Talanta 128 (2014) 15–22

Contents lists available at ScienceDirect

Talanta

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

Geographical characterization of Spanish PDO paprika by multivariate analysis of multielemental content

Ana Palacios-Morillo, José Marcos Jurado*, Ángela Alcázar, Fernando de Pablos

Department of Analytical Chemistry, University of Seville, c/ Profesor García González 1, E-41012 Seville, Spain

ARTICLE INFO

Article history:
Received 10 October 2013
Received in revised form
11 April 2014
Accepted 16 April 2014
Available online 24 April 2014

Keywords:
Paprika
Inductively coupled plasma atomic emission spectrometry
Pattern recognition
Linear discriminant analysis
Support vector machines
Artificial neural network
Soft independent modeling of class analogy

ABSTRACT

A multielemental analytical method has been proposed to determine the contents of Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr and Zn in paprika samples from the two Protected Designations of Origin recognized in Spain, such as Murcia and La Vera (Extremadura). The samples are mineralized by acid wet digestion using a mixture of perchloric and nitric acids and analyzed by means of inductively coupled plasma atomic emission spectroscopy. The method performance has been checked studying the absence of matrix effect, trueness, precision, linearity, limit of detection and limit of quantification. The proposed method has been applied to analyze samples of sweet, hot and hot/sweet paprika from the considered production areas. Differences between paprika samples from Murcia and Extremadura were found and pattern recognition methods, such as linear discriminant analysis, linear support vector machines, soft independent modeling of class analogy and multilayer perceptrons artificial neural networks, has been used to obtain classification models. Sweet and hot/sweet paprika types were differentiated by means of linear models and hot paprika was differentiated by using artificial neural networks. A model based on artificial neural networks is proposed to differentiate the geographical origin of paprika, with independence of the type, leading to an overall classification performance of 99%.

1. Introduction

Paprika is a red powder made from grinding the dried pepper pods of some varieties of *Capsicum annuum* L. [1]. This natural food product is commonly used as spice and natural colorant in cookery and to provide redness to meat products and commercial sauces [2]. There are different types of paprika according to its pungency, named sweet, sweet/hot and hot. This variety of tastes makes this product very common in different traditional cuisines. The adequate consumption of pepper fruits is also related to potential health effects due to the presence of antioxidant compounds [3]. Although paprika is original from America, it is also produced in Europe, particularly in Hungary, Turkey and Spain. In Spain, paprika is called "pimentón" and it is cultivated and produced in two main areas namely Murcia and La Vera [4]. Paprika from both regions is recognized under Protected Designation of Origin (PDO) by the European Commission of Agriculture and Rural Development [5], being these products of great importance for the local economies. The production region of the paprika protected by La Vera PDO is located in the centre-west of Spain. This PDO includes some towns of the province of Cáceres in the region of Extremadura. La Vera PDO uses peppers belonging to the variety Bola and Ocales from the species *C. annuum* L. and *C. longum* L., respectively [6]. The other pepper growing region which is currently protected under the Murcia PDO comprises the homonymous region, located to the south east of Spain. In this case, the peppers used belong to the variety Bola from *C. annuum* L. [7].

Nowadays, consumers relate quality of foodstuff to certain characteristics influenced by the production areas, the harvesting practices and the raw materials. For this reason, it is of great importance the development of suitable methodologies allowing the characterization of different products. Within this context, several studies have been focused on the compositional profiling of paprika samples. Mateo et al. [4] identified volatile compounds influencing the overall flavor of paprika from La Vera, whilst volatile composition of paprika from Murcia has been studied by Guadayol et al. [8]. Kocsis et al. [9] researched on the volatile component composition of Hungarian red paprika as an important parameter of quality and identity. Due to the use of red paprika as a source of pigments to enhance or change food color, the study of constituents related to this characteristics, such as carotenoids, is valuable. In fact, color is used as a parameter to monitoring the quality changes during the elaboration or storage process. Techniques based on the spectrophotometric measurement of color

---

* Corresponding author. Tel.: +34 954556442; fax: +34 954557168.
E-mail address: jmjurado@us.es (J.M. Jurado).

http://dx.doi.org/10.1016/j.talanta.2014.04.025
0039-9140/© 2014 Elsevier B.V. All rights reserved.

and HPLC separation with UV–vis detection have been developed aiming this purpose [10,11]. On the other hand, antioxidant activity attributed to paprika powder is mainly related to the presence of polyphenols and several authors consider these compounds as useful variables to characterize paprika [3,11]. Metals have been also used to characterize paprika. The content of Cu, Fe, Mn, K and Na, determined by absorption and emission flame atomic spectrometry, has been used to characterize paprika samples cultivated in Venezuela [12]. Arc atomic emission spectrometry has been also used to determine B, P, Fe, Mg, Si, Mn, Al, Ca and Cu with the aim to differentiate parts of the paprika plant [13]. Multielemental profile, determined by inductively coupled plasma-mass spectrometry (ICP-MS) has been used to establish the authenticity of Hungarian paprika Szegedi Füszerpaprika (PDO) by means of pattern recognition techniques [14]. Trace elements are important parameters for establishing the PDO of foods by using appropriate chemometric data. In this realm, ICP-MS and inductively coupled plasma optical emission spectrometry (ICP-OES) are powerful analytical tools due to their low detection limits and feasibility to perform multicomponent determinations in a relatively short time. Gonzalvez et al. [15] and Jakubowski et al. [16] have reviewed the usefulness of these techniques to authenticate the origin of food matrices, including vegetable-type products.

Spanish paprika from La Vera and Murcia PDO has not been yet characterized according to their elemental composition. The metal content in paprika could be influenced by different factors such as the level of these elements in soil, fertilizing practices and the processing conditions and some differences are expected between these two PDOs. The presence of some major elements such as Ca, K, Mg, P, Al, Fe and Na in vegetable-type products is highly related to those factors. These elements are valuable chemical descriptors to perform the geographic differentiation of tea, coffee and other food products [15–18]. Consequently, it could be expected that these elements would be useful for paprika geographical differentiation. Other minor constituents, such as B, Cu, Mn, Ni, Pb, Sr or Zn, are also influenced by geographical factors and could be relevant to obtain classification models [17,18].

