received August 2, 2016
Accepted September 9, 2016
Published November 3, 2016
DOI http://dx.doi.org/10.4238/gmr15049049

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

**ABSTRACT.** Cowpea (*Vigna unguiculata*) is grown in three Brazilian regions: the Midwest, North, and Northeast, and is consumed by people on low incomes. It is important to investigate the genotype x environment (GE) interaction to provide accurate recommendations for farmers. The aim of this study was to identify cowpea genotypes with high adaptability and phenotypic stability for growing in the Brazilian Cerrado, and to compare the use of artificial neural networks with the Eberhart and Russell (1966) method. Six trials with upright cowpea genotypes were conducted in 2005 and 2006 in the States of Mato
Grosso do Sul and Mato Grosso. The data were subjected to adaptability and stability analysis by the Eberhart and Russell (1966) method and artificial neural networks. The genotypes MNC99-537F-4 and EVX91-2E-2 provided grain yields above the overall environment means, and exhibited high stability according to both methods. Genotype IT93K-93-10 was the most suitable for unfavorable environments. There was a high correlation between the results of both methods in terms of classifying the genotypes by their adaptability and stability. Therefore, this new approach would be effective in quantifying the GE interaction in upright cowpea breeding programs.

**Key words:** Artificial intelligence; Genotype x environment interaction; Vigna unguiculata

**INTRODUCTION**

Cowpea [Vigna unguiculata (L.) Walp.] is one of the most important and strategic food sources in tropical and subtropical regions of the world (Torres et al., 2015a). Brazil is the third-largest producer of this crop in the world, which is grown in the Midwest, North, and Northeast, and is consumed by people on low incomes (Oliveira et al., 2013). However, Almeida et al. (2012) reported that a supply deficit often occurs in these regions, because the average Brazilian yield is extremely low (300 kg/ha). One way of increasing yield is to identify genotypes with a high yield that are suitable for Brazilian soil and climatic conditions (Santos et al., 2014a).

Crop production depends on genetic and environmental factors, in addition to interactions between them, which when significant, result in differential genotype behavior in different environmental conditions (Cruz et al., 2012). Therefore, when quantifying the magnitude of the genotype x environment interaction (GE), we should identify stable genotypes with wide adaptation capacities that can be grown in a range of environments, i.e., genotypes adapted to unfavorable environments that are suitable for small farmers using low-tech equipment, and genotypes responsive to improved environments that are suitable for high-tech equipment.

Previous studies have attempted to select cowpea genotypes with both a wide adaptability and a high phenotypic stability in different Brazilian regions (Santos et al., 2014a,b). Several statistical methods have been used, including additive main effect and multiplicative interaction (Santos et al., 2015), a Bayesian approach (Teodoro et al., 2015a,b; Barroso et al., 2016), restricted maximum likelihood/best linear unbiased prediction (Torres et al., 2015b, 2016), and the Eberhart and Russell (1966) method, which is based on linear regression (Almeida et al., 2012; Barros et al., 2013; Nunes et al., 2014). These studies have assisted in the introduction and improvement of cowpea cultivars in several tropical regions, such as the Brazilian Cerrado (Teodoro et al., 2015a,b).

The Eberhart and Russell (1966) method is widely used in genetic assessments of stability and adaptability because of its easy application, use, and interpretation of results. However, when the number of environments assessed in a breeding program is low (usually less than six), the method is inconsistent, because it can result in a failure to reject the null hypothesis. In order to solve this problem, Nascimento et al. (2013) used artificial neural networks (ANNs) in combination with the Eberhart and Russell (1966) method to classify alfalfa genotypes. Following this approach, we simulated genotypes belonging to the
Neural networks for selection of cowpea genotypes

phenotypic adaptability and stability classes defined by Eberhart and Russell (1966), which were subsequently used in the training and validation of ANNs.

