Food Research International 178 (2024) 113975

Contents lists available at ScienceDirect

Food Research International

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

Discriminative power of DNA-based, volatilome, near infrared spectroscopy, elements and stable isotopes methods for the origin authentication of typical Italian mountain cheese using sPLS-DA modeling

Marco Cardin [a,b], Jérôme Mounier [b], Emmanuel Coton [b], Barbara Cardazzo [a], Matteo Perini [c], Daniela Bertoldi [c], Silvia Pianezze [c], Severino Segato [d], Barbara Di Camillo [e,a], Marco Cappellato [e], Monika Coton [b], Lisa Carraro [a], Sarah Currò [a], Rosaria Lucchini [f], Hooriyeh Mohammadpour [a], Enrico Novelli [a,*]

[a] Department of Comparative Biomedicine and Food Science, University of Padova, Viale Università 16, 35020, Legnaro, PD, Italy
[b] Univ Brest, INRAE, Laboratoire Universitaire de Biodiversité et Écologie Microbienne, F-29280 Plouzané, France
[c] Centro Trasferimento Tecnologico, Fondazione Edmund Mach, Via E. Mach, 1, 38098 San Michele all'Adige, Italy
[d] Department of Animal Medicine, Production and Health, University of Padova, Viale Università 16, 35020 Legnaro, PD, Italy
[e] Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova, Italy
[f] Italian Health Authority and Research Organization for Animal Health and Food Safety (Istituto zooprofilattico sperimentale delle Venezie), Viale Università 10, 35020 Legnaro, PD, Italy

ARTICLE INFO

Keywords:
Geographical origin
Cheese authentication
Shotgun metagenomics
Near Infrared Spectroscopy
Element analysis
Stable isotopes

ABSTRACT

Origin authentication methods are pivotal in counteracting frauds and provide evidence for certification systems. For these reasons, geographical origin authentication methods are used to ensure product origin. This study focused on the origin authentication (*i.e.* at the producer level) of a typical mountain cheese origin using various approaches, including shotgun metagenomics, volatilome, near infrared spectroscopy, stable isotopes, and elemental analyses. DNA-based analysis revealed that viral communities achieved a higher classification accuracy rate (97.4 ± 2.6 %) than bacterial communities (96.1 ± 4.0 %). Non-starter lactic acid bacteria and phages specific to each origin were identified. Volatile organic compounds exhibited potential clusters according to cheese origin, with a classification accuracy rate of 90.0 ± 11.1 %. Near-infrared spectroscopy showed lower discriminative power for cheese authentication, yielding only a 76.0 ± 31.6 % classification accuracy rate. Model performances were influenced by specific regions of the infrared spectrum, possibly associated with fat content, lipid profile and protein characteristics. Furthermore, we analyzed the elemental composition of mountain Caciotta cheese and identified significant differences in elements related to dairy equipment, macronutrients, and rare earth elements among different origins. The combination of elements and isotopes showed a decrease in authentication performance (97.0 ± 3.1 %) compared to the original element models, which were found to achieve the best classification accuracy rate (99.0 ± 0.01 %). Overall, our findings emphasize the potential of multi-omics techniques in cheese origin authentication and highlight the complexity of factors influencing cheese composition and hence typicity.

1. Introduction

Dairy products are among the most common products concerned by food frauds. In particular, protected land- and tradition-related labeled cheeses (*e.g.* Protected Designation of Origin or "mountain product") are subjected to food fraud (mainly mislabeling and fraudulent documentation) due to their high economic value. Indeed, consumers are more willing to pay higher prices for traditional and typical mountain cheeses due to their distinctive organoleptic traits but also their more natural and animal-friendly production attributes (Menozzi et al., 2022). For these reasons, geographical authentication methods are used to ensure product origin. In addition to chemical analyses, including

* Corresponding author.
E-mail address: enrico.novelli@unipd.it (E. Novelli).

https://doi.org/10.1016/j.foodres.2024.113975
Received 20 October 2023; Received in revised form 29 December 2023; Accepted 3 January 2024
Available online 5 January 2024

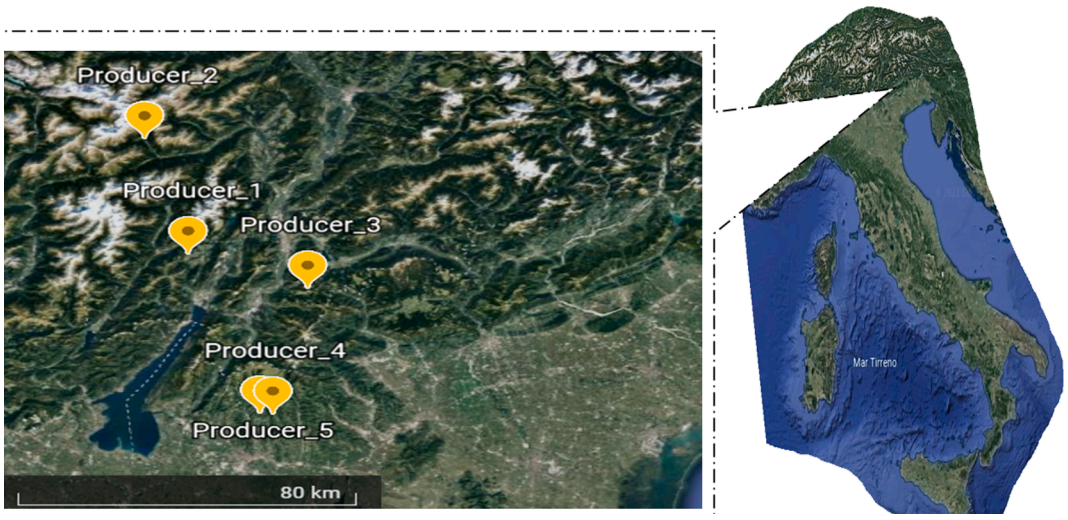

Fig. 1. Map of the origin of mountain Caciotta producers. Producer 1 is located in Giudicarie Esteriori area while Producer 2 and Producer 3 are located in Val di Pejo and Altipiani Cimbri respetively. Producer 4 and Producer 5 are located in the Alti Pascoli della Lessinia area.

stable isotope ratios, trace elements, and fatty acid profiles, emerging methods such as DNA-based methodologies and Near Infrared Spectroscopy are also investigated in this field (Cardin et al., 2022). However, to our knowledge, no comparative study has been conducted on the performances of chemico-physical analysis and DNA-based methods in authenticating cheese origin.

The stable isotope ratios of hydrogen (δ^2H), carbon (δ^{13}C), nitrogen (δ^{15}N), oxygen (δ^{18}O) and sulfur (δ^{34}S) of feed and water are correlated with the local environment, therefore have been widely used to differentiate cheese originating from distinct geographical areas (Bontempo et al., 2011; Bontempo et al., 2019; Camin et al., 2004; Camin et al., 2008; Pianezze et al., 2020). On the other hand, the elemental composition of cheese, in addition to animal breed, plant feed and mineral supplementation, is reflective of geological and pedological traits of the environment (Bontempo et al., 2011; Camin et al., 2007; Camin et al., 2012) as well as of the production process as observed for Greek Graviera (Danezis et al., 2020). These aspects make this approach in combination with stable isotope ratio the current reference method for cheese origin authentication (Camin et al., 2012; Camin et al., 2015).

Near Infrared Spectroscopy (NIRS) is an emerging technique in the origin authenticating field (Segato et al., 2019; Currò et al., 2022; Sammarco et al., 2023; Silva et al., 2022). It generates spectral data ranging from 800 to 2500 nm (near infrared) to 25–1000 μm (far infrared) that contains diverse chemical and physical information (Lei and Sun, 2019). This enables the use of NIRS for authenticating the cheese origin, such as geographical origin, type of milk used, manufacturing process, quality parameters, composition data, and detecting any potential fraud through adulteration (Abbas et al., 2018; Medina et al., 2019).

Lastly, the other two prominent analytical methods in origin authentication are volatilome and microbiota analysis. Volatile organic compound (VOC) analysis has emerged as a prominent method for authenticating cheese origin due to the significance of volatile compounds in defining cheese typicity (Pillonel et al., 2003). These compounds are produced by the metabolic activities of cheese microbiota during glycolysis, proteolysis, and lipolysis, resulting in a diverse array of VOCs. The metabolic activities of cheese microbiota are crucial in shaping the distinct characteristics of cheese types.

Cheese harbors a diverse microbiome that is composed of distinct and complex bacterial, viral, and for some fungal, communities. The diversity of microbial communities is influenced by the type of ecosystem in which they reside (Fierer and Jackson, 2006). Although cheese style is the primary predictor of rind microbiota (Wolfe et al., 2014), dairy farms and cheese-producing plants also play critical roles in defining cheese microbiota, which ultimately affects the quality of traditional cheeses (Goerges et al., 2008; Vacheyrou et al., 2011; Frétin et al., 2018). Microbial ecology studies have also highlighted how the combinations of different environmental factors, and cheese-making conditions and traditional know-how select specific microorganisms. Thus, DNA-based methods applied to microbiome analysis have also been suggested as potential tools to authenticate cheese geographical origin (Kamilari et al., 2019).

In order to discriminate cheese origin based on various qualitative and quantitative data, statistical analyses have to be used. In this context, sparse partial least squares discriminant analysis (sPLS-DA) is a statistical method that can be used to analyze datasets with high dimensionality (large number of features) and identify patterns that can be used to differentiate between different groups or classes (Lê Cao et al., 2011). It is based on partial least squares (PLS) regression, which is a technique that is used to model the relationship between a set of predictor variables (also known as features or characteristics) and a response variable. sPLS-DA is particularly useful when there are more predictor variables than observations, which can be a common problem in -omics datasets.

sPLS-DA works by projecting the data into a lower-dimensional space (called a latent space) in a way that maximizes the separation between the different classes or groups. This is achieved by minimizing the residuals between the observed response variables and the predicted response variables, while also maximizing the variance explained by the latent variables (Chin, 1998). The resulting model can then be used to classify new observations into one of the pre-defined groups based on their feature values. sPLS-DA is a relatively simple and computationally efficient method that has been applied to a wide range of data type, including metagenomics, metagenetics and multi-omics data (Lê Cao et al., 2016). sPLS-DA presents the advantage of being able to handle high-dimensional data and missing values and is relatively easy to interpret compared to some other machine learning techniques (Chung and Keles, 2010).

In this context, the aim of this study was to evaluate the discriminative power of DNA shotgun metagenomics (both bacterial and viral community profiling), volatilome, NIRS, elemental profile and stable isotopes ratio for cheese origin authentication. To do so, we used sPLS-DA models and adopted a case study approach to typical semi-hard raw milk Italian mountain cheese (Caciotta).

2. Methods

2.1. Cheese sampling

In total, 42 Caciotta cheeses were collected in triplicate from five closely situated producers (within a range of 51 ± 26 km) located in the mountainous regions of Trentino Alto-Adige and Veneto. These regions include Alti Pascoli della Lessinia, Giudicarie esteriori, and Trento province areas (Fig. 1). Producer 1 was 38 km, 30 km, 55 km, and 53 km away from producers 2, 3, 4, and 5, respectively. The longest distances were observed for producer 2, who was 60 km, 92 km, and 94 km away from producers 3, 5, and 4, respectively. Lastly, producer 3 was 40 km and 42 km away from producers 4 and 5. Considering the origin altitudes, producers 2 and 3 were located at 1162 m and 1169 m above sea level, respectively, while producers 1, 4, and 5 were located at 628 m, 640 m, and 588 m above sea level, respectively. The geographical coordinates are as follow, 46.02 N and 10.822 E; 46.353 N and 10.69 E; 45.918 N and 11.189 E; 45.556 N and 11.077 E; 45.558 N and 11.038 E, for producers 1 to 5 respectively. The sites 4 and 5 have an average distance from the Garda lake and from the Adriatic sea of 27 km and 100 km respectively. For the sites 1–3 the aforesaid average distance (calculated from the barycenter of a polygon that connects them) is of 28 km and 135 km, respectively.

Cheeses, of 2020 and 2021, were sampled at the end of the ripening period (60 ± 14 days). Approximately 500 g of cheese were sampled from each producer. The obtained sample was divided into aliquots; 300 g were employed for NIR spectroscopy analysis, 150 g for stable isotope and trace elements analyses. The remaining core cheese was homogenized using sterile equipment (*i.e.* knifes and cutting boards) and 2.5 g and 150 mg weighed for volatile compounds and DNA analyses, respectively. Three technical replicates were obtained for each of the above-mentioned analyses. The samples were stored at −80 °C until analysis.

2.2. Shotgun metagenomics

Shotgun metagenomics data obtained from the previous characterization of typical mountain Caciotta cheese (Cardin et al., 2023) were employed to develop predictive sPLS-DA model to authenticate cheese origin. Briefly, for both Caciotta cheese and starters, total DNA was extracted from the samples using the DNeasy PowerSoil kit according to the manufacturer's instructions (Qiagen, Hilden, Germany). The quantity of DNA was assessed using a Qubit dsDNA HS Assay (Invitrogen, Life Technologies, Italy). Libraries were constructed with the Nextera XT DNA Sample Preparation Kit (Illumina, Inc., San Diego, USA) and IDT for Illumina Nextera DNA UD Indexes. Libraries were combined in equimolar amounts and assessed for quality and quantity using the Agilent 2100 Bioanalyzer and Qubit Assay Kit HS, respectively. Sequencing was performed by UC DAVIS Genome Center (California, US) on a NOVASEQ Sp500 platform, generating 1.1 billion reads for cheese and 340 million reads for starters (Cardin et al., 2023). The FASTQC software was used to assess the quality of the raw reads (v.0.11.9, Brown et al., 2017), which were then processed with the bioBakery3 platform for quality control, contaminant depletion, and taxonomic assignment using KneadData and MetaPhlAn 3 (Beghini et al., 2021). KneadData was used to remove low-quality, repetitive sequence, and adapter sequences with a quality score cut-off of 35. High quality microbial reads were taxonomically profiled using MetaPhlAn3, an assembly free taxonomic profiler (Segata et al., 2012; Beghini et al., 2021). All raw sequence data in read-pairs format were deposited in the National Centre for Biotechnology Information (NCBI) in the Sequence Read Archive (SRA) under the project PRJNA922379 and PRJNA922380, for cheese samples and starter cultures, respectively.

2.3. Volatilome analysis

The HiSorb probes were used in conjunction with UNITY-xr, both from Markes International (UK), to perform headspace thermal desorption coupled with gas chromatography-mass spectrometry (5977B GC–MS Agilent Technologies, US). The cheese samples were placed in 2.5 ml vials and headspace was sampled using an HiSorb Agitator (Markes International UK) at 40 °C and 200 rpm for one hour. The probes were thermodesorbed using UNITY-xr at 280 °C for 12 min, and a purge flow of 50 ml/min for 1 min was used. The flow path was set at 200 °C with a trap low of 25 °C and a trap high of 290 °C, and injection in the GC was performed using a low split 5 ml/min flow. A DB-5 ms capillary column 60 m × 250 μm × 0,25 μm (Agilent Technologies, US) was used for the analysis. The oven temperature program was set to initial 40 °C held for 2 min, then ramped 3 °C/min up to 180 °C, and again ramped 20 °C/min up to 260 °C for 5 min, and finally held for 6 min. The constant flow rate of helium carrier gas was set to 1 ml/min. The MS analyses were done in a full scan mode (TIC mode), with a scan range of 33 to 350 amu. To validate detected peaks, forty-four standard molecules were injected. The MassHunter quantitative analysis workstation (v.11.1, Agilent Technologies, US) was used for semi-quantitative analysis, with total peak area used for statistical analyses (Cardin et al., 2023).

