



Elemental tracer determination and modeling for geographical origin designation of sweet oranges

Melisa Jazmin Hidalgo^a, Michael Pérez-Rodríguez^{b,*}, José Emilio Gaiad^a, Héctor Casimiro Goicoechea^c, Alberto Mendoza^b, Roberto Gerardo Pellerano^a

^a Instituto de Química Básica y Aplicada del Nordeste Argentino (IQUIBA-NEA), UNNE-CONICET, Facultad de Ciencias Exactas y Naturales y Agrimensura, Ave. Libertad 5400, Corrientes, 3400, Argentina

^b Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Eugenio Garza Sada 2501, Monterrey, 64849, N.L., Mexico

^c Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Santa Fe, 3000, Argentina

ARTICLE INFO

Keywords:

Sweet orange fruits
Elemental tracers
Geographical origin designation
MP-AES
Classification modeling

ABSTRACT

Sweet oranges have long been an integral part of global health and culinary practices, offering a wealth of nutrients and bioactive compounds. Ensuring the authenticity of these citrus fruits is essential for maintaining consumer confidence, promoting transparency in sourcing, and protecting producers' reputations in the marketplace. In this study, we explored the feasibility of using multi-element profiling combined with pattern recognition algorithms to trace the origin of sweet orange samples. To achieve this, we employed an optimized microwave plasma atomic emission spectroscopy (MP-AES) method to analyze the elemental composition (Al, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Li, Mg, Mn, Mo, Na, Ni, Pb, Sr, and Zn) of 183 orange samples from four production regions in northeastern Argentina. Support vector machine (SVM), random forest (RF), and gradient boosting tree (GBT) models were then built using the collected data to identify elemental tracer's indicative of origin. Based on a comprehensive evaluation of overall accuracy, receiver operating characteristic (ROC) curves, and area under the curve (AUC), the GBT model demonstrated the best classification performance, achieving a 96.5 % correct prediction rate on test samples, as confirmed by the ROC curve (AUC = 0.973). Consequently, this approach provides compelling evidence for the potential utility of MP-AES combined with supervised modeling to determine the geographic origin of sweet oranges produced in Argentina, thereby contributing to consumer protection against fraud.

1. Introduction

Citrus fruits have been integral to human history, significantly impacting both traditional and modern medicine in efforts to enhance health. Their broad utility ranges from fresh consumption to a variety of products like juices, desserts, and processed goods, thereby enriching global culinary and nutritional practices while supporting the economic well-being of the regions where they are cultivated [1]. Among the citrus genus, the sweet orange (*Citrus sinensis* L.) is particularly notable, enjoying immense popularity due to its broad consumer appeal. Renowned for its high nutraceutical value, the sweet orange is a rich source of essential nutrients and bioactive compounds, including dietary fibers, vitamins, minerals, sugars, organic acids, and antioxidants like polyphenols, flavonoids, and carotenoids [2]. In addition, organoleptic

properties like aroma, flavor, texture, appearance, along with color and seed presence are key quality attributes influencing in the commercial viability of sweet oranges [3].

The increasing significance of citrus authentication research reflects the essential need to maintain crop integrity and bolster consumer confidence in farm-grown foods [4]. As advancements in this field progress, the focus has broadened to encompass crucial aspects such as determining geographical origin, evaluating cultivation practices, and ensuring compliance with specific standards or certifications [5]. This progress has greatly improved the ability to promote transparency within the supply chain, fostering greater accountability and ethical sourcing practices. With consumers placing a higher value on agricultural products with verified origins, accurate labeling not only boosts market credibility but also supports premium pricing [6,7]. While both

* Corresponding author.

E-mail addresses: michaelp1984@gmail.com, michaelp1984@tec.mx (M. Pérez-Rodríguez).

producers and consumers may face financial losses owing to fraud, counterfeiting, and mislabeling, effectively verifying the authenticity of critics like sweet oranges helps producers mitigate these risks and protect their brand reputation, thereby upholding high standards of quality.

The chemical composition and biological characteristics of citrus fruits, such as sweet oranges, are shaped by a variety of factors including soil properties, agricultural practices, environmental conditions, geographic location, and genetic differences among cultivars [8]. Analyzing how these factors interact is essential for determining the authenticity of sweet oranges from different regions, posing a significant analytical challenge. Researchers have applied a broad spectrum of analytical techniques to assess orange quality. These include evaluations of phytochemical and nutritional properties [9], aroma profiles [10], polyphenol content [11], genetic diversity [12,13], morphological features [14], conventional physicochemical parameters, flavonoids, and volatile compounds [15,16]. Furthermore, advanced methods such as multi-elemental isotopic analysis [17], trace element analysis [18–23], and rare earth element profiling [24] have been employed to provide a thorough understanding of these fruits.

By evaluating the authenticity of oranges, experts have explored various analytical techniques to address this challenge. Notable approaches include high-resolution nuclear magnetic resonance (NMR) for identifying saccharides, amino acids, organic acids, alcohols, ketones, and flavanones [25], as well as inductively coupled plasma mass spectrometry (ICP-MS) for analyzing multi-elemental isotope ratios [26]. Additionally, Centonze et al. [27] have explored patterns of volatile compounds using an MS-based electronic noise to discriminate the geographical origin of these fruits. Although these techniques provide valuable insights, they often require specialized equipment, involve time-consuming and costly analyses, demand labor-intensive sample preparation, and can generate toxic waste. Furthermore, these methods may face limitations in sensitivity and selectivity, potentially leading to false positives or misinterpretation of results. As a result, while these approaches are informative, their drawbacks highlight the need for developing complementary methods that allow achieving a more comprehensive analysis of orange authenticity.