ANNs are computational techniques that create a model that simulates a neural network, which is able to quickly process a large amount of data and recognize patterns based on self-learning (Haykin, 2009). After training the ANNs, we evaluated the genotypes for phenotypic stability and adaptability. This assessment was not only executed based on the genotypes studied, but on a large collection of simulated genotypes according to predefined classes (Nascimento et al., 2013). The aims of this study were to identify cowpea genotypes with high phenotypic adaptability and stability for growing in the Brazilian Cerrado and to compare the use of ANNs with the Eberhart and Russell (1966) method.

MATERIAL AND METHODS

Six trials were conducted in 2005 and 2006 in the municipalities of Aquidauana, Chapadão do Sul, and Dourados in the State of Mato Grosso do Sul and the municipality of Primavera do Leste, Mato Grosso (Table 1). The experiment had a randomized block design with 17 treatments and four replicates. The experimental unit consisted of four 5.0-m long rows that were spaced 0.5 m apart, with 0.25 m between plants within each row. In each experimental unit, grain yield was evaluated in the two central rows, and was corrected for 13% moisture and extrapolated to kg/ha.

The treatments consisted of 17 lines (MNC99-537F-1, MNC99-537F-4, MNC99-541-F5, MNC99-541-F8, IT93K-93-10, Pretinho, Fradinho-2, MNC99-519D-1-1-5, MNC00-544D-10-1-2-2, MNC00-544D-14-1-2-2, MNC00-553D-8-1-2-2, MNC00-553D-8-1-2-3, MNC00-561G-6, EVX63-10E, MNC99542F-5, EVX91-2E-2, and MNC99-557F-2) and three cultivars (BRS Guariba, Patativa, and Vita-7), totaling 20 genotypes.

The data were subjected to individual analyses of variance (ANOVA) for each environment, with the genotype effect fixed and the other effects random (Cruz et al., 2012), according to the following model:

\[ Y_{ij} = \mu + B_j + G_i + \varepsilon_{ij} \]  

(Equation 1)

where \( Y_{ij} \) is the value of the \( i \)th genotype in the \( j \)th block (\( i = 1, \ldots, g \) and \( j = 1, \ldots, b \), \( g \) and \( b \) being the number of genotypes and blocks, respectively); \( \mu \) is the overall mean; \( B_j \) is the effect of the \( j \)th block; \( G_i \) is the effect of the \( i \)th genotype; and \( \varepsilon_{ij} \) is the random error. A joint analysis of the trials was performed that included the effect of genotype as fixed and the other effects as random, according to the following model:
where $Y_{ijk}$ is the value of the $i$th genotype in the $j$th block in the $k$th environment ($k = 1, ..., e$, $e$ being the number of environments); $\mu$ is the overall mean; $B_{jk}$ is the effect of the $j$th block in $k$ environment; $G_i$ is the effect of the $i$th genotype; $GE_{ik}$ is the effect of the GE interaction; and $\varepsilon_{ijk}$ is the random error. Subsequently, the data were submitted to adaptability and stability analysis by the Eberhart and Russell (1966) method and ANNs (Nascimento et al., 2013).

The method proposed by Eberhart and Russell (1966) is based on linear regression analysis, which measures the response of each genotype to environmental variation. Therefore, for an experiment with $g$ genotypes, $e$ environments, and $r$ repetitions, we define the following statistical model:

$$Y_{ij} = \beta_{0i} + \beta_{1i}I_j + \psi_{ij}$$  \hspace{1cm} (Equation 3)

where $Y_{ij}$ is the mean of genotype $i$ in environment $j$; $\beta_{0i}$ is the linear coefficient of the $i$th genotype; $\beta_{1i}$ is the regression coefficient that measures the response of the $i$th genotype to variation in environment $j$; and $I_j$ is defined as the environmental index, by the following equation:

$$I_j = \frac{\sum_{i} Y_{ij}}{g} - \frac{\sum_{i} Y_{i\cdot}}{ge}$$  \hspace{1cm} (Equation 4)

