Journal of Agricultural Science; Vol. 10, No. 4; 2018
ISSN 1916-9752 E-ISSN 1916-9760
Published by Canadian Center of Science and Education

Separation of Cultivars of Soybeans by Chemometric Methods Using Near Infrared Spectroscopy

João S. Panero[1,6], Henrique E. B. da Silva[2], Pedro S. Panero[1,6], Oscar J. Smiderle[3], Francisco S. Panero[4,6], Fernando S. E. D. V. Faria[5,6] & Anselmo F. R. Rodriguez[5,6]

[1] Federal Institute of Science and Technology of Roraima, Boa Vista, Brazil

[2] Federal University of Rio Grande do Norte, Natal, Brazil

[3] Brazilian Agricultural Research Corporation (Embrapa), Boa Vista, Brazil

[4] Chemistry Department, Federal University of Roraima, Boa Vista, Brazil

[5] Science and Technology Innovation Program, Federal University of Acre, Rio Branco, Brazil

[6] Biodiversity and Biotechnology Network of the Legal Amazon, Biotechnology Department, Federal University of Acre, Rio Branco, Brazil

Correspondence: Anselmo F. R. Rodriguez, Biodiversity and Biotechnology Network of the Legal Amazon, Biotechnology Department, Federal University of Acre, Rio Branco, Road BR 364, Km 04, CEP 69915-900, Brazil. Tel: 55-683-901-2719. E-mail: ruiz@ufac.br

Received: January 3, 2018 Accepted: February 11, 2018 Online Published: March 15, 2018

doi:10.5539/jas.v10n4p351 URL: https://doi.org/10.5539/jas.v10n4p351

Abstract

Near Infrared (NIR) Spectroscopy technique combined with chemometrics methods were used to group and identify samples of different soy cultivars. Spectral data, collected in the range of 714 to 2500 nm (14000 to 4000 cm^{-1}), were obtained from whole grains of four different soybean cultivars and were submitted to different types of pre-treatments. Chemometrics algorithms were applied to extract relevant information from the spectral data, to remove the anomalous samples and to group the samples. The best results were obtained considering the spectral range from 1900.6 to 2187.7 nm (5261.4 cm^{-1} to 4570.9 cm^{-1}) and with spectral treatment using Multiplicative Signal Correction (MSC) + Baseline Correct (linear fit), what made it possible to the exploratory techniques Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) to separate the cultivars. Thus, the results demonstrate that NIR spectroscopy allied with de chemometrics techniques can provide a rapid, nondestructive and reliable method to distinguish different cultivars of soybeans.

Keywords: hierarchical cluster analysis, NIR spectroscopy, principal component analysis, soybean classification

1. Introduction

It is of significant value to the seed industry the development and adaptation of new technologies that improve the system of identification and evaluation of seed lot purity. Soybean is widely desired for its nutritional value. Thus, interest in a particular soybean variety may be tied to its iron or protein or oil content, for example. Identification of the soybean variety may be necessary, among other reasons, to avoid fraud, checking the purity of a lot, since the grains/seeds may be difficult to distinguish visually, with the possibility of mixing different soybean varieties, where a cultivar of high productivity and low nutritional value may be giving volume to a lot of variety of greater importance, it becomes necessary a technique to make the identification of quickly and without destruction of the sample.

Considering spectral data from the NIR region can carry information about the composition (qualitatively and quantitatively) of the samples and these compositions may vary as cultivar varies, these associations suggest the possibility of using chemometric tools to extract and relate such information.

NIR spectroscopy can detect the main structural changes related to composition, coming about as consequence of the changes in the DNA structure, since that the phenotypic changes reflect the changes on the genotypic structure (Alishahi, Farahmand, Prieto, & Cozzollino, 2010; Munck, Møller, Jacobsen, & Søndergaard, 2004).

In this context, considering that seed composition can be affected by genetic factor and that many compounds present in soybeans have one or more functional groups that absorb in the NIR region, the soybean becomes liable to be investigated by NIR spectroscopy. Thus, this study evaluates the potential of the NIR technique combined with chemometrics methods to identify and separate different soybean cultivars.

1.1 Soybean

Soybean (*Glycine max* [L.]) is the most important agricultural product in Brazil, which is the second largest producer in the world, behind only the USA. According to the United States Department of Agriculture (USDA) and National Food Supply Company (CONAB—*Companhia Nacional de Abastecimento*), the 2016/2017 soybean crop in Brazil is estimated at 114 million tons, for the USA is 117 million tons, and for Argentina, the third largest producer, is estimated 58 million tons (CONAB, 2017; USDA, 2017).

