UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN
"Science degli Alimenti"

CICLO XXXV

INNOVATIVE ANALYTICAL STRATEGIES COUPLED WITH MULTIVARIATE ANALYSIS TO DEAL WITH COMPLEX ISSUES IN FOOD INTEGRITY, FOOD AUTHENTICITY AND SENSOMICS

Coordinatore:
Chiar.mo Prof. Chiara Dall'Asta

Tutori:
Chiar.mo Prof. Chiara Dall'Asta
Prof. Michele Suman

Dottorando: Giuseppe Sammarco

Anni 2019-2020 / 2021-2022

Table of contents

Preface

Food is authentic or genuine if the product corresponds to the original conditions and the information on the label. Authentic commodities are free from adulteration, in terms of composition, nature, varietal purity, geographical origin and manufacturing method (Eurofins, s.d.). A recent European Committee for Standardization (CEN) standard defines food authenticity as the congruity between the *food product characteristics and the corresponding food claims* (CEN WS86). Food integrity has been described as the state of safety, quality, authenticity, traceability and genuineness of a food product, in all its aspect (Elliott, 2014). Thus, food integrity can be considered a multidisciplinary concept, including the entire food chain, from producer to consumer, and comprehending all the steps of food production, from the sourcing to the distribution (Alrobaish, Jacxsens, Luning & Vlerick, 2021). Illegal activities, that are threatening the authenticity and integrity of several commodities, are nowadays more and more common. Food fraud is defined as the commercialization of food to gain financial or economic profit by intentional deception of the consumers (European Commission, s.d.). To achieve a valuable level of health protection, to guarantee the consumers' right to information, and to protect the business of honest suppliers/producers from unfair competition, the European Union Regulation (EU) No. 1169/2011 (No 1169/2011 of the European Parliament and of the council of 25 October 2011 on the provision of food information to consumers, 2011) has required that the consumers should be properly informed about the food they consume. The labelling requirements, which are legally defined and vary according to the product, may include several features:

- The scientific name or breed
- The production method
- Ethical/religious issues (halal, vegetarian, vegan…)
- Nutritional composition (vitamins, omega3…)
- The production area, for both sustainability reasons and the EU legislation about the Protected Designation of Origin (PDO), Protected Geographical Indication (PGI), Traditional Specialties Guaranteed (TSG) …
- The status of the product (i.e., previously frozen and defrosted)
- The presence of undeclared ingredients potentially harmful for the consumers (allergens such as gluten, nuts…)

(Haynes, Jimenez, Pardo, & Helyar, 2019). The present Ph.D. work focused on the geographical origin assessment of strategic food commodities, widely employed by the main companies. Besides the authenticity, the geographical origin could also impact the food quality, since the production area could imply unique characteristics, such as special taste and flavour, that can be related to the history

of the product, the traditional practices, the processing methods, and the climatic conditions. Zero km foods from local producers and short supply chains for sustainable commodities represent two aspects, that are nowadays more and more relevant, related to food provenience and traceability (Stein & Santini, 2022). On the other hand, geographical indications could also be linked with negative aspects of human nutrition. In fact, several regions present a worrying amount of contaminants, such as arsenic, that are not strictly due to pollution, but the natural environment (Danezis & Georgiou, 2022). According to the Institute of Food Science Technology (IFST), detecting frauds and/or verifying the information's correctness involves an analytical test to evaluate the composition, the processing, the geographical origin, and compliance with certification systems. There are typical ways to classify the analytical approach:

- Targeted vs untargeted analysis
- Single-variate vs Multi-variate Analysis (MVA)
- Laboratory vs point-of-use testing

Targeted analyses refer to an approach where it is already pre-defined what the analyst wants to measure (i.e., specific adulterant, a particular DNA section…), while untargeted analyses have no pre-defined list of parameters, there is only the pattern of results, multiple data points collected from the samples (pattern of proteins, metabolites, or genes). Targeted approaches are mainly adopted for quantification analyses, as they focus on specific compounds, and this permits the analyst to obtain a reliable quantification. On the other hand, untargeted approaches aim at analysing all the detectable compounds, also with the scope of identifying unknown molecules. This aspect cannot guarantee the same quantification reliability provided by the targeted methods, but it allows the researches to unravel new chemical structures (Cajka & Fiehn, 2016). Behind all the untargeted analyses, and also the targeted analyses where more pre-defined parameters are measured, there is the MVA, which allows us to get an indication of the result from the overall pattern. Regarding laboratory testing, many industries used to send samples to be analysed by an external laboratory. Nowadays, this is changing, because it is a clear advantage for them to conduct the analyses at a point of use, gaining results in real time.

The techniques mainly employed for food authenticity testing include:

- Mass Spectrometry (MS)
- Stable Isotope Ratio Mass Spectrometry (IRMS)
- DNA analysis (real-time polymerase chain reaction (PCR), barcoding, Random Amplification of Polymorphic DNA (RAPD) …)
- Nuclear Magnetic Resonance (NMR)
- Spectroscopy (Infra-red (IR), Near Infra-red (NIR), or UV-Visible)

(IFST – Food authenticity testing part 1: The role of analysis; IFST – Food authenticity testing part 2: Analytical techniques). MS and NMR techniques can be exploited for untargeted and targeted approaches since they both permit to identify unknown compounds as well as to specifically quantify them.

The present Ph.D. work was structured by deepening all the techniques exploited for the geographical origin assessment of two main food matrices, hazelnut, considering its chain (fresh, roasted, paste), and dehydrated apple. Each of them was analysed through one or more rapid approaches, for screening/fingerprinting, and then one confirmatory was applied. In this way, it was possible to develop fast strategies, useful for the industrial environment, where they can be transferred on the production chain line, to directly detect frauded unit, lots. Contextually, the robust technique allows us to confirm the fast method's outcomes, as well as permits the author to gain more chemical information, eventually with molecule/markers identification.

The impact of these technologies on the industries was crucial, as this research was developed in strict collaboration with an international food company (Barilla G. & R. Fratelli S.p.A.).

REFERENCES

W. S. Alrobaish, L. Jacxsens, P.A. Luning & P. Vlerick, «Food Integrity Climate in Food Businesses: Conceptualization, Development, and Validation of a Self-Assessment Tool,» *Foods*, vol. 10, pp. 1302, 2021.

T. Cajka & O. Fiehn, «Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics,» *Analytical Chemistry*, vol. 88, pp. 524-545, 2016.

G. P. Danezis & C. A. Georgiou, «Elemental metabolomics: food elemental assessment could reveal geographical origin,» *Current Option in Food Science*, vol. 44, p. 100812, 2022.

C. Elliott, «Elliott Review into the Integrity and Assurance of Food Supply Networks. Final Report: A National Food Crime Prevention Framework,» Crown Copyright, 2014, Available online: https://bit.ly/2NA2CGP (accessed on 14 June 2020).