The main objectives of this work are the development and validation of a multielemental analytical method to determine the mineral content of paprika samples and the use of that chemical information to obtain adequate classification models to authenticate PDO Spanish paprika samples. Accordingly, an ICP-OES method has been proposed and the contents of Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr and Zn have been determined in sweet, sweet/hot and hot paprika from the two Spanish PDOs. In order to differentiate the geographical origin of the considered paprika samples, pattern recognition techniques such as principal component analysis (PCA), linear discriminant analysis (LDA), support vector machines (SVM), soft independent modeling of class analogy (SIMCA) and multilayer perceptron artificial neural networks (MLP-ANN) have been applied.

2. Materials and methods

2.1. Chemicals and reagents

Nitric (65%), sulfuric (96%), perchloric (60%) acids and hydrogen peroxide (30%) were used in the mineralization of samples. All of them were of analytical grade and obtained from Merck (Darmstadt, Germany). Standard solutions of 1000 mg l^{-1} (Merck) were used to prepare working solutions. Ultrapure water (Milli-Q, Millipore, Bedford, MA) was used throughout.

2.2. Samples

Samples of paprika belonging to two different origins: La Vera (Extremadura) ($n=72$) and Murcia ($n=72$) were obtained from local stores. Within these classes there are three different types of paprika: sweet, hot/sweet and hot, being 24 the number of samples of each pair of type-origin.

2.3. Apparatus and methods

ICP-OES analysis was performed using an ULTIMA 2 atomic emission spectrometer (Horiba Jobin Yvon, Kyoto, Japan). The ICP operating conditions are shown in the electronic Supplementary material (Table S1).

Three mixtures of acids were tested to carry out wet ashing mineralization of paprika samples: M1 (2 ml of H_2SO_4 and 15 ml of HNO_3), M2 (5 ml of H_2O_2 and 20 ml of HNO_3) and M3 (2 ml of $HClO_4$ and 20 ml of HNO_3). The acids were gradually added to 1 g of paprika (weighted with precision of 0.1 mg) placed in a beaker and the mixture was heated till complete mineralization. All obtained solutions were cooled to room temperature, filtered (0.45 μm) and transferred to 50 ml volumetric flasks. Adequate blanks were prepared for each method.

2.4. Chemometrics

A data matrix consisting of 14 columns (the determined elements) and 144 rows (the paprika samples) was created for the chemometrics calculations. Two-way nested analysis of variance (ANOVA) was used to compare recoveries of the mineralization methods. Non-parametric comparison test such as Kruskal–Wallis and Mann–Whitney tests were applied in order to draw attention to significant difference in elemental content between the paprika types and origins, respectively. PCA was used to visualize data trends and to get a first evaluation of the discriminant power of the variables. LDA, SVM, SIMCA and MLP-ANN were applied to obtain classification models. The STATISTICA 8.0 software package (StatSoft, 2007) was used for the statistical analysis.

3. Results and discussion

3.1. Methods comparison

Three mineralization methods were proposed with the aim of determining metals in paprika samples by means of ICP-OES. In order to compare among the selected methods of mineralization, and taking into account that there are no CRMs available for the trueness study, recovery assays with spiked samples was carried out [19]. A control sample was prepared by mixing paprika samples from each considered origin and type. This control sample was used to prepare the spiked samples at the expected concentrations for each element [20]. The control sample was mineralized according to the method i, fortified in j levels and analyzed in k replicates. The corresponding elemental recovery of each ijk value, R_{ijk}, was obtaided. A two-way nested ANOVA was performed for each element. The fortification level was nested in the oxidant mixture used for digestion. Variance due to the factor oxidant mixture was compared with pure error variance and Tukey-HSD *post-hoc* analysis [21] was used to detect the methods responsible of bias. The obtained results are shown in Table 1. There are significant differences for many of the determined elements for at least a pair of methods. The three methods are equivalent in the case of B, Ca, K, Mn and Ni. There are significant differences between all the possible comparisons for Sr. For the remaining elements one of the methods differs for the other two, except for Na and P, which only shows differences in the recoveries obtained by M1 and M2. In order to select the best mineralization method, the uncertainty associated to each method, $u(\overline{R}_i)$, was evaluated and their recoveries statistically compared to 100. The uncertainty

Table 1
Two-way nested ANOVA results for method comparison.

ANOVA calculations				Result of *post hoc* analysis		
Element	S^2_{method}	S^2_{rep}	F_{exp}	M1–M2	M1–M3	M3–M2
Al	27.4	5.8	4.72	■		■
B	139.7	96	1.46			
Ca	42.7	59	0.72			
Cu	162.4	8.5	19.11		■	■
Fe	1423	24.6	57.85	■	■	
K	3.6	9.4	0.38			
Mg	193.8	6.6	29.36	■	■	
Mn	108.9	74.4	1.46			
Na	558.4	5.8	96.28	■		
Ni	74.4	85.6	0.87			
P	88.3	21.7	4.07	■		
Pb	270.6	15.1	17.92		■	■
Sr	489.5	16.9	28.96	■	■	■
Zn	815.6	123.5	6.6		■	■

S^2_{method}, variance due to use of different oxidant mixtures; S^2_{pe}, pure error variance; $F(0.05,2,18)=3.55$.

Table 2
Recoveries and expanded uncertainties obtained for the studied elements for the compared mineralization methods.

