

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Short communication

Using near infrared spectroscopy to classify soybean oil according to expiration date



Gean Bezerra da Costa ^a, David Douglas Sousa Fernandes ^a, Adriano A. Gomes ^b, Valber Elias de Almeida ^a, Germano Veras ^{a,*}

^a Laboratório de Química Analítica e Quimiometria, Centro de Ciências e Tecnologia, Universidade Estadual da Paraíba, 58.429-500 Campina Grande, PB, Brazil

^b Departamento de Química, CCEN, Universidade Federal da Paraíba, 58.051-970 João Pessoa, PB, Brazil

ARTICLE INFO

Article history:

Received 1 April 2013

Received in revised form 14 September 2015

Accepted 20 September 2015

Available online 28 September 2015

Keywords:

Classification

Conservation state

Linear Discriminant Analysis

Successive Projection Algorithm

Discriminant Analysis by Partial Least

Squares

ABSTRACT

A rapid and non-destructive methodology is proposed for the screening of edible vegetable oils according to conservation state expiration date employing near infrared (NIR) spectroscopy and chemometric tools. A total of fifty samples of soybean vegetable oil, of different brands and lots, were used in this study; these included thirty expired and twenty non-expired samples. The oil oxidation was measured by peroxide index. NIR spectra were employed in raw form and preprocessed by offset baseline correction and Savitzky–Golay derivative procedure, followed by PCA exploratory analysis, which showed that NIR spectra would be suitable for the classification task of soybean oil samples. The classification models were based in SPA–LDA (Linear Discriminant Analysis coupled with Successive Projection Algorithm) and PLS–DA (Discriminant Analysis by Partial Least Squares). The set of samples (50) was partitioned into two groups of training (35 samples: 15 non-expired and 20 expired) and test samples (15 samples: 5 non-expired and 10 expired) using sample-selection approaches: (i) Kennard–Stone, (ii) Duplex, and (iii) Random, in order to evaluate the robustness of the models. The obtained results for the independent test set (in terms of correct classification rate) were 96% and 98% for SPA–LDA and PLS–DA, respectively, indicating that the NIR spectra can be used as an alternative to evaluate the degree of oxidation of soybean oil samples.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Vegetable oils are complex mixtures of organic compounds. The main constituents are triglycerides in addition to other minor compounds. This composition gives the food a high energetic value, also being a source of vitamins; moreover, it gives a good appearance to food. However, oxidation is a spontaneous, unavoidable and most important process, caused mainly by oxygen from air, heat or light exposure, that modifies oil properties, causing changes in color, taste and odor, which makes it unsuitable for consumption and decreases the product value (Conde, Gordon, Moure, & Dominguez, 2011; Fadda et al., 2012; Kim, Yeo, Kim, Kim, & Lee, 2013; Lerma-García, Simó-Alfonso, Bendini, & Cerretani, 2011; Ramalho & Jorge, 2006; Silvagni, Franco, Bagno, & Rastrelli, 2012).

The quality of an oil sample can be assessed by various physicochemical analyses, acid value, peroxide value or other methods such as Ramciat. High values of peroxide indicate that the oil is oxidized,

while low values show that the sample is in an acceptable condition, or at least in the initial stage of degradation (Silva, Borges, & Ferreira, 1999).

In Brazil, unlike other countries, the marketing of products with expired shelf life is prohibited. Several chemical and physical characterizations can be used to indicate the conservation state of the vegetable oil samples. Acidity and peroxide levels are important quality indexes of oils, and are used as parameters by the Brazilian National Agency for Sanitary Vigilance (Brasil, 2005). However, this methodology is laborious and requires a large volume of chemical reagents (Frankel, 1993; Osawa, Gonçalves, & Ragazzi, 2006; Presa-Owens, López-Sabater, & Rivero-Urgell, 1995).