Eurofins, [Online]. Available: https://www.eurofins.de/food-analysis/analytical-testing/food-authenticity/.

European Commission, [Online]. Available: https://food.ec.europa.eu/safety/agri-food-fraud_en.

No 1169/2011 of the European Parliament and of the council of 25 October 2011 on the provision of food information to consumers, 2011.

E. Haynes, E. Jimenez, M. A. Pardo & S. J. Helyar, «The future of NGS (Next Generation Sequencing) analysis in testing food authenticity,» *Food Control*, vol. 101, pp. 134-143, 2019.

A. J. Stein & F. Santini, «The sustainability of "local" food: a review for policy-makers,» *Review of Agricultural, Food and Environmental Studies*, pp. 77-89, 2022.

Introduction

Issues and alerts related to the authenticity of food commodities are nowadays very common. False origin declarations and other mislabelling illegal activities could affect the food quality, as well as the consumers' safety. In recent years, many techniques for authenticity assessment were considered, but still, there are no harmonised methods for several food matrices. The increasing demand for fighting/preventing this type of fraud has led to the introduction of innovative technologies and/or the application of renowned ones in this field. Both rapid and confirmatory strategies were exploited, mostly at the academic level, whereas food companies are not well-aligned, as the small-medium enterprises cannot invest in facilities and expertise, and the big ones started to work on this aspect, but the gap with the university research groups is still quite wide.

This introductive chapter is divided into three parts: the first one is about one of the most employed approaches for fine fraud detection, chromatography hyphenated with Mass Spectrometry (MS). It reports several types of technologies, from liquid to gas chromatography, from high-res MS to rapid spectrometric strategies (Ambient MS, Isotope Ratio MS), applied on various matrices. For each of them, it compares their applications in both academic and industrial environments, highlighting the lack of communication and harmonisation between these two realities.

The second part focuses on other analytical technologies, successfully applied in the food authenticity area, and a brief overview of the chemometric tools used to extract useful chemical information for statistical analysis. Moreover, the basis and principle of the sensomics approach were discussed. The applicability of these approaches will be underlined, considering diverse commodities. The third part will be about the matrices of interest, on whom the abovementioned analytical strategies have been applied, pointing out the need for these specific methods to correctly assess their authenticity.

The following review was published in "*TrAC - Trends in Analytical Chemistry"*. For additional details see the section "Author".

Fighting Food Frauds exploiting Chromatography-Mass Spectrometry Technologies: Scenario Comparison between Solutions in Scientific Literature and Real Approaches in place in Industrial Facilities

Michele Suman[a], Daniele Cavanna[a,b] , Giuseppe Sammarco[a,b], Francesca Lambertini[a], Cecilia Loffi[a,b]

[a]Barilla G. R. F.lli SpA, Advanced Laboratory Research, via Mantova 166, 43122, Parma, Italy
[b]Department of Food and Drug, University of Parma, Parco Area delle Scienze 95/A, 43124, Parma, Italy

INTRODUCTION

Recently, consumers have increased attention to the safety, authenticity and quality of food commodities (Borras, et al., 2015), and the development of reliable methods to detect fraud could better preserve the quality of production, with an advantage for the industry and the final consumer, who would have more confidence in the quality and safety of the product. Current governance of food supply chains, through certifications, inspections and audit controls, has historically been indeed developed, with a focus on food safety. However, criminal actions not only involve the adulteration of raw materials or finished products, but also the fraudulent declaration of the geographical origin of commodities (Bontempo, Camin, Paolini, Micheloni, & Laursen, 2016).

Understanding the root causes of food integrity issues, behaviours that drive certain decisions and activities, is crucial and the data that could help in the identification and prevention of fraud is therefore very important. There is a close link between criminal activities, geographic regions and specific agrifood resources and businesses: in this sense, criminals become real entrepreneurs, looking mainly to increase profits on the black market. Local complicity by a "corrupted political environment" that acts in order to facilitate illegal actions also cannot be excluded (Ponzi, 2017). In addition, speculation and unjustifiable low prices and promotional initiatives indirectly favour the work of illegal fraudulent networks (Elliott, 2019).

State-of-the-art instrumentation employed in —omics approaches represents a powerful tool to face the challenge of detecting potential adulterations, contaminations (Gilbert-Lopez, et al., 2017) and other several frauds. Especially chromatographic applications, hyphenated with MS, may stimulate interactions between researchers and industrial experts, and this could be the challenge of providing affordable methodologies for the identification of food fraud in some specific raw materials and commodities.

However, sometimes it seems that the new analytical methods developed and presented in peer-reviewed journals are not applied or even not known by the industrial quality control laboratories.

In specific circumstances, their lack of application can be justified by limited applicability only to specific fields: for instance, the use of molecular markers limits (e.g. geographical origin, organic origin), requires extensive up-to-date databases (Cubero-Leon, De Rudder, & Maquet, 2018).

The present study stemmed from the activities included in the framework of the FoodIntegrity European Funded EU-FP7 Project (Ensuring the Integrity of the European food chain). Specifically, FoodIntegrity was a 5-year/12 million Euro project, recently completed (2014-2018), with associated partners including regulators, consumers, academia and the food industry. The main aims and key activities of the project were: (i) to provide Europe with a state-of-the-art and integrated capability for detecting fraud and assuring the integrity of the food chain, (ii) to provide a sustainable body of expertise, (iii) to share data and knowledge, (iv) to develop methods and systems for industry, (v) to develop new methods of analysis, (vi) to develop early warning systems, (vii) to understand consumer behaviour for export.

The FoodIntegrity consortium devoted specific attention to selecting, identifying and evaluating the industrial impact of relevant food chains that can be affected by food fraud issues. Some of these matrices are widely studied and reviews about frauds on them are already available, such as in the case of milk (Nascimento, Santos, Rodrigues Pereira-Filho, & Rocha, 2016), and spices (Galvin-King, Haughey, & Elliott, 2017).

In the present paper, four of these food chains of interest (wheat/cereals, fruits, nuts and nut products, eggs & egg products), strategically relevant for food industries worldwide, were considered. Despite the presence of effective spectroscopic methods for fraud detection, such as Nuclear Magnetic Resonance (NMR), Near Infrared (NIR), and molecular biology strategies as well, like Polymerase Chain Reaction (PCR), both a large amount of information about the application of chromatographic-MS techniques on the abovementioned food chains and their capabilities to provide for high detection sensitivity and molecular specificity outcomes have determined the choice to focus on this approach on these matrices. On the other side, high costs, relatively long analysis time and the need for skilled scientists for data acquisition and elaboration must be considered as cons of these techniques (Stachniuk & Fornal, 2016). Relevant information about the existence in the scientific literature of the application of main chromatographic methods (such as Gas Chromatography (GC) and Liquid Chromatography (LC)) interfaced with several (MS) techniques was then collected. The information was combined with feedback directly collected from many industry representatives (through informal interviews and official open days & workshops executed during the FoodIntegrity EU project), about their internal approaches to prevent food fraud. Therefore, what could be done and what actually is done represents the final goal of this work: a comparison between the chromatographic-MS approaches suggested in scientific papers to detect food frauds and the current scenario within food

companies is presented through an analysis of these different types of raw materials/commodities. This comparison could be a valuable indication for companies in approaching several modern frauds. Simultaneously, it could be an interesting point for academies to better understand how to create connections and collaborations with industries. In this way, exploiting industrial facilities and academical competences will be a relevant perspective in the food fraud fight.

