

# DETERMINATION OF THE INFLUENCE OF THE VARIATION OF REDUCING AND NON-REDUCING SUGARS ON COFFEE QUALITY WITH USE OF ARTIFICIAL NEURAL NETWORK

JOSÉ A. T. MESSIAS<sup>1</sup>, EVANDRO DE C. MELO<sup>2</sup>, ADÍLIO F. DE LACERDA FILHO<sup>3</sup>, JOSÉ L. BRAGA<sup>4</sup>, PAULO R. CECON<sup>5</sup>

**ABSTRACT:** The present study aimed at evaluating the use of Artificial Neural Network to correlate the values resulting from chemical analyses of samples of coffee with the values of their sensory analyses. The coffee samples used were from the *Coffea arabica* L., cultivars Acaia do Cerrado, Topázio, Acaia 474-19 and Bourbon, collected in the southern region of the state of Minas Gerais. The chemical analyses were carried out for reducing and non-reducing sugars. The quality of the beverage was evaluated by sensory analysis. The Artificial Neural Network method used values from chemical analyses as input variables and values from sensory analysis as output values. The multiple linear regression of sensory analysis values, according to the values from chemical analyses, presented a determination coefficient of 0.3106, while the Artificial Neural Network achieved a level of 80.00% of success in the classification of values from the sensory analysis.

**KEYWORDS:** coffee quality, chemical analysis, classification, computer systems.

## DETERMINAÇÃO DA INFLUÊNCIA DA VARIAÇÃO DE AÇÚCAR REDUTOR E NÃO REDUTOR NA QUALIDADE DA BEBIDA DO CAFÉ COM UTILIZAÇÃO DE REDE NEURAL ARTIFICIAL

**RESUMO:** Este trabalho teve como objetivo avaliar a utilização de Rede Neural Artificial para correlacionar os valores resultantes de análises químicas de amostras de café com os valores de sua análise sensorial. As amostras de café utilizadas foram referentes ao café *Coffea arabica* L., cultivar Acaia do Cerrado, Topázio, Acaia 474-19 e Bourbon, coletados na região sul de Minas Gerais. As análises químicas foram de açúcar não redutor e açúcar redutor, sendo a qualidade da bebida avaliada pela análise sensorial. O método de Rede Neural Artificial utilizou os valores das análises químicas como variáveis de entrada e os valores da análise sensorial como valores de saída. A regressão linear múltipla dos valores de análise sensorial, em função dos valores das análises químicas, teve coeficiente de determinação de 0,3106, enquanto o método de Rede Neural Artificial empregado obteve um nível de acerto na classificação dos valores da análise sensorial de 80,00%.

**PALAVRAS-CHAVE:** qualidade de café, análises químicas, classificação, sistemas de computação.

### INTRODUCTION

The study of processes that aim to determine the factors affecting the beverage quality of coffee is the focus of numerous academic papers. These studies establish several characteristics that interfere directly or indirectly in the quality of the beverage. The analysis of these variables is traditionally performed with the use of statistical tools. These used to obtain a model that explains the variation in characteristics but typically they do not provide a model to efficiently explain the correlation between these characteristics and beverage classification.

<sup>1</sup> Eng<sup>o</sup> Civil, Prof. Adjunto, Departamento de Computação, UFOP, Ouro Preto - MG.

<sup>2</sup> Eng<sup>o</sup> Agrícola, Professor Associado, Departamento de Engenharia Agrícola, UFV, Viçosa - MG.

<sup>3</sup> Eng<sup>o</sup> Agrônomo, Professor Associado, Departamento de Engenharia Agrícola, UFV, Viçosa - MG.

<sup>4</sup> Eng<sup>o</sup> Eletricista, Professor Titular, Departamento de Informática, UFV, Viçosa - MG.

<sup>5</sup> Eng<sup>o</sup> Agrônomo, Professor Adjunto, Departamento de Estatística, UFV, Viçosa - MG.

Recebido pelo Conselho Editorial em: 22-7-2010

Aprovado pelo Conselho Editorial em: 10-11-2011

The question analyzed in this study was to establish the correlation between characteristic variations, as determined by chemical analyses, presented by the coffee beans due to crop management practices, harvesting, post-harvest processing and storage, as a result of beverage quality.