Other analytical techniques for example, chromatography, mid-infrared spectroscopy, Raman scattering, fluorescence, chemiluminescence and square wave voltammetry have been used for quality control of oils (Gambarra-Neto et al., 2009; Sikorska, Górecki, Khmelinskii, Sikorski, & Koziol, 2005). In this context, near infrared spectroscopy has been reported in the literature as being among the fast, non-destructive and non-invasive techniques (Cooper, Larkin, & Abdelkader, 2011; Pereira et al., 2008; Wang, Lee, Wang, & He, 2006). These advantages have motivated the

* Corresponding author.

E-mail address: germano.veras@pq.cnpq.br (G. Veras).

development of analytical methodologies employing this spectral range for quality control of several complex matrices, for example, fuels (Balabin & Safieva, 2008; Breitzkreitz et al., 2003; Veras et al., 2010), food (Chen et al., 2011; Olivieri, Di Egidio, Woodcock, & Downey, 2011), drinks (Pontes et al., 2006) and drugs (Bittner et al., 2011), among others.

The analytical signal in near infrared corresponds to overtones of fundamental transitions, thus the information is found to be overlapping (Pasquini & Braz, 2003), which makes necessary the application of chemometric tools for multivariate data analysis.

In this work, a rapid, non-expensive and non-destructive methodology is proposed for soybean oil screening with respect to the conservation state of expired and non-expired shelf life, based on the expiration date indicated on the label, and peroxide index, using near infrared spectroscopy. Hard and soft modelling techniques SPA-LDA and PLS-DA, respectively were used. The method robustness was tested using several approaches for sample selection.

2. Experimental

2.1. Sample acquisition

A total of thirty samples from different brands and lots of soybean oil were acquired and stored until the expiration date indicated on the label. Later, twenty more samples were acquired to compose the non-expired sample class. The samples were acquired in the city of Campina Grande, Brazil.

2.2. Oxidation state of oils

The oxidation state of soybean oil samples, the principal characteristic of oil degradation, was evaluated by peroxide index. The experimental procedure for evaluating the peroxide index was as follows: 5 g of oil was weighted and 30 mL acetic acid/chloroform 3:2 was added. Then, 0.5 mL saturated KI was added with 1 min resting in the dark. In sequence, 30 mL deionized water and 1 mL starch were added as indicator. The solution was titrated with sodium thiosulfate (0.01 mol/L) until the dark color disappeared.

2.3. Spectra acquisition

The spectra of vegetable oil samples were recorded in triplicate, ranging from 1100 to 1600 nm, with 1 nm resolution in Perkin Elmer spectrophotometer 750 Lambda, equipped with 1 cm optical path quartz cell, tungsten source and R928 photomultiplier tube and Peltier-cooled PbS detection systems.

2.4. Data analysis and software

The data set is composed of 50 samples, each with 501 wavelengths spectrum, resulting in a 50×501 dimension matrix. The spectra were used in raw form and pre-treated using a Savitzky-Golay algorithm with the first derivative, window with 17 points and second-order polynomial. Initial exploratory analysis was carried out by PCA for each preprocessing, for a preliminary assessment of the discriminant power of NIR spectra. For constructing the classification models for SPA-LDA (Moreira, Pontes, Galvão, & Araújo, 2009; Pontes et al., 2005; Soares, Galvão, Pontes, & Araújo, 2014) and PLS-DA (Barker & Rayens, 2003), the fifty sample set was partitioned into two groups – 35 samples for training (15 non-expired and 20 expired) and 15 for testing (5 non-expired and 10 expired) – using three different approaches, Kennard–Stone (Kennard & Stone, 1969), Duplex (Snee, 1977), and Random, in order to evaluate the robustness of the models.

3. Results/discussion

3.1. Peroxide value

The fifty samples studied in this paper were titrated according to the experimental procedure described in Section 2.2. Peroxide values are showed in Fig. 1. The non-expired samples show an average value for peroxide equal to 7.06 with low variability (standard deviation of 1.7) while the expired samples present peroxide rates in a wide range (from 16.5 to 348.6, with an average equal to 124.2 and a standard deviation of 99.3). This wide variation may be linked to several factors, such as the type of antioxidant used by each manufacturer and the form of storage employed.

