# Assessment of macadamia nut quality defects by means of near infrared spectroscopy (NIRS) and nuclear magnetic resonance (NMR)

Lívia Cirino de Carvalho[a], Fabíola Manhas Verbi Pereira[b], Camilo de Lelis Medeiros de Morais[c], Kássio Michell Gomes de Lima[d], Gustavo Henrique de Almeida Teixeira[e,*]

[a]Universidade Estadual Paulista (UNESP), Faculdade de Ciências Farmacêuticas (FCFAR), Departamento de Alimentos e Nutrição, Campus de Araraquara, Rodovia Araraquara-Jaú, km 1 – CP 502, CEP: 14.800-903, Araraquara, São Paulo, Brazil.

[b]Universidade Estadual Paulista (UNESP), Instituto de Química (IQ), Campus de Araraquara. Rua Prof. Francisco Degni, 55, CEP: 14800-060, Araraquara – SP, Brazil.

[c]University of Central Lancashire, School of Pharmacy and Biomedical Sciences, Preston, Lancashire, PR1 2HE, United Kingdom.

[d]Universidade Federal do Rio Grande do Norte (UFRN), Instituto de Química, Química Biológica e Quimiometria, Avenida Senador Salgado Filho, n° 3000, Bairro de Lagoa Nova, CEP: 59.078-970, Natal, Rio Grande do Norte, Brazil.

[e]Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agrárias e Veterinárias (FCAV), Campus de Jaboticabal. Via de Acesso Prof. Paulo Donato Castellane s/n, CEP: 14.884-900, Jaboticabal, São Paulo, Brazil.

*****Corresponding author: gustavo@fcav.unesp.br

**Abstract**

Macadamia kernels are visually sorted based on the presence of quality defects by specialized labors. However, this process is not as accurate as non-destructive methods such as near infrared spectroscopy (NIRS) and nuclear magnetic resonance (NMR). Thus, NIRS and NMR in combination with chemometrics have become established non-destructive method for rapid assessment of quality parameters in the food and agricultural sectors. Therefore, the quality of macadamia nuts was assessed by NIRS and NMR using chemometric tools such as PCA-LDA and GA-LDA to evaluate kernel defects. Macadamia kernels were classified as: 1=good, marketable kernels without defects; 2=kernels with discoloration; 3=immature kernels; 4=kernels affected by mold; and 5=kernels with insect damage. Using NIRS, the GA-LDA resulted in an accuracy and specificity of 97.8 % and 100 %, respectively, to classify good kernels. On the other hand, PCA-LDA technique resulting in an accuracy higher than 68 % and specificity of 97.2 % to classify immature kernels. For NMR, PCA-LDA resulted in an accuracy higher than 83% and GA-LDA resulted in an accuracy of 100%, both to classify kernels with insect damage. NIRS and NMR spectroscopy can be successfully used to classify unshelled macadamia nuts based on the defects. However, NIRS out-performed NMR based on the higher accuracy results.

**Keywords:** *Macadamia integrifolia* Maiden & Betche, TD–NMR, PCA-LDA, GA-LDA, chemometrics.

## 1. Introduction

Macadamia (*Macadamia integrifolia* Maiden & Betche) nut growers are keen to continuously improve nut quality as this is the main characteristic required by the final consumers. Nogueira (2008) mentioned that the quality of macadamia fruit is associated with favorable climatic conditions, planning and orchard management, varieties, pest control, plant nutrition, harvest and post-harvest practices. All these factors are decisive for macadamia development and nut quality.

According to O'Hare et al. (2004), the main defects that can be observed in macadamia nuts are immaturity; small nuts; cracks in the shell that allow the occurrence of biological and chemical contamination; lipids oxidation, which result in unpleasant odor and taste; bruises, and high moisture. Guthrie et al. (2004) reported other defects that may be considered, as such: fungal growth, decomposition, germination, and discoloration of macadamia nuts. Therefore, sound and/or good macadamia nuts must have light cream color, no signs of mold, decay, insect scars, blemishes, hollow centers, dark centers, shriveling, off-odors, adhering shells, and loose of extraneous material (Wall, 2013).

Macadamia industry has developed various parameters of quality standards. The Southern African Macadamia Growers' Association (SAMAC) classifies macadamia nuts into three classes: first grade, commercial grade, and local market. These classes are established based on kernel color, flavor and odor, kernel dust, insect infestation, foreign material. A limit of 1.5 % is used reject the nuts based on the presence of insect damage, discoloration, and immaturity (SAMAC, 2018). On the other hand, the United Nations Economic Commission for Europe (UNECE) has a higher tolerance (5 %) for the presence of these defects (UNECE, 2010).

The sorting process of macadamia kernels in the industry can be carried out manually (Piza, 2005) or electronically (France, 2007), but both present flaws, since manual sorting of defective kernels can decrease dramatically with the use of inadequate lighting and untrained personnel, and the electronic selection uses color to sort kernels, which may lead to improper selection, since immature kernels can only be identified based on the deformed, wrinkled, and shrunken kernel (SAMAC, 2018).

The increasing requirements of consumers, regulatory agencies, and competitors have been an impulse for the development of more accurate quality assessment techniques in the food industry. In this regard, near infrared spectroscopy (NIRS) in combination with chemometric modelling have become an established method for rapid assessment and non-destructive quality parameters in the food and agricultural sectors (Abbott, 1999; Jensen et al., 2001), since it is fast, safe, relatively inexpensive technique and provides automation of quality control processes in products of agroindustry (Pasquini, 2003).