2.4. Near infrared spectral acquisition

A 300 g slice of cheese was ground with a Retsch Grindomix (Retsch GmbH, Haan, Germany) at 4000 rpm for 10 s after removal of 2 cm of crust all around. Ground cheese samples were analyzed using a FOSS DS-2500 scanning monochromator (FOSS NIRSystem, Hillerød, Denmark). Scans were recorded in reflectance mode (850–2500 nm at 0.5-nm intervals) using a slurry cup with a quartz window (12.6 cm^2 area) in 30 g aliquots. Spectral data were recorded as absorbance (A) calculated as log (1/R), where R represents reflectance, using WinISI4 software V4.10.0.15326 (FOSS Analytical A/S, Hillerød, Denmark). Before statistical analysis, spectra were exported to an Excel (Microsoft Office, USA) spreadsheet and averaged before further chemometric modeling.

2.5. Element analysis

About 0.5 g of fresh cheese were acid digested using an UltraWAVE System (Milestone, Shelton, CT, USA) equipped with PTFE vials and following the method reported in Muñoz-Redondo et al. (2022). The mineral element profile (Li, Be, B, Na, Mg, Al, P, K, Ca, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As, Se, Rb, Sr, Y, Mo, Pd, Ag, Cd, Sn, Sb, Te, Cs, Ba, La, Ce, Pr, Nd, Sm, Eu, Gd, Dy, Ho, Er, Tm, Yb, Re, Hg, Tl, Pb, Bi and U) was determined using an ICP-MS (Agilent 7800, Agilent Technologies, Tokyo, Japan). Helium was used as collision gas in the Octopole Reaction System for the effective analysis of Na, Mg, K, Ca, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, Ge, As and Pd with a flow of 5 ml/min while for Se, Sn, Eu a flow of 10 ml/min was used. Instrumental parameters were optimized to maximize sensitivity and reduce spectral interferences at each analytical batch following manufacturing guidelines. A solution of Sc, Rh and Tb was used as internal standard online for the correction of signal drift whereas a solution of a known concentration was added to each sample before mineralization for volume correction. Each batch included, together with samples, a blank sample (only reagents) to ensure cleanliness. Accuracy was verified using samples spiked with a known amount of a standard solution. The calculated recoveries ranged from 83 to 117 % and were considered acceptable for the aim of this study. For determination of the limits of quantification (LOQ), 10 blank samples were prepared and analyzed in a sequence and the calculated standard deviation obtained for each element was multiplied by 10. All the materials in contact with standard or samples during mineralization and analysis were washed with a 5 % HNO_3 solution and rinsed with ultrapure water (18.2 MΩ-cm, Millipore, Bedford, MA, USA). Moisture

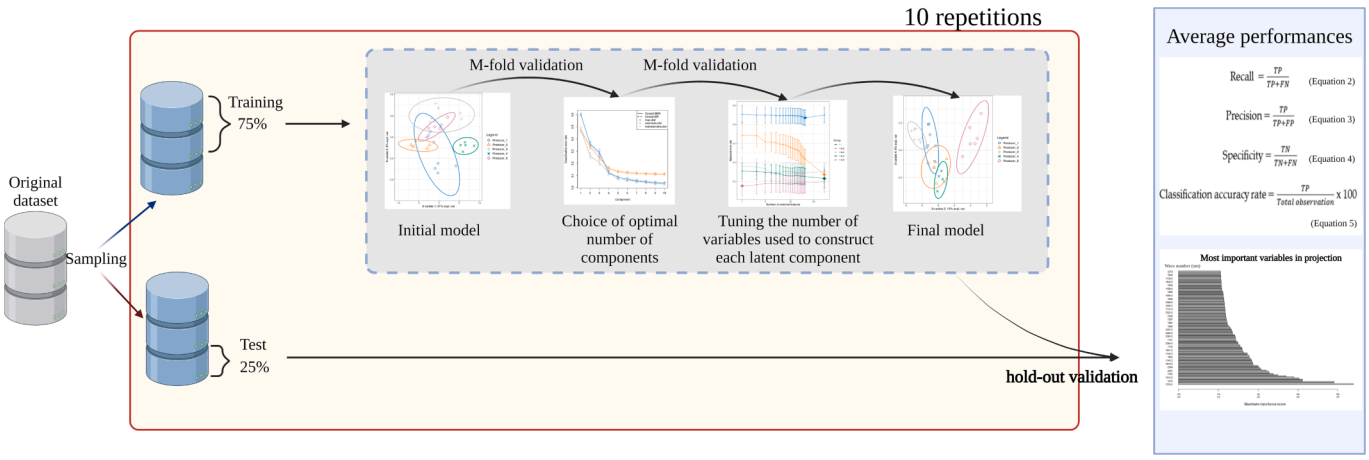

Fig. 2. Procedure used to obtain and compare validated sPLS-DA models. Each dataset was sampled 10 times obtaining representative origins class for both training (75% of samples) and test (25% of samples) dataset. Each model was validated with M−fold validation and tested with hold-out validation.

of each sample was quantified as the loss of weight in an oven at 105 °C. Elemental contents were expressed as percentage of dry matter (dm).

2.6. Stable isotopes ratio

Four g of freeze-dried cheese were extracted 3 times with 30 ml of a petroleum ether:diethyl ether (2:1) mixture, homogenizing with an Ultraturrax device (IKA T25, IKA GmbH, Staufen, Germany) (11500 rpm for 3 min) and using a centrifuge (ALC PK 131R, Thermo Scientific, Bremen, Germany) (*e.g.* 4100 rpm for 6 min) to separate the ether from the residue. A Soxhlet extractor (Merck KGaA, Darmstadt, Germany) was used as an alternative to extract fat. After lipid extraction, the skimmed cheese was warmed to 40 °C to remove any possible residual ether. Then, the residue was washed twice with 20 ml of water, centrifuging each time at 4100 rpm for 3 min. The residue, mainly made up mainly of casein, was lyophilised (LIO5P, 5Pascal, Milan, Italy) and kept at room temperature until analysis.

All samples were weighted in silver and tin capsules for OH − and CNS − isotope measurements, respectively. $^{15}N/^{14}N$, $^{13}C/^{12}C$ and $^{34}S/^{32}S$ ratios were determined using an isotope ratio mass spectrometer (IRMS) (Isoprime, AP2003, GV Instruments Ltd, Manchester, UK) equipped with an elemental analyser (Vario EL III Elementar Analysensysteme GmbH, Hanau, Germany), while the $^{18}O/^{16}O$ and $^{2}H/^{1}H$ ratios were determined with an IRMS (Flash EA1112, Thermo Fisher Scientific, Bremen, Germany) equipped with a pyrolyzer (TC/EA, Thermo Fisher Scientific, Bremen, Germany).

In agreement with the IUPAC protocol (Brand et al., 2014), the isotopic values were expressed in delta in relation to the international standard V − PDB (Vienna − Pee Dee Belemnite) for $\delta^{13}C$, V − SMOW (Vienna − Standard Mean Ocean Water) for $\delta^{2}H$ and $\delta^{18}O$, V − CDT (Vienna − Canyon Diablo Troilite) for $\delta^{34}S$ and Air (atmospheric N_2) for $\delta^{15}N$, following equation (1):

$$\delta_{ref}\left({}^{i}E/{}^{j}E, sample\right) = \frac{R\left({}^{i}E/{}^{j}E, sample\right)}{R\left({}^{i}E/{}^{j}E, ref\right)} - 1 \quad (1)$$

where *ref* is the international measurement standard, *sample* is the analyzed sample and ${}^{i}E/{}^{j}E$ is the isotope ratio between heavier and lighter isotopes. The delta values are multiplied by 1000 and expressed commonly in units "per mil" (‰) or, according to the International System of Units (SI), in unit 'milliurey' (mUr).

Isotopic values were calculated against two standards through the creation of a linear equation. The standards used in the isotopic analyses were international reference materials or in − house working standards that had been calibrated against them. In particular, the international standards used were: for $^{13}C/^{12}C$, fuel oil NBS − 22 ($\delta^{13}C = -30.03 \pm 0.05$ ‰), sucrose IAEA − CH − 6 ($\delta^{13}C = -10.45 \pm 0.04$ ‰) (IAEA − International Atomic Energy Agency, Vienna, Austria), and L − glutamic acid USGS 40 ($\delta^{13}C = -26.39 \pm 0.04$ ‰) (U.S. Geological Survey, Reston, VA, USA); for $^{15}N/^{14}N$, L − glutamic acid USGS 40 ($\delta^{15}N = -4.52 \pm 0.06$ ‰) (U.S. Geological Survey, Reston, VA, USA), ammonium sulfate salts IAEA − N − 1 ($\delta^{15}N = +0.43 \pm 0.07$ ‰), IAEA − N − 2 ($\delta^{15}N = +20.41 \pm 0.12$ ‰) and potassium nitrate IAEA − NO3 ($\delta^{15}N = +4.7 \pm 0.2$ ‰); for $^{34}S/^{32}S$, USGS 42 ($\delta^{34}S = +7.84 \pm 0.25$ ‰), USGS 43 ($\delta^{34}S = +10.46 \pm 0.22$ ‰), barium sulphate IAEA − SO − 5 ($\delta^{34}S = +0.5 \pm 0.2$ ‰) and NBS 127 ($\delta^{34}S = +20.3 \pm 0.4$ ‰); for $^{2}H/^{1}H$ fuel oil NBS − 22 ($\delta^{2}H = -119.6 \pm 0.6$ ‰), keratins CBS (Caribou Hoof Standard $\delta^{2}H = -157 \pm 2$ ‰) and KHS (Kudu Horn Standard $\delta^{2}H = -35 \pm 1$ ‰) from U.S. Geological Survey; for $^{18}O/^{16}O$ benzoic acid IAEA 601 ($\delta^{18}O = +23.14 \pm 0.19$ ‰) and benzoic acid IAEA 602 ($\delta^{18}O = +71.28 \pm 0.36$ ‰) from IAEA.

Each reference material was measured in duplicate at the start and end of each daily group of analyzed samples (each sample was also analyzed in duplicate). A control material was also included in the analyses of each sample group, to check the efficiency of the measure. The maximum standard deviations of repeatability accepted were 0.3 ‰ for $\delta^{13}C$ and $\delta^{15}N$, 0.4 ‰ for $\delta^{34}S$, 0.5 ‰ for $\delta^{18}O$ and 3 ‰ for $\delta^{2}H$.

2.7. Statistical analysis

Statistical analysis of microbial communities, volatile organic compounds, NIRS, trace elements analyses and stable isotopes ratio were performed in R version 4.0.5 (R Foundation for Statistical Computing, Vienna, Austria) and IBM SPSS Statistics (v. 28, Segrate, Italy). Significance of the median was obtained with Kruskal-Wallis Rank Sum Test while multiple comparison was performed with pairwise Wilcoxon Rank Sum Tests. Obtained *p*-values were adjusted with Benjamini & Hochberg (1995) correction method. For stable isotopes ratio, mean significance was obtained with ANOVA and Tukey's honest significance test. Pearson's r correlation coefficient was used to test correlations between stable isotopes and altitude, latitude and longitude of the studied producers. Trace elements distribution presented challenging problems connected with LOQ. In this instance, we employed missing value substitution according to the results presented by Farnham et al. (2002). LOQ was divided by two to insert missing values.

Unsupervised principal component analysis was used to investigate the datasets. sPLS-DA was employed to evaluate authentication performance (Lê Cao et al., 2011). Centered log-ratio transformed relative abundance for bacterial and viral communities, peak abundances for volatilome analysis, wavelength absorbance for near infrared spectroscopy, elements and isotope concentration for trace elements and

Table 1
Performance of the sPLS-DA model in classifying caciotta cheese producers' based on bacterial community through hold-out validation on the test sets.

Origin/ Parameter	Recall	Precision	Specificity	Classification accuracy rate (%)
Producer 1	0.92	0.88	0.98	
Producer 2	0.97	0.98	0.99	
Producer 3	0.91	0.98	0.98	96.13 ± 4.02
Producer 4	1.00	0.98	1.00	
Producer 5	1.00	0.96	1.00	

isotopes analysis were used with sPLS-DA model using the *mixOmics* R package (Rohart et al., 2017). To avoid sampling biases and represent the true performances of the models, each dataset was sampled 10 times obtaining representative origins class for both training (75 % of samples) and test (25 % of samples) dataset. M−fold cross-validation of the training dataset (*i.e.* the process of dividing of dataset into M subsets and then, iteratively, using some of them to train the model while exploiting the others to evaluate its performance) was performed using 10 folds and 100 repeats. The obtained models were used to predict the origin of the tested dataset in a hold-out validation. Average predictive performances were compared considering true positive (TP), true negative (TN), false positive (FP) and false negative (FN) ratios expressed through recall (Fig. 2, Equation 2), precision (Fig. 2, Equation 3), specificity (Fig. 2, Equation 4) classification accuracy rate (Fig. 2, Equation 5) and classification error rate (Ting, 2011; Kassambara, 2018; Bisutti et al., 2019). Fig. 2, summarize the procedure used to obtain the validated models.

Recall, precision, specificity and classification accuracy rate metrics are commonly used to compare the performance of predictive models. Each metric captures different aspects of the model's performance, and they are used in combination to provide a comprehensive evaluation of the model's effectiveness. Recall measures the proportion of true positive cases that were correctly identified by the model. Precision evaluates the proportion of true positive cases among all cases predicted as positive by the model. Specificity measures the proportion of true negative cases that were correctly identified by the model. It is a useful metric when the cost of a false positive error is high. Classification accuracy rate measures the proportion of correct predictions made by the model.

In sPLS-DA models, loading is a vector that represents the relationship between the predictor variables (*e.g.* taxa relative abundance) and

Table 2
Performance of the sPLS-DA model in classifying caciotta cheese producers based on viral community through hold-out validation on the test sets.