Although some authors show that the peroxide values present a variation over time on a Gaussian shape, Fig. 1 shows two groups of samples with different behavior, depending on the oxidation state, which in this case was assessed by the peroxide value. This result suggests that this information on the oxidation state of soybean oil samples may be obtained by means of spectroscopic techniques like NIR.

3.2. NIR spectra

Fig. 2a presents the raw spectra of fifty soybean oil samples ranging between 1100 and 1600 nm. The spectra also present in the baseline feature have been treated as described in Section 2.4. The corrected spectra were displayed in Fig. 2b and c.

The blue line spectra are from the expired samples while the red lines are from non-expired samples. It is possible to observe a strong overlap among the spectra (blue and red lines) and a visual inspection of the spectra does not allow an easy identification of the samples' conservation state. However, in this spectral range there are signals associated with first overtones of RC₂OR and RCOR, which are functional groups directly involved with vegetable oil's oxidation process. In this range there are also associated information with the combination bands of C–C, CH₂, CH₃ and CHO (Xiaobo, Jiewen, Povey, Holmes, & Hanpin, 2010). This suggests that NIR data contains information associated with the degradation stage of soybean oil samples and can be assessed through a multivariate approach.

3.3. Exploratory analysis by PCA

Fig. 3a –c present the PC2 \times PC1 score plot resulting from the application of PCA to raw data, corrected by offset and first derivative, respectively. As can be seen, although with a slight overlap, for all cases the result of PCA showed the formation of two sample

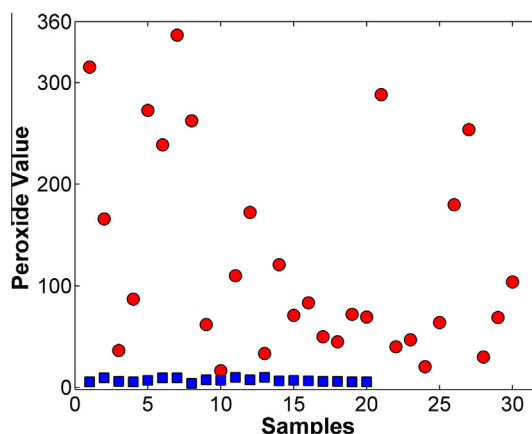


Fig. 1. Peroxide value versus index samples for all data set.

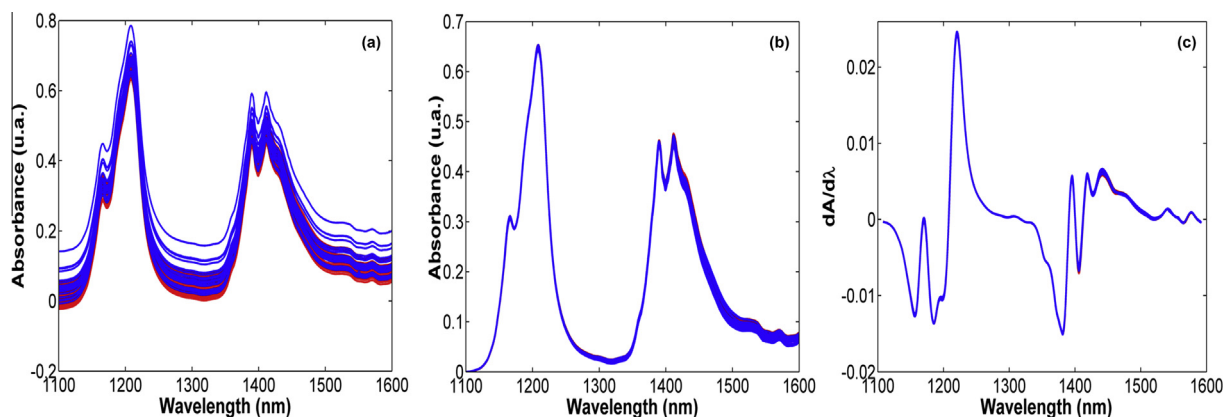


Fig. 2. Spectra of the fifty soybean oil samples (a) raw, (b) corrected by offset and (c) derivative.