and $\Psi_{ij}$ are random errors, in which each component can be decomposed as the following equation:

$$\Psi_{ij} = \delta_{ij} + \varepsilon_{ij}$$  \hspace{1cm} (Equation 5)

where $\delta_{ij}$ is the regression deviation and $\varepsilon_{ij}$ is the mean experimental error. Estimators of the adaptability and stability parameters are respectively given by:

$$\hat{\beta}_i = \frac{\sum_{j} Y_{ij}}{\sum_{j} I_j}$$  \hspace{1cm} (Equation 6)

and:

$$\hat{\sigma}_i^2 = \frac{MSD_i - MSR}{r}$$  \hspace{1cm} (Equation 7)

where $MSD_i$ is the mean square of deviations of genotype $i$ and $MSR$ is the mean squared residue. The hypotheses of interest were $H_0: \beta_{1i} = 1$ versus $H_1: \beta_{1i} \neq 1$ and $H_0: \sigma_i^2 = 0$ versus $H_1: \sigma_i^2 > 0$. These hypotheses were evaluated by a Student $t$-test and an $F$-test, respectively.

For evaluating the adaptability and stability of genotypes by ANNs, two datasets are required: the training set and the testing set. To obtain these sets according to the classes
Neural networks for selection of cowpea genotypes

defined, 1500 genotypes were simulated according to statistical model 1, and were evaluated in seven environments. The parameter values used for obtaining the genotypes of classes 1, 2, and 3 (Table 2), each consisting of 500 genotypes, were as follows: Class 1: \( \beta_i = \bar{x}_i, \beta_i \sim U[0.90; 1.10], \) and \( \sigma^2 = 250, \) i.e., \( \beta_i \) is considered equal to 1 if \( \beta_i \in [0.90; 1.10]; \) Class 2: \( \beta_i = \bar{x}_i, \beta_i \sim U[1.11; 2.00], \) and \( \sigma^2 = 250, \) i.e., \( \beta_i \) is considered greater than 1 if \( \beta_i \in [1.11; 2.00]; \) Class 3: \( \beta_i = \bar{x}_i, \beta_i \sim U[0.00; 0.89], \) and \( \sigma^2 = 250, \) i.e., \( \beta_i \) is considered lower than 1 if \( \beta_i \in [0.00; 0.89]. \) In addition, \( U[a;b] \) was the continuous uniform probability distribution, with parameters a and b. For obtaining the three remaining classes (4, 5, and 6) in order to linearize the set of values, the simulated values were transformed for the logarithmic scale, i.e., for classes 4, 5, and 6 we had \( \sigma^2 = 0. \) Thus, in the same manner as in study conducted by Finlay and Wilkinson (1963), the concept of stability was linked to the capacity of the genotypes to present a predictable response, according to the environment stimulus.

<table>
<thead>
<tr>
<th>Class</th>
<th>Practical classification</th>
<th>Parametric value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General adaptability and low predictability</td>
<td>( \beta_i = 1 ) and ( \sigma^2 &gt; 0 )</td>
</tr>
<tr>
<td>2</td>
<td>Specific adaptability to favorable environments and low predictability</td>
<td>( \beta_i &gt; 1 ) and ( \sigma^2 &gt; 0 )</td>
</tr>
<tr>
<td>3</td>
<td>Specific adaptability to unfavorable environments and low predictability</td>
<td>( \beta_i &lt; 1 ) and ( \sigma^2 &gt; 0 )</td>
</tr>
<tr>
<td>4</td>
<td>General adaptability and high predictability</td>
<td>( \beta_i = 1 ) and ( \sigma^2 = 0 )</td>
</tr>
<tr>
<td>5</td>
<td>Specific adaptability to favorable environments and high predictability</td>
<td>( \beta_i &gt; 1 ) and ( \sigma^2 = 0 )</td>
</tr>
<tr>
<td>6</td>
<td>Specific adaptability to unfavorable environments and high predictability</td>
<td>( \beta_i &lt; 1 ) and ( \sigma^2 = 0 )</td>
</tr>
</tbody>
</table>

In the same way as Nascimento et al. (2013), after obtaining 3000 genotypes (representatives of the six classes), the dataset was partitioned in two: the training set and the testing set. The training set was composed of 2400 genotypes, and was obtained by the random selection of 400 genotypes within each class. The testing set was composed of the remaining 600 genotypes (100 in each class), and was used for testing the network.

