JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

Chemometric Classification of Potatoes with Protected Designation of Origin According to Their Producing Area and Variety

Carlos Herrero Latorre,* Julia Barciela García, Sagrario García Martín, and Rosa M. Peña Crecente

Departamento Química Analítica, Nutrición y Bromatología, Facultad de Ciencias, Universidad de Santiago de Compostela, Alfonso X el Sabio s/n, 27002 Lugo, Spain

ABSTRACT: Potatoes from Galicia (northwestern Spain) are subjected to a Protected Geographic Indication (PGI) according to European legislation. Ten trace elements (Li, Na, K, Rb, Ca, Fe, Mg, Mn, Cu, and Zn) have been determined by atomic spectrometry in two sets of potato samples: Geo-Origin.set and Variety.set. The first data set is composed of samples of the only variety authorized by PGI (Kennebec) with two geographical origins: Galician and non-Galician. The second set corresponds to samples from different varieties but with only Galician geographical origin. Chemometric pattern recognition techniques have been applied to the study of potato geographical and varietal origins in relation to their capability for translocating metals from soil to tuber. Also, authentication models for classifying potato samples with Galician PGI based on metal fingerprints have been developed. The results obtained showed that samples of the same variety, Kennebec, have different metal fingerprints when they have been produced in different geographic locations. Also, diverse potato varieties cultivated on equal geographic Galician origin presented different metal profiles as well. Therefore, it can be concluded that classification studies on the differentiation of geographical origin of foods should take into account information of production area together with varietal data. Otherwise, classification obtained on the basis of the geographical origin could be due to the different variety or vice versa. Finally, two models were constructed for Kennebec Galician samples against Kennebec from other origins as well as against other varieties cultivated in Galicia (Liseta and Baraka). Both models achieved adequate classification rates (93−100%), good sensitivities, and total specificities (100%), allowing the fraud detection in the PGI label.

KEYWORDS: *potatoes, metal fingerprint, geographical and varietal origin, chemometric techniques*

INTRODUCTION

Authentication of geographical and varietal origins of commodities and foods constitutes, in recent years, a very active area in research.[1] European Union (EU) regulations[2] established the Protected Denomination of Origin (PDO), Protected Geographical Indication (PGI), and Traditional Specialty Guaranteed (TSG) to encourage diverse agricultural production, to protect product names from misuse and imitation or falsification, and to protect consumers. PDO identifies quality agricultural products and foodstuffs that are produced, processed, and packed in a specific geographical area following traditional methods. PGI covers agricultural products and foodstuffs closely linked to the geographical area, which means that at least one of the stages of production, processing, or preparation must take place in the geographical area of PGI. TSG is a label for highlighting the traditional character of a food product (both in compositions and in means of production). Therefore, these labels help to ensure the quality and origin of food, facilitate the preservation of the product's image, contribute to guaranteeing its quality (avoiding adulterations and falsifications), aid in preventing economic injuries for producers, and protect the consumers from overpayment. Recently, the EU has promulgated the regulation on quality schemes for agricultural products and foodstuffs,[3] in which PDO and PGI must assist producers of products linked to a geographical area by (a) securing fair returns for the quality of their products, (b) ensuring uniform protection of the names as an intellectual property right in the territory of the Union, and (c) providing clear information on the value-adding attributes of the product to consumers.

Galicia is an autonomous region located in northwestern Spain that is well-known for its quality food productions (including wine, alcoholic distillates, honey, meet, cheese, and potatoes), which have a fundamental importance in the economy of the territory. Potato cultivation is essential for the agricultural activity in Galicia because it achieves a total cultivated area of 20,000 ha, with an annual production around 400,000 tonnes, according to the regional government's data.[4] Therefore, following the European legislation, the EU recognized this product with a PGI "*Pataca de Galicia*"[5] and, consequently, the political authorities and the PGI certification council established the criteria that must be met in terms of quality, geographical origin, process, and label.[6] Potatoes must be of the sole variety authorized (Kennebec); furthermore, the product must be cultivated only in controlled Galician geographical areas following the traditional practices described in the normative of PGI, including fertilization, irrigation procedure, and harvesting time. Finally, harvested potatoes are subjected to several physicochemical analyses to check the compliance with the required characteristics. PGI regulations also indicated the packaging system and the storage conditions to ensure product quality. The verification of compliance with the product specifications indicated by a PGI corresponds to the public authorities designated by the country and/or the product

Received: May 7, 2013
Revised: July 31, 2013
Accepted: August 2, 2013
Published: August 2, 2013

ACS Publications © 2013 American Chemical Society

certification bodies. Therefore, modern systems based on nonsubjective parameters to determine the authenticity of a product are widely demanded by sectors involved in production, processing, and packaging of the different foods as well as for regulatory bodies and consumers. The most frequently employed approach for such tasks is the determination of the compositional profile of the product by means of appropriate analytical techniques and the subsequent modelization of this chemical information using statistical multivariate chemometric procedures to develop classification systems to assign samples according to their belonging or not to a particular class (i.e., PDO, PGI, ...).[7] Using this strategy, successful authentication was achieved for numerous food products such as wine,[8,9] alcoholic distillate and liquors,[10,11] honey,[12,13] coffee,[14,15] oil,[16,17] vinegar,[18,19] meat,[20] and others, to guarantee their quality, genuineness, and geographical origin. Different chemical and spectral fingerprints have been employed as information data (vitamins, amino acids, aroma compounds, phenols, elements, near-infrared spectra, magnetic nuclear resonance spectra, among others), and also diverse mathematical methods have been used to develop models (linear discriminant analysis, *K*-nearest neighbors, soft independent modeling of class analogy, partial least-squares discriminant analysis, and several types of neural networks, among others). However, in many papers, the metal profile of the samples was selected as the chemical information employed. This can be explained because (i) the metal composition of food products reflects the characteristics of the soil and the environment in which the foodstuff has been produced, (ii) mineral and trace element composition is stable and not influenced by storage conditions, and (iii) the determination of major and trace levels of metal contents in food is important for both food safety and nutritional information. In addition, it can be stated that the metal profile of food could be in certain cases an adequate indicator of environmental pollution. Therefore, in many cases, the mineral composition of food has been considered an appropriate primary fingerprint for developing authentication and classification procedures. Recently, two excellent reviews on the use of trace element composition and stable isotope ratio for the geographical origin determination of food with PDO or PGI in the EU demonstrated the validity and usability of this line of attack.[21,22]

