



## Analytical Methods

# A simple and practical control of the authenticity of organic sugarcane samples based on the use of machine-learning algorithms and trace elements determination by inductively coupled plasma mass spectrometry



Rommel M. Barbosa<sup>a</sup>, Bruno L. Batista<sup>b</sup>, Camila V. Barião<sup>b</sup>, Renan M. Varrique<sup>b</sup>, Vinicius A. Coelho<sup>b</sup>, Andres D. Campiglia<sup>c</sup>, Fernando Barbosa Jr.<sup>b,\*</sup>

<sup>a</sup> Instituto de Informática, Universidade Federal de Goiás, Goiânia-Go, Brazil

<sup>b</sup> Laboratório de Toxicologia e Essencialidade de Metais, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Avenida do Café s/n, Monte Alegre, 1404903 Ribeirão Preto, SP, Brazil

<sup>c</sup> Department of Chemistry, University of Central Florida, Orlando, USA

## ARTICLE INFO

## Article history:

Received 6 April 2013

Received in revised form 25 January 2015

Accepted 27 February 2015

Available online 6 March 2015

## Keywords:

Machine learning

Chemometric

Food samples

ICP-MS

Sugarcane

Trace elements

## ABSTRACT

A practical and easy control of the authenticity of organic sugarcane samples based on the use of machine-learning algorithms and trace elements determination by inductively coupled plasma mass spectrometry is proposed. Reference ranges for 32 chemical elements in 22 samples of sugarcane (13 organic and 9 non organic) were established and then two algorithms, Naive Bayes (NB) and Random Forest (RF), were evaluated to classify the samples. Accurate results (>90%) were obtained when using all variables (i.e., 32 elements). However, accuracy was improved (95.4% for NB) when only eight minerals (Rb, U, Al, Sr, Dy, Nb, Ta, Mo), chosen by a feature selection algorithm, were employed. Thus, the use of a fingerprint based on trace element levels associated with classification machine learning algorithms may be used as a simple alternative for authenticity evaluation of organic sugarcane samples.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Sugarcane is one of the most consumed food commodities worldwide. It is grown primarily in the tropics and subtropics, and Brazil is by far the world's largest producer, accounting for one third of world production. Increasing world demands for organic food products have stimulated organic sugarcane production by several producers.

“Organic” food indicates a food has been produced according to specific rules, which forbid the use of pesticides and inorganic fertilizers, and is certified by a constituted agency. Certification requires a series of steps that lead to a considerable increase of the value of the product. Efficient control of authenticity of organic food products is, however, still considered challenging since conventional and organic food products cannot be distinguished visually meaning specific analysis is mandatory. Methodologies for the authentication of organic food are a matter of great interest (Capuano, Boerrigter-Eenling, Van der Veer, Van Ruth, 2013; Kahl

et al., 2012). Considering the significant differences between organic and conventional sugarcane cultivation systems, a distinguishable pattern in mineral concentrations should also exist (Yadav, Jain, & Rai, 2010). Based on this, previously, major and trace element profiling has been used to distinguish between organic and conventionally cultivated barley, coffee, fava bean, potatoes, tomato and wheat samples (Fernandes, Tagliaferro, Azevedo, & Bode, 2002; Kelly & Bateman, 2010; Laursen et al., 2011). Previous studies have also demonstrated that nitrogen isotope composition may be used to distinguish between crops grown under conventional and organic conditions (Choi, Ro, & Lee, 2003; Kelly & Bateman, 2010). However, the time of application and the chemical form of synthetic fertilizer are important in determining how fertilizer  $\delta^{15}\text{N}$  impacts crop  $\delta^{15}\text{N}$  (Kelly & Bateman, 2010).

Trace elements can easily be determined in sugarcane samples with the use of atomic spectrometry techniques including atomic absorption spectrometry (AAS) (Segura-Muñoz et al., 2006), atomic emission spectrometry with inductively coupled plasma (ICP-OES) (Mohamed, 1999) or inductively coupled plasma mass spectrometry (ICP-MS) (Nardi et al., 2009). However, ICP-MS has numerous distinct advantages compared with AAS or ICP-OES, including

\* Corresponding author. Tel.: +55 1636024701; fax: +55 16 36024725.

E-mail address: [fbarbosa@fcrp.usp.br](mailto:fbarbosa@fcrp.usp.br) (F. Barbosa Jr.).

measurement of multiple elements coupled with very low detection limits (Parsons & Barbosa, 2007). Moreover, it offers a wider linear dynamic range which allows the determination of dozens of chemical elements in the same sample injection (Parsons & Barbosa, 2007). ICP-MS can also be considered a high throughput technique enabling large quantities of data to be generated rapidly (Kelly & Bateman, 2010; Parsons & Barbosa, 2007). However, this analytical approach also demands proper statistical treatment of data.

