



# Artificial intelligence in the selection of common bean genotypes with high phenotypic stability

A.M. Corrêa<sup>1</sup>, P.E. Teodoro<sup>2</sup>, M.C. Gonçalves<sup>3</sup>, L.M.A. Barroso<sup>4</sup>,  
M. Nascimento<sup>4</sup>, A. Santos<sup>5</sup> and F.E. Torres<sup>1</sup>

<sup>1</sup>Departamento de Fitotecnia, Universidade Estadual do Mato Grosso do Sul, Aquidauana, MS, Brasil

<sup>2</sup>Departamento de Biologia Geral, Universidade Federal de Viçosa, Viçosa, MG, Brasil

<sup>3</sup>Departamento de Estatística, Universidade Federal da Grande Dourados, Dourados, MS, Brasil

<sup>4</sup>Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, Brasil

<sup>5</sup>Departamento de Melhoramento Genético, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campo dos Goytacazes, RJ, Brasil

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

Genet. Mol. Res. 15 (2): gmr.15028230

Received December 8, 2015

Accepted January 18, 2016

Published April 27, 2016

DOI <http://dx.doi.org/10.4238/gmr.15028230>

**ABSTRACT.** Artificial neural networks have been used for various purposes in plant breeding, including use in the investigation of genotype x environment interactions. The aim of this study was to use artificial neural networks in the selection of common bean genotypes with high phenotypic adaptability and stability, and to verify their consistency with the Eberhart and Russell method. Six trials were conducted using 13 genotypes of common bean between 2002 and 2006 in the municipalities of Aquidauana and Dourados. The experimental design was a randomized block with three replicates. Grain yield data were submitted to individual and joint variance analyses. The data were then

submitted to analysis of adaptability and stability through the Eberhart and Russell and artificial neural network methods. There was high concordance between the methodologies evaluated for discrimination of phenotypic adaptability of common bean genotypes, indicating that artificial neural networks can be used in breeding programs. Based on both approaches, the genotypes Aporé, Rudá, and CNFv 8025 are recommended for use in unfavorable, general and favorable environments, respectively by the grain yield above the overall average of environments and high phenotypic stability.

**Key words:** Artificial neural networks; Eberhart and Russell method; Genotype x environment interaction; *Phaseolus vulgaris* L.

## INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is one of the main crops in Brazil. Genetic breeding programs of common bean have provided the Brazilian market new cultivars with desirable traits, such as yield stability, and have contributed to the increase in average yield from 0.75 Mg/ha in 1997 to 1.25 Mg/ha in 2007 (Chiorato et al., 2010). In this regard, some studies have been conducted to select genotypes of common bean that are superior for both adaptability and phenotypic stability. Several methodologies have been used, including the AMMI (Additive Main Effects and Multiplicative Interaction) and Eberhart and Russell (1966) methods, which are based on linear regression (Pereira et al., 2009; Ribeiro et al., 2009).

Of these, the method proposed by Eberhart and Russell (1966) is widely used owing to its ease of application. However, in studies evaluating a small number of environments ( $n < 6$ ), the results obtained by this method are not consistent, which may cause non-rejection of the null hypotheses. To address this problem, artificial neural networks (ANNs) have been used for alfalfa genotype classification in accordance with the method described by Eberhart and Russell (1966). In this approach, which was developed by Nascimento et al. (2013), genotypes with adaptability and phenotypic stability as defined by Eberhart and Russell (1966) are simulated. Subsequently, the simulated genotypes are used to train and validate neural networks. Thus, by the training of ANNs, the identifying the stable genotypes is not only executed based on the studied genotypes, but for a large collection of simulated genotypes according to predefined groups (Nascimento et al., 2013; Silva et al., 2014; Sant'Anna et al., 2015).

ANNs are models that emulate a network of biological neurons, which are able to quickly process a large amount of data and recognize patterns based on self-learning (Haykin, 2009). As a result, this method has been employed for breeding in crops such as alfalfa (Nascimento et al., 2013), sugarcane (Brasileiro et al., 2015), eucalyptus (Bhering et al., 2015), and papaya (Barbosa et al., 2011). The aim of the present study was to use ANNs to select common bean genotypes that have high adaptability and phenotypic stability, and to verify their consistency with the Eberhart and Russell (1966) method.

