Food Chemistry: X 20 (2023) 101040

Contents lists available at ScienceDirect

Food Chemistry: X

journal homepage: www.sciencedirect.com/journal/food-chemistry-x

Geographical origin identification of mandarin fruits by analyzing fingerprint signatures based on multielemental composition

Melisa Jazmin Hidalgo [a,*], José Emilio Gaiad [a], Héctor Casimiro Goicoechea [b], Alberto Mendoza [c], Michael Pérez-Rodríguez [c,*], 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] *Laboratorio de Desarrollo Analítico y Quimiometría (LADAQ), Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Ciudad Universitaria, Santa Fe 3000, Argentina*

[c] *Tecnologico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Eugenio Garza Sada 2501, Monterrey 64849, N.L., Mexico*

ARTICLE INFO

Keywords:
Mandarin juices
Element profiles
MP-AES, Geographical origin authentication
Predictive modeling
Support vector machine

ABSTRACT

Given rising traders and consumers concerns, the global food industry is increasingly demanding authentic and traceable products. Consequently, there is a heightened focus on verifying geographical authenticity as food quality assurance. In this work, we assessed pattern recognition approaches based on elemental predictors to discern the provenance of mandarin juices from three distinct citrus-producing zones located in the Northeast region of Argentina. A total of 202 samples originating from two cultivars were prepared through microwave-assisted acid digestion and analyzed by microwave plasma atomic emission spectroscopy (MP-AES). Later, we applied linear discriminant analysis (LDA), k-nearest neighbor (k-NN), support vector machine (SVM), and random forest (RF) to the element data obtained. SVM accomplished the best classification performance with a 95.1% success rate, for which it was selected for citrus samples authentication. The proposed method highlights the capability of mineral profiles in accurately identifying the genuine origin of mandarin juices. By implementing this model in the food supply chain, it can prevent mislabeling fraud, thereby contributing to consumer protection.

1. Introduction

The global food industry is undergoing a growing demand for authentic and traceable products, driven by consumers' increasing concerns about food safety, quality, and ethical sourcing (Aung & Chang, 2014; King et al., 2017). Particularly, citrus production has gained notable attention within the agricultural sector due to their widespread consumption and economic importance. Consequently, the geographical authenticity verification of citrus has emerged as a pivotal endeavor for upholding the integrity of supply chains and nurturing consumer trust (Rosales Martínez et al., 2020). This practice is an initiative adopted by many countries to ensure the quality of foodstuffs originating from specific regions by establishing controlled labeling. The protected designation of origin depicts a challenge involving establishing a clear link between the geographical region where a crop is harvested and its unique characteristics, qualities, or attributes (Dias & Mendes, 2018).

Mandarin (*Citrus reticulata*), one of the original citrus ancestors, is distinguished by its unique flavor and numerous nutritional benefits (Goldenberg et al., 2018). In Argentina, for example, significant emphasis falls on growing mandarin varieties like 'Okitsu' (*Citrus unshiu* Marc) and 'Murcott' (*Citrus reticulata* Blanco). These cultivars hold great importance for the regional population, predominating in the provinces of Entre Ríos, Misiones, and Corrientes. While mandarins are widely enjoyed as fresh fruits, specific attention has been paid to their processed products, attributing it to the market gradual saturation regarding the fresh citrus. Mandarin juice is a processed product that has gained worldwide popularity and acclaim because of its distinctive flavor and aroma, nutrient-rich composition, antioxidant properties, hydration capacity, revitalizing nature, and freshness (Cheng et al., 2020). This processed product represents a testament to the versatility and desirability of mandarins in elaborating healthy commodities.

The chemical composition of mandarins, like those of any other crop, is strictly underlying the environmental conditions prevailing in farming

* Corresponding authors.
E-mail addresses: melujaz1806@gmail.com (M. Jazmin Hidalgo), michaelpr@tec.mx, michaelpr1984@gmail.com (M. Pérez-Rodríguez).

https://doi.org/10.1016/j.fochx.2023.101040
Received 25 September 2023; Received in revised form 21 November 2023; Accepted 28 November 2023
Available online 1 December 2023

regions. Therefore, factors encompassing soil characteristics, climate, watering management, and fertilizer usage play a fundamental role in shaping of this citrus, consequently influencing its overall quality (Beltrán-González et al., 2008; Elbasiouny et al., 2022; Schoonover & Crim, 2015). Numerous analytical techniques have been explored to develop methodologies centered on evaluating the quality of mandarins and their processed products. Within this framework, the correlation between flavor composition and sensory attributes in thermally processed mandarin juices was investigated using multidimensional gas chromatography-mass spectrometry (MDGC-MS) (Cheng et al., 2023). Approaches based on electronic nose (E-nose) and electronic tongue (E-tongue) have been employed to monitor the quality status of mandarins (Qiu & Wang, 2015). The applications of the latter extend to differentiate between high pressure processed mandarin juices under varying pressures (Qiu et al., 2017) and to ascertain influence of storage temperature and time on the intrinsic quality of Satsuma mandarins (Qiu & Wang, 2015).

Regarding authenticity concerns, Markus Jungen et al. (2020) studied the effect of species and processing techniques of mandarins on phlorin using 1H NMR spectroscopy. This enabled the differentiation of juices from peel extracts and whole-processed products according to phlorin profiles. In turn, Jandrić and Cannavan (2017) employed ultra-high performance liquid chromatography with quadrupole time-of-flight mass spectrometry (UHPLC–Q-TOF–MS) to identify the provenance and variety of diverse citrus fruits such as orange, grapefruit, pomelo, and mandarin. Furthermore, research into mineral contents and nutritional attributes encompassing the pulp and peel of various mandarin cultivars were undertaken by inductively coupled plasma-optical emission spectrometry (ICP OES) (Czech et al., 2020; Guo et al., 2023; Z. Li et al., 2021). Pursuant to this perspective, the development of analytical protocols exploring the intrinsic molecular and elemental fingerprint of foodstuffs has been progressively increasing to unveil their true origin.

Microwave plasma atomic emission spectroscopy (MP-AES) is a user-friendly and cost-effective analytical technique compared to ICP-AES or ICP-MS, offering broad element coverage. It harnesses self-sustained atmospheric pressure microwave energy to generate a plasma discharge using nitrogen supplied from either a gas cylinder or a nitrogen generator via ambient air. The latter makes it cheaper and safer than previous techniques requiring flammable gases. Minimal sample preparation and the high temperature reaching by microwave plasma ensure rapid and efficient elemental analysis, providing quantitative and accurate results by enhancing atomization and reducing spectral and chemical interferences (Balaram, 2020). Its simplified design and compatibility with diverse complex matrices make it suitable for applications in mineral exploration (Helmeczi et al., 2018), energy fuels (Poirier et al., 2017), and agriculture (W. Li et al., 2013), as well as in the areas of health (Carter et al., 2019), environment (Kamala et al., 2014), and foods (Gallego Ríos et al., 2017).