In the particular case of potatoes, despite the fact that other data such as isoelectrophoretic focusing patterns,[23] volatile compounds,[24] and physicochemical composition[25] have been used for classification, metal and elemental compositions are the dominant ones used as chemical information for authentication purposes. From the year 1999, in which Anderson et al.[26] developed a method to confirm the geographical authenticity of Idaho-labeled potatoes based on a set of 18 metal and trace elements using neural network classifiers, different authors have published characterization and geographical or varietal classification studies on potatoes from diverse origins. Casañas Rivero et al.[27] studied potatoes cultivated in Canary Islands (Spain) on the basis of the content of nine metals (Na, K, Rb, Ca, Mg, Fe, Cu, Zn, and Mn). Using factor analysis, an appropriate separation of samples according to their location and irrigation procedure was achieved; also, traditional varieties were well distinguished from imported potatoes. Italian potatoes of the Fucino origin were successfully discriminated from ones produced in the region of Abruzzo and in other regions of Italy.[28] The samples were characterized by 10 mineral and trace elements (Mg, Cr, Mn, Fe, Ni, Cu, Zn, Sr, Cd, and Ba), and the classification models were made employing linear discriminant analysis. Adamo et al.[29] developed a partial least-squares discriminant model to authenticate potatoes from three different Italian regions (Apulia, Campania, and Sicily), but in this case, in addition to the elemental concentrations of Mn, Cu, Zn, Rb, Sr, and Ca, other biological data of 11 secondary DNA metabolites were also considered. In relation to Galician potatoes, in previous works,[30,31] we demonstrated the possibility of identifying potatoes from Galicia with Certified Brand of Origin and Quality (CBOQ) on the basis of a set of 10 metal and trace elements using diverse chemometric tools, comparing classical pattern recognition techniques with new neural network approaches. As indicated above, different studies are directed to the geographical and varietal origins of the potatoes. However, in most cases, due to the cultivar characteristics and also the sample availability, potatoes of different varieties and from diverse geographical origins were taken into account for this task. This fact makes it difficult to appropriately judge the individual influences of both factors, soil and variety. With the use of two sets of samples, the objective of this work was double: (i) to evaluate the influence of different soils in the capability of metal movement from the soil to the potato for the Kennebec variety and (ii) to study the capability of Kennebec compared to other varieties (such as Liseta and Baraka) for metal movement from soil into tubers when they are cultivated in similar soils. Additionally, assignation pattern recognition models have been developed to detect fraud in geographical or varietal origins of samples labeled under the Galician PGI.

■ MATERIALS AND METHODS

Potato Samples and Sample Preparation. Two different sets of potato samples of the same harvest were considered; the first one (Geo-Origin.set) was used for geographical origin modelization, whereas the second (Variety.set) was employed for varietal study. Geo-Origin.set is composed of 90 samples, 45 of these samples are from guaranteed Galician origin, and the remaining 45 samples come from other Spanish regions except Galicia. All of them belong to the Kennebec variety (the only authorized by the PGI normative). Variety.set includes 75 samples; 45 samples are Kennebec variety, 15 samples correspond to Baraka variety, and 15 samples are Liseta variety. All of these 75 samples were cultivated in Galician locations. Because one of the most important criteria to be considered in authentication studies is not having any doubt about the origin and variety of the samples, the potato samples for this study were obtained as follows. Galician origin samples were provided by Certification Council of PGI "Pataca de Galicia" with varietal and geographical origin guaranteed. Non-Galician samples from other Spanish areas are Kennebec samples with undoubtful geographical origin obtained from local producers and retailers.

The treatment of potato samples before analysis was the following. Five tubers of each sample were rinsed with water and skinned. Cross sections were obtained from the tubers, minced, and freeze-dried by employing a Labconco Freeze-Dry system (Labconco, Kansas, MO, USA). Two grams of the lyophilized potato sample was reduced to ash (550 ± 25 °C) following the procedure described by AOAC.[32] Working solutions were prepared by dissolving the obtained ashes in 10 mL of 0.6 M HCl followed by dilution to 25 mL with ultrapure water obtained from a Milli-Q water purification system (Millipore, Bedford, MA, USA).

Metal and Trace Element Determinations. After the sample pretreatment described above, 10 metals and trace elements were determined in triplicate for the analyzed potato samples, K, Na, Rb, Li, Zn, Fe, Mn, Cu, Mg, and Ca, using a Varian (Palo Alto, CA, USA) flame atomic absorption spectrometer model AA10-Plus. Na, K, Li, and Rb were measured by flame atomic emission spectrometry (FAES), whereas the remaining elements were determined in atomic absorption spectrometry (FAAS). Experimental conditions have been published elsewhere.[30−33]

Statistical and Chemometric Procedures. The Geo-Origin.set is a 90 × 10 matrix where the rows correspond to the 90 Kennebec samples studied and the columns correspond to the content of the 10 metal and trace elements determined (K, Na, Rb, Li, Zn, Fe, Mn, Cu, Mg, and Ca). For differentiation purposes, Galician samples were coded as class 1, and non-Galician ones were considered as class 2. The Variety.set is a 75 × 10 matrix where the rows correspond to the 75 Galician samples described and the columns are the content of the above-mentioned metals. In this case, for differentiation purposes, Kennebec samples were coded as class 1, whereas Baraka and Liseta potatoes were considered as class 2.