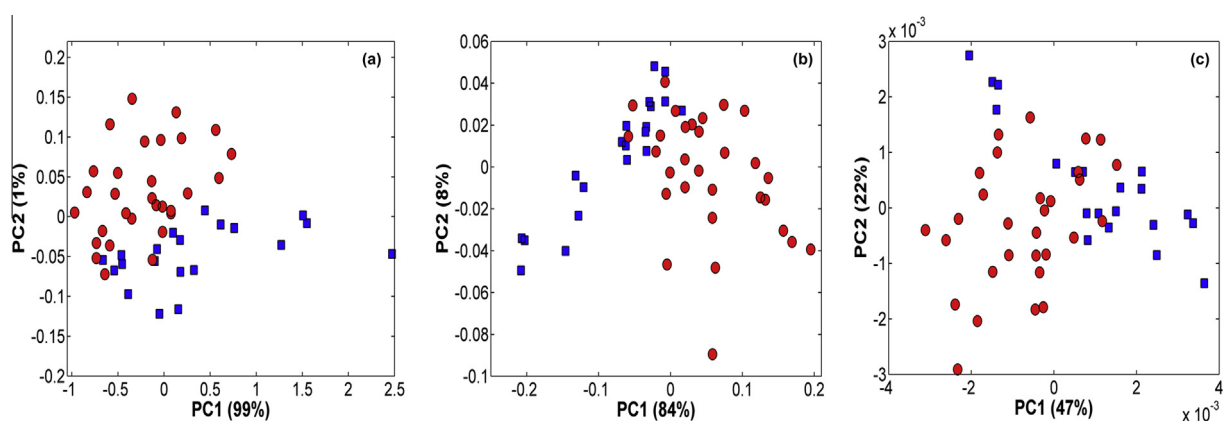


Fig. 3. PC1 versus PC2 score plot for all data (a) raw, (b) corrected by offset and (c) derivative.

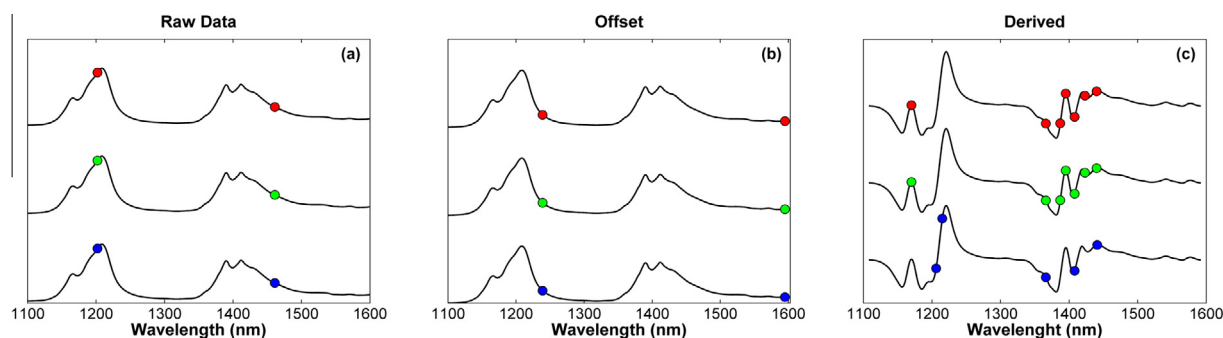


Fig. 4. Selected variables by SPA-LDA for (●) Kennard–Stone, (●) Duplex and (●) Random.

groups (expired and non-expired), confirming that the NIR spectra have the proper information for the task of classifying soybean oil samples.

Regardless of the data processing type, a similar result is observed, samples are discriminated by PC1 and PC2 bisectrix, showing that the information contained in both PCs are important for this study.