Recognizing the pivotal role of pattern recognition algorithms that utilizes multiple data sources in overcoming challenges related to the quality, authenticity, and traceability of citrus fruits, the current work is focused to identifying the geographical origin of mandarin juices from various farming regions in northeastern Argentina by modeling mineral profiles (Al, Ca, Co, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, Sr, and Zn) obtained using MP-AES after microwave-assisted acid digestion.

2. Materials and methods

2.1. Chemicals and standards

Analytical reagent grade HNO_3 65 % and ultra-pure grade 30 % (w/w) H_2O_2 were acquired from Sigma (St. Louis, MO, USA). A TraceCERT® CRM multi-element standard solution purchased from Merck (Darmstadt, Germany) was employed to prepare the calibration standard solutions by aqueous dilution with 10 % nitric acid. Working aqueous solutions were prepared using ultrapure water (18 MΩ cm^{-1} at 25 °C) obtained through a Milli-Q® Water purification system (Millipore, Bedford, MA, USA). Yttrium internal standard was acquired from Sigma-Aldrich (St. Louis, MO, USA) and argon (99.998 % purity) was supplied by Praxair (Córdoba, Argentina). All glassware and plastic bottles were cleaned with a 10 % (m/v) HNO_3 solution and rinsed several times with ultrapure water.

2.2. Sampling and sample preservation

A total of 202 mandarin fruit samples were collected from three distinct citrus-producing zones located in the Northeast region of Argentina. The botanical cultivars subjected to scrutiny were: 'Okitsu' (*Citrus unshiu* Marc) and 'Murcott' (*Citrus reticulata* Blanco). Sampling was directly undertaken by seasoned producers, which ensured information concerning the specific cultivar, geographical location, and production system. All mandarin fruits were harvested over two consecutive years (2021/2022), comprising the harvest of 100 fruits in 2021 and 102 fruits in 2022. According to the Argentine Citrus Federation (Federcitrus), the sampling region was divided into three mandarin-producing areas: Misiones (MIS, $n = 70$), Corrientes (CTE, $n = 56$), and Entre Ríos (ETR, $n = 76$). This division was employed for authenticating the samples based on their geographical origin. Fresh juices were obtained by squeezing the fruits using an electric juicer. After juice extraction, samples were homogenized, freeze-dried, labeled, and stored in polyethylene zipper bags, preserving them at −18 °C until analysis.

2.3. Sample preparation and analysis

An Ethos One microwave oven (Milestone Laboratory Systems, Italy) equipped with a high-pressure rotor was employed for sample acid digestions. Approximately 200–300 mg of freeze-dried fresh juice samples were weighted and placed within Teflon closed vessels to later add 2 mL of HNO_3 (65 %, w/v) and 1 mL of H_2O_2 (30 %, w/v). The microwave digester was operated under the following conditions: an initial 1-min interval at 250 W, followed by 1 min with zero power; then 5 min at 250 W, succeeded by 5 min at 450 W, 3 min at 600 W, and concluding with 5 min at 300 W. Afterwards, the resulting colorless solution was transferred into a 25 mL volumetric flask and made up to the volume with ultrapure water.

Elemental measurements (Al, Ca, Co, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, Sr, and Zn) were conducted utilizing an Agilent 4210 microwave plasma atomic emission spectrometer (MP-AES, Agilent Technologies, Japan). The sample introduction system comprised solvent-resistant tubing, a double-pass cyclonic chamber, and an inert flow-blurring nebulizer (OneNeb). The operating parameters were configured as follows: microwave frequency at 2450 MHz, magnetron power set at 1 kW, employment of nitrogen as the plasma gas with a flow rate of 20 L/min, nebulizer pressure ranging from 140 to 240 kPa, sample uptake duration of 8 s, stabilizing time of 15 s, rinsing time of 30 s, and a reading time of 5 s. All instrumental parameters were optimized daily while aspirating the tuning solution. The spectral lines selected for the measurement of the individual elements were as follows: 396.152 for Al, 393.366 for Ca, 340.512 for Co, 425.433 for Cr, 324.754 for Cu, 371.993 for Fe, 766.491 for K, 285.213 for Mg, 403.076 for Mn, 379.825 for Mo, 588.995 for Na, 352.454 for Ni, 407.771 for Sr and 213.857 for Zn. A Genius SQ 24 nitrogen gas generator (Peak Scientific Instruments Ltd., Scotland, UK) was used to produce high-purity nitrogen gas from ambient air. Background correction procedures were automatically and individually optimized for each element using the Agilent MP Expert software (Agilent Technologies, Santa Clara, CA, USA), provided alongside the MP-AES instrument.

2.4. Method validation

The proposed MP-AES method was validated by assessing analytical performance parameters, including linearity, sensitivity, precision, and recovery. To achieve this, we designed an experimental plan involving the preparation of five calibration curves at different concentration levels, which were obtained by diluting a TraceCERT® multi-element standard solution with 1 % nitric acid. The coefficients of determination (R^2) were obtained by least squares linear regression analysis. The limits of detection (LOD) and limits of quantification (LOQ) were calculated following the guidelines outlined by AOAC (Latimer, 2016). Specifically, LOD = SD × 3/S and LOQ = SD × 10/S, where SD represents the standard deviation from measurements of 10 blank solutions, and S signifies the slope of the calibration curve. Concerning precision and recovery assays, samples were fortified with the target elements at concentration levels of 2.5 and 5.0 mg kg^{-1}, and subsequently analyzed by MP-AES after undergoing digestion. The experiments were conducted within a single day for the intra-day repeatability test and on three separate days for the inter-day repeatability test. The variability of determinations was expressed as the relative standard deviation (% RSD). Recovery, expressed as a percentage (%), was calculated by determining the ratio between the measured quantity of elements and the total fortified amount. All results were expressed as an average of five measurements.

To verify the reliability of the multi-elemental data obtained by MP-AES, we fortified and analyzed a total of 10 randomly selected samples using an ICP-MS method previously reported by Gao et al. (2022), which served as a reference technique. The data obtained by both analytical methodologies were statistically compared using a one-way ANOVA analysis with the aim of proposing the MP-AES method as a feasible alternative for mineral determination in mandarin juices.