In the recent years, advances in chemometric techniques in quality control of food products have gained considerable attention from groups worldwide (Arvanitoyannis & Vlachos, 2009; Barbosa et al., 2014; Drivelos & Georgiou, 2012; Fabani, Raverac, & Wunderlin, 2013). In this context, several machine-learning techniques have been proposed such as Support Vector Machine (SVM), Multilayer Perceptron (MLP) and Random Forest (RF) (Aguilar et al., 2012; Alcazar, Jurado, Palacios-Morillo, de Pablos, & Martín, 2012; Batista et al., 2012; Bereton & Loyd, 2010; Jurado, Alcázar, Palacios-Morillo, & de Pablos, 2012). These data mining tools are supervised learning models with associated learning algorithms that analyze data and recognize patterns, and are used for classification and regression analysis (Koitsiantis, Zaharakis, & Pintelas, 2006). They can be used to identify the group to which a new sample belongs after a preliminary group classification. (Batista et al., 2012) Classification is made, for instance, after analyzing chemical components in a matrix and establishing a pattern (i.e., a chemical fingerprint). However, to our knowledge, these tools have not been used to control the quality and authenticity of organic food products so far.

Thus, the aim of this study was to apply the machine-learning techniques Naive Bayes (NB) and Random Forest (RF) to classify organic and conventional sugarcane samples based on a multi-mineral composition database obtained using inductively coupled plasma mass spectrometry, which might be used to control the authenticity of organic sugarcane samples.

## 2. Material and methods

### 2.1. Instruments

The determination of trace elements in sugarcane samples was carried out by using an ICP-MS (ELAN DRCII, PerkinElmer, CT, USA) with high-purity argon (99.999%, White Martins, Brazil). The instrumental parameters and optimized conditions are provided by Nardi et al. (2009).

### 2.2. Reagents

All reagents used were of analytical-reagent grade except for HNO<sub>3</sub>, which was previously purified in a quartz sub-boiling still

(Kürner Analysentechnik). High purity deionized water (resistivity 18.2 MΩ cm) obtained using a Milli-Q water purification system (Millipore, Bedford, MA, USA) was used. Multi-element (10 mg L<sup>-1</sup>) and rhodium (1000 mg L<sup>-1</sup>) solutions were obtained from PerkinElmer (Shelton, CT, USA).

### 2.3. Sampling and analytical procedures

Certified organic sugarcane ( $n = 13$ ) and non-organic sugarcane ( $n = 9$ ) samples from different brands were obtained in supermarkets from the São Paulo state (southeast region of Brazil). To avoid differences in metal levels due to geography variation between samples, it was certified that all samples were originated from sugarcane plants cultivated in the São Paulo state region. Then, five grams of each sample was put in propylene metal-free Falcon® tubes (Becton Dickinson) before analysis. The method proposed by Nardi et al. (2009) was used with some modifications to determine trace elements in sugarcane samples. Briefly, samples (0.10 g) were weighed accurately into a PFA digestion vessel, and 5 ml of nitric acid 14 mol/L + 2 mL of 30% (v/v) H<sub>2</sub>O<sub>2</sub> were added. The bomb was placed in a microwave, and decomposition carried out according to the following heating program: (a) step 1 (power 700 W, 4.5 min, 160 °C); (b) step 2 (power 0 W, 0.5 min, 160 °C); (c) step 3 (power 800 W, 5.0 min, 230 °C); (d) step 4 (power 0 W, 20 min, 35 °C) (Nardi et al., 2009).

After that, the samples were left to cool and the volume made up to 50 mL with Milli-Q water. Then, rhodium was added as internal standard to a final concentration of 10 µg/L. Using ICP-MS, 32 chemical elements (Al, Be, Bi, Ce, Co, Dy, Er, Eu, Gd, Ge, La, Mn, Mo, Nb, Nd, Ni, Pb, Pr, Rb, Se, Sm, Sr, Ta, Tb, Th, Ti, Tl, U, Y, Yb, W and Zr) were determined.

### 2.4. Analytical quality control

The quality control of data was guaranteed by analyzing NIST Standard Reference material NIST SRM 1515 apple leaves, NIST 1547 peach leaves, and NIST 1515 Wheat Flour acquired from the National Institute of Standards and Technology (NIST, USA). Reference samples were analyzed before and after sample determinations. Values were in good agreement with reference values.

### 2.5. Chemometric studies

Machine learning can be defined as methods from mathematics, computer science and statistics using data collected to make accurate predictions or classification.

In supervised learning, the learning scheme is presented with a set of classified examples from which it can classify unseen examples. Instances in a dataset are characterized by values (in our case,

**Table 1**  
Descriptive statistical analysis of the concentration of 32 elements in conventional Brazilian sugarcane samples.