Because for both data files Geo-Origin.set and Variety.set 10 metals were determined, each potato sample is characterized by a 10-dimensional vector. Therefore, each sample is defined by a point in a multidimensional space of 10 dimensions. A primary evaluation of the two data files was carried out by means of unsupervised chemometric techniques: principal component analysis (PCA) and hierarchical cluster analysis (HCA), with the aim of exploring the relationship among samples and varieties in the 10-multidimendional space. PCA is a dimension reduction and display chemometric technique that transforms the original data matrix $X_{n\times m}$ into a product of two matrices, one of which contains information about the variables (score matrix $S_{n\times \mathrm{PC}}$), and the other about variables (loadings matrix $L_{\mathrm{PC}\times n}$). If the number of factors PC (principal components) selected is small compared to m (number of original variables), PCA supplies a considerable reduction and simplification of the original data matrix X. This chemometric technique was employed to perform a study on the data structure in a reduced dimension preserving the maximum variability contained in data.[34] HCA is a technique (commonly employed complementarily with PCA) applied to the data to find for natural groups of samples a way to discover the structure residing in them. In the present case, samples were hierarchically clustered according to the Ward[35] agglomerative procedure on the basis of the distance among them measured as the squared Euclidean distance. The result is a graphical presentation of the sample clusters in the form of dendrogram, a tree diagram frequently used to illustrate the arrangement of the clusters produced by cluster analysis. To avoid the influence of the diverse size of variables, for both PCA and HCA the original variables were autoscaled (subtracting the mean of the variable and dividing it by the standard deviation), producing new variables with zero mean and unity variance.

The supervised pattern recognition procedure applied for establishing classification models for geographical origin and variety was soft independent modeling of class analogy (SIMCA).[36] SIMCA models a class with a number of principal components selected by cross-validation. Each class in the data set was then described by a model constructed on the basis of a limited number of principal components defining a class subspace in the multidimensional space. Unknown samples will be assigned to a considered class if they are inside the model of this class. The use of SIMCA presents two great advantages over other pattern recognition procedures: first, each data class is modeled separately from the rest of the classes, avoiding influence of the class "B", "C", etc., in the definition of the model for class "A"; second, SIMCA achieves outlier detection. This means that this technique can identify samples not belonging to any of the established classes. Data pretreatment applied prior to SIMCA was an autoscaling performed separately for each class. Once the classification model has been developed, it is important to extend the validation to further data, other than those used for model construction. The ideal situation is when there are enough samples available to create separate (independent) training and test sets. When this is not possible because the number of available samples is low (as is the case at hand), the strategy commonly applied, named cross-validation, consists of dividing the available samples between training and test sets. The first one is used to develop the decision model, whereas the second is employed for validation of the decision rule obtained. As the decision rule and the validation obtained depend on the objects in each set, this process must be repeated many times (with different constitutions of both sets) to guarantee that all samples, at least once, belong to the test set. When only one sample at a time is selected as training set, the validation strategy was named the leave-one-out cross-validation (LOOCV) method, and the procedure must be repeated a number of times equal to the number of samples. In the case at hand, a single sample was considered as the validation set and the remaining observations comprised the training data. This arrangement was repeated 89 and 74 times, respectively, in Geo-Origin.set and Variety.set, so that each sample was used once to validate the model.[37] The success in classification and the prediction for the classification models developed using the SIMCA were evaluated according to their sensitivity and specificity in addition to their classification abilities. Sensitivity and specificity are adequate statistical measurements for the performance of binary classification systems. Sensitivity of a model for class "A" is the proportion of genuine A samples that the model recognizes as belonging to this class. Specificity of a model for a class "A" is the proportion of non-A samples that the model refuses as not belonging to this model. It is clear that these two measures are related to type I and type II errors, and they can be denoted in terms of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Sensitivity is TP/(TP + FN), and specificity is calculated as TN/(TN + FP). A model with 100% sensitivity and 100% specificity for a class is capable of accepting all genuine samples of this class and refusing all foreign samples. The statistical packages employed in this work were the following: PCA and HCA were performed with the use of Statgraphics Centurion XVI (Statistical Graphics Corp., Rockville, MD, USA), and SIMCA was carried out using V-Parvus (University of Genoa, Genoa, Italy).

RESULTS AND DISCUSSION

Determination of Metal Elements Concentration. The results obtained for the 10 analyzed elements in potato samples according to the sample set Geo-Origin.set and Variety.set are summarized in Table 1. As it can be seen, similar results were obtained in both sets. The levels for all metals were in the range reported by other authors in potato samples from diverse origins such as the United States,[26] Denmark,[38] and Spain.[27,30,31] Despite certain differences detected intragroups for both data sets, construction of box−whisker plots for each individual variable showed and overlapped between the ranges of concentrations for samples with diverse geographical and varietal origins. In addition, ANOVA tests carried out on the data showed no significant differences in metal composition for both sets (95% confidence level). Therefore, none of the variables, by itself, is capable of performing discrimination between classes. Consequently, the multivariate approach needs to be explored.

Table 1. Results (Mean and Standard Deviation (SD)) for Metal Elements Determined According to the Data File[a]

	Geo-Origin.set n = 90		Variety.set n = 75	
	mean	SD	mean	SD
K	468	72	469	83
Na	7.30	2.17	8.14	1.99
Rb	0.21	0.12	0.20	0.14
Li	0.12	0.09	0.11	0.09
Zn	0.49	0.14	0.45	0.14
Fe	0.62	0.10	0.48	0.17
Mn	0.17	0.07	0.16	0.06
Cu	0.11	0.04	0.14	0.03
Mg	26.5	3.8	25.7	4.66
Ca	10.1	2.50	11.2	3.59

[a]All results are in mg/100 g.

