

## ARABICA COFFEE CLASSIFICATION USING NEAR INFRARED SPECTROSCOPY AND TWO-STAGE MODELS

Izabele Marquetti<sup>2</sup>; Jade Varaschim Link<sup>3</sup>; André Luis Guimarães Lemes<sup>4</sup>; Maria Brígida dos Santos Scholz<sup>5</sup>; Patrícia Valderrama<sup>6</sup>; Evandro Bona<sup>7</sup>

<sup>1</sup> Trabalho financiado pelo CNPq, CAPES e Fundação Araucária

<sup>2</sup> Mestranda, MS, Programa de Pós-Graduação em Tecnologia de Alimentos, UTFPR-CM, ibmarquetti@gmail.com

<sup>3</sup> Doutorando, MS, Programa de Pós-Graduação em Engenharia de Alimentos, UFSC, jadevl@hotmail.com

<sup>4</sup> Iniciação Científica, BS, Curso de Engenharia de Alimentos, UTFPR-CM, alglemes@hotmail.com

<sup>5</sup> Pesquisadora, DR, Instituto Agrônômico do Paraná, IAPAR, mbsholz@iapar.br

<sup>6</sup> Pesquisadora, DR, Programa de Pós-Graduação em Tecnologia de Alimentos, UTFPR-CM, patriciav@utfpr.edu.br

<sup>7</sup> Pesquisador, DR, Programa de Pós-Graduação em Tecnologia de Alimentos, UTFPR-CM, ebona@utfpr.edu.br

**ABSTRACT:** Coffee quality depends on the environmental conditions of the growing area. Factors such as climate, soil type and altitude, associated with agricultural practices, directly influence the chemical composition of the coffee beans. This study developed two-stage models to determine the geographic and genotypic origin of the grain. For the first stage, the partial least squares with discriminant analysis (PLS-DA) and principal component analysis (PCA) models were tested. Then, two artificial neural network (ANN) non-linear models, i.e. multilayer perceptron (MLP) and the radial-basis function (RBF), were evaluated as the second stage. Samples from four genotypes, cultivated in four different cities within Parana State in Brazil, were analyzed using near infrared spectroscopy (NIRS) in the 1100 to 2498 nm range. Three preprocessing techniques were tested on the spectra, i.e. multiplicative scatter correction (MSC); the Savitzky-Golay second-derivative and both combined. The best models were obtained with the spectra treated using MSC plus the second-derivative, with PLS-DA as first stage followed by the RBF network. For geographic and genotypic classification the sensitivity and specificity values of 100% were obtained for the training and test sets. The NIRS spectra presented better class separation when compared with the FTIR spectra used in a previous work. These results demonstrate that NIRS spectra, allied with the right pattern recognition techniques, can be used as a quick and efficient technique to distinguish green coffee samples both geographically and genotypically.

**KEYWORDS:** multiplicative scatter correction, principal component analysis, PLS-DA, artificial neural network, sequential simplex, Bayesian approach.

## CLASSIFICAÇÃO DE CAFÉ ARÁBICA USANDO ESPECTROSCOPIA DE INFRAVERMELHO E MODELOS DE DOIS ESTÁGIOS

**RESUMO:** A qualidade do café depende das condições ambientais da região de cultivo. Fatores como clima, tipo de solo e altitude, associados às práticas agrícolas influem diretamente na composição do grão de café. Neste estudo foram construídos modelos de dois estágios para a origem geográfica e genotípica do grão. No primeiro estágio, foram testados os métodos de mínimos quadrados parciais com análise discriminante (PLS-DA) e de análise das componentes principais (PCA). No estágio, que é não linear, foram avaliadas duas redes neurais artificiais (ANN), o perceptron de múltiplas camadas (MLP) e a rede de funções de base radial (RBF). Amostras de quatro genótipos cultivadas em quatro cidades diferentes do estado do Paraná (Brasil) foram avaliadas usando espectroscopia de infravermelho próximo (NIRS) na faixa de 1100 até 2500 nm. Para o pré-processamento dos espectros foram testadas três técnicas, correção do espalhamento multiplicativo (MSC), 2ª derivada e a combinação de ambas as técnicas. Os melhores modelos foram obtidos com o espectro tratado pelo MSC + 2ª derivada, com o PLS-DA no primeiro estágio e a rede RBF no segundo. Tanto para a classificação geográfica quanto genotípica foram obtidas sensibilidades e especificidades de 100% para o conjunto de treinamento e de teste. O espectro NIRS aliado com uma técnica de reconhecimento de padrões adequada pode ser usado como uma ferramenta eficiente e rápida para segmentação de amostras de café verde por origem geográfica e genotípica.