NIRS has been used to evaluate macadamia nut quality. Guthrie et al. (2004) developed modified partial least squares regression (MPLS) models for oil content determination in intact macadamia kernels with a root mean square error of calibration (RMSEC) of 2.4 % and discriminated intact kernels with brown centers or rancidity from each other and from sound kernels using PCA. Canneddu et al. (2016) developed models for predicting peroxide value (PV) and acidity index (AI) using PLSR and classification models to discriminate defects present on shelled macadamia nuts using FT-NIR. The best model for PV prediction resulted in a coefficient of determination ($R_p^2$) of 0.72, and for AI prediction a SEP of 0.14 % and a $R_p^2$ of 0.80. Adequate classification models (93.2 %) for defects was possible using principal component analysis linear discriminant analysis (PCA-LDA). Carvalho et al. (2017) classified

intact macadamia nuts according to cultivars using PCA-LDA and genetic algorithm with linear discriminant analysis (GA-LDA), reporting an accuracy higher than 94.4 % and a value of 82.7 % for sensitivity using GA-LDA, respectively. The better performance of GA-LDA can be due to that GA algorithm selects several wavenumbers in a single band, due to collinearity problems. Carvalho et al. (2019) evaluated the oxidative stability in intact macadamia nuts during drying process and reported a SEP of 0.55 meq.kg$^{-1}$ and $R^2c$ of 0.57 for PV prediction, and SEP of 0.14 % and $R^2c$ of 0.29 for AI prediction. These results demonstrate that NIRS can be used to assess the oxidative stability of intact macadamia nuts.

Nuclear magnetic resonance (NMR) has also been stated as an alternative method among non-destructive techniques to evaluate fruit quality (Abbott, 1999). TD-NMR has wide applications for qualitative and quantitative in food analysis (Conalgo, 1996). In this regard, Pedersen et al. (2000) combined low-field nuclear magnetic resonance (LF-NMR) and PCA to classify rape and mustard seeds according on the type of seed, obtaining two distinct groups and 100 % of explained variance. This technique was also applied to evaluate the efficacy of hydrophobic coatings as a barrier to the oxidation of macadamia nuts (Colzato et al., 2009).

Although some results can be found regarding the use of NIRS to assess macadamia quality defects (Canneddu et al., 2016), this study was performed evaluating the macadamia in nut not the kernel (unshelled), and no reports were found on using NMR to evaluate macadamia kernel defects. Therefore, the objective of this study was to develop NIRS and NMR calibration models to evaluate macadamia kernels based on the most common defects aiming to improve the quality control process in the macadamia industry.

## 2. Material and Methods

### 2.1. Plant material

Macadamia (*Macadamia integrifolia* Maiden & Betche) kernels were obtained in a commercial orchard located in Dois Córregos, São Paulo, Brazil (22º 37' S, latitude, 48º 38' W, longitude, 753 m altitude) in 2017 harvest season. Nuts were harvested three times during the season (April, June, and August) and kernels were visually sorted by the industry personnel based on their quality attributes, as such: 1=good, marketable kernels without defects; 2=kernels with discoloration; 3=immature kernels; 4=kernels affected by mold; and 5=kernels with insect damage. (Figure 1). These quality attributes represented the five studied classes (model). It is important to state that the nuts were dried by the processing industry and used in the analyses without any previous treatment.

### 2.2. NIR spectra acquisition

On the surface of each macadamia kernel two Fourier Transformed (FT) NIR reflectance spectra (11,544 – 3,952 $cm^{-1}$, nm, resolution of 16 $cm^{-1}$, and 64 scans) were collected using a Bruker NIR spectrometer (Tango, Ettlingen, Germany) after temperature stabilization at ~25°C. The two replica spectra measured per nut were averaged, so the model is made on a sample basis. Samples were collected in three different harvests, where 20 nuts were sorted and used for spectra acquisition for each defect class. This resulted in a total of 300 measured samples (20 nuts x 5 classes x 3 harvests).

### 2.3. Time domain (TD) NMR measurements

TD-NMR measurements of macadamia kernels (n=100) were carried out at 22 °C in a 0.27 T (11.3 MHz for $^{1}H$) benchtop SLK200 Spinlock instrument (Spinlock Magnetic Resonance Solutions, Cordoba, Argentina). The measurements were

performed using the standard CPMG sequence to obtain the exponential decay signal that is governed by the transverse relaxation time (T2). The sequence used $\pi/2$ and $\pi$ of 11.6 and 19.6 μs, respectively, an echo time of 600 μs, 4 scans and 1500 echoes. Samples harvested in June 2017 were used and for each defect class 20 nuts were sorted and used for spectra acquisition, totaling 100 spectra. The mass of the samples ranged from 14 to 24 g depending on the sample density. The samples were the same used to collect the NIRS spectra, but the spectra were collected on different days.

**2.4. Chemometrics**

Data analysis of NIR and TD-NMR were performed within MATLAB R2014b environment (MathWorks Inc., USA) using PLS Toolbox version 7.9.3 (Eigenvector Research Inc., USA) and lab-made routines. Three different pre-processing methods were applied to test the averaged sample spectrum (average of 10 spectra per sample): (1) only mean-centering; (2) standard normal variate (SNV) followed by mean-centering; (3) Savitzky-Golay second derivative (window of 5 points, $2^{nd}$ order polynomial function) followed by mean-centering. The data was split into training (70 %, 210 samples), validation (15 %, 45 samples) and test (15 %, 45 samples) sets using the Kennard-Stone sample selection algorithm (Kennard and Stone, 2012). The training and validation sets were used for model construction and internal optimization, respectively; while the test set was used to evaluate the final predictive performance of the classification models built towards external samples.