Analyte	Pb <sup>α</sup>	Ni <sup>α</sup>	Mn <sup>α</sup>	Se <sup>α</sup>	Co <sup>α</sup>	Rb <sup>α</sup>	U <sup>α</sup>	Al <sup>β</sup>
Mean ± SD	16.1 ± 7.3	15.3 ± 4.5	206 ± 108	48 ± 14.9	1.6 ± 2.4	23.7 ± 11.7	4.2 ± 1.7	0.77 ± 0.36
(min–max)	8.4–36	10.2–29.1	88–511	20.1–74	0.4–8.5	9.8–77	1.0–5.9	0.42–2.0
	Be <sup>α</sup>	Bi <sup>α</sup>	Sr <sup>α</sup>	Ti <sup>α</sup>	Ce <sup>α</sup>	Dy <sup>α</sup>	Er <sup>α</sup>	Eu <sup>α</sup>
Mean ± SD	2.2 ± 1.3	8.9 ± 12.0	265 ± 84	0.19 ± 0.12	10.4 ± 4.7	1.3 ± 0.6	0.87 ± 0.32	0.50 ± 0.17
(min–max)	0.1–4.5	0.8–57.5	39–595	0.13–0.54	2.0–14.5	0.3–2.0	0.14–1.20	0.13–0.72
	Gd <sup>α</sup>	La <sup>α</sup>	Nd <sup>α</sup>	Pr <sup>α</sup>	Sm <sup>α</sup>	Tb <sup>α</sup>	Th <sup>α</sup>	Y <sup>α</sup>
Mean ± SD	2.1 ± 1.0	5.4 ± 2.8	4.7 ± 2.4	1.3 ± 0.7	1.8 ± 1.8	0.37 ± 0.07	11.5 ± 5.7	3.5 ± 1.7
(min–max)	0.3–3.5	0.9–8.3	0.6–7.5	0.3–3.2	0.1–8.1	0.13–0.40	0.9–18.5	0.5–5.5
	Yb <sup>α</sup>	Ge <sup>α</sup>	Nb <sup>α</sup>	Ta <sup>α</sup>	Ti <sup>α</sup>	W <sup>α</sup>	Zr <sup>α</sup>	Mo <sup>α</sup>
Mean ± SD	0.77 ± 0.41	0.52 ± 0.23	4.3 ± 1.8	0.58 ± 0.15	134 ± 98	1.6 ± 0.3	13.4 ± 4.6	1.8 ± 0.6
(min–max)	0.13–1.73	0.13–1.08	1.0–6.4	0.13–0.94	24.5–574	1.0–2.4	3.1–18.2	0.9–3.4

Notes: α: elements in ng g<sup>-1</sup>; β: elements in µg g<sup>-1</sup>; SD: standard deviation.

**Table 2**  
Descriptive statistical analysis of the concentration of 32 elements in organic Brazilian sugarcane samples.

Analyte	Pb <sup>α</sup>	Ni <sup>α</sup>	Mn <sup>α</sup>	Se <sup>α</sup>	Co <sup>α</sup>	Rb <sup>α</sup>	U <sup>α</sup>	Al <sup>β</sup>
Mean ± SD	16.0 ± 13.2	19.2 ± 12.0	227 ± 25	47.7 ± 12.3	1.3 ± 0.5	112 ± 26	0.42 ± 0.14	1.5 ± 0.4
(min–max)	6.1–82.3	8.7–73.2	179–288	16.7–73.5	0.5–3.4	78–168	0.14–0.78	1.0–2.8
	Be <sup>α</sup>	Bi <sup>α</sup>	Sr <sup>α</sup>	Tl <sup>α</sup>	Ce <sup>α</sup>	Dy <sup>α</sup>	Er <sup>α</sup>	Eu <sup>α</sup>
Mean ± SD	1.9 ± 1.1	10.4 ± 15.8	155 ± 27	0.24 ± 0.13	6.1 ± 3.2	0.64 ± 0.30	0.24 ± 0.10	0.16 ± 0.06
(min–max)	0.1–4.2	0.7–88.2	104–218	0.13–0.54	2.9–19.7	0.26–1.58	0.13–0.52	0.13–0.38
	Gd <sup>α</sup>	La <sup>α</sup>	Nd <sup>α</sup>	Pr <sup>α</sup>	Sm <sup>α</sup>	Tb <sup>α</sup>	Th <sup>α</sup>	Y <sup>α</sup>
Mean ± SD	1.1 ± 0.5	2.7 ± 1.7	2.9 ± 1.1	0.7 ± 0.3	0.75 ± 0.38	0.16 ± 0.08	1.6 ± 0.9	0.95 ± 0.44
(min–max)	0.4–3.0	1.0–9.2	1.3–6.2	0.3–2.0	0.26–2.02	0.13–0.52	0.7–4.6	0.40–2.36
	Yb <sup>α</sup>	Ge <sup>α</sup>	Nb <sup>α</sup>	Ta <sup>α</sup>	Ti <sup>α</sup>	W <sup>α</sup>	Zr <sup>α</sup>	Mo <sup>α</sup>
Mean ± SD	0.21 ± 0.08	0.63 ± 0.23	0.28 ± 0.13	0.14 ± 0.03	97 ± 64	2.2 ± 0.6	4.8 ± 3.8	13.9 ± 5.9
(min–max)	0.13–0.41	0.13–1.02	0.13–0.58	0.12–0.29	40–387	1.1–3.9	2.0–20.3	1.4–37