Being one of the most complete food known to man, soy is considered a functional food, provides nutrients to the body and brings health benefits. It is a source rich in proteins, in energy, and has isoflavones and polyunsaturated fatty acids that have action in reducing the risks of chronic-degenerative diseases. It is also excellent source of minerals (such as iron, potassium, phosphorus and calcium), fibers, vitamins, and other nutrients. Soybean ingestion is advised not only for people with allergic reactions to animal milk, but also recommended to prevent heart disease, obesity, hypercholesterolemia, cancer, diabetes, kidney disease, and osteoporosis (García, Marina, Laborda, & Torre, 2008; Messina, 1999; Zarkadas et al., 2007).

The three main components of soybean seeds are protein, lipid and carbohydrate. Usually, the composition in a typical cultivar is 35-40% of protein, 18-20% of oil and 12% carbohydrate (of the dry weight), besides others compounds—like phospholipids, phytic acid, isoflavones, vitamins, and saponins (Krishnan, 2001; Zarkadas et al., 2007). Soybeans contain four groups of proteins: the structural proteins (including both ribosomal and chromosomal), the enzymes involved in metabolism, the membrane proteins; and the storage proteins [predominating salt-soluble globulins, β-conglycinins (7S) and glycinins (11S)] (Krishnan, 2001; Zarkadas et al., 2007).

Studies have already concluded that seed composition can be strongly affected by genetic and environmental factors and that the genetic background of each variety controls seed composition and how much of the multi-subunit storage proteins contribute for total seed protein of soybeans (Krishnan, 2001; Nielsen et al., 1989; Zarkadas et al., 2007).

1.2 NIR Spectroscopy

The near infrared (NIR) covers the region of the electromagnetic spectrum from 12500 (or 14000, also described in the literature) to 4000 cm^{-1} (780-2500 nm) and is dominated main by overtones and combinations of O-H, N-H, C-H vibrations—in addition to others C-O, C=O, S-H, C-C, C-Cl (Burns & Ciurczak, 2007; Roggo et al., 2007; Sablinskas, Steiner, & Hof, 2003; Sun, 2009; Workman & Weyer, 2007).

NIR spectroscopy is increasingly used in process and environmental analysis, the food industry, agriculture, the pharmaceutical industry and polymer analysis. Attributes such as ease of sample handling, speed, nondestructive analysis (using common transmission and reflection techniques), possibility of routine and in-line measurements with optical probes and the combination with chemometric methods aiming at qualitative and quantitative analyses have made NIR analysis more versatile by increasing its use (Burns & Ciurczak, 2007; Ozaki, McClure, & Christy, 2007; Roggo et al., 2007; Sablinskas et al., 2003; Sun, 2009).

The NIR spectrum carries valuable information of the composition of the sample, however, this information is not easily perceptible without the help of computation, and since only one spectral data can have a table with more than 2500 columns, multiplying this by several samples, gets a large database that needs mathematical treatments in order to identify redundant, relevant and anomalous information, besides correcting scattering of signals, noise, identifying patterns, among others.

1.3 Exploratory Analysis Techniques

With the development of chemometrics, the data handling became less difficult and more fast and sophisticated, enabling that small changes in the spectral data deriving of the different absorptions to be exploited by multivariate data analysis, in addition, the NIR spectroscopy allows non-destructive analysis of raw samples (Christy, Kasemsumran, Du, & Ozaki, 2004; Ozaki et al., 2007).

Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) are exploratory analysis techniques classified as methods of recognition of unsupervised patterns used to examine similarities or differences between samples (Roggo et al., 2007). The exploratory analysis uses algorithms that reduce large and complex multivariate data sets into sets that can be better interpreted, deciphering patterns associated with independent variables (such as spectral data), and providing information and correlations between the samples and/or the independent variables.

PCA provides the best possible view of variability in the matrix of independent variables, which reveals whether there is any natural grouping of the data and whether discrepant samples exist. It may also be possible to assign chemical (or biological or physical) meaning to the data patterns that appear on the PCA chart (Arvanitoyannis, 2006; Pirouette, 2003).

During the execution of the PCA technique, linear combinations are found in the independent variables and/or in the dependent variables and the original matrix is represented by new variables, orthogonal to each other and directed according to the spatial distribution of the data. These new variables are called Principal Component or PC (Arvanitoyannis, 2006; Luna, Silva, Pinho, Ferré, & Boqué, 2013; Pirouette, 2003).

HCA aims to construct a group division, where the variables are joined together in a hierarchical way from the nearest one (which resemble the most) to the farthest, and then expressed in a dendrogram or tree.