**PALAVRAS-CHAVE:** correção do espalhamento multiplicativo, análise de componentes principais, PLS-DA, redes neurais artificiais, simplex sequencial, abordagem Bayesiana.

## INTRODUCTION

The coffee beans from the *Coffea arabica* species are known for their intense aroma and sweetness and less bitter taste, producing a better quality beverage when compared to *robusta* coffee; *arabica* coffee is more appreciated by consumers and has a higher aggregate price [1,2]. In addition to adequate coffee genotype selection, the environmental conditions of the growing location, such as the climate, soil type and altitude, as well as the agronomic practices used directly

influence beverage quality because they are responsible for the final chemical composition of the bean [3,4]. The highest quality coffees are associated with higher sucrose, lipid, amino acid, and trigonelline contents, and lower chlorogenic acid and caffeine contents, which *are* responsible for contributing to the bitter taste of coffee [2,5]. The chemical composition and the quality of the growing location are ways to aggregate value to the coffee in highly competitive markets [6,7]. To guarantee to the consumer the genotype and geographic origin of coffee, efficient and fast analytic methods are frequently required. A commonly used technique in coffee analysis is near infrared spectroscopy (NIRS), which requires minimal sample preparation and allows for simultaneous analysis [8]. Correctly classifying *arabica* coffee samples by geographic and genotype origin is a complex task due to the high numbers of independent variables. So, it is necessary to use pattern recognition and multivariate classification tools. Principal component analysis (PCA) is commonly used for this purpose, as it is able to reduce data dimensionality by grouping highly related information. However, due to the fact that data are described by linear combinations, non-linear systems are not well represented. Thus, the quality of the result can be influenced by discrepant samples [9]. Partial least squares with discriminant analysis is another linear method that distinguishes by being supervised, that is it uses the class information from previous samples in data decomposition [10]. Even so, there is the possibility that these methods may not present satisfactory results; if the significant component numbers are high, it is hard to extract useful information from data [9]. Artificial neural networks (ANN) are a set of mathematic methods commonly used for classification and pattern recognition. They are non-linear computational tools capable of modeling extremely complex functions, where the knowledge is acquired by training during a learning process [11,12]. ANN are used for mapping the input data into output data and are implemented with electronic components or by simulation in a computer [9,13,14]. They have been applied in different areas including food authentication, sensory analysis, and consumer preference mapping [12]. This study aimed to establish methods for geographic and genotypic classification of green *arabica* coffee beans. For this, NIRS spectra were analyzed in two-stage models (TSM), first using a linear method, i.e. PCA or PLS-DA, then employing a non-linear method based on ANN, i.e. multilayer perceptron and radial-basis function.