Element	**Recovery, %**		
	$\overline{R}_{M1} \pm U(\overline{R}_{M1})$	$\overline{R}_{M2} \pm U(\overline{R}_{M2})$	$\overline{R}_{M3} \pm U(\overline{R}_{M3})$
Al	99 ± 4	102 ± 5	99 ± 1
B	98 ± 13	104 ± 15	97 ± 3
Ca	102 ± 15	103 ± 11	99 ± 2
Cu	91 ± 6	89 ± 8	97 ± 3
Fe	82 ± 6	106 ± 14	101 ± 3
K	100 ± 4	101 ± 10	100 ± 2
Mg	108 ± 6	100 ± 5	101 ± 2
Mn	93 ± 23	91 ± 11	98 ± 4
Na	98 ± 5	84 ± 11	98 ± 3
Ni	93 ± 17	93 ± 15	98 ± 4
P	96 ± 10	102 ± 8	100 ± 1
Pb	90 ± 8	92 ± 13	101 ± 3
Sr	92 ± 14	82 ± 9	97 ± 3
Zn	110 ± 13	114 ± 24	96 ± 4

M1, H_2SO_4/HNO_3; M2, HNO_3/H_2O_2; M3, $HNO_3/HClO_4$.

corresponding to $(\overline{R}_i)$ values is computed from two terms, combined uncertainty and pure repetition uncertainty [19]. Uncertainties (u) are converted to expanded one (U) multiplying by a coverage factor $k=2$ [22]. Recoveries can be considered free from bias when the inequality $|\overline{R}_i - 100| \leq U(\overline{R}_i)$ holds true. As can be seen in Table 2 mixture M3 leads to recoveries whose confidence interval includes the 100% for all the elements. In addition, the uncertainties of M3 are generally lower than those of M1 and M2.

Furthermore, there are some practical considerations to take into account to select the oxidant mixture to be used for paprika mineralization. The presence of high concentrations of acids during the measurement could lead to a decrease of sensitivity due to the higher viscosity of the solution. For this reason an evaporation step is needed in order to eliminate the excess of acids. The vapor pressure of sulfuric acid is very low in comparison to those for perchloric acid, nitric acid and hydrogen peroxide aqueous solutions. This means that the elimination of sulfuric acid requires higher temperatures and longer heating times. In addition the mixture of nitric and sulfuric acids may yield a precipitate of compounds like lead sulfate. Perchloric acid may react explosively with organic compounds but it can be used safely after a first action of nitric acid. When hydrogen peroxide is used, longer heating time and higher volume is needed to mineralize the samples. Therefore, from recovery assays results and practical considerations, the mixture of nitric and perchlroic acids was selected to carry out the mineralization of paprika samples.

3.2. Analytical performance

The performance characteristic of the method, such as presence of matrix effect, trueness, repeatability, intermediate precision, linearity, limit of detection (LOD) and limit of quantification (LOQ) were evaluated.

The presence of matrix effect was studied by comparing external calibration and standard addition calibration. In absence of matrix effect the ratio (R) of the slopes of external calibration (b_{EC}) and standard addition method (b_{SAM}) must be equal to 1. This can be checked statistically by means of *t*-student test [23]. The computed *t*-value was compared with the tabulated one for $\alpha=0.05$ and the effective degrees of freedom of the standard deviation of *R*, calculated according to the Welch–Satterthwaite expression [24]. According to the results obtained (Table S2) no matrix effect was found for any of the studied elements.

Trueness of the method was evaluated by means of recovery assays. As can be seen in Table 2 recoveries obtained with the selected M3 method are statistically equal to 100%, and accordingly, the trueness is assured.

Precision was evaluated for all the determined elements by analyzing in triplicate the control sample in repeatability and intermediate precision conditions. One-way ANOVA calculations were used to obtain the relative standard deviation of repeatability and intermediate precision from within-condition variance and between-condition (between-days) variance, respectively [23]. Results, expressed as relative standard deviation (RSD), are shown in Table 3. Repeatability varied from 0.72% and 10.21% and intermediate precision from 1.17% to 11.43%. These results are in accordance to those obtained by using the Horwitz function [25] depending on the analyte level.

Linearity of the calibration range was computed from the external calibration parameters as $100(1-s_{\text{bEC}}/b_{\text{EC}})$ [26]. As can be seen in Table 3 linearity for all the elements is higher than 95%.

Limit of detection (LOD) and limit of quantification (LOQ) for each element were obtained from repeated measurements of a sample blank. LOD and LOQ were calculated as the concentration

Table 3
Repeatability, intermediate precision, linearity, limit of detection and limit of quantification obtained for the analyzed elements.

Element	**RSD_{repeat} (%)**	**RSD_{ip} (%)**	**% *L***	**LOD[a] (mg kg^{-1})**	**LOQ[a] (mg kg^{-1})**
Al	0.88	1.86	98.8	0.1	0.35
B	6.37	8.38	96.78	3.2	10.5
Ca	2.44	3.50	99.1	0.1	0.35
Cu	3.07	3.97	99.0	0.45	1.5
Fe	2.25	3.89	98.87	0.05	0.15
K	2.10	2.79	98.6	2.25	7.5
Mg	2.04	3.16	95.3	0.05	0.15
Mn	8.26	11.43	97.2	0.05	0.15
Na	4.34	4.69	98.4	0.4	1.35
Ni	7.93	10.00	98.7	0.3	1
P	0.72	1.17	99.1	0.55	1.85
Pb	4.19	4.74	96.9	0.4	1.5
Sr	1.57	8.30	98.4	0.05	0.15
Zn	10.21	11.69	98.3	0.05	0.15

RSD_{repeat}, repeatability; RSD_{IP}, intermediate precision; % *L*, linearity; LOD, limit of detection; LOQ, limit of quantification.

[a] Limits were estimated considering samples of 1 g of paprika and a sample volume of 50 ml.

Table 4
Median, minimum and maximum (in brackets) metal content in Spanish paprika. Concentrations expressed in mg kg^{-1} (dry weight).