Multivariate classification was performed by means of principal component analysis linear discriminant analysis (PCA-LDA) and genetic algorithm linear discriminant analysis (GA-LDA). PCA-LDA performs a feature extraction using principal component analysis (PCA) followed by a linear discriminant classifier (LDA) (Morais and Lima, 2018) For this, PCA is applied to the pre-processed data reducing

the original number of variables (i.e., wavelengths) to a few number of principal components (PCs) accounting for the majority of the original data variance. Each PC is composed by scores and loadings, where the first represents the variance between the samples and the latter the variance on wavelength direction (Bro and Smilde, 2014). LDA is applied to the PCA scores in a non-Bayesian form as follows (Dixon and Brereton, 2009; Wu et al, 1996).

$$L(\mathbf{x}_i) = (\mathbf{x}_i - \bar{\mathbf{x}}_k)^{\mathrm{T}} \mathbf{C}_{\mathbf{pooled}}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_k) \quad (1)$$

where $L(\mathbf{x}_i)$ represents the LDA classification scores for sample *i*; $\mathbf{x}_i$ is the input vector (i.e., the PCA scores) for sample *i*; $\bar{\mathbf{x}}_k$ is the average vector of class *k*; $\mathbf{C}_{\mathbf{pooled}}$ is pooled covariance matrix; and T represents the matrix transpose operation.

GA-LDA is feature selection technique followed by an LDA classifier. Initially, a genetic algorithm (GA) is applied to reduce to the spectral data into a few number of variables based on an evolutionary process (Bro and Smilde, 2014); then LDA is applied to these variables according to Eq. 1. These variables are in the same scale of the original spectral data and are selected according to the lowest risk of miss classification $G$. $G$ is calculated in the validation set as (Carvalho et al. 2017).

$$G = \frac{1}{N_{\mathrm{v}}} \sum_{n=1}^{N_{\mathrm{v}}} g_n \quad (2)$$

where $N_{\mathrm{v}}$ is the number of validation samples and $g_n$ is defined as:

$$g_n = \frac{r^2(x_n, m_{I(n)})}{\min_{I(m) \neq I(n)} r^2(x_n, m_{I(m)})} \quad (3)$$

in which the numerator is the squared Mahalanobis distance between sample $x_n$ (of class index $I(n)$) and the mean $m_{I(n)}$ of its true class; and the denominator represents the squared Mahalanobis distance between sample $x_n$ and the mean $m_{I(m)}$ of the closest wrong class. GA was performed through 100 generations, having 200 chromosomes each. Cross-over and mutation probabilities were set at 60% and 1%, respectively. The algorithm was repeated three times and the best result was chosen.

### 2.5. Figures of merit

The classification performance of each algorithm was evaluated according to the quality parameters of accuracy (total number of samples correctly classified considering true and false negatives), sensitivity (proportion of positives correctly identified) and specificity (proportion of negatives correctly identified). These parameters are calculated as follows (Morais and Lima, 2017):

$$\text{Accuracy}\ (\%) = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \times 100 \quad (4)$$

$$\text{Sensitivity}\ (\%) = \frac{\text{TP}}{\text{TP}+\text{FN}} \times 100 \quad (5)$$

$$\text{Specificity}\ (\%) = \frac{\text{TN}}{\text{TN}+\text{FP}} \times 100 \quad (6)$$

where TP stands for true positives; TN for true negatives; FP for false positives; and FN for false negatives.

## 3. Results and Discussion

209 **3.1. NIR spectra**

210 The raw FT-NIR spectra obtained from all macadamia kernels and the average
211 spectra from each quality attribute class can be seen in Figure 2. It was not possible to
212 observe spectral differences between the quality attributes when all macadamia kernels
213 were assessed (Figure 2A). On the other hand, the mean spectra were quite different for
214 each defect category (Figure 2B), especially at the wavelength 1,900 nm to 2,500 nm.

215 The FT-NIR spectra presented absorption bands at 1,200 nm, which are related
216 to CH stretch second overtone (Cozzolino et al., 2005), while those at 1,700 – 1,800 nm
217 are associated to the first overtones of CH stretching vibrations of $–CH_3$, $–CH_2–$ and –
218 HC=CH (Armenta and La Guardia, 2007). Absorption bands at 1,350 – 1,600 nm and
219 1,950 nm and 2,100 nm are related to the presence of glucose, sucrose, and fructose
220 (Lanza and Li, 1984) and immature kernels have higher sucrose and reducing sugar
221 contents than fully mature kernels (Wall, 2013). In Figure 2B can be seen that at 1,350
222 – 1,600 nm the immature kernels exhibit a higher absorption intensity, since maturity is
223 inversely related to sugar content (Ripperton et al., 1938).

224 The wavelength region situated at 2,200 – 2,500 nm is mainly related to the
225 oxidation and hydrolytic degradation of lipids (Cozzolino et al. 2005). It is possible to
226 observe that the immature kernels, classified as kernel which is misshapen, abnormally
227 small or partially aborted, including shriveled and shrunken kernels (SAMAC, 2016)
228 present a lower absorption band (2,200 nm - 2,500 nm) (Figure 2B). This result might
229 be due to the fact that maturity is correlated with oil content (Cavaletto, 1985),
230 consequently with less lipid degradation.

231 *3.1.1. Model development*

To correlate the FT-NIR spectra to the quality categories, discriminant classifications based on PLS-DA and GA-LDA were used and compared and evaluated in terms of sensitivity, specificity, accuracy, separately for each category.

Regarding pre-processing, SNV lead to best results using PCA-LDA, resulting in an accuracy of 68 % and a specificity of 97 % for immature kernels (Table 1). The accuracy shows the proportion of samples correctly grouped, while specificity represents the probability of a sample without the desired characteristic to be given a negative test result (Amodio et al., 2017). However, the sensitivity presented low values (67 %), and this parameter describes the model ability to correctly recognize samples belonging to a class (Ballabio and Consonni, 2013). For example, if none of the marketable kernels were classified as other class (FN is equal to zero), the sensitivity for the marketable kernels class would have been equal to 100 %.