Microwave Plasma Atomic Emission Spectroscopy (MP-AES) offers several advantages for determining the geographical origin of sweet orange fruits through elemental composition analysis. It is both environmentally friendly and cost-effective, as it eliminates the need for hazardous or expensive gases to create plasma discharge [28]. MP-AES is also simpler to operate and maintain compared to other atomic emission techniques [29]. It produces a stable and efficient plasma at low temperatures and atmospheric pressure using microwave energy, offering a wide dynamic range that allows the detection of both trace and major elements in a single analysis with minimal sample preparation. This approach ensures high sensitivity, precision, and reproducibility, making MP-AES an excellent choice for comprehensive and reliable elemental profiling of sweet oranges.

The present research focused on determining the mineral composition (Al, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, Pb, Sr, and Zn) of sweet oranges using MP-AES for designating the geographical origin of the samples. To achieve this, principal component analysis, support vector machine (SVM), random forest (RF), and gradient boosting tree (GBT) models were constructed based on the obtained data aiming to pinpoint elemental markers indicative of geographical provenance. This research makes a significant contribution to the field of citrus authentication by introducing and validating an MP-AES-based method for identifying elemental tracers associated with the geographical origin of oranges. It also enhances the reliability of citrus fruit authentication by demonstrating MP-AES's potential for assessing their quality and certification. This advancement is crucial for verifying the authenticity of farm products, ensuring compliance with labeling standards, and protecting both consumers and producers from fraud and misrepresentation. Consequently, the study not only supports the development of more robust food authentication protocols but also contributes to the

broader goal of maintaining transparency and trust in the global food supply chain.

2. Materials and methods

2.1. Sweet orange samples

In this study, a total of 183 fully ripened sweet orange fruits (*Citrus sinensis*) grown in Argentina were analyzed using MP-AES. These fruits were sourced from agricultural cooperatives and producers between June and November 2021/2022. A minimum of six fruits were randomly collected from each tree, with at least three trees sampled per orchard. The sampling included two botanical varieties of significant commercial importance in the region: Valencia Late ($n = 93$) and Salustiana ($n = 90$). The samples were collected from four production regions, namely Bella Vista (BVS), Monte Caseros (MCS), Misiones (MNS), and Entre Ríos (ERE), spanning 16 different locations across three northeastern provinces of Argentina. These production areas were selected for sampling because the three northeastern provinces of Argentina, namely Entre Ríos, Corrientes, and Misiones, are among the most important citrus-producing regions in the country. Along with the northwestern region (Tucumán, Salta, and Jujuy), this area is particularly renowned for its production of sweet oranges [30,31].

2.2. Chemicals

Ultrapure deionized water (with a resistivity of 18.2 MΩ cm at 25 °C) utilized for the preparation of both working and standard solutions was procured through a Milli-Q Plus water purification system (Millipore, Bedford, MA, USA). Analytical grade HNO₃ 65 % (w/w) and H₂O₂ 30 % (w/w) reagents were sourced from Sigma (St. Louis, MO, USA). Calibration standard solutions were prepared by aqueous dilution with 10 % HNO₃, using a TraceCERT® CRM multi-element standard solution purchased from Merck (Darmstadt, Germany). Argon gas (99.998 % purity) was provided by Praxair (Córdoba, Argentina), and yttrium internal standard was acquired from Sigma-Aldrich (St. Louis, MO, USA). To ensure the purity of our experiments, all glassware and plastic bottles were thoroughly cleansed using a 10 % (m/v) HNO₃ solution. Prior to usage, these vessels underwent multiple rinses with ultrapure water.

2.3. Apparatus

A Taurus Citrix electric juicer (Taurus Group, Lleida, Spain) was utilized to extract fresh juices from sweet oranges. The extracted fruit juices were dried using a Chamber Freeze Dryer (Lab1st, USA). Sample acid digestion was carried out using an Ethos One microwave oven (Milestone Laboratory Systems, Italy) equipped with a high-pressure rotor.

An Agilent 4210 microwave plasma atomic emission spectrometer (MP-AES, Agilent Technologies, Japan) was used for the quantification of four macro elements (Na, K, Ca, Mg) and thirteen microelements (Al, Ba, Cd, Co, Cr, Cu, Fe, Mn, Mo, Ni, Pb, Sr, and Zn) in the digested samples. The sample introduction system consisted of solvent-resistant tubing, a double-pass cyclonic chamber, and an inert flow-blurring nebulizer (OneNeb). Aiming to provide high-purity nitrogen gas for measurements, a Genius SQ 24 nitrogen gas generator (Peak Scientific Instruments Ltd., Scotland, UK) was coupled to the MP-AES spectrometer, producing nitrogen gas from ambient air.

