

ARTICLE

<http://dx.doi.org/10.1590/1984-70332015v15n2a14>

Selection in sugarcane families with artificial neural networks

Bruno Portela Brasileiro^{1*}, Caillet Dornelles Marinho¹, Paulo Mafra de Almeida Costa¹, Cosme Damião Cruz², Luiz Alexandre Peternelli¹ and Márcio Henrique Pereira Barbosa³

Received 4 March 2014

Accepted 7 November 2014

Abstract – *The objective of this study was to evaluate Artificial Neural Networks (ANN) applied in an selection process within sugarcane families. The best ANN model produced no mistake, but was able to classify all genotypes correctly, i.e., the network made the same selective choice as the breeder during the simulation individual best linear unbiased predictor (BLUPIS), demonstrating the ability of the ANN to learn from the inputs and outputs provided in the training and validation phases. Since the ANN-based selection facilitates the identification of the best plants and the development of a new selection strategy in the best families, to ensure that the best genotypes of the population are evaluated in the following stages of the breeding program, we recommend to rank families by BLUP, followed by selection of the best families and finally, select the seedlings by ANN, from information at the individual level in the best families.*

Key words: *Saccharum spp, artificial intelligence and breeding.*

INTRODUCTION

One of the most important steps in sugarcane breeding is the initial phase (T1), with the first selections of plants or families (Oliveira et al. 2008). Through the vegetative propagation process, the genotypes selected in this phase are forwarded to the following stages and planted in replicated designs to better identify the potentially superior, to include them in evaluation experiments (experimental phase, EF) in different locations and successive years. After phase T1, no new genotypes are introduced, that is, the genotypes of the stages T2, T3, FS, and FM are a subset of those in phase T1. Thus, the selection performed in T1 is crucial for the success of the program.

Although mass selection is routinely applied in the early stages of breeding programs, this type of selection has been criticized for its inefficiency due to the lack of replications, the competition between plants and the genotype-environment interaction (Kimbeng and Cox 2003).

The selection within families with high genotypic values may result in a greater probability of finding superior clones among the progenies (Barbosa et al. 2005). Based on this premise, family selection was routinely applied before developing clones in several sugarcane breeding programs

(Kimbeng and Cox 2003, Stringer et al. 2011, Barbosa et al. 2012).

Some alternative strategies to mass selection were suggested in the literature. Resende (2007) showed that the ideal selection strategy for sugarcane would be to predict genotypic values by the individual Best Linear Unbiased Predictor (BLUPI). This procedure uses data of both the family and plants for selection. However, the method is rarely used in breeding programs because of the operational problems related to the data acquisition at the individual plant level.

In practice, the selection of sugarcane families in the early breeding stages is based on the means or the sum of all plots. To circumvent the difficulties related to assessments at the individual level, Resende and Barbosa (2006) proposed to select the best families first and, in a second step, identify the best plants within these best families. This strategy selects families with genotypic values above the overall mean, followed by the simulation of the number of plants to be selected in each family according to the ratio between their genotypic values and the number of plants to be selected in the best family, resulting in a procedure called simulated individual BLUP (BLUPIS).

¹ Universidade Federal de Viçosa (UFV), Departamento de Estatística, 36.570-900, Viçosa, MG, Brazil. *E-mail: brunobiogene@hotmail.com

² UFV, Departamento de Biologia Geral

³ UFV, Departamento de Fitotecnia

The process of identifying the best genotypes suitable for selection based on BLUPIS can be enhanced by using Artificial Neural Networks (ANN) as a selection strategy after the individual assessment in the best families. These ANN are processing models that emulate a network of biological neurons, able to quickly recover a large amount of data and recognize patterns based on experience (Haykin 2009).

Neural networks offer an interesting alternative for acting as universal approximators with complex functions, capturing the nonlinear relationships between the explanatory variables and the response variable, learning about functional forms by adaptation, due to a sequence of transformations by parameter-controlled activation functions (Gianola et al. 2011).

Due to these capacities, neural networks have been used mainly in agronomic studies on pattern recognition for germplasm classification and selection (Pandolfi et al. 2009, Barbosa et al. 2011, Zhou et al. 2011), for adaptability and stability evaluation of genotypes (Nascimento et al. 2013), for yield prediction (Kaul et al. 2005, Ji et al. 2007, Zhang et al. 2010) and of complex quantitative traits (Gianola et al. 2011, Ventura et al. 2012).

