



## Artificial neural networks reveal efficiency in genetic value prediction

L.A. Peixoto, L.L. Bhering and C.D. Cruz

Departamento de Biologia Geral, Universidade Federal de Viçosa,  
Viçosa, MG, Brasil

Corresponding author: L.A. Peixoto  
E-mail: [leoazevedop@gmail.com](mailto:leoazevedop@gmail.com)

Genet. Mol. Res. 14 (2): 6796-6807 (2015)

Received September 5, 2014

Accepted March 12, 2015

Published June 18, 2015

DOI <http://dx.doi.org/10.4238/2015.June.18.22>

**ABSTRACT.** The objective of this study was to evaluate the efficiency of artificial neural networks (ANNs) for predicting genetic value in experiments carried out in randomized blocks. Sixteen scenarios were simulated with different values of heritability (10, 20, 30, and 40%), coefficient of variation (5 and 10%), and the number of genotypes per block (150 and 200 for validation, and 5000 for neural network training). One hundred validation populations were used in each scenario. Accuracy of ANNs was evaluated by comparing the correlation of network value with genetic value, and of phenotypic value with genetic value. Neural networks were efficient in predicting genetic value with a 0.64 to 10.3% gain compared to the phenotypic value, regardless the simulated population size, heritability, or coefficient of variation. Thus, the artificial neural network is a promising technique for predicting genetic value in balanced experiments.

**Key words:** Backpropagation; Heritability; Testing; Programming; Improvement

## INTRODUCTION

Identification of superior genotypes requires selection methods able to efficiently exploit the available genetic material, maximizing the genetic gain in relation to the characteristics of interest (Oda et al., 2007). Several screening methods have been used in breeding programs, especially selection among and within families (Paula et al., 2002), combined selection (Martins et al., 2005), and selection by mixed models through the best linear unbiased prediction method (BLUP) (Garcia and Nogueira, 2005).

Additional genetic gains that enable the improvement of lines, hybrids, and commercial varieties have become increasingly difficult when considering species that undergo long selection processes. Thus, besides resources related to the determination of genetic designs, selection methods, and good agricultural experimentation, there is a new trend: the use of more refined analytical procedures, such as linear mixed models (Hiraoka et al., 2011) and artificial neural networks (Mugnai et al., 2008), for detailed studies of components of the mean and variance of a character, and in order to predict heritable variance, i.e., the genetic value.

In general, the biggest concern in relation to genetic improvement is the genetic value of the individual, so that selection is carried out with maximum accuracy. For better predicting the genetic value of a characteristic, it is possible to use phenotypic information about the individual itself or its relatives (descendants or ancestors), or information about other correlated characteristics. The combination of all this information has been studied by several biometricians.

Genetic value is based on the additive model, and has played an important role in selection gain of complex characteristics in plants and animals (Crossa et al., 2010). Besides this additive model, Bayesian interactions - BLUP (Piepho et al., 2008) and genomic selection (Jannink et al., 2010) - have been used in plants and animals (González-Recio et al., 2008).

A method that has recently been used for more efficient selection of families is the artificial neural network (ANN). It has been used by several researchers for image classification in remote sensing (Aitkenhead and Aalders, 2008), analysis of genetic diversity (Barbosa et al., 2011), identification of superior genotypes (Mugnai et al., 2008), and prediction of genetic value in animals (Ventura et al., 2012). However, there is no report on the use of ANNs for predicting genetic value in balanced experiments in plants. Therefore, the aim of this study was to evaluate the efficiency of ANNs in predicting genetic value in experiments with a randomized block structure.

## MATERIAL AND METHODS

### Simulation data

The design used for simulation of experiments was randomized blocks with six replications. Sixteen scenarios were simulated (experiments). For each scenario, there was a set of values for heritability (10, 20, 30, and 40%), simulated coefficient of variation (CV; 5 and 10%), and number of genotypes per block (150 or 200 genotypes) (Table 1). In order to validate neural networks, 100 populations were simulated (repetitions), keeping constant the genetic structure (variance and covariance matrix, mean heritability, and CV). The statistical model used in the simulation process was:

**Table 1.** Estimates of the parameters established for defining simulated populations. Experiments with two population sizes were simulated in randomized blocks with six replications (150 and 200 genotypes per block).