To validate the multi-elemental data obtained through the MP-AES method, we utilized an Agilent 7700 Series inductively coupled plasma mass spectrometer (Agilent Technologies, USA), as our reference technique. This ICP-MS system was equipped with a radiofrequency solid-state generator operating at 1500 W and 27.12 MHz. The setup incorporated a MicroMist™ nebulizer, a cooled double-pass quartz spray chamber, and a Fassel-type ICP torch comprising a three-cylinder assembly with a 2.5 mm injector diameter. Operational parameters were

configured according to specifications outlined by Turra et al. [18]. These included a plasma gas flow rate of 15 L min⁻¹, collision and reaction gas flow rate for helium (99.999 % purity) at 4 mL min⁻¹, an argon (99.999 % purity) carrier gas flow rate of 0.8 L min⁻¹, a sample lifting rate of 0.3 r s⁻¹ and an atomization chamber temperature of 2 °C. Yttrium served as an internal standard, manually introduced into all solutions until reaching a concentration of 10 µg L⁻¹. To maintain optimal performance, we conducted daily optimization of instrumental parameters using the tuning solution during sample aspiration.

2.4. Sample preparation

Fresh sweet orange juices were initially extracted using an electric juicer, followed by homogenization and freeze-drying. The resulting samples were then digested using a microwave oven decomposition system, following the procedure outlined by Donato et al. [32] with certain modifications. Specifically, 100–250 mg of the obtained extract was weighed and transferred into Teflon vessels along with 5 mL of HNO₃ (65 %, w/v) and 2 mL of H₂O₂ (30 %, w/v). Digestion occurred at a constant power of 1000 W at the temperature of 160 °C, with a 15-min ramp, maintaining for an additional 40 min. Subsequently, the clear solutions obtained were left to cool at room temperature before being quantitatively transferred into 25 mL volumetric flasks and diluted with ultrapure deionized water.

2.5. MP-AES measurements

The concentrations of 17 elements in digested sweet orange juices were quantified by MP-AES. The instrument's operating parameters were set according to the manufacturer's recommendations, including 2450 MHz microwave frequency, 1 kW magnetron power, and 20 L min⁻¹ plasma gas (nitrogen) flow rate. The nebulizer pressure ranged from 140 to 240 kPa, with an 8-sec sample uptake duration, a 15-sec stabilizing time, a 30-sec rinsing time, and a 5-sec reading time. Instrumental parameters were optimized daily during the aspiration of the tuning solution. For the determination of individual elements, specific spectral lines were selected, including 396.152 for Al, 455.403 for Ba, 393.366 for Ca, 228.802 for Cd, 340.512 for Co, 425.433 for Cr, 324.754 for Cu, 371.993 for Fe, 766.491 for K, 383.829 for Mg, 403.076 for Mn, 379.825 for Mo, 588.995 for Na, 352.454 for Ni, 405.781 for Pb, 407.771 for Sr, and 213.857 for Zn. These emission lines were chosen based on their optimal signal-to-noise ratio and minimal spectral interferences. To ensure the sample introduction system cleanliness, a blank solution (ultrapure water) was analyzed after every ten samples, along with a multi-element standard solution containing Al, Ba, Cd, Co, Cu, Mn, Mo, Ni, Sr, and Zn at 20 µg L⁻¹, Ca, Fe, and Na at 50 µg L⁻¹, and Mg and K at 150 µg L⁻¹. Background corrections were automatically performed for each element using Agilent MP Expert software (Agilent Technologies, Santa Clara, CA, USA).

2.6. Statistic analytics and classification modeling

To streamline statistical processing and pattern recognition analysis, the multi-elemental data obtained by MP-AES was structured into a rectangular matrix. The matrix rows represented the 183 orange fruit samples, each labeled according to its origin designation, and the columns correspond to the concentrations of the elements quantified in the samples.

Firstly, the dataset underwent preprocessing involving autoscaling and mean centering procedures. This allowed normalizing the data and lessen disparities in concentration scales. Subsequently, we applied the nonparametric version of the ANOVA test, known as the Kruskal-Wallis's test, to identify potential discriminant variables based on statistical differences among sample origins. As a complementary tool, principal component analysis (PCA) was conducted to investigate the inherent distribution of samples within a reduced dimensional space and identify

clustering trends among sample classes [33].

Lastly, three machine learning algorithms, namely random forest (RF), support vector machine (SVM), and gradient boosting tree (GBT), were employed to develop classification models of sweet orange juices. Support Vector Machine (SVM) is a prominent machine learning technique typically used for binary classification. Its primary objective is to identify the optimal hyperplane in a multi-dimensional space that separates different classes of objects or samples by employing maximization techniques. SVM theoretically simplifies the complexities associated with high-dimensional spaces and addresses classification challenges through internal processing functions known as kernel functions. It also facilitates the construction of multi-class classifiers by combining multiple binary classifiers. In this work, SVM was implemented using two distinct kernels to explore different outcomes: a linear kernel (SVM-1) and a radial basis function kernel (SVM-2). Additionally, two ensemble methods were employed: gradient boosting trees (GBT) and random forests (RF). Although both methods excel in solving complex classification issues, they differ in their training processes and how they aggregate individual tree outputs, resulting in models with unique characteristics. GBT builds trees sequentially, with each new tree correcting errors made by the previously trained trees. In contrast, RF trains each tree independently, using a random sample of the data. This randomness aids in creating a more robust model compared to a single decision tree, reducing the likelihood of overfitting [34].

To build the classifiers, the entire dataset was partitioned into training and testing subsets using the Kennard-Stone algorithm. The training set was used to optimize the hyperparameters of each model, while the testing set was employed to evaluate the final model's performance and generalization capabilities. Hyperparameter optimization was performed using a 10-fold cross-validation method, in which the dataset was divided into ten groups, with nine groups used for model training and one for testing.