Origin/ Parameter	Recall	Precision	Specificity	Classification accuracy rate (%)
Producer 1	1.00	0.91	1.00	
Producer 2	0.95	0.98	0.98	
Producer 3	0.97	0.98	0.99	97.42 ± 2.58
Producer 4	1.00	0.98	1.00	
Producer 5	0.95	1.00	0.98	

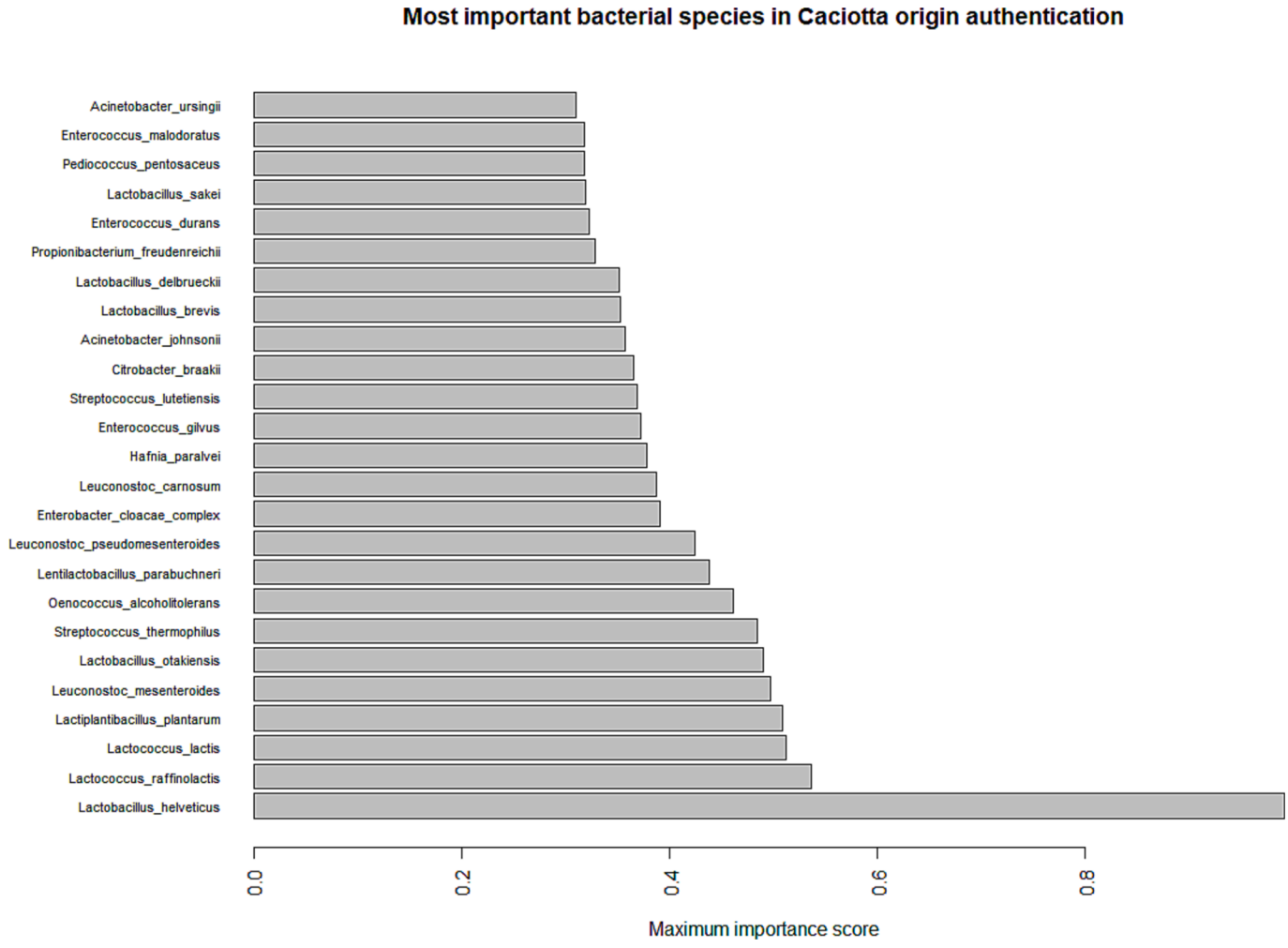

Fig. 3. Mean maximum importance score of the most important 25 bacterial species used in the sPLS-DA models.

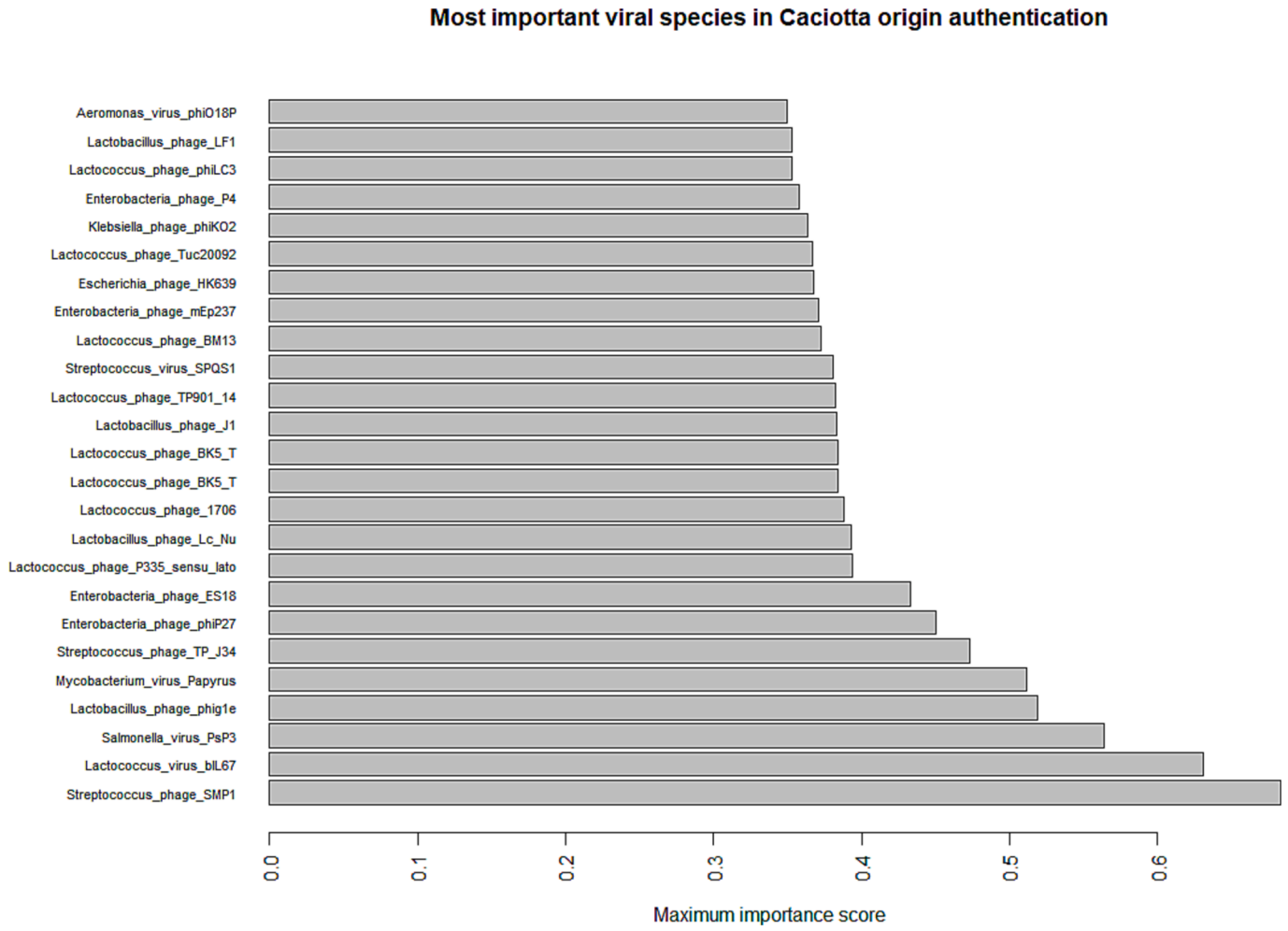

Fig. 4. Mean maximum importance score of the most important 25 viral species used in the sPLS-DA models.

Table 3
Performance of the sPLS-DA model in classifying caciotta cheese producers based on volatile organic compounds (VOCs) through a hold-out validation on the test sets.

Origin/ Parameter	Recall	Precision	Specificity	Classification accuracy rate (%)
Producer 1	0.98	0.82	0.99	
Producer 2	0.9	0.98	0.97	
Producer 3	0.86	0.91	0.96	90.0 ± 11.11
Producer 4	0.93	0.82	0.98	
Producer 5	0.86	0.98	0.96	

the response variable (*e.g.* origin) that defines the class labels. Loading was calculated during the model fitting process and used to identify the variables that contribute the most to the separation of the classes in the data. The average loading importance was computed by extracting the loading value from each model and then calculating the absolute value of each. Finally, the average of these absolute values was used to represent the 25 most important variables for the prediction.

3. Results

3.1. Bacterial and viral communities

Caciotta microbiota showed a complex profile for both bacterial and viral communities. The univariate and multivariate analysis of the microbiota of typical Caciotta cheese has been previously outlined, including the results of the Kruskal-Wallis test, visual representation through non-metric multidimensional scaling, and permutational multivariate analysis of variance (Cardin et al., 2023). Briefly, 45 bacterial and 44 viral species had significantly different mean relative abundance according to their origin. Starter (*Steptococcus thermophilus* and *Lactobacillus delbrueckii*) and non-starter lactic acid bacteria (*e.g. Lactococcus raffinolactis, Lentilactobacillus parabuchneri, Lactiplantibacillus paraplantarum,* and *Propionibacterium freudenreichii*) showed significant relative abundance differences. Some secondary microbiota components were only found in specific origins and a clear link between the cheese and its producer for several *S. thermophilus* strains was observed. Starter lactic acid bacteria were the most abundant, with associated bacteriophages like Streptococcus phage TP 778l, Streptococcus virus DT1, Streptococcus virus phiAbc2, and Lactobacillus phage A2 exhibiting the higher abundance and significant variations based on Caciotta origin. Both bacterial and viral communities formed clusters according to producer location (data not shown). However, viral communities showed narrower clusters for each cheese origin and demonstrated significant differences in permutation analysis of variance. Bacterial communities tended to cluster based on regional area (*i.e.* producer 1, 2, 3, were separated from producer 4 and 5) as confirmed by permutation analysis (data not shown).

The potential of bacterial communities for authenticating Caciotta origin was evaluated using sPLS-DA models. The training models for the bacterial communities used a variable number of components, ranging from 9 to 15 and achieved a correct classification accuracy rate of 99.89 ± 0.20 %. The validated outcomes are presented in Table 1.

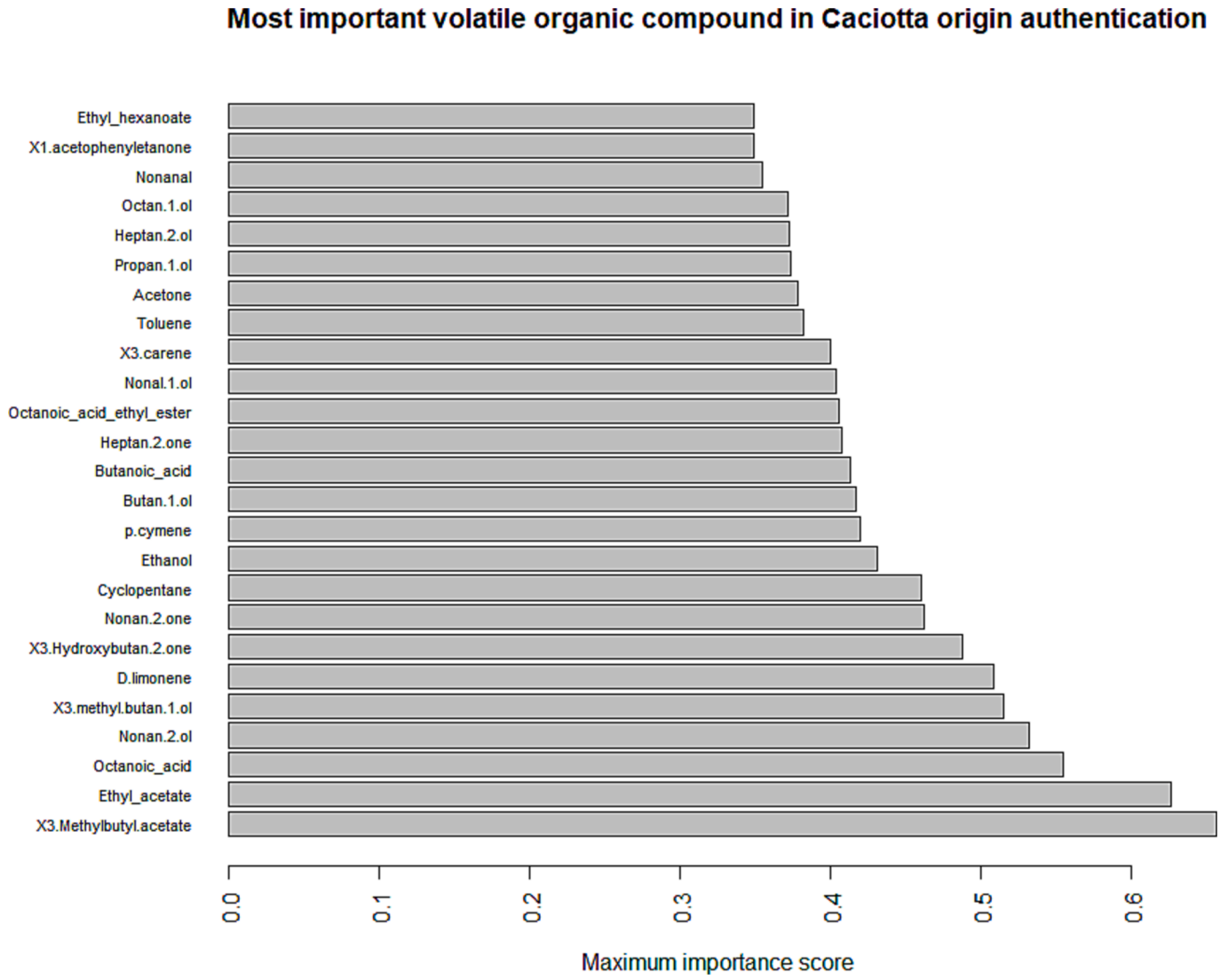

Fig. 5. Mean maximum importance score of the most important 25 VOCs used in the sPLS-DA models.

Hold-out validated models exhibited a slight decrease in performances compared to the training ones. Generally, recall, precision, and specificity presented high values for bacterial communities of Caciotta cheese. Producer 1 showed the lowest performances in origin authentication driven by the lowest value of precision while producer 3 had the lowest recall. Overall, bacterial communities yielded a high value of classification accuracy rate 96.13 ± 4.02 %. The most important bacterial species in Caciotta origin authentication are shown in Fig. 3.

Most of the taxa with the highest scores were attributed to lactic acid bacteria. Among them the majority showed a significant difference in relative abundance connected to origin. Moreover, as previously noticed (Cardin et al., 2023), origin specific non-starter lactic acid bacteria like *Lactobacillus helveticus, Lactococcus raffinolactis* and *Propionibacterium freudenreichii* presented the highest score for origin authentication.

Viral communities constituted the second portion of the microbiota investigated to authenticate mountain Caciotta cheese. The training models for viral communities presented a lower number of components, from 7 to 12, and higher classification accuracy rate (99.92 ± 0.16 %) than the bacterial one. Also, the hold-out validated performances showed higher values of recall, precision, specificity and classification accuracy rate than the bacterial one (Table 2).

Overall, viral communities' models showed the highest performances for the authentication of producer 1, 4 and 5, while the lowest performances were observed for producer 2. The obtained classification accuracy rate was higher than bacterial communities yielding a correct classification of 97.42 ± 2.58 % of the samples. The most important viral species used in the loadings of the model are shown in Fig. 4.

Generally, *Streptococcus*, *Lactococcus*, and *Lactobacillus* phages were the taxa obtaining the highest score for origin prediction. Other important taxa corresponded to *Salmonella*, *Mycobacterium* and *Enterobacteriaceae* associated viruses.