Study of Geographical Origin. The study of the geographical origin of samples was performed on the chemical information contained in the Geo-Origin.set data file, comprising samples of only the Kennebec variety cultivated in different geographical areas. Primary examination of the data set was performed by means of PCA that was carried out on the autoscaled data. On the basis of the eigenvalues of the principal components (PCs), the selection of the three first PCs provides a cumulative explained variance of 63.24% (see Table 2). This indicates that a graphical representation of the samples in the space defined for these three PCs allows a visualization of the 10-multidimensional structure preserving this percentage of total information contained in the data. By examination of the loadings of the PCs listed in Table 3, for principal component 1 (which represented 34.32% of total variance), it can be seen that Rb and K are the dominant features in the positive part of the axix, whereas Ca and Li are the most influent variables in the negative part of PC 1. This factor can be associated with the soil characteristics and agricultural fertilization practices. PC 2 (16.92% of total variance) is dominated by Fe and Cu in positive and negative parts, respectively. Chemically, this second factor could be interpreted in relation to anthropogenic influences. The evaluation of the scores of the samples has been performed by constructing a score plot in the PC space, accounting for 63.24% of total data variance (Figure 1). As it can be seen, interesting results have been achieved: a natural separation between Galician and non-Galician potato samples was detected. This means that both groups are located in separate regions of the 10-multidimensional space of the variables. This result has been also confirmed by HCA. The dendrogram obtained with this technique using squared Euclidean distance as between-sample similarity measure and clustering Ward method is shown in Figure 2. In the dendrogram, from left to right, two clusters named "A" and "B" are identified. The first cluster, A, is composed of only non-Galician samples and the second cluster, B, is formed from Galician origin potatoes (except one sample). This fact constitutes new evidence of the natural separation between samples from Galician and foreign origins. It corroborates that the metal profile of potato samples contains useful information, which is enough to differentiate between two categories even when the same variety was considered. Therefore, it is clear that the same variety is capable of translocating different metal contents from the diverse soils in which they have been cultivated to tubers.

Table 2. Eigenvalues and Variance Explained for the First Three Principal Components in Data Files Geo-Origin.set and Variety.set

data set	PC	eigenvalue	variance explained (%)	cumulative variance explained (%)
Geo-Origin.set	1	3.432	34.32	34.32
n = 90	2	1.692	16.92	51.24
	3	1.199	11.99	63.24
Variety.set	1	5.255	52.55	52.55
n = 75	2	1.405	14.06	66.61
	3	0.989	9.89	76.50

Table 3. Loadings of the Variables in the First PCs, Geo-Origin.set Data File

variable	PC 1	PC 2	PC 3
K	0.404	0.193	−0.040
Na	0.197	−0.382	−0.380
Rb	0.447	−0.127	0.093
Li	−0.104	0.107	0.551
Zn	0.380	0.273	−0.190
Fe	0.142	0.549	0.045
Mn	0.354	−0.320	0.232
Cu	0.335	−0.412	0.036
Mg	0.329	0.368	−0.267
Ca	−0.272	−0.049	−0.612

Taking into account this result, the following step was the development of a supervised pattern recognition classification system to verify the Galician geographical origin of unknown samples. For this task, SIMCA has been applied to the Geo-Origin.set file, under the LOOCV validation conditions explained under Statistical and Chemometric Procedures. Data were separately autoscaled for both Galician and non-Galician categories, and normal range SIMCA models based on three principal components with a significant level for critical distance equal to 5% were built for each class. The results obtained for correct classification, sensitivity, and specificity are listed in Table 4. Recognition ability is the correct assignation rate in the training set, whereas prediction ability is the correct classification in the validation set obtained by LOOCV procedure. As it can be seen in Table 4, the result for Galician class is highly satisfactory at 97.8 and 93.3%, respectively. The non-Galician model

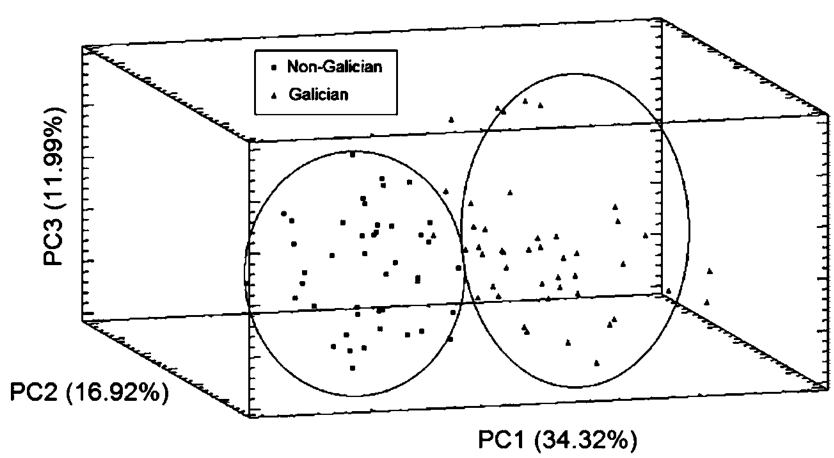

Figure 1. Score plot of the potato samples from Geo-Origin.set in the space of the three first principal components (63.24% variance explained).