Notes: α: elements in ng g<sup>-1</sup>; β: elements in μg g<sup>-1</sup>; SD: standard deviation.

the minerals), or variables, which measure different aspects of the instance (Witten, Frank, & Hall, 2011).

Classification algorithms were applied to the dataset obtained and accuracy was analyzed. *k*-Fold cross validation was used for model building and performance. The following classification algorithms were used in this study:

1. *Naïve Bayes*: It is a classification algorithm based on applying Bayes's rule that discovers the most feasible of the potential classifications. It calculates the prior probabilities of each attribute in each class. Probabilities are assumed to be independent from one another. The classification is made by using the known probabilities of each class and the known probabilities of each attributes. That is, let {C<sub>1</sub>, C<sub>2</sub>} be the two given classes (organic sugar and conventional sugar), which have foregoing probabilities P(C<sub>1</sub>) and P(C<sub>2</sub>), respectively, and *r* attributes a<sub>1</sub>, a<sub>2</sub>, ..., a<sub>r</sub> which for an example X have values x<sub>1</sub>, x<sub>2</sub>, ..., x<sub>r</sub>, respectively, the subsequent probability of class C<sub>i</sub> occurring for the specified example is given by: P(C<sub>i</sub>) × P(a<sub>1</sub> = x<sub>1</sub> and a<sub>2</sub> = x<sub>2</sub> ... and a<sub>r</sub> = x<sub>r</sub>|C<sub>i</sub>). Supposing that the attributes are independent, the value of this expression can be achieved through the expression:

$$P(C_i) \times P(a_1 = x_1|C_i) \times P(a_2 = x_2|C_i) \times \dots \times P(a_r = x_r|C_i).$$

Thus we calculate the product above for *i* = 1, 2 and pick the classification that has the greatest value (Bramer, 2013).

2. *Random Forest*: The RF algorithm generates multiple decision trees using bootstrap samples from the original training data. Then the set of trees is used for classification of an example based on the most frequent classification among them. The RF algorithm is an example of an ensemble classifier, because each tree can be considered as an individual classifier. High levels of accuracy are achieved, usually much higher than the accuracy obtained with a single decision tree (Breiman, 2001).

Afterwards, we applied a feature selection algorithm to exclude the least important variables, and applied the same classification algorithms to this dataset with fewer variables. We used Weka to execute the machine-learning algorithms. Weka is an open source software issued under the GNU General Public License (Witten, 2011).

**Table 3**  
The results of behavior of classification algorithms according to the number variables.

Method	Accuracy	
	NB (%)	RF (%)
Using the original 32 attributes	90.9	95.4
Using the 8 minerals selected by the CFS algorithm	95.4	95.4

NB = Naïve Bayes, RF = Random Forest.

### 3. Results and discussion

#### 3.1. Trace elements in organic and conventional sugarcane samples

Tables 1 and 2 show the mean values and ranges for the elements determined in conventional and organic sugarcane samples, respectively. Considering the potentially toxic elements evaluated, aluminum presented the highest levels 0.42–2.0 μg/g and 1.0–2.8 μg/g for conventional and organic sugarcane, respectively. The levels of essential elements Mn, Se, Co and Mo, were 88–511 ng/g, 20.1–74 ng/g, 0.4–8.5 ng/g, 0.9–3.4 ng/g, respectively, in conventional sugarcane samples and 179–288 ng/g, 16.7–73.5 ng/g, 0.5–3.4 ng/g, 1.4–37 ng/g in organic samples. Our reference ranges were very similar to those previously found by Rodushkin et al. (2011) in sugarcane samples from Argentina, Costa Rica and USA.

For two thirds of the elements analyzed in the presented study, the levels were not statistically different between organic and conventional sugarcane samples. However, the levels of uranium, thorium, dysprosium, niobium, samarium, strontium, yttrium and ytterbium were statistically higher in conventional sugarcane samples compared with organic samples.