Number of individuals per population						
150			200			
Mean	h <sup>2</sup>	CV (%)	Mean	h <sup>2</sup>	CV (%)	
20	10	5	20	10	5	
20	10	10	20	10	10	
20	20	5	20	20	5	
20	20	10	20	20	10	
20	30	5	20	30	5	
20	30	10	20	30	10	
20	40	5	20	40	5	
20	40	10	20	40	10	

h<sup>2</sup> = heritability; CV = coefficient of variation.

$$Y_{ij} = \mu + G_i + B_j + \varepsilon_j \quad (\text{Equation 1})$$

where  $Y_{ij}$ : simulated observation of a given characteristic;  $\mu$ : overall mean of the characteristic, whose values are shown in Table 1;  $G_i$ : effect associated with the  $i^{\text{th}}$  genotype, being  $G_i \sim N(0, \sigma^2g)$ ;  $B_j$ : effect associated with the  $j^{\text{th}}$  block;  $\varepsilon_{ij}$ : Random error, being. For data simulation with the normal distribution, we used the variables proposed by the Box-Muller transform:

$$\varepsilon_j \sim N(0, \sigma^2) \quad x = \sqrt{-2\log_e(\text{RND})}V\text{Cos}(2\pi\text{RND}) \quad (\text{Equation 2})$$

$$y = \sqrt{-2\log_e(\text{RND})}V\text{Sen}(2\pi\text{RND}) \quad (\text{Equation 3})$$

with RND being an ANN random number. It has been shown that the x and y values obtained with the transform have a normal distribution with zero mean and variance V. As the simulation process demanded to obtain n data points with mean  $\mu$  and variance  $\sigma^2$ , a strategy was used to generate data (z) by the following expression:

$$z = \mu + \frac{1}{2\theta} \sum_{i=1}^{\theta} (x_i + y_i) \quad (\text{Equation 4})$$

$$V = 2\theta\sigma^2 \quad (\text{Equation 5})$$

where  $\theta$  is the repeatability of each simulated point. As the  $\theta$  value is set higher by the program's user, the simulation becomes more accurate but slower.

### Simulation of block effects

The data set contained  $n$  values in arithmetic progression, with ratio  $r$  and mean  $\bar{X}$ , in which the first term is denoted by  $X_1$  and the last by  $X_n$ , and with variance given by:

$$S^2 = \frac{n(n+1)}{3(n-1)^2} (X_n - \bar{X})^2 \quad (\text{Equation 6})$$

Thus, for estimating block effects, there existed  $b$  fixed effects, whose values constitute an arithmetic progression of the ratio  $r$ , with the particularity that  $B_1 = -B_b$  and . Therefore, the value  $B_b$  is estimated by:

$$B_b = \frac{(n-1)\sqrt{3\phi_b}}{\sqrt{n(n+1)}} \quad (\text{Equation 7})$$

and the other effects were established considering the ratio of the arithmetic progression are given by:

$$r = \frac{B_b - B_1}{b-1} \quad (\text{Equation 8})$$

### Simulation of effects of genotype

For estimating the effects of genotype, it was necessary to know the value of genetic variance, which was obtained from the information of heritability ( $h^2$ ), and the coefficient of variation (CV). Thus, we first obtained the environmental variance value by:

$$\sigma^2 = \left( \frac{\mu CV}{100} \right)^2 \quad (\text{Equation 9})$$

where  $\sigma^2$  is the environmental variance. Heritability ( $h^2$ ) is calculated by:

$$h^2 = \frac{100\sigma_G^2}{\sigma_G^2 + \frac{1}{b}\sigma^2} \quad (\text{Equation 10})$$

where  $\sigma_G^2$  is the genotypic variance. Consequently,

$$\sigma_G^2 = \frac{\sigma^2 h^2}{b(100 - h^2)} \quad (\text{Equation 11})$$

### Random effect of genotype

The effect associated with the  $i^{\text{th}}$  genotype was defined as  $G_i \sim NID(0, \sigma_G^2)$ . Once the

value of  $\sigma_g^2$  was known, effects were estimated using the random function described above.

### Simulation of random errors

Random errors are considered that  $\theta$ . Since the CVE and characteristic mean are available, consequently the value of  $\theta$  becomes known, and therefore, independent and random errors can be estimated using the random function described above.

All simulations were performed using the GENES software (Cruz, 2013).