## MATERIAL AND METHODS

Twelve trials (environments) were performed between 2000 and 2006. These were composed of combinations of local environments, growing seasons, and agricultural years

(Table 1). Thirteen genotypes (Rudá, Aporé, Xamego, sPérola, Ouro Negro, Diamante Negro, IAPAR 14, ENGOPA 201, and IAC-Carioca Eté and the lines BambuÍ, CNF 4999, CNF 4129 A 54, and CNFv. 8025) were evaluated in a randomized block design with three replicates. Experimental units consisted of four rows of plants 1.50 m long, spaced 0.50 m apart. Grain yield was evaluated in the central rows, and was expressed in Mg/ha, considering the area harvested and adjusting the data for 13% moisture.

**Table 1.** Environments, site, climate, latitude, longitude, altitude, harvest, and agricultural year of trials conducted in the State of Mato Grosso do Sul.

Environment	Site	Climate <sup>1</sup>	Latitude	Longitude	Altitude (m)	Harvest	Agricultural year
1	Dourados	Cwa	22°12'	54°48'	452	Rainy	2000/2001
2	Dourados	Cwa	22°12'	54°48'	452	Drought	2000/2001
3	Aquidauana	Aw	20°20'	55°48'	207	Drought	2000/2001
4	Dourados	Cwa	22°12'	54°48'	452	Rainy	2001/2002
5	Dourados	Cwa	22°12'	54°48'	452	Drought	2001/2002
6	Aquidauana	Aw	20°20'	55°48'	207	Drought	2001/2002
7	Aquidauana	Aw	22°12'	54°48'	207	Drought	2002/2003
8	Aquidauana	Aw	20°20'	55°48'	207	Drought	2003/2004
9	Aquidauana	Aw	22°12'	54°48'	207	Drought	2004/2005
10	Dourados	Cwa	22°12'	54°48'	452	Rainy	2005/2006
11	Dourados	Cwa	22°12'	54°48'	452	Drought	2005/2006
12	Aquidauana	Aw	22°12'	54°48'	207	Drought	2005/2006

<sup>1</sup>According to Köppen-Geiger classification.

Data were subjected to individual analysis of variance, considering the effects of treatments as fixed and the other effects as random. The relationship between the highest and lowest mean square of residue from an individual analysis of variance did not exceed the ratio 7:1, allowing the joint analysis of trials (Sprent and Smeeton, 2007). Subsequently, data were submitted to adaptability and stability analysis using the Eberhart and Russell (1966) and ANNs methods (Nascimento et al., 2013).

The method proposed by Eberhart and Russell (1966), which is based on linear regression analysis, measures the response of each genotype to environmental variations. Thus, for an experiment with  $g$  genotypes,  $e$  environments and  $r$  repetitions, we defined the following statistical model 1:  $Y_{ij} = \beta_{0i} + \beta_{1i}I_j + \Psi_{ij}$  where in  $Y_{ij}$  is the mean of genotype  $i$  in the environment  $j$ ;  $\beta_{0i}$  is the linear coefficient of the  $i$ -th genotype;  $\beta_{1i}$  is the regression coefficient that measures the response of  $i$ -th genotype to variation of the environment  $j$ ; and  $I_j$  is defined as the environmental index by:

$$I_j = \frac{\sum_j Y_j}{g} - \frac{\sum_i \sum_j Y_{ij}}{ge} \quad (\text{Equation 1})$$

where in  $\Psi_{ij}$  are random errors, in which each component can be decomposed as:

$$\Psi_{ij} = \delta_{ij} + \bar{\epsilon}_{ij} \quad (\text{Equation 2})$$

with  $\delta_{ij}$  being the regression deviation and  $\bar{\epsilon}_{ij}$  the mean experimental error.

The estimators of adaptability and stability parameters are respectively given by:

$$\hat{\beta}_{li} = \frac{\sum_j Y_{ij}}{\sum_j I_j^2} \quad (\text{Equation 3})$$

and:

$$\sigma_{di}^2 = \frac{\text{MSD}_i = \text{MSR}}{r} \quad (\text{Equation 4})$$