APPROACH TO DATA COLLECTION

In order to highlight new analytical strategies to detect food fraud, research on scientific literature was carried out with a focus on potential applicability and technology transfer, taking into account the most relevant papers published from 2010 to 2020. The focus is theoretically on works with a linear and robust workflow, which usually starts with training dataset building, where several variables (season and yearly variabilities, harvesting and climatic conditions, …) are considered for proper sampling. Afterward, a validation dataset (with fewer, but significant samples) is built, to confirm the previous data; lastly, the same method is applied to real (or independent) samples, to assess application robustness. Eventually, methods developed could also be transferred to other labs to evaluate inter-lab reproducibility. However, other studies, that do not completely follow this procedure, have been considered, because of their relevant and potential application in the food fight, highlighting anyway limitations and necessity of further insights.

Mainly ScienceDirect-Elsevier (www.sciencedirect.com) & CAS-SciFinder (www.scifinder.cas.org) databases were searched: a preliminary literature screening was executed combining the keywords “food fraud”, “food adulteration”, “food traceability” for each commodity. The results from extracted literature were then further refined using the keyword “LC-MS”, coming to a final number of approximately 54 records. In parallel, another refinement was made using the keyword “GC-MS”. In this case, the final number of extracted records reached 97.

Finally, a subsequent selection of the most relevant papers was evaluated within the Food Integrity Consortium, following the suggestions of different experts, mainly coming from the industrial integration work-package.

The critical literature revision was combined with a rational discussion (single interviews were conducted, face to face or by dedicated conference calls) with 22 industrial experts from the different selected food chains: senior scientists and managers that cover correlated roles within Quality Assurance, Vendor Assurance, Purchasing, R&D departments. In addition, information collected from literature and interviews was compared with outcomes derived from official panel discussions at two main events in 2015: Food Integrity EXPO UK Science Innovation Network Workshop (Milan, Lodi, September 2015) & Food Integrity Recent Advances in Food Analysis Open Days 2015

(Prague, November 2015) [more details, information and documentation about these events can be found on the FoodIntegrity website www.foodintegrity.eu)].

The involved industrial experts were encouraged to consider vulnerabilities to specific supply chains, to protect their businesses by using information and intelligence, also providing information about their risk assessments and assurance actions.

COMMODITY "JOURNEY"

Each food chain is analysed separately as follows. There is an introduction with the list of the most diffused frauds for each commodity, with selected examples of LC-MS and GC-MS approaches presented in the literature to manage these problems and, lastly, the industrial approach to the same topics. A comprehensive overview is reported in the summary tables.

Wheat and other cereals (Table 1)

Most diffused fraud

One of the most common perpetuated frauds regarding cereals is the addition of common wheat to durum wheat. In Italy, dried pasta must be made exclusively of durum wheat, and a maximum contamination of 3% from common wheat flour in durum wheat flour is allowed (Presidential Decree no. 187, 9 February 2001, Suppl. Mat.). This restriction is due to the rheological properties of durum wheat flour that are ideally suited to the pasta manufacturing process (Dexter & Matsuo, 1980). However, the common wheat cost, is 25 % (an average) lower than that of durum wheat (Institute of Food Services for the Agricultural Market (ISMEA). www.ismea.it), favoring fraudulent actions. Another type of fraud involves the adulteration of bread, with refined wheat flours, or other flour not indicated on the label.

False declarations of geographical and botanical origin are popular as well in the cereals market, since, nowadays, these are required claims to certify food quality.

About adulterations, in this field, it is quite diffused the addition of exogenous nitrogen-rich adulterants, such as melamine, ameline, ammelide, etc. This fraud occurs to raise the protein content for economic purposes.

False "gluten-free" declaration represents a diffused fraud, since a mislabeling not indicating the presence of traces of wheat flour in other flours may be dangerous for coeliac people. This, indeed, represents not only a classic illegal action, but it is a serious concern for public health; both intentional and unintentional contamination can occur leading to an authenticity issue for such products.

Finally, today's consumers are increasingly interested in organically grown products. For a cereal-derived product to be labelled as organic, the producer must follow and comply with specific rules

laid down in international regulations. This will inevitably lead to higher costs for producing organic products compared to conventional ones: an authenticity issue will therefore arise when cheaper non-organic product is sold as organic.

Approaches presented in the literature

Concerning the adulteration of durum wheat with common wheat, LC-High Resolution (HR)MS and LC-MS/MS are the most diffused approaches to detect this fraud. In particular, different C17:0 to C21:0 alkylresorcinol homologue ratios in the two cereals may be an ideal marker to determine the presence of common wheat flour in that of the durum (Knödler, Most, Schieber, & Carle, 2010). A similar approach involved the evaluation of two peptides ratio, one of them present in all wheat samples, used as a marker of total wheat content, and the other one present only in samples without durum wheat. Thus, the ratio of their chromatographic areas reflects the proportion of common wheat in the sample, using a calibration curve made with standards of known composition (Prandi, et al., 2012).

An untargeted lipidomic approach is evaluated as well to discriminate between common and durum wheat; furthermore, digalactosyl diglyceride (DGDG) 36:4 is considered an effective marker to distinguish between authentic durum wheat and its adulterated admixture (Righetti, et al., 2018). Other studies are focused on discriminant compounds such as heterocyclic amines, polyketides, phospholipids (Matthews, et al., 2012) and a peptide of Puroindoline-a (Pin-a) (Russo, et al., 2014). Therefore, both untargeted and targeted approaches are largely adopted for durum wheat adulteration. They could provide for a wide and heterogeneous strategy (studying peptides, lipids, small molecules…) in fraud identification.

Bread adulteration is mainly detected with LC-HRMS and LC-MS/MS as well: for instance, ultrahigh-performance LC with high-resolution accurate multistage MS (UHPLC-HRAM-MS^n) with atmospheric pressure chemical ionisation (APCI) and chemometric analysis (PCA and SIMCA) is a valuable strategy to establish differences between whole wheat flour and refined wheat flour (Geng, Harnly, & Chen, 2016). Multivariate statistical analysis is indeed a valuable tool to elaborate and interpret data from LC-MS.