Element	MURCIA			EXTREMADURA		
	Sweet	Hot	Hot/Sweet	Sweet	Hot	Hot/Sweet
Al	140 (95–276)	136 (43–353)	42 (26–137)	132 (80–246)	117 (83–160)	128 (91–227)
B	14.2 (7.4–20.8)	12.2 (0.26–22.7)	7.9 (5.5–23.6)	14.8 (7.3–27.9)	13.5 (7.0–23.5)	10.5 (5.8–23.8)
Ca	1946 (1498–2442)	2191 (1012–2790)	1494 (1031–2174)	1895 (1063–2465)	2044 (833–2437)	1862 (1113–2422)
Cu	20.3 (9.3–22.2)	12.9 (9.1–22.6)	11.1 (9.2–22.6)	11.6 (6.0–20.0)	12.7 (7.4–20.6)	9.8 (6.5–20.7)
Fe	229 (183–403)	201 (123–418)	106 (85–264)	178 (84–295)	168 (123–236)	183 (150–247)
K	15687 (9169–20158)	14538 (9226–19765)	14387 (7691–17547)	14734 (11175–17581)	14261 (10640–23416)	15206 (10318–20213)
Mg	2012 (1758–2657)	2256 (1703–2718)	1785 (1642–2190)	2189 (1483–2461)	2174 (1867–2390)	2194 (1951–2346)
Mn	24.9 (10.7–35.8)	20.5 (12.3–53.1)	10.2 (9.1–27.7)	26.9 (19.7–97.6)	26.1 (15.3–55.6)	28.7 (19.1–68.5)
Na	509 (419–881)	441 (168–784)	546 (278–948)	328 (155–531)	333 (255–497)	377 (21–560)
Ni	0.93 (0.81–1.43)	1.06 (0.83–1.56)	0.61 (0.54–0.88)	1.00 (0.17–2.94)	1.00 (0.52–2.13)	1.02 (0.49–3.40)
P	3396 (3081–3575)	3555 (2685–4093)	3644 (3154–4226)	3613 (2499–4266)	3568 (3170–3950)	3619 (3263–4140)
Pb	0.49 (0.03–10.7)	0.40 (0.09–11.1)	0.10 (0.01–9.1)	0.40 (0.03–19.0)	1.12 (0.10–16.8)	0.42 (0.04–21.3)
Sr	29.8 (9.6–38.5)	15.1 (8.5–38.4)	12.2 (9.0–32.1)	8.7 (5.7–29.1)	9.4 (5.5–21.3)	9.2 (6.6–16.3)
Zn	43.3 (14.9–62.2)	15.1 (14.2–60.8)	30.5 (23.9–69.0)	35.7 (18.8–69.6)	32.0 (18.9–62.2)	33.4 (18.3–60.5)

corresponding to a signal 3 and 10 times the standard deviation of the blank, respectively. LODs vary in the range 0.05–3.2 mg kg^{-1}. Fe, Mg and Mn present the lowest values (0.05 mg kg^{-1}) and B and K present values of 3.2 and 2.25 mg kg^{-1}, respectively. The same pattern is obtained for LOQ, ranging from 0.15 to 10.5 mg kg^{-1}.

3.3. *Mineral content of paprika*

Paprika samples were analyzed by triplicate and their mineral content was determined. Table 4 summarizes the median value and range for samples of sweet, hot and hot/sweet paprika from the two considered origins. As can be seen, the most abundant elements are K, P, Mg and Ca, with mean contents of 14908, 3558, 2117 and 1866 mg kg^{-1}, respectively. The ranges observed for Na, Fe and Al are 155–948, 84–418 and 26–276 mg kg^{-1}, respectively. The mean contents of B, Cu, Mn, Sr and Zn range from 12.3 mg kg^{-1} for B to 36.0 mg kg^{-1} in the case of Zn. Ni and Pb present mean contents of 0.95 and 1.81 mg kg^{-1}, respectively. These contents are in accordance to those found by Brunner et al. for several European paprika samples [14]. Sweet/hot variety from Murcia presents the lowest median contents of Al (42 mg kg^{-1}), B (7.9 mg kg^{-1}), Ca (1494 mg kg^{-1}), Fe (106 mg kg^{-1}), mg (1785 mg kg^{-1}) and Mn (10.2 mg kg^{-1}). The highest median contents of Cu (20.3 mg kg^{-1}), Sr (29.8 mg kg^{-1}) and Zn (43.3 mg kg^{-1}) are found in sweet paprika from Murcia. The lowest value for this last element (15.1 mg kg^{-1}) is found for hot Murcian paprika. The three varieties of paprika from Extremadura present the lowest median concentrations in the case of Na (328–377 mg kg^{-1}) and Sr (8.7–9.4 mg kg^{-1}), whilst Murcian paprika ranges from 509 to 546 mg kg^{-1} and 12.2 to 29.8 mg kg^{-1}, respectively. The highest median content of Pb (1.12 mg kg^{-1}) is present in hot paprika from Extremadura. As some differences can be glimpsed, a comprehensive statistical analysis has been performed.

3.4. *Geographical differentiation of paprika*

The study to differentiate the geographic origin of Spanish paprika has been carried out considering the three types sweet, hot and hot/sweet. It is expectable finding out some differences among these varieties or at least, an important variability within the metal contents for each geographical origin due to this factor. Variables with high variability involve noisy information in the data set which can lead to bad performance of the classification models. For this reason, Kruskal–Wallis multiple comparison test [27] is applied to check differences between paprika types. Significant differences between hot/sweet paprika and the other two varieties are found for Al, B, Fe and Ni. Elements such as Ca, Mn, Ni and Pb show significant differences between hot and hot/sweet paprika. Sweet and hot/sweet paprika present statistical differences in the case of P. Taking into account these results, it is plausible to construct a classification model for each of the considered types of paprika. Additionally, a global model has been also built for comparison with those obtained for each paprika type.

Initially, Mann–Whitney *U*-test [27] is performed to highlight significant differences between paprika samples from the two origins. Sweet paprika samples present significant differences between their geographical origins in the case of Cu, Fe, Mn, Na, P and Sr. In the case of hot paprika differences in the contents of Al, Cu, Fe, Na, Pb and Sr were found. Concentrations of Al, Ca, Fe, Mg, Mn, Na, Ni, P, Pb and Sr show differences for hot/sweet variety. If the three varieties are considered overall, significant differences are found for Cu, Mg, Mn, Na, P, Pb and Sr. In light of these results, pattern recognition techniques are used to obtain adequate classification models.