3.2. Volatilome

VOC analysis in typical Caciotta cheese (both univariate and multivariate approaches) was already described in our previous work (Cardin et al., 2023). These analyses included outcomes from the Kruskal-Wallis test, visual representations via principal components analysis, and permutational multivariate analysis of variance. Mountain Caciotta cheese presented a complex volatilome from which prominent levels of alcohols and ketones, accompanied by lower proportions of terpenes were observed. Cheese origin led to significant variations in the relative abundances of most investigated volatile organic compounds (*i.e.* ethanol, butan-2-ol, 3-methylbutan-1-ol, benzaldehyde, ethyl acetate, 3-methylbutyl acetate, butan-2-one, heptan-2-one, octanoic acid, D-limonene) with the exclusion of 3-hydroxybutan-2-one and 1-

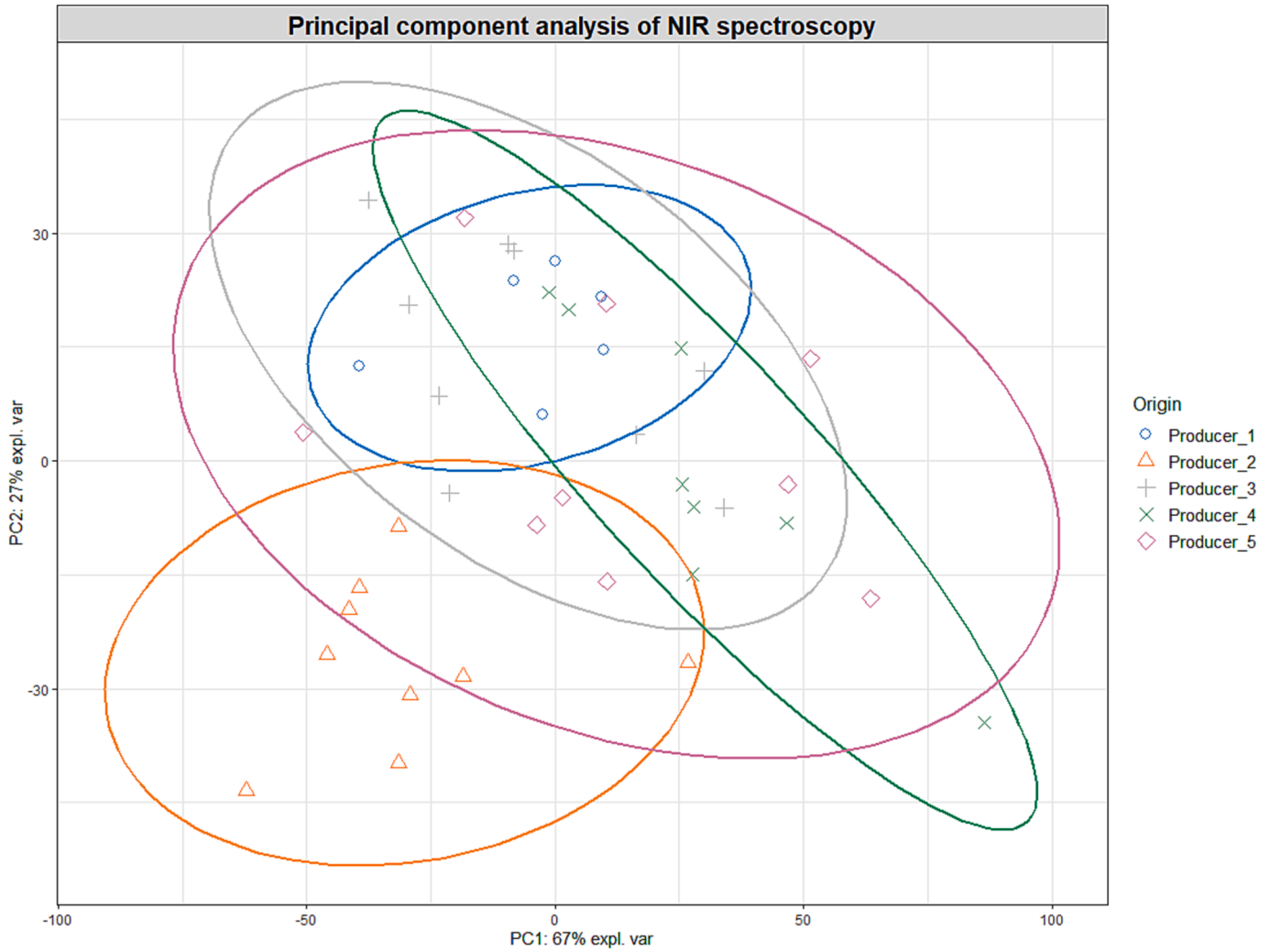

Fig. 6. Principal component analysis was conducted on the absorbance of Caciotta in the Near Infrared (NIR) region (850–2500 nm). The absorbance data in the NIR were transformed using the centered log-ratio method.

Table 4
Performance of the sPLS-DA model in classifying caciotta cheese producers based on NIRs through hold-out validation on the test sets.

Origin/ Parameter	Recall	Precision	Specificity	Classification accuracy rate (%)
Producer 1	0.95	0.68	0.98	
Producer 2	0.95	1.00	0.98	
Producer 3	0.55	0.65	0.88	76.0 ± 31.57
Producer 4	0.7	0.64	0.91	
Producer 5	0.65	0.93	0.9	

acetophenyletanone. Additionally, certain terpenoids like *p*-cymene and 3-carene were not consistently found across all origins. The pairwise permutational analysis of variance revealed that each producer of typical mountain Caciotta cheese exhibited a distinctive VOC profile, varying to some degree from one another.

The training models for the volatile organic compounds used a variable number of components, ranging from 9 to 15 and achieving a correct classification accuracy rate 99.87 ± 0.18 %. The hold-out validated outcomes are presented in Table 3.

Producer 2 showed the highest values of recall, precision and specificity compared to the other origins. Producers 1 and 4 had the lowest values of precision while producers 5 and 3 had the lowest values of recall. Overall, high specificity values were observed for all the origins. The hold-out validated model obtained a high value of correct accuracy rate 90.0 ± 11.11 % but it was characterized by a high accuracy error as well. Fig. 5 reports the mean maximum score of the most important volatiles used in Caciotta origin authentication.

Among the most important VOCs for authenticating mountain Caciotta origin, alcohols and ketones were the most prevalent classes, followed by esters, terpenes, fatty acids and hydrocarbons. Esters, such as 3-methylbutyl acetate and ethyl acetate, showed the highest scores. Terpenes, like D-limonene, p-cymene and 3-carene, significantly contributed to the model performance, along with alcohols, such as nonan-2-ol, 3-methylbutan-2-ol and ethanol. Surprisingly, 3-hydroxybutan-2-one and 1-acetophenyletanone, which did not show significant differences based on origin, were also identified as important compounds for the authentication models.

3.3. *Near infrared spectroscopy*

Absorbance values of typical Caciotta cheese were firstly investigated with principal components analysis (Fig. 6).

Although the unsupervised analysis of Caciotta cheese explained 94 % of the variance in the first and second components, it revealed overlapping clusters of samples related to their origin. The ability of NIRs data to authenticate Caciotta origin was further investigated with sPLS-Da models. The training models exhibited a wide range of components, varying from 3 to 14. Nonetheless, they displayed encouraging

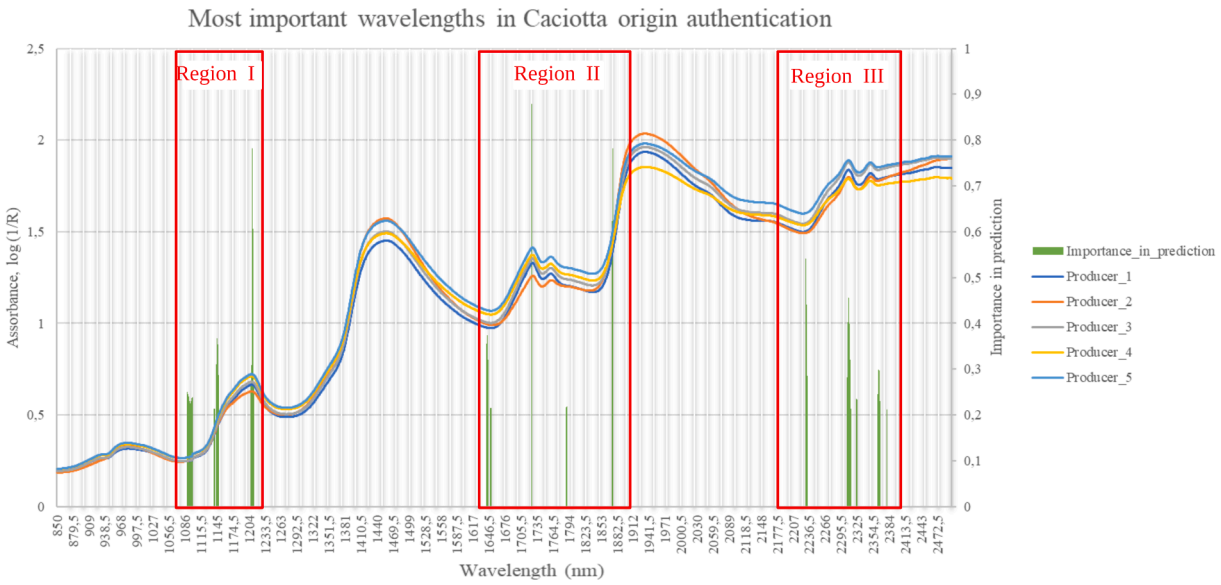

Fig. 7. Average absorbance spectra of typical Caciotta cheese based on origin. Green bars, indicating the average importance value in prediction, were grouped in Region I, II and III. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

classification performance, resulting in an average classification accuracy rate of 92.67 ± 7.93 %. Table 4 reports the average performance of the hold-out validated models.

Hold-out validated models exhibited a considerable decrease in performances compared to the training one. Producer 3 reported the lowest values of recall, precision and specificity that connected to the precision and recall of producer 4 and 5 led to a classification accuracy rate of 76 ± 31.57 %. This model exhibited a high error in accuracy, which could be attributed to the low precision values observed for producer 1, 3, and 4. Given that a singular wavelength lacks significant information in comparison to NIR analysis regions, we assessed the 100 most crucial wavelengths for predicting the origin of Caciotta (Fig. 7).

The wavelengths with the highest mean maximum importance score in the prediction of Caciotta origin formed three regions. Regions I and III were characterized by three clusters around 1090, 1145, 1204 and 2230, 2310 and 2360 nm, respectively. Finally, Region II presented scattered important wavelengths around 1646, 1735, 1780 and 1870 nm.

3.4. Element analysis

The elemental analysis of Caciotta cheese showed different concentration profiles for 39 elements (Table 5) while Be, Ga, Ge, Pd, Ag, Sn, Te, Dy, Ho, Tm, Hg and Tl which presented concentrations inferior to LOQ for 90 % of observations were excluded from the following analysis. The results of the univariate analysis showed significant variations in element concentrations across different producers of Caciotta cheese.

Each producer exhibited unique elemental signatures whereas the major differences were observed for producers 1, 2, and 5. The elements can be categorized into four groups according to their potential ties to dairy equipment (*i.e.* Fe, Cu, Al, Ni, Cr), their classification as macronutrient (*i.e.* P, K, Ca, Mg, Na) or rare earth (*i.e.* Ce, Er, Eu, La, Yb), and "other elements". Producer 1 exhibited the highest concentration of copper (Cu), phosphorus (P), barium (Ba), lead (Pb) and uranium (U); producer 2 had the highest concentrations of iron (Fe) and nickel (Ni), sodium (Na) and strontium (Sr); producer 3 exhibited the highest cerium (Ce), manganese (Mn) and rubidium (Rb); producer 4 showed the highest concentration of chromium (Cr) while producer 5 had the highest concentration of aluminum (Al), europium (Eu), strontium (Sr) as well as Ba, Pb and U.

We further examined the elemental composition of Caciotta cheese with unsupervised multivariate analysis based on principal component analysis (Fig. 8).

Caciotta cheeses from the same origin tended to cluster together. However, while the confidence intervals of producer 2 and 3, as well as producer 1 and 4, presented minor areas of overlap, the confidence interval of producer 5 overlapped significantly with that of all other producers. The possibility of authenticating Caciotta origin was tested with the sPLS-DA model for which the cross-validated results are reported in Table 6. The number of components in the training models varied from 9 to 15, suggesting a potential impact of dataset sampling. Nevertheless, this effect did not seem to affect the performances of these models, as they all achieved a 100 % correct classification rate.

The hold-out validated models (Table 6), on average, demonstrated a slightly lower but still excellent correct classification rate, achieving 99 ± 0.01 % accuracy in correctly classifying Caciotta origin. For producers 1, 4, and 5, the model achieved 100 % recall, precision, and specificity, demonstrating its accurate classification of all samples from these areas. However, producers 2 and 5 had recall, precision, and specificity values ranging from 0.95 to 0.99, which led to a slight decrease in the model's performance. We further investigate which element characterized the performance of the obtained models through maximum importance score (Fig. 9).

3.5. Stable isotopes

Mean values with standard deviation of $\delta^{13}C$, $\delta^{15}N$, $\delta^{34}S$, $\delta^{18}O$ and $\delta^{2}H$ determined in the mountain Caciotta cheese are shown in Table 7.

Producer 1 showed a notably lower $\delta^{13}C$ value compared to the rest of the producers. On the other hand, producers 2, 4 and 5 exhibited higher $\delta^{15}N$ values when compared to producers 1 and 3. Overall, the $\delta^{34}S$ values ranged from 3.32 ‰ to 5.30 ‰, but no significant differences were observed. Differently, $\delta^{18}O$ showed higher values for producers 4 and 5 and lower for 2 and 3 whereas producer 1 had intermediate value. Lastly, we noticed significant differences on the $\delta^{2}H$ values for all the considered origins. Specifically, producer 2 had the lowest value, while producers 4 and 5 had the highest value. Producers 1 and 3 had intermediate values. Significant correlations for $\delta^{18}O$ and $\delta^{2}H$ versus both the latitude and the altitude were observed (Pearson's r −0.503, $p < 0.001$; Pearson's r −0.497, $p < 0.001$) (Pearson's r −0.763, $p < 0.001$; Pearson's r −0.587, $p < 0.001$) The unsupervised analysis of stable isotope ratio is shown in Fig. 10.

Table 5
Pairwise Wilcoxon Rank Sum of element concentrations expressed as dry matter concentration in Caciotta cheese.