Another interesting work focuses on the identification and quantification of rye, spelt and wheat peptide markers in bread, through LC-MS/MS, in order to avoid mislabelling issues (Bonick, Huschek, & Rawel, 2017).

The assessment of durum wheat geographical and botanical origins employs both chromatographic techniques, GC and LC, coupled with MS. Headspace-Solid Phase Microextraction (HS-SPME) GC-MS approach, for example, was considered to detect 32 main volatile compounds to discriminate

between 10 different durum wheat cultivars of three Chinese regions (Wadood, Boli, Xiaowen, Raza, & Yimin, 2019). Furthermore, a non-targeted HRMS approach allows providing for a selection of markers related to the geographical origin of Italian/European/Not European durum wheat: samples related to the 2016 harvest were used to set up the model and select the markers, samples collected in 2018 were then used for model and markers validation (Cavanna, Loffi, Dall'Asta, & Suman, 2020). Moving from an untargeted to a targeted approach is getting relevance since it allows one to set up a model and to confirm then it, validating simultaneously the markers previously selected.

About botanical origin, a non-target metabolomic strategy is presented in the literature as a possible way to distinguish between three ancient Triticum varieties (emmer, einkorn and spelt): a UHPLC-QTOF analysis shows how alkylresorcinols are important markers to identify Triticum varieties and as cultivar markers as well since they are not affected by environmental changes (Righetti, et al., 2016). In this way, the variability is reduced, hence the selection of the marker is less biased by other factors.

Moving to the GC approaches, a study on the botanical origin of Serbian cereals and oilseeds is conducted, employing a GC-MS instrument: several fatty acids have proven to be valuable markers for developing authentication methods in that field (Pastor, Ilić, Vujić, Jovanović, & Ačanski, 2019). Another GC-MS approach is employed to authenticate small grain and corn species, comparing their sugar and lipid profiles (Pastor, et al., 2016) (Pastor, Acanski, Vujic´, Jovanovic´, & Wienkoop, 2016). It could be interesting to perform an inter-lab method transfer, to assess the botanical origin of cereals and oilseeds also from other geographies.

Discrimination of seven species (Barley, Wheat, Einkorn, Emmer, Oats, Rye, Spelt) is carried out with GC-HRMS strategy as well, highlighting differences in the patterns of alkylresorcinols, free sterols, glycosylated sterols and sterol esters, tocopherols, traces of tocomonoenols, diunsaturated and methyl-alkylresorcinols between the cereals (Hammann, Korf, Bull, Hayen, & Cramp, 2019).

Concerning the adulterants, nitrogen-rich compounds are added to increase the protein content for economically motivated adulterations: a paper describes an LC-MS/MS method able to detect 14 nitrogen-rich molecules and to quantify them, using an isotopic dilution, reaching the Limit of Quantitation (LOQ) of 0.05-0.20 mg/kg in cereals and other matrices (Frank, Bessaire, Tasser, Goyion, & Delatour, 2017). As previously mentioned, these techniques allow the authors to reach even traces of fraudulent additives and this explains their importance in fraud detection.

Other works focus on specific nitrogen-rich markers of adulteration, such as melamine and its related compounds, ammeline, ammelide and cyanuric acid, analyzed through LC-APCI-MS, LC-Diode Array Detection (DAD) and isotope dilution GC-MS (ID-GC-MS) in cereals and other various matrices (Vinas, Campillo, Ferez-Melgarejo, & Hernandez-Cordoba, 2012) (Wong & Mok, 2013).

The last fraud considered on cereals regards the adulteration of white rice with its corresponding blended product. A study about Korean white rice is carried out by combining targeted lipidomics with powerful supervised learning methods, and, as result, 17 lysoglycerophospholipids are found to be key compounds to discern between Korean white rice and mixtures with Chinese white rice (Lim, et al., 2017).

Preliminary metabolomic analyses seem promising as reliable screening techniques to distinguish between organic and conventional crops; however, to better assess this distinction, high sensitivity and specificity of LC-HRMS measurements are required: interesting examples on cabbage, strawberries and tomatoes are already present in the scientific literature (Mie, et al., 2014) (D'Urso, d'Aquino, Pizza, & Montoro, 2015) (Bueno, Diaz-Galliano, Rajski, Cutillas, & Fernandez-Alba, 2018).

Industrial approach and needs

To semi-quantitatively determine common wheat (Triticum aestivum) content in durum wheat semolina (Triticum durum), the food/cereals industry environment has been internally exploiting since a long time immunoassay strategies, such as the so-called Durotest-S Kit, which apply the use of a monoclonal antibody, specific for the protein friabilin, (present in common wheat but not in durum wheat) (R-Biopharm AG, 2020).

Then, the grains industry over the last decade has mainly explored molecular diagnostic techniques potentialities (progressively substituting previously applied electrophoretic approaches (Bonetti, et al., 2004) for the detection of common wheat adulterations in durum wheat. Both endpoint and real-time approaches are indicated by the interviewed experts.

The substitution of higher value species with lower value flours is monitored internally again with immunoassay tests, while LC/GC-MS strategies can find their application mainly in pesticide residue analysis to provide supplementary evidence in suspected cases of organic misdescription. The potential use of melamine is further evaluated with HPLC-MS methods.

As regards the specific case of wheat varieties, the food industry usually follows the International Union for the Protection of New Varieties of Plants (UPOV) guidelines for the “examination of distinctness, uniformity, stability and the development of harmonized descriptions of new varieties of plants” (UPOV, 1994).

There are relevant potentialities that instrumental analysis can offer in the organic scenario if they are accepted as being part of the overall certification process attested for the food industry, which is at the moment only sporadically exploited in outsourced external labs: Isotopic Ratio Mass Spectrometry (IRMS) and or pesticide residue analysis. N & O isotopes analysis can provide evidence

to corroborate whether chemical fertiliser has been applied to a crop, multi-element analysis improves classification rates (Laursen, et al., 2013).

Fruits (Table 2)

Most diffused fraud

Fruit is a wide and heterogenous matrix that encompasses various types of products. Thus, it is easy to think about several frauds in this field.

A valuable number of works is focused on defining aromas or flavours typical of different fruits, others are aimed to gain information about metabolic profiles and to evaluate different stages of ripening. This sometimes helps in protecting Designation Origin and avoiding fraud.

Fraud related to origin is perpetrated as well, mainly as regard geography. Geographical claim, like in many other foods, is a fundamental point, economically speaking: thus, it is very common to unravel frauds origin related. Different strategies are employed to univocally define not only the geographical origin but also the botanical one since the variety is often linked to the territory of cultivation.

