Food Chemistry 184 (2015) 154–159

Contents lists available at ScienceDirect

Food Chemistry

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

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 [a], Bruno L. Batista [b], Camila V. Barião [b], Renan M. Varrique [b], Vinicius A. Coelho [b], Andres D. Campiglia [c], Fernando Barbosa Jr. [b,*]

[a] *Instituto de Informática, Universidade Federal de Goiás, Goiânia-Go, Brazil*
[b] *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*
[c] *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.

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 $\delta^{15}N$ impacts crop $\delta^{15}N$ (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@fcfrp.usp.br (F. Barbosa Jr.).

http://dx.doi.org/10.1016/j.foodchem.2015.02.146
0308-8146/

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 Randon Forest (RF) (Aguiar 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 multimineral 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_3, 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 ($10\ mg\ L^{-1}$) and rhodium ($1000\ mg\ L^{-1}$) 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_2O_2 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$^{\alpha}$	Ni$^{\alpha}$	Mn$^{\alpha}$	Se$^{\alpha}$	Co$^{\alpha}$	Rb$^{\alpha}$	U$^{\alpha}$	Al$^{\beta}$
Mean ± SD (min–max)	16.1 ± 7.3 8.4–36	15.3 ± 4.5 10.2–29.1	206 ± 108 88–511	48 ± 14.9 20.1–74	1.6 ± 2.4 0.4–8.5	23.7 ± 11.7 9.8–77	4.2 ± 1.7 1.0–5.9	0.77 ± 0.36 0.42–2.0
	Be$^{\alpha}$	Bi$^{\alpha}$	Sr$^{\alpha}$	Tl$^{\alpha}$	Ce$^{\alpha}$	Dy$^{\alpha}$	Er$^{\alpha}$	Eu$^{\alpha}$
Mean ± SD (min–max)	2.2 ± 1.3 0.1–4.5	8.9 ± 12.0 0.8–57.5	265 ± 84 39–595	0.19 ± 0.12 0.13–0.54	10.4 ± 4.7 2.0–14.5	1.3 ± 0.6 0.3–2.0	0.87 ± 0.32 0.14–1.20	0.50 ± 0.17 0.13–0.72
	Gd$^{\alpha}$	La$^{\alpha}$	Nd$^{\alpha}$	Pr$^{\alpha}$	Sm$^{\alpha}$	Tb$^{\alpha}$	Th$^{\alpha}$	Y$^{\alpha}$
Mean ± SD (min–max)	2.1 ± 1.0 0.3–3.5	5.4 ± 2.8 0.9–8.3	4.7 ± 2.4 0.6–7.5	1.3 ± 0.7 0.3–3.2	1.8 ± 1.8 0.1–8.1	0.37 ± 0.07 0.13–0.40	11.5 ± 5.7 0.9–18.5	3.5 ± 1.7 0.5–5.5
	Yb$^{\alpha}$	Ge$^{\alpha}$	Nb$^{\alpha}$	Ta$^{\alpha}$	Ti$^{\alpha}$	W$^{\alpha}$	Zr$^{\alpha}$	Mo$^{\alpha}$
Mean ± SD (min–max)	0.77 ± 0.41 0.13–1.73	0.52 ± 0.23 0.13–1.08	4.3 ± 1.8 1.0–6.4	0.58 ± 0.15 0.13–0.94	134 ± 98 24.5–574	1.6 ± 0.3 1.0–2.4	13.4 ± 4.6 3.1–18.2	1.8 ± 0.6 0.9–3.4

Notes: α: elements in $ng\ g^{-1}$; β: elements in $\mu g\ g^{-1}$; SD: standard deviation.