The objective was to propose a method using Artificial Neural Network (ANN), which, from experimental data, may establish a model able to correlate the input and output variables and generate classes according to output variable values. In this particular case, the input values are the results of chemical analysis of coffee samples and the output value, the quality of the beverage, according to sensory analysis of beverage samples.

Due to its modeling characteristics, the use of ANN enables the establishment and recognition of classes according to their characteristics. The use of ANN involves the solution of the problem under study, namely to establish a method that determines the correspondence between the variations of sensory analysis values and the variations of chemical analyzes values.

The ANN is used today in various areas of knowledge, and this broad spectrum is increasing ANN research and development. In the studies by CHERKASSKY & MA (2009) and KHOSHGOFTAAR & GAO (2007), target function in regression problems and classification were studied, while studies by HERVÁS-MARTÍNEZ et al. (2008), CIURANA et al. (2008), SOUSA et al. (2007), and VERLINDEN et al. (2008) compared the performance between ANN and MLR. RECAMALES et al. (2007) established a correlation process between wine color stored with different wooden barrels, DUTTA et al. (2003) established the process for determining tea quality, and PARDO et al. (2006), KERMANI et al. (2005) and CHARUMPORN et al. (2003) developed study to identify volatile chemicals with the use of "electronic nose".

SUCHACZ & WESOLOWSKI (2006) analyzed the concentrations of zinc, copper, lead and cadmium in medicine, using them as input variables for neural network models to recognize the taxonomy of plants used in their composition. Research developed by ANDERSON & SMITH (2002) demonstrated the feasibility of a method to distinguish coffee from different geographic regions. Elemental analyzes of K, Mg, Ca, Na, Al, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Mo, S, Cd, Pb and P from coffee bean samples served as input in the system using methods of statistical pattern recognition, including principal component analysis, discriminant function, as well as analysis and modeling of neural networks, to determine the geographical origin of food.

When verifying the influence of the drying process, in different types of patios, on the coffee quality during storage, REINATO et al. (2007) used the analysis of total sugars, among other compounds, and reported the decrease of total sugars in coffee dried in a thick layer, that provided the best conditions for microorganisms, which utilize sugar as a major substrate for development.

In order to determine the coffee quality at different temperatures, airflow and period of pre-drying, BORÉM et al. (2006) used sensory analysis and chemical analyzes of total and reducing sugars. The highest values of total and reducing sugars were indicators of better quality of coffee beverage. These sugars specified flavor and aroma of roasted coffee.

## **MATERIAL AND METHODS**

The data used in this study were provided by the team of the Professor Flavio Meira Borém, from the Department of Engineering, at the Federal University of Lavras, State of Minas Gerais, Brazil, whose data are characterized in the references: ISQUIERDO (2008), 30 sets; RIBEIRO (2009), 84 sets; NOBRE (2009), 70 sets, FIGUEIREDO (2009), 115 sets, and FERNANDES (2008), 96 sets. In total, 395 sets of data were used for implementation of the ANN. These data are from *Coffea arabica* L.; cultivars Acaia do Cerrado, Topázio, Acaia 474-19, and Bourbon group. The coffee beans were derived predominantly from the southern region of Minas Gerais, in the municipalities of Lavras, Três Pontas, Santo Antônio do Amparo, Guaxupé, and Machado. The procedures for harvesting, drying and storage were specific for each experiment. The analyses were

appropriate for the purposes of each experiment, but the chemical analyses of non-reducing sugars were common to all. The sensory analysis, either BSCA or SCAA methods, were also common to all experiments.

The linear regression analysis was performed with the SAEG program (2007). The linear regression of the sensory analysis results was carried out as a function of the results from the chemical analyses.

The ANN was successfully implemented, trained and tested using MATLAB<sup>®</sup>, version 6, Math Works Inc., with the neural networks tool kit, to create a network object in MATLAB<sup>®</sup>.

Theoretically, the value of sensory evaluation is between 0 and 100. For implementation of the method, values of the sensory analysis were grouped as follows: class 1 for values greater than 79; class 2 to values less than or equal to 79 and greater than 74; class 3 to less than or equal to 74 and greater than 69; and class 4 to less than or equal to 69. In the 395 data sets, chemical composition and sensory analysis used to implement the ANN were divided randomly into three groups: the first one, used for training the network (around 80% of data), a second group for validation during a network training (around 10% of the data), and a third group to test the network generated (around 10% of the data).