The metrics used to evaluate classification performance included global accuracy, precision, recall, and F1 score. Accuracy measures the proportion of correct predictions to the total number of cases examined. Precision assesses the accuracy of positive predictions, representing the ratio of true positive predictions to the sum of true positive and false positive predictions. Conversely, recall measures the model's ability to identify all positive instances, calculated as the ratio of true positive predictions to the sum of true positive and false negative predictions. The F1 score, serving as a balance between precision and recall, is the harmonic mean of the two metrics. It offers a comprehensive evaluation of the model's performance, particularly valuable when handling imbalanced datasets due to its consideration of both false positives and false negatives [35]. Furthermore, the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) were also computed. The ROC curve visually illustrates the classifiers' performance in distinguishing between classes considering various threshold settings. It aids in comparing different models and selecting the optimal threshold for a specific application. AUC is a single metric quantifying the model's ability to discriminate between the positive and negative classes, regardless of the chosen threshold. Ranging from 0.0 to 1.0, an AUC of 0.5 indicates performance equivalent to random chance, an AUC >0.5 suggests some degree of discrimination ability, with higher values indicating better performance, and an AUC of 1 signifies perfect discrimination, achieving a perfect balance between sensitivity (true positive rate) and specificity (true negative rate) [36].

Data analysis was conducted using Orange Data Mining Library in Python and R-software version 4.4.0 with the Caret package.

3. Results and discussion

3.1. Analytical parameters and validation

The MP-AES method for analyzing sweet orange juice samples was developed and validated considering various analytical performance

parameters. Table 1 presents the main figures of merit obtained for the proposed method. Linearity and sensitivity were determined from the calibration curves. The determination coefficients (R^2) of the analytical curves were greater than 0.9980, which indicated good linearity for all elements across their respective concentration ranges covered in the calibration.

The calculation of limits of detection (LOD) and limits of quantification (LOQ) was based on the criteria outlined by AOAC [37]. The proposed methodology yielded LOD values between 0.07 and 3.71 $\mu\text{g g}^{-1}$ and LOQ values ranging from 0.21 to 11.1 $\mu\text{g g}^{-1}$. Measurement variability was expressed as the relative standard deviation (% RSD). Intra-day RSD (%) values ranged from 0.6 to 3.6 %, with Cu exhibiting the best precision. The average recoveries fell within the range of 87.2–100.2 %, indicating no matrix effects post-sample microwave-assisted acid digestion. These results meet the acceptance criteria established for intermediate precision (≤ 15 % RSD) and recovery (80–110 %) as per AOAC regulations [37].

To assess the accuracy of data acquired by MP-AES as a practical method for mineral determination in sweet orange juices, we fortified 10 randomly selected samples with known concentrations of various elements. The spiking levels were set at 5 $\mu\text{g g}^{-1}$ for Ba, Co, Cr, Cu, Fe, Mn, Mo, Ni, Sr, and Zn; 25 $\mu\text{g g}^{-1}$ for Al, Ca, and Na; and 250 $\mu\text{g g}^{-1}$ for K and Mg. These fortified samples were then analyzed using the ICP-MS technique, as described by Turra et al. [18], which served as the reference method. We then compared the results obtained from both methods using a Student's t-test at a 95 % confidence level. The calculated t-values, ranging from 0.314 to 2.578, were all below the critical value ($t = 2.776$; $p = 0.05$), indicating no significant differences between the MP-AES and ICP-MS data.

In this study, we compared the data obtained by MP-AES with those obtained by ICP-MS to evaluate their respective strengths in determining the elemental profiling. While ICP-MS is widely recognized for its exceptional sensitivity and ability to detect a broader range of elements at trace levels, MP-AES technique offers several advantages, including cost-effectiveness, ease of use, and sufficient accuracy for the specific needs of our study. Our findings suggest that MP-AES is an effective and viable alternative to ICP-MS technique for designating the geographic origin of sweet orange fruits, making it a practical option for routine applications in the field of food authentication.

3.2. Content and variability of elements in sweet orange juices

The concentrations of seventeen elements determined in sweet orange juices are shown in Table 2. Our findings revealed that over 95 % of the fruit samples contained three elements, i.e., Cd, Ni, and Pb, below the method detection limits, hence excluding them from further

Table 1
Figures of merit, precisions, and recoveries obtained for the MP-AES method.

Element	R^2	LOD ($\mu\text{g g}^{-1}$)	LOQ ($\mu\text{g g}^{-1}$)	RSD (%)	Recovery (%)
Al	0.9989	0.92	2.70	0.8	88.1
Ba	0.9991	1.15	3.32	3.6	87.2
Ca	0.9988	2.03	6.14	2.2	93.5
Cd	0.9998	0.07	0.21	1.3	97.0
Co	0.9998	0.09	0.27	0.6	92.6
Cr	0.9980	0.08	0.24	1.0	94.5
Cu	0.9999	0.14	0.42	0.6	100.2
Fe	0.9988	0.28	0.84	2.0	98.0
K	0.9998	2.80	8.42	0.9	94.5
Mg	0.9997	3.71	11.1	1.8	86.8
Mn	0.9998	0.07	0.21	0.6	98.7
Mo	0.9996	0.07	0.21	3.1	92.3
Na	0.9986	2.54	7.54	0.7	91.5
Ni	0.9999	0.12	0.36	0.8	97.4
Pb	0.9996	0.13	0.38	2.9	87.9
Sr	0.9998	0.10	0.30	1.2	96.1
Zn	0.9996	0.12	0.36	1.7	97.8

Table 2
Concentrations of elements (average \pm standard deviation) quantified in sweet orange juice samples.