The ANNs used in this study, as denoted by a back-propagation hidden layer, are described by Nascimento et al. (2013). After training and testing the ANNs, which had a maximum error of 2% for the testing set, a cotton dataset was subjected to ANNs for classification. Subsequently, classification based on adaptability and stability was conducted; for comparison, this was also performed by the Eberhart and Russell (1966) method. The ANNs were implemented in R (R Development Core Team, 2011), and the Genes software (Cruz, 2013) was used for the Eberhart and Russell (1966) method.

RESULTS AND DISCUSSION

The individual ANOVAs revealed a significant block effect in all of the environments (Table 3), demonstrating that this design should be used in these types of experiments in order to control this source of heterogeneity. There were significant differences between the genotypes in all of the trials. The coefficients of variation obtained by the individual ANOVAs ranged between 23.08 and 34.08%, which were similar to those reported in other studies on cowpea (Rocha et al., 2007; Almeida et al., 2012; Santos et al., 2014a,b; Torres et al. 2015a,b).
The ratio between the highest (E1) and lowest (E5) residual mean square of the trials was 12.97, which indicates variance heterogeneity according to the Banzatto and Kronka (2006) criterion, which considers values greater than 7.0 indicative of variance heterogeneity. Therefore, we adjusted the degrees of freedom of the mean error and the GE interaction, according to the Cochran (1954) method.

A summary of the joint ANOVA results is presented in Table 4. The genotype effect was not significant (P > 0.05), suggesting an absence of genetic variability among the genotypes. However, Cruz et al. (2012) reported that when the genotype effect is significant in individual ANOVAs but not in a joint ANOVA, the genetic variability present is consumed by the magnitude of the GE interaction effect.

Environment and GE interaction effects were significant (P < 0.01), indicating that the environments significantly differed and there were differential genotype responses in the different environments. This can be explained by the edaphic and climatic features of each environment (Table 1), which differed in altitude, latitude, longitude, climate, and soil type, in addition to climatic variables such as rainfall and temperature. Similar results were obtained by Rocha et al. (2007), Barros et al. (2013), Torres et al. (2015b), and Santos et al. (2015). Torres et al. (2016) also reported significant environment and GE interaction effects when evaluating cowpea genotypes in multi-environment trials in Brazil. A significant GE interaction indicates that phenotypic stability and adaptability analyses are required, because edaphoclimatic factors affect grain yield more than any other parameters.

Table 4 shows the mean grain yield and phenotypic adaptability and stability of the genotypes using the Eberhart and Russell (1966) method and ANNs. Genotypes MNC99-537F-4 and EVX91-2E-2 had higher grain yields than the overall average for the environments,
and were highly stable according to both methods of analysis. Therefore, these genotypes are the most suitable for favorable environments and can be used by farmers that use high-tech equipment and procedures, because they can respond to environmental improvements in terms of fertilization and irrigation, among other practices. Low-tech farmers should grow the IT93K-93-10 genotype, which despite not having a higher grain yield than the overall average, was highly predictable according to both methods of analysis. Our results suggest that this genotype should maintain its production level under different environmental conditions.