The network performance was verified by comparing the expected result for the third data group, with the ANN results. Each data set used results of chemical composition data as input for the ANN, and the results of sensory analysis as output values. The use of the sensory analysis results was done during the training phase to determine the error to be used in the measurement of training, and in the testing phase, to determine the ANN performance.

After all initial tests had been performed, a set of parameters were defined for routine in the ANN implementation during the performance tests, with the use of:

- Multilayer Progressive Neural Network (feed forward) with back propagation learning algorithm for the error.
- Six input values, one for each associated variable of the chemical analysis value.
- Two intermediate layers, the first with 5; 7 or 9 neurons and the second with 5; 7; 9 or 13 neurons.
- Output layer with a neuron associated with the value of sensory analysis.
- *Logsig* function as activation function of hidden layers and *purelin* as last layer.
- Learning algorithm *trainlm*;
- 300 cycles (epocs) for each execution.
- Accuracy of the mean squared error for stop criterion to 0.0001.
- 20 executions for each network configuration by selecting the configuration with best performance.

## RESULTS AND DISCUSSION

Data analysis was executed in order to verify the influence of data variation of chemical analysis values and sensory analysis values of coffee. Data analysis was performed using statistical analysis and ANN. Initially, interference of variation was observed for each chemical analysis in the sensory analysis, and then the aggregate interference of two chemical analyses in the sensory analysis.

The sensory analysis showed the following values: lowest value (44), highest value (82), average (69.56) and standard deviation (12.11).

**Analysis of reducing sugar**

Results of reducing sugar analyses showed the lowest value: 0.15, highest value: 0.71, average value: 0.3819, and standard deviation: 0.1254.

Linear regression of the results for reducing sugar (ARd) and sensory analysis (ASe) of the 395 sets of experimental data resulted in eq.(1), which coefficient of determination  $r^2$  was 0.1400. Figure 1 represents the results of linear regression.

$$\hat{ASe} = 83.3774 - 36.1581 \cdot ARd \tag{1}$$

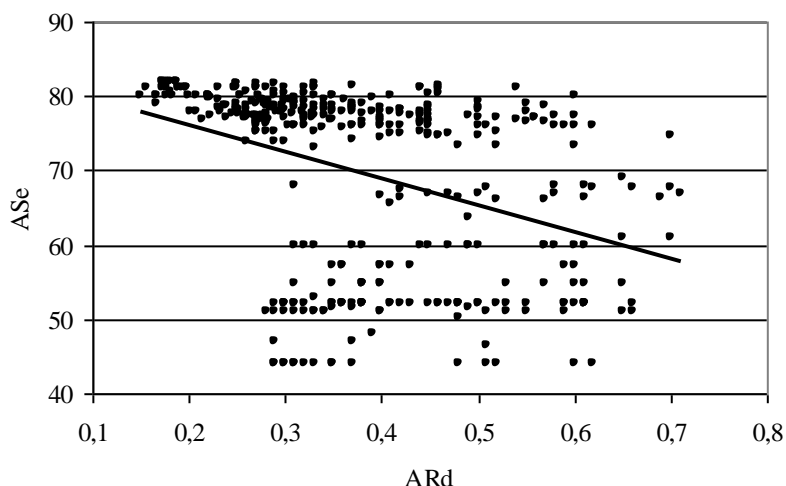


FIGURE 1. Linear Regression  $ASe = f(ARd)$ .

The ANN implemented with the results from the reducing sugar analysis and sensory analysis presented the best accuracy classification for the value of 42.50% in the configuration of (7; 13; 1) neurons per layer.

The ANN performed better than linear regression, since the coefficient of determination indicates that only 14% of the results of sensory analysis are explained by the regression. The network performance, despite indicating a small success rate, with less than half of the cases, was still better than the linear regression.

**Analysis of Non-Reducing Sugar**

Data analyses of non-reducing sugar showed the following values: lowest value of 1.95; highest value of 10.38; average values of 6.5498, and standard deviation of 1.6184.