where in  $\text{MSD}_i$  is the mean square of deviations of the genotype  $i$ ; MSR is the mean square of residue. The hypotheses of interest are  $H_0: b_{li} = 1$  versus  $H_1: b_{li} \neq 1$  and  $H_0: \sigma_{di}^2 = 0$  versus  $H_1: \sigma_{di}^2 = 0$ . These hypotheses were evaluated by the  $t$ -test and F-test, respectively. In order to evaluate the adaptability and stability of genotypes by ANNs, two data sets are required, which are the training set and the testing set. To obtain these data sets according to the classes defined, 1500 genotypes were simulated, according to statistical model 1, which were evaluated in seven environments. The parameter values used to obtain the genotypes of classes 1, 2, and 3 (Table 2), each consisting of 500 genotypes, were: Class 1:  $\beta_{0i} = \bar{X}_G$ ,  $\beta_{ii} \sim U[0.90; 1.10]$  and  $\sigma_{\psi}^2 = 250$ , i.e.,  $\beta_{li}$  will be considered equal to 1 if  $\beta_{ii} \in [0.90; 1.10]$ ; Class 2:  $\beta_{0i} = \bar{X}_G$ ,  $\beta_{ii} \sim U[1.11; 2.00]$  and  $\sigma_{\psi}^2 = 250$ , i.e.,  $\beta_{li}$  will be considered greater than 1 if  $\beta_{ii} \in [1.11; 2.00]$ ; Class 3:  $\beta_{0i} = \bar{X}_G$ ,  $\beta_{ii} \sim U[0.00; 0.89]$  and  $\sigma_{\psi}^2 = 250$ , i.e.,  $\beta_{li}$  will be considered lower than 1 if  $\beta_{ii} \in [0.00; 0.89]$ . In addition,  $U[a;b]$  is the continuous uniform probability distribution with parameters  $a$  and  $b$ . In order to obtain the remaining three classes (4, 5, and 6), which aimed to linearize the set of values, the simulated values were transformed into a logarithmic scale, i.e., for classes 4, 5 and 6, we have  $\sigma_{\psi}^2 = 0$ . Thus, this is consistent with the study of Finlay and Wilkinson (1963), in which the concept of stability was linked to the capacity of the genotypes present to obtain a predictable response according to the environment stimulus.

Consistent with the study of Nascimento et al. (2013), after 3000 genotypes were obtained, which represented the six classes, the data set was partitioned into two, namely the training set and the ANN testing set. The ANN training set was composed of 2400 genotypes and was obtained by randomly selecting 400 genotypes within each class. The testing set was composed of the remaining 600 genotypes, which consisted of 100 from, each class, and was used to test the network.

The ANNs used in this study, which were denoted by back-propagation hidden layer, was described in the study of Nascimento et al. (2013). After the training and testing stage of ANN, in which a maximum error of 2% was considered for the testing set, the cotton data set was submitted to ANNs for classification. The adaptability and phenotypic stability were classified. For the purpose of comparison, evaluation of the adaptability and stability of genotypes was also carried out by the method proposed by Eberhart and Russell (1966).

**Table 2.** Genotype classes as defined by the Eberhart and Russell method (1966) and their respective parametric values.

Class	Practical classification	Parametric values
1	General adaptability low predictability	$\beta_{hi} = 1$ and $\sigma_{di}^2 > 0$
2	Specific adaptability to favorable environments and low predictability	$\beta_{hi} > 1$ and $\sigma_{di}^2 > 0$
3	Specific adaptability to unfavorable environments and low predictability	$\beta_{hi} < 1$ and $\sigma_{di}^2 > 0$
4	General adaptability and high predictability	$\beta_{hi} = 1$ and $\sigma_{di}^2 = 0$
5	Specific adaptability to favorable environments and high predictability	$\beta_{hi} > 1$ and $\sigma_{di}^2 = 0$
6	Specific adaptability to unfavorable environments and high predictability	$\beta_{hi} < 1$ and $\sigma_{di}^2 = 0$

The ANNs used was implemented in R software (R Development Core Team, 2015) using the codes available at <http://www.det.ufv.br/~moyses/links.php>. To obtain results consistent with those reported by Eberhart and Russell (1966), the Genes program was used (Cruz, 2013).

## RESULTS AND DISCUSSION

All effects were found to be significant ( $P < 0.01$ ) on joint analysis (Table 3), which indicates a contrast between the environments and the occurrence of differential responses of genotypes to the environmental effects. This can be confirmed by observing the features of each environment, which show differences in climate, altitude, latitude, longitude, and harvest. This reflects the unpredictable effects of rainfall and temperature within each environment. Similar results were obtained by Coimbra et al. (1999), Carbonell et al. (2004), Ribeiro et al. (2009), Pereira et al. (2009, 2011), and Torga et al. (2013), who also found significant differences for the effects of genotypes, environments and genotype x environment interaction, when evaluating common bean genotypes in multienvironment trials in Brazil. The existence of significant genotype x environment interactions for grain yield indicates that the analyses of stability and adaptability are appropriate, since they are the edaphoclimatic factors that most influence the adaptability and stability of genotypes.