The ANN models can increase the efficiency of the selection process in stage T1 of sugarcane breeding programs, due to their high capacity of discriminating genotypes with high and low yield potential (Zhou et al. 2011). In addition, the ANN consider all traits simultaneously during the selection process, be they continuous and/or categorical, whereas the visual selection poses a risk of using the selection criteria independently when judging the merit of some genotypes.

The selection in the T1 phase can be performed after a complete evaluation of the experiments and the visual inspection of the genotypes of the best families. The technical data such as total reducing sugars, sucrose content and fiber content, along with the yield traits stalk diameter, stalk number and stalk height, in addition to morphological and physiological traits such as stalk pith, flowering, size of axial bud and growth habit constitute the input to train the neural network for the selection of the best genotypes under evaluation.

The accumulation of experimental information, as well as meteorological data, pedigree information (genealogy), and molecular data over the years compose the inputs for training a neural network (Kaul et al. 2005, Ji et al. 2007, Gianola et al. 2011).

The purpose of ANNs is to establish a configuration that fits the training data appropriately, while preserving a high predictive capacity of the validation data. This can be accomplished by limiting the magnitude of the strength of

the network connections, for example through shrinkage, by a process known as regulation. Mackay (1992) and Titterton et al. (2004) proposed the application of Bayesian Regularization (BR) in neural networks. The BR circumvents the difficulties generated by the increased complexity of the neural network due to the large number of neurons involved in solving the problem, when estimating the effective number of network parameters (Mackay 1992).

Logistic regression analysis can also be used to predict the plants to be selected, aside from evaluating the effect of each agronomic trait used in decision making (Agresti 2007). Therefore, logistic regression, a conventional statistical method, may be useful in assessing the efficiency of ANNs in the selection process of sugarcane.

The purpose of this study was to evaluate the use of Artificial Neural Networks (ANN) by Bayesian Setting in the selection process of individual sugarcane plants within the best families.

MATERIAL AND METHODS

Plant material

The 128 families of half-sibs used in this study were derived from crosses performed at the Experimental Station of the Serra do Ouro, of the Federal University of Alagoas, in the municipality of Murici, Alagoas, in 2010.

After acclimatization, the seedlings were planted in the experiments at two locations: at the Center for Research and Improvement of Sugarcane of the Federal University of Viçosa, in the municipality of Oratory, Minas Gerais (lat 20° 25' S, long 42° 48' W; 494m asl); and at the plant Usina Coruripe in the municipality of Limeira do Oeste, MG (lat 19° 33' S, long 50° 34' W, 428m asl).

The 128 families were distributed in 7 experiments per location. Each experiment consisted of 20 families in a randomized block design with four replications. Cluster analysis was performed with two families (RB011532 x ? e SP80-3250 x ?) represented in all experiments. The 560 evaluated plots contained one 5-m long row each with 10 plants, spaced 1.40 m apart from each other, totaling 8,400 plants.

Phenotypic evaluation

The families were evaluated by estimating the yield in tons of cane per hectare, by weighing a sample of 10 stalks per plot: $TCH = (SN \times MSW \times 10)/AP$, where SN is the number of stalks per plot, MSW is the mean stalk weight and PA is the plot area in m² $PA = 7$.

The traits evaluated in plants of the selected families and underlying decision-making of the breeders were: stalk height (SH), stalk diameter (SD), stalk number (SN), internode length (IL) bud prominence (BP), presence of cracks (PC), smut incidence (SI), and plant vigor (PV).

The trait SH was evaluated (in m) in one stalk per plant, measuring from the ground to the first leaf of which the section between the leaf blade and sheath was visible. The SD (in mm) was measured at the third internode from the ground with a digital caliper, of one stalk per plant. The SN of each plant per plot was counted. For the traits IL, BP, PC as well as for SH and SD, the plants were classified as good (1) fair (0.5) and poor (0). For SI, the plants were separated in healthy (1) and diseased (0). For SN, the plants were grouped for good (1) and poor tillering (0). For PV, the plants were scored from 1 (lowest vigor) to 5 (highest vigor).