Element	Producer 1	Producer 2	Producer 3	Producer 4	Producer 5
Al (µ/kg dm)	324.92 ± 21,62^{a}	504.15 ± 81,75^{ab}	143.47 ± 12.88^{c}	372.86 ± 26.33^{ab}	938.35 ± 300.06^{b}
As (µg/kg dm)	0.5 ± 0^{a}	1.05 ± 0.28^{a}	0.5 ± 0^{a}	0.5 ± 0^{a}	1.20 ± 0.34^{a}
B (µg/kg dm)	134.28 ± 35.58^{ab}	223.79 ± 27.22^{a}	143.52 ± 11.94^{ab}	125.74 ± 23.08^{b}	177.41 ± 28.89^{ab}
Ba (µg/kg dm)	1556.72 ± 14.52^{a}	1366.10 ± 193.27^{ab}	975.13 ± 87.36^{bc}	963.81 ± 23.08^{c}	1045.59 ± 111.25^{bc}
Bi (µg/kg dm)	0.1 ± 0^{a}	0.1 ± 0^{a}	3.11 ± 0.86^{b}	0.1 ± 0^{a}	0.1 ± 0^{a}
Ca (g/kg dm)	12.31 ± 0.08^{a}	10.84 ± 0.29^{ab}	12.42 ± 0.11^{a}	11.59 ± 0.13^{a}	10.60 ± 0.26^{b}
Cd (µg/kg dm)	0.31 ± 0.02^{a}	0.37 ± 0.02^{a}	0.32 ± 0.01^{a}	0.30 ± 0.01^{a}	0.32 ± 0.01^{a}
Ce (µg/kg dm)	0.25 ± 0.04^{a}	0.36 ± 0.06^{ab}	0.04 ± 0^{c}	0.32 ± 0.04^{ab}	0.69 ± 0.22^{b}
Co (µg/kg dm)	0.752 ± 0.01^{a}	1.52 ± 0.09^{b}	1.40 ± 0.12^{b}	1.85 ± 0.09^{c}	1.35 ± 0.17^{b}
Cr (µg/kg dm)	5.32 ± 0.46^{a}	53.00 ± 4.48^{b}	10.56 ± 1.76 ^{cd}	7.38 ± 0.76^{ac}	16.78 ± 0.74^{d}
Cs (µg/kg dm)	8.25 ± 1.16^{a}	10.98 ± 1.43^{a}	6.35 ± 0.77^{a}	1.62 ± 0.07^{b}	3.24 ± 0.74^{b}
Cu (µg/kg dm)	12240.97 ± 1096^{a}	496.19 ± 131.38^{b}	307.96 ± 21.92^{b}	6078.50 ± 221.11^{c}	250.47 ± 23.46^{b}
Er (µg/kg dm)	0.02 ± 0^{a}	0.03 ± 0.006^{a}	0.02 ± 0^{a}	0.02 ± 0^{a}	0.04 ± 0.01^{a}
Eu (µg/kg dm)	0.09 ± 0.006^{a}	0.06 ± 0.01^{ab}	0.03 ± 0^{b}	0.03 ± 0^{b}	0.07 ± 0.01^{ab}
Fe (µg/kg dm)	1036.84 ± 36.94^{a}	1687.78 ± 120.29^{b}	1078.63 ± 82.21^{a}	1148.50 ± 50.54^{a}	1511.33 ± 122.53^{b}
Gd (µg/kg dm)	0.03 ± 0^{a}	0.03 ± 0^{a}	0.03 ± 0^{a}	0.03 ± 0^{a}	0.08 ± 0.02^{a}
K (mg/kg dm)	1651.22 ± 44^{ab}	2424.96 ± 180^{c}	2107.02 ± 159^{ac}	1529.64 ± 23^{d}	1637.53 ± 146^{bd}
La (µg/kg dm)	0.12 ± 0.01^{a}	0.21 ± 0.04^{ab}	0.07 ± 0.47^{c}	0.23 ± 0.03^{ab}	0.39 ± 0.11^{b}
Li (µg/kg dm)	2.75 ± 0.16^{a}	18.99 ± 6.14^{b}	1.36 ± 0.47^{a}	1.65 ± 0.01^{a}	3.03 ± 0.87^{ab}
Mg (mg/kg dm)	0.43 ± 0.01^{ab}	0.49 ± 0.01^{a}	0.49 ± 0.01^{a}	0.44 ± 0.01^{a}	0.40 ± 0.01^{b}
Mn (µg/kg dm)	198.15 ± 5.99^{a}	284.46 ± 23.13^{b}	262.56 ± 7.00^{b}	235.19 ± 13.97^{b}	213.28 ± 22.63^{ab}
Mo (µg/kg dm)	252.45 ± 12.49^{a}	402.35 ± 32.77^{b}	249.34 ± 12.00^{a}	251.97 ± 14.52^{a}	257.83 ± 10.08^{a}
Na (g/kg dm)	8.09 ± 0.30^{a}	11.69 ± 0.80^{b}	8.36 ± 1.25^{ab}	9.12 ± 0.12^{a}	9.31 ± 0.49^{ab}
Nd (µg/kg dm)	0.11 ± 0.01^{ab}	0.15 ± 0.02^{ac}	0.03 ± 0^{b}	0.14 ± 0.17^{ac}	0.34 ± 0,12^{c}
Ni (µg/kg dm)	1.25 ± 0^{a}	20.62 ± 2.18^{b}	3.58 ± 1.07^{ac}	4.33 ± 0.61^{c}	5.73 ± 0.66c
P (g/kg dm)	7.34 ± 0.10^{a}	6.58 ± 0.15^{ab}	7.31 ± 0.16^{ab}	7.20 ± 0.19^{ab}	6.83 ± 0.09^{b}
Pb (µg/kg dm)	6.64 ± 0.10^{a}	2.59 ± 0.27^{b}	3.23 ± 0.15^{b}	3.55 ± 0.25^{b}	10.81 ± 1.96^{a}
Pr (µg/kg dm)	0.02 ± 0^{a}	0.04 ± 0.007^{b}	0.02 ± 0^{a}	0.02 ± 0^{a}	0.08 ± 0.03^{b}
Rb (µg/kg dm)	3375.65 ± 276^{a}	3888.21 ± 626^{ab}	5207.71 ± 761^{b}	1298.42 ± 48^{c}	1479.29 ± 259^{c}
Re (µg/kg dm)	0.35 ± 0.06^{a}	0.57 ± 0.21^{a}	0.06 ± 0.004^{bc}	0.03 ± 0^{b}	0.17 ± 0.05^{ac}
Sb (µg/kg dm)	0.21 ± 0.01^{ab}	0.21 ± 0.01^{ab}	0.18 ± 0.01^{a}	0.18 ± 0.01^{a}	0.28 ± 0.01^{b}
Se (µg/kg dm)	93.40 ± 8.13^{ab}	160.86 ± 23.89c	117.88 ± 3.91^{a}	94.56 ± 5.71^{b}	109.55 ± 13,7^{ab}
Sm (µg/kg dm)	0.03 ± 0^{a}	0.03 ± 0^{a}	0.03 ± 0^{a}	0.03 ± 0^{a}	0,087 ± 0.02^{b}
Sr (µg/kg dm)	3803.73 ± 336^{ab}	5359.01 ± 271^{c}	2789.52 ± 41^{a}	2848.11 ± 52^{a}	3641.18 ± 257^{b}
U (µg/kg dm)	0.22 ± 0.03^{a}	0.09 ± 0.02^{ab}	0.15 ± 0.02^{b}	0.08 ± 0.01^{b}	0.35 ± 0,04^{a}
V (µg/kg dm)	0.25 ± 0^{a}	2.07 ± 0.71^{b}	0.79 ± 0.05^{ab}	0.57 ± 0.12^{a}	1.43 ± 0.58^{ab}

Table 5 (*continued*)

Element	Producer 1	Producer 2	Producer 3	Producer 4	Producer 5
Y (µg/kg dm)	0.11 ± 0.002^{a}	0.15 ± 0.03^{ab}	0.04 ± 0^{c}	0.33 ± 0.05^{b}	0.32 ± 0.08^{b}
Yb (µg/kg dm)	0.02 ± 0^{b}	0.04 ± 0.004^{a}	0.02 ± 0^{b}	0.02 ± 0^{b}	0.04 ± 0.01^{a}
Zn (mg/kg dm)	43.03 ± 1.19^{a}	43.42 ± 1.24^{ab}	43.07 ± 1,37^{a}	51.85 ± 1.20^{c}	38.29 ± 0.86^{b}

The first two components of the principal component analysis accounted for 87 % of the variability, but no separate clusters of Caciotta cheese related to its origin were discernible. We proceeded to assess the potential for authenticating Caciotta origin using sPLS-DA. The training models showed a slight variation in components, ranging from 3 to 4. However, upon comparison with the previously reported analysis, it became evident that the training model achieved the lowest classification performance, resulting in an average classification accuracy rate of 86.94 ± 5.34 %. Table 8 reports the classification performances of the hold-out validated models.

We observed a dichotomy in the performance of sPLS-DA for cheese origin authentication. Cheeses from producers 4 and 5 demonstrated high recall, precision, and specificity values, indicating accurate classifications. However, as we moved to producers 1, 2 and 3, the classification performances gradually decreased. In particular, producer 3 exhibited a recall and precision of zero, indicating that the models were unable to correctly identify or predict any cheeses from this producer. However, the specificity was 0.76, suggesting that the model performed comparatively well in correctly identifying cheeses that did not originate from producer 3. These factors influenced the classification accuracy rate, resulting in the lowest value and the highest accuracy error when compared to the other models. Fig. 11 shows the mean maximum importance score of the stable isotopes used in the sPLS-DA model.

3.6. Combination of element analysis and stable isotopes

Scientific literature reports that the combination of elements and stable isotopes can increase the origin authentication performance of cheese. Consequently, we merged the two datasets and proceeded to assess the performance of the models.

The unsupervised analysis, conducted through principal component analysis, revealed significant differences when compared to individual analyses (Fig. 12).

The clusters observed in the elemental analysis were better separated. This effect was more appreciable for producers 2 and 3 than for producers 1 and 4. Nevertheless, the explained variability of the analysis decreased from 87 % to 54 % when both analyses were combined, with trace elements alone accounting for only 53 %.

In the supervised analysis, the training models showed a decrease in the number of optimal components, ranging from 5 to 8, while achieving the same performance of element analysis (100 %). However, we did notice variations in the hold-out validated models (Table 9).

The origin authentication performance for producers 1 and 4 matched that of the element analysis models. As for producer 2, there was an improvement in recall and specificity, but the precision of the models decreased from 1 to 0.9. A similar trend was noticed for producer 3. The most significant decrease in performance was observed for producer 5. Overall, there was a slight decline in the classification accuracy rate, which was 97 ± 3.09 %. The most important variables for the prediction (Fig. 13) were similar to the element analysis models. We noticed a decrease in the importance of some elements such as Pb and Co while other elements (*e.g.* Cu and Se) reported higher scores. There were some differences in the importance of δ^{15}N, δ^{18}O and δ^{34}S as well. δ^{15}N achieved a higher score than δ^{18}O, while δ^{18}O obtained a higher score than δ^{34}S. However, the reverse order was observed for the stable isotopes models.

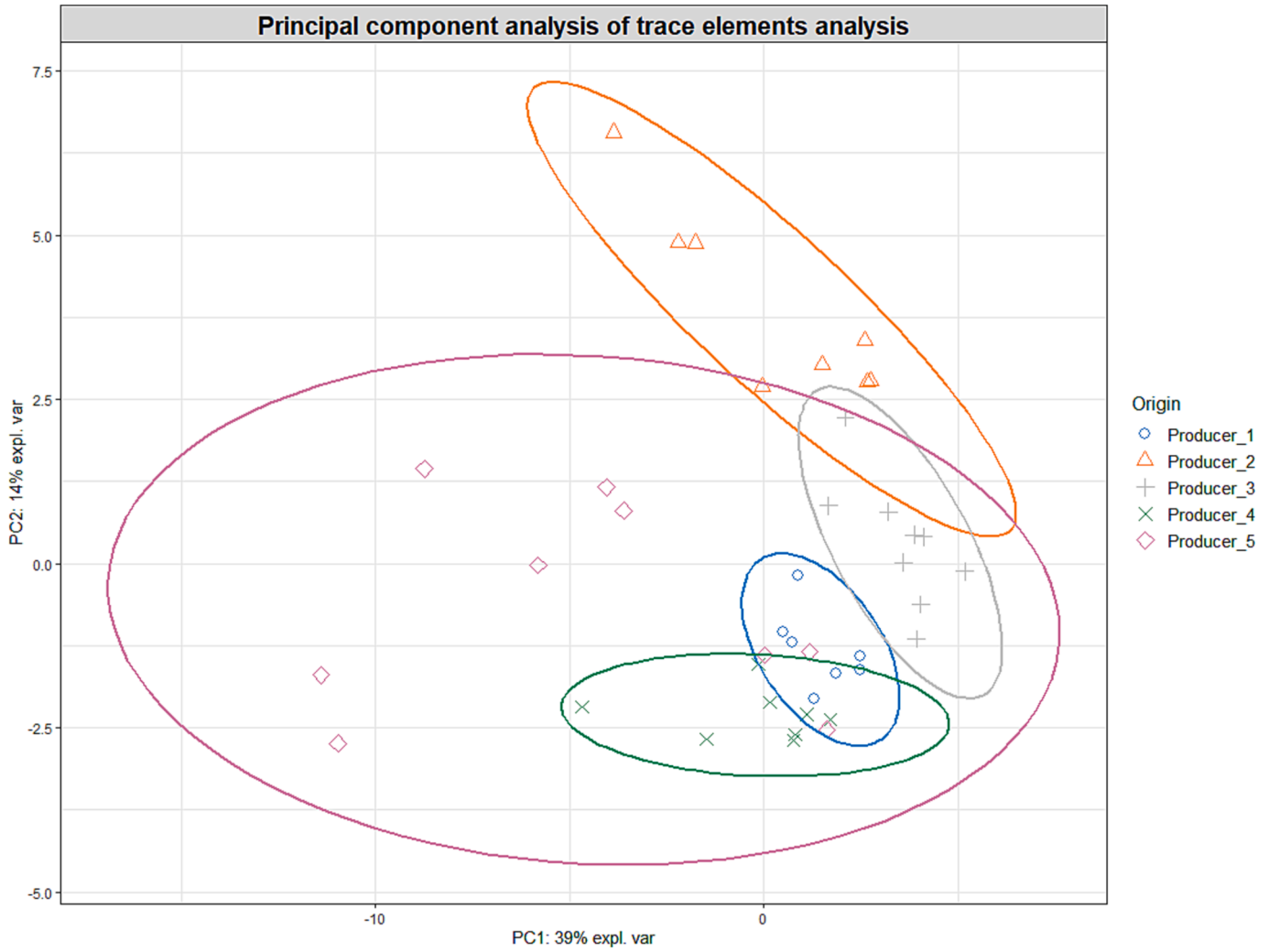

Fig. 8. Principal component analysis of Caciotta elemental composition. Elemental composition data were transformed using the centered log-ratio method.

Table 6
Performance of the sPLS-DA model in classifying caciotta cheese producers based on elements composition through hold-out validation on the test sets.

Origin/ Parameter	Recall	Precision	Specificity	Classification accuracy rate (%)
Producer 1	1.00	1.00	1.00	
Producer 2	0.95	1.00	0.99	
Producer 3	1.00	0.95	1.00	99.00 ± 0.01
Producer 4	1.00	1.00	1.00	
Producer 5	1.00	1.00	1.00	

4. Discussion

Cheese origin (*i.e.* the combination of geographical origin and local environmental variables) has been reported to significantly affect cheese microbiota (Sun and D'Amico, 2021; Kamilari et al., 2022; Reuben et al., 2023). In our previous study (Cardin et al., 2023), we assessed the effect of different biotic and abiotic factors that shaped the microbiome of typical Caciotta cheese, finding that origin was a major contributor to the observed differences. A comparable effect was observed concerning the VOCs, highlighting the importance of microbiota's metabolic activity in the development of Caciotta's typicity. In the present study, we further evaluate the use of the microbiome as a mean to authenticate cheese origin through a comparison of multiple methods, including -omics analyses, such as metagenomics and metabolomics, encompassing both volatilome and elemental metabolomics. To achieve this, we compared it with both established reference analyses and emerging methods. Additionally, we used a two-year sampling period to assess the temporal stability of each method (Riedl et al., 2015).

4.1. Bacterial and viral communities

Among DNA-based methods, the classification accuracy rate of the viral communities outperformed that of the bacterial ones (97.42 ± 2.58 vs 96.13 ± 4.02 %). For both communities, we noticed origin specific non-starter lactic acid bacteria, like *Lactobacillus helveticus, Lactococcus raffinolactis* and *Propionibacterium freudenreichii,* or phage, like Salmonella virus PsP3 and Enterobacteria phage ES18. Corroborating these findings, Dugat-Bony et al. (2016) reported the presence of specific operational taxonomic units in one or several of the 12 analyzed cheese varieties, forming distinct patterns with the cheese production facility. Similarly, the taxa identified in our study showed the highest score in the prediction, suggesting their discrimination ability for origin authentication.