**Table 3.** Summary of joint analysis of variance for grain yield, in Mg/ha, of 13 common bean genotypes, evaluated in six environments of the State of Mato Grosso do Sul.

Sources of variation	Degrees of freedom	Mean square
Blocks/environment	12	0.0218
Genotypes	12	0.2298*
Environments	5	0.7258*
Genotypes x environments	60	0.5823*
Error	132	0.0566
Mean	-	1.89
Coefficient of variation (%)	-	4.48

\*Significant at 5% probability by the F-test.

The genotypes Aporé, Rudá, and CNFv 8025 stood out by exhibiting grain yield that was above the overall average of environments and high phenotypic stability based on both methods being recommended for unfavorable, general, and favorable environments, respectively (Table 4). Similar results were obtained by Coimbra et al. (1999), who verified adaptability to general environments and high predictability for the genotype Rudá.

**Table 4.** Mean grain yield (Mg/ha), classification of 13 common bean genotypes for adaptability and phenotypic stability, evaluated in six environments in the State of Mato Grosso do Sul, by Eberhart and Russell (1966) and artificial neural networks methods.

Genotype	Mean (Mg/ha)	Eberhart and Russell (1966)		Artificial neural networks	
		Adaptability	Stability	Adaptability	Stability
Rudá	2.18	General	High	General	High
Aporé	2.20	Unfavorable	High	Unfavorable	High
Xamego	1.86	General	High	General	High
Pérola	1.69	Unfavorable	High	Unfavorable	High
Ouro Negro	1.74	General	High	General	High
Diamante Negro	1.68	General	High	General	High
IAPAR 14	2.06	Unfavorable	Low	Unfavorable	High
EMGOPA-201	1.95	Favorable	High	Favorable	High
IAC Carioca Ete	1.70	Unfavorable	High	Unfavorable	High
Bambuí	2.11	Favorable	Low	Favorable	High
CNF 4999	1.76	Favorable	High	Favorable	High
CNF 4129 A 54	1.48	Favorable	Low	Favorable	Low
CNFv 8025	2.15	Favorable	High	Favorable	High
Agreement percentage		Adaptability: 100%		Stability: 85%	

The results obtained here were 100% consistent between the Eberhart and Russell (1966) and ANN methods regarding the discrimination of phenotypic adaptability of common bean genotypes. For phenotypic stability, there was 85% similarity in the classification of genotypes, probably because stability in the ANN method is based on the method of Finlay and Wilkinson (1963), who consider stability as an invariance and therefore not predictable, which differs from the Eberhart and Russell method (1966). Similar results were obtained by Nascimento et al. (2013) and Teodoro et al. (2015), who showed an average agreement of 93 and 85% between the Eberhart and Russell (1966) and ANN methods regarding adaptability and phenotypic stability of alfalfa and cowpea genotypes, respectively.

Due to the high rates of agreement among the evaluated methods, ANNs can be considered an effective alternative to measure the adaptability and phenotypic stability of genotypes in breeding programs. In addition, due to their non-linear structure (Haykin, 2009) ANNs can capture more complex features of data sets and do not require detailed information about the process to be modeled due to its self-learning (Nascimento et al., 2013; Silva et al., 2014; Sant'Anna et al., 2015).

### Conflicts of interest

The authors declare no conflict of interest.

### ACKNOWLEDGMENTS

We thank the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) and the National Council for Scientific and Technological Development (CNPq) for financial support.

### REFERENCES

Barbosa CD, Viana AP, Quintal SSR and Pereira MGP (2011). Artificial neural network analysis of genetic diversity in *Carica papaya* L. *Crop Breed. Appl. Biotechnol.* 11: 224-231.