The similarity of the samples is found considering the “distance” between them. This distance is calculated mathematically. There are several algorithms to define the proximity between samples and/or groups, such as single linkage or nearest neighbour, complete link or farthest neighbour, average link, incremental, etc. The degree of similarity ranges from 0 to 1, the closer to 1, the more identical. Thus, the similarity represents the degree of correspondence between two objects considering all the characteristics used in the analysis (Arvanitoyannis, 2006; Pirouette, 2003).

The main objective of HCA is to present data in a way that emphasizes natural groupings since the variables assembled in the same group have similar attributes (Pirouette, 2003). As observed in the work of Crupi et al. (2015), where was used PCA and HCA tools to differentiate seven grape varieties used for juice production. Considering its contents of flavonoids, it was possible to separate them into two groups, thus tracing the origin of the cultivar used in grape juice.

The separation of six extracts of vegetable tannins was also possible using HCA, in addition to separating the six extracts according to the plant origin, the HCA also separated them into two groups, hydrolysable tannins and condensed tannins. The analysis was based on spectral data in the NIR region, in the spectral range of 1200-2500 nm with first derivative of the spectra (Grasel & Ferrão, 2016).

Luna et al. (2013) applied chemometric tools in NIR spectroscopic data and demonstrated that NIR spectroscopy can provide a rapid, nondestructive and reliable method to distinguish non-transgenic and transgenic soybean oils.

Munck et al. (2004) reported the use of PCA in NIR spectral data as screening method to characterize barley endosperm mutants and concluded that genetic diversity such as from gene regulated polysaccharide and storage protein pathways in the endosperm tissue can be discovered directly from the phenotype by chemometric classification of a spectral library, where the spectra were subjected to multiplicative signal correction (MSC).

Alishahi et al. (2010) presented a review where have investigated the results of various researches involving application of NIR spectroscopy technology to identify transgenic products, highlighting the advantages in relation to the antecedent methods such as PCR and ELISA. Among the researches described, comment about the application of NIR spectroscopy to distinguish transgenic plants, transgenic corns, types of teas, coffee varieties and, allied with PCA technique, to separate transgenic foods, as soybeans, barley flour and tomatoes.

2. Materials and Methods

2.1 Sampling and Spectral Data Set

For this study, four different soybean cultivars donated by Brazilian Agricultural Research Corporation (Embrapa Roraima) were used. The soybeans were cultivated in the Experimental Field Água Boa, under the same conditions, same kind of soil, harvested in mature stage in the same year (September/October) and conditioned at 25±3 ^{o}C and approximately 75% relative humidity. The cultivars were: BRS 252 Serena, Embrapa 63 Mirador, BRS MG Nova Fronteira and Celeste. Three sets of samples of raw soybean seed of each cultivar were used (from which the sticks, pods, leaves and other interferers were removed before acquiring the spectra).

Fifteen spectral samples were acquired from each cultivar, totalizing 60 spectra, which resulted in a matrix of 60 × 2595. The soybean spectra were acquired using a BOMEM MD-160 spectrophotometer, equipped with a diffuse reflectance accessory, with signals expressed in log (1/R) and the program used was Win-Bomem Easy 3.04b. For the acquisition of the spectra, the raw soybeans seeds samples were placed in glass flasks of 30 mL with transparent and uniform walls. Under controlled temperature, each spectrum was obtained as the average of

100 scans (the flasks were spinning during the acquisition of the spectra) in the range 14 000 to 4000 cm^{-1} (714 to 2500 nm) with a resolution of 8 cm^{-1}.

2.2 Software and Chemometric Methods

The spectral matrix was assembled using the computer program OriginPro 8.5 and for the application of the chemometric methods (PCA, HCA and spectral pre-treatments) the Pirouette program was used.

As analytical signals from instruments are usually mixed with noises and with non-relevant information, making it difficult to interpret and model the properties associated with them. In addition, the response of the spectral data to the physical effects can cause significant baseline changes and the size and packaging of the particles can cause light scattering effects and higher signal intensity, it becomes necessary test preprocessing tools to correct this problems in the NIR spectra (Ozaki et al., 2007). Thus, several preprocessing methods (centering the spectral data on the average; Variance and self-scaling) and spectral pre-treatments (normalization, baseline correction, 1st and 2nd derivative, smoothing, standard normal variate (SNV) and multiplicative signal (scatter) correction (MSC)) were tested with the aim of improving the efficiency of the method. The data treatment uses mathematical processes and is applied to spectra prior to the application of PCA or HCA

3. Results and Discussion

3.1 Collected Soybean Spectra and Detailed Analysis of the Best Performing Spectral Region

Figure 1 shows the NIR spectrum, ranging of 714 to 2500 nm (14000-4000 cm^{-1}), of soybean seed. In order to obtain a better performance in soybean classification, the whole spectra and several spectral regions and combinations of these were studied. Thus, the spectrum was arbitrarily divided into eight regions (A, B, C, D, E, F, G and H) and removed the band with noise from 714 to 1105.5 nm.