### Establishment of phenotypic and genotypic values

Once the mean value of the characteristic and the effects involved were known, the phenotypic values of each variable were estimated by the model:

$$Y_{ij} = \mu + g_i + b_j + \varepsilon_{ij} \quad (\text{Equation 12})$$

Genotypic values were obtained as follows:

$$Z_i = \mu + G_i \quad (\text{Equation 13})$$

### Simulation of data for network training

For network training, simulated data were generated with the same characteristics as those for validation populations, defined by the mean, heritability, and CV.

For training purposes in this study, experiments with the same number of blocks (six) as the original experiments were simulated, but with an enhanced number of genotypes (5000). A larger number of genotypes will certainly benefit the process of predicting the correct value.

For this simulation, it was taken into account that each block represented a vector of phenotypic values to be reproduced with the same mean, variance, and covariance. The vector of genotypic values was also reproduced, considering that its variance and covariance together with the values of phenotypic vectors of blocks were represented by the estimate of the genetic variance. Thus, the concern was to obtain an enhanced data set [ $Z_{ij} \sim N(\mu, \sigma^2)$ , where  $i = 1, 2, \dots, 5000, j = 1, 2, \dots, 7$ , and a variance and covariance matrix sized  $7 \times 7$ , with  $Z_{i7}$  being the real genotypic value] from the original data [ $Y_{ij} \sim N(\mu, \sigma^2)$ , where  $i = 1, 2, \dots, 150$  or  $200, j = 1, 2, \dots, 6$ , and a variance and covariance matrix sized  $6 \times 6$ ].

The effectiveness of the training process was assessed by the correlation between network values and genetic value (output) of evaluated genotypes.

### Simulation of data for network validation

We used a procedure similar to that described in the simulation for training purposes; however, using a number of genotypes closer to the real test situation. We also used information from new experiments simulated with the same characterization of the mean, heritability, CVE, and variance and covariance matrix. However, only 150 or 200 genotypes were used. It is noteworthy that tests for validation purposes, although also derived from the simulation, are

not subsamples of the data set; i.e., they are new tests with environmental and genetic structure preserved.

### Architecture of the neural network

In this study, we used multilayer neural networks. The proposed neural network has 1 input layer, 3 intermediate layers, and 1 output layer. The input layer has 6 inputs (phenotypic value of the individual in the block). The input layer was based on a “ $n \times m$ ” matrix, where  $n$  is the number of blocks ranging from 1 to 6, and  $m$  the number of individuals in each block, which was 5000 in training experiments, and 150 or 200 in validation experiments. In the intermediate layer, the number of neurons per layer varied between one and 10 neurons in the first layer, one and 20 in the second, and one and 8 and in the third. The output layer consisted of one neuron, and the output was the genetic value of the population. This value was known in the training, but unknown in the validation. The best network architecture was established by the one with higher mean accuracy, considering the 43,200 possibilities, calculated by multiplying the number of neurons in each layer and the possible activation functions ( $10 \times 20 \times 8 \times 3 \times 3 \times 3$ ). The activation functions used were linear (purelin), hyperbolic tangent (tansig), logarithmic (Logsig), and Trainbr. Training lasted for 1000 epochs (iterations). The MATLAB 2011a software was used. The commands used for this ANN are found in [Supplementary material](#).

### Effectiveness of the ANN in genetic studies

The efficiency of the ANN was estimated by comparing the significance of correlation between network and genetic values in relation to the correlation between phenotypic means and genetic values. The square of the latter correlation represents the heritability of the characteristic analyzed.

## RESULTS AND DISCUSSION

### Obtaining the experimental data

Applying spectral decomposition was effective in simulating populations with the same structure as the original ones, such as heritability, CV, mean, and variance and covariance matrix. In a population with defined CV,  $h^2$ , and mean, it was possible to simulate the population with characteristics very similar to these ones (Table 2).

The simulated mean was identical to the expected mean in all simulated populations. Thus, the simulation process ensures the preservation of the expected mean in 100% of the cases.

The simulated CV ranged from 4.97 to 5.4 for an expected CV of 5. For an expected CV of 10, the simulated CV ranged from 9.91 to 10.02. Thus, simulation produced a population where the simulated CV ranged from 1.2 to 1.8% for an expected CV of 5 and 10%, respectively.