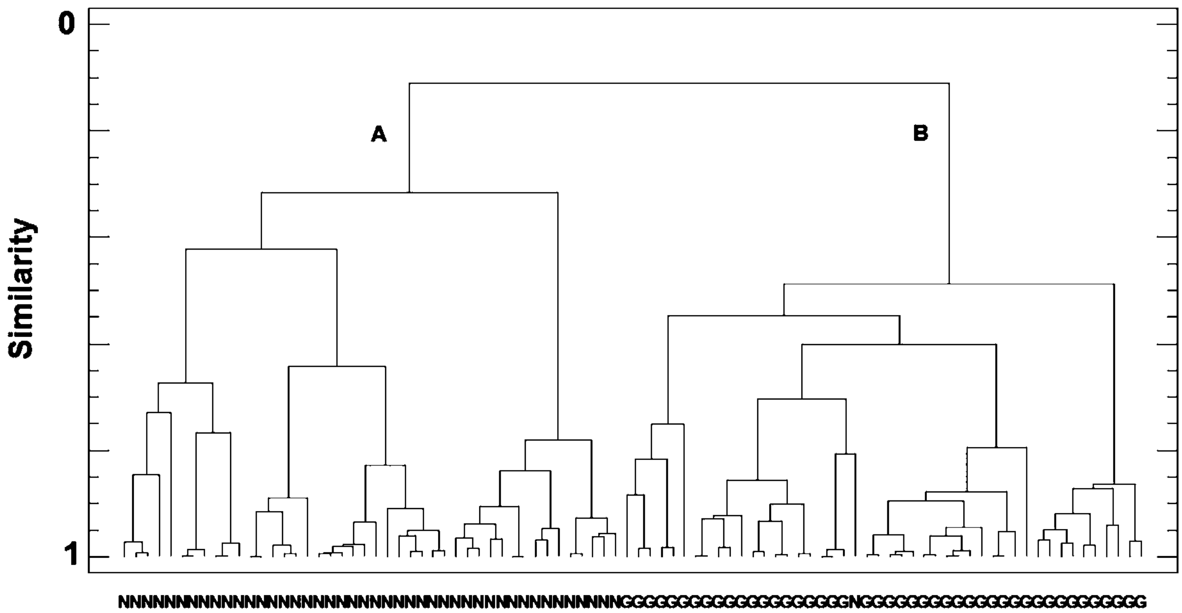

Figure 2. HCA dendrogram of potato samples from Geo-Origin.set. G, Galician origin; N, non-Galician origin.

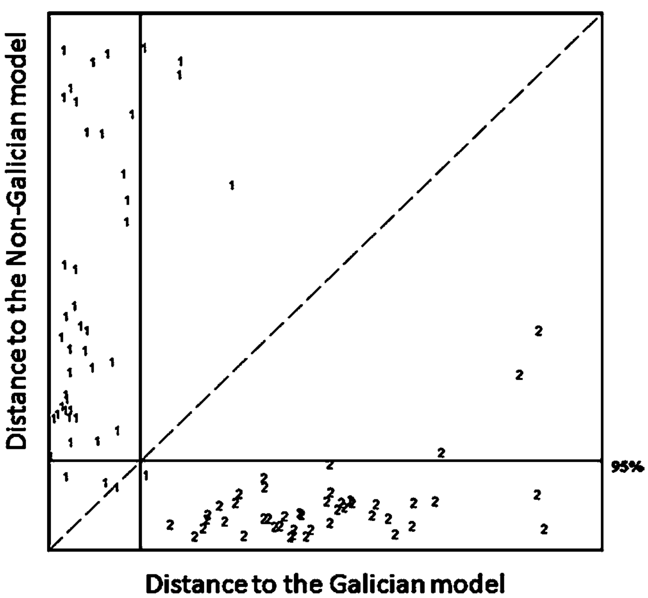

Figure 3. Coomans plot of the squared SIMCA distances for geographical origin models developed. 1, Galician origin; 2, non-Galician origin.

achieved a total classification success with 100% of right assignations. The correct concordance between two recognition and prediction abilities means that the classification model developed is stable and can be generalized. However, despite the good results, the evaluation of the Galician and non-Galician models developed to authenticate samples from this geographical origin must be performed in terms of their sensitivities and specificities. The Galician model attained a sensitivity of 91.1% (the model accepted as Galician 41/45 genuine samples) and a specificity of 100% (reject 45/45 foreign samples). Thus, the Galician SIMCA model built was reasonably satisfactory to authenticate Galician potatoes, but there is a risk of 8.9% of false negatives rejecting a genuine Galician sample as false. However, the model was demonstrated to be highly useful to discover fraud, because it detects all foreign samples trying to be marketed as genuine Galician ones.

These results can be properly visualized using a Coomans plot (as presented in Figure 3), in which the distance of the sample to the Galician model was plotted against the distance to the non-Galician model. In the case at hand, the rectangle in the left side of the plot represents the model for Galician samples because this area shows low distance to this model and high distance to the non-Galician model. Consequently, the rectangle at the bottom of the plot defined the non-Galician model with low distance to this model and high distance to the Galician model. The square in the left-bottom corner corresponds to the overlapping area between two models. As it can be seen, the model for Galician samples does not include any sample of non-Galician origin (coded 2), confirming the 100% specificity previously indicated that allows the complete rejection of foreign samples.

As a conclusion it can be affirmed that the same potato variety produced different metal profiles in the tubers on the basis of the different soils in which they were cultivated. Therefore, the metal fingerprint of potatoes could be appropriate chemical information to develop authentication and fraud detection

Table 4. Classification Abilities, Sensitivities, and Specificities for the SIMCA Models Developed for Geographical and Varietal Origins

model and data file	class	recognition ability (%)	prediction ability (%)	sensitivity (%)	specificity (%)
SIMCA model for geographical origin	Galician	97.8	93.3	91.1 (41/45)	100 (45/45)
(Geo-Origin.set)	non-Galician	100	100	93.3 (45/45)	91.1 (41/45)
SIMCA model for varietal origin	Kennebec	100	100	88.9 (40/45)	100 (30/30)
(Variety.set)	Liseta + Baraka	100	100	90.0 (27/30)	100 (45/45)

procedures combined with pattern recognition techniques, but always concerning the same variety.