## MATERIALS AND METHODS

Four *Coffea arabica* genotypes were analyzed: IPR 99, IPR 105, IPR 106, and IA 59. These genotypes were developed by the Agronomic Institute of Parana (IAPAR). The IA 59 and IPR 99 genotypes originated from a crossing between the *Coffea arabica* varieties “Villa Sarchi 971/10” and “Híbrido de Timor 832/2”; both are resistant to all known kinds of rust [15]. The IPR 105 genotype is derived from the “Catuai” genotype, while the IPR 106 genotype originated from the “Icatu” genotype. Among these genotypes, only the IA 59 and IPR 99 genotypes are available to farmers [16]. Eighteen samples of coffee, carefully selected, cultivated in four different cities were analyzed: Cornélio Procópio-CP (23°10'51"S 50°38'48"W; altitude 658 m, mean annual temperature 21–22°C), Paranavaí-PV (23°04'22"S 52°27'55"W; altitude 470 m, mean annual temperature 22–23°C), Mandaguari-MD (23°32'52"S 51°40'15"W; altitude 650 m, mean annual temperature 20–21°C), and Londrina-LD (23°18'36"S 51°09'56"W; altitude 585 m, mean annual temperature 21–22°C), all in Parana State in Brazil [17]. One sample of each genotype was used per city, except for the IA 59 samples cultivated in PV and CP, where two samples were made available for each city. The PV and CP samples were harvested in 2008, and the LD and MD samples were harvested in 2010. Different years and harvest location is a complicating factor during classification because the NIRS spectral signature is strongly affected by environmental conditions [18]. The beans were collected and transported to IAPAR in Londrina, Paraná (Brazil). The samples were placed in wooden boxes with a mesh bottom and moved eight times per day until bean moisture of 11–12% was reached. Then, the samples were processed (removal of the husk and parchment). Finally, the beans were ground in a disk mill (0.5 mm) and stored in a freezer at -18°C for subsequent analysis [19].

The analysis of the green coffee spectra was performed using a near infrared spectroscopy NIRSystem 5000-M (Foss Tecator AB, Höganäs, Sweden) in the wavelength range of 1100–2498 nm at 2 nm intervals, with five replicates for each individual sample, totaling 90 spectra (room temperature 23°C). The software WinISI III version 1.50e (Foss NIRSystems/Tecator Infracore International, LLC, Silver Spring, MD, USA) was used to save the spectra. The absorbance was obtained as decimal logarithm of the inverse of transmittance, i.e.,  $\log(1/T)$ . For the TSM, 72 preprocessed spectra (80%) were used as training samples and 18 spectra (20%) as test samples (one repetition, chosen randomly, of each coffee studied). Three transformations were performed on the data matrix of the original spectra: multiplicative scatter correction (MSC) [20]; the Savitzky-Golay second-derivative [21] (second order polynomial with a seven point window) and both combined. MSC uses linear regression of the spectra variables versus the average spectra to simultaneously correct the multiplicative and additive scatter effects [20]. The second-derivative removes occasional problems due to slope changes between samples [21].

After spectral data preprocessing, PCA [22, 23] and PLS-DA [10, 24–26] were used as the first stage of the classification model to evaluate the best preliminary model to provide the input for the ANN (second stage). In this work, the quantity of PC utilized as input for the second stage, just like the quantity of LV, was optimized by the sequential simplex method. Before the input vectors (PC scores or LV scores) were fed into the ANN, they were normalized to avoid an overflow error or prevent the saturation of the activation functions in the artificial neurons [9]. For this purpose, three normalization methods were tested: minimum and maximum (minimax), which transform the value into a range between -1 and 1; autoscale, in which the mean centered data are divided by the variance, so after the

transformation each variable will have a null average and variance equal to 1, limiting the values to -3 to +3; or unit vector, in which the data vector is divided by its Euclidean norm, such that after transformation the norm of each variable is 1 [27-30].

For the second stage of the classification model, two artificial neural networks were tested, i.e. multilayer perceptrons (MLP) [31, 32] and radial basis functions (RBF) [33-35], which are non-linear models. Normalized PCA scores or PLS-DA scores were used as network inputs. In both networks, MLP and RBF, some parameters were optimized using the sequential simplex method to maximize the correct classification percentage and reduce the mean square error for the validation sample set with the lowest model possible [28, 36-38]. The classification model reliability was evaluated using a threshold value that separates the classes; i.e. the sample belongs to class  $C_k$  or the sample do not belongs to  $C_k$ . The threshold value for the class separation is based on Bayes' Theorem [33]. The evaluation of classification performance was performed using the confusion table to calculate the sensitivity and specificity [40].