The simulated heritability ranged from 8.65 to 11.19, 19.3 to 21.67, 28.73 to 31.04, and 39.05 to 40.86% for an expected heritability of 10, 20, 30, and 40%, respectively. Thus, we observed a variation in simulated  $h^2$  of 25.54, 11.84, 7.7, and 4.525%, for expected  $h^2$  of 10, 20, 30, and 40%, respectively. Heritability was the parameter that presented the greatest variation. This fact is due to greater difficulty in keeping genetic variance constant. Once the expected  $h^2$  increases, the variation of simulated  $h^2$  decreases. Therefore, the present popula-

tion simulation is more efficient for qualitative characteristics, i.e., it is less influenced by the environment since it is possible to keep the expected variances, and thus keep constant the heritability value.

The size of the simulated populations (150 or 200 individuals) did not influence the simulation process. The mean did not change for either population size. The simulated CV ranged from 4.97 to 5.04 and 4.98 to 5.4% in populations with 150 and 200 genotypes, respectively, for an expected CV of 5%. With an expected CV of 10%, variation ranged from 9.93 to 1.10 and 9.91 to 2.10% for populations with 150 and 200 genotypes, respectively.

**Table 2.** Characterization of original and simulated populations.

Pop	$h^2_{exp}$	$CV_{exp}$	$X_{exp}$	$h^2_{sim}$	$CV_{sim}$	$X_{sim}$
150	10	5	20	11.02	5.01	20
150	10	10	20	9.21	10.01	20
150	20	5	20	21.67	4.97	20
150	20	10	20	20.28	10.01	20
150	30	5	20	30.82	5.03	20
150	30	10	20	30.18	10.00	20
150	40	5	20	39.05	5.01	20
150	40	10	20	40.86	9.93	20
200	10	5	20	8.65	5.04	20
200	10	10	20	11.19	10.01	20
200	20	5	20	21.12	4.99	20
200	20	10	20	19.3	10.02	20
200	30	5	20	31.04	5.00	20
200	30	10	20	28.73	10.01	20
200	40	5	20	40.00	4.98	20
200	40	10	20	39.38	9.91	20

Pop = number of genotypes per block;  $h^2_{exp}$  = expected heritability;  $CV_{exp}$  = expected coefficient of variation;  $X_{exp}$  = Expected mean;  $h^2_{sim}$  = simulated heritability;  $CV_{sim}$  = simulated coefficient of variation;  $X_{sim}$  = simulated mean of the characteristic.

## Obtaining enhanced data for network training

Obtaining training data becomes an extremely important practice for ensuring good efficiency in validating neural network estimations since networks require training data as a standard (Mather and Koch, 2011). Characteristics of training data selected for analysis are considerably important for ANN performance. Training data must be defined for analysis as being representative data of its population. Quality and size of data sets for training are essential for the efficiency of ANNs (Kavzoglu, 2009). Therefore, it is very important that the training data used for predicting the genetic value are representative of the original experiment, since the network learns from these data for further estimation of genetic value in other experiments.

The size of the training set is crucial for the performance of networks. There is a direct relationship between the size of a dataset for training and the reliability of network validation estimate data. Sample size is related mainly to the statistical characteristics of the neural networks used for training. A sample with few individuals is not large enough for a neural network to recognize all possible classes. A larger sample can make the network more specific and improve the reliability of results; however, it requires longer computational time to perform network training (Kavzoglu, 2001).

There are two ways of obtaining an efficient training data set for use in an ANNs: the first is the use of time series, and the second is data simulation. The use of time series is

reported by Ventura et al. (2012), who worked with data from 19,240 Tabapuã cattle from 152 farms located in several states between 1976 and 1995. They used the data to predict the genetic value of weight at 205 days of age, and used the Levenberg-Marquardt algorithm for network training. The authors concluded that genetic values obtained by ANNs were highly correlated with those obtained by BLUP.

However, when it is not possible to obtain a time series for composing a set of training data, it is possible to obtain these data through the simulation process, as performed in this study.

Aiming at achieving a training population that retained the genetic structure of the initial experiment, the genetic structure of the initial experiment was used in the simulation process of populations in the GENES software (Cruz, 2013). For simulation, the genetic structure of the initial experiment was maintained, i.e., mean, heritability, and CV were kept constant. Furthermore, it was possible to keep the matrix of variance and covariance consistent by using multivariate spectral decomposition. Thus, the training population was simulated with the same genetic structure as the initial population (Table 3).