For the prediction of genotypic values of families for the trait TCH, data from the two locations (Oratórios, MG and Limeira do Oeste, MG) were used. However, individual assessments of plants were only performed in the experiments of Limeira do Oeste.

Selection by BLUPIS

Data of tons of cane per hectare (TCH) were analyzed by the mixed models REML/BLUP, using a statistical model associated with the evaluation of half-sib families in an incomplete block design with the plot mean, based on the statistical model described below by Resende (2007): $y = Xr + Zg + Wb + Ti + e$, where: y = data vector; r = Vector of replication effects added to the overall mean (assumed as fixed); g = Vector of genotypic effects (assumed as random); b = Vector of block effects (random); i = Vector of genotype \times environment interaction (random) and e = error vector (random). The letters X, Z, W, and T represent the incidence matrix, respectively, for the effects of r , g , b , and i .

Although the BLUPIS procedure indicates the selection of all families with means above the overall mean, we decided to select only 10% of the evaluated families, corresponding to 13 families with highest means for TCH. According to Simmonds (1996), approximately 60% of the best genotypes are concentrated in 10% of the best families in a population under evaluation.

The number of selected plants per family k ($k = 1, 2, \dots, j, \dots, 13$) was calculated as $n_k = (\hat{g}_k / \hat{g}_j) / n_j$, where \hat{g}_j is the genotypic value of the best family and n_j the number of selected plants in the best family, as recommended for the BLUPIS procedure (Resende and Barbosa 2006). In this study, n_j was considered equal to 16 plants. The selection

within families was performed based on the consensus of three breeders.

The analysis of the mixed models REML/BLUP described above was performed using software Selegen (Resende 2007).

Selection by the artificial neural network

The Artificial Neural Network (ANN) was trained with the data inputs of the traits used as selection criteria (SN, SD, SH, IL, BP, CR, SI and PV). Two ANN models were assessed; in model 1, quantitative data of yield components (SN, SH and SD) were used and in model 2 categorical data of these components.

The output of the network was the decision of three breeders together to select (1) or not select (0) within the best families (BLUPIS). The set of training and validation data of the neural network was also used in logistic regression.

The training data set consisted of 51 plants evaluated in replication 1 and the validation data set of 235 plants evaluated in 3 other replications of the best 13 families.

The inputs of ANN were normalized for the range [-1, 1], to improve the numeric stability, making the training network more efficient (Gianola et al. 2011, Ventura et al. 2012). The standardization of each variable is given by: $x_{\text{norm}} = -1 + \frac{(x - x_{\text{min}})}{(x_{\text{max}} - x_{\text{min}})} \cdot 2$, where: x_{norm} = normalized value of: x and x_{max} and x_{min} are the maximum and minimum values of the non-standardized data.

The criterion used for the network training stop depended on the number of iterations (samples 500) or on the occurrence of convergence on the surface of the sum of squared error ($E_D \leq 0.001$). The best network architecture consisted of three hidden layers, with six neurons per layer and using a logistic sigmoid activation function in the hidden layers and in the neuron of the network output.

The network was trained by the function 'trainbr' of software Matlab R2013b version 8.2. The 'trainbr' function updates the weight and bias values (bias is the term used in the machine learning literature for the intercept) according to the Levenberg-Marquardt training algorithm (LM) (Demuth et al. 2009). This function minimizes the combination of square errors and weights, and immediately determines the correct combination so as to produce a network with good generalization capability. This entire process is called Bayesian Regularization (BR) (Demuth et al. 2009).

With the Bayesian neural network, the individuals i , ($i = 1, 2, 3, \dots, n$) can be classified into selected (1) or discarded (0) by the following model:

$$\hat{y}_i = \sum_{k=1}^s w_k g_k(b_k + \sum_{j=1}^p a_{ij} \beta_j^{[k]}) + e_i,$$

where $e_i \sim N(0, \sigma_e^2)$; b_k = Bias of neuron k , ($k = 1, 2, 3, \dots$); W_k = Weight of each neuron k ; $\beta_j^{[k]}$ = Weight of input j , ($j = 1, 2, 3, \dots, p$) in neuron k of the network; a_{ij} = network inputs (independent variables); and g_k = activation function of neuron k . In this study, we used the logistic sigmoid activation function: $g_k = \frac{1}{1 + \exp^{-x_i}}$, where $x_i = (b + \sum_{j=1}^p a_{ij} \beta_j^{[k]}) + e_i$. The Bayesian Neural Network minimizes $F = \beta E_D + \alpha E_w$, where $E_D = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ = sum of squared error; E_w = quadratic sum of network parameters (weights and biases); $\beta = \frac{1}{2 \sigma_e^2}$; $\alpha = \frac{1}{2 \sigma_\theta^2}$; where σ_θ^2 = parameter of weight and bias dispersion.