All the spectra pre-processing, PCA, PLS-DA, ANN and the sequential simplex optimization method were performed in MATLAB R2008b software (The MathWorks Inc., Natick, USA).

## RESULTS AND DISCUSSION

After the sequential simplex optimization of the selected parameters for the MLP network, one thousand TSM were created for each spectral treatment and for each first stage method, i.e. PCA and PLS-DA. The best TSM was chosen according to the lowest mean square error and the highest correct classification percentage for test samples.

Analyzing the results, it can be seen that all TSM with MLP networks correctly classified 100% of the training and test samples by geographic origin. However, the TSM that obtained the lowest mean square error with fewer free parameters was TSM04, which used MSC and PLS-DA. MLP network architecture parameters are dependent on the problem, so the application of sequential simplex optimization allows for an automated choice of the best value set for each case [39]. For all models using the MLP as second stage, except TSM05, the number of input variables (scores from PC or LV) conducted to accumulated variances greater than 93% for X matrix. The number of free parameters, network weights, was lower when PLS-DA was used as the first stage, this result was expected because PLS-DA model incorporate information about Y too; so in second stage it was necessary a smaller ANN to model the remaining information from first stage. The regularization parameter values were high, even for a small number of hidden neurons, showing that was necessary to smooth the ANN model. The number of synaptic weights in all networks was lower than the size of training set (72 samples), indicating that there were enough degrees of freedom to avoid the overfitting.

For genotype classification, three TSM correctly classified 100% of the training and test samples, i.e. TSM08, TSM09 and TSM11. However, all the others presented a high correct classification percentage, even for genotypes with a narrow genetic base, i.e., a similar genetic composition [41]. TSM09, in general, presented satisfactory performance due to the high classification percentage with fewer free parameters when compared to the others. As in geographic classification, it was observed high values for regularization parameter and smaller ANN when using PLS-DA as the first stage. For genotypes, all the best networks were those with the hyperbolic tangent sigmoid as activation function in the hidden layer. However, for the geographic classification, the logistic functions were best suited. Regarding normalization, for both classifications, the autoscale method presented the best results. The networks trained for geographic classification presented lower numbers of synaptic weights and input variables when compared to the genotype classification; this occurred because the samples were from the same coffee species, which was a more difficult problem to model. This behavior has been observed previously [31].

After the sequential simplex optimization of the selected parameters for the RBF networks, one thousand TSM were created for each spectra treatment presented and for each first stage, i.e. PCA and PLS-DA. The best TSM for each type was chosen according to the lowest mean square error and the highest correct classification percentage for the test samples.

For the geographic classification, four TSM correctly classified 100% of the training and test samples, i.e. TSM15, TSM16, TSM17 and TSM18. However, the TSM that obtained the lowest mean square error for the training and test samples with fewer free parameters was TSM17, which was performed using MSC plus 2<sup>nd</sup> derivative and PLS-DA. Again, simplex optimization proved to be effective to chosen network's architecture parameters. A reduction was also observed, in some cases, in the input and radial-basis quantity when compared PLS-DA with PCA. In contrast to the MLP network, for RBF, the regularization parameters were small, because in this type of network, basis width already controls the mapping smoothness. All the best networks for this classification were Gaussian.

In the same way, for genotype segmentation, two TSM correctly classified 100% of the training and test samples, i.e. TSM23 and TSM24. However, TSM23, trained with the scores of PLS-DA and spectra treated with MSC and the second-derivative simultaneously, has a reduced number of free parameters when compared to the others. Besides, it presented the lowest mean square error for the training and test samples. The best networks trained with the scores of PCA used, for genotypic classification, multiquadrics radial basis, while the ones with PLS-DA used Gaussian functions. Among the distance function for the K-means algorithm, the cosine function presented superiority toward the other tested functions.