2.5. Reference method

An Agilent 7700 Series ICP-MS system (Agilent Technologies, USA), equipped with a 27.12 MHz radiofrequency solid-state generator operating at 1500 W, was employed as a reference technique to verify the accuracy of the multi-elemental data obtained by the MP-AES method. The setup included a MicroMist™ nebulizer, a cooled double-pass quartz spray chamber, and a Fassel-type ICP torch composed of a three-cylinder assembly with a 2.5 mm injector diameter. Operating conditions were set as follows: a plasma gas flow rate of 15 L min^{-1}, collision and reaction gas flow rate for helium at 4 mL min^{-1}, a carrier gas flow rate of 0.8 L min^{-1} (argon purity exceeding 99.99 %), a sample lifting rate of 0.3 r s^{-1}, and an atomization chamber temperature of 2 °C (Gao et al., 2022). Yttrium was used as internal standard, manually added to all solutions to get a concentration of 10 μg L^{-1}. Instrumental parameters were optimized daily while aspirating the tuning solution.

2.6. Elemental data modeling

Elemental data obtained by MP-AES were fitted into a matrix format X (202 × 14), in which the rows represented the 202 mandarin fruit samples, and the columns, the concentrations of elements analyzed (Al, Ca, Co, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, Sr, and Zn). The data matrix was initially preprocessed through autoscaling and mean centering to mitigate issues associated with the varying concentration scales of the elemental variables. The non-parametric Kruskal-Walli's test was also employed to identify potential discriminant variables depending on the statistical differences observed among mandarin provenances. Later, the unsupervised dimension-reducing principal component analysis (PCA) was carried out to explore the natural distribution of samples within a reduced dimensional space and look for clustering trends among classes of samples (Bro & Smilde, 2014).

For further analysis, we employed the hold-out method to partition the database into two representative subsets, assigning a 70/30 ratio for the training and testing stages, respectively. The training set was utilized for constructing the classifiers, while the test set served to validate them. This method is a straightforward and computationally efficient approach widely chosen for assessing the performance of classification models. Besides, it involves an iterative process that can be executed multiple times by randomly selecting distinct training and testing subsets in a stratified manner. This ensures that the distributions of the subsets closely resemble those of the original dataset, thereby guaranteeing the development of a generalized model that robustly fits the input data.

Considering the previous partitioning of subsets, we developed and validated four classification models, including linear discriminant analysis (LDA), k nearest neighbor (k-NN), support vector machine (SVM), and random forest (RF). LDA is a supervised algorithm designed to maximize the variance between categories while minimizing the variance within categories. The method is founded on estimating several canonical or discriminant functions, representing linear combinations of the original variables that allow achieving separation. K-NN is a discrimination algorithm focusing on distances between objects, particularly the closest ones. It classifies unknown samples by projecting them into a multivariate space and assigning them to their nearest neighbor class. The variable k is a hyperparameter requiring optimization since it stands the number of neighbors considered to assign the class of unknown samples through majority voting. The selection of the k value significantly influences the model's outcomes. SVM is a supervised machine learning algorithm that operates by identifying a hyperplane in a high-dimensional feature space maximizing the margin, which is the distance between the hyperplane and the nearest data points from each class, known as support vectors. The model classifies the data into different categories by separating it through this hyperplane. To accomplish robust predictions, the process typically involves optimizing internal hyperparameters significantly affecting the outcomes. RF is an ensemble learning algorithm based on the building of multiple decision trees that employ a majority voting approach for classification predictions. It assesses the number of votes for each class across all decision trees, and the class with the highest number of votes is selected as the final prediction for the new sample. The choice of the number of decision trees is a crucial hyperparameter to consider in the model for avoiding data overfitting or underfitting.

Classification modeling performance was assessed by calculating the ratios among all correct predictions and the total number of examined cases (accuracy), correct positive predictions and the number of positive cases (sensitivity), and correct negative predictions and number of negative cases (specificity) (Ballabio et al., 2018). All these metrics were expressed as a percentage (%). R-project software version 4.2.3 (Team, 2023) was employed for elemental data analysis.

3. Results and discussion

3.1. Analytical performance features

Table 1 summarizes the main figures of merit and elemental composition resulting from the analysis of mandarin juices using MP-AES. The calibration curves exhibited coefficients of determination (R^2) exceeding 0.9990, signifying a strong linear relationship within the concentration range established for each element. The LOD and LOQ values were in the ranges of 0.003–2.081 and 0.013–5.848 mg kg^{-1}, respectively. These results underscore the proposed MP-AES method's good linearity and sensitivity.

Recovery data from known samples fortified with standard elements are presented in Table 2. The average recoveries obtained at two concentration levels ranged from 93.35 to 108.67 % for the MP-AES method and from 94.87 to 106.01 % for the ICP-MS method. In terms of measurement precision with MP-AES, intra-day RSD (%) values were below 4.7 %, while those inter-day were less than 6.5 %. In accordance with AOAC regulations (Latimer, 2016), these results meet the acceptance

Table 1
Limits of detection (LOD), limits of quantification (LOD), and elemental concentrations (mg kg^{-1}) determined in mandarin juice samples by MP-AES method.

Element	LOD	LOQ	Sampling mandarin fruits (mg kg^{-1})					
	(mg kg $^{-1}$)	(mg kg $^{-1}$)	CTE (n = 56)		ETR (n = 76)		MIS (n = 70)	
			Average	Min–Max	Average	Min–Max	Average	Min–Max
Al	0.23	0.74	10.7	4.2–15.8	9.9	4.1–16.7	5.6	3.9–10.2
Ca	2.08	5.85	82.4	50.1–108.0	73.2	38.2–104.0	55.1	39.1–71.3
Co	0.007	0.013	0.17	0.14–0.21	0.13	0.05–0.21	0.20	0.14–0.28
Cr	0.004	0.011	0.81	0.49–1.06	0.90	0.48–1.21	1.03	0.71–1.32
Cu	0.05	0.13	2.07	0.87–3.36	1.73	0.52–3.16	2.01	0.99–3.28
Fe	0.39	0.51	1.38	0.84–1.75	1.12	0.64–1.69	1.32	0.92–1.72
K	0.19	0.73	1881	1326–2518	2048	1346–2598	1916	1352–2558
Mg	0.08	0.11	288.3	168.2–382.1	227.0	125.0–373.2	400.4	317.1–489.2
Mn	0.009	0.014	0.70	0.29–1.03	0.54	0.17–1.01	0.67	0.33–1.09
Mo	0.003	0.019	0.08	0.05–0.11	0.07	0.04–0.11	0.10	0.08–0.13
Na	0.22	0.54	4.51	2.73–6.01	4.33	2.51–5.93	4.52	2.60–5.91
Ni	0.010	0.013	0.04	0.02–0.23	0.05	0.02–0.42	0.04	0.02–0.42
Sr	0.029	0.083	4.26	1.76–7.01	1.90	0.65–6.77	1.12	0.46–1.96
Zn	0.20	0.47	1.97	1.27–2.54	1.21	0.70–2.55	2.23	1.91–2.74

Table 2
Recovery data obtained using the proposed MP-AES and reference ICP-MS methods.