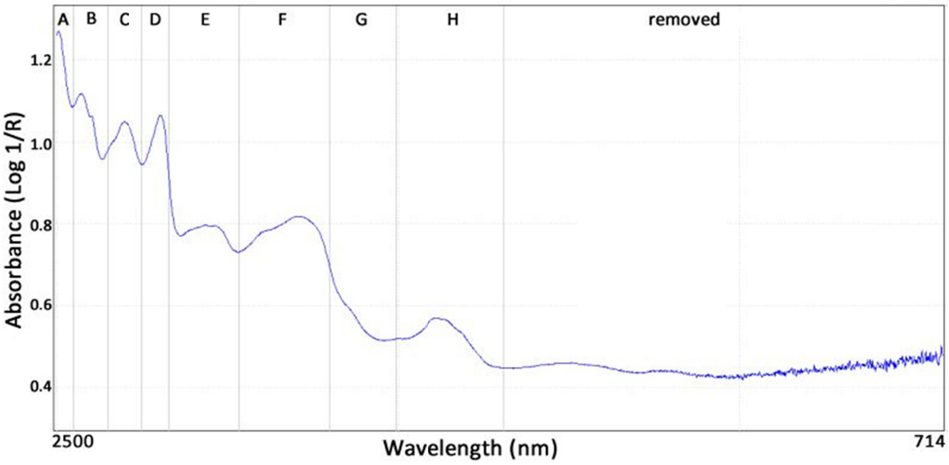

Figure 1. Near-infrared spectrum of a raw soybean seed. Letters delimit sectors of the spectrum. The wavelengths of the sectors appear in Table 1

Table 1. Wavelengths of the studied regions

Region	Spectral range
A	(2500-2389 nm)
B	(2389-2187.7 nm)
C	(2187.7-2016 nm)
D	(2016-1900.6 nm)
E	(1900.6-1652 nm)
F	(1652-1411 nm)
G	(1411-1277 nm)
H	(1277-1105.5 nm)

Table 1 shows the wavelengths of the studied regions. The original spectra of the samples are shown in Figure 2. After the removal of the spectral range of a lot of instrumental noise, eight regions and combinations of these were tested, taking into account the wavelengths where the main absorptions (of OH, NH, CH and other bonds) are characteristic of proteins, lipids, carbohydrates, water and other substances present in soybeans.

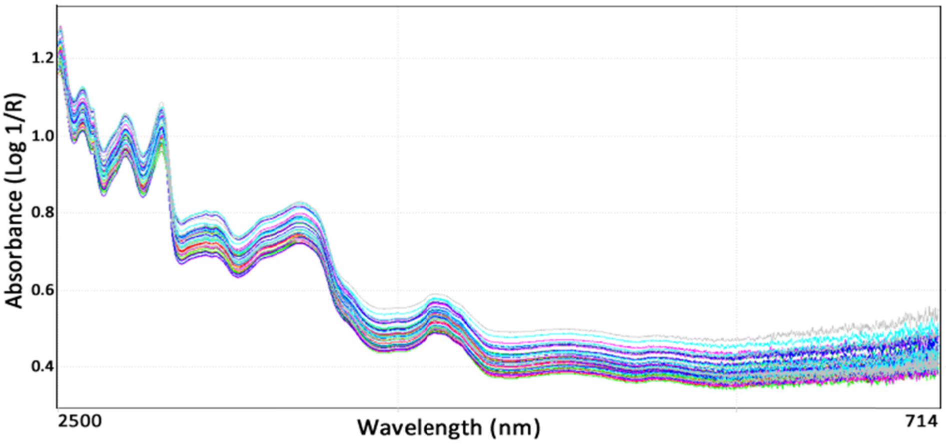

Figure 2. Sixty original spectra in the range near-infrared acquired of the samples

The main NIR band assignments for proteins, oils/lipids, water/moisture and carbohydrates collected from literature (Burns & Ciurczak, 2007; Sablinskas et al., 2003; Sun, 2009; Workman & Weyer, 2007), which justify the characteristic peaks and bands contained in the soybean spectra, the spectral areas selected for the study and support the explanation of the spectral bands responsible for the separation of soybean cultivars, are given in Table 2, Table 3, Table 4 and Table 5, respectively, being ν: stretching and δ: bending.