Cannedu et al. (2016) classified marketable macadamia kernels in relation to non-marketable kernels using PLS-DA and reported percentages of 88 % for calibration and 87 % for prediction. These results were inferior than what we obtained, probably because we used more samples (n = 300) than Cannedu et al. (2016) (n = 100). Therefore, the inclusion of more data into the dataset improved the robustness and increase the classification accuracy.

Marketable kernels and kernels with defects (immature, insect damage, mold, and discoloration) could be discriminated from each other using GA-LDA (Figure 3). The accuracy and specificity of GA-LDA for marketable kernels achieved a value of 97.8 % and 100 %, respectively (Table 2).

To perform the GA-LDA, some of the wavelengths were selected (Table 3). This selection was based on compounds of particular interest, e.g., 1,020 nm and 1,173 nm, representing the C–H groups from lipids; 1,485 nm and 1,789 nm, related to the

first overtone of stretching and anti-symmetric O–H bond and second overtone of stretching O–H bend, respectively. Absorption bands at the wavelength near 1,450 and 1,940 nm are related to the presence of water in foods (Moscetti et al., 2014) and this explains why the wavelengths 1,485 nm, 1,975 nm and 1,987 nm were selected by GA.

It is possible to observe that the kernels with discoloration had a higher moisture content than the others (Figure 2B), and these moisture contents correspond to water activities (aw) greater than 0.8 at which browning reaction rates are high (Wall, 2013), and maintaining nuts-in-shell at high moisture content can cause discoloration (Walton et al., 2013).

**3.2. TD-NMR**

The typical curves of the CPMG decays for the different defects found in macadamia kernels can be seen in Figure 4. It can be observed that kernels with insect damage presented a faster settling time compared to the others, whereas the kernels with presence of fungi (moldy) showed the slower signal decay (Figure 4).

The intensity of the TD-NMR signals from relaxation (our case) and diffusion measurements is related to the water content related to water status, water compartmentalization and molecular mobility in the food sample (Kirtil et al., 2017). In order to evaluate the influence of the water content on the nutrient content of the food, it is important to note that there are variations in the moisture content of the kernels, since these moisture contents correspond to water activities at which microbial growth rates are high (Wall, 2013). This explains the fact that moldy kernels have a higher moisture content.

In Figure 5 it is possible to observe that there was not a clear separation between the defect classes. However, in Figure 5A there was a tendency of separation between the good and immature kernels. Probably because there are differences in the

decay time between these classes (Figure 4), with showed that the most rapid decay is due to solid components, mainly composed of proteins and carbohydrates (Prestes et al., 2007) and immature kernels present a higher carbohydrate concentration, represented by sucrose and fructose higher than mature kernels (Wall, 2013).

*3.2.1. Model development*

The best TD-NMR classification models were obtained using the PCA-LDA and GA-LDA without pre-processing the signals (Table 4). Using PCA-LDA, it was possible to achieve 86 % accuracy for the training set and 83.3 % for the validation set to classify kernels with insect damage. On the other hand, the GA-LDA analysis obtained 64 % for the calibration set and 100 % for the validation set, allowing the use of this model to classify kernels with insect damages.

TD-NMR has been used to classify other oleaginous produces including nuts. Di Caro et al. (2017) studying not damaged and moldy hazelnuts kernels highlighted that NMR might be used to discriminate oils extracted from both kernel classes. Di Caro (2018) also reported that using NMR was possible to obtain values of 97 % for sensitivity and 81 % for specificity to classify in-shell damaged hazelnuts. Therefore, NMR might be a useful analytical tool for quality control in nut industry.

**3.3. NIRS versus TD-NMR**

The results obtained from both techniques for the development of the classification models for macadamia kernels quality defects can be seen in Table 1, 2, and 4. Overall, the NIRS showed better classification capability as higher values of accuracy were obtained using GA-LDA models. The lower performance of the classification models developed using the TD-NMR signals might be related to the number of samples, as just the kernels harvested in June 2017 were used.

NIRS and TD-NMR present many similarities as they are fast non-destructive analytical methods, do not need sophisticated sample preparation, and the results can be collected, processed, and stored directly in a microcomputer (Colnago, 1996; Pasquini, 2003). However, when it comes to NMR spectroscopy, high cost is normally considered as one of the most serious drawbacks and this technique requires special skills to interpret the spectra acquisition (Xu et al., 2015). Another limitation of NMR spectroscopy is the insensitivity to minor fat component detection (Kucha et al., 2018). These suggest that, due the fact that NIRS is useful for detecting components with up to 0.1 % concentration (Xu et al., 2015) and NMR presents lower sensitivity, NIRS models presented more satisfactory results.

## 4. Conclusions

NIRS and TD-NMR combined with chemometric methods proved to be powerful tools to classify macadamia kernels based on their quality defects. However, NIRS out-performed TD-NMR based on the higher accuracy results.

NIRS and TD-NMR spectroscopy can be successfully used to evaluate the quality of unshelled macadamia nuts and have potential to improve the existing postharvest techniques used in the macadamia industry.

## Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

## References

Abbott, J.A. (1999). Quality measurement of fruit and vegetables. *Postharvest Biology and Technology, 15*, 207-223.

Amodio, M.L., Ceglie, F., Chaudhry, M.M., Piazzolla, F. & Colelli, G. (2017). Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. *Postharvest Biology and Technology, 125*, 112– 121. Doi.Org/10.1016/J.Postharvbio.2016.11.013.