To the best of our knowledge, only one other study has evaluated the performances of cheese microbiota in authenticating cheese origin (Kamilari et al., 2022). Our results showed lower prediction performances than those reported using Random Forest algorithm. In particular, for the former authors, the area under the curve (*i.e.* the ratio between specificity and sensitivity) was found to be equal to 1, meaning that all the tested samples were correctly classified according to their

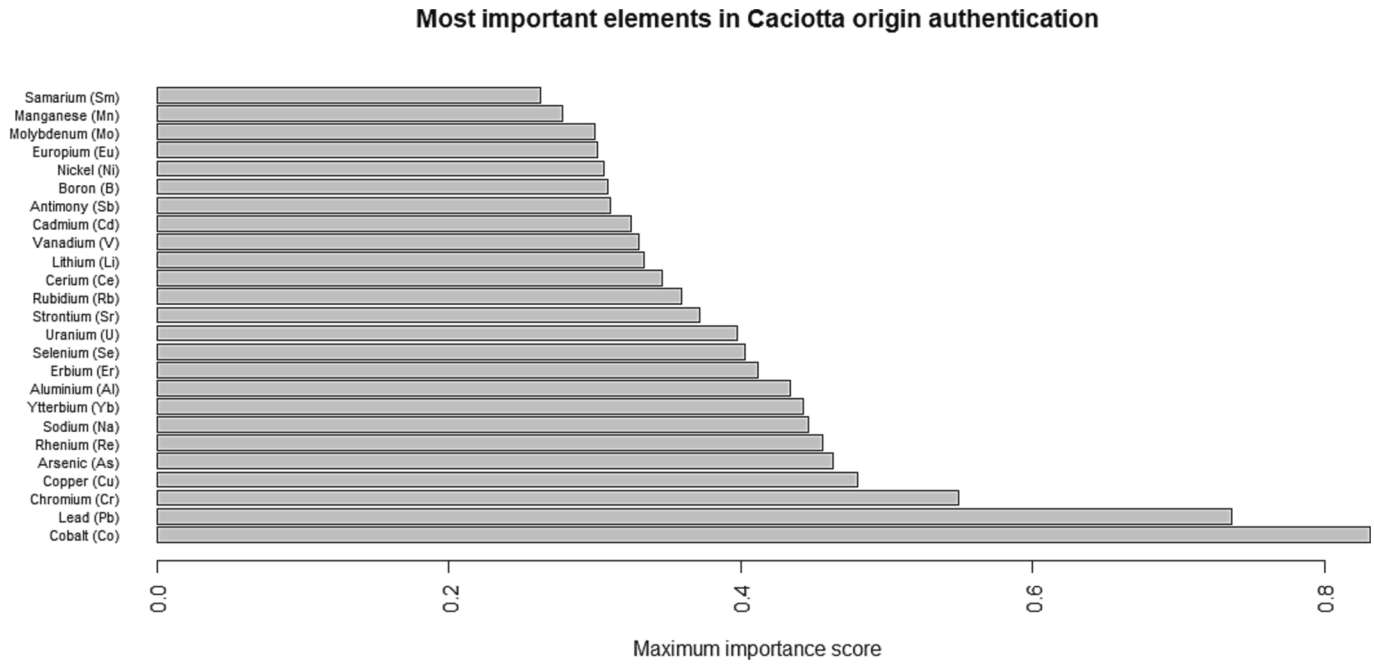

Fig. 9. Mean maximum importance score of the most important 25 elements used in the sPLS-DA models.

Table 7
Tukey's honest significance test of isotopes concentration in Caciotta cheese.

Origin	$\delta^{13}C$ (‰, vs V-PDB)	$\delta^{15}N$ (‰, vs AIR)	$\delta^{34}S$ (‰, vs V-CDT)	$\delta^{18}O$ (‰, vs V-SMOW)	$\delta^{2}H$ (‰, vs V-SMOW)
Producer 1	−25.50 ± 0.52^{a}	3.15 ± 1.23^{ac}	4.98 ± 1.75^{a}	7.50 ± 1.64^{ab}	−114.32 ± 4.64^{a}
Producer 2	−23.65 ± 0.68^{b}	5.39 ± 1.03^{b}	3.32 ± 0.96^{a}	5.82 ± 1.84^{a}	−120.02 ± 4.15^{b}
Producer 3	−24.18 ± 0.37^{ab}	3.73 ± 0.59^{ac}	5.30 ± 1.35^{a}	6.71 ± 1.31^{ab}	−114.05 ± 3.66^{a}
Producer 4	−22.88 ± 1.76^{b}	4.59 ± 0.63^{abc}	4.84 ± 1.42^{a}	8.21 ± 1.10^{b}	−109.02 ± 2.07^{ac}
Producer 5	−24.27 ± 1.12^{ab}	4.69 ± 0.72^{abc}	4.32 ± 1.51^{a}	7.86 ± 0.87^{b}	−108.60 ± 3.80^{c}

origin. The differences between the two studies could be connected to potential overfitting of the developed model and/or excessively small training dataset. More studies are needed to really assess the performances of DNA-based authentication methods.

4.2. Volatilome

Volatile organic compounds showed potential clusters according to cheese origin. Among others, two ketones, namely, 3-hydroxybutan-2-one (acetoin) and 1-acetophenone, were identified as significant compounds for predicting origin. Acetoin was extensively studied (Lo et al., 2018), and its formation involves the conversion of pyruvate from citrate metabolism, followed by the synthesis and decarboxylation of acetolactate, leading to acetoin. On the other hand, acetophenone can be produced as an intermediate product through the deamination of phenylalanine (Sieber et al., 1995), the microbial breakdown of tyrosine (Parliment et al., 1982), or as a minor product from the Strecker degradation of phenylalanine (Griffith and Hammond, 1989). A study on Pecorino Romano cheese showed that 87.5 % of samples were correctly classified after cross-validation while another study on Graviera cheese reported only 47.5 % of correctly classified samples (Di Donato et al., 2021; Vatavali et al., 2020) whereas in the present study 90.0 ± 11.11 % of mountain Caciotta cheese was correctly classified suggesting that the discriminative power of VOC profile for cheese authentication could vary considerably according to the considered cheese type and methodological approach as mentioned by Cardin et al. (2022).

4.3. Near infrared spectroscopy

To our knowledge, no comparison of the performances of NIRS and DNA-based methods or others reference analysis exists for cheeses. In this study the classification accuracy rate obtained using NIRS to authenticate cheese origin was 76.0 ± 31.57 % highlighting a considerable error in prediction's accuracy. The performance was primarily influenced by three areas of the IR spectrum, where important wavelengths for clustering were observed. In region I and III the importance of the selected wavelength might be associated with different fat content and/or different lipidic profile. Indeed, Silva et al. (2022) reported that vibrational and rotational motions of C–H/C = O groups could be associated with different wavelengths ranging from 1200 to 1214 and 2234 to 2348 nm. On the other hand, region II could be associated with a different protein profile/proteolytic activity, since the range from 1620 to 1700 nm is associated with the N–H bond (Alinovi et al., 2019). Considering the model performances, our results led to lower accuracy compared to the reported authentication efficacy of Emmental PDO (85.7 %) or the discrimination of cheeses from different dairy systems (67.1 %) (Karoui et al., 2005; Bergamaschi et al., 2020) possibly due to the limited size of the investigated samples (Niemöller and Holroyd, 2019).

4.4. Element analysis

Elements in mountain Caciotta cheese showed characteristic signature for each producer. The elemental pathway from farm to the final food product is regulated by the bioavailability of elements, which is influenced by the elemental composition of soil, water, feed, and, subsequently, by animals and cheesemaking process (Danezis and Georgiou, 2022). In the present study, this complex interplay could have been influenced by differences in soil composition between the Veneto and Trentino regions. For example, the Pedemontana area of the Veneto region (producers 4 and 5) lies on limestone rocks and is characterized

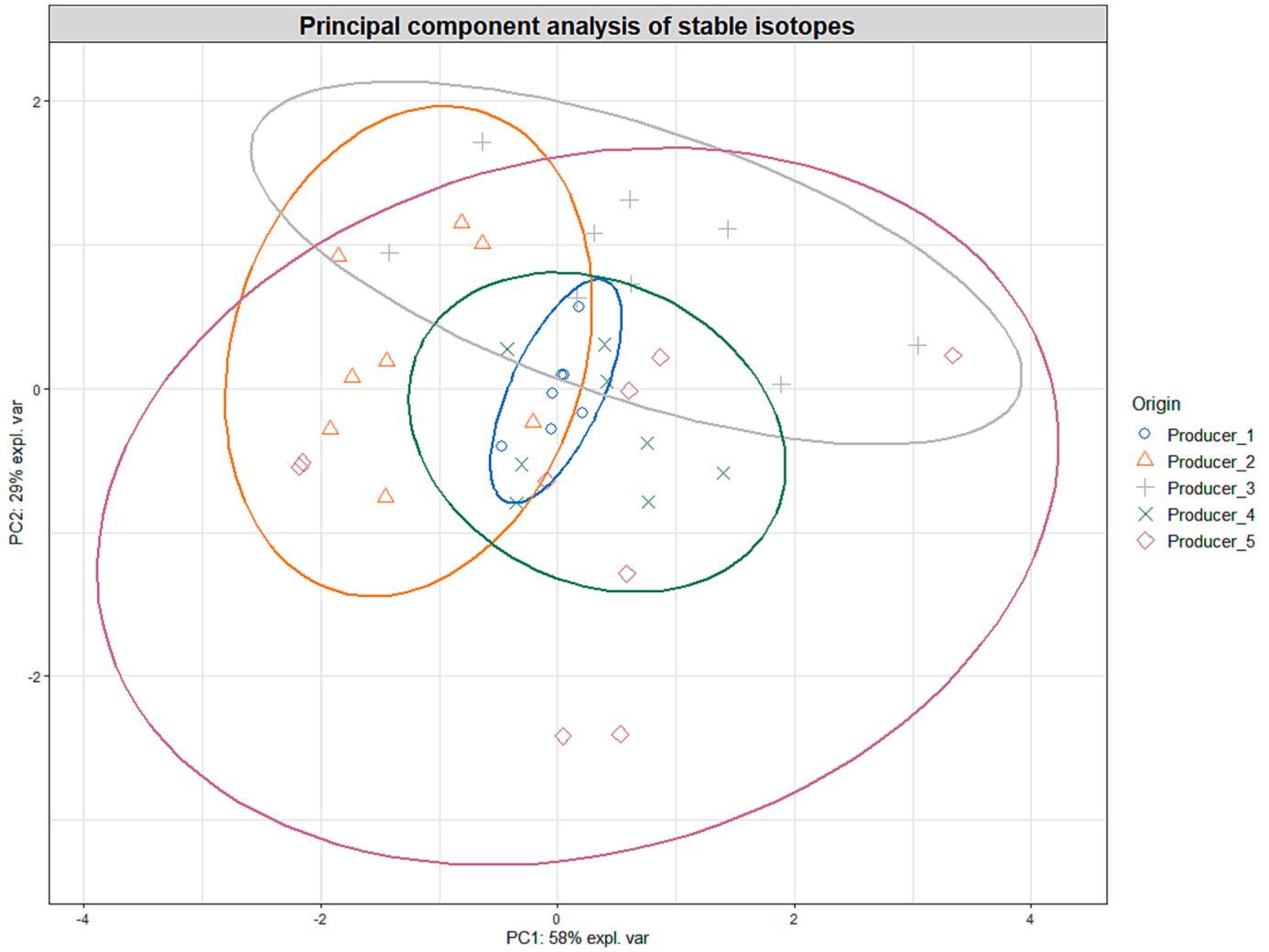

Fig. 10. Principal component analysis of Caciotta stable isotopes composition. Isotopes composition data were transformed using the centered log-ratio method.

Table 8
Performance of the sPLS-DA model in classifying caciotta cheese producers based on stable isotopes through hold-out validation on the test sets.

Origin/ Parameter	Recall	Precision	Specificity	Classification accuracy rate (%)
Producer 1	0.5	0.33	0.85	
Producer 2	0.90	0.62	0.96	
Producer 3	0.00	0.00	0.76	65.00 ± 53.84
Producer 4	1.00	0.91	1.00	
Producer 5	0.85	1.00	0.94	

by normal salinity and alkaline/subalkaline acidity (ARPAV, 2020). These factors are known to affect element availability for plants (Tyler and Olsson, 2001) and could therefore affect the composition of milk and cheese. On the other hand, the area of Trento (producers 1, 2 and 3) is characterized by galena, chalcopyrite, and sphalerite but also accessory minerals, such as tetrahedrite, tennantite, acanthite and sulfosalts rich of As, Bi and Sb that are associated with Pb–Cu–Zn mineralization (Bianchini et al., 2019). Some toxic metalloids had concentrations above the limit of quantification (LOQ), which further prompted their selection for prediction purposes. The benchmark dose lower bound for a 0.5 % response of inorganic arsenic is 3 μg/kg body weight for lung cancer, whereas the tolerable daily intake of antimony is 6 μg/kg body weight (WHO, 2011; WHO, 2022). Considering that the highest concentrations observed for arsenic and antimony were 1.2 and 0.28 μg/kg of dry matter, respectively and the average moisture was 37.57 %, we can exclude the potential harmful effects connected to the consumption of these cheeses.

Element hold-out validated models demonstrated an excellent correct classification rate, achieving 99.00 ± 0.01 % accuracy in correctly classifying Caciotta origin. Among the most important elements for the prediction, we found a majority of those classified in the "other elements" category, such as cobalt, lead and arsenic, elements connected with dairy equipment (*e.g.* copper and aluminum) and rare earths such as cerium and europium which were found to significantly differ among origins. The high correct classification rate and low accuracy error were similar to those observed in the studies of Danezis et al. (2019 and 2020) in which classification accuracy rate of 95.9 % and 92.1 % were achieved. Nevertheless, while rare earths had critical importance in these studies' models, their relevance in Caciotta authentication was lower than that of other elements.

4.5. *Stable isotopes*

Stable isotope ratios of mountain Caciotta was investigated as a reference method for origin authentication. As is known, the botanical origin influences the isotopic composition of the plant, and consequently, it also affects the isotopic composition of the animals that consume it as part of their diet, as well as that of the products obtained from the animals, such as milk and meat. C4 plants like maize or cane have $\delta^{13}C$ values between −14 and −12 ‰ while C3 plants like grass or

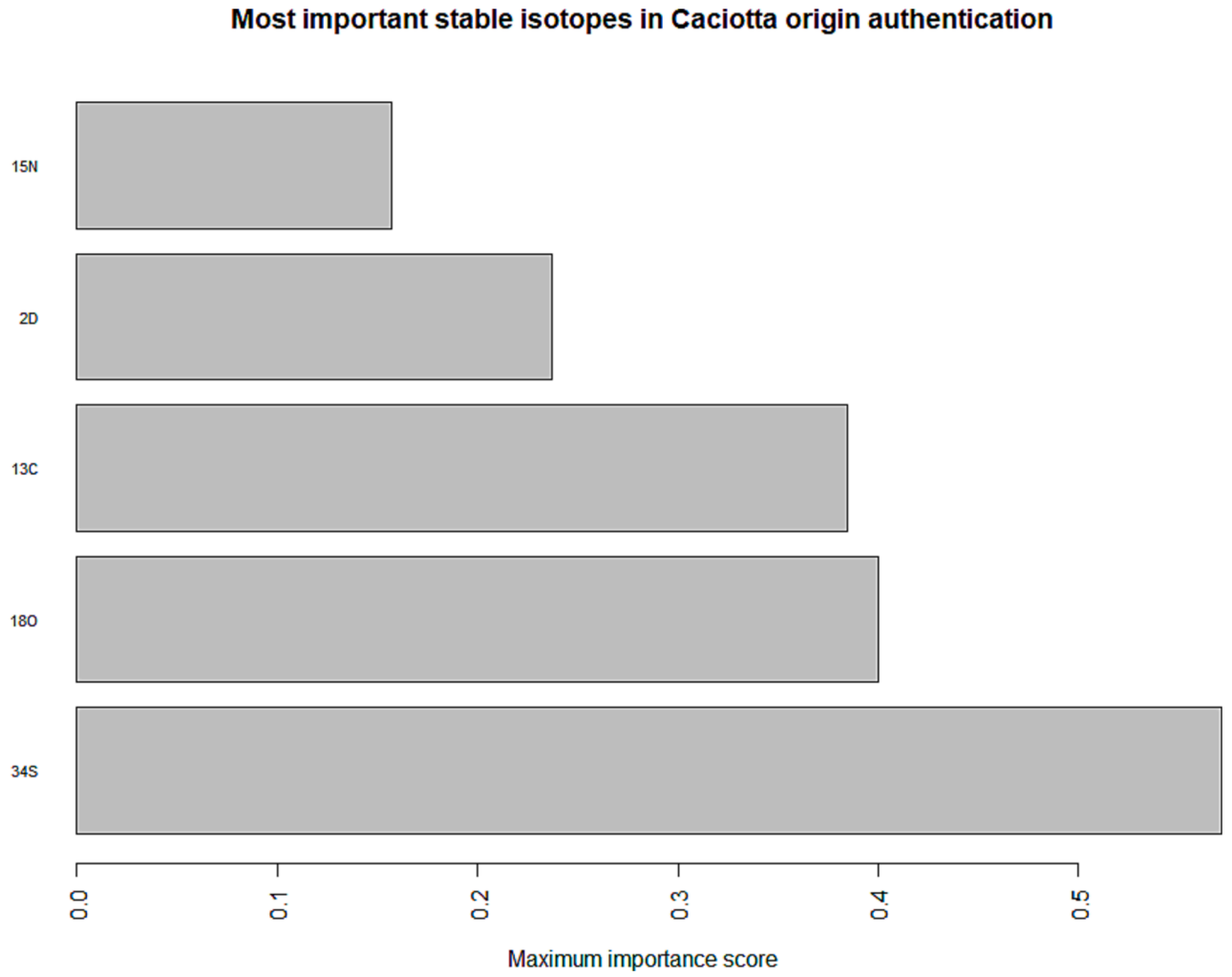

Fig. 11. Mean maximum importance score of the stable isotopes used in the sPLS-DA models.

fruit plants range from −30 to –23 ‰ (O'Leary, 1988). For $\delta^{13}C$ the values observed for producers 1,3 and 5 reflect a diet made of C3 forage and grass while those of producers 2 and 4 seems to indicate that some supplements in the cow diet (*e.g.* corn, a C4 cycle plant) would be commonly used for milk production.