Table 2. Spectral bands characterizing proteins in the near-infrared region

Protein	
Wavelength (nm)	Bond vibration
1471, 1490	N-H ν amide with N-R group
1463, 1483, 1500-1530	N-H ν amide or NH_2
1510	N-H ν 1st overtone
1550	N-H ν from secondary amide
1570	N-H ν amide with N-R group
1600	N-H ν amide II (1st harmonic) comb.
1690	due to peptide N-H and C=O groups
1738	due to C=O hydrogen bonded to the N-H
1980	N-H ν (asym) and N-H in-plane δ comb.
2050, 2055	N-H comb. band + C=O amide I band
2055	Sym N-H ν amide I combination
2055	N-H ν and C=O ν (amide I) combination
2060	N-H δ 2nd overtone or N-H δ/N-H ν combination
2060	N-H δ and N-H ν (amide) combination
2060	N-H combination band from secondary amides in proteins
2180	N-H δ 2nd overtone
2180	C-H ν C=O ν combination
2180	C=O ν amide III combination
2300	C-H δ 2nd overtone
2352	CH_2 δ 2nd overtone
2470	Sym C-N-C ν 1st overtone
2530	Asym C-N-C ν 1st overtone

Table 3. Spectral bands characterizing oils/lipids in the near-infrared region

Oils/Lipids	
Wavelength (nm)	Bond vibration
1208	C-H ν (2nd harmonic)
1410	O-H 1st overtone
1416	2x C-H ν + C-H δ combination
1724, 1760	C-H ν (1st harmonic)
2070	O-H combination
2140	C-H deformation
2140	C-H ν and C=O ν combination
2144	=C-H ν + C=C ν combination
2190	CH_2 asym ν + C=C ν combination
2304	C-H ν + C-H δ of CH_2 group combination
2310	C-H δ 2nd overtone
2348	C-H sym ν of CH_2 + =CH_2 δ combination
2380	C-H ν C-C ν combination
2470	C-H combination

Table 4. Spectral bands characterizing water/moisture in the near-infrared region

Water/Moisture	
Wavelength (nm)	Bond vibration
1450	O-H ν 1st overtone
1790	O-H combination
1940	O-H δ 2nd overtone
2090	O-H combination

Table 5. Spectral bands characterizing carbohydrates in the near-infrared region

Carbohydrates	
Wavelength (nm)	Bond vibration
1450, 1490, 1540	O-H ν 1st overtone
1780	C-H ν 1st overtone
1780	C-H ν HOH deformation combination
1820	O-H ν C-O ν 2nd overtone
1930	O-H ν HOH deformation combination
1960	O-H ν O-H δ combination
2100	O-H δ C-O ν combination
2100	O-H δ and C-O ν combination
2100	Asym C-O-O ν 3rd overtone
2100	C=O-O polymeric (C=O and C-O ν) C=O-O
2200	C-H ν and C=O combination
2270, 2273	O-H ν C-O νcombination
2280	C-H νCH_2 deformation
2322, 2330, 2335	C-H ν CH_2 deformation combination.
2352	CH_2 δ 2nd overtone
2488	C-H ν C-C ν combination
2500	C-H ν C-C and C-O-C ν combination

With this, after tests, the spectral range that allowed better results was that encompassing regions C and D, which together comprise of 1900.6 to 2187.7 nm (5261.4 to 4570.9 cm^{-1}).

In this range, there are absorption bands characteristic of proteins, where the mainly bands assigned are at 2060 nm and 2180 nm (Burns & Ciurczak, 2007; O'Sullivan, O'Connor, Kelly, & McGrath, 1999; Sun, 2009; Workman & Weyer, 2007) which are characteristic absorption bands of the peptide bond, but in addition, others combinations absorptions involving NH bonds of amines/amides and C=O and C-N bonds appear of 1980 nm to 2183 nm region (Sun, 2009; Workman & Weyer, 2007).

What's more, in this selected range are also found assignment of spectral bands of lipids at about 2140 nm (Sun, 2009; Workman & Weyer, 2007) and 2190 nm (Sun, 2009) of carbohydrate at 1960nm and 2100 nm (Burns & Ciurczak, 2007; Workman & Weyer, 2007) and of water at about 1940 nm (Burns & Ciurczak, 2007; O'Sullivan et al., 1999; Sun, 2009; Workman & Weyer, 2007).

Among the types of pre-treatments tested, the one that presented the best result, making promising the application of PCA and HCA techniques, was the combination of MSC + Baseline Correct (linear fit). Figure 3 shows the spectra (C + D range) of the samples before and after the best performance spectral treatment.