Armenta, S., Garrigues, S. & De la Guardia, M. (2007). Determination of edible oil parameters by near infrared spectrometry. *Analytica Chimica Acta, 596*, 330-337. Doi:10.1016/j.aca.2007.06.028

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods, 5*, 3790–3798. Doi: 10.1039/c3ay40582f

Bro, R., & Smilde A. K. (2014). Principal component analysis. *Analytical Methods, 6*, 2812-2831. Doi:10.1039/c3ay41907j

Canneddu, G., Júnior, L.C. & Teixeira, G.H.A. (2016). Quality evaluation of shelled and unshelled macadamia nuts by means of near-infrared spectroscopy (NIR). *Journal of Food Science, 81*, 1613-1621. Doi.org/10.1111/1750-3841.13343

Carvalho, L.C., Morais, C.L., Lima, K.M.G., Cunha Júnior, L.C., Nascimento, P.A., Faria, J.B. & Teixeira, G.H.A. (2016). Determination of the geographical origin and ethanol content of Brazilian sugarcane spirit using near-infrared spectroscopy coupled with discriminant analysis. *Analytical Methods, 8*, 5658-5666.

Carvalho, L.C., Morais, C.L.M, Lima, K.M.G., Leite, G.W.P., Oliveira, G.S., Casagrande, I.P., Santos Neto, J.P. & Teixeira, G.H.A. (2018). Using intact nuts and near infrared spectroscopy to classify macadamia cultivars. *Food Analytical Methods, 11*, 1857–1866. https://doi.org/10.1007/s12161- 017-1078-9

Carvalho, L.C., Marcondes, L.L., Morais, C.L.M., Lima, K.M.G., & Teixeira, G.H.A. (2019). Non-destructive assessment of the oxidative stability of intact macadamia nuts during the drying process by near-infrared spectroscopy. *LWT – Food Science and Technology, 103*, 101-107. Doi.org/10.1016/j.lwt.2018.12.056

Colnago, L.A. (1996). *Análise do teor de óleos em sementes por RMN*. São Carlos: Embrapa Instrumentação Agropecuária. Circular Técnica, 3, 14p.

Cozzolino, D., Murray, I., Chree, A. & Scaife, J.R. (2005). Multivariate determination of free fatty acids and moisture in fish oils by partial least-squares regression and near-infrared spectroscopy. *LWT – Food Science and Technology*, 38, 821-828.

Di Caro, D. (2018). NMR measurements for hazelnuts classification. PhD dissertation. Università Degli Studi Di Salerno. Salerno, Italy.

Liguori, C., Pietrosanto, A., Sommella, P. (2017). Hazelnut oil classification by NMR techniques. *IEEE transactions on instrumentation and measurement, 66,* 928-934. Doi: 10.1109/TIM.2016.2640438

Di Caro, D., Liguori, C., Pietrosanto, A., Sommella, P. (2017). Hazelnut oil classification by NMR techniques. *IEEE transactions on instrumentation and measurement, 66,* 928-934. Doi: 10.1109/TIM.2016.2640438

Dixon, S.J., & Brereton, R.G. (2009). Comparison of performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure. *Chemometrics and Intelligent Laboratory Systems, 95,* 1-17.

Jensen, P.N., Sorensen, G., Soren, B.E. & Bertelsen, G. (2001). Evaluation of quality changes in walnut kernels (*Juglans regia* L.) by Vis/NIR spectroscopy. *Journal of Agricultural and Food Chemistry,* 49, 5790–5796. Doi: 10.1021/jf010509t

Kennard, R.W., & Stone, L.A. (1969). Computer aided design of experiments. *Technometrics, 11,* 137–148. Doi: 0.1080/00401706.1969.10490666

Kirtil, E., Cikrikci, S., Mccarthy, M. J., & Oztop, M.H. (2017). Recent advances in time domain NMR & MRI sensors and their food applications. *Current Opinion in Food Science, 17,* 9-15.

Kucha, C.T., Liu, L., & Ngadi, M.O. (2018). Non-destructive spectroscopic techniques and multivariate analysis for assessment of fat quality in pork and pork products: a review. *Sensors, 18,* 1-23.

Lanza, E. & Li, B.W. (1984). Applications of near infrared spectroscopy for detecting the sugar content of fruit juices. *Journal of Food Science*, *49*, 995-998.

McCall, J. (2005). Genetic algorithms for modelling and optimization. *Journal of Computational and Applied Mathematics – Special Issue: Mathematics applied to immunology, 184*, 205-222.

Medeiros, E. P., Feliz, P.H.D., Melo, J.K.A., Barreto, A. N., & Nascimento, J.J.V.R. (2008). Um método não destrutivo por RMN de baixo campo e quimiometria para classificação de sementes de mamona em função do perfil espectral do óleo. III Congresso brasileiro de Mamona, energia e recinoquímica.

Morais, C.L.M. & Lima, K.M.G. (2017). Comparing unfolded and two-dimensional discriminant analysis and support vector machines for classification of EEM data. *Chemometrics and Intelligent Laboratory Systems*, *170*, 1-12

Morais, C.L.M. & Lima, K. M. G. (2018) Principal component analysis with linear and quadratic discriminant analysis for identification of cancer samples based on mass spectrometry. *Journal of the Brazilian Chemistry Society*, *29*, 472-481.

Nogueira, N.R. (2008). O agronegócio da macadâmia: uma breve análise dos aspectos de produção, econômico e social. *Interface tecnológica, 5*, 21-28.

Pasquini, C. (2003). Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *Journal of Brazilian Chemistry Society, 14*, 198-219.

Pedersen, H., Munck, L., & Engelsen, S. (2000). Low-field $^{1}$H nuclear magnetic resonance and chemometrics combined for simultaneous determination of water, oil, and protein contents in oilseeds. *Journal of the American Oil Chemists' Society*, *77*, 1069–1077.

Ripperton, J.C., Moltzau, R.H., & Edwards, D.W. (1938). *Methods of evaluating the macadamia nut for commercial use and the variation occurring among seedling plantings in Hawaii*. Washington: Hawaii Agricultural Experiment Station Bulletin, N°. 79.