The TSM developed with RBF and MLP presented similar performance to the geographic and genotypic classification, for both 100% classification was achieved for the test samples. According to [9], for all MLP, there is an RBF with

equivalent performance. However, the TSM using the RBF networks trained with PLS-DA and, MSC plus the second-derivative together, can be considered as the best option for geographic classification (TSM17) and for genotypic classification (TSM23). Because they have less free parameters when compared to the others, indicating a simpler structure and faster training for second stage. PCA is commonly used as first stage to reduce data dimensionality. However, the better results obtained with PLS-DA show the advantages of the application of a linear supervised method in first stage. Currently, studies that use PLS-DA and ANN combined are scarce, in [42] was compared to PCA and PLS-DA as first stage. It was obtained 100% classification in TSM that used PLS-DA and ANN approach, therefore, PLS-DA utilization must be considered as an reliable option for two-stage models. In addition, analyzing the *a posteriori* probabilities curves (Figure 1), only TSM17 did not present any samples in the rejection region for geographic classification, for other models the *a posteriori* probabilities curves were omitted for lack of space. This confirms that for this model the classes were far from the threshold, represented by the dashed line in Figure 1, which guarantees greater reliability in the model.

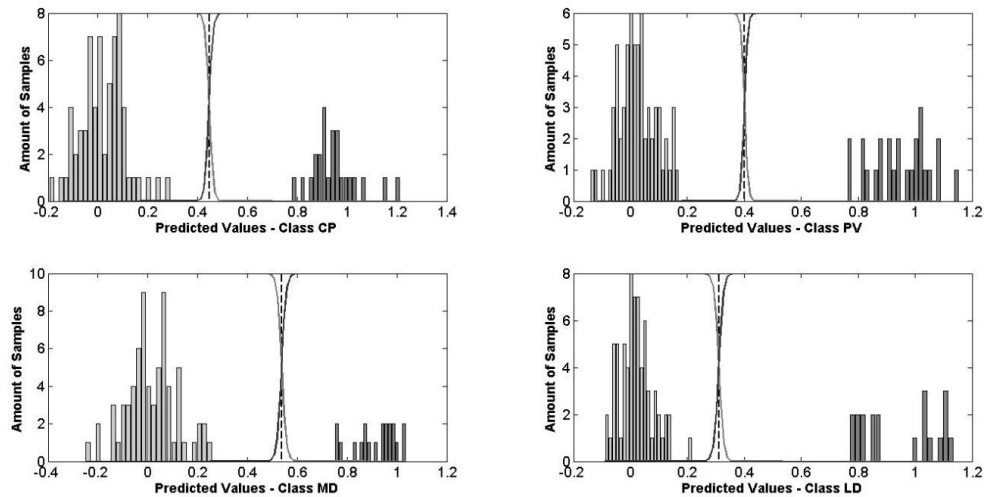


Figure 1. *A posteriori* probability curve for the best two-stage model for geographic classification (TSM17). The vertical dashed line shows de Bayesian threshold, lighter bars represents the samples that do not belongs to the class and the darker ones the samples that belongs to the class.

Figure 2 presents the outputs generated by class for TSM23, the Bayesian threshold values were 0.3821, 0.5251, 0.4603 and 0.5104 for classes IPR 105, IPR 106, IPR 99 and IA 59, respectively, and are represented by the horizontal dashed line. Like in geographic classification, no sample was below the threshold line, indicating that all the classes obtained sensitivity equal to 1, and no sample from the other classes was above the threshold line, showing that all classes obtained specificity equal to 1. Comparing with the PLS-DA model alone, one test sample from class IPR 106 being wrongly classified and was observed specificity below 1 for classes IPR 105, IPR 106 and IPR 99 during training and for class IPR 105 in the test [43]. In a study performed using only PLS-DA for classification, the models correctly classified 94.4% of samples by genotype and growing region [43]. Using TSM, it was possible to increase the classification percentage and improve the parameters that evaluated the model reliability, i.e. sensitivity and specificity. For both the MLP and RBF networks, trained with PLS-DA, using spectra treated with MSC and the second-derivative together, the sensitivity and specificity values were 1 for all classes.