Approaches presented in the literature

The authenticity of fruits is a relevant requirement, in order to avoid fraud related to false protected denominations (PDO or PGI) declarations or to mislabelling of fruit products in general. Various chromatographic approaches, coupled with mass spectrometry, are applied. A study assesses the authenticity of apple aromas through GC-C-IRMS analysis of key volatile compounds sampled using HS-SPME. A database of $\delta^{13}C$ values of 16 compounds is established, hence the identification of fraudulent aromatic substances is now reachable (Strojnik, et al., 2018).

Metabolomics is a considerable field of study to evaluate the authenticity since the heterogeneity of fruits provides them with different metabolic profiles: in literature it is possible to find many works concerning non-targeted, targeted, or both, approaches in authenticity assessment. Indian citrus fruits were studied to prevent fraud in the above-mentioned area; a metabolomic strategy was applied to unravel potential authenticity markers, involving an LC-quadrupole time of flight-MS (LC-QTOF-MS) analysis. 9 compounds of interest were found and, subsequently, a targeted study was conducted to optimize the analysis of these markers (Jandric, Islam, Singh, & Cannavan, 2015). Another paper on metabolomics aims to compare the non-volatile metabolite of five different Tanzanian tropical fruits, employing a broad-spectrum GC-MS new methodology. This technique reveals a different composition in carbohydrates, fatty acids and organic acids of these tropical fruits, becoming a potential application for large-scale studies not only on authenticity/variety but also on fruit quality

(Khakimov, et al., 2016). An interesting work presented in the literature is focused on the aroma of red grapes, aiming to characterize aroma metabolites and their precursors in ten selected genotypes, including six *Vitis vinifera* cultivars, two American species and two interspecific crosses. Two orthogonal techniques were considered, before and after enzymatic hydrolysis: GC-MS was used to detect 66 aroma-free volatile compounds, while LC-MS was used to identify 15 volatile compounds precursors (Ghaste, et al., 2015). This orthogonal analysis could allow the authors to build a sort of metabolic pathway, identify both volatile compounds and their precursors. It could be applied to a wider set of samples, from several geographical regions, to assess its reproducibility and robustness.
Amongst various metabolites, polyphenols are very common in different species and their profile could have a key role in authenticity evaluation. An HS-SPME-GC-MS approach was conducted also to protect a European Designation of Origin of 'Melocoton de Canada', studying four different varieties of peach cultivars and comparing their volatile profiles (Montero-Prado, Bentayeb, & Nerin, 2012). A similar approach, using GC-MS techniques, was employed for the detection of lactones at three progressive post-harvest stages of fully ripe air-freighted and green ripe sea-freighted pineapples (Steingass, Langen, Carle, & Schmarr, 2014).
Geographical and botanical origins are fundamental claims in food quality, so economically motivated frauds are quite common in the fruit market. The strategies here applied for geographical and botanical origin investigations are basically like the ones used to assess authenticity, hence metabolomics, determination of flavor and aroma molecules, or analysis of specific fractions. The geographical origin of apples in Northeast Italy was defined by the characterization of their flavor composition, through HS-SPME-GC-MS analysis (Giannetti, Mariani, Mannino, & Marini, 2017). The same approach was employed to discern between several varieties of Madeira Island passion fruits (Porto-Figueira, Freitas, Cruz, Figueira, & Camara, 2015). GC-MS technique was also exploited in studying the volatile profile of five different banana varieties and 68 volatile organic metabolites identified have a key role in this work (Pontes, Pererira, & Camara, 2012). Furthermore, volatile molecules were identified and semi-quantified to discern four sweet cherry varieties, according to both botanical and geographical origin, through SPME-GC-MS analysis (Papapetros, et al., 2018).
Two metabolomics works concerning the geographical origin of goji berries focused on metabolic profiles of Asian goji berries (Bondia-Pons, et al., 2014), and specific carotenoid profiles to distinguish between Italian and Asian ones (Bertoldi, et al., 2018) both of them using an LC-MS technique. It is interesting to notice that LC-MS is used both in 2014 and 2018 studies, whereas, in the 2018's article, IRMS is also exploited, as a recent tool for origin assessment. A similar approach, employing QTOF-MS was also put in action to differentiate Spanish oranges from Argentinean,

Brazilian and South African ones (Diaz, Pozo, Sancho, & Hernandez, 2014). Non-targeted and targeted metabolomics strategies were applied to wild strawberries as well, to assess the South Italian geographical origin: the non-targeted approach is useful to select markers of origin, while the targeted one is to quantitatively analyze them (D'Urso, et al., 2016).

Industrial approach and needs

The industry is increasing the use of rapid NIR analysis for fingerprinting and calibration as official methods for parameters such as acidity, sugar content and alcoholic strength, which could represent a first sign of manipulation: those parameters are in any case usually measured, mainly exploiting IRMS (Camin, et al., 2017) or NMR techniques not directly in fruit but in fruit-juices or other fruit-derived products (where dilution with water, the addition of exogenous sugars, etc. are quite common fraudulent procedures…).

PCR, NMR and isotopic techniques are also applied in industry for adulterations detection but outsourced to external laboratories. Chromatographic-MS strategies are not so exploited for food detection on these matrices. However, the progressive request for high-quality food (i.e. with a certification of the geographical origin, bio food, organic crops) will likely claim for techniques more sensitive and with higher specificity for markers identification.

Studies about aromas and volatile profiles are another relevant topic regarding food quality and safety; the so-called “Sensomics” could be a valuable interface between analytical and sensorial data. Thus, a state-of-art approach to screen and/or specifically identify volatile compounds, such as GC-MS, can perfectly fit the possibility of merging its results with olfactive and gustatory analysis from panels.

Nuts (Table 3)

Most diffused fraud

The most popular fraud regards basically the origin since the country of origin is a mandatory indication on the label. The indications on the label can also include the crop year (older nuts are more prone to infestation and rancidity) and the variety.

Furthermore, defining precisely the species is another fundamental indication, particularly for problems related to allergies. In fact, the substitution of nuts from different species can affect allergy patients in case of lack of declaration.

Concerning authenticity, PDO and PGI for cultivars of some edible nuts are particularly appreciated and can be subjected to fraud, whether in the “as is” form or as ingredients in products as well (Maestri, Imperiale, & Marmiroli, 2018). One example is the hazelnut "Tonda Gentile delle Langhe"

(Nocciola Piemonte, PGI since 1996) from Italy; Other relevant PDO examples in Europe concern: chestnuts (Portugal and Italy), almonds (Portugal), walnuts (France, Italy).
A popular ingredient in confectionery is hazelnut paste, where the percentage of hazelnuts is a critical quality issue and dilution or substitution with artificial compounds or with other ingredients are quite common fraudulent practices.