3.4. Supervised pattern recognition

To develop a method for classifying soybean oil samples, two approaches for supervised pattern recognition were employed: SPA-LDA and PLS-DA. Raw data, spectra corrected by offset and derivative were used combined with three different methods of

partitioning samples sets (Kennard–Stone, Duplex and Random) to evaluate the robustness of the proposed method. The combination of different methods of data processing with different methods of samples selection result in nine SPA-LDA and nine PLS-DA models.

3.4.1. SPA-LDA classification

LDA is a hard method for classification. It employs Mahalanobis distance, and needs a preview step of the variable selection. In this work, this step was carried out by SPA. The selected wavelengths by SPA using raw data (Fig. 4a), spectra corrected by offset (Fig. 4b), and derivative spectra (Fig. 4c), are displayed in average spectrum in the form of colored circles. Red, green and blue circles

Sample Select Algorithm

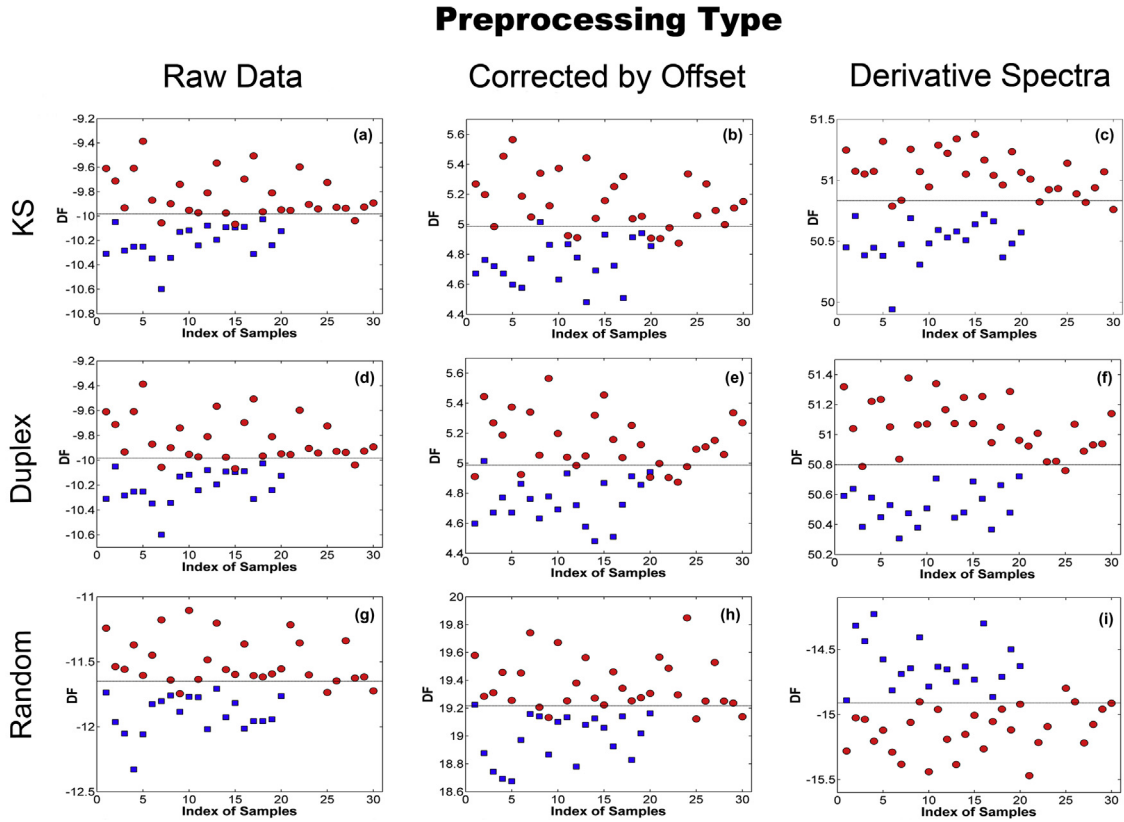


Fig. 5. Discriminant function for each LDA model, (●) expired samples and (■) non-expired samples.