For the same green *arabica* coffee beans, MLP and RBF networks combined with PCA were also applied in a previous work to classify the samples using Fourier transform infrared spectroscopy (FTIR) [31,39]. NIR spectra lead to the best results with a lower mean square error and were able to correctly classify 100% of test samples by genotype (using MLP in second stage) and geographic origin. The results indicate that NIRS is an adequate technique to be employed in the geographic and genotypic identification of green coffee beans, and can be used as an alternative analysis method in the industry. Due to the fact that minimal sample preparation is necessary, experimental errors that can occur during the analysis are avoided, so the results are more reliable. Additionally, it has the advantage of being a fast technique and allows reuse of samples after analysis.

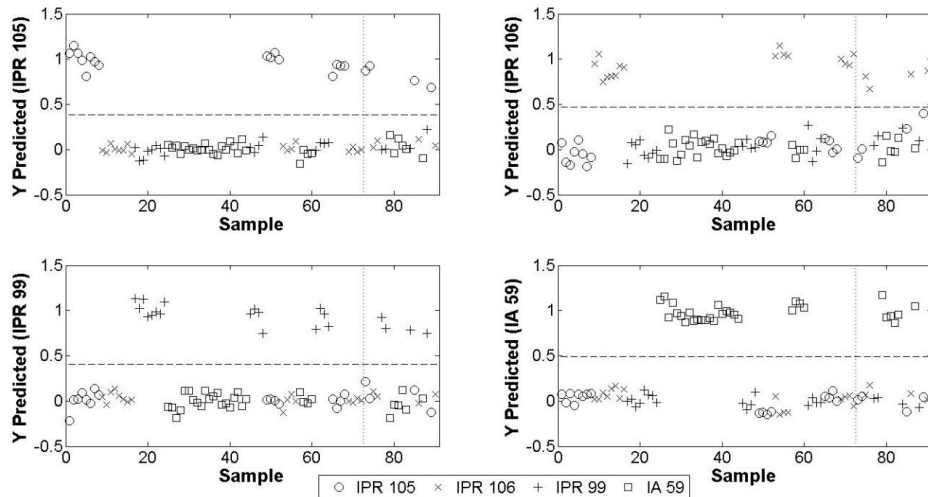


Figure 2. The best two-stage model output for genotypic classification (TSM23), spectra treated with MSC plus second-derivative, PLS-DA as first stage and RBF in second stage. The vertical dotted line separates the training samples from the test ones and the horizontal dashed line shows de Bayesian threshold.

## CONCLUSION

The optimized TSM were able to correctly classify all *arabica* coffee samples geographically and genotypically. However, radial-basis functions were distinguished over the others, mainly those trained with the PLS-DA latent variables, which had satisfactory performance, with a low mean square error value and reduced numbers of free parameters in the networks. The greater amount of information present in the PLS-DA model makes the network need fewer synaptic weights to perform the classification, thus providing evidence for the benefits of using a supervised method as the network input. Sequential simplex proved to be an efficient methodology to determine the ANN parameters, thereby maximizing the performance and minimizing the network size. NIR spectra, processed with PCA, presented better results to classify with both networks, i.e. MLP and RBF, when compared with FTIR spectra. MLP networks trained with FTIR did not correctly classify the samples by genotype, while those that used the NIR spectra achieved 100% classification. Thus, near infrared spectroscopy can be used as an alternative technique in coffee analysis, and can be performed in an efficient and fast manner with minimal sample preparation.