Element	Fortified amount	Recovery (%)[a]	
	(mg kg^{-1})	MP-AES	ICP-MS
Al	2.5	93.3 ± 1.2	95.0 ± 1.1
	5.0	108.7 ± 0.6	106.0 ± 1.4
Ca	2.5	102.9 ± 1.5	100.3 ± 0.9
	5.0	100.3 ± 0.9	104.6 ± 1.4
Co	2.5	99.0 ± 1.1	96.7 ± 1.9
	5.0	101.6 ± 0.7	98.2 ± 1.8
Cr	2.5	104.6 ± 0.5	102.1 ± 1.6
	5.0	96.5 ± 1.2	94.8 ± 0.9
Cu	2.5	102.1 ± 1.2	100.3 ± 2.0
	5.0	104.3 ± 0.4	103.9 ± 1.6
Fe	2.5	106.8 ± 1.4	102.1 ± 2.0
	5.0	101.1 ± 1.3	99.2 ± 1.0
K	2.5	99.9 ± 1.8	103.9 ± 0.9
	5.0	104.9 ± 1.2	106.0 ± 0.8
Mg	2.5	98.8 ± 1.0	95.1 ± 2.2
	5.0	101.0 ± 0.7	104.0 ± 1.3
Mn	2.5	99.0 ± 0.8	97.9 ± 1.8
	5.0	100.2 ± 0.9	99.8 ± 1.9
Mo	2.5	100.0 ± 1.3	104.8 ± 0.9
	5.0	103.9 ± 1.0	102.0 ± 1.4
Na	2.5	102.9 ± 1.1	99.4 ± 2.0
	5.0	101.1 ± 0.9	103.5 ± 1.1
Ni	2.5	104.8 ± 1.5	101.0 ± 2.2
	5.0	100.7 ± 0.8	99.0 ± 1.8
Sr	2.5	102.4 ± 1.3	100.7 ± 2.8
	5.0	98.5 ± 0.9	99.6 ± 1.0
Zn	2.5	99.4 ± 1.4	103.9 ± 0.9
	5.0	98.9 ± 1.3	102.1 ± 2.0

[a] Average ± SD corresponding to five measurements (n = 5).

criteria established for intermediate precision (≤15 % RSD) and recovery (80–110 %). The obtained data highlights the effectiveness and reliability of the proposed method (MP-AES) for the multielemental analysis of mandarin juices when compared to the reference method (ICP-MS).

3.2. Elemental composition

The elemental profiles of mandarin juice samples from three different production regions are shown in Table 1. The concentrations of measured elements can be arranged in the following abundance order: K > Mg > Ca > Al > Na > Sr > Cu > Zn > Fe > Cr > Mn > Co > Mo > Ni. The multi-element data obtained indicates that potassium was the most abundant element in the samples, with concentrations ranging from 1326 to 2598 mg kg^{-1}. These results align with previous studies discussing the mineral content in citrus juice samples (Pérez-López et al., 2007; Ruggiero et al., 2021). The highest levels of potassium were observed in the ETR juice samples (2048 mg kg^{-1}), while the lowest levels were found in the CTE juice samples (1881 mg kg^{-1}). Potassium plays a crucial role in the production and quality of citrus fruits, including their juice. It regulates various enzyme functions and plant water relations, as well as the electrochemical balance of cells, protein synthesis, and the maintenance of cell turgor and extensibility, which impact fruit size (Alva et al., 2006).

Magnesium serves as a cofactor and activator for various enzymes, prominently regulating those involved in glycolysis. Additionally, it contributes to the protein and DNA synthesis (Baaij et al., 2015). The highest Mg content was observed in the juice samples from MIS (400 mg kg^{-1}), while the lowest was found in juices from ETR (227 mg kg^{-1}) samples. Notably, Mg content in mandarin juice is lower than that reported in a previous study on different Spanish mandarin varieties (Beltrán-González et al., 2008). Calcium is essential as a structural component for bones and teeth, and it is vital in muscle contraction and blood clotting (Beto, 2015). The juice from CTE samples had the highest Ca content (82 mg kg^{-1}), followed by ETR samples (73 mg kg^{-1}) and MIS samples (55 mg kg^{-1}). These findings are consistent with a previous work on mandarin juice conducted by Beltrán-González et al. (2008).

The average values of minor and trace elements, such as Na, Sr, Cu, Zn, and Fe, ranged from 1 to 10 mg kg^{-1}, while the concentrations of Cr, Mn, Co, Mo, and Ni were below 1 mg kg^{-1}. The CTE samples exhibited the highest concentrations of Cu, Fe, Mn, and Sr, whereas the MIS samples showed the highest concentrations of Co, Cr, Mo, and Zn. By comparing with other natural or commercial citrus juices, the concentrations of Cr, Mn, Co, Ni, Cu, Zn, and Sr were consistent with those reported in the literature (Ajai et al., 2014).

The content of both macro and microelements in mandarin juice can vary significantly depending on its geographical origin. These differences arise from a complex interplay of factors. Soil composition plays a pivotal role, with regions featuring distinct mineral-rich soils likely to yield fruits with higher concentrations of those minerals. Climate and weather patterns further contribute to these variations, affecting nutrient absorption and health of plants. Additionally, irrigation water quality, agricultural practices, and botanic distinctions influence mineral content. Geological features and environmental pollution levels in an area also impact fruit quality, as do harvesting and post-harvest handling practices (Elbasiouny et al., 2022; Schoonover & Crim, 2015).