Selection by logistic regression

Logistic regression was applied to determine the relationship between the characters used as selection criteria (explanatory variables) with the fact that the plant is or is not selected according to the application of BLUPIS. The set of training and validation data of logistic regression analyses was the same as for the ANN analyses. To validate the model, a cutoff of 0.5 was adopted, i.e., individuals i , ($i = 1, 2, 3, \dots, n$) with a selection probability above 0.5 were selected. The probability of selection was estimated using the following logistic regression model:

$$\hat{y}_i = \exp(\sum_{j=0}^k \beta_j a_{ij}) / (1 + \exp(\sum_{j=0}^k \beta_j a_{ij})),$$

where: β_j = regression coefficients; a_{ij} = explanatory variables (SN, SD, SH, IL, BP, CR, SI, and PV).

As in the analysis via artificial neural network, two logistic regression models were also evaluated. For Model 1, we used quantitative data of yield components (SN, SH and SD) and in Model 2 categorical data of these components,

as defined above in this paper. Logistic regression was performed with software R (<http://www.R-project.org>).

RESULTS AND DISCUSSION

A total of 130 plants were selected from the 13 best families by BLUPIS. Model 1 of ANN prediction produced no misclassification, but was able to classify all genotypes correctly, i.e., the ANN made the same selection choice as the breeder during the implementation of BLUPIS. In Model 2 of ANN prediction, 12 plants were misclassified, with an apparent error rate (AER) of 5.10% (Table 1).

In the selection using the best logistic regression model (model 1), the probability of selection of 80 plants of the validation population exceeded 0.5. Twenty-five other plants selected by BLUPIS did not reach the selection threshold, while 26 genotypes that were not selected by BLUPIS were eventually selected by the logistic model.

In total, 51 plants were misclassified by the best logistic regression model. In Model 2 of logistic regression, 55 classifications were incorrect. The apparent error rate (AER) of Model 1 and 2 was 21.70% and 23.40%, respectively (Table 1). These results indicate a slight superiority of one model to identify the best plants.

According to the values of the odds ratios, the variables PV, BP and SI were the most important traits in the selection process (Table 2). Bud prominence (BP) is an important trait for clone selection, since cultivars with large buds are undesirable for the mechanical planting process and selection of genotypes with this trait is therefore avoided. Smut incidence (SI) was important in this study because of the high disease incidence in the population. In the case of plant vigor (PV), the great importance of this variable is

Table 1. Data validation (235 plants) by different logistic regression models and via artificial neural network, where the dependent variable was the within family selection practiced as BLUPIS

Logistic regression				
BLUPIS	Model 1*		Model 2	
	Select	Discard	Select	Discard
Select	80	25	77	28
Discard	26	104	27	103
AER%**	21.70		23.40	
Artificial Neural Network				
BLUPIS	Model 1		Model 2	
	Select	Discard	Select	Discard
Select	105	0	97	8
Discard	0	130	4	126
AER%**	0.00		5.10	

* Model 1 used quantitative data of the traits SN, SH, and SD; Model 2 used categorical data of the traits SN, SH and SD; **AER = $\frac{1}{N} \sum_{j=1}^g m_j$, where N = total number of observations, g = number of responses (Select or Discard), m_j = number of misclassifications per response.

easily explained, since plants with greater vigor are the most interesting for breeders because they usually have high SN, SH and SD, aside from being disease-free.

In model 1, according to the odds ratio a plant with grade 4 is $1.60 \times 10^{+17}$ times more likely to be selected than a plant with a grade below 4. If a plant has a bud considered good (small bud), the chances of being selected are 371 times higher than a plant with poor buds. And a plant without smut incidence (SI) is eight times more likely to be selected than a plant with the disease (Table 2).