- Bhering LL, Cruz CD, Peixoto LA, Rosado AM, et al. (2015). Application of neural networks to predict volume in eucalyptus. *Crop Breed. Appl. Biotechnol.* 15: 125-131. <http://dx.doi.org/10.1590/1984-70332015v15n3a23>
- Brasileiro BP, Marinho CD, Costa PMA, Cruz CD, et al. (2015). Selection within families of sugarcane via artificial neural networks. *Crop Breed. Appl. Biotechnol.* 15: 72-78. <http://dx.doi.org/10.1590/1984-70332015v15n2a14>
- Carbonell SAM, Azevedo Filho JA, Dias LAS, Garcia AAF, et al. (2004). Common bean cultivars and lines interactions with environments. *Sci. Agric.* 61: 169-177. <http://dx.doi.org/10.1590/S0103-90162004000200008>
- Chiorato AF, Carbonell SAM, Vencovsky R, Fonseca NS, Junior., et al. (2010). Genetic gain in the breeding program of common beans at IAC from 1989 to 2007. *Crop Breed. Appl. Biotechnol.* 10: 329-336.
- Coimbra HLM, Carvalho FIF, Hemp S and Silva SA (1999). Adaptability and phenotypic stability of collar bean genotypes (*Phaseolus vulgaris* L.) in three different environments. *Cienc. Rural* 29: 441-448. <http://dx.doi.org/10.1590/S0103-84781999000300010>
- Cruz CD (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Sci. Agron.* 35: 271-276. <http://dx.doi.org/10.4025/actasciagron.v35i3.21251>
- Eberhart SA and Russell WA (1966). Stability parameters for comparing varieties. *Crop Sci.* 6: 36-40. <http://dx.doi.org/10.2135/cropsci1966.0011183X000600010011x>
- Finlay KW and Wilkinson GN (1963). The analysis of adaptation in a plant-breeding programme. *Crop Pasture Sci.* 14: 742-754. <http://dx.doi.org/10.1071/AR9630742>
- Haykin S (2009). Neural networks and learning machines. Prentice Hall, New Jersey.
- Nascimento M, Peternelli LA, Cruz CD, Nascimento ACC, et al. (2013). Artificial neural networks for adaptability and stability evaluation in alfalfa genotypes. *Crop Breed. Appl. Biotechnol.* 13: 152-156. <http://dx.doi.org/10.1590/S1984-70332013000200008>
- Pereira HS, Melo LC, Faria LC, Peloso MJD, et al. (2009). Adaptability and stability of common bean genotypes with carioca grain type for central Brazil. *Pesquisa Agropecu. Bras.* 44: 29-37. <http://dx.doi.org/10.1590/S0100-204X2009000100005>
- Pereira HS, Melo LC, Peloso MJD, Faria LC, et al. (2011). Complex interaction between genotypes and sowing seasons of carioca common bean in Goiás/Distrito Federal. *Crop Breed. Appl. Biotechnol.* 3: 207-215.
- R Development Core Team (2015). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at [<http://www.R-project.org>]. Accessed April 15, 2015.
- Ribeiro NF, Souza JF, Antunes IF and Poersch NL (2009). Yield stability of common bean cultivars of different commercial groups in Rio Grande do Sul State. *Bragantia* 68: 339-346. <http://dx.doi.org/10.1590/S0006-87052009000200007>
- Sant'Anna IC, Tomaz RS, Silva GN, Nascimento M, et al. (2015). Superiority of artificial neural networks for a genetic classification procedure. *Genet. Mol. Res.* 14: 9898-9906. <http://dx.doi.org/10.4238/2015.August.19.24>
- Silva GN, Tomaz RS, Sant'Anna IC, Nascimento M, et al. (2014). Neural networks for predicting breeding values and genetic gains. *Sci. Agric.* 71: 494-498. <http://dx.doi.org/10.1590/0103-9016-2014-0057>
- Sprent P and Smeeton NC 2007. Applied nonparametric statistical methods. Chapman and Hall/CRC, Boca Raton.
- Teodoro PE, Barroso LMA, Nascimento M, Torres FE, et al. (2015). Redes neurais artificiais para identificar genótipos de feijão-caupi semiprostrado com alta adaptabilidade e estabilidade fenotípicas. *Pesquisa Agropecu. Bras.* 50: 1054-1060. <http://dx.doi.org/10.1590/S0100-204X2015001100008>
- Torga PP, Melo PGS, Pereira HS, Faria LC, et al. (2013). Interaction of common beans cultivars of the black group with years, locations and sowing seasons. *Euphytica* 189: 239-248. <http://dx.doi.org/10.1007/s10681-012-0793-y>