3.3. Exploratory data analysis

To explore variations in mineral content among mandarin juice samples originating from different production regions, we conducted several statistical analyses, including the Kruskal-Walli's test, one-way ANOVA, and PCA, considering the multielemental data obtained. The ANOVA test revealed significant differences ($p < 0.001$) in the concentrations of Al, Ca, Co, Cr, Fe, Mg, Mo, Sr, and Zn ($p < 0.001$), Cu, Mn ($p < 0.01$), and K ($p < 0.05$) in the juice depending on the geographical origins studied. In contrast, the levels of Na and Ni were not significantly different ($p > 0.05$). Further statistic tests, using Kruskal–Wallis all-pairwise comparisons ($\alpha = 0.05$), confirmed that the concentrations of Ca, Co, Cr, Mg, Mo, Sr, and Zn in juice samples from the three production regions exhibited significant differences. These findings suggest that the elemental content in mandarin juice samples from various sources can provide valuable insights for developing models focused on sample authentication.

We then conducted PCA to identify patterns or relationships among samples and variables within the original dataset. PCA facilitates data projection from a higher-dimensional space to a lower-dimensional one and enables reconstruction of them without prior assumptions about their distribution. The cumulative variance contribution of the first seven principal components (PCs) was 73.7 %.

Fig. 1 presents the biplot generated using PC1 and PC2, offering a clearer visualization of the distribution of mandarin juice samples in relation to their geographical origin. The dominant elements influencing PC1 were Mg, Zn, Mo, and Co, while PC2 was mainly influenced by Sr and Cr. As illustrated in the biplot, the samples tend to cluster based on their origin, forming three distinct groups. MIS samples are more concentrated, while ETR and CTE samples partially overlap. In general, MIS samples are characterized by elevated concentrations of Mg, Mo, Zn, Cr, and Co, and low levels of Al and Ca. CTE samples exhibit higher levels of Al, Sr, Fe, Cu, Mn, and Ca, while ETR samples were distinguished by lower concentrations of Fe, Mn, and Cu, alongside higher levels of K. Although PCA effectively visually discriminates the geographical origin of mandarin juices, employing supervised predictive models is essential for achieving a comprehensive classification of the samples into their three regional origins.

3.4. Geographical origin prediction of mandarin juices

As an approximation method to enhance the distinguishability of mandarin juices, PCA was applied to the input dataset. This technique extracts categorical data and compresses the feature space by projecting samples into a discriminant vector space. It maximizes the inter-class distance and minimizes the intra-class distance in this new subspace, ensuring optimal discrimination among different samples. Finally, samples are classified depending on their distances from the center point of each class.

LDA discriminant function was established using seven characteristic elements: Al, Cr, Fe, K, Mg, Mo, and Zn, which were selected through PCA screening. Function 1 accounted for 87.0 % of the variance, while function 2 accounted for 13.0 %. Fig. 2 depicted scatter plot made considering function 1 vs. function 2. The results showed that among the 81 samples from ETR, 16 were misclassified as CTE, and one as MIS, resulting in a correct discrimination rate of 79.0 %. Among the 51 samples from CTE, 12 were misclassified as ETR, and 2 as MIS, leading to a success rate of 72.5 %. MIS exhibited the highest prediction rate at 98.6 %, with only one sample misclassified as CTE among 70 samples. The overall classification accuracy for the three regions was 83.3 %. These results suggest that mandarin juice samples from the MIS province can be classified with an acceptable level of accuracy using the LDA method. However, it's important to note that some samples from the CTE and ETR provinces were not exclusively distinguishable based on their origin. This observation may be attributed to the geographical proximity of these regions and their shared climatic conditions during fruit production. These similarities make it challenging to detect characteristic differences in elemental levels. As an outcome, there is a need to explore nonparametric classification methods to address the complexity of high-dimensional space and the presence of overlapping samples.

Implementing non-linear pattern recognition tools is a common approach for addressing food authentication challenges (Rocha et al., 2020). In this context, we evaluated k-NN, SVM, and RF algorithms for classifying mandarin juices based on their multi-element composition. These methods can handle the complexity of high-dimensional data and

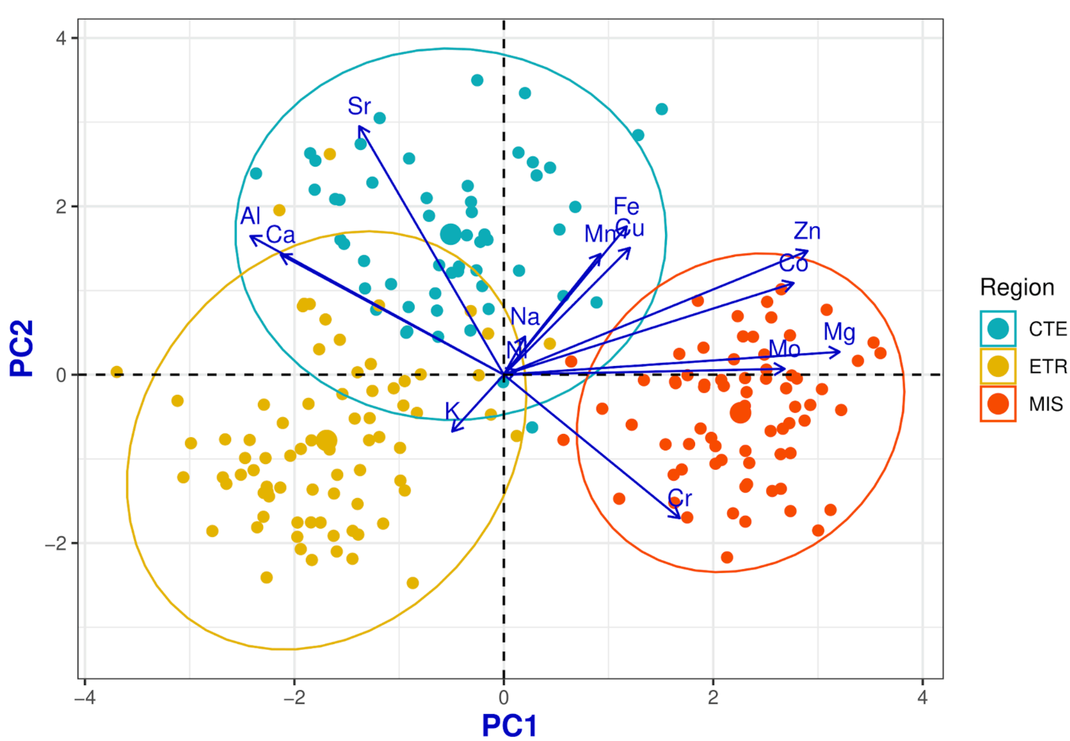

Fig. 1. Projection of scores and loadings of the original variables onto the PC1-PC2 plane constructed from the mandarin juice samples.