Sample Select Algorithm

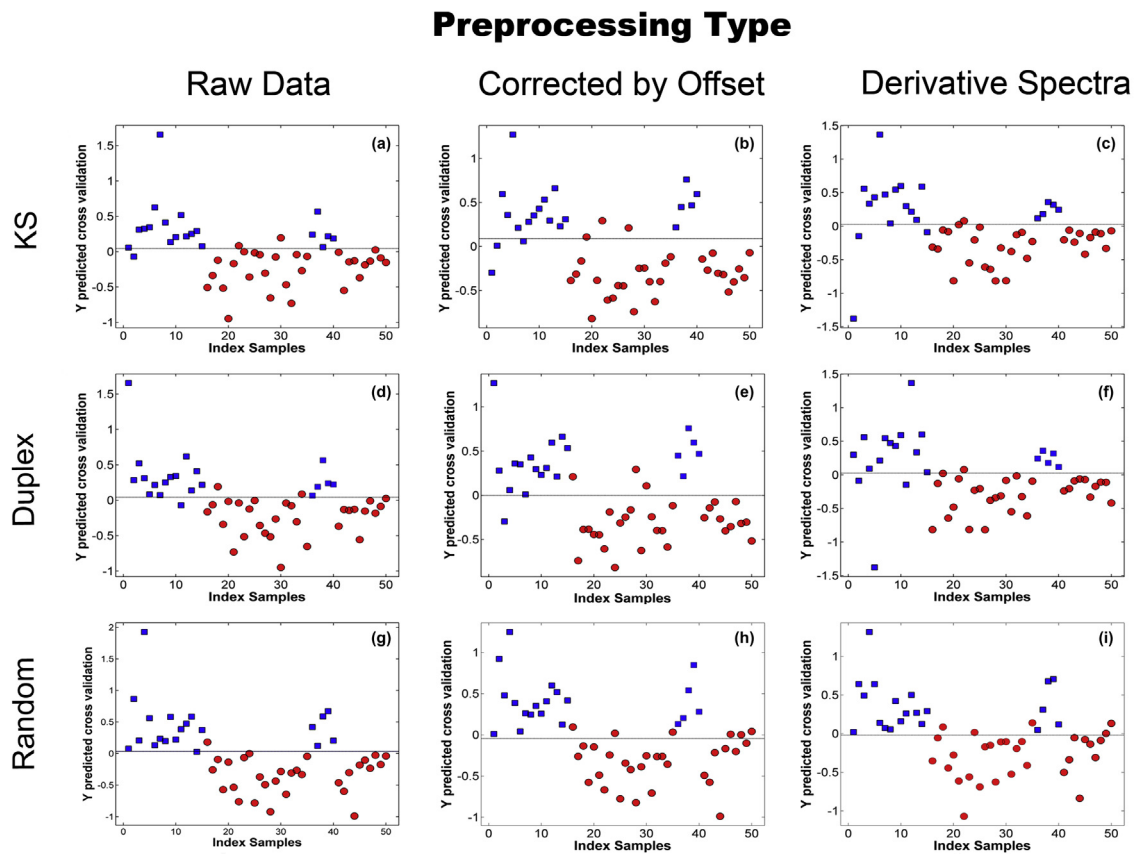


Fig. 6. Y predicted by PLS-DA model, (●) expired samples and (■) non-expired samples.

indicate models built with the dataset partition using Kennard–Stone, Duplex, and Random methods, respectively.

For all cases, the subset of the selected variable by SPA does not change depending on the type of sample selection algorithm (training and testing), but change according to the type of processing applied to the data, showing which SPA–LDA is robust with respect to the composition of the training set. For raw data and the data corrected by offset, the minimum value for cost function associated with variable selection by SPA was obtained for two wavelengths: 1202 and 1461 nm for raw data. For the spectra corrected by offset, these values were 1239 and 1595 nm. For derivative spectra the minimum value for cost function was obtained with more variables concentrated around 1400 nm (see Fig. 4c).

Fig. 5 shows the discriminant functions computed as a linear combination of variables selected by SPA. As can be seen, SPA–LDA models show a good performance in classifying soybean oil samples.