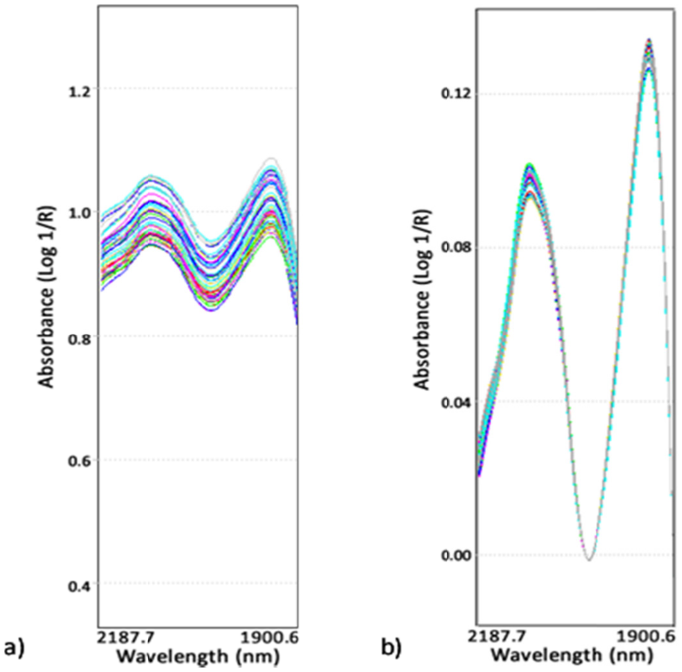

Figure 3. Spectral range with better chemometric results (regions C and D from 1900.6 to 2187.7 nm): a) without treatment and b) with treatment using Multiplicative Signal Correction + Baseline Correct (linear fit)

3.2 Identification of the Anomalous Samples by the PCA

Before the application of the separation of the soybean cultivars with the HCA and PCA techniques, three spectral samples were removed because they were discordant to the others when compared in the anomalous samples identification graph, in a previous analysis by PCA, which shows in Figure 4 the residues of the samples versus the Mahalanobis distance. Thus, the HCA and PCA analyzes were performed on a 57 × 180 data matrix, with 180 spectral signals (ranging from 4570.9 nm to 5261.4 nm) for each of the 57 corresponding spectral samples from the four soybean cultivars.

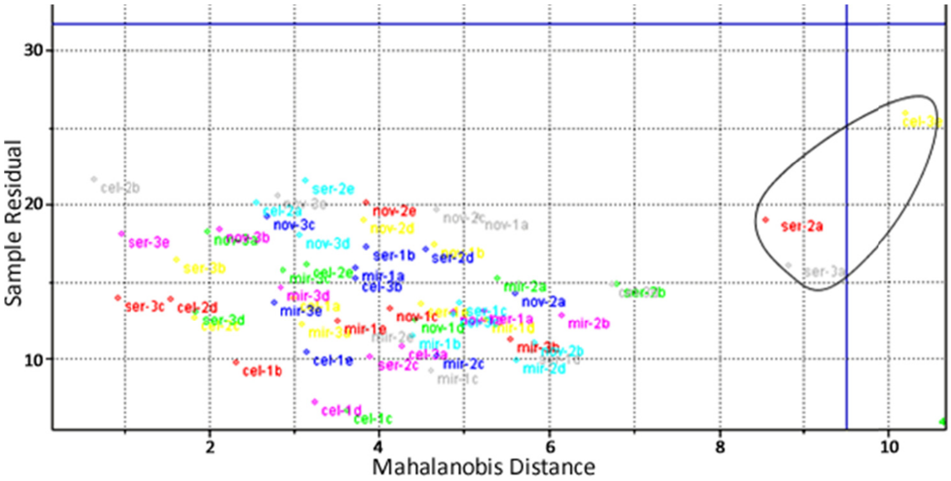

Figure 4. Identification of the anomalous samples by the PCA

3.3 Separation of Cultivars by HCA and PCA Techniques

With the application of the HCA technique in the NIR spectra of the soybean grains, it was possible to separate four cultivars. The technique was successful considering the data centered on the mean, the Euclidean metric distance and the Incremental Linkage algorithm as the clustering rule. The resulting dendrogram is presented in Figure 5, showing the separation of the four soybean cultivars, considering the spectral range of 4570.9 nm at 5261.4 nm, with better spectral treatment using the MSC + baseline correct).