Linear regression analysis results for non-reducing sugar (AnR) and sensory analysis (ASe) of the 395 sets of experimental data resulted in equation 2, which coefficient of determination  $r^2$  was 0.1016. In Figure 2, the results of linear regression are shown.

$$\hat{ASe} = 53.93 + 2.38766 \cdot AnR \tag{2}$$

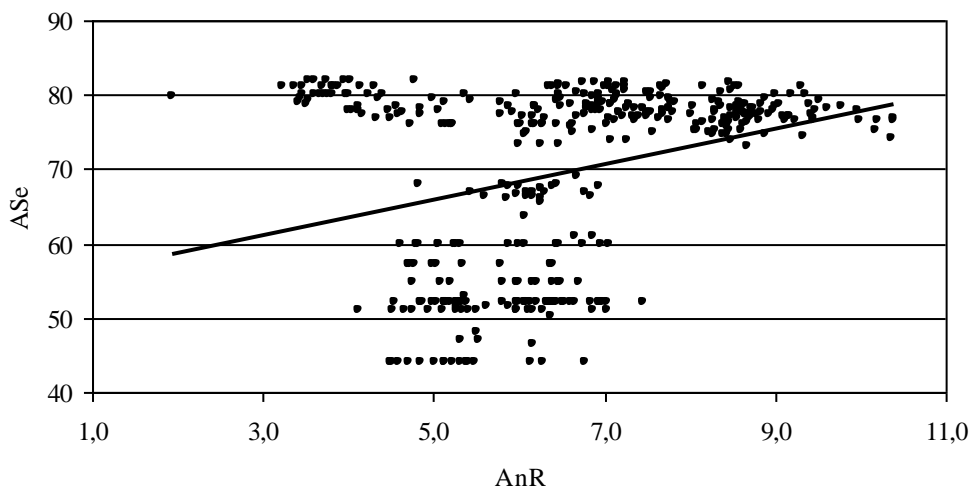


FIGURE 2. Linear Regression  $ASe = f(AnR)$ .

The ANN implemented with the results of analysis of non-reducing sugar and sensory analysis presented the best accuracy classification for the value of 62.50% in the configuration of (7; 7; 1) and (9; 7; 1) neurons per layer.

In this case also, the ANN performed better than the linear regression, since the coefficient of determination indicated that only 10.16% of sensory evaluation results were explained by the regression.

### Analysis of Reducing and Non-Reducing Sugars

The variation of sensory analysis values as a function of variations in the values of reducing and non-reducing sugar analyses assessed by use of multiple linear regression analysis in the 395 sets of experimental data, resulted in the equation 3, which coefficient of determination  $R^2$  was 0.3106.

$$\hat{ASe} = 66.0923 - 45.3181 \times ARd + 3.17312 \times AnR \quad (3)$$

The ANN implemented with the results of non-reducing sugar and sensory analyses resulted in best classification accuracy of 80% in the configuration (7; 13; 1) neurons per layer.

In this case, the ANN also had better performance than the linear regression, since the coefficient of determination indicated that only 31.06% of sensory evaluation results were explained by the regression. This combined analysis resulted in significant improvement, since the performance result of the ANN classification was 80% correct, being significantly superior to previous implementations, which showed best performance of 62.50%.

### CONCLUSION

This work demonstrated the potential of the ANN method to correlate analysis of reducing and non-reducing sugars results with sensory evaluation.

Data analysis of values for each chemical analysis and sensory analysis using simple linear regression and ANN was not efficient.

The method of using ANN to evaluate data resulted from chemical and sensory analyses showed correspondence between the variation of chemical analyses to those obtained by sensory analysis, when analyzed together.

There is a direct association between the simulation results of linear regression analysis and application of the ANN method in establishing the correspondence between the variation of chemical analysis and sensory analysis. The best performance resulted from the application of linear regression occurred in most cases when the best performance was also achieved by the ANN method.

## ACKNOWLEDGMENTS

To CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*), for granting of scholarships. To the Department of Computation, Federal University of Ouro Preto, for releasing a full-time faculty member for doctorate studies, which allowed this work to be developed. To the Department of Agricultural Engineering at the Federal University of Viçosa, for technical support.