A better approach was achieved using SPA–LDA with derivative spectra, independent of the sample set partition (see Fig. 5). All samples in accordance (non-expired) were classified correctly and only two of the expired samples were misclassified. This represents a 96% rate of correct classification.

3.4.2. PLS–DA classification

As an alternative method, PLS–DA was performed on NIR data using full spectrum. The results are presented in Fig. 6. With soft modelling by PLS–DA of NIR spectra using raw data and random selection for training samples, it was possible to correctly classify 98% of the whole samples set (see Fig. 6g).

PLS–DA showed better results in comparison to SPA–LDA; however, these results changed as a function of the training set, while SPA–LDA showed better robustness with respect to different training sets.

4. Conclusion

This work describes a rapid and non-destructive methodology for the evaluation of the oxidative state of soybean oil, employing NIR spectroscopy and chemometric tools. This study has great potential as an alternative to conventional techniques that require much effort and reagent consumption. PLS–DA achieved better results in terms of correct classifications, but SPA–LDA showed a simpler and more robust model employing small sets of wavelengths.

Acknowledgments

The authors are grateful to CAPES and CNPq for scholarships.

References

- Balabin, R. M., & Safieva, R. Z. (2008). Gasoline classification by source and type based on near infrared (NIR) spectroscopy data. *Fuel*, 87, 1096–1101.
- Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, 17, 166–173.
- Bittner, L. K. H., Heigl, N., Petter, C. H., Noisternig, M. F., Griesser, U. J., Bonn, G. K., & Huck, C. W. (2011). Near-infrared reflection spectroscopy (NIRS) as a successful tool for simultaneous identification and particle size determination of amoxicillin trihydrate. *Journal of Pharmaceutical and Biomedical Analysis*, 54, 1059–1064.
- Brasil (2005). *Diário Oficial da União*, 184, 373.
- Breitkreitz, M. C., Raimundo Junior, I. M., Jarbas, J. R., Pasquini, C., Dantas Filho, H. A., José, G. E., & Araújo, M. C. U. (2003). Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration. *Analyst*, 128, 1204–1207.
- Chen, L., Xue, X., Ye, Z., Zhou, J., Chen, F., & Zhao, J. (2011). Determination of Chinese honey adulterated with high fructose corn syrup by near infrared. *Food Chemistry*, 128, 1110–1114.
- Conde, E., Gordon, M. H., Moure, A., & Dominguez, H. (2011). Effects of acid and bovine serum albumin in reducing the rate of development of rancidity in oil-in-water-in-oil emulsions. *Food Chemistry*, 129, 1652–1659.
- Cooper, J. B., Larkin, C. M., & Albdelkader, M. F. (2011). Virtual standard slope and bias calibration transfer of partial least squares jet fuel property models to multiple near infrared spectroscopy instruments. *Journal of Near Infrared Spectroscopy*, 19, 139–150.
- Fadda, C., Del Caro, A., Sanguinetti, A. M., Urgoghe, P. P., Vacca, V., Arca, P. P., & Piga, A. (2012). Changes during storage of quality parameters and in vitro antioxidant activity of extra virgin monovarietal oils obtained with two extraction technologies. *Food Chemistry*, 134(3), 1542–1548.
- Frankel, E. N. (1993). In search of better methods to evaluate natural antioxidants and oxidative stability in food lipids. *Trends in Food Science & Technology*, 4, 220–225.
- Gambarra-Neto, F. F., Araújo, M. C. U., Galvão, R. K. H., Pontes, M. J. C., Medeiros, E. P., & Lima, R. S. (2009). Classification of edible vegetable oils using square wave voltammetry with multivariate data analysis. *Talanta*, 77, 1660–1666.
- Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics*, 11, 137–148.