Data trends are studied by means of PCA, computing a number of principal components (PCs) explaining as much variance as possible of the original data and reducing the dimensionality of the data matrix [28]. In the case of the global comparison, the three first PCs explain the 64.7% of the original variance. PC1 is highly correlated to the variables Al, B, Ca, Cu, Fe, Mn and Zn and PC3 is correlated to Na content. No significant correlations were found in the case of PC2. Fig. 1A shows the distribution of data in the space formed by the three first PCs. Two groups of paprika from Extremadura can be seen, one at negative values of PC1 and positive of PC2 and other in the opposite quadrant. Murcian paprika present dispersed PC2 scores and two groups are found at positive and negative values of PC1. Most of samples from Extremadura appear at positive values of PC3, whilst Murcian paprika present negative PC3 scores (Fig. 1B). When the comparison is performed considering sweet paprika, the three first PCs account for the 64.0% of total variance. In this case Al, B, Ca, Cu and Sr present high factor loadings for PC1, Fe is correlated to PC2 and no significant correlations are found for PC3. Most of samples from Extremadura appear at negative values of PC2 and positive values of PC3, whilst samples from Murcia principally present positive and negative scores for PC2 and PC3, respectively. In the case of hot paprika, 71.3% of total variance is explained with the three first PCs. Paprika samples from Extremadura are mostly distributed at positive values of PC2 while samples from Murcia do at negative values (Fig. 1C). The most contributing elements to PC1 are Al, Ca, Cu, Fe, Mn and Zn. Na and P are highly correlated to PC2 and PC3,

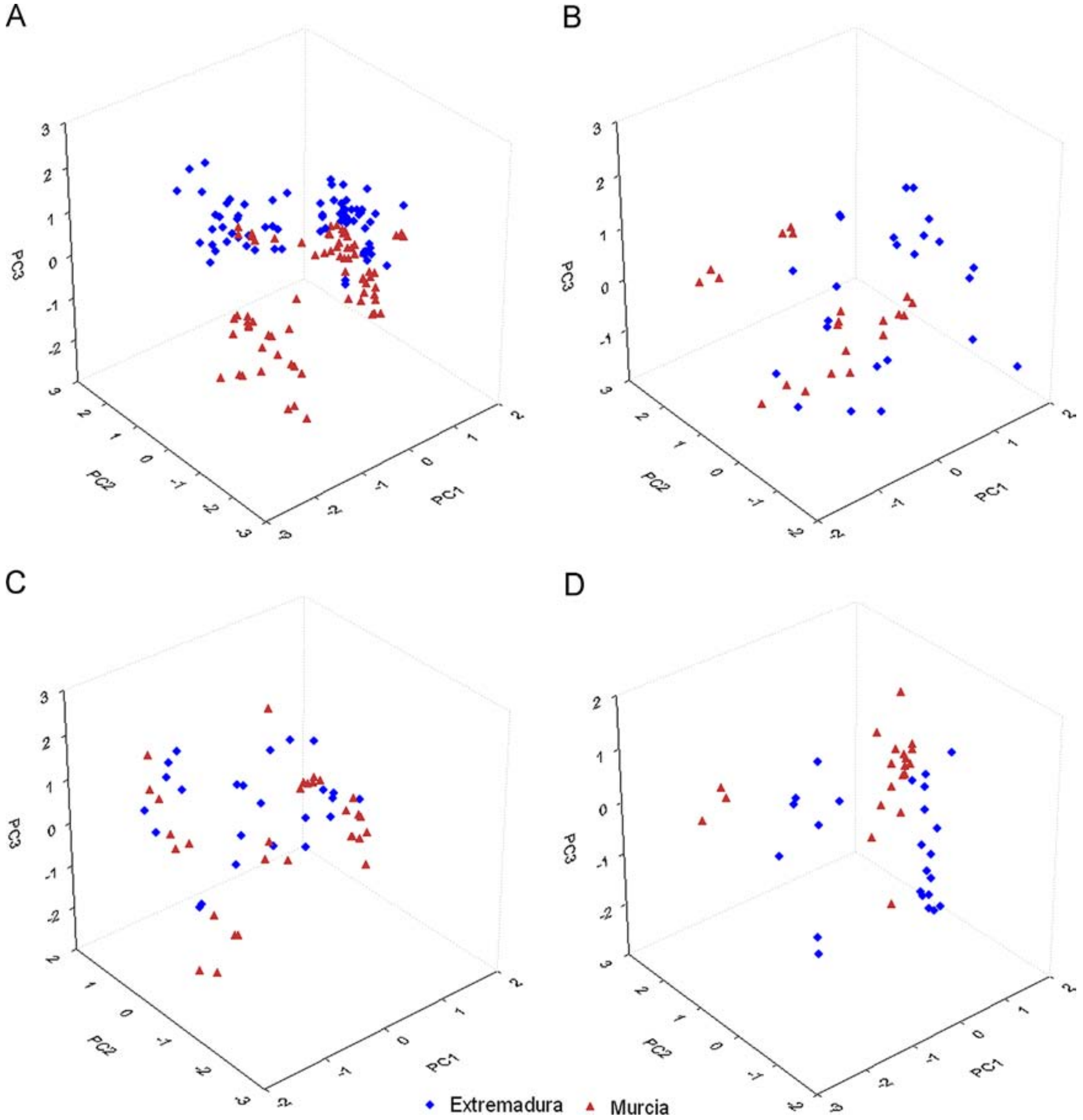

Fig. 1. Scatterplot of paprika in the space formed by the three first PCs for (A) global comparison; (B) sweet paprika; (C) hot paprika and (D) hot/sweet paprika.

respectively. When hot/sweet paprika is considered, 73.0% of original variance is explained by the three first PCs. The most contributing elements to PC1 are Al, B, Ca, Cu, Fe, Mn, P and Zn. Na and Sr contributes to PC2 and P to PC3. Most of samples from Murcia have positive scores of PC1 and PC2, being distinguishable from paprika from Extremadura, distributed at lower values of PC2 (Fig. 1D).

The models to be built in this work are numbered as 1 for global comparison, 2 for sweet paprika, 3 for hot paprika and 4 for hot/sweet paprika. Data matrix is divided in a training set (75% of the cases) to build the models and a test set (25%) to compute their classification performance. In the case of MLP-ANN models, as they are trained by back propagation, a verification set is needed and then the division is 50%, 25% and 25% for training, verification and test sets, respectively. All the proposed models were cross-validated by means of a stratified delete-a-group jackknifing (DAGJK) [29]. This procedure consists of obtaining a number of replication models (nine in this case) with the same percentage of training and test cases in each considered classes. The cases included in each set are randomly changed for each replication model. The models are built and optimized with the training set and the test set is used to obtain the classification performance. Finally a mean classification performance is obtained for each class and overall. The classification performance is evaluated by means of sensitivity (SENS) and specificity (SPEC) [30], being SENS the percentage of cases of a class correctly classified into this class and SPEC the percentage of cases not belonging to a class that are correctly classified as not pertaining to this class.