Phosphorus fertilizers utilized in conventional sugarcane crop systems contain variable amounts of thorium, uranium and rare earth elements (REEs) as contaminants from either phosphate rock ores or other ingredients used in the fertilizer industry (Abdel-Haleem, Sroor, El-Bahi, & Zohny, 2001; Otero, Vitoria, Soler, & Canals, 2005; Turra, Fernandes, & Bacchi, 2011; Turra et al., 2013). This fact could explain the higher levels of dysprosium, samarium, thorium, yttrium, ytterbium and uranium in conventional sugarcane samples compared with organic samples.

On the other hand, organic samples presented statistically higher levels of rubidium, aluminum and molybdenum. Kelly and Bateman (2010) found much higher levels of rubidium in organic tomatoes compared with conventional samples, corroborating our findings.

After establishing reference ranges for 32 metals in organic and conventional sugarcane samples, a subset of relevant features was selected (elements) for construction of the machine-learning models (reduction of variables).

#### 3.2. Reduction of variables

Feature selection is a method for selecting a subset of relevant variables from original feature set. The quantity of variables in a problem influences the computational time and the accuracy of the classification algorithm. An important notion when using a feature selection algorithm is that the data have several redundant or irrelevant variables. Redundant variables are those that provide no

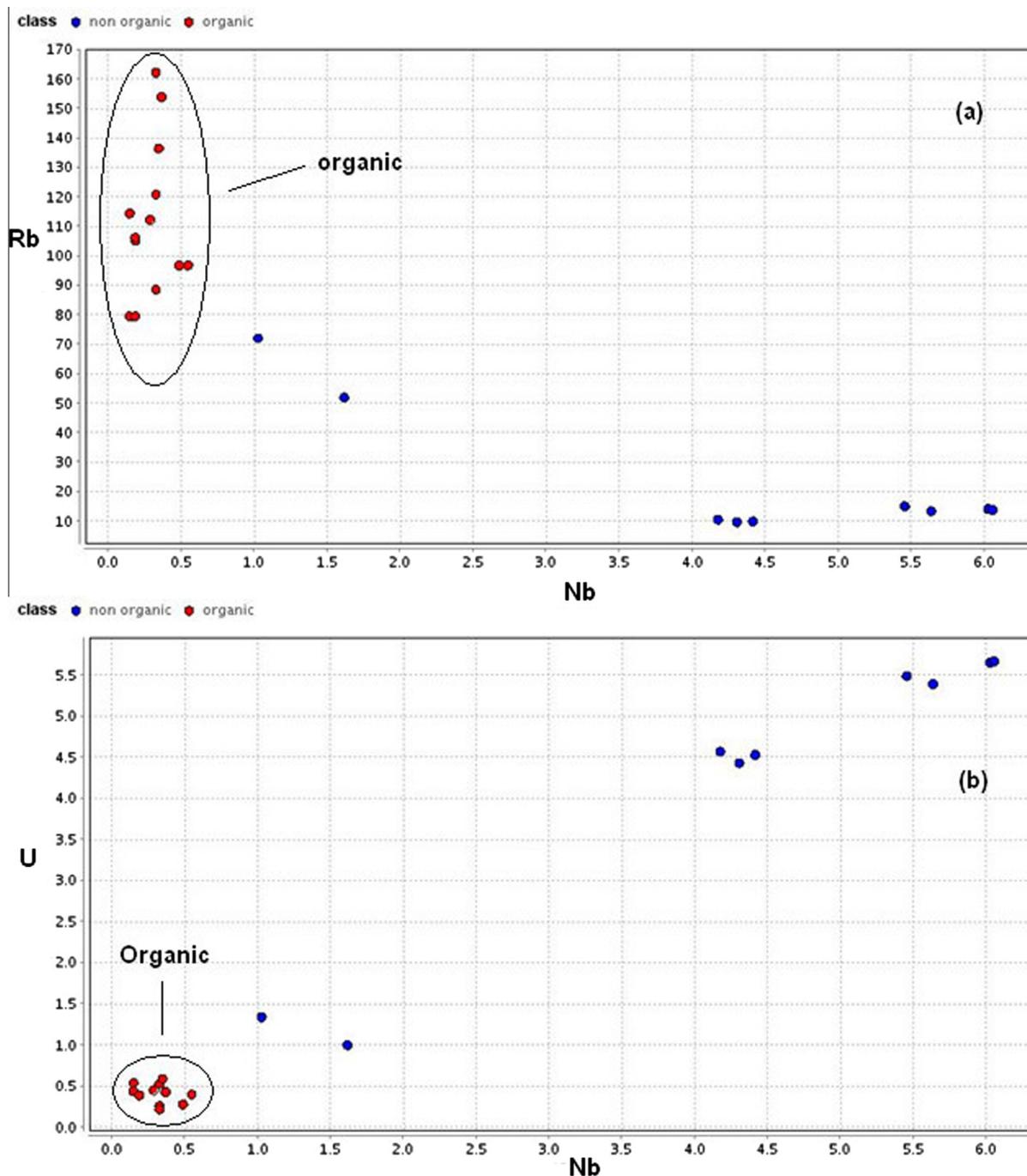


Fig. 1. Scatter plot of some variables selected by the CFS algorithm. (a) Nb  $\times$  Rb; (b) Nb  $\times$  U.

additional information than the currently selected ones, and irrelevant variables offer no useful information in any circumstance. In our study, we used the Correlation Feature Selection (CFS) Subset Algorithm for feature selection.