We also observed that the training population simulated with 5000 genotypes per block kept the genetic structure of all populations initially simulated with 150 and 200 genotypes, i.e., heritability, mean, and CV were kept constant despite an increase in the number of genotypes in the experiment. Therefore, it is possible through an experiment carried out either in the field or greenhouse, to obtain through simulation an experiment with a larger number of genotypes that retains the genetic structure of the initial experiment.

A remarkable characteristic of the simulation process, besides retaining genetic structure, is that the value of the training population is also simulated. Through simulated genetic value, the network will be able to learn about that experiment and predict the real genetic value of the initial experiment.

**Table 3.** Genetic structure of training population with 5000 genotypes per block.

Pop	$h^2_{sim}$	$CV_{SIM}$	$X_{SIM}$	$h^2_{sim}$	$CV_{SIM}$
150	11.02	5.01	20	10.78	4.99
150	9.21	10.01	20	11.84	9.96
150	21.67	4.97	20	21.13	4.93
150	20.28	10.01	20	19.29	9.99
150	30.82	5.03	20	29.19	5.01
150	30.18	10.00	20	31.42	9.89
150	39.05	5.01	20	38.18	5.04
150	40.86	9.93	20	38.03	10.05
200	8.65	5.04	20	10.77	4.99
200	11.19	10.01	20	11.84	9.96
200	21.12	4.99	20	18.98	5.01
200	19.30	10.02	20	18.51	10.00
200	31.04	5.00	20	30.84	4.95
200	28.73	10.01	20	30.38	10.00
200	40.00	4.98	20	40.10	4.97
200	39.38	9.91	20	39.95	9.98

POP = number of genotypes per block in the initial experiment;  $h^2_{sim}$  = simulated heritability;  $CV_{SIM}$  = simulated coefficient of variation;  $X_{SIM}$  = characteristic simulated mean.

### Performance of neural networks for predicting genetic value

For predicting genetic value, we found that neural networks were superior to the use of the mean phenotypic value (Table 4). In relation to the phenotypic value, the gain was



greater than 80% in 75% of simulated networks. Therefore, using neural networks for predicting genetic value is better than using the phenotypic value since the network can reduce the environmental effect (noise) on a given experiment. Thus, the network value is closer to the real breeding value.

**Table 4.** Percentage of network gain (G) compared to the mean of 100 validation populations.

POP	h <sup>2</sup>	CV	G*
150	10	5	82
150	10	10	100
150	20	5	61
150	20	10	90
150	30	5	66
150	30	10	93
150	40	5	99
150	40	10	96
200	10	5	83
200	10	10	100
200	20	5	100
200	20	10	94
200	30	5	100
200	30	10	87
200	40	5	71
200	40	10	70

POP = number of individuals in each population; h<sup>2</sup> = heritability; CV = coefficient of variation. \*The gain of the network was calculated as the percentage of simulations in which the network value was higher than phenotypic value; G = 100\* (number of populations where the network was superior/total number of populations).

Chen et al. (2010) verified the accuracy of ANNs in the range of 88-94% for the classification of maize seed. The high accuracy verified with the use of ANNs can be explained through the use of nonlinear models to generate an output (response) (Dai et al., 2011).

The use of ANNs is important in agriculture due to its potential in solving complex problems for conventional computational and mathematical techniques (Huang et al., 2010), and particularly in the selection of genotypes since environmental effects may disguise the potential of some genotypes; thus, selection is erroneous. ANNs can decrease these environmental effects (noise) by backpropagation of error, which is a technique with great potential for prediction of genetic value.

We observed that there was no difference between simulated experiments with different values of heritability (10, 20, 30, and 40%) and CV (5 and 10%). Therefore, even if the environmental effect is high in the experiment, the network value can approach the real breeding value. Predictions of genetic value at different levels of heritability have been carried out by many researchers, especially for detecting quantitative trait loci (Meuwissen et al., 2001; Bernardo and Yu, 2007).

ANNs were also efficient for predicting genetic value for low (10 and 20%) and moderate (30 and 40%) heritability. Therefore, networks can be used for quantitative characteristics (low heritability), being a promising alternative for predicting genetic value. Some authors have also found high efficiency in predicting genetic value using neural networks (Cavero et al., 2008; Ventura et al., 2012).

The correlation between genetic value and network value was greater than the correlation between phenotypic value and network value in all simulated scenarios. The difference between these correlations ranged from 0.64 to 10.3% (Table 5).

**Table 5.** Correlation between genetic value and network value (r VG x VR), genetic value and phenotypic value (r VG x VF), and network value and phenotypic value (r VR x VF).