LDA computes discriminant functions as a linear combination of the original variables in order to differentiate the considered classes by minimizing the within-class and between-class ratio [31]. In this case, LDA models have been obtained by means of forward stepwise analysis, including only the most differentiating variables to compute the discriminant function. When global comparison is performed, the model (LDA1) is built with all the variables except K. In the case of sweet paprika, Sr and Zn are not used in the model (LDA2). Al, Na, Ni and P are not used in the model (LDA3) to differentiate hot paprika, whilst only Al is not used in the case of hot/sweet samples (model LDA4). As can be seen in Fig. 2A, corresponding to model LDA1, samples from Extremadura are distributed at negative values of the discriminant function whilst paprika samples from Murcia do at positive values, although some overlapping is observed. The distribution of data when the geographical origin comparison is performed by types is

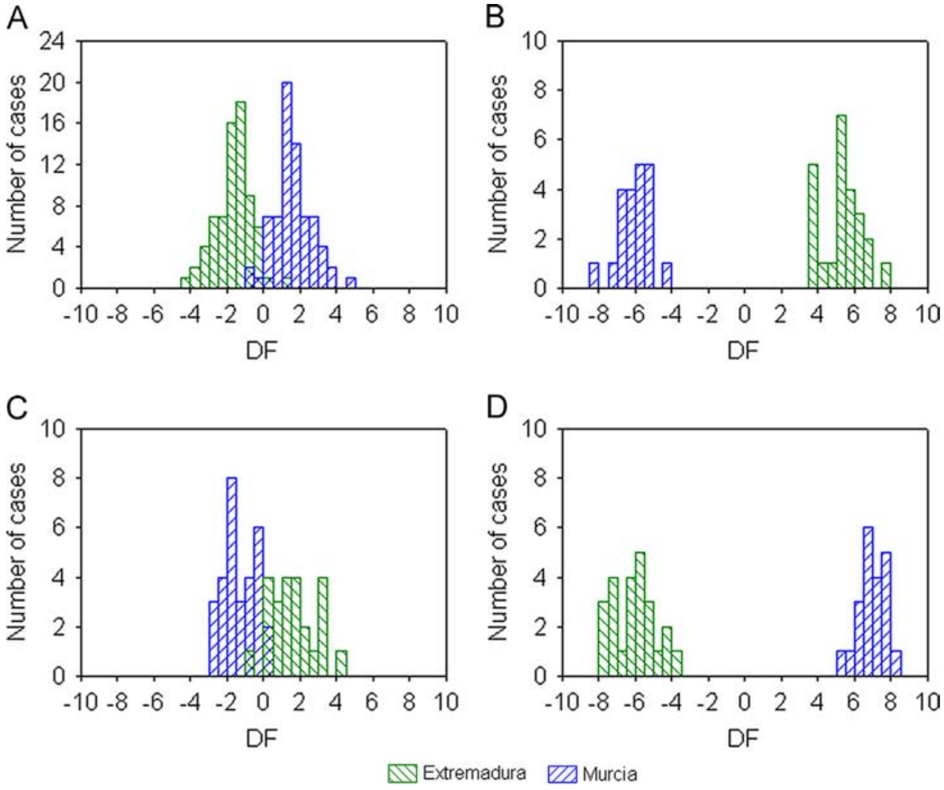

Fig. 2. Distribution of the samples according to their scores for the computed discriminant function for (A) global comparison; (B) sweet paprika; (C) hot paprika and (D) hot/sweet paprika.

Table 5
SENS and SPEC results (%) obtained by the built differentiation models for (1) global comparison, (2) sweet paprika, (3) hot paprika and (4) hot/sweet paprika.

Model	Extremadura		Murcia		Overall	
	SENS	SPEC	SENS	SPEC	SENS	SPEC
LDA1	98 ± 3	91 ± 7	91 ± 7	98 ± 3	95 ± 4	94 ± 5
LDA2	100	100	100	100	100	100
LDA3	93 ± 12	89 ± 13	89 ± 13	93 ± 12	91 ± 10	91 ± 10
LDA4	100	100	100	100	100	100
SVM1	94 ± 4	87 ± 9	87 ± 9	94 ± 4	91 ± 3	91 ± 4
SVM2	100	98 ± 7	98 ± 7	100	99 ± 3	99 ± 4
SVM3	83 ± 12	87 ± 18	87 ± 18	83 ± 12	85 ± 9	85 ± 8
SVM4	100	100	100	100	100	100
SIMCA1	84 ± 6	84 ± 6	87 ± 6	64 ± 7	86 ± 3	74 ± 5
SIMCA2	76 ± 15	100	89 ± 15	100	83 ± 10	100
SIMCA3	78 ± 20	88 ± 20	89 ± 16	35 ± 21	85 ± 11	59 ± 19
SIMCA4	91 ± 15	100	73 ± 24	100	86 ± 12	100
MLP-ANN1	99 ± 2	99 ± 4	99 ± 4	99 ± 2	99 ± 2	99 ± 2
MLP-ANN2	100	100	100	100	100	100
MLP-ANN3	96 ± 7	99 ± 4	99 ± 4	96 ± 7	98 ± 4	97 ± 4
MLP-ANN4	100	100	100	100	100	100

different in the case of sweet and hot paprika, Fig. 2B and C, respectively. It can be seen that Murcia samples are distributed at negative values of the discriminant function and samples from Extremadura do at positive values. In the case of sweet type the classes appear well differentiated, but hot paprika samples overlaps. Hot/sweet paprika samples appear separately, being cases from Murcia distributed at positive values of DF and cases from Extremadura at negative values. The results of the models are shown in Table 5, sweet and hot/sweet paprika present SENS and SPEC of 100%, whilst hot paprika perform with overall SENS an SPEC of 91%. Samples corresponding to this type seem to be responsible of the results obtained for the global LDA model, with SENS of 95% and SPEC of 94%. These results could be improved by the use of other pattern recognition techniques and in this way SVM, SIMCA and ANN are applied.