CFS Subset Evaluator: as in many feature selection algorithms, CFS uses a search algorithm in conjunction with a function to figure out the quality of feature subsets (Hall, 1998).

### 3.3. *k*-Fold cross validation

The holdout method for model building and accuracy evaluation of the algorithms consists of selecting one subset of data for training and another one for testing. The training phase is used

to identify sample characteristics. In the testing phase, the algorithm is checked to see whether it is able to identify patterns for the samples to be classified. The training and testing samples were selected to be as representative as possible to ensure accurate results. Here we used an approach called *k*-fold cross-validation in which a fixed number of folds of data are chosen. The data are split into *k* approximately equal partitions. Therefore, the holdout method is the simplest type of *k*-fold cross validation in which  $k = 2$ . We used  $(k - 1)$  data subsets for training and the remaining subset for testing, and this procedure was repeated *k* times so every instance was used once for testing. Since our subset of samples was small, we used the leave-one-out cross-validation (LOOCV) approach. LOOCV is a *k*-fold cross-validation, where *k* is

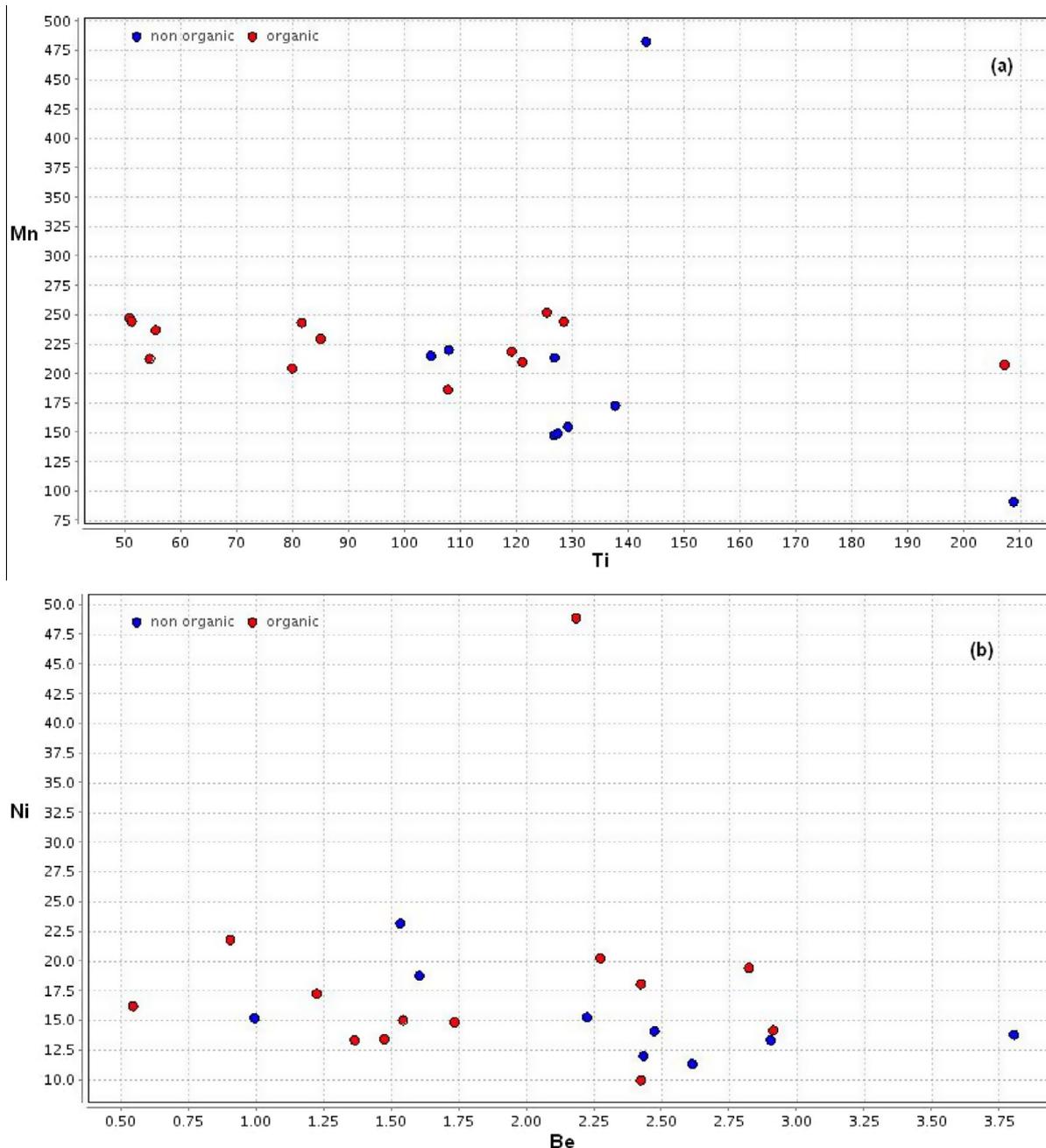


Fig. 2. Scatter plots for some variables not selected by the CFS algorithm. (a) Mn  $\times$  Ti; (b) Ni  $\times$  Be.

the number of samples in the data set. Each instance in turn is left out, and the learning scheme is trained on all the remaining samples. Since no random sampling is involved in the LOOCV, it is a deterministic method (Witten, 2011).

Two tests were performed to measure algorithm performance during training to identify patterns in the samples, as follows:

- Test 1: The original 32 variables (elements) were used;
- Test 2: 8 best ranked variables selected by CFS Subset Eval procedure were used (Rb, U, Al, Sr, Dy, Nb, Ta, Mo).

Table 3 shows the results of behavior from classification algorithms according to the number variables. Accurate results (>90%) were obtained when using all the variables (i.e., 32 elements). However, accuracy (95.4% for NB) improved with eight minerals (Rb, U, Al, Sr, Dy, Nb, Ta, Mo) chosen by a feature selection

algorithm. It can be pointed out that by analyzing separately the 8 minerals selected by the CFS algorithm, we also realized that 3 of them, Rb, U and Nb, are classifiers for the two classes (organic, non organic), see Fig. 1. The same is not true for the remaining 29 minerals, see for example Fig. 2 which are scatter plots for Mn  $\times$  Ti and Ni  $\times$  Be.

#### 4. Conclusion

We successfully applied machine learning to classify Brazilian organic and non-organic sugarcane samples. Two algorithms for classification (Naïve Bayes and Random Forest) were used. Accurate results (>90%) were obtained when using all the variables (i.e., 32 elements). However, accuracy (95.4% for NB and RF) was improved when only eight minerals (Rb, U, Al, Sr, Dy, Nb, Ta, Mo)

chosen by a feature selection algorithm was used. The use of a fingerprint, based on trace element levels in association with classification machine-learning algorithms, may be used as a simple alternative for authenticity evaluation of organic sugarcane samples.

## Acknowledgements

The authors are grateful to Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financial support. We also thank Suzanne Franks for language assistance.