SAMAC. Southern African Macadamia Growers' Association. (2018). Raw macadamia products quality specifications. http://www.samac.org.za/index.php/standards/ Acessed 12 August 2018.

O'Hare, P., Hidden, D.V., Burton, D. & Salmon, T. (2004). Australia macadamia industry code of sound orchard practices. Lismore: Australian macadamia society.

Xu, J., Riccioli, C., & Sun D. (2015). An overview on destructive spectroscopic techniques for lipid and lipid oxidation analysis in fish and fish products. *Comprehensive Reviews in Food Science and Food Safety*, *14*, 466-477.

Wall, M.M. (2013). Improving the quality and safety of macadamia nuts. In L.J. Harris, (Ed.), *Improving the quality and safety of nuts* (pp.274-296). Cambridge: Woodhead Publishing Series in Food Science and Technology and Nutrition.

Walton, D.A., Randall, B.W., Le Lagade, M.D. & Wallace, H.M. (2013). Maintaining high moisture content of macadamia nuts-in-shell during storage induces brown centres in raw kernels. *Journal of the Science of Food and Agriculture*, 93, 2953–295.

Wu, W., Mallet, Y., Walczak, B., Penninckx, W., Massart, D.L., Heuerding, S. & Erni, F. (1996). Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data. *Analytica Chimica Acta*, 329, 254-265.

## Tables

**Table 1**. Values of accuracy, sensitivity and specificity to classify macadamia kernels based on quality defects using PCA-LDA and NIRS.

| **Classes** | | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|---|
| **Pre-Processing** | | | | | | |
| **Raw** | AC(%) | 88.9 | 84.4 | 75.6 | 82.2 | 75.6 |
| | SENS(%) | 88.9 | 66.7 | 44.4 | 44.4 | 22.2 |
| | SPEC(%) | 88.9 | 88.9 | 83.3 | 91.7 | 88.9 |
| **SNV** | AC(%) | 80.0 | 68.9 | 88.9 | 75.6 | 75.6 |
| | SENS(%) | 66.7 | 55.6 | 55.6 | 11.1 | 33.3 |
| | SPEC(%) | 83.3 | 72.2 | 97.2 | 91.7 | 86.1 |
| **2$^{nd}$ Derivative** | AC(%) | 82.2 | 73.3 | 86.7 | 88.9 | 75.6 |
| | SENS(%) | 66.7 | 44.4 | 77.8 | 66.7 | 11.1 |
| | SPEC(%) | 86.1 | 80.6 | 88.9 | 94.4 | 91.7 |

SNV= standard normal variate; AC= accuracy; SENS= sensitivity; SPEC= specificity. 1=good, marketable kernels without defects; 2=kernels with discoloration; 3=immature kernels; 4=kernels affected by mold; and 5=kernels with insect damage.

441

442

443 **Table 2**. Values of accuracy, sensitivity and specificity to classify macadamia kernels
444 based on quality defects using GA-LDA and NIRS.

| **Classes** | | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|---|
| **Pre-Processing** | | | | | | |
| **Raw** | AC(%) | 86.7 | 82.2 | 86.7 | 86.7 | 82.2 |
| | SENS(%) | 66.7 | 66.7 | 55.6 | 66.7 | 55.6 |
| | SPEC(%) | 91.7 | 86.1 | 94.4 | 91.7 | 88.9 |
| **SNV** | AC(%) | 97.8 | 84.4 | 88.9 | 91.1 | 84.4 |
| | SENS(%) | 88.9 | 88.9 | 55.6 | 77.8 | 55.6 |
| | SPEC(%) | 100 | 83.3 | 97.2 | 94.4 | 91.7 |
| **2$^{nd}$ Derivative** | AC(%) | 91.1 | 75.6 | 84.4 | 86.7 | 68.9 |
| | SENS(%) | 66.7 | 44.4 | 44.4 | 55.6 | 55.6 |
| | SPEC(%) | 97.2 | 83.3 | 94.4 | 94.4 | 72.2 |

445 SNV= standard normal variate; AC= accuracy; SENS= sensitivity; SPEC= specificity.
446 1=good, marketable kernels without defects; 2=kernels with discoloration; 3=immature
447 kernels; 4=kernels affected by mold; and 5=kernels with insect damage.

448

449

450

451 **Table 3.** Selected variables for GA-LDA to classify macadamia kernels using different

452 pre-processing.

| Pre-processing | Selected variables (nm) |
|---|---|
| **Raw** | 882; 886; 946; 990; 1171; 1395; 1429; 1511; 1622; 1664; 1942; 1979; 2075; 2187; 2260; 2328 |
| **SNV** | 866; 1020; 1173; 1280; 1485; 1578; 1789; 1975; 1987; 2083; 2170; 2277; 2300; 2388; 2451 |
| **2nd Derivative** | 894; 898; 1078; 1251; 1335; 1436; 1488; 1952; 1964; 2126; 2328; 2356 |

453 SNV=standard normal variate

454

455

456 **Table 4**. Values of accuracy to classify macadamia kernels based on quality parameters
457 using PCA-LDA, GA-LDA and TD-NMR spectroscopy.