Study of Varietal Origin. The Variety.set data file containing metal profiles for potato samples of different varieties (Kennebec, Liseta, and Baraka) cultivated in the same soil type has been used to study if the metal profile information could be useful in the construction of classification models according to variety. In this case, the same chemometric tools of the previous geographical study were utilized. Once PCA was carried out, the study of eigenvalues reveals that three principal components are enough for such data, accounting for the 76.50% of total variance explained (Table 2). Examination of samples' scores in the space of two factors, presented in Figure 4, showed that varieties are in

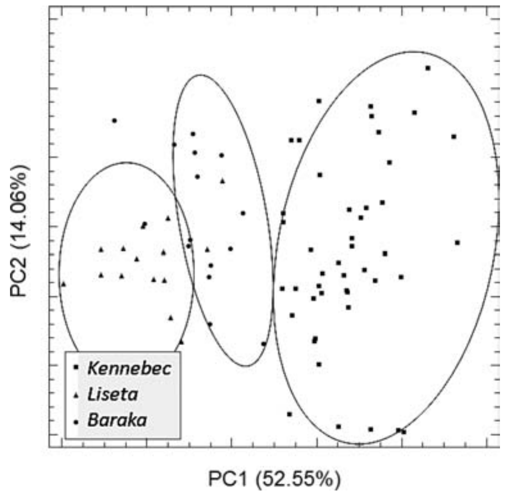

Figure 4. Score plot of the potato samples from Variety.set in the space of the two first principal components (66.61% variance explained).

different areas of the multidimensional space. However, despite the fact that Kennebec samples form a clearly separate group, Liseta and Baraka varieties presented a certain degree of overlapping. This same conclusion has been confirmed by HCA (see Figure 5), in which cluster A was composed of only Kennebec samples, whereas the second cluster, B, was formed of samples corresponding to Liseta and Baraka varieties. This means that, in the same Galician soil, Kennebec potatoes have a different mineral fingerprint from Liseta and Baraka, and, in addition, the latter varieties share similar metal movement capabilities, which differ from that of Kennebec. In the previous study of the Geo-Origin.set, the first two principal components were associated with soil characteristics, agricultural fertilization practices, and anthropogenic influence, respectively. However, in this data set, from the study of loadings of the first principal components it can be concluded that a clear relationship between each component and (only one) a unique factor influencing metal content does not exist. The three first principal components are all mixed factors influenced by soil, agricultural practices, variety, etc.

Taking into account the data set structure revealed by display techniques and bearing in mind that Kennebec is the only variety authorized by PGI "Pataca de Galicia", the SIMCA supervised pattern recognition procedure was employed to establish a model for this variety compared to the other two varieties. For this objective and only with modelization purposes, in the Variety.set data the following varietal arrangement was considered: Kennebec was class 1, and Liseta plus Baraka was class 2. SIMCA was developed under the same conditions as given under Study of Geographical Origin, and the results obtained are summarized in Table 4. Models for Kennebec variety and Liseta and Baraka varieties achieved good classification and prediction rates, confirming that metal content constitutes useful information for varietal classification. Examination of sensitivities showed that there is a certain possibility in the range of 10−11% of false negatives in both models; however, the 100% specificity of the Kennebec model allows the detection of unauthorized foreign varieties (see the Coomans plot for varietal models in Figure 6).

It was demonstrated that the metal profile of potato samples with Galician PGI combined with display and supervised pattern recognition chemometric procedures constitutes an appropriate approach to develop classification systems for geographical origin authentication because the metal fingerprint reflects the soil in which the samples are cultivated. On the other hand, it was also proven that different varieties produced in the same soil type produced different metal profiles due to the different metal movement capability from soil to tuber of the diverse varieties. Among other size and quality requirements, the PGI regulations establish that the potatoes labeled under this quality denomination should be of the only authorized variety Kennebec and that they must be produced in geographic areas of Galicia. Hence, two SIMCA classification models were developed to classify (i) Kennebec potato samples of Galician geographical origin against

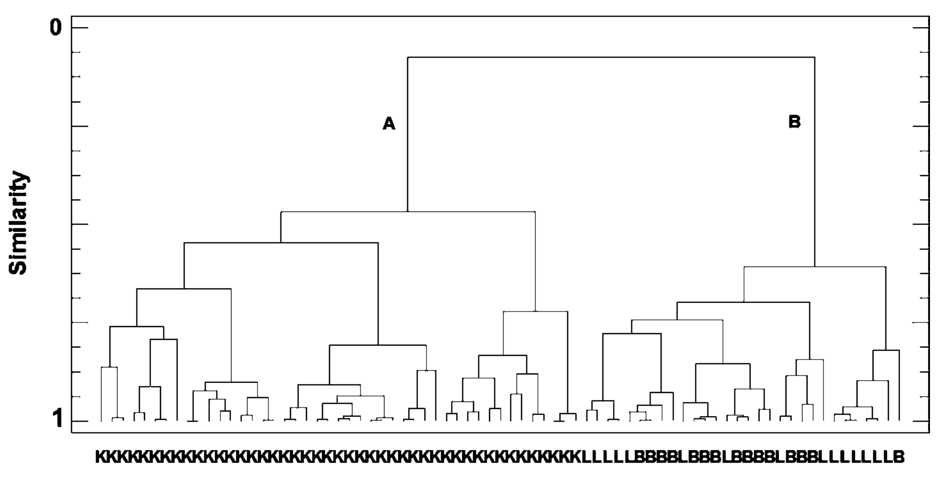

Figure 5. HCA dendrogram of potato samples from Variety.set. K, Kennebec; L, Liseta; B, Baraka.