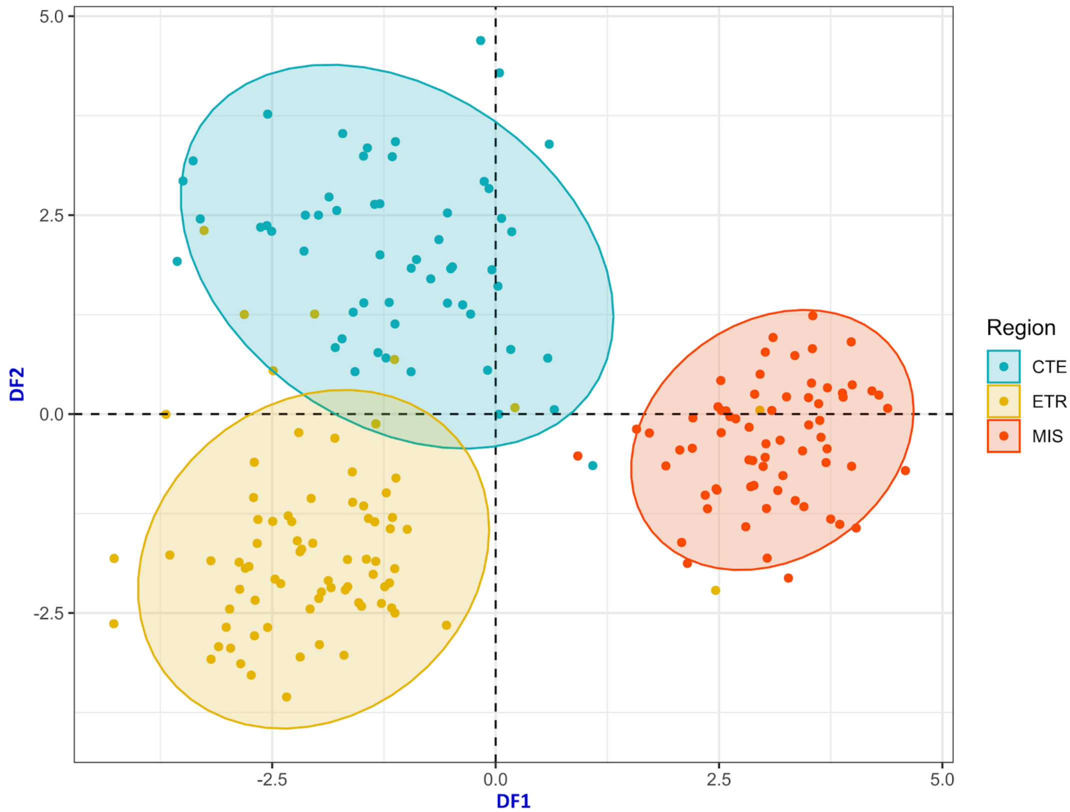

Fig. 2. Separation chart of the mandarin juice samples based on their geographical origin considering the space defined by the first two discriminant functions.

enhance the performance of classification models. For comparison purposes, we performed 10-fold cross-validation (10-fold CV) method. This approach is effective for estimating the predictive capabilities of classification models on unseen data. By using this method, the data is divided into subsets, with one subset used for training while the model is tested on the remaining data. This process is repeated multiple times, each time using a different subset as the testing set. The results from each test step are averaged to obtain a more reliable estimate of the model's performance. This helps us assess how the model will generally perform when making predictions for data that were not part of the model's training set. Additionally, this iterative process allowed us to optimize the hyperparameters for each algorithm. Table 3 presents the overall accuracy obtained through 10-fold CV for different classification models. Herein, it is evident that there are significant differences in the global classification accuracy of the three models: 64.0 % for k-NN, 96.4 % for SVM, and 93.2 % for RF. The optimized hyperparameters for each algorithm were as follows: k = 5 for k-NN, penalty factor = 32 and insensitive loss function = 0.025 for SVM, and number of trees = 500 and number of features tried at any given split = 3.74 for RF.

Finally, we applied the built classifiers to new subsets obtained by splitting the original data into a 70/30 ratio for training and testing, respectively. The results of the geographical origin classification of mandarin juices are presented in Table 4. All three algorithms achieved accuracies ranging between 94.9 and 100 % on the training samples. Likewise, when applied to the test samples, the overall classification performance consistently exceeded 90 %, with SVM being the most promising model upon achieving a prediction rate of 95.1 %. The k-NN model maintained consistent behavior in both calibration and validation predictions. Despite the RF model being excellently trained, it exhibited the lowest classification rate on the test samples. On his part, sensitivities ranged from 82.4 to 100 %, and specificities ranged from 92.5 to 100 %. Sensitivity, i.e., the true positive rate or recall, measures the model's ability to correctly identify positive instances among all actual positives, while specificity, the true negative rate, evaluates the model's

Table 3
Confusion matrix obtained for the classification of mandarin juices using the studied algorithms via 10-fold cross-validation method.

	Region	Predicted group membership for k-NN			Accuracy (%)
		Predicted class			
		CTE	ETR	MIS	
Observed class	CTE	17	22	17	30.4
	ETR	14	60	2	78.9
	MIS	9	3	58	82.8
	Overall accuracy (%)				64.0
	Region	Predicted group membership for SVM			Accuracy (%)
		Predicted class			
		CTE	ETR	MIS	
Observed class	CTE	53	3	–	94.6
	ETR	4	72	–	94.7
	MIS	–	–	70	100
	Overall accuracy (%)				96.4
	Region	Predicted group membership for RF			Accuracy (%)
		Predicted class			
		CTE	ETR	MIS	
Observed class	CTE	52	3	1	92.9
	ETR	9	67	–	88.2
	MIS	–	1	69	98.6
	Overall accuracy (%)				93.2

capacity to accurately identify negative instances among all actual negatives. These metrics are crucial for a comprehensive assessment of classification models, because in addition to offering a nuanced understanding of their effectiveness in handling positive and negative class

Table 4
Performance of the classification models for determining the geographical origin of mandarin juice samples.