Element	Sampling sweet oranges fruits ($\mu\text{g g}^{-1}$)			
	BVS ($n = 44$)	ERE ($n = 66$)	MCS ($n = 49$)	MNS ($n = 24$)
Al	5.2 \pm 1.1	12.8 \pm 11.7	14.8 \pm 8.6	74.8 \pm 4.2
Ba	3.3 \pm 0.5	3.8 \pm 0.90	8.3 \pm 0.9	9.2 \pm 0.6
Ca	22.9 \pm 8.3	25.0 \pm 3.2	12.5 \pm 10.7	18.7 \pm 3.6
Cd ^a	< LOD	< LOD	< LOD	< LOD
Co	0.5 \pm 0.6	0.3 \pm 0.5	0.2 \pm 0.2	0.7 \pm 0.5
Cr	0.2 \pm 0.1	0.3 \pm 0.08	1.1 \pm 0.6	0.2 \pm 0.4
Cu	2.8 \pm 0.5	9.5 \pm 0.3	10.0 \pm 5.5	1.8 \pm 0.3
Fe	0.6 \pm 0.3	0.8 \pm 0.2	0.7 \pm 0.2	0.9 \pm 0.4
K ^b	1.1 \pm 0.5	1.7 \pm 0.6	1.1 \pm 0.5	0.9 \pm 0.3
Mg	138.2 \pm 45.4	130.1 \pm 25.2	112.0 \pm 65.2	146.4 \pm 52.0
Mn	0.5 \pm 0.2	3.4 \pm 0.1	0.9 \pm 0.3	0.6 \pm 0.5
Mo	0.7 \pm 0.5	0.9 \pm 0.4	0.7 \pm 0.6	0.9 \pm 0.2
Na	6.5 \pm 1.2	3.1 \pm 2.4	2.8 \pm 1.5	8.3 \pm 1.3
Ni ^a	< LOD	< LOD	< LOD	< LOD
Pb ^a	< LOD	< LOD	< LOD	< LOD
Sr	2.3 \pm 0.4	2.5 \pm 0.8	3.1 \pm 0.3	2.8 \pm 0.1
Zn	1.2 \pm 0.5	1.5 \pm 0.7	2.2 \pm 0.6	1.8 \pm 0.8

^a Concentration values under limit of detection (<LOD).

^b Concentration values (mg g^{-1}) refer to dry sample.

investigation in this study.

Among the quantified elements, the average concentrations followed the order: $\text{K} > \text{Mg} > \text{Ca} > \text{Al} > \text{Na} > \text{Cu} > \text{Ba} > \text{Sr} > \text{Zn} > \text{Fe} > \text{Mn} > \text{Mo} > \text{Co} > \text{Cr}$. Potassium (K) exhibited the highest average concentrations in the sweet orange fruit samples, ranging from 0.89 mg g^{-1} in the MNS region to 1.72 mg g^{-1} in the ERE region. Potassium (K) is essential for citrus plants, serving various functions such as neutralizing organic acids, influencing enzymatic reactions, and facilitating the synthesis of proteins, sugars, and starch. It also contributes to fruit formation by enhancing size, color, and flavor. Notably, potassium (K) levels reported in commercial orange juice (ranging from 1.44 to 13.57 mg g^{-1}) were higher than those observed in this study [17]. The second most abundant element in the samples was magnesium (Mg), with a range of 132 $\mu\text{g g}^{-1}$ in MCS to 164 $\mu\text{g g}^{-1}$ in MNS. This element plays a crucial role as an activator of various enzymes in citrus plants and is involved in processes such as photosynthesis, carbohydrate metabolism, and nucleic acid synthesis [38]. The observed Mg levels exceeded those previously reported in orange juice from Brazil, which ranged from 46.2 to 100.3 $\mu\text{g g}^{-1}$ [20].

In the different regions studied, aluminum (Al) and calcium (Ca) content ranked third and fourth, respectively, following potassium (K) and magnesium (Mg), except in the MCS region. The highest levels of these elements were detected in the MNS region, measuring 74.8 $\mu\text{g g}^{-1}$ for Al and 18.7 $\mu\text{g g}^{-1}$ for Ca. The latter is a vital mineral for citrus plants that plays a structural role in membrane systems, being also essential for cell division and chromosome stability occurring in the cell wall [39]. The aluminum (Al) levels in orange juice analyzed in our study were consistent with those reported in orange juice from Brazil [18]. However, calcium (Ca) concentrations were lower compared to another study conducted in Brazil [20].

The overall average values of Ba, Cu, Mn, Sr, Na, and Zn ranged from 1 to 10 $\mu\text{g g}^{-1}$. Samples from the MNS region exhibited the highest levels of Ba and Na, while those from the ERE region showed the highest Mn contents. Conversely, samples from the MCS region displayed elevated mean values of Cu, Sr, and Zn. In particular, manganese (Mn) is essential for nitrogen metabolism and respiration in citrus plants [40]. Cristea et al. [17] reported that the concentrations of Mn, Cu, Zn, and Sr determined in commercial orange juice agreed with our results, except for Na levels, which were lower in our study.