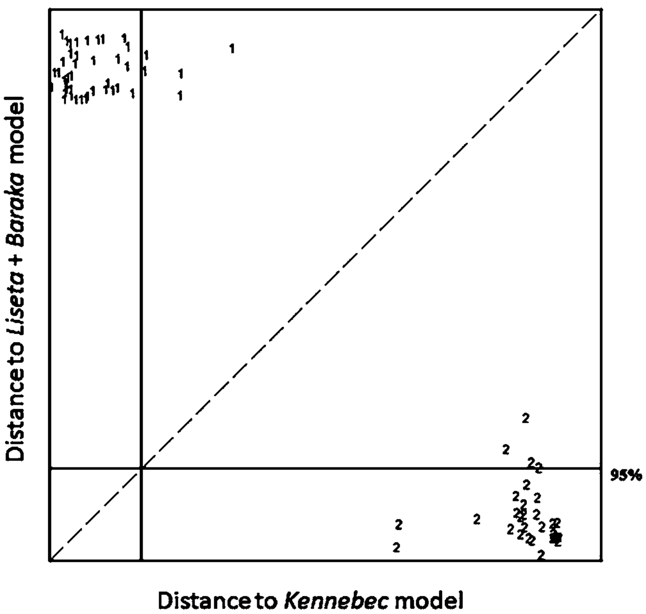

Figure 6. Coomans plot of the squared SIMCA distances for variety models developed. 1, Kennebec; 2, Liseta + Baraka.

Kennebec samples from other Spanish sources and (ii) Galician Kennebec samples against other varieties such as Liseta and Baraka cultivated in geographic Galician authorized areas. The high specificity of the models developed (100%) detected frauds produced by the inappropriate employment and commercialization of foreign and non-Kennebec samples. In contrast, these models have a moderate risk (≈10%) to produce false negative, rejecting as false genuine Galician Kennebec samples.

As a final conclusion of this work, it can be indicated that in this classification study of three potato genotypes (Kennebec, Liseta, and Baraka) from Galician and non-Galician origins, an absolutely reliable knowledge of both origin and variety of the samples is necessary because both factors influence the metal content profile employed as chemical data for classification. For this case, the differentiation of geographical origin should be taken into account together with the varietal data. Otherwise, classifications obtained on the basis of the geographical origin (by influence of soil and climate) could be due to the different variety or vice versa.

■ AUTHOR INFORMATION

Corresponding Author

*(C.H.L.) Phone: + 34 98 282 40 64. Fax: +34 98 228 58 72. E-mail: carlos.herrero@usc.es.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Sun, D. W. In *Modern Techniques for Food Authentication*; Academic Press: Burlington, MA, USA, 2008.

(2) Council Regulation (EC) No. 510/2006 of 20 March 2006 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs. *Off. J. Eur. Union* **2006**, *L93*, 12−25.

(3) Regulation (EU) No. 1151/2012 of the European Parliament and of the Council of 21 November 2012 on quality schemes for agricultural products and foodstuffs. *Off. J. Eur. Union* **2012**, *L343*, 1−25.

(4) Datos Estadísticos 2012. Consellería do Medio Rural e do Mar. Xunta de Galicia; http://www.medioruralemar.xunta.es/institucional/estatisticas/medio_rural/.

(5) Commission Regulation (EC) No. 148/2007 of 15 February 2007 registering certain names in the Register of protected designation of origin and protected geographical indications. *Off. J. Eur. Union* **2007**, *L46*, 14−17.

(6) Orden 29 de octubre del 2001 por la que se aprueba el Reglamento de la Indicación Geográfica Protegida Pataca de Galicia y la de su Consejo Regulador. *DOG* **2011**, *217*, 14468−14482.

(7) Munck, L.; Nørgaard, L.; Engelsen, S. B.; Bro, R.; Andersson, C. A. Chemometrics in food science, a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemom. Intell. Lab.* **1998**, *44*, 31−60.

(8) Rodrigues, S. M.; Otero, M.; Alves, A. A.; Coimbra, J.; Coimbra, M. A.; Pereira, E.; Duarte, A. C. Elemental analysis for categorization of wines and authentication of their certified brand of origin. *J. Food Compos. Anal.* **2011**, *24*, 548−562.

(9) Salvatore, E.; Cocchi, M.; Marchetti, A.; Marini, F.; de Juan, A. Determination of phenolic compounds and authentication of PDO Lambrusco wines by HPLC-DAD and chemometric techniques. *Anal. Chim. Acta* **2013**, *761*, 34−45.

(10) García-Martín, S.; Herrero, C.; Peña, R. M.; Barciela, J. Solid-phase microextraction gas chromatography-mass spectrometry (HS-SPME-GC-MS) determination of volatile compounds in orujo spirits: multivariate chemometric characterisation. *Food Chem.* **2010**, *118*, 456−461.

(11) Sádecká, J.; Tóthová, J.; Májek, P. Classification of brandies and wine distillates using front face fluorescence spectroscopy. *Food Chem.* **2009**, *117*, 491−498.

(12) Escriche, I.; Kadar, M.; Juan-Borrás, M.; Domenech, E. Using flavonoids, phenolic compounds and headspace volatile profile for botanical authentication of lemon and orange honeys. *Food Res. Int.* **2011**, *44*, 1504−1513.

(13) Jasicka-Misiak, I.; Poliwoda, A.; Dereń, M.; Kafarski, P. Phenolic compounds and abscisic acid as potential markers for the floral origin of two Polish unifloral honeys. *Food Chem.* **2012**, *131*, 1149−1156.

(14) Esteban-Díez, I.; González-Sáiz, J. M.; Sáenz-González, C.; Pizarro, C. Coffee varietal differentiation based on near infrared spectroscopy. *Talanta* **2007**, *71*, 221−229.