SVMs compute an optimal separation hyperplane by means of an iterative algorithm learning the sample distribution in the boundaries of each considered class. The complexity of the model is controlled by a penalty error function in order to avoid overtraining [32]. In this study, the SVM models are obtained by using the same variables selected by previous LDA and the results can be seen in Table 5. The results are similar to those obtained by LDA in the case hot/sweet paprika. SENS obtained for Murcia sweet paprika is slightly lower when SVMs are used instead of LDA. The same occurs in the case of hot paprika, but also SENS for samples from Extremadura decays to 83%. For this reason, results of SVM in the case of global comparison do not improve those obtained by LDA.

SIMCA performs a PCA analysis for each considered class determining the number of PCs needed to describe the structure of each class. The boundaries of each class are computed and objects are classified in a group if they fall into the n-dimensional boxes limited by these boundaries. SIMCA is a soft modeling technique, so that samples can be included or not included into a class and also can fall in the confluence space of the classes [33]. The results of this technique can be projected in a Cooman's plot (Fig. 3). For global comparison and hot paprika there are a number of samples from Murcia and Extremadura in the confluence space of both classes (bottom left corner), what decreases SPEC for both groups (Table 5). In the case of sweet and hot/sweet paprika there is no any sample assigned to both classes at the same time and SPEC results are comparable to those obtained by LDA and SVM. Cases not pertaining to any of the classes lead to lower SENS values.

Linear models, such as LDA, linear SVM and SIMCA, do not solve the classification problem and therefore MLP-ANNs are applied. MLP-ANN are feedforward multilayer networks consisting on

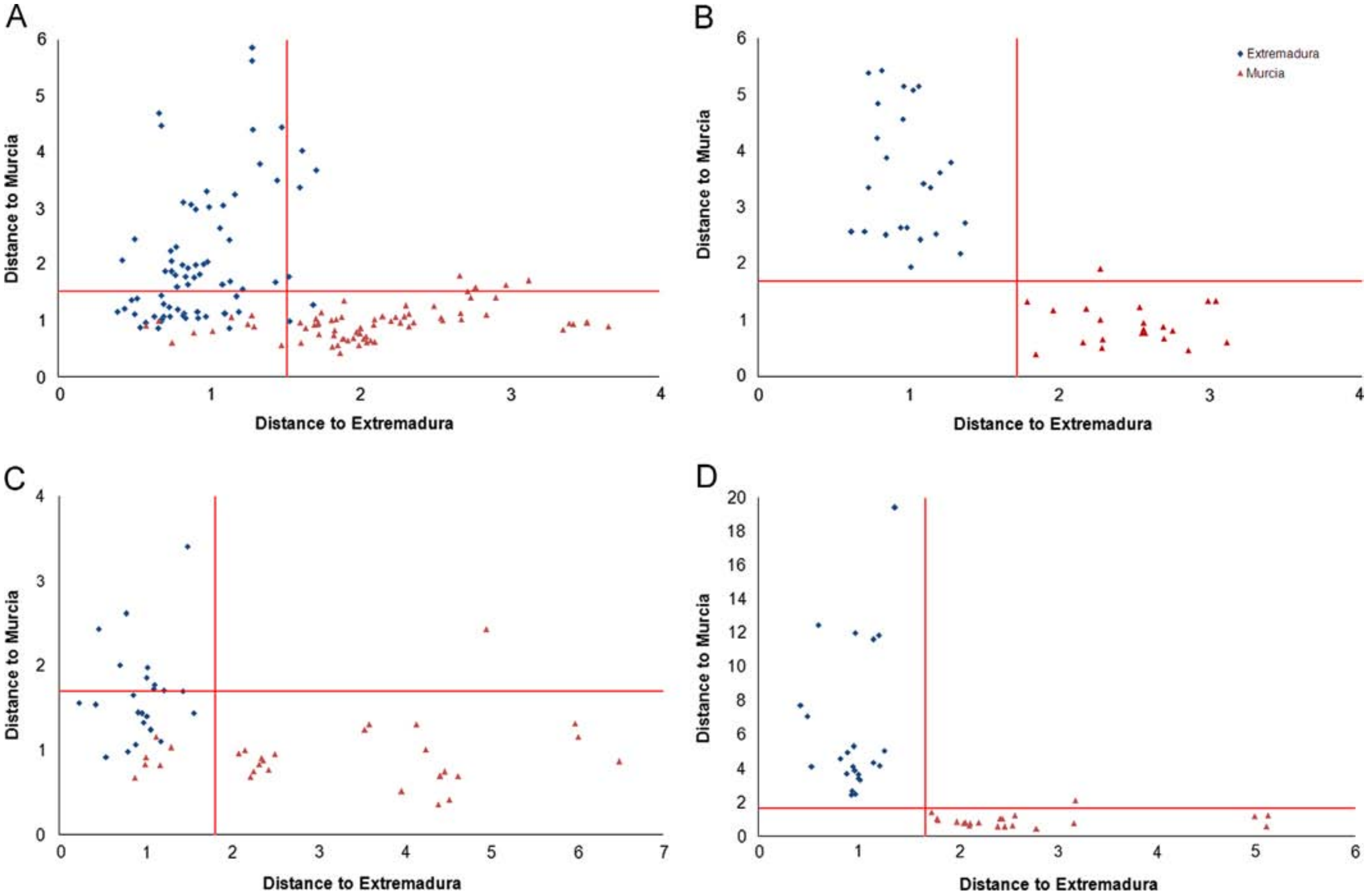

Fig. 3. Cooman's plots obtained from SIMCA models for (A) global comparison; (B) sweet paprika; (C) hot paprika and (D) hot/sweet paprika.

neurons arranged in layers. An input layer read the values of the variables used in the model. In a hidden layer, each neuron performs a weighted sum of their inputs and transforms it with an activation function to produce their output [34]. A output layer computes the probability of pertaining at each class. In this case, the variables used for each comparison are the same selected by LDA and logistic functions are used as activation function in the hidden layer. In the case of global comparison the model is built with architecture 13:11:2. This model performs better than linear ones with SENS and SPEC of 99% for both classes. For sweet paprika 100% of samples appear correctly assigned to their corresponding class using a 12:7:2 structure. Hot paprika is classified with overall SENS of 98% and SPEC of 97%. In this case the network architecture presents 10 neurons in the input layer, 7 in the hidden layer and 2 in the output layer. When the comparison is carried out with only hot/sweet paprika, 100% of performance efficiency is obtained with a 13:9:2 network. In light of such results, MLP-ANN models improve the results of linear models and can be used to differentiate between paprika samples from the two considered PDO, with independence of their pungency.