Among the trace elements, Co, Cr, Fe, and Mo were detected in concentrations below 1 $\mu\text{g g}^{-1}$. Despite their low abundance, these elements play significant roles in various physiological processes of citrus

plants. For instance, iron (Fe) is involved in photosynthetic enzyme systems and catalyzes chlorophyll production [41]. In our study, Fe levels were lower than those reported by Cristea et al. [17] for commercial orange juice, whereas Co levels were higher, underscoring the intricate balance of these elements within the fruit.

3.3. Exploratory elemental composition analysis

The Kruskal-Wallis's test identified significant differences in six elements, namely Al, Ba, Fe, K, Mg, and Mn, across the four regions ($p < 0.05$). Consequently, PCA was conducted to discern similarities and disparities among the multi-elemental profiles of the examined sweet orange samples (Fig. 1). This statistical tool is adept at exploring data by unveiling the latent structure of high-dimensional datasets without prior information. The cumulative variance contribution of the first five principal components (PCs) reached 95.0 %. Notably, PC1 (39.8 %) was primarily influenced by the concentrations of Al, Ba, Ca, Mg, and Cu, while PC2 (20.2 %) was predominantly driven by the levels of Fe, Sr, Cr, Mn, and K. Fig. 1 depicts the biplot generated by PC1 vs. PC2, wherein orange samples from distinct geographical origins tended to cluster, facilitating the identification of four distinct sample groups. The clusters associated with MCS, ERE, and BVS samples exhibited greater overlap compared to those from MNS samples along PC1. Particularly, the MNS samples displayed relatively higher concentrations of Al, Ba, and Mg, along with lower levels of Cu. This characteristic low Cu content was notably observed in the MCS samples. In addition, the MCS and ERE samples showed considerable overlap, with slight separation along PC2, characterized by differences in Fe, Sr, Mn, and K contents. These results indicate a relationship between specific elements and the samples' geographical origin.

3.4. Origin discrimination of sweet orange fruits

Data-driven machine learning involves identifying patterns within observed data samples and leveraging these patterns to forecast future data or infer unobservable data. It is a fundamental approach in modern artificial intelligence, finding widespread applications across various

burgeoning fields, particularly in food authenticity testing [42,43].

To determine the geographical origin of the sweet orange fruits studied, three machine learning algorithms (SVM, GBT, and RF) were employed to construct four classification models for performance comparison. The classifiers were trained, validated, and evaluated using a training set (70 %) and a test set (30 %). During model training, an exhaustive grid search method was applied to identify the optimal combination of hyperparameters. This process involved iterative 10-fold cross-validation, with the best hyperparameters selected based on the training set results. The test set was then used to evaluate the performance of the proposed models [44].

Table 3 presents the classification results of the proposed models. All models achieved prediction rates above 95 % on the training sets. However, to accurately compare the classifiers, it is crucial to evaluate their performance on the test data (unseen data). The overall accuracies obtained were as follows: 93.3 % for SVM-1, 91.0 % for SVM-2, 92.3 % for RF, and 96.5 % for GBT. Therefore, the GBT model demonstrated the best performance, achieving a perfect accuracy rate on the training samples. Additionally, this model yielded the highest values for other classification metrics, including precision (0.949), recall (0.949), and F1-score (0.952).

To further assess the discrimination performance of the GBT and RF models for each geographical origin, ROC curves were depicted in Fig. 2 and AUC values were computed. The AUC is a key single-value metric used to compare the overall performance of different classifiers through ROC curve analysis [45]. The results indicated that the ensemble

Table 3

Performance of the supervised learning models in discriminating sweet orange samples from different geographical origin.

Models	Training set	Testing set	Precision	Recall	F1-score	AUC
	Accuracy	Accuracy				
SVM-1	0.962	0.933	0.883	0.880	0.880	0.701
SVM-2	0.956	0.910	0.880	0.880	0.880	0.811
RF	1.000	0.923	0.915	0.915	0.915	0.947
GBT	1.000	0.965	0.949	0.949	0.952	0.973