(15) Alves, R. C.; Casal, S.; Alves, M. R.; Oliveira, M. B. Discrimination between arabica and robusta coffee species on the basis of their tocopherol profiles. *Food Chem.* **2009**, *114*, 295−299.

(16) Bosque-Sendra, J. M.; Cuadros-Rodríguez, L.; Ruiz-Samblás, C.; de la Mata, A. P. Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data. A review. *Anal. Chim. Acta* **2012**, *724*, 1−11.

(17) Dais, P.; Hatzakis, E. Quality assessment and authentication of virgin olive oil by NMR spectroscopy: a critical review. *Anal. Chim. Acta* **2013**, *765*, 1−27.

(18) Consonni, R.; Cagliani, L. R.; Benevelli, F.; Spraul, M.; Humpfer, E.; Stocchero, M. NMR and chemometric methods: a powerful combination for characterization of balsamic and traditional balsamic vinegar of Modena original. *Anal. Chim. Acta* **2008**, *611*, 31−40.

(19) Consonni, R.; Cagliani, L. R.; Rinaldini, S.; Incerti, A. Analytical method for authentication of traditional balsamic vinegar of Modena. *Talanta* **2008**, *75*, 765−769.

(20) Ballin, N. Z. Authentication of meat and meat products. *Meat Sci.* **2010**, *86*, 577−587.

(21) Gonzalvez, A.; Armenta, S.; de la Guardia, M. Trace-element composition and stable isotope ratio for discrimination of foods with Protected Designation of Origin. *TRAC, Trends Anal. Chem.* **2009**, *28*, 1295−1311.

(22) Drivelos, S. A.; Georgiou, C. A. Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. *TRAC, Trends Anal. Chem.* **2012**, *40*, 38−51.

(23) Jensen, K.; Tygesen, T. K.; Kesmir, C.; Skovgaard, I. M.; Søndergaard, I. Classification of potato varieties using isoelectrophoretic patterns, neural nets and statistical methods. *J. Agric. Food Chem.* **1997**, *45*, 158−161.

(24) Longobardi, F.; Casiello, G.; Sacco, D.; Tedone, L.; Sacco, A. Characterisation of the geographical origin of Italian potatoes, based on

stable isotope and volatile compound analysis. *Food Chem.* **2011**, *124*, 1708−1713.
(25) Rodríguez Galdón, B.; Hernández Rodríguez, L.; Ríos Mesa, D.; Lorenzo León, H.; Luna Pérez, N.; Rodríguez Rodríguez, E. M.; Díaz Romero, C. Differentiation of potato cultivars experimentally cultivated based on their chemical composition and by applying linear discriminant analysis. *Food Chem.* **2012**, *133*, 1241−1248.
(26) Anderson, K. A.; Magnuson, B. A.; Tschirgi, M. L.; Smith, B. Determining the geographic origin of potatoes with trace metal analysis using statistical and neural network classifiers. *J. Agric. Food Chem.* **1999**, *47*, 1568−1575.
(27) Casañas Rivero, R.; Suárez Hernández, P.; Rodríguez Rodríguez, E. M.; Darias Martín, J.; Díaz Romero, C. Mineral concentrations in cultivars of potatoes. *Food Chem.* **2003**, *83*, 247−253.
(28) Di Giacomo, F.; del Signore, A.; Giaccio, M. Determining the georaphic origin of potatoes using mineral and trace element content. *J. Agric. Food Chem.* **2007**, *55*, 860−866.
(29) Adamo, P.; Zampella, M.; Quétel, C. R.; Aversano, R.; del Piaz, F.; de Tomasi, N.; Frusciante, L.; Lepore, L.; Carputo, D. Biological and geochemical markers of the geographic and genetic identity of potatoes. *J. Geochem. Explor.* **2012**, *121*, 62−68.
(30) Padín, P. M.; Peña, R. M.; García, S.; Iglesias, R.; Barro, S.; Herrero, C. Characterization of Galician (N.W. Spain) quality brand potatoes: a comparison study of several pattern recognition techniques. *Analyst* **2001**, *126*, 97−103.
(31) Peña, R. M.; García, S.; Iglesias, R.; Barro, S.; Herrero, C. Authentication of Galician (NW Spain) quality brand potatoes using metal analysis. Classical pattern recognition techniques versus a new vector quantization based procedure. *Analyst* **2001**, *126*, 2186−2193.
(32) AOAC. *Official Methods of Analysis of the AOAC International*, 19th ed.; AOAC International: Arlington, VA, USA, 2012.
(33) Latorre, M. J.; García-Jares, C.; Mèdina, B.; Herrero, C. Pattern recognition analysis applied to classification of wines form Galicia (NW Spain) with certified brand of origin. *J. Agric. Food Chem.* **1994**, *42*, 1451−1455.
(34) Joliffe, I. T. *Principal Component Analysis*; Springer: New York, USA, 1986.
(35) Massart, D. L.; Kaufman, L. In *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*; Wiley: New York, USA, 1983; pp 75−101.
(36) Wold, S. Pattern recognition by means of disjoint principal component models. *Pattern Recognit.* **1976**, *8*, 127−139.
(37) Vandeginste, B. G.; Massart, L.; Buydens, L. M.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. In *Handbook of Chemometrics and Qualimetrics: Part B*; Elsevier: Amsterdam, The Netherlands, 1998; pp 207−241.
(38) Bibak, A.; Stürup, S.; Haahr, V.; Gundersen, P.; Gundersen, V. Concentrations of 50 major and trace elements in Danish agricultural crops measured by inductively coupled plasma mass spectrometry. 3. Potato (*Solanum tuberosum* Folva). *J. Agric. Food Chem.* **1999**, *47*, 2678−2684.