## REFERENCES

- 1 Lashermes P, Anthony F (2007) In Kole C (ed) Genome mapping and molecular breeding in plants. Springer Berlin Heidelberg, Berlin.
- 2 Ky C-L, Louarn J, Dussert S, Guyot B, Hamon S, Noiro M (2001) Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. Food Chemistry 75: 223-230.
- 3 Farah A. (2009). Coffee as speciality and functional beverage. *Functional and speciality beverage technology*, 370-395.
- 4 Farah A, Donangelo CM (2006) Phenolic compounds in coffee. *Braz J Plant Physiol* 18: 23-36.
- 5 Stalmach A, Mullen W, Nagai C, Crozier A (2006) On-line HPLC analysis of the antioxidant activity of phenolic compounds in brewed, paper-filtered coffee. *Braz J Plant Physiol* 18: 253-262.
- 6 Haiduc A, Gancel C, Leloup V (2010) NIR-based Determination of differences in green coffee chemical composition due to geographical. [http://www.heliospir.net/medias/upload/8eme\\_heliospir\\_Haiduc.pdf](http://www.heliospir.net/medias/upload/8eme_heliospir_Haiduc.pdf). Accessed 16 Feb 2014.
- 7 Teuber R (2010) Geographical indications of origin as a tool of product differentiation: the case of coffee. *Journal of International Food & Agribusiness Marketing* 22: 277-298.
- 8 Esteban-Díez I, González-Sáiz JM, Sáenz-González C, Pizarro C (2007) Coffee cultivar differentiation based on near infrared spectroscopy. *Talanta* 71: 221-229.
- 9 Haykin S (2001) *Redes Neurais: Princípios e Práticas*, 2nd ed. Bookman, Porto Alegre.
- 10 Barker M, Rayens W (2003) Partial least squares for discrimination. *Journal of Chemometrics* 17: 166-173.
- 11 Graupe D (2007) *Principles of Artificial Neural Networks: Advanced Series on Circuits and Systems*, 2nd ed. World Scientific, Chicago.
- 12 Marini F (2009) Artificial Neural Networks in foodstuff analyses: Trends and perspectives - A review. *Analytica Chimica Acta* 635: 121-131.
- 13 Priddy KL, Keller PE (2005) *Artificial Neural Network: an introduction*. SPIE, Washington.
- 14 Marini F, Bucci R, Magri AL, Magri AD (2008) Artificial neural networks in chemometrics: History, examples and perspectives. *Microchemical Journal* 88: 178-185.