Model		k-NN					
	Region	Training			Test		
		Predicted class			Predicted class		
		CTE	ETR	MIS	CTE	ETR	MIS
Observed class	CTE	36	1	1	14	–	–
	ETR	2	50	–	2	21	–
	MIS	1	2	48	1	2	21
Sensitivity (%)		92.3	94.3	98.0	82.4	91.3	100
Specificity (%)		98.0	97.7	96.7	100	94.7	92.5
Accuracy (%)		94.9			91.2		

Model		SVM					
	Region	Training			Test		
		Predicted class			Predicted class		
		CTE	ETR	MIS	CTE	ETR	MIS
Observed class	CTE	39	1	1	16	2	–
	ETR	–	52	–	–	21	–
	MIS	–	–	49	1	–	21
Sensitivity (%)		100	98.1	98.0	94.1	91.3	100
Specificity (%)		98.1	100	98.9	95.5	100	97.4
Accuracy (%)		98.7			95.1		

Model		RF					
	Region	Training			Test		
		Predicted class			Predicted class		
		CTE	ETR	MIS	CTE	ETR	MIS
Observed class	CTE	39	–	–	14	2	1
	ETR	–	53	–	2	19	–
	MIS	–	–	49	1	2	20
Sensitivity (%)		100	100	100	82.4	82.6	95.2
Specificity (%)		100	100	100	93.2	94.7	92.5
Accuracy (%)		100			86.7		

distinctions, they provide noteworthy insights into the model performance in distinguishing between classes. This is particularly relevant in scenarios with imbalanced class distributions, as observed in our work.

The discrepancies observed in calculating the previously discussed metrics can be attributed to the high degree of overlap in the chemical information of the samples from distinct provinces. Consequently, SVM emerged as the optimal classifier for authenticating the geographical origin of mandarin juice samples. This model effectively detected variations in the elemental composition of the samples, attributing them to specific production regions and contributing to the protection of their designation of origin. These findings highlight the discriminative capabilities of the MP-AES technique in identifying the provenance of mandarin juice samples, presenting it as a viable alternative for addressing issues related to geographic origin authentication.

Over the last few years, MP-AES has gained increasing prominence as a powerful analytical tool for diverse material analyses. In comparison to ICP-MS, MP-AES offers distinct advantages. It is cost-effective, user-friendly, and less susceptible to matrix effects, making it suitable for complex samples. MP-AES can simultaneously analyze multiple elements, reducing both analysis time and consumable usage. Furthermore, it operates at lower temperatures and power levels, enhancing safety and environmental friendliness while demanding less maintenance. These distinctions fundamentally arise from ionization/atomization sources required per each technique. Particularly, MP-AES employs nitrogen plasma, while ICP-MS uses argon plasma. Nitrogen plasma is generated at atmospheric pressure through a lower-pressure microwave source, utilizing readily available atmospheric air via a compressor. In contrast, argon, especially when high-purity argon is required, is a relatively expensive gas, significantly impacting laboratory operational costs. Despite the above, it is important to note that the choice between MP-AES and ICP-MS should be guided by specific analytical needs, sensitivity requirements, and budget constraints.

4. Conclusions

This paper proposes using the mineral profiles obtained through MP-AES for the geographical origin identification of mandarin juices via pattern recognition modeling. The authenticity of juices was assessed by constructing multiple models, including PCA, LDA, k-NN, SVM, and RF, designed for discriminating samples based on their multielemental compositions. Notably, overall classification rates ranged from 86.7 to 95.1 % in the test subsets, with SVM emerging as the most proficient classifier. Furthermore, it exhibited sensitivities and specificities within an acceptable range, highlighting its robustness in authenticating mandarin juice samples. The proposed model demonstrates the capability of mineral profiles in discerning the genuine origin of mandarin juices, which prevents mislabeling fraud and contributes to consumer protection.

CRediT authorship contribution statement

Melisa Jazmin Hidalgo: Investigation, Methodology, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **José Emilio Gaiad:** Methodology, Formal analysis, Data curation. **Héctor Casimiro Goicoechea:** Project administration, Resources, Writing – review & editing. **Alberto Mendoza:** Funding acquisition, Supervision, Writing – review & editing. **Michael Pérez-Rodríguez:** Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Roberto Gerardo Pellerano:** Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors are grateful for the financial assistance provided by the Secretaría General de Ciencia y Técnica (SGCyT - UNNE), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), and Tecnologico de Monterrey, which greatly supported our work.

References

Ajai, A. I., Ochigbo, S. S., Abdullahi, Z., & Anigboro, P. I. (2014). Determination of trace metals and essential minerals in selected fruit juices in Minna, Nigeria. *International Journal of Food Science*, 1–5.
Alva, A. K., Mattos, D., Paramasivam, S., Patil, B., Dou, H., & Sajwan, K. S. (2006). Potassium management for optimizing citrus production and quality. *International Journal of Fruit Science, 6*, 3–43.
Aung, M. M., & Chang, Y. S. (2014). Traceability in a food supply chain: Safety and quality perspectives. *Food Control, 39*, 172–184.
Baaij, J. H., Hoenderop, J. G. J., & Bindels, R. J. M. (2015). Magnesium in Man: Implications for Health and Disease. *Physiological Reviews, 95*, 1–46.
Balaram, V. (2020). Microwave plasma atomic emission spectrometry (MP-AES) and its applications – A critical review. *Microchemical Journal, 159*, Article 105483.
Ballabio, D., Grisoni, F., & Todeschini, R. (2018). Multivariate comparison of classification performance measures. *Chemometrics and Intelligent Laboratory Systems, 174*, 33–44.
Beltrán-González, F., Pérez-López, A. J., López-Nicolás, J. M., & Carbonell-Barrachina, Á. A. (2008). Effects of agricultural practices on instrumental colour,

mineral content, carotenoid composition, and sensory quality of mandarin orange juice, cv. Hernandina. *Journal of the Science of Food and Agriculture, 88*(10), 1731–1738.

Beto, J. A. (2015). The Role of Calcium in Human Aging. *Clinical Nutrition Research, 4*(1), 1–8.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods, 6*(9), 2812–2831.

Carter, J. A., Long, C. S., Smith, B. P., Smith, T. L., & Donati, G. L. (2019). Combining elemental analysis of toenails and machine learning techniques as a non-invasive diagnostic tool for the robust classification of type-2 diabetes. *Expert Systems with Applications, 115*, 245–255.

Cheng, C., Jia, M., Gui, Y., & Ma, Y. (2020). Comparison of the effects of novel processing technologies and conventional thermal pasteurisation on the nutritional quality and aroma of Mandarin (Citrus unshiu) juice. *Innovative Food Science and Emerging Technologies, 64*, Article 102425.

Cheng, Y., Han, L., Huang, L., Tan, X., Wu, H., & Li, G. (2023). Association between flavor composition and sensory profile in thermally processed mandarin juices by multidimensional gas chromatography and multivariate statistical analysis. *Food Chemistry, 419*, Article 136026.