Approaches presented in the literature

Methods based on chromatography-MS solutions devoted to the analysis of fatty acids or other metabolites have been developed to check for the geographic origin and authenticity of the species or the cultivar, even in correspondently processed foods (Cizkova, Rajchl, Snebergrova, & Voldrich, 2013) (Barreira, et al., 2009).
Characterization of differences in lipids, proteins, carbohydrates through non-targeted analyses are being developed, for instance, exploiting those correspondent chemical profiles to discriminate geographic origin. (Esteki, et al., 2017)
Characterization of volatile profiles carried out with bidimensional GC × GC - MS is another approach presented in the literature for the geographical origin assessment as in the case of a study on the volatile fraction of Italian and Turkish roasted hazelnuts (Cordero, et al., 2010). Also, in this case, it might be interesting to apply the abovementioned bidimensional strategy to other geographical locations, to assess its reproducibility and robustness.
Geographical origin is really a popular claim nowadays, hence a valuable number of works is conducted to assess the origin of nuts and nut products: the most preferred approach is metabolomics which provides for the characterization of metabolic profiles of matrices from different origins, in order to favour the discrimination among several countries. Thus, non-targeted and/or targeted LC-MS metabolomic strategies were applied for various types of nuts, to classify geographically and botanically almonds (Gil Solsona, Boix, Ibanez, & Sancho, 2017) and to define the provenience of hazelnuts from some European countries (Klockmann, Reiner, Bachmann, Hackl, & Fischer, 2016) (Klockmann, Reiner, Cain, & Fischer, 2017), focusing particularly also on phenolic and sterolic compounds (Ghisoni, et al., 2020). This approach was also used for origin discrimination of pistachio nuts, by analyzing some key molecules, such as phenols, anthocyanin, catechin, fatty acids, the latter analyzed also through GC-MS (Mannino, Gentile, & Maffei, 2019).
As regards species origin, a peculiar adulteration is represented by pistachio nut powder, mixed not openly declared with green pea: this fraud was characterized via HPLC-QTOF-MS (Cavus, Us, & Guzelsoy, 2018).

The sale of nuts with no controlled levels of contaminants is another fraudulent action that has required the development of several techniques able to detect even minimum amounts of these compounds. The most present contaminants are mycotoxins, secondary metabolites produced by fungi like *Aspergillus* (aflatoxins) and *Penicillium* (Sataque Ono, Hirooka, Rossi, & Ono, 2011). MS, hyphenated with LC, is a state-of-art tool able to face this fraud by detecting mycotoxins with low Limit of Detection (LOD) and LOQ, also thanks to the technological progress of this instrumentation achieved in the last years. In particular, LC-MS/MS strategies were applied to detect and profile aflatoxins and other mycotoxins in cashew nuts from different African areas (Lamboni, et al., 2016) (Adetunij, et al., 2019). MS/MS was also used to detect *Aspergillus* and *Penicillium* toxins in fresh and dried chestnuts from Northern Italy (Bertuzzi, Rastelli, & Pietri, 2014). A study on hazelnut samples aimed to investigate mycotoxins by extracting them with two different extraction mixtures and then analyzing them through LC-MS/MS. This work divided hazelnuts for geographical origin, agricultural practices adopted and commercial typology (Prelle, Spadaro, Garibaldi, & Gullino, 2012). GC-MS tool was exploited to detect other dangerous compounds. Particularly, Taghizadeh et al. used the abovementioned technique for cumulative risk assessment of pesticide residues in different Iranian pistachio cultivars (Taghizadeh, et al., 2019).

Other dangerous and threatening molecules are pesticides and allergens, both not and wrong declared on the label, as a fraudulent act. Peanuts are the most dangerous nuts concerning allergy; a proteomic study focused on the detection of their protein allergens, Ara h 1, Ara h 2 and Ara h 3, and Advanced Glycation Endoproducts (AGE) in fresh and roasted peanuts. This proteomic approach consists of protein separation with Sodium Dodecyl Sulphate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) and identification by Western blotting and LC-MS/MS (Hebling, McFarland, Callahan, & Ross, 2012). Another allergens-related work was aimed to evaluate changes in solubility and detectability of allergens in roasted walnuts, prepared with different roasting times and temperatures, using an LC-MS/MS method (Downs, Baumert, Taylor, & Mills, 2016).

False authenticity declaration is also considered for nuts as a fraudulent act. Therefore, several chromatographic-MS methods were progressively developed and improved to assess this claim. A study on hazelnut samples was demonstrated to authenticate 29 different European cultivars, including also some PGI and PDO varieties: this result was achieved by phenotypic features study, phenolic profiling and MS-based proteomic approach (Ciarmiello, et al., 2014). Authenticity assessment was performed by GC-MS as well, through volatile molecules investigation and flavor profiles characterization of Turkish fresh pistachio nut varieties (Kendirci & Onogur, 2011).

Industrial approach and needs

The food manufacturing industry commonly buys nuts, as ingredients, to be processed and/or re-packaged. The chocolate industry is the largest user of nuts, together with the snack industry specifically for large quantities of peanuts as well. Considering this abundant use of edible nuts, they have been progressively subjected to revision and adjustments of the labels, regarding health attributes and claims. Increasingly there are requirements for sustainable products in appealing to consumers, particularly when edible nuts come from developing countries. Most industrial standards on nuts imply direct visual inspection and morphological evaluations: the U.S. Food and Drug Administration (FDA) has a list of methods for the analysis of nuts and nut products addressing mainly defects, infestation, and presence of foreign material. (FDA, 2017)

For the recognition of specific cultivars, checking for PGI or PDO varieties, visual inspection or imaging techniques are often applied in industry. (Rabadan, Pardo, Gomez, Alvarruiz, & Alvarez-Ortì, 2017) Visual and sensorial inspection are, according to some interviewed retailers, methods of election for fraud prevention. It is indeed quite easy for experts to visually detect the type of variety and the geographical origin of the raw product. Moreover, it is more difficult to make frauds on raw nuts, whereas it is common to find fraudulent products after industrial processing, like in some nuts paste.

Enzyme-Linked Immunosorbent Assay (ELISA) and Lateral Flow Devices (LFD) are in general used in industrial quality control facilities as rapid testing for the presence of specific nuts in food products because of potential allergenic risks; sometimes, due to potential degradation of allergenic proteins in processing conditions, there is the need to confirm that presence through the analysis of DNA markers in collaboration with external laboratories. (Ding, et al., 2020).

Concluding, despite the relevant technological progress in this food sector, the lack of knowledge about the potentialities and the parallel current use of cheap techniques (like the abovementioned visual and sensorial inspection) are still determining a significant gap: the industrial direct exploitation or outsourcing of chromatography-MS solutions is far from been stably applied yet. As mentioned for fruit matrices, the progressive demand for high-quality food could probably emphasize the application of GS/LC-MS strategies, taking into consideration that traces of contaminants and/or adulterants can be revealed mainly by more efficient chromatographic-MS approaches. Moreover, processed and semi-processed nuts products (roasted grains, paste, etc.) are often subjected to fraudulent adulteration (i.e. chickpea flour added to hazelnut paste); this illegal addiction might be detected by LC/GC-MS techniques, considering both untargeted and targeted approaches.