The input data used in the different logistic regression models and by ANN had great influence on the results. Both in the training of the neural network as in logistic regression, the continuous use of SN, SH, SD data resulted in better adjusted models (Table 1).

The inferior results of model 2 by both logistic regression and ANN may be related to the difficulty in classifying the yield components (Figure 1).

Some plants that should be classified as good for SN were misclassified with the rating poor. For SD and SH was also confusion in the separation of plants classified as good, fair and poor. Only very low or very high plants were classified correctly, as also observed for stalk diameter (SD), where only plants with a very large or small diameter were

classified appropriately. For SD and SH the difficulty was in defining the category fair, since plants that should receive this rating were misclassified as good or poor (Figure 1).

The use of categorical variables only as inputs for the neural network training (model 2) was an attempt to facilitate the individual assessment, since the acquisition of quantitative data of the traits SN, SD and SH requires more time and labor. Although the ANN model adjusted for categorical data produced AER, the error was low, with only 12 misclassified plants. The improvement of the classification process of the traits SN, SD and SH, along

Table 2. Coefficients and odds ratio estimated by logistic regression based on data of Model 1, where quantitative data of the traits SN, SH and SD were used.

Trait	Coefficient (β)	Odds Ratio = exp (β)
SN*	0.0178	1.018
SH	0.2128	1.2371
SD	0.0517	1.053
IL=0	-0.8759	0.4165
IL=0.5	0.8832	2.4186
IL=1	0.0377	1.0384
BP=0	-21.1216	0.000
BP=0.5	5.6288	278.3209
BP=1	5.9173	371.4035
PC=0	-0.1841	0.8319
PC=0.5	-2.0535	0.1283
PC=1	0.5362	1.7095
SI	2.1366	8.4708
PV=1	-7.4434	0.0006
PV=2	-5.5383	0.0039
PV=3	-1.3432	0.261
PV=4	39.6131	$1.60 \times 10^{+17}$
PV=5	32.6124	$1.46 \times 10^{+14}$
Intercept	-6.412	

* SN: Stalk number, SH: stalk height (m), SD: stalk diameter (mm), IL: internode length, BP: bud prominence, PC: presence of cracks, SI: smut incidence and PV: plant vigor.

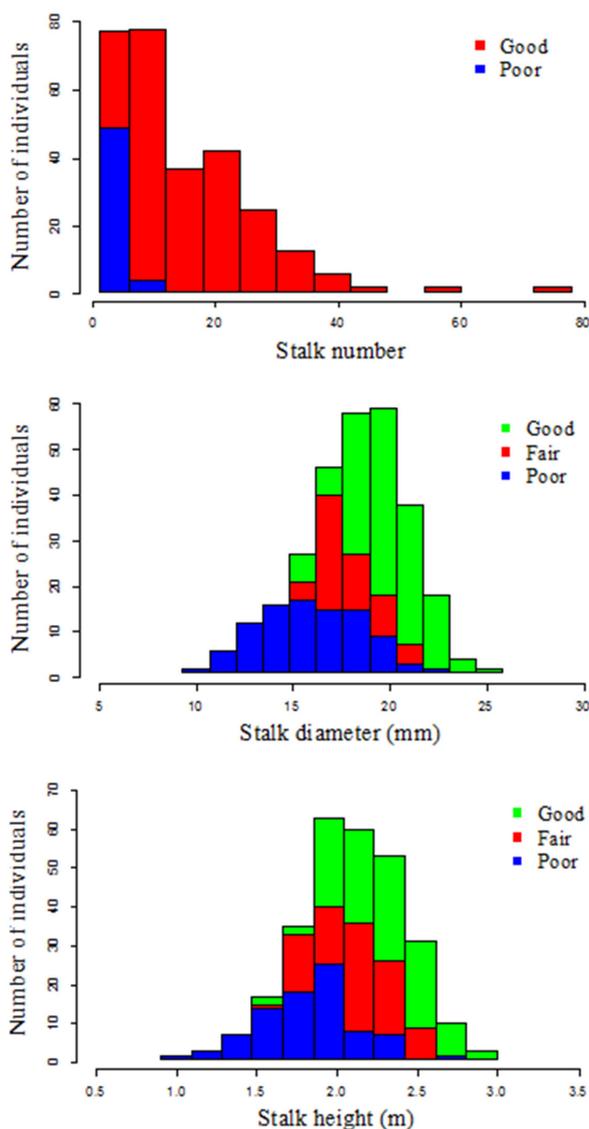


Figure 1. Histograms with continuous data distribution and the respective classification (good, fair and poor) of each plant for the traits stalk number (SN), stalk diameter (SD) and stalk height (SH).

with the greater data accumulation to train the ANN, could make selection via artificial intelligence a reality in sugarcane breeding programs.