- Kim, T. S., Yeo, J. D., Kim, J. Y., Kim, M. J., & Lee, J. H. (2013). Determination of the degree of oxidation in highly-oxidised lipids using profile changes of fatty acids. *Food Chemistry*, 138, 1792–1799.
- Lerma-García, M. J., Simó-Alfonso, E. F., Bendini, A., & Cerretani, L. (2011). Rapid evaluation of oxidised fatty acid concentration in virgin olive oil using Fourier-transform infrared spectroscopy and multiple linear regression. *Food Chemistry*, 124, 679–684.
- Moreira, E. D. T., Pontes, M. J. C., Galvão, R. K. H., & Araujo, M. C. U. (2009). Near infrared reflectance spectrometry classification of cigarettes using the successive projections algorithm for variable selection. *Talanta*, 79, 1260–1264.
- Olivieri, P., Di Egidio, V., Woodcock, T., & Downey, G. (2011). Application of class-modelling techniques to near infrared data for food authentication purposes. *Food Chemistry*, 125, 1450–1456.
- Osawa, C. C., Gonçalves, L. A. G., & Ragazzi, S. (2006). Titulação potenciométrica aplicada na determinação de ácidos graxos livres de óleos e gorduras comestíveis. *Química Nova*, 29, 593–599.
- Pasquini, C., & Braz, J. (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14, 198–219.
- Pereira, A. F. C., Pontes, M. J. C., Gambarra-Neto, F. F., Santos, S. R. B., Galvão, R. K. H., & Araújo, M. C. U. (2008). NIR spectrometric determination of quality parameters in vegetable oils using PLS and variable selection. *Food Research International*, 41, 341–348.
- Pontes, M. J. C., Galvão, R. K. H., Araujo, M. C. U., Moreira, P. N. T., Pessoa Neto, O. D., Joseph, G. E., & Saldanha, T. C. B. (2005). The successive projections algorithm for spectral variable selection in classification problems. *Chemometrics and Intelligent Laboratory Systems*, 78, 11–18.
- Pontes, M. J. C., Santos, S. R. B., Araújo, M. C. U., Almeida, L. F., Lima, R. A. C., Gaião, E. N., & Souto, U. T. C. P. (2006). Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectroscopy. *Food Research International*, 39, 182–189.
- Presas-Owens, S., López-Sabater, M. C., & Rivero-Urgell, M. (1995). Shelf-life prediction of an infant formula using an accelerated stability test (Rancimat). *Journal of Agricultural and Food Chemistry*, 43, 2879–2882.
- Ramalho, V. C., & Jorge, N. (2006). Antioxidantes utilizados em óleos, gorduras e alimentos gordurosos. *Química Nova*, 29, 755–760.
- Sikorska, E., Górecki, T., Khmelinskii, I. V., Sikorski, M., & Koziol, J. (2005). Classification of edible oils using synchronous scanning fluorescence spectroscopy. *Food Chemistry*, 89, 217–225.
- Silva, F. A. M., Borges, M. F. M., & Ferreira, M. (1999). Métodos para avaliação do grau de oxidação lipídica e da capacidade antioxidante. *Química Nova*, 22, 94–103.
- Silvagni, A., Franco, L., Bagno, A., & Rastrelli, F. (2012). Thermo-induced lipid oxidation of a culinary oil: The effect of materials used in common food processing on the evolution of oxidised species. *Food Chemistry*, 133, 754–759.
- Snee, R. D. (1977). Validation of regression models: methods and examples. *Technometrics*, 19, 415–428.
- Soares, S. F. C., Galvão, R. K. H., Pontes, M. J. C., & Araújo, M. C. U. (2014). A new validation criterion for guiding the selection of variables by the successive projections algorithm in classification problems. *Journal of the Brazilian Chemical Society*, 25, 176–181.
- Veras, G., Gomes, A. A., Silva, A. C., Brito, A. L. B., Almeida, P. B. A., & Medeiros, E. P. (2010). Classification of biodiesel using NIR spectrometry and multivariate techniques. *Talanta*, 83, 565–568.
- Wang, L., Lee, F. S. C., Wang, X., & He, Y. (2006). Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR. *Food Chemistry*, 95, 529–536.
- Xiaobo, Z., Jiwen, Z., Povey, M. J. W., Holmes, M., & Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, 667, 14–32.