POP	h <sup>2</sup>	CV	r VG x VR	r VG x VF	r VR x VF
150	10	5	0.1536	0.1254	0.7758
150	10	10	0.1205	0.0984	0.8124
150	20	5	0.1972	0.1683	0.5962
150	20	10	0.2338	0.2111	0.9414
150	30	5	0.4069	0.3689	0.6699
150	30	10	0.3506	0.3313	0.9376
150	40	5	0.4283	0.3982	0.9334
150	40	10	0.4501	0.4075	0.8927
200	10	5	0.2242	0.1574	0.5496
200	10	10	0.1714	0.1173	0.6295
200	20	5	0.2142	0.1985	0.9184
200	20	10	0.2505	0.2302	0.9346
200	30	5	0.3385	0.3175	0.9414
200	30	10	0.3436	0.2989	0.8273
200	40	5	0.4278	0.4071	0.9179
200	40	10	0.3402	0.3294	0.9669

POP = number of individuals in each population; h<sup>2</sup> = heritability; CV = coefficient of variation.

Heritability is defined as the square of the genetic correlation. Thus, heritability calculated by the network was higher than that calculated by the phenotypic mean. Therefore, selection gain will be higher if the correlation between network value and genotypic value is used since selection gain is correlated with heritability.

There was no difference between experiments containing 150 or 200 genotypes per block (Table 5). Thus, it is possible to use artificial neural networks for predicting genetic value in experiments with a reduced number of genotypes (such as experiments at the final stage of a breeding program), or even with an enhanced number of genotypes (as in pre-breeding experiments that evaluate a large number of families).

Crossa et al. (2010) found correlations between BLUP and phenotypic values of 0.41 to 0.51. In this study, the correlation between network value and phenotypic value ranged between 0.5 and 0.98. This high correlation value is important since it was found that the network value is very close to the real value measured in the field. Heffner et al. (2009) concluded that the correlation between real and estimated genetic values is enough to consider the selection of the best genotypes in a breeding program, especially using genomic selection through molecular markers.

Crossa et al. (2010) concluded that the greatest gain selection, achieved by greater efficiency in the prediction of genetic value, may occur for the largest number of characteristics, or evaluated markers, or for the improvement of methods used for the prediction of genetic value. In the present study, ANNs were effective due to their high correlation with genetic value. Therefore, using ANNs for predicting genetic value can more accurately evaluate each genotype with no environmental effect. Thus, selection of the superior progenies in a breeding program will be more efficient using the superior genotypic value, and not the superior phenotypic value.

There was no difference in the prediction of genetic values for ANNs when the CV varied (5:10%) (Table 5). Therefore, it is possible to estimate the genetic value through neural networks, even when the variation within the block is high. Efficiency of selection in a breeding program depends on the prediction of genetic value from phenotypic value or from the efficiency of another criterion used (Lorenzana and Bernardo, 2009). Thus, ANNs may ease

the selection of genotypes based on genetic value, and reduce the error caused by the effect of environment on phenotype.

We found that in simulated experiments with lower heritability (10 and 20%), the difference between the correlation between network value and genetic value and the correlation between phenotypic value and genetic value was higher (Table 5). This is because when heritability is lower, the environmental effect of this characteristic is increased, i.e., noise is increased. Thus, the network can reduce the noise that occurs for characteristics with low heritability. When heritability is higher (30 and 40%), the environmental effect is lower, as is the noise. Thus, even if the network reduces noise, such a reduction is miniscule.

The standard genetic model, resulting in a phenotypic value, is the sum of genetic and environmental values (Crossa et al., 2010). Thus, since the value obtained in experiments is the phenotypic value, it is difficult to obtain the real genetic value if the environmental value is too high. For a breeding program, which aims at selecting the best genotypes, estimated genetic value is extremely important for decision making, with no influence from the environment. As ANNs use nonlinear equations for estimating genetic value, it is possible to get closer to the real genetic value.

## CONCLUSIONS

Artificial neural networks were effective at predicting genetic value in balanced experiments in random blocks for quantitative characteristics (heavily influenced by the environment).

## ACKNOWLEDGMENTS

We thank the editor, associate editor, and reviewers for their thoughtful and constructive comments. Research supported by CAPES, FAPEMIG, and CNPq.