One way to overcome the difficulties of classification of SN, SH and SD would be by counting the stalk number and by using devices to facilitate the genotype classification for stalk diameter and height. For example, forks with spaces between the teeth can be used to separate the stalks in thin (< 20mm), medium (20-25mm), thick (26-30mm), and very large (> 30mm). Short, medium and tall stalks can also be separated easily by measuring tapes with marks of the three possible categories, according to the ranges to be defined previously for each class. The establishment of ranges for each class may vary depending according to the weather conditions of each growing season and the type of environment in which families are evaluated.

Input data used in the neural network, consisting of traits usually analyzed during visual selection, are easily and quickly assessed, even by staff with little experience. In addition, the possibility of using continuous, binary and/or multicategorical traits as inputs for training the network makes the application of artificial intelligence even more interesting in sugarcane breeding programs selecting potential clones, in view of the importance of many of the qualitative traits.

The results demonstrate the ability of ANN to make the best decision using the same criteria used by the professionals simultaneously responsible for selection in phase T1 of sugarcane breeding programs. Moreover, the neural network can be trained according to the type of selection determined by the breeder (selection with higher or lower standards).

Since mass selection is performed when not all plants were evaluated, but at the moment when each plant is

inspected during the visual assessment of the population under selection, given the quantity of plants taken to the field (~200,000 seedlings per year) and the time required for evaluation of the population, superior genotypes can be discarded and undesirable genotypes can be selected. In addition, this selection is not done by a single professional, but by a team of technical and practical staff, which hampers a standardization of the selection process. Therefore, the use of ANN based on an effective training ensures better decisions, namely, the selection of the best genotypes based on pre-defined standards.

To develop new selection strategies in the best families, so that the best genotypes of a population are evaluated in the next stages of the breeding program, we recommend ranking the families by BLUP using the plot means for TCH, followed by the selection of the best families and finally, select the genotypes via artificial neural networks, from data at the individual level collected in only the best families.

In the future, neural networks can also be used for the selection among families and can come to substitute conventional methods, e.g., analysis of variance and mixed model procedures by REML/BLUP, to predict the genotypic values of families. Studies addressing this issue are also being carried out in our research group and preliminary results have demonstrated the potential of ANN in the classification of the best families.

CONCLUSION

Artificial Neural Networks (ANN) have a great potential in the individual selection process within the best sugarcane families, contributing to the standardization of the selection process and therefore the identification of the best genotypes, increasing the efficiency of sugarcane breeding programs.

Seleção dentro de famílias de cana-de-açúcar via redes neurais artificiais

Resumo – O objetivo desse trabalho foi avaliar o uso de Redes Neurais Artificiais (RNA) na seleção dentro de famílias de cana-de-açúcar. O melhor modelo de RNA não apresentou erro, sendo capaz de classificar corretamente todos os genótipos, ou seja, a rede tomou a mesma decisão de seleção realizada pelo melhorista durante a aplicação do BLUP individual simulado (BLUPIS), demonstrando a capacidade de aprendizado da RNA a partir das entradas e saídas informadas nas fases de treinamento e validação. Tendo em vista que a seleção via RNA facilita a identificação dos melhores indivíduos e visando desenvolver uma nova estratégia de seleção dentro das melhores famílias, de forma a garantir que os melhores genótipos da população sejam avaliados nas próximas fases do programa de melhoramento, recomendamos, ranquear as famílias via BLUP, selecionar as melhores e realizar a seleção individual via RNA, a partir das informações coletadas em nível individual nas melhores famílias.

Palavras-chave: *Saccharum spp.*, inteligência artificial e melhoramento genético.

REFERENCES

Agresti A (2007) **An introduction to categorical data analysis**. 2nd edn, John Wiley & Sons, New Jersey, 372p.