The main factor affecting $\delta^{15}N$ in cultivated plants is the fertilization process used. Synthetic fertilizers, produced from atmospheric nitrogen via the Haber process, have $\delta^{15}N$ values between −4 and + 4 ‰, while organic fertilizers are characterized by values between + 0.6 and + 36.7 ‰ (Vitòria et al., 2004). Based on the isotopic composition of nitrogen, producers 2, 4 and 5 seem to be the ones that make the most use of manure as a pasture amending technique. However, it should be noted that within each farm the values were highly variable.

The sulfur isotopic ratio is influenced by many different factors, such as abundance of sulfides in soil, plants' aerobic and anaerobic growth, local bedrocks (Rubenstein and Hobson, 2004), active microbial process in the soil, fertilization procedures, active deposition (Krouse and Mayer, 2000). In this case, the isotopic composition of the plants is reflected in that of the products (such as milk and, consequently, cheese) obtained by the animal that consumed those plants as part of its diet. The producer 3 is close to the peat bog of Echen, consisting of glacial deposits of late Pleistocene and on a cretaceous limestone substrate that is an intermittent glacio-karst basin. When atmospheric sulfate settles above on water-logged peat bogs and sulfate ions reach the anaerobic layer, sulfate reducing bacteria produce sulfide, causing S and O isotope fractionation (dissimilatory reduction). The residual sulfate becomes progressively enriched in the heavy isotopes $\delta^{18}O$ and $\delta^{34}S$ (Novák et al., 2005). The $\delta^{34}S$ mean value found for the producer 3 was tendentially the higher among the five sites of study, two deltas higher than producer 2, that could be attributed to the peculiar wetland area where it is. However, any significant difference between the $\delta^{34}S$ stable isotopes was observed.

The hydrogen isotope ratio in animal proteins is mainly correlated with that of the feed consumed. However, Hobson et al. (1999) have shown that approximately 30 % of it comes from drinking water. A connection has been established between the isotopic signature of meteoric water and the $\delta^{18}O$ and $\delta^{2}H$ values of local water. This relationship depends on factors such as latitude, altitude, distance from the sea or lake, and temperature (Camin et al., 2012; Giustini et al., 2016). Furthermore, due to plant transpiration, the water in livestock feed is significantly enriched in its isotopic composition compared to the water that the plant absorbs from the soil (Roden and Ehleringer, 2000). All these factors impact the two isotopic ratios $\delta^{18}O$ and $\delta^{2}H$ and are the basis of their use as geographic tracers. The producers 4 and 5 had $\delta^{18}O$ values significantly higher than those of producers 2 and 3. The former are located at lower latitude and altitude whereas the latter are both at altitude higher than 1160 m a.s.l. and latitude close to or higher than 46°. The producer 1, whose $\delta^{18}O$ value is between the two groups, is at latitude above 46° but an altitude similar to that of producers 4 and 5. In the case of $\delta^{2}H$ the producers 4 and 5 had values significantly higher than those of producers 1 and 2. while the producer 3, although located at 1169 m a.s.l., had $\delta^{2}H$ levels relatively high probably due to the

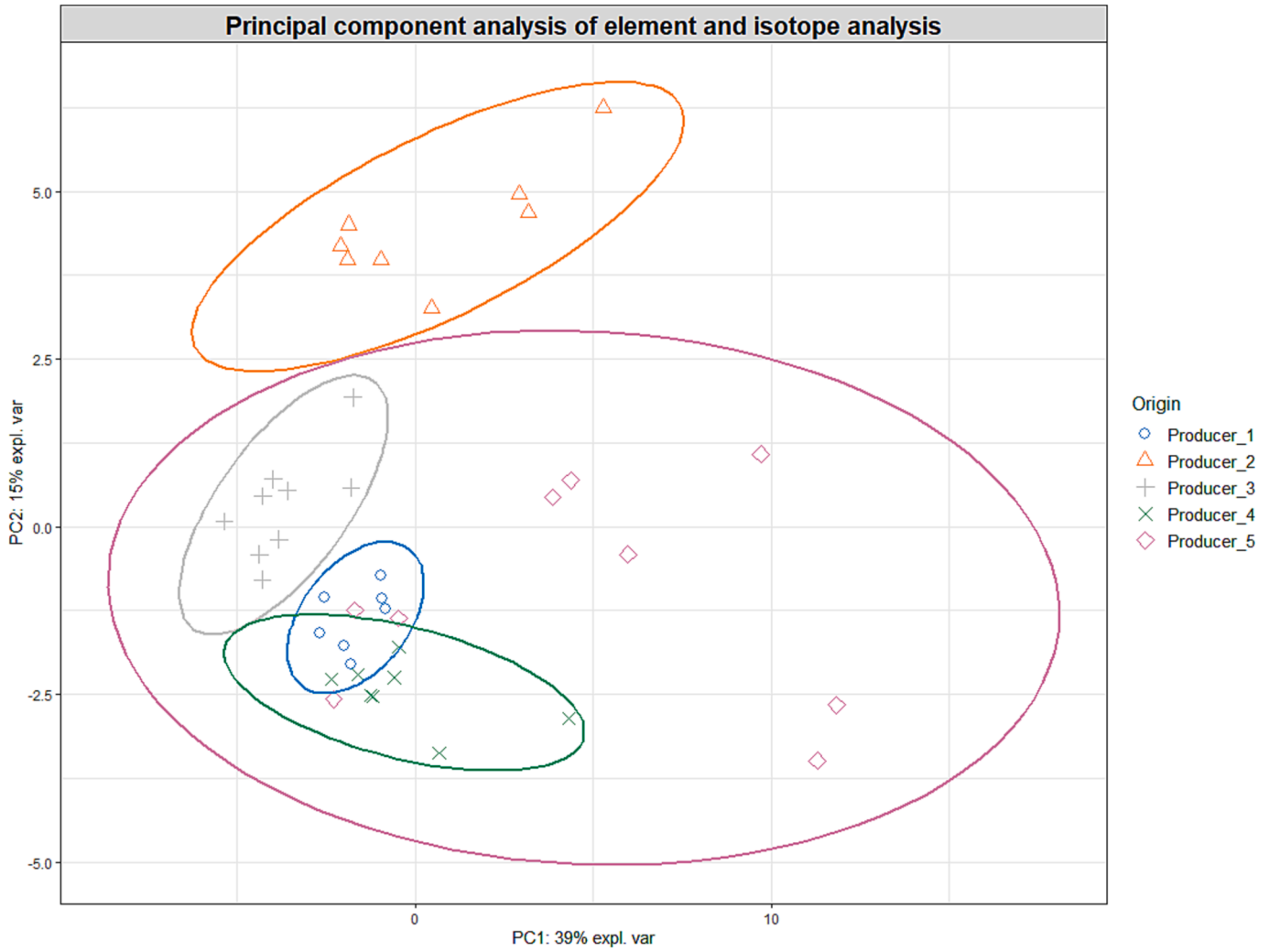

Fig. 12. Principal component analysis of Caciotta stable isotopes and elements composition. Isotopes and elements composition data were transformed using the centered log-ratio method.

Table 9
Performance of the sPLS-DA model in classifying caciotta cheese producers based on elements and stable isotopes through hold-out validation on the test sets.

Origin/ Parameter	Recall	Precision	Specificity	Classification accuracy rate (%)
Producer 1	1.00	1.00	1.00	
Producer 2	1.00	0.90	1.00	
Producer 3	0.95	1.00	0.98	97.00 ± 3.09
Producer 4	1.00	1.00	1.00	
Producer 5	0.9	0.94	0.97	

proximity to the wetland, whose peculiar microenvironment can impact on isotopes fractionation.

The hold-out validated sPLS-DA models yielded poor performances in classification accuracy rate (65.00 ± 53.84 %), characterized by a high error in accuracy. Bontempo et al. (2012), using a similar combination of isotopes, studied the origin of cheese produced in two geographical sites highly different each other in terms of distance, latitude and altitudes achieving a classification accuracy rate of 80 %.

Lastly, upon combining elements and isotopes, we observed a decrease in the authentication performances (97.00 ± 3.09 %) compared to the original element models. These differences could be probably explained by the short distance both in terms of kilometers and in terms of latitude/altitude among the investigated sites of Caciotta cheese making.

5. Conclusions

An approach based on five different analytical methods (DNA shotgun metagenomics for bacterial and viral community profiling, volatilome, stable isotopes ratio, elements and near infrared spectroscopy) has been used to highlight significant differences connected to origin and/or origin-specific features that could potentially discriminate against 5 different sites of cheesemaking located in a restricted geographical area. The overall classification accuracy rate varied from stable isotope ratio (65.00 ± 53.84 %) to element analysis (99.00 ± 0.01 %). DNA-based methods obtained appreciable classification accuracy, with the viral communities reaching the second most accurate method (97.42 ± 2.58 %), followed by the combination of stable isotopes and element analysis (97.00 ± 3.09 %), the bacterial communities (96.13 ± 4.02 %) and the volatilome (90.0 ± 11.11). Average and below-average performances were obtained for near-infrared spectroscopy and stable isotope ratio analysis, possibly connected to the limited number of analyzed samples and the close proximity of cheese origin, which can be a starting point for future in-depth analyses. Therefore, at least some of the methods described here can be used to support the traceability procedures of food production as required by Reg. EU 178/2002 and product authenticity in terms of geographical origin as

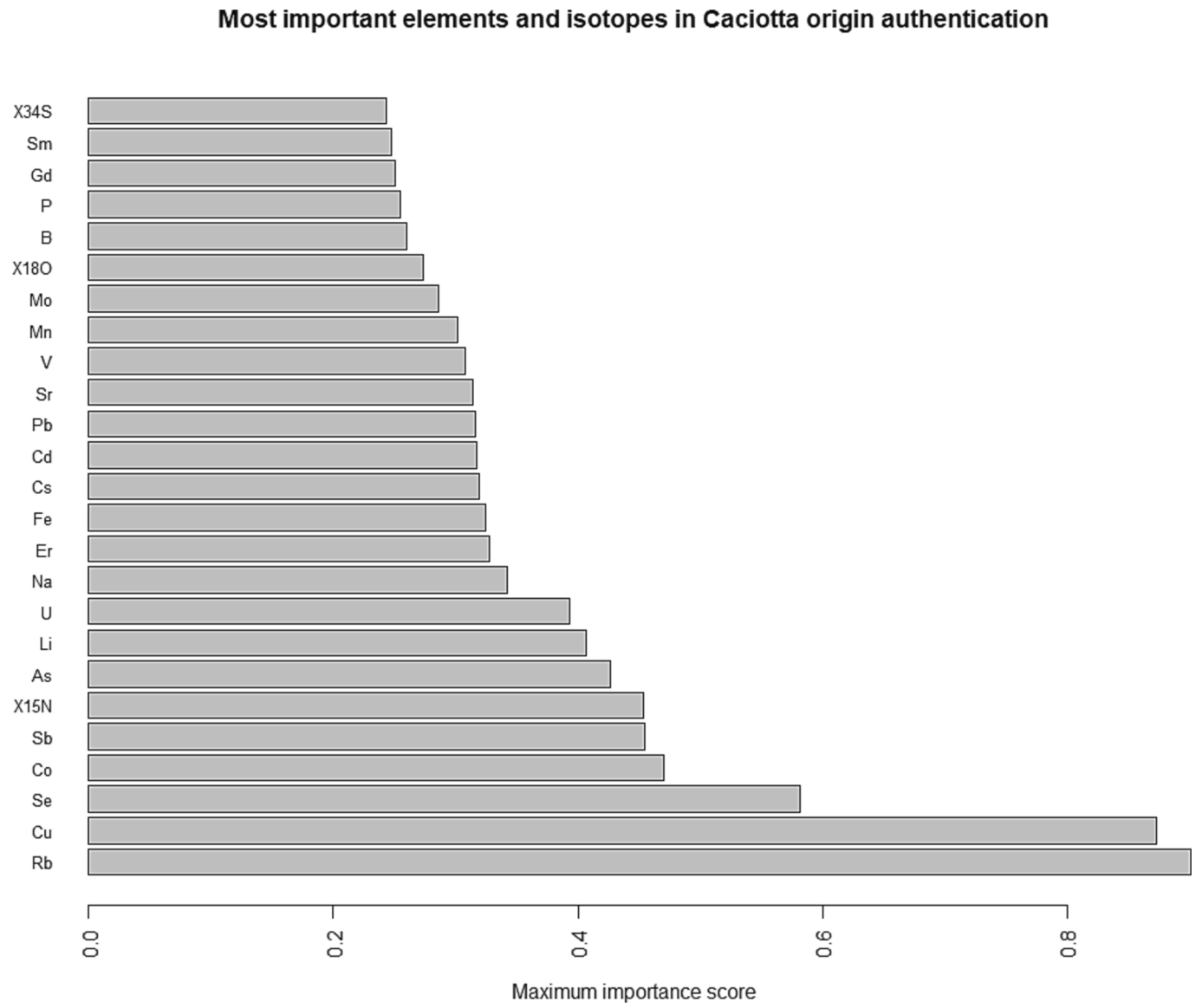

Fig. 13. Mean maximum importance score of the stable isotopes and elements sPLS-DA models.

required by Reg. EU 1169/ 2011.

CRediT authorship contribution statement

Marco Cardin: Resources, Investigation, Data Curation, Formal analysis, Visualization, Writing - Original Draft, Writing - review & editing. **Jérôme Mounier:** Conceptualization, Supervision, Writing - review & editing. **Emmanuel Coton:** Conceptualization, Supervision, Funding acquisition, Writing - review & editing. **Barbara Cardazzo:** Conceptualization, Supervision. **Matteo Perini:** Methodology, Data Curation. **Daniela Bertoldi:** Data curation, Methodology, Writing – review & editing. **Silvia Pianezze:** Data curation, Methodology. **Severino Segato:** Methodology, Data Curation. **Barbara Di Camillo:** Conceptualization, Supervision. **Marco Cappellato:** Methodology. **Monika Coton:** Methodology, Investigation, Writing - review & editing. **Lisa Carraro:** Methodology. **Sarah Currò:** Methodology, Data Curation, Writing - review & editing. **Rosaria Lucchini:** Resources. **Hooriyeh Mohammadpour:** Methodology. **Enrico Novelli:** Conceptualization, Supervision, Funding acquisition, 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.