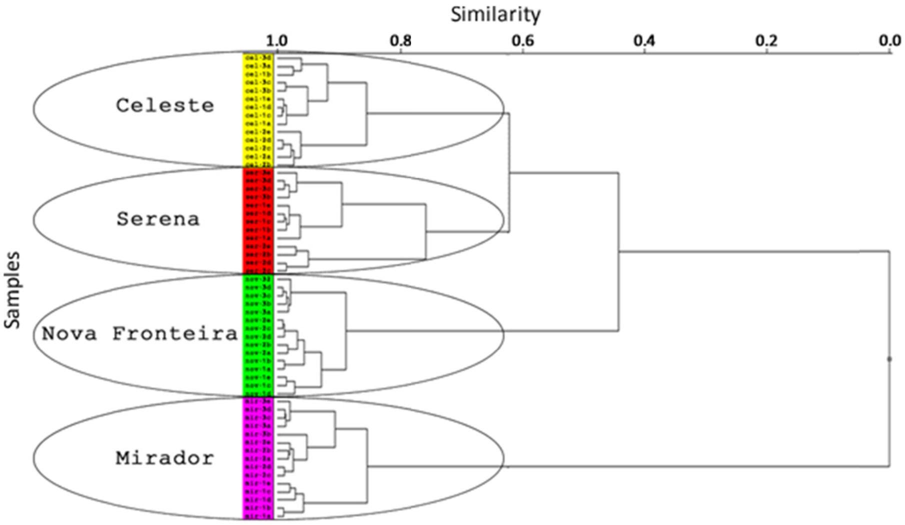

Figure 5. Separation of the four soybean cultivars by the HCA technique

The Celeste and Serena cultivars presented greater similarity among themselves, whereas the Mirador cultivar presented less similarity when compared to the others by the analysis of the HCA technique.

For the identification of groups between the samples, PCA was also applied in the spectral data, considering the spectral range from 4570.9 nm to 5261.4 nm, with better spectral treatment using MSC + baseline correct (linear fit) and the use of four PCs explaining 99.4% of the total variance, and PC1 explained 82.26%; and accumulated with PC2 explained 97.89%. As a result, the graph is shown in Figure 6, which shows the separation of four groups, which refer to the four cultivars, according to the result obtained by the HCA.

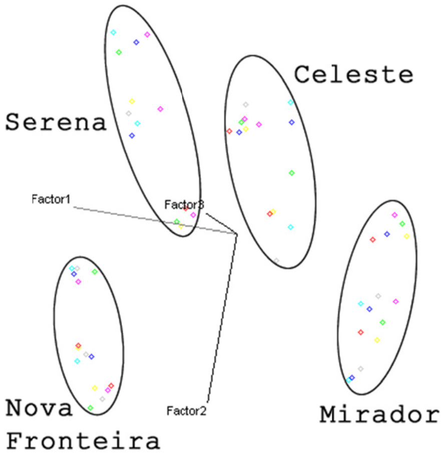

Figure 6. Separation of the four soybean cultivars by the PCA technique

The success obtained with the MSC application can be explained because this treatment has been developed to reduce the effect of scattered light on diffuse reflection and transmission NIR spectra, since the scattering should have a multiplicative effect on reflection spectra, where the observed spectra can contain a changing background from differential scattering at each wavelength (Burns & Ciurczak, 2007).

The differences "measured" by the PCA and HCA techniques in the spectral data are probably due to differences in sample composition, not only in the total content of proteins, lipids, moisture, carbohydrates and other compounds likely to absorb NIR radiation, but, possibly also of the protein and lipid profile affected by the genetic background of each cultivar.

Thus, the feasibility of combining NIR spectroscopy and chemometrics techniques to separate different soybean cultivars, originates from the assumption that NIR spectroscopy can detect the main structural changes related to composition, coming about as consequence of the changes in the DNA structure, since that the phenotypic changes reflect the changes on the genotypic structure (Alishahi et al., 2010; Munck et al., 2004).

4. Conclusion

After testing, it was possible to use the PCA and HCA techniques to separate samples from four different soybean cultivars, showing four different groups of samples, where each group represented each cultivar, using only near-infrared spectral data and studies involving selection of spectral bands and spectral pre-treatments.

The combination of the NIR spectroscopy with the chemometric techniques of exploratory analysis makes possible and suggests the application of this methodology as an alternative for comparison of soybeans samples and for possibility of the identification of different cultivars soybean, being a fast and efficient way and providing advantages such as the low manipulation and not destruction of the samples, without use reagent and waste generation.

References

Alishahi, A., Farahmand, H., Prieto, N., & Cozzollino, D. (2010). Identification of transgenic foods using NIR spectroscopy: A review. *Spectrochimica Acta Part A, 75*, 1-7. https://doi.org/10.1016/j.saa.2009.10.001

Arvanitoyannis, I. S. (2006). Multivariate Analysis. In S. S. Sablani, A. K. Datta, M. S. Rehman, & A. S. Mujumdar (Eds.), *Handbook of Food and Bioprocess Modeling Techniques* (pp. 323-356). Boca Raton: CRC Press. https://doi.org/10.1201/9781420015072.ch10

Burns, D. A., & Ciurczak, E. W. (2007). *Handbook of Near-Infrared Analysis* (3rd ed., p. 834). Boca Raton: CRC Press.