## References

- Abdel-Haleem, A. S., Sroor, A., El-Bahi, S. M., & Zohny, E. (2001). Heavy metals and rare earth elements in phosphate fertilizer components using instrumental neutron activation analysis. *Applied Radiation and Isotopes*, 55(4), 569–573.
- Aguiar, G. F. M., Batista, B. L., Rodrigues, J. L., Silva, L. R. S., Campiglia, A. D., Barbosa, R. M., et al. (2012). Determination of trace elements in bovine semen samples by inductively coupled plasma mass spectrometry and data mining techniques for identification of bovine class. *Journal of Dairy Science*, 95, 7066–7073.
- Alcazar, A., Jurado, J. M., Palacios-Morillo, A., de Pablos, F., & Martín, M. J. (2012). Recognition of the geographical origin of beer based on support vector machines applied to chemical descriptors. *Food Control*, 23, 258–262.
- Arvanitoyannis, I. S., & Vlachos, A. (2009). Maize authentication: Quality control methods and multivariate analysis (chemometrics). *Critical Reviews in Food Science Nutrition*, 49, 501–537.
- Barbosa, R. M., Batista, B. L., Varrigue, R. M., Coelho, V. A., Campiglia, A. D., & Barbosa, F. (2014). The use of advanced chemometric techniques and trace element levels for controlling the authenticity of organic coffee. *Food Research International*, 2013. <http://dx.doi.org/10.1016/j.foodres.2013.07.060>. in press.
- Batista, B. L., da Silva, L. R. S., Rocha, B. A., Rodrigues, J. L., Berretta-Silva, A. A., Bonates, T. O., et al. (2012). Multi-element determination in Brazilian honey samples by inductively coupled plasma mass spectrometry and estimation of geographic origin with data mining techniques. *Food Research International*, 49(1), 209–215.
- Bereton, R. G., & Loyd, G. R. (2010). Support Vector Machines for classification and regression. *Analyst*, 135, 230–267.
- Bramer, M. (2013). *Principles of Data Mining*. Springer.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Capuano, E., Boerrigter-Eenling, Van., der Veer, G., & Van Ruth, S. M. (2013). Analytical authentication of organic products: An overview of markers. *Journal of Science and Food Agriculture*, 93, 12–28.
- Choi, W., Ro, H. M., & Lee, S. M. (2003). Natural <sup>15</sup>N abundances of inorganic nitrogen in soil treated with fertilizer and compost under changing soil moisture regimes. *Soil Biology and Biochemistry*, 35, 1289–1298.
- Drivelos, S. A., & Georgiou, C. A. (2012). Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. *TrAC Trends in Analytical Chemistry*, 40, 38–51.
- Fabani, M. P., Raverac, M. J. A., & Wunderlin, D. A. (2013). Markers of typical red wine varieties from the Valley of Tulum (San Juan-Argentina) based on VOCs profile and chemometrics. *Food Chemistry*, 141, 1055–1062.
- Fernandes, E. A. D., Tagliaferro, F. S., Azevedo, A., & Bode, P. (2002). Organic coffee discrimination with INAA and data mining; KDD techniques: New perspectives for coffee trade. *Accreditation and Quality Assurance*, 7(10), 378–387.
- Hall, M. A. (1998). Correlation-based Feature Subset Selection for Machine Learning. PhD Thesis. University of Waikato.
- Jurado, J. M., Alcázar, A., Palacios-Morillo, A., & de Pablos, F. (2012). Classification of Spanish DO white wines according to their elemental profile by means of support vector machines. *Food Chemistry*, 135, 898–903.
- Kahl, J., Baars, T., Bugel, S., Busscher, N., Huber, M., Kusche, D., et al. (2012). Organic food quality: A framework for concept, definition and evaluation from the European perspective. *Journal of the Science of Food and Agriculture*, 92, 2760–2765.
- Kelly, S. D., & Bateman, A. S. (2010). Comparison of mineral concentrations in commercially grown organic and conventional crops—Tomatoes (*Lycopersicon esculentum*) and lettuces (*Lactuca sativa*). *Food Chemistry*, 119, 738–745.
- Koitsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Reviews*, 26, 159–190.
- Laursen, K. H., Schjoerring, J. K., Olesen, J. E., Askegaard, M., Halekoh, U., & Husted, S. (2011). Multielemental fingerprinting as a tool for authentication of organic wheat, barley, faba bean, and potato. *Journal of Agriculture and Food Chemistry*, 59, 4385–4396.
- Mohamed, A. E. (1999). Environmental variations of trace element concentrations in Egyptian cane sugar and soil samples (Edfu factories). *Food Chemistry*, 65, 503–507.
- Otero, N., Vitoria, L., Soler, A., & Canals, A. (2005). Fertiliser characterisation: Major, trace and rare earth elements. *Applied Geochemistry*, 20, 1473–1488.
- Parsons, P. J., & Barbosa, F. Jr., (2007). Atomic spectrometry and trends in clinical laboratory medicine. *Spectrochimica Acta Part B*, 62, 992–1003.
- Rodushkin, I., Baxter, D. C., Engstrom, E., Hoogewerff, J., Horn, P., Papesch, W., et al. (2011). Elemental and isotopic characterization of cane and beet sugars. *Journal of Food Composition and Analysis*, 24, 70–78.
- Nardi, E. P., Evangelista, F. S., Tormen, L., SaintPierre, T. D., Curtius, A. J., de Souza, S. S., et al. (2009). The use of inductively coupled plasma mass spectrometry (ICP-MS) for the determination of toxic and essential elements in different types of food samples. *Food Chemistry*, 112, 727–732.
- Segura-Muñoz, S. I., da Silva Oliveira, A., Nikaido, M., Trevilato, T. M. B., Bocio, A., Takayanagui, A. M. M., et al. (2006). Metal levels in sugar cane (*Saccharum* spp.) samples from an area under the influence of a municipal landfill and a medical waste treatment system in Brazil. *Environment International*, 32, 52–57.
- Turra, C., Fernandes, E. A. N., Bacchi, M. A., Sarries, G. A., Barbosa, F., Reyes, A. E. L., et al. (2013). Rare earth elements in citrus production systems. *Journal of Plant Nutrition*. <http://dx.doi.org/10.1080/01904167.2012.754040>. in press.
- Turra, C., Fernandes, E. A. N., & Bacchi, M. A. (2011). Evaluation on rare earth elements of Brazilian agricultural supplies. *Journal of Environmental Chemistry and Ecotoxicology*, 3(4), 86–92.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
- Yadav, D. V., Jain, R., & Rai, R. K. (2010). Impact of heavy metals on sugarcane. *Soil Biology*, 19, 339–367.