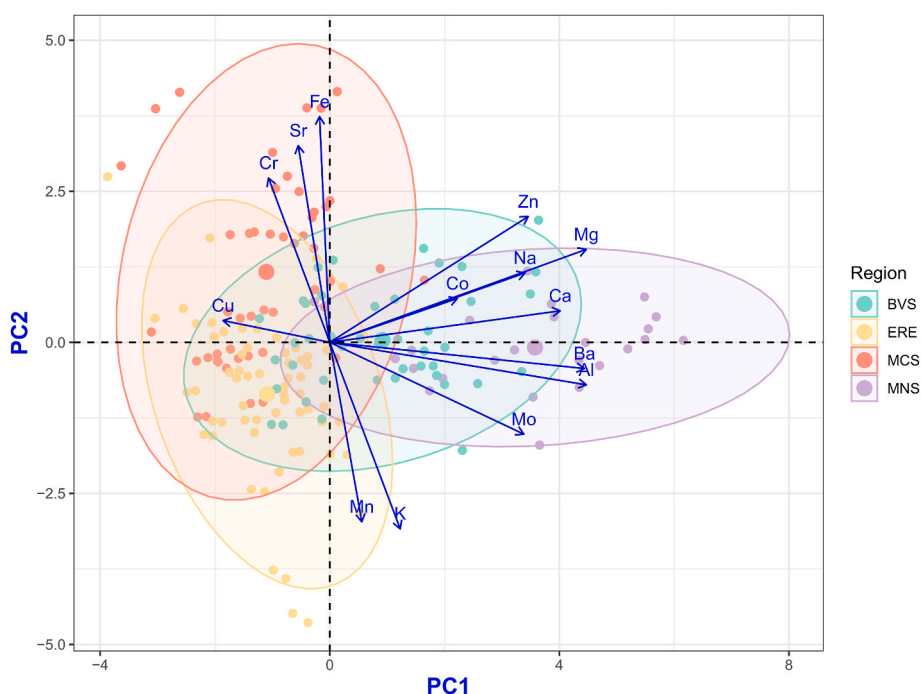


Fig. 1. Biplot of scores and loadings of the elemental predictors onto the PC1-PC2 plane build from the sweet orange samples. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

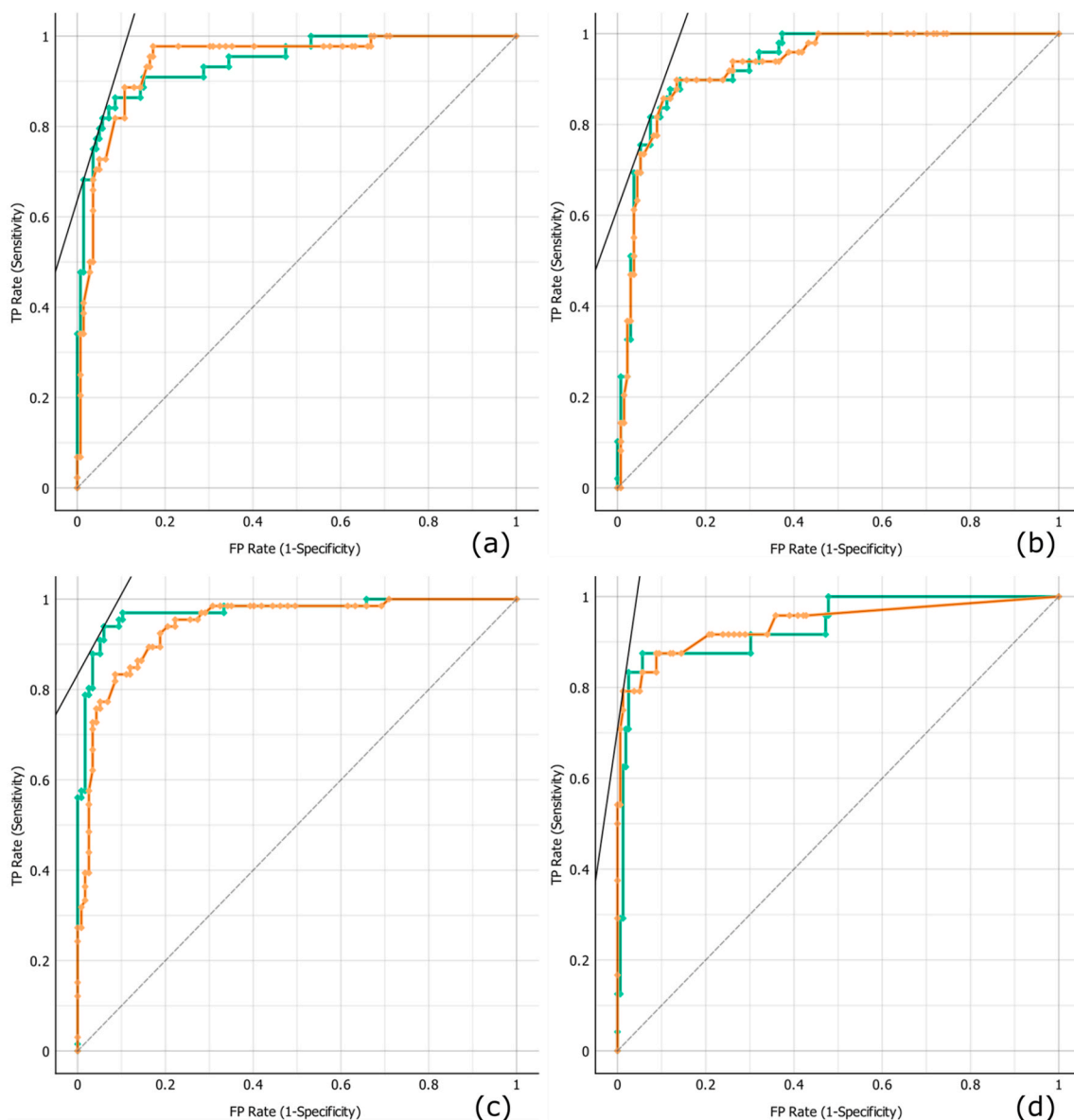


Fig. 2. ROC curves of the ensemble models for classification of sweet oranges based on geographical origin: BVS (a), ERE (b), MCS (c), and (d) MNS. Green line (GBT), Orange line (RF). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

methods achieved higher AUC values than those obtained by the SVM models with different kernels. Furthermore, Fig. 2 shows that the GBT model demonstrates better sensitivity and specificity, particularly in the samples from MCS (Fig. 2-c).