Eggs & Egg products (Table 4)

Most diffused fraud

Frauds can be potentially found in all different types of egg products: eggs, shelled eggs, egg powder. At present, one of the most critical issues is egg freshness, which makes a major contribution to the value of the product, because consumers may perceive variability in freshness as a lack of quality (Lin, Zhao, Sun, Chen, & Zhou, 2011).

Furthermore, another well-known type of fraud is the addition of incubated eggs to fresh eggs, together with the addition of dyes.

The emerging risks related to this commodity are the use of conventional products fraudulently declared as organic and the introduction of melamine (which determines the apparent increase of the protein content) in eggs.

Approaches presented in the literature

Egg freshness is evaluated mostly by employing an LC-MS-based metabolomic approach. In particular, a work present in literature had the scope to determine the differences in small molecule profiles of egg yolks; subsequently, it focuses on how choline is involved in egg aging (Johnson A. E., Sidwick, Pirgozliev, Edge, & Thompson, 2019). Another study, through a UHPLC-HRMS method with multivariate analysis identified and selected 31 compounds discerning between fresh and not fresh eggs (Cavanna, Catellani, Dall'Asta, & Suman, 2018).

The addition of incubated eggs to fresh ones is a fraud quite common. A dedicated study evaluated the efficiency of European legislative indices, such as β-hydroxybutyric acid and lactic acid, used to detect incubator-rejected eggs (IRE) to egg products; this work suggests a revision of the current limits of the legislated markers, proposing also the introduction of uracil as future additional legal parameter: this assessment was performed using UHPLC-HRMS analysis (Hidalgo, Galbiati, Cavanna, & Suman, 2019)

Other types of fresh egg adulteration are the addition of illegal dyes, which are easily detected using LC-MS techniques: in the work of Chen *et al*, for example, the determination of illegal Sudan dyes and their metabolites was reached by ultra-sound assisted extraction and LC-ESI-MS/MS analysis (Chen, et al., 2013).

Another fraud is the introduction of melamine that, as explained in previous paragraphs, illegally increases the apparent protein content: its detection in eggs and egg products may be achieved with chromatography coupled with MS as well. With this approach, also melamine analogues, like cyanuric acid, can be quantified. For example, a peculiar work demonstrates the ability of a cleanup-free LC-MS/MS method to quantify these adulterants in egg powder and soy protein (Mondal,

Desmarchelier, Konings, Acheson-Shalom, & Delatour, 2010). The detection of these molecules can be carried out also by GC-MS/MS (Miao, et al., 2010).

Concerning the authenticity of eggs and egg products, some frauds may occur about the misrepresentation of the egg farming system. A study presented in the literature focuses on this illegal act, highlighting potential differences between lipid profiles of eggs from cage and barn housing systems, and tentatively identifying a compound (1,2-dipalmitoyl-glycero-3-phosphocholine) that could have potential as a biomarker of the egg housing system. Analysis was performed through HPLC-QToF-MS metabolomic approach and multivariate analysis (Johnson A. E., Sidwick, Pirogzliev, Edge, & Thompson, 2019).

Industrial approach and needs

In this food chain, the application of paper trails and documental strategies to look for potential traceability incongruences is particularly relevant (e.g. mass balance, etc…).

The prohibited addition of water can be easily checked by industries through dry matter testing.

Even though several rapid methods are available for the assessment of eggs freshness, industries do not generally implement these techniques in their routine controls, instead preferring to evaluate this type of fraud using the more common lactic acid quantification, as required by EU legislation (Reg. CE 853/2004).

Also, the illegal addition of incubated eggs is usually detected through enzymatic assays with a specific molecular marker target (3-hydroxybutyric acid) (Reg. CE 853/2004).

In agreement with the information presented in the literature, the addition of dyes and false declaration of organic eggs are detected using chromatographic and MS techniques.

On the contrary, due to a lack of expertise and the relatively high cost of the equipment, the industry usually outsources the analyses of melamine to externally certified laboratories.

PRODUCERS FINAL REMARKS

The open discussion with food industrial experts highlighted some interesting key points that could be helpful for the global scientific community: the vulnerability of “highly fragmented” chains (with a high number of steps, players and geographical areas) is perhaps the most crucial issue for food fraud, followed in most cases by the absence of implementation of constantly applied rapid screening procedures. This second point is in part the consequence of a limited level of awareness of risks and threats in the food fraud scenario within the stakeholder groups, in particular in small-medium enterprises.

The sensitivity of the food industry for investing money into analytical monitoring plans with a set of rapid and confirmatory analyses, optimized for different chains, needs to be amplified, concretely transmitting the perception that this investment is a "sort of insurance" to prevent much more serious economic-image damages that could arise after a public domain food fraud case.

The virtuous path and counterattack that industries would like to put in place are based on different strengths. Firstly, today many food tests that are still at the "research stage" should be implemented into a "routine scenario" for more effective and wider prevention and control, adopting appropriate certification standards and accreditations in the meantime. A relevant starting point to gain knowledge, know-how and to reach the so called "routine scenario" could be a direct connection between the academic and industrial worlds, where industries provide insights, applied scientific knowledge, economical and human resource, while academic research groups have the leading role into the project, transmitting skills and experience. Therefore, spin-offs and active collaboration may fill the academical-industrial gap, without excessive costs for small-medium enterprises. Further, this can be also seen as a growth opportunity for students and PhDs to get oriented to a professional/industrial environment, having a sort of pre-induction through those connections.

Thinking specifically to a proper future industrial exploitation of non-targeted MS strategies, a well-accepted/established harmonized approach does not exist up to now (especially for the validation of the results), so their robustness is not always certified: inspiration could be taken from a recently published scientific opinion, which proposes a workflow for the validation of non-targeted approaches with a specific focus on the industrial perspective of the topic. (Cavanna, Righetti, Elliott, & Suman, 2018)

Other crucial points that industries highlighted are the need to better establish separate assessment & prevention models for safety and authenticity issues, and the desire to achieve a standardized database for food profiling, sharing data in the scientific community while still ensuring appropriate confidentiality channels.

Their suggestion on this last topic is to make efforts on reducing (through regulated sharing) the economic impact of "databases re-construction" year by year.

CONCLUSIONS

This paper presented a comparison of chromatography-MS analytical approaches suggested in the scientific literature for fraud detection in several food chains with the real actions implemented by producers in their production plants.

Besides the clear fact that there is still a significant lack of awareness, especially in small-medium enterprises, food producers declared to know a relevant number of frauds studied in the literature, but

sometimes they do not apply concrete methods in their facilities to identify all the corresponding sophistication/adulterations. On the other hand, small-medium enterprises will never be able to afford expensive-complex instrumentations, like HRMS. In this situation, commercial laboratories, which on the contrary usually have this kind of accessibility and potentialities, could play a relevant role.