Christy, A. A., Kasemsumran, S., Du, Y., & Ozaki, Y. (2004). The detection and quantification of adulteration in olive oil by near-infrared spectroscopy and chemometrics. *Analytical Sciences, 20*, 935-940. https://doi.org/10.2116/analsci.20.935

CONAB (Companhia Nacional de Abastecimento). (2017). Acomp. safra bras. grãos, V. 5-SAFRA 2017/18-N. 2, *Segundo levantamento* (pp. 1-120). Brasília: CONAB.

Crupi, P., Bergamini, C., Perniola, R., Dipalmo, T., Clodoveo, M. L., & Antonacci, D. (2015). A chemometric approach to identify the grape cultivar employed to produce nutraceutical fruit juice. *European Food Research and Technology, 241*, 487-496. https://doi.org/10.1007/s00217-015-2478-y

García, M. C., Marina, M. L., Laborda, F., & Torre, M. (1998). Chemical characterization of commercial soybean products. *Food Chemistry, 62*(3), 325-331. https://doi.org/10.1016/S0308-8146(97)00231-8

Grasel, F. S., & Ferrão, M. F. (2016). A rapid and non-invasive method for the classification of natural tannin extracts by near infrared spectroscopy and PLS-DA. *Analytical Methods, 8*, 644-649. https://doi.org/10.1039/C5AY02526E

Krishnan, H. B. (2001). Biochemistry and Molecular Biology of Soybean Seed Storage Proteins. *Journal of New Seeds, 2*(3), 1-25. https://doi.org/10.1300/J153v02n03_01

Luna, A. S., Silva, A. P., Pinho, J. S. A., Ferré, J., & Boqué, R. (2013). Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy. *Spectrochimica Acta Part A, 100*, 115-119. https://doi.org/10.1016/j.saa.2012.02.085

Messina, M. J. (1999). Legumes and soybeans: Overview of their nutritional profiles and health effects. *American Journal of Clinical Nutrition, 70*(Suppl.), 439S-450S.

Munck, L., Møller, B., Jacobsen, S., & Søndergaard, I. (2004). Near infrared spectra indicate specific mutant endosperm genes and reveal a new mechanism for substituting starch with (1→3, 1→4)-β-glucan in barley. *Journal of Cereal Science, 40*, 213-222. https://doi.org/10.1016/j.jcs.2004.07.006

Nielsen, N. C., Dickinson, C. D., Cho, T.-J., Thanh, V. H., Scallon, B. J., Fischer, R. L., … Goldberg, R. B. (1989). Charcterization of the glycinin gene family in soybean. *The Plant Cell, 1*, 313-328. https://doi.org/10.1105/tpc.1.3.313

O'Sullivan, A., O'Connor, B., Kelly, A., & McGrath, M. J. (1999). The use of chemical and infrared methods for analysis of milk and dairy products. *International Journal of Dairy Technology, 52*, 139-148. https://doi.org/10.1111/j.1471-0307.1999.tb02856.x

Ozaki, Y., McClure, W. F., & Christy, A. A. (2007). *Near-Infrared Spectroscopy in Food Science and Technology* (p. 406). Hoboken, NJ: John Wiley & Sons, Inc.

Pirouette$^{®TM}$. (2003). *Multivariate Data Analysis, Pirouette User Guide, Software Version 3.11*. Infometrix Inc., Woodinville, Washington.

Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., & Jent, N. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis, 44*, 683-700. https://doi.org/10.1016/j.jpba.2007.03.023

Sablinskas, V., Steiner G., & Hof, M. (2003). Applications. In G. Gauglitz & T. Vo-Dinh (Eds.), *Handbook of Spectroscopy* (pp. 89-168). Weinheim: Wiley-VCH. https://doi.org/10.1002/3527602305.ch6

Sun, D.-W. (2009). *Infrared Spectroscopy for Food Quality Analysis and Control* (1st ed., p. 448). San Diego: Elsevier Academic Press.

USDA (United States Department of Agriculture). (2017) *World Agricultural Supply and Demand Estimates* (pp. 1-40). WASDE-572.

Workman, J, Jr., & Weyer, L. (2007). *Practical Guide to Interpretive Near-Infrared Spectroscopy* (p. 344). Boca Raton: CRC Press.

Zarkadas, C. G., Gagnon, C., Gleddie, S., Khanizadeh, S., Cober, E. R., & Guillemette, R. J. D. (2007). Assessment of the protein quality of fourteen soybean [*Glycine max* (L.) Merr.] cultivars using amino acid analysis and two-dimensional electrophoresis. *Food Research International, 40*, 129-146. https://doi.org/10.1016/j.foodres.2006.08.006

Copyrights