Barbosa CD, Viana AP, Quintal SSR and Pereira MG (2011) Artificial

neural network analysis of genetic diversity in *Carica papaya* L. **Crop Breeding and Applied Biotechnology** 11: 224-231.

Barbosa MHP, Resende MDV, Bressiani JA, Silveira LCI and Peternelli LA (2005) Selection of sugarcane families and parents by Reml/Blup. **Crop Breeding and Applied Biotechnology** 5: 443-450.

- Barbosa MHP, Resende MDV, Dias LAS, Barbosa GVS, Oliveira RA, Peternelli LA and Daros E (2012) Genetic improvement of sugar cane for bioenergy: the Brazilian experience in network research with RIDESA. **Crop Breeding and Applied Biotechnology S2**: 87-98.
- Demuth H, Beale M and Hagan M (2009) **Neural network toolbox™ 6 user's guide**. The MathWorks, Natick, 405p.
- Gianola D, Okut H, Weigel KA and Rosa GJM (2011) Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. **BMC Genetics 12**: 1-14.
- Haykin S (2009) **Neural networks and learning machines**. 3rd edn, Prentice Hall, New Jersey, 936p.
- Ji B, Sun Y, Yang S and Wan J (2007) Artificial neural networks for rice yield prediction in mountainous regions. **Journal of Agricultural Science 145**: 249-261.
- Kimbeng CA and Cox MC (2003) Early generation selection of sugarcane families and clones in Australia: a review. **Journal American Society of Sugarcane Technologists 23**: 20-39.
- Kaul M, Hill RL and Walthall C (2005) Artificial neural networks for corn and soybean yield prediction. **Agricultural System 85**: 1-18.
- Mackay DJC (1992) Bayesian Interpolation. **Neural Computation 4**: 415-447.
- Nascimento M, Peternelli LA, Cruz CD, Nascimento ACC, Ferreira RP, Bhering LL and Salgado CC (2013) Artificial neural networks for adaptability and stability evaluation in alfalfa genotypes. **Crop Breeding and Applied Biotechnology 13**: 152-156.
- Oliveira RA, Daros E, Bepalhok-Filho JC, Zambon JLC, Ido OT, Weber H, Resende MDV and Zeni-Neto H (2008) Seleção de famílias de cana-de-açúcar via modelos mistos. **Scientia Agraria 9**: 269-274.
- Pandolfi C, Mugnai S, Bergamasco S, Masi E and Mancuso S (2009) Artificial neural networks as a tool for plant identification: A case study on Vietnamese tea accessions. **Euphytica 166**: 411-421.
- R Development Core Team (2014): A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <<http://www.R-project.org>> Accessed on Jan 15, 2014.
- Resende MDV (2007) **Software SELEGEN-REML/BLUP: sistema estatístico e seleção genética computadorizada via modelos lineares mistos**. Embrapa Florestas, Colombo, 359p.
- Resende MDV and Barbosa MHP (2006) Selection via simulated Blup base on family genotypic effects in sugarcane. **Pesquisa Agropecuária Brasileira 41**: 421-429.
- Simmonds NW (1996) Family selection in plant breeding. **Euphytica 90**: 201-208.
- Stringer JK, Cox MC, Atkin FC, Wei X and Hogarth DM (2011) Family selection improves the efficiency and effectiveness of selecting original seedlings and parents. **Sugar Tech 13**: 36-41.
- Titterton DM (2004) Bayesian methods for neural networks and related models. **Statistical Science 19**: 128-139.
- Ventura RV, Silva MA, Medeiros TH, Dionello NL, Madalena FE, Fridrich AB, Valente BD, Santos GG, Freitas LS, Wenceslau RR, Felipe VPS and Corrêa GSS (2012) Uso de redes neurais artificiais na predição de valores genéticos para peso aos 205 dias em bovinos da raça Tabapuã. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia 64**: 411-418.
- Zhang L, Zhang J, Kyei-Boahen S and Zhang M (2010) Simulation and prediction of soybean growth and development under field conditions. **American-Eurasian Journal of Agricultural & Environmental Sciences 7**: 374-385.
- Zhou MM, Collins A, Kimbeng CA, Tew TL, Gravois KA and Pontif MJ (2011) Artificial neural network models as a decision support tool for selection in sugarcane: a case study using seedling populations. **Crop Science 51**: 21-31.