## REFERENCES

- ANDERSON, K. A.; SMITH, B. W. Chemical profiling to differentiate geographic growing origins of coffee. *Journal of Agricultural and Food Chemistry*, Easton, v.50, p.2068-2075, 2002.
- BORÉM, F. M.; RIBEIRO, D. M.; PEREIRA, R. G. F. A.; ROSA, S. D. V. F.; MORAIS, A. R. Qualidade do café submetido a diferentes temperaturas, fluxos de ar e períodos de pré-secagem. *Coffee Science*, Lavras, v.1, n.1, p.55-63, Lavras, abr./jun. 2006.
- CHARUMPORN, B.; YOSHIOKA, M.; FUJINAKA, T.; OMATU, S. Identify household burning smell using an electronic nose system with artificial neural networks. *Computational Intelligence in Robotics and Automation*, Kobe, p.1070-1074, 2003.
- CHERKASSKY, V.; MA, Y. Another look at statistical learning theory and regularization. *Neural Networks*, New York, v.22, p.958-969, 2009.
- CIURANA, J.; QUINTANA, G.; GARCIA-ROMEU, M. L. Estimating the cost of vertical high-speed machining centres, a comparison between multiple regression analysis and the neural networks approach. *International Journal of Production Economics*, Amsterdam, v.115, p.171-178, 2008.
- DUTTA, R.; KASHWAN, K.R.; BHUYAN, M.; HINES, E.L.; GARDNER J.W. Electronic nose based tea quality standardization. *Neural Networks*, New York, v.16, p.847-853, 2003.
- FERNANDES, M. *Uso de cloreto de benzalcônio na pré-colheita e suas iterações com a colheita e pós-colheita na qualidade do café*. 2008. 91f. Tese (Doutorado em Ciência dos Alimentos) – Universidade Federal de Lavras, Lavras, 2008.
- FIGUEIREDO, L. P. *Perfil sensorial, químico e proteômico relacionados ao processamento pós-colheita de cultivares de cafeeiro*. Lavras: UFLA, 2009. (Projeto de Pesquisa).
- HERVÁS-MARTÍNEZ, C.; MARTÍNEZ-ESTUDILLO, F. J.; CARBONERO-RUZ, M. Multilogistic regression by means of evolutionary product-unit neural networks. *Neural Networks*, New York, v.21, p.951-961, 2008.
- ISQUIERDO, E. P. *Secagem e qualidade do café cereja desmucilado submetido a diferentes períodos de repouso*. 2008. 98 f. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Federal de Lavras, Lavras, 2008.
- KERMANI, B. G.; SCHIFFMAN, S. S.; NAGLE, H. T. Performance of the Levenberg–Marquardt neural network training method in electronic nose applications. *Sensors and Actuators B Chemical*, Lausanne, v.110, p.13-22, 2005.
- KHOSHGOFTAAR, T. M.; GAO, K. Count models for software quality estimation. *IEEE Transactions on Reliability*, New York, v.56, n.2, Jun. 2007.

- NOBRE, G. W. *Processamento e qualidade de frutos verdes de café arábica*. 2009. 85f. Tese (Doutorado em Agronomia) – Universidade Federal de Lavras, Lavras, 2009.
- PARDO, M. ; SISK, B. C. ; SBERVEGLIERI, G.; LEWIS, N. S. Comparison of Fisher's linear discriminant to multilayer perceptron networks in the classification of vapors using sensor array data. *Sensors and Actuators B Chemical*, New York, v.115, p.647-655, 2006.
- RECAMALES, A. F.; HERNANZ, D.; ÁLVAREZ, C.; GONZÁLEZ-MIRET, M. L.; HEREDIA, FRANCISCO J. Colour of Amontillado wines aged in two oak barrel types. *European Food Research and Technology*, Berlin, v.224, n.3, p.321–327, jan. 2007.
- REINATO, C. H. R.; BORÉM, F. M.; SILVA, PABLO J.; OLIVEIRA, E. C. Influência da secagem, em diferentes tipos de terreiro, sobre a qualidade do café ao longo do armazenamento. *Coffee Science*, Lavras, v.2, n.1, p.48-60, jan./jun. 2007.
- RIBEIRO, F. C. *Análise Técnica da utilização do silo bolsa no armazenamento de cafés especiais*. Lavras, MG: UFLA, 2009. (Projeto de Pesquisa).
- SAEG - *Sistema para análises estatísticas*. Versão 9.1: Viçosa: UFV/Fundação Arthur Bernardes, 2007.
- SOUSA, S. I. V.; MARTINS, F. G.; ALVIM-FERRAZ, M. C. M.; PEREIRA, M. C. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*, Oxford, v.22, p.97-103, 2007.
- SUCHACZ, B; WESOLOWSKI, M. The recognition of similarities in trace elements content in medicinal plants using MLP and RBF neural networks. *Talanta*, London, v.69, p.37-42, 2006.
- VERLINDEN, B.; DUFLOU, J. R.; COLLIN, P.; CATTRYSSE, D. Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study. *International Journal of Production Economics*, Amsterdam, v.111, p.484-492, 2008.