Finally, the previous results indicate that the performance of the two proposed ensemble methods is significantly superior to that of the SVM models in designating the geographic origin of sweet orange fruits. Notably, the GBT model showed exceptional performance in some regions, demonstrating high robustness in distinguishing among sample sets with high similarities, which makes discrimination challenging. However, it is important to acknowledge that ensemble methods also have limitations, such as longer training times and a potential tendency toward overfitting and instability. These concerns have been thoroughly addressed through the cross-validation studies presented in this work, reinforcing the validity of our findings.

4. Conclusions

Elemental profiling measured by MP-AES was employed to designate the geographical origin of sweet orange fruits produced in Argentina.

PCA-based exploratory data analysis revealed patterns reflecting geographical similarities among the samples. Subsequently, SVM, RF, and GBT models were constructed, using the region of origin as a classification factor, to identify elemental tracer's indicative of provenance. In testing samples, the GBT model achieved a success rate of 96.5 %, outperforming SVM-1 at 93.3 % and RF at 92.3 %. Moreover, the GBT model demonstrated the highest reliability in distinguishing sweet oranges, as confirmed by ROC curve analysis. These findings underscore the potential of the MP-AES technique for verifying the origin of orange samples through elemental profile modeling. The proposed approach significantly enhances food authentication efforts and contributes to protecting consumers against mislabeling and fraud in the supply chain. Future research in this area could explore factors such as seasonality, climate, pollution, contact materials, and varying stages of fruit maturation for linking elemental content more accurately to geographic origin. Additionally, ensuring an adequate sample size and random sampling is essential to fully understand the insights obtained.

- [34] P. Filzmoser Varmuza, Introduction to Multivariate Statistical Analysis in Chemometrics, CRC Press, 2009, <https://doi.org/10.1201/9781420059496>.
- [35] D. Ballabio, F. Grisoni, R. Todeschini, Multivariate comparison of classification performance measures, *Chemometr. Intell. Lab. Syst.* 174 (2018) 33–44, <https://doi.org/10.1016/j.chemolab.2017.12.004>.
- [36] C. Gigliarano, S. Figini, P. Muliere, Making classifier performance comparisons when ROC curves intersect, *Comput. Stat. Data Anal.* 77 (2014) 300–312, <https://doi.org/10.1016/j.csda.2014.03.008>.
- [37] G.W. Latimer Jr., Guidelines for Standard Method Performance Requirements: Official Methods of Analysis, AOAC International, 2016, pp. 1–18. Available from: <https://www.aoac.org/>.
- [38] I.E. Papadakis, C. Antonopoulou, T. Sotiropoulos, C. Chatzissavvidis, I. Therios, Effect of magnesium on mineral nutrition, chlorophyll, proline and carbohydrate concentrations of sweet orange (*Citrus sinensis* cv. Newhall) plants, *Appl. Sci.* 13 (2023) 7995, <https://doi.org/10.3390/app13147995>.
- [39] D. Eticha, A. Kwast, T.R. de S. Chiachia, N. Horowitz, H. Stützel, Calcium nutrition of orange and its impact on growth, nutrient uptake and leaf cell wall, *Citrus Res. Technol.* 38 (2017) 62–70, <https://doi.org/10.4322/crt.ICC096>.
- [40] S. Kwakye, D.M. Kadyampakeni, E. van Santen, T. Vashisth, A. Wright, Variable manganese rates influence the performance of Huanglongbing-affected citrus trees in Florida, *Hortscience* 57 (2022) 360–366, <https://doi.org/10.21273/HORTSCI16337-21>.
- [41] S.F. El-Gioushy, Z. Ding, A.M.E. Bahloul, M.S. Gawish, H.M. Abou El Ghit, A.M.R. A. Abdelaziz, H.S. El-Desouky, R. Sami, E. Khojah, T.A. Hashim, A.M.S. Kheir, R.M. Y. Zewail, Foliar application of nano, chelated, and conventional iron forms enhanced growth, nutritional status, fruiting aspects, and fruit quality of Washington navel orange trees (*Citrus sinensis* L. Osbeck), *Plants* 10 (2021) 2577, <https://doi.org/10.3390/plants10122577>.
- [42] Z. Deng, T. Wang, Y. Zheng, W. Zhang, Y.H. Yun, Deep learning in food authenticity: recent advances and future trends, *Trends Food Sci. Technol.* 144 (2024) 104344, <https://doi.org/10.1016/j.tifs.2024.104344>.
- [43] R. Goyal, P. Singha, S.K. Singh, Spectroscopic food adulteration detection using machine learning: current challenges and future prospects, *Trends Food Sci. Technol.* 146 (2024) 104377, <https://doi.org/10.1016/j.tifs.2024.104377>.
- [44] B.V. Canizo, L.B. Escudero, M.B. Pérez, R.G. Pellerano, R.G. Wuilloud, Intra-regional classification of grape seeds produced in Mendoza province (Argentina) by multi-elemental analysis and chemometrics tools, *Food Chem.* 242 (2018) 272–278, <https://doi.org/10.1016/j.foodchem.2017.09.062>.
- [45] J.T. Wixted, L. Mickes, S.A. Wetmore, S.D. Gronlund, J.S. Neuschatz, ROC analysis in theory and practice, *J. Applied Research in Memory and Cognition* 6 (2017) 343–351, <https://doi.org/10.1016/j.jarmac.2016.12.002>.