References

Abbas, O., Zadravec, M., Baeten, V., Mikuš, T., Lešić, T., Vulić, A., & Pleadin, J. (2018). Analytical methods used for the authentication of food of animal origin. *Food Chemistry, 246*, 6–17.

Alinovi, M., Mucchetti, G., & Tidona, F. (2019). Application of NIR spectroscopy and image analysis for the characterisation of grated Parmigiano-Reggiano cheese. *International Dairy Journal, 92*, 50–58.

ARPAV 2020. Legenda della carta dei suoli in scala 1:50.000. Available at https://gaia.arpa.veneto.it/documents/772 (accessed July 26, 2023).

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife, 10*, e65088.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal statistical society: Series B (Methodological), 57*(1), 289–300.
Bergamaschi, M., Cipolat-Gotet, C., Cecchinato, A., Schiavon, S., & Bittante, G. (2020). Chemometric authentication of farming systems of origin of food (milk and ripened cheese) using infrared spectra, fatty acid profiles, flavor fingerprints, and sensory descriptions. *Food chemistry, 305*, Article 125480.
Bianchini, G., Natali, C., Ferretti, P., Casagrande, L., Conedera, M., & Marchina, C. (2019). Trace-Element Distribution on Sulfide Mineralization in Trento Province. *NE Italy. Minerals, 9*, 736.
Bisutti, V., Merlanti, R., Serva, L., Lucatello, L., Mirisola, M., Balzan, S., & Capolongo, F. (2019). Multivariate and machine learning approaches for honey botanical origin authentication using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 27*(1), 65–74.
Bontempo, L., Barbero, A., Bertoldi, D., Camin, F., Larcher, R., Perini, M., Sepulcri, A., Zicarelli, L., & Piasentier, E. (2019). Isotopic and elemental profile of Mediterranean buffalo milk and cheese and authentication of Mozzarella di Bufala Campana PDO: An initial exploratory study. *Food Chemistry, 285*, 316–323.
Bontempo, L., Lombardi, G., Paoletti, R., Ziller, L., & Camin, F. (2012). H, C, N and O stable isotope characteristics of alpine forage, milk and cheese. *International dairy journal, 23*(2), 99–104.
Bontempo, L., Larcher, R., Camin, F., Hölzl, S., Rossmann, A., Horn, P., & Nicolini, G. (2011). Elemental and isotopic characterisation of typical Italian alpine cheeses. *International dairy journal, 21*(6), 441–446.
Brand, W. A., Coplen, T. B., Vogl, J., Rosner, M., & Prohaska, T. (2014). Assessment of international reference materials for isotope-ratio analysis (IUPAC Technical Report). *Pure and Applied Chemistry, 86*(3), 425–467.
Brown, J., Pirrung, M., & McCue, L. A. (2017). FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics, 33*(19), 3137–3139.
Camin, F., Bertoldi, D., Santato, A., Bontempo, L., Perini, M., Ziller, L., & Larcher, R. (2015). Validation of methods for H, C, N and S stable isotopes and elemental analysis of cheese: Results of an international collaborative study. *Rapid Communications in Mass Spectrometry, 29*(5), 415–423.
Camin, F., Wehrens, R., Bertoldi, D., Bontempo, L., Ziller, L., Perini, M., & Larcher, R. (2012). H, C, N and S stable isotopes and mineral profiles to objectively guarantee the authenticity of grated hard cheeses. *Analytica Chimica Acta, 711*, 54–59.
Camin, F., Perini, M., Colombari, G., Bontempo, L., & Versini, G. (2008). Influence of dietary composition on the carbon, nitrogen, oxygen and hydrogen stable isotope ratios of milk. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry, 22*(11), 1690–1696.
Camin, F., Bontempo, L., Heinrich, K., Horacek, M., Kelly, S. D., Schlicht, C., & Rossmann, A. (2007). Multi-element (H, C, N, S) stable isotope characteristics of lamb meat from different European regions. *Analytical and bioanalytical chemistry, 389*, 309–320.
Camin, F., Wietzerbin, K., Cortes, A. B., Haberhauer, G., Lees, M., & Versini, G. (2004). Application of multielement stable isotope ratio analysis to the characterization of French, Italian, and Spanish cheeses. *Journal of agricultural and food chemistry, 52*(21), 6592–6601.
Cardin, M., Cardazzo, B., Coton, M., Carraro, L., Lucchini, R., Novelli, E., Coton, E., & Mounier, J. (2023). Ecological diversity and associated volatilome of typical mountain Caciotta cheese from Italy. *International Journal of Food Microbiology, 110523*.
Cardin, M., Cardazzo, B., Mounier, J., Novelli, E., Coton, M., & Coton, E. (2022). Authenticity and Typicity of Traditional Cheeses: A Review on Geographical Origin Authentication Methods. *Foods, 11*(21), 3379.
Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research, 295*(2), 295–336.
Chung, D., & Keles, S. (2010). Sparse partial least squares classification for high dimensional data. *Statistical applications in genetics and molecular biology, 9*(1).
Currò, S., Fasolato, L., Serva, L., Boffo, L., Ferlito, J. C., Novelli, E., & Balzan, S. (2022). Use of a portable near-infrared tool for rapid on-site inspection of freezing and hydrogen peroxide treatment of cuttlefish (Sepia officinalis). *Food Control, 132*, Article 108524.
Danezis, G. P., & Georgiou, C. A. (2022). Elemental metabolomics: Food elemental assessment could reveal geographical origin. *Current Opinion in Food Science, 44*, Article 100812.
Danezis, G. P., Pappas, A. C., Tsiplakou, E., Pappa, E. C., Zacharioudaki, M., Tsagkaris, A. S., & Georgiou, C. A. (2020). Authentication of Greek Protected Designation of Origin cheeses through elemental metabolomics. *International Dairy Journal, 104*, Article 104599.
Di Donato, F., Biancolillo, A., Mazzulli, D., Rossi, L., & D'Archivio, A. A. (2021). HS-SPME/GC–MS volatile fraction determination and chemometrics for the discrimination of typical Italian Pecorino cheeses. *Microchemical Journal, 165*, Article 106133.
Dugat-Bony, E., Garnier, L., Denonfoux, J., Ferreira, S., Sarthou, A. S., Bonnarme, P., & Irlinger, F. (2016). Highlighting the microbial diversity of 12 French cheese varieties. *International Journal of Food Microbiology, 238*, 265–273.
Farnham, I. M., Singh, A. K., Stetzenbach, K. J., & Johannesson, K. H. (2002). Treatment of nondetects in multivariate analysis of groundwater geochemistry data. *Chemometrics and Intelligent Laboratory Systems, 60*(1–2), 265–281.
Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences, 103*(3), 626–631.
Frétin, M., Martin, B., Rifa, E., Isabelle, V. M., Pomiès, D., Ferlay, A., & Delbès, C. (2018). Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems. *Scientific reports, 8*(1), 200.
Giustini, F., Brilli, M., & Patera, A. (2016). Mapping oxygen stable isotopes of precipitation in Italy. *Journal of Hydrology: Regional Studies, 8*, 162–181.
Goerges, S., Mounier, J., Rea, M. C., Gelsomino, R., Heise, V., Beduhn, R., & Scherer, S. (2008). Commercial ripening starter microorganisms inoculated into cheese milk do not successfully establish themselves in the resident microbial ripening consortia of a South German red smear cheese. *Applied and environmental microbiology, 74*(7), 2210–2217.
Griffith, R., & Hammond, E. G. (1989). Generation of Swiss cheese flavor components by the reaction of amino acids with carbonyl compounds. *Journal of Dairy Science, 72*(3), 604–613.
Hobson, K. A., Atwell, L., & Wassenaar, L. I. (1999). Influence of drinking water and diet on the stable-hydrogen isotope ratios of animal tissues. *Proceedings of the National Academy of Sciences, 96*(14), 8003–8006.
Kamilari, E., Tsaltas, D., Stanton, C., & Ross, R. P. (2022). Metataxonomic mapping of the microbial diversity of Irish and eastern Mediterranean cheeses. *Foods, 11*(16), 2483.
Kamilari, E., Tomazou, M., Antoniades, A., & Tsaltas, D. (2019). High throughput sequencing technologies as a new toolbox for deep analysis, characterization and potentially authentication of protection designation of origin cheeses?. International Journal of Food Science, 2019.
Karoui, R., Dufour, É., Pillonel, L., Schaller, E., Picque, D., Cattenoz, T., & Bosset, J. O. (2005). The potential of combined infrared and fluorescence spectroscopies as a method of determination of the geographic origin of Emmental cheeses. *International Dairy Journal, 15*(3), 287–298.
Kassambara, A. (2018). Machine learning essentials: Practical guide in R. *Sthda, 26*, 129–135.
Krouse, H. R., & Mayer, B. (2000). Sulphur and oxygen isotopes in sulphate. In *Environmental tracers in subsurface hydrology* (pp. 195–231). Boston, MA: Springer, US.
Lê Cao, K. A., Boitard, S., & Besse, P. (2011). Sparse PLS discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems. *BMC bioinformatics, 12*(1), 1–17.
Lê Cao, K. A., Costello, M. E., Lakis, V. A., Bartolo, F., Chua, X. Y., Brazeilles, R., & Rondeau, P. (2016). MixMC: A multivariate statistical framework to gain insight into microbial communities. *PloS one, 11*(8), e0160169.
Lei, T., & Sun, D. W. (2019). Developments of nondestructive techniques for evaluating quality attributes of cheeses: A review. *Trends in Food Science & Technology, 88*, 527–542.
Lo, R., Bansal, N., & Turner, M. S. (2018). The genetic basis underlying variation in production of the flavour compound diacetyl by Lactobacillus rhamnosus strains in milk. *International journal of food microbiology, 265*, 30–39.
Medina, S., Perestrelo, R., Silva, P., Pereira, J. A., & Câmara, J. S. (2019). Current trends and recent advances on food authenticity technologies and chemometric approaches. *Trends in Food Science & Technology, 85*, 163–176.
Menozzi, D., Yeh, C. H., Cozzi, E., & Arfini, F. (2022). Consumer Preferences for Cheese Products with Quality Labels: The Case of Parmigiano Reggiano and Comté. *Animals, 12*(10), 1299.
Muñoz-Redondo, J. M., Bertoldi, D., Tonon, A., Ziller, L., Camin, F., & Moreno-Rojas, J. M. (2022). Multi-element and stable isotopes characterization of commercial avocado fruit (Persea americana Mill) with origin authentication purposes. *Food Control, 137*(2022), Article 108975.
Niemöller, A., & Holroyd, S. (2019). Applications of near infrared spectrometry for the analysis of milk and milk products. *Bulletin of the International Dairy Federation, 497*.
Novák, M., Vile, M. A., Bottrell, S. H., Štěpánová, M., Jačková, I., Buzek, F., … Newton, R. J. (2005). Isotope systematics of sulfate-oxygen and sulfate-sulfur in six European peatlands. *Biogeochemistry, 76*, 187–213.
O'Leary, M. H. (1988). Carbon isotopes in photosynthesis. *Bioscience, 38*(5), 328–336.
Parliment, T. H., Kolor, M. G., & Rizzo, D. J. (1982). Volatile components of Limburger cheese. *Journal of Agricultural and Food Chemistry, 30*(6), 1006–1008.
Pianezze, S., Bontempo, L., Perini, M., Tonon, A., Ziller, L., Franceschi, P., & Camin, F. (2020). δ34S for tracing the origin of cheese and detecting its authenticity. *Journal of Mass Spectrometry, 55*(7), e4451.
Pillonel, L., Ampuero, S., Tabacchi, R., & Bosset, J. (2003). Analytical methods for the determination of the geographic origin of Emmental cheese: Volatile compounds by GC/MS-FID and electronic nose. *European Food Research and Technology, 216*, 179–183.
Reuben, R. C., Langer, D., Eisenhauer, N., & Jurburg, S. D. (2023). Universal drivers of cheese microbiomes. *Iscience, 26*(1).
Riedl, J., Esslinger, S., & Fauhl-Hassek, C. (2015). Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Analytica Chimica Acta, 885*, 17–32.
Roden, J. S., & Ehleringer, J. R. (2000). There is no temperature dependence of net biochemical fractionation of hydrogen and oxygen isotopes in tree-ring cellulose. *Isotopes in environmental and health studies, 36*(3), 303–317.
Rohart, F., Gautier, B., Singh, A., & Lê Cao, K. A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS computational biology, 13*(11), e1005752.
Rubenstein, D. R., & Hobson, K. A. (2004). From birds to butterflies: Animal movement patterns and stable isotopes. *Trends in ecology & evolution, 19*(5), 256–263.
Sammarco, G., Bardin, D., Quaini, F., Dall'Asta, C., Christmann, J., Weller, P., & Suman, M. (2023). A Geographical Origin assessment of Italian Hazelnuts: Gas Chromatography-Ion mobility spectrometry coupled with Multivariate Statistical Analysis and Data Fusion approach. *Food Research International, 113085*.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., & Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature methods, 9*(8), 811–814.

Segato, S., Caligiani, A., Contiero, B., Galaverna, G., Bisutti, V., & Cozzi, G. (2019). 1H NMR metabolic profile to discriminate pasture based alpine Asiago PDO cheeses. *Animals, 9*(10), 722.

Sieber, R., Bütikofer, U., & Bosset, J. O. (1995). Benzoic acid as a natural compound in cultured dairy products and cheese. *International Dairy Journal, 5*(3), 227–246.

Silva, L. K., Santos, L. S., & Ferrão, S. P. (2022). Application of infrared spectroscopic techniques to cheese authentication: A review. *International Journal of Dairy Technology, 75*(3), 490–512.

Sun, L., & D'Amico, D. J. (2021). Composition, succession, and source tracking of microbial communities throughout the traditional production of a farmstead cheese. *Msystems, 6*(5), e00830–e00921.

Ting, K. M. (2011). Error Rate. In C. Sammut, & G. I. Webb (Eds.), *Encyclopedia of Machine Learning (pp 331)*. Boston, MA: Springer.

Tyler, G., & Olsson, T. (2001). Plant uptake of major and minor mineral elements as influenced by soil acidity and liming. *Plant and Soil, 230*, 307–321.

Vacheyrou, M., Normand, A. C., Guyot, P., Cassagne, C., Piarroux, R., & Bouton, Y. (2011). Cultivable microbial communities in raw cow milk and potential transfers from stables of sixteen French farms. *International journal of food microbiology, 146*(3), 253–262.

Vatavali, K., Kosma, I., Louppis, A., Gatzias, I., Badeka, A. V., & Kontominas, M. G. (2020). Characterisation and differentiation of geographical origin of Graviera cheeses produced in Greece based on physico-chemical, chromatographic and spectroscopic analyses, in combination with chemometrics. *International Dairy Journal, 110*, Article 104799.

Vitòria, L., Otero, N., Soler, A., & Canals, A. (2004). Fertilizer characterization: Isotopic data (N, S, O, C, and Sr). *Environmental science & technology, 38*(12), 3254–3262.

Wolfe, B. E., Button, J. E., Santarelli, M., & Dutton, R. J. (2014). Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. *Cell, 158*(2), 422–433.

World Health Organization. (2022). Guidelines for drinking-water quality: fourth edition incorporating the first and second addenda. Geneva. Licence: CC BY-NC-SA 3.0 IGO, 339-340.

World Health Organization. (2011). Evaluation of certain contaminants in food. *World Health Organization technical report series, 959*, 21–22.