## [Supplementary material](#)

## REFERENCES

- Aitkenhead M and Aalders I (2008). Classification of Landsat Thematic Mapper imagery for land cover using neural networks. *Int. J. Remote Sensing* 29: 2075-2084.
- Barbosa CD, Viana AP, Quintal SSR and Pereira MG (2011). Artificial neural network analysis of genetic diversity in *Carica papaya* L. *Crop Breed. Appl. Biotech.* 11: 224-231.
- Bernardo R and Yu J (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47: 1082-1090.
- Cavero D, Tölle KH, Henze C, Buxadé C, et al. (2008). Mastitis detection in dairy cows by application of neural networks. *Livestock Sci.* 114: 280-286.
- Chen X, Xun Y, Li W and Zhang J (2010). Combining discriminant analysis and neural networks for corn variety identification. *Computers Electronics Agric.* 71: S48-S53.
- Crossa J, Campos G de L, Perez P, Gianola D, et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics* 186: 713-724.
- Cruz CD (2013). GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scient. Agronomy* 35: 271-276.
- Dai X, Huo Z and Wang H (2011). Simulation for response of crop yield to soil moisture and salinity with artificial neural network. *Field Crops Res.* 121: 441-449.
- Garcia C and Nogueira M (2005). Utilização da metodologia REML/BLUP na seleção de clones de eucalipto. *Scie. Forestalis* 68: 107-112.
- González-Recio O, Gianola D, Long N, Weigel KA, et al. (2008). Nonparametric methods for incorporating genomic information into genetic evaluations: an application to mortality in broilers. *Genetics* 178: 2305-2313.

- Heffner EL, Sorrells ME and Jannink JL (2009). Genomic selection for crop improvement. *Crop Sci.* 49: 1-12.
- Hiraoka Y, Kuramoto N, Ohira M, Okamura M, et al. (2011). Estimation of genetic data and breeding values of traits related to wax production in *Rhus succedanea* L. clones using the REML/BLUP method. *J. Forest Res.* 16: 509-517.
- Huang Y, Lan Y, Thomson SJ, Fang A, et al. (2010). Development of soft computing and applications in agricultural and biological engineering. *Comput. Electr. Agr.* 71: 107-127.
- Jannink JL, Lorenz AJ and Iwata H (2010). Genomic selection in plant breeding: from theory to practice. *Brief. Func. Gen.* 9: 166-177.
- Kavzoglu T (2001). An investigation of the design and use of feed forward artificial neural networks in the classification of remotely sensed images. University of Nottingham.
- Kavzoglu T (2009). Increasing the accuracy of neural network classification using refined training data. *Environ. Model. Soft.* 24: 850-858.
- Lorenzana RE and Bernardo R (2009). Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *TAG. Theoretical and applied genetics. Theor. Angewandte Genetik* 120: 151-161.
- Martins IS, Cruz CD, Rocha MdGB, Regazzi AJ, et al. (2005). Comparação entre os processos de seleção entre e dentro e o de seleção combinada, em progênies de *Eucalyptus Grandis*. *Cerne* 11: 16-24.
- Mather P and Koch M (2011). Computer processing of remotely-sensed images: an introduction. John Wiley & Sons, Chichester.
- Meuwissen R, Linn SC, van der Valk M, Mooi WJ, et al. (2001). Mouse model for lung tumorigenesis through Cre/lox controlled sporadic activation of the K-Ras oncogene. *Oncogene* 20: 6551.
- Mugnai S, Pandolfi C, Azzarello E, Masi E, et al. (2008). *Camellia japonica* L. genotypes identified by an artificial neural network based on phyllometric and fractal parameters. *Plant System. Evol.* 270: 95-108.
- Oda S, Mello EJ, Silva JF and Souza ICG (2007). Melhoramento Florestal. In: Biotecnologia Florestal (A. Borém, ed.). UFV, Viçosa. 51-71.
- Paula RC, Pires IE, Borges RdCG and Cruz CD (2002). Predição de ganhos genéticos em melhoramento florestal. *Pesq. Agropec. Bras. Brasília* 37: 159-165.
- Piepho H, Möhring J, Melchinger A and Büchse A (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161: 209-228.
- Ventura R, Silva M, Medeiros T, Dionello N, et al. (2012). Use of artificial neural networks in breeding values prediction for weight at 205 days in Tabapuã beef cattle. *Arq. Bras. Med. Vet. Zoo.* 64: 411-418.