**Table 5.** Mean grain yield and classification of 20 upright cowpea (*Vigna unguiculata*) genotypes based on phenotypic adaptability and stability by the Eberhart and Russell (1966) method and artificial neural networks in four environments in Mato Grosso do Sul, Brazil.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MNC99-537F-1</td>
<td>725.58</td>
<td>Overall</td>
<td>Low</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>MNC99-537F-4</td>
<td>891.92</td>
<td>Favorable</td>
<td>High</td>
<td>Favorable</td>
<td>High</td>
</tr>
<tr>
<td>MNC99-541F-5</td>
<td>716.75</td>
<td>Overall</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>MNC99-541F-8</td>
<td>651.01</td>
<td>Favorable</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>IT93K-93-10</td>
<td>514.18</td>
<td>Unfavorable</td>
<td>High</td>
<td>Unfavorable</td>
<td>High</td>
</tr>
<tr>
<td>Pretinho</td>
<td>433.20</td>
<td>Overall</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>Fradinho-2</td>
<td>638.64</td>
<td>Overall</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>MNC99-519D-1.1.1.5</td>
<td>671.86</td>
<td>Overall</td>
<td>Low</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>MNC00-544D-10.1.2.2</td>
<td>602.69</td>
<td>Overall</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>MNC00-544D-14.1.2.2</td>
<td>722.08</td>
<td>Overall</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>MNC00-553D-8.1.2.2</td>
<td>641.91</td>
<td>Overall</td>
<td>Low</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>MNC00-553D-8.1.2.3</td>
<td>650.44</td>
<td>Overall</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>MNC00-561G-6</td>
<td>690.61</td>
<td>Favorable</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>EVX63-10E</td>
<td>682.57</td>
<td>Overall</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>MNC99542F-5</td>
<td>882.23</td>
<td>Overall</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>EVX91-2E-2</td>
<td>722.23</td>
<td>Favorable</td>
<td>High</td>
<td>Favorable</td>
<td>High</td>
</tr>
<tr>
<td>MNC99-557F-2</td>
<td>494.64</td>
<td>Overall</td>
<td>Low</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>BRS Guariba</td>
<td>667.20</td>
<td>Overall</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>Patativa</td>
<td>753.34</td>
<td>Overall</td>
<td>High</td>
<td>Overall</td>
<td>High</td>
</tr>
<tr>
<td>Vita-7</td>
<td>496.39</td>
<td>Unfavorable</td>
<td>Low</td>
<td>Unfavorable</td>
<td>High</td>
</tr>
</tbody>
</table>

Agreement: Adaptability 90% Stability 75%

According to Eberhart and Russell (1966), an ideal genotype should maintain its constant production potential when grown in unfavorable environments, and increase productivity in favorable environments. Therefore, the ideal genotype is one that has a high yield, good adaptability, and high predictability. In this study, we identified (by using two analytical methods) the following ideal genotypes for growing in the State of Mato Grosso do Sul, Brazil: MNC99-541-F5, MNC00-544D-14.1.2.2, EVX63-10E, MNC99542F-5, BRS Guariba, and Patativa. These results should be used to guide producers in this region, as well as to increase cowpea cultivation in the Brazilian Cerrado.

There was 90% agreement between the Eberhart and Russell (1966) method and ANNs in terms of the phenotypic adaptability of the genotypes (Table 5), and 75% agreement in terms of the phenotypic stability; this was lower than the adaptability value, probably because ANN stability is based on the Finlay and Wilkinson (1963) method, which differs from the Eberhart and Russell (1966) method by considering stability, invariance, and non-predictability. The strong agreement between the traditional Eberhart and Russell (1966) method and ANNs has been reported in studies that evaluated the GE interaction in genotypes of alfalfa (Nascimento et al., 2013), semi-prostate cowpea (Teodoro et al., 2015a), and common bean (Correa et al.,
2016). This new approach is an effective method of quantifying the adaptability and stability of different genotypes in upright cowpea breeding programs. The main advantage of ANNs over the Eberhart and Russell (1966) method is that because of their non-linear structure (Haykin, 2009), they can capture the most complex features of a dataset without requiring detailed information about the process to be modeled, because they are self-learning (Nascimento et al., 2013).

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support.

REFERENCES


Neural networks for selection of cowpea genotypes


Genetics and Molecular Research 15 (4): gmr15049049