- 15 Sera T, Shigueoka LH, Sera GH, Azevedo JA, Carvalho FG, Andreazi E (2011) Nova seleção da cultivar de café Iapar 59 com grãos mais graúdos. In *Simpósio de Pesquisa dos Cafés do Brasil*, Araxá, BR, (2011).
- 16 Sera GH, Sera T, Fonseca ICdB, Ito DS (2010) Resistance to leaf rust in coffee cultivars. *Coffee Science* 5: 59-66.
- 17 Caviglione JH, Wrege MS, Gonçalves SL, Faria RT, Androcioli AF, Sera T, Chaves JCD, Kogushi MS (2001) Climatic risk zoning for coffee (*Coffea arabica* L.) in Paraná state, Brazil. *Revista Brasileira de Agrometeorologia* 9: 486-494.
- 18 Posada H, Ferrand M, Davrieux F, Lashermes P, Bertrand B (2009) Stability across environments of the coffee variety near infrared spectral signature. *Heredity* 102: 113-119.
- 19 Brasil, Regulamento técnico de identidade e de qualidade para a classificação do café beneficiado e de grão verde, Instrução Normativa no 8 de 11 jun. 2003, Ministério da Agricultura Pecuária e Abastecimento. [http://www.claspar.pr.gov.br/arquivos/File/pdf/cafefenef008\\_03.pdf](http://www.claspar.pr.gov.br/arquivos/File/pdf/cafefenef008_03.pdf) Accessed 10 Sep 2011.
- 20 Isaksson T, Næs T (1988) The Effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. *Applied Spectroscopy* 42: 1273-1287.
- 21 Savitzky A, Golay MJE (1964) Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36: 1627-1639.
- 22 Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 2: 37-52.
- 23 Matos GD, Pereira-Filho ER, Poppi RJ, Arruda MAZ (2003) Análise exploratória em química analítica com emprego de quimiometria: PCA e PCA de imagens. *Revista Analytica* 3: 38-50.
- 24 Valderrama P (2005) Avaliação de figuras de mérito em calibração multivariada na determinação de parâmetros de controle de qualidade em indústria alcooleira por espectroscopia no infravermelho próximo. Unicamp, Campinas.
- 25 Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. *Analytica Chimica Acta* 185: 1-17.
- 26 Bassbasi M, De Luca M, Ioele G, Oussama A, Ragno G (2014) Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data. *Journal of Food Composition and Analysis* 33: 210-215.
- 27 Pérez-Magariño S, Ortega-Heras M, González-San José ML, Boger Z (2004) Comparative study of artificial neural network and multivariate methods to classify Spanish DO rose wines. *Talanta* 62: 983-990.
- 28 Bona E, Silva RSSF, Borsato D, Bassoli DG (2011) Optimized Neural Network for Instant Coffee Classification through an Electronic Nose. *International Journal of Food Engineering* 7.
- 29 Borsato D, Pina MVR, Spacino KR, Scholz MBS, Androcioli AF (2011) Application of artificial neural networks in the geographical identification of coffee samples. *Eur Food Res Technol* 233: 533-543.
- 30 Galão OF, Borsato D, Pinto JP, Visentainer JV, Carrão-Panizzi MC (2011) Artificial neural networks in the classification and identification of soybean cultivars by planting region. *J Braz Chem Soc* 22: 142-147.
- 31 Link JV, Lemes ALG, Sato HP, Scholz MBS, Bona E. (2012) Optimized multilayer perceptron for the geographical and genotypic classification of four genotypes of arabica coffee. *Revista Brasileira de Pesquisa em Alimentos (REBRAPA)* 3: 72-81.
- 32 Debska B, Guzowska-Swider B (2011) Application of artificial neural network in food classification. *Analytica Chimica Acta* 705: 283-291.
- 33 Bishop CM (2006) *Pattern Recognition and Machine Learning*. Springer, New York.
- 34 Buhmann MD, Ablowitz MJ (2003) *Radial basis functions: Theory and implementations*. Cambridge University Press, United Kingdom.
- 35 Tudu B, Jana A, Metla A, Ghosh D, Bhattacharyya N, Bandyopadhyay R (2009) Electronic nose for black tea quality evaluation by an incremental RBF network. *Sensors and Actuators B: Chemical* 138: 90-96.
- 36 Beveridge GSG, Schechter RS (1987) *Optimization theory and practice*. Mac Graw-Hill & Sons, Tokyo.
- 37 Splendley W, Himsforth FR, Hext GR (1962) Sequential application of simplex designs in optimization and evolutionary operation. *Technometrics* 4: 441-461.
- 38 Gao F, Han L (2012) Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Comput Optim Appl* 51: 259-277.
- 39 Link JV, Lemes ALG, Marquetti I, Scholz MBS, Bona E (2014) Geographical and genotypic classification of arabica coffee using Fourier transform infrared spectroscopy and radial-basis function networks. *Chemometrics and Intelligent Laboratory Systems* 135: 150-156.
- 40 Almeida MR, Fidelis CHV, Barata LES, Poppi RJ (2013) Classification of Amazonian rose wood essential oil by Raman spectroscopy and PLS-DA with reliability estimation. *Talanta* 117: 305-311.
- 41 Sera T (2001) Coffee genetic breeding at IAPAR. *Crop Breeding and Applied Biotechnology* 1: 179-199.
- 42 Ciosek P, Brzozka Z, Wroblewski W, Martinelli E, Di Natale C, D'Amico A (2005) Direct and two-stage data analysis procedures based on PCA, PLS-DA and ANN for ISE-based electronic tongue - Effect of supervised feature extraction. *Talanta* 67: 590-596.
- 43 Marquetti, Izabele. Coffee arabica genotype classification using near infrared spectroscopy. 2014. 79 f. Dissertação – Programa de Pós-Graduação em Tecnologia de Alimentos, Universidade Federal Tecnológica do Paraná. Campo Mourão, 2014.