Czech, A., Zarycka, E., Yanovych, D., Zasadna, Z., Grzegorczyk, I., & Kłys, S. (2020). Mineral Content of the Pulp and Peel of Various Citrus Fruit Cultivars. *Biological Trace Element Research, 193*(2), 555–563.

De Carvalho Rocha, W. F., Do Prado, C. B., & Blonder, N. (2020). Comparison of chemometric problems in food analysis using non-linear methods. *Molecules, 25*(13), 3025.

Dias, C., & Mendes, L. (2018). Protected Designation of Origin (PDO), Protected Geographical Indication (PGI) and Traditional Speciality Guaranteed (TSG): A bibiliometric analysis. *In Food Research International, 103*, 492–508.

Elbasiouny, H., El-Ramady, H., Elbehiry, F., Rajput, V. D., Minkina, T., & Mandzhieva, S. (2022). Plant Nutrition under Climate Change and Soil Carbon Sequestration. *Sustainability, 14*(2), 914.

Gallego Ríos, S. E., Peñuela, G. A., & Ramírez Botero, C. M. (2017). Method Validation for the Determination of Mercury, Cadmium, Lead, Arsenic, Copper, Iron, and Zinc in Fish Through Microwave-Induced Plasma Optical Emission Spectrometry (MIP OES). *Food Analytical Methods, 10*(10), 3407–3414.

Gao, F., Hao, X., Zeng, G., Guan, L., Wu, H., Zhang, L., … Li, H. (2022). Identification of the geographical origin of Ecolly (Vitis vinifera L.) grapes and wines from different Chinese regions by ICP-MS coupled with chemometrics. *Journal of Food Composition and Analysis, 105*, Article 104248.

Goldenberg, L., Yaniv, Y., Porat, R., & Carmi, N. (2018). Mandarin fruit quality: A review. *Journal of the Science of Food and Agriculture, 98*(1), 18–26.

Guo, H., Zheng, Y. J., Wu, D. T., Du, X., Gao, H., Ayyash, M., … Gan, R. Y. (2023). Quality evaluation of citrus varieties based on phytochemical profiles and nutritional properties. *Frontiers in Nutrition, 10*, 1165841.

Helmeczi, W., Helmeczi, E., Baker, L. A., Wang, Y., & Brindle, I. D. (2018). Development of a general acid method for the digestion of gold ore samples together with a comparison of extraction solvents for gold and determination by microwave-induced plasma-atomic emission spectrometry (MIP-AES). *Journal of Analytical Atomic Spectrometry, 33*(8), 1336–1344.

Jandrić, Z., & Cannavan, A. (2017). An investigative study on differentiation of citrus fruit/fruit juices by UPLC-QToF MS and chemometrics. *Food Control, 72*, 173–180.

Jungen, M., Schütz, B., & Schweiggert, R. (2020). Influence of species and processing techniques on phlorin in Citrus juices as quantified by 1H-NMR spectroscopy. *LWT, 134*, Article 109949.

Kamala, C. T., Balaram, V., Dharmendra, V., Subramanyam, K. S. V., & Krishnaiah, A. (2014). Application of Microwave Plasma Atomic Emission Spectrometry (MP-AES) for environmental monitoring of industrially contaminated sites in Hyderabad City. *Environmental Monitoring and Assessment, 186*(11), 7097–7113.

King, T., Cole, M., Farber, J. M., Eisenbrand, G., Zabaras, D., Fox, E. M., & Hill, J. P. (2017). Food safety for food security: Relationship between global megatrends and developments in food safety. *Trends in Food Science and Technology, 68*, 160–175.

Latimer, G. W., Jr (2016). Guidelines for standard method performance requirements: Official methods of analysis. *AOAC International*, 1–18.

Li, W., Simmons, P., Shrader, D., Herrman, T. J., & Dai, S. Y. (2013). Microwave plasma-atomic emission spectroscopy as a tool for the determination of copper, iron, manganese and zinc in animal feed and fertilizer. *Talanta, 112*, 43–48.

Li, Z., Jin, R., Yang, Z., Wang, X., You, G., Guo, J., … Pan, S. (2021). Comparative study on physicochemical, nutritional and enzymatic properties of two Satsuma mandarin (Citrus unshiu Marc.) varieties from different regions. *Journal of Food Composition and Analysis, 95*, Article 103614.

Pérez-López, A. J., López-Nicolás, J. M., & Carbonell-Barrachina, A. A. (2007). Effects of organic farming on minerals contents and aroma composition of Clemenules mandarin juice. *European Food Research and Technology, 225*(2), 255–260.

Poirier, L., Nelson, J., Gilleland, G., Wall, S., Berhane, L., & Lopez-Linares, F. (2017). Comparison of Preparation Methods for the Determination of Metals in Petroleum Fractions (1000°F+) by Microwave Plasma Atomic Emission Spectroscopy. *Energy and Fuels, 31*(8), 7809–7815.

Qiu, S., & Wang, J. (2015). Effects of storage temperature and time on internal quality of satsuma Mandarin (Citrus unshiu marc.) by means of E-nose and E-tongue based on two-way MANOVA analysis and random forest. *Innovative Food Science and Emerging Technologies, 31*, 139–150.

Qiu, S., Wang, J., & Du, D. (2017). Assessment of high pressure processed mandarin juice in the headspace by using electronic noise and chemometric analysis. *Innovative Food Science and Emerging Technologies, 42*, 33–41.

Rosales Martínez, V., Francisco Rubio, A., Casanova Pérez, L., Fraire Cordero, S., Flota Bañuelos, C., & Galicia Galicia, F. (2020). Citrus farmers' perception of the effect of climate change in Campeche. *Revista mexicana de ciencias agrícolas, 11*(4), 727–740.

Ruggiero, L., Fontanella, M. C., Amalfitano, C., Beone, G. M., & Adamo, P. (2021). Provenance discrimination of Sorrento lemon with Protected Geographical indication (PGI) by multi-elemental fingerprinting. *Food Chemistry, 362*, Article 130168.

Schoonover, J. E., & Crim, J. F. (2015). An Introduction to Soil Concepts and the Role of Soils in Watershed Management. *Journal of Contemporary Water Research & Education, 154*(1), 21–47.

Team, R. C. (2023). R: A Language and Environment for Statistical Computing. The R Foundation - The R Project for Statistical Computing. Available online: http://www.R-project.org/.