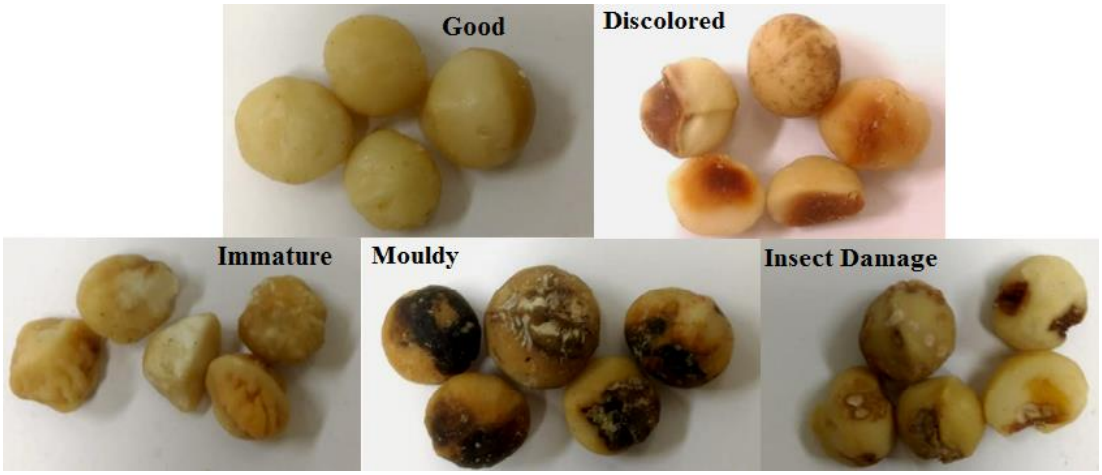
| **Classes** | | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|---|
| **Pre-Processing** | | | | | | |
| **Nil** | PCA-LDA | | | | | |
| | Training (%) | 64.3 | 35.7 | 42.9 | 85.7 | 64.3 |
| | Validation (%) | 16.7 | 33.3 | 16.7 | 66.7 | 83.3 |
| | GA-LDA | | | | | |
| | Training (%) | 64.3 | 50.0 | 35.7 | 64.3 | 50.0 |
| | Validation (%) | 66.7 | 16.7 | 66.7 | 66.7 | 100 |

458 1=good, marketable kernels without defects; 2=kernels with discoloration; 3=immature
459 kernels; 4=kernels affected by mold; and 5=kernels with insect damage.

460

461

462 **Figures**




463

464 **Figure 1.** Macadamia kernels quality defects: 1=good, marketable kernels without

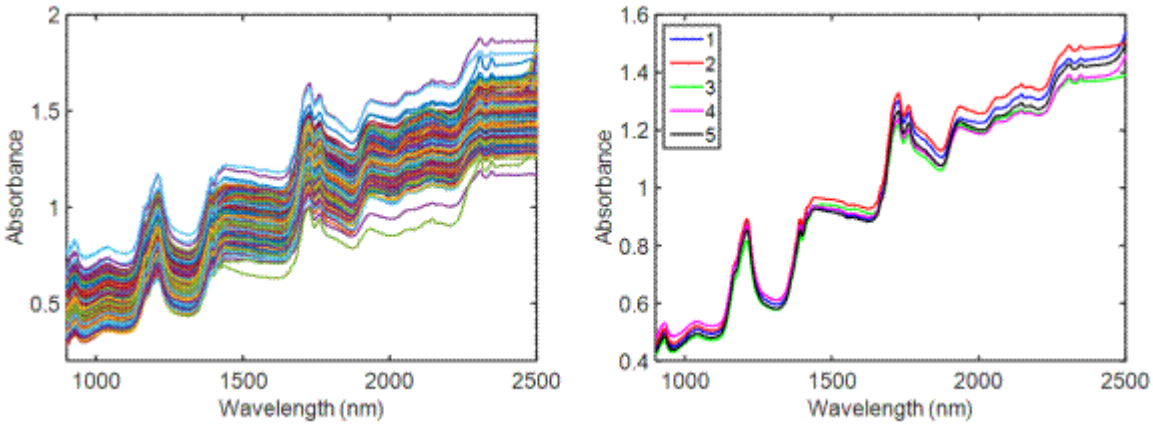
465 defects; 2=kernels with discoloration; 3=immature kernels; 4=kernels affected by mold;

466 and 5=kernels with insect damage.