4. Conclusions

A method to perform the determination of Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr and Zn in paprika samples has been proposed. Different mixtures of acids to perform the mineralization have been tested, being the mixture of perchloric and nitric acids which leads to adequate recoveries for all the considered elements and also presents practical advantages. The method performance has been checked considering the absence of matrix effect, trueness, repeatability, intermediate precision, linearity and limits of detection and quantification. The obtained results show the adequacy of the method to carry out the determination of the studied elements in paprika samples in their expected levels of concentration. ICP-OES presents the advantage of being a multi-elemental determination technique, with relatively short time of analysis, leading to an easy implementation of the proposed methodology in quality control laboratories.

The contents of Al, B, Ca, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr and Zn has been determined in paprika samples from the two Spanish PDO, Extremadura and Murcia. Using the contents of these elements and applying LDA, SVM, SIMCA and MLP-ANN classifications models have been obtained to differentiate the two Spanish PDO of paprika. A classification efficiency of 99 ± 2% has been achieved with the MLP-ANN model. The results obtained prove the usefulness of the metallic profile to differentiate PDOs of paprika and the applicability of this methodology in the quality control of this product.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.talanta.2014.04.025.

References

[1] G. Reineccius, Sourcebook of Flavors, second ed., Aspen Publisher, Gaitherburg, USA, 1994.
[2] M. Attokaran, Natural Food Flavors and Colorants, Wiley-Backwell, Ames, USA, 2011.
[3] J.M. Navarro, P. Flores, C. Garrido, V. Martinez, Food Chem. 96 (2006) 66–73.

[4] J. Mateo, M. Aguirrezábal, C. Dominguez, J.M. Zumalacárregui, J. Food Compos. Anal. 10 (1997) 225–232.

[5] European Commission, Agricultural and Rural Development, European Union Agricultural Product Quality Policy, Specific EU Quality Schemes E Guaranteeing Quality, ⟨http://ec.europa.eu/agriculture/quality/⟩.

[6] Official Journal of the European Union, Information and Notices, 2006/C 287/02, 24.11.2006.

[7] Official Journal of the European Communities, Information and Notices, 2000/C 173/06, 22.6.2000.

[8] J.M. Guadayol, J. Caixach, J. Ribé, J. Cabañas, J. Rivera, J. Agric. Food Chem. 45 (1997) 1868–1872.

[9] N. Kocsis, M. Amtmann, Z. Mednyánszky, K. Korány, J. Food Compos. Anal. 15 (2002) 195–203.

[10] D. Hornero-Méndez, M.I. Mínguez-Mosquera, J. Agric. Food Chem. 49 (2001) 3584–3588.

[11] J. Park, S. Kim, B. Moon, J. Food Sci. 76 (2011) C1075–C1080.

[12] A.C. Vasquez, A.M. Ocando, J.C. Torres, M.C. Rodriguez, V.A. Granadillo, Ciencia 8 (2000) 85–92.

[13] J. Borszeki, J. Kepes, J. Inczedy, L. Koltay, E. Gegus, Z. Lebensm. Unters. Forsch. 180 (1985) 292–296.

[14] M. Brunner, R. Katona, Z. Stefánka, T. Prohaska, Eur. Food Res. Technol. 231 (2010) 623–634.

[15] A. Gonzalvez, S. Armenta, M. de la Guardia, Trends Anal. Chem. 28 (2009) 1295–1311.

[16] N. Jakubowski, T. Prohaska, F. Vanhaecke, P.H. Roos, T. Lindemann, J. Anal. At. Spectrom. 26 (2011) 727–757.

[17] R. Muñiz-Valencia, J.M. Jurado., S.G. Ceballos-Magaña, A. Alcázar, J. Reyes, Food Anal. Methods 103 (2013) 79–83.

[18] J. McKenzie, J.M. Jurado, F. Pablos, Food Chem. 123 (2010) 2914–2923.

[19] J.M. Jurado, A. Alcazar, F. Pablos, M.J. Martín, A.G. Gonzalez, Talanta 63 (2004) 297–302.

[20] A.G. Gonzalez, M.A. Herrador, A.G. Asuero, Talanta 65 (2005) 1022–1030.

[21] J.A. Rafter, M.L. Abell, J.P. Braselton, SIAM Rev. 44 (2002) 259–278.

[22] EURACHEM/CITAC Guide, Quantifying Uncertainty in Analytical Measurement, second ed., 2000.

[23] A.G. González, M.A. Herrador, Trends Anal. Chem. 26 (2007) 227–238.

[24] B.L. Welch, Biometrika 34 (1947) 28–35.

[25] W. Horwitz, Anal. Chem. 54 (1982) 67A–76A.

[26] L. Cuadros, A.M. García, J.M. Bosque, Anal. Lett. 29 (1996) 1231–1239.

[27] J. Muth, Basic Statistic and Pharmaceutical Statistical Applications, Marcel Dekker, NY, USA, 1999.

[28] I.T. Jolliffe, Principal Components Analysis, second ed., Springer, New York, 2002.

[29] P.S. Kott, J. Off. Stat. 17 (2001) 521–526.

[30] M. Forina, C. Armanino, R. Leardi, G. Drava, J. Chemometr. 5 (1991) 435–453.

[31] W.P. Gardiner, Statistical Analysis Methods for Chemists, Royal Society of Chemistry, Cambridge, 1997.

[32] R.G. Brereton, G.R. Lloyd, Analyst 135 (2010) 230–267.

[33] D.L. Massart, Handbook of Chemometrics and Qualimetrics—Part B, Elsevier, Amsterdam, 1998.

[34] W.S. Sarle, Neural networks and statistical models, in:Proceedings of the Nineteenth Annual SAAS Users Group International Conference, 1994.