467

468

469



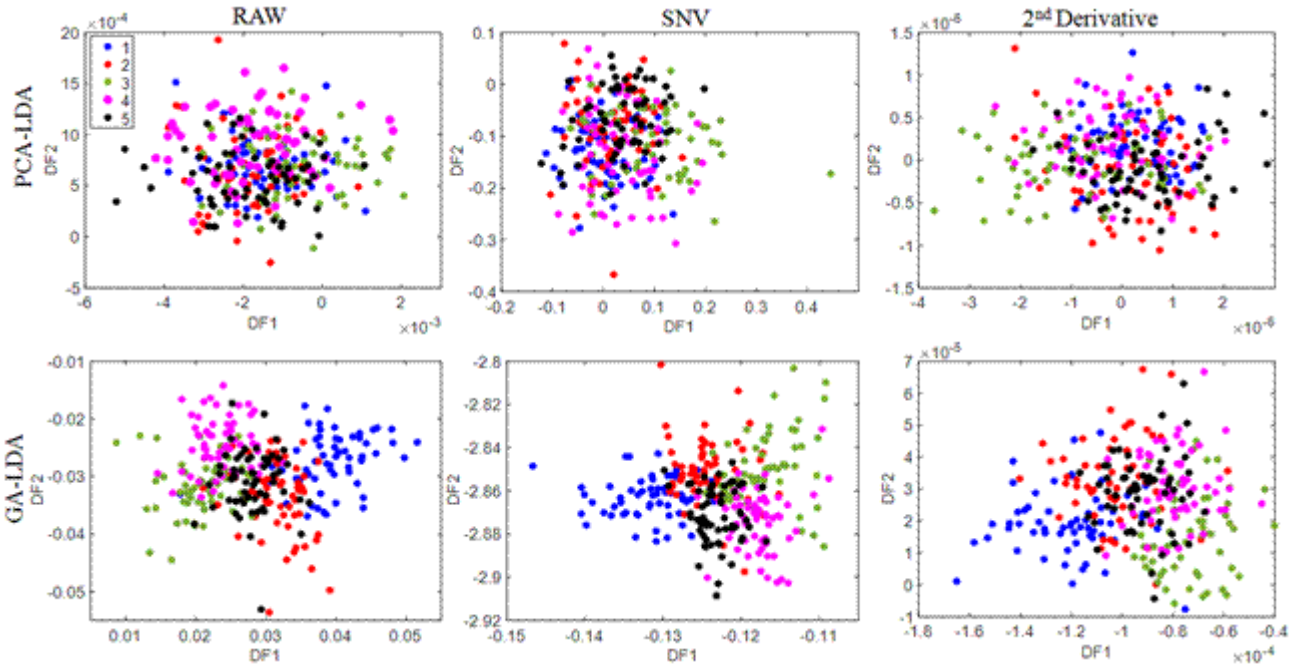

470

471 **Figure 2.** Raw NIR spectra (a) and average NIR spectra (b) of macadamia kernels.
472 1=good, marketable kernels without defects; 2=kernels with discoloration; 3=immature
473 kernels; 4=kernels affected by mold; and 5=kernels with insect damage.

474

475

476




477

478 **Figure 3.** Discriminant function (DF) plot of PCA-LDA and GA-LDA with raw NIR

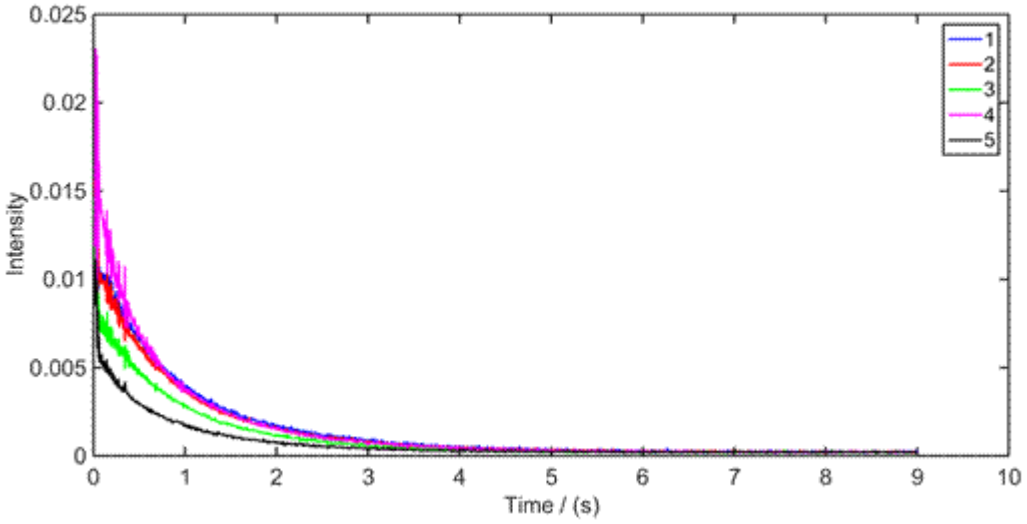
479 spectra of macadamia kernels, SNV and 2nd derivative Savitzky-Golay. 1=good,

480 marketable kernels without defects; 2=kernels with discoloration; 3=immature kernels;

481 4=kernels affected by mold; and 5=kernels with insect damage.

482

483




484

485 **Figure 4.** Typical CPMG decay curves of macadamia kernels with different quality
486 defects. 1=good, marketable kernels without defects; 2=kernels with discoloration;
487 3=immature kernels; 4=kernels affected by mold; and 5=kernels with insect damage.

488

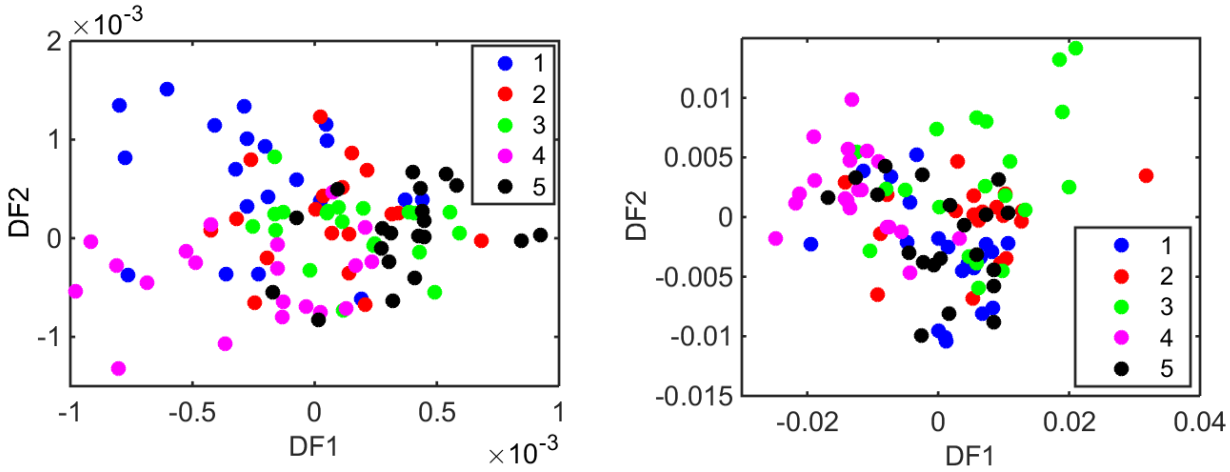



**Figure 5.** Discriminant function (DF) of PCA-LDA (A) and GA-LDA (B) with raw TD-NMR spectra of macadamia kernels. 1=good, marketable kernels without defects; 2=kernels with discoloration; 3=immature kernels; 4=kernels affected by mold; and 5=kernels with insect damage.