It seems that there are only a few situations where the academic and industrial production worlds are aligned, and in most cases, the methods applied by academic researchers are not implemented in the plants.

This relates to the fact that not all food producers stay updated on what scientific literature recommends; at the same time, some methods, even if known, are not designed for routine application to meet industry needs (due to skills/costs that are not always affordable): at present these are only efficient in a research laboratory. It is also important to mention that some discussed approaches, such as -omics ones, could have a potential leading role in food fraud fighting. Despite that, they still need to be optimised and it is necessary to make them suitable for practical industrial application, in terms of reliability, harmonization and know-how.

In the future, there is an outlook for these two worlds to continue to interact and "contaminate/cross fertilize" each other: therefore, food industries will have more sophisticated techniques available where MS plays a "king role", and researchers will be encouraged to develop reliable methods that could also be applied to routine analyses, and not only used from a research standpoint.

ACKNOWLEDGEMENTS

This work was supported by the European Union, Seventh Framework Program (FP7/2007-2013), under Grant Agreement 613688 (Ensuring the Integrity of the European food chain _ FoodIntegrity).

BIBLIOGRAPHIC REFERENCES

Adetunij, M. C., Aroyeun, S. O., Osho, M. B., Sulyok, M., Krska, R., & Mwanza, M. (2019). Fungal metabolites and mycotoxins profile of cashew nuts from selected locations in two African countries. *Food Additives & Contaminants: Part A*, 1847-1859.

Akkak, A., Boccacci, P., & Botta, R. (2008). AN EFFICIENT DNA-EXTRACTION PROTOCOL FOR NUT SEEDS. *Journal of Food Quality*.

Barbosa, S., Pardo-Mates, N., Hidalgo-Serrano, M., Saurina, J., Puignou, L., & Nunez, O. (2018). Detection and Quantitation of Frauds in the Authentication of Cranberry-Based Extracts by UHPLC-HRMS (Orbitrap) Polyphenolic Profiling and Multivariate Calibration Methods. *Journal of Agricultural and Food Chemistry*.

Barreira, J. C., Alves, R. C., Casal, S., Ferreira, I. C., Olivera, M. P., & Pereira, J. (2009). Vitamin E Profile as a Reliable Authenticity Discrimination Factor between Chestnut (Castanea sativa Mill.) Cultivars. *Journal of Agricultural and Food Chemistry*, 5524-5528.

Bertoldi, D., Cossignani, L., Blasi, F., Perini, M., Barbero, A., Pianezze, S., & Montesano, D. (2018). Characterisation and geographical traceability of Italian goji berries. *Food Chemistry*, 585-593.

Bertuzzi, T., Rastelli, S., & Pietri, A. (2014). Aspergillus and Penicillium toxins in chestnuts and derived products produced in Italy. *Food Control*, 876-880.

Bodia-Pons, I., Savolainen, O., Torronen, R., Martinez, J., Poutanen, K., & Hanhineva, K. (2014). Metabolic profiling of Goji berry extracts for discrimination of geographical origin by non-targeted liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. *Food Research International*, 132-138.

Bondia-Pons, I., Savolainen, O., Torronen, R., Martinez, J. A., Poutanen, K., & Hanhineva, K. (2014). Metabolic profiling of Goji berry extracts for discrimination of geographical origin by non-targeted liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. *Food Research International*, 132-138.

Bonetti, A., Marotti, I., Catizone, P., Dinelli, G., Maietti, A., Tedeschi, P., & Brandolini, V. (2004). Copared use of HPLC and FZCE for cluster analysis of Triticum spp and for the identification of T. durum adulteration. *Journal of Agricultural and Food Chemistry*, 4080-4089.

Bonick, J., Huschek, G., & Rawel, H. (2017). Determination of wheat, rye and spelt authenticity in bread by targeted peptide biomarkers. *Journal of Food Composition and Analysis, 58*, 82-91.

Bonick, J., Huschek, G., & Rawel, H. M. (2017). Determination of wheat, rye and spelt authenticity in bread by targeted peptide biomarkers. *Journal of Food Composition and Analysis*, 82-91.

Bontempo, L., Camin, F., Paolini, M., Micheloni, C., & Laursen, K. (2016). Multi-isotopic signatures of organic and conventional italian pasta along the production chain. *Journal of Mass Spectrometry , 51*, 675-683.

Borras, E., Ferrè, J., Boqué, R., Mestres, M., Acena, L., & Busto, O. (2015). Data fusion methodologies for food and beverages authentication and quality assessment- A review. *Analytica Chimica Acta , 891*, 1-14.

Bueno, M. J., Diaz-Galliano, F. J., Rajski, L., Cutillas, V., & Fernandez-Alba, A. R. (2018). A non-targeted metabolomic approach to identify food markers to support discrimination between organic and conventional tomato crops. *Journal of Chromatography A*, 66-76.

Cagliero, C., Bicchi, C., Cordero, C., Rubiolo, P., Sgorbini, B., & Liberto, E. (2011). Fast headspace-enantioselective GC–mass spectrometric-multivariate statistical method for routine authentication of flavoured fruit foods. *Food Chemistry*.

Camin, F., Boner, M., Bontempo, L., Fauhl-Hassek, C., Kelly, S. D., Riedl, J., & Rossmann, A. (2017). Stable isotope techniques for verifying the declared geographical origin of food in legal cases. *Trends in Food Science & Technology*, 176-187.

Cavanna, D., Catellani, D., Dall'Asta, C., & Suman, M. (2018). Egg product freshness evaluation: A metabolomic approach. *Journal of Mass Spectrometry*, 849-861.

Cavanna, D., Loffi, C., Dall'Asta, C., & Suman, M. (2020). A non-targeted high-resolution mass spectrometry approach for the assessment of the geographical origin of durum wheat. *Food Chemistry, 317*, 126366.

Cavanna, D., Righetti, L., Elliott, C., & Suman, M. (2018). The scientific challenges in moving from targeted to non-targeted mass spectrometric methods for food fraud analysis: A proposed validation workflow to bring about a harmonized approach. *Trends in Food Science and Technology*, 223-241.

Cavus, F., Us, M. F., & Guzelsoy, N. A. (2018). Assesing pistachio nut (Pistacia vera L.) adulteration with green pea (Pisum sativum L.) by untargeted liquid chromatography- (quadrupole-time of flight)-mass spectrometry method and chemometrics. *Journal of Food and Feed Science - Technology*, 34-41.

Cavus, F., Us, M. F., & Guzelsoy, N. A. (2018). Assesing Pistachio Nut (Pistacia vera L.) Adulteration with Green Pea (Pisum sativum L.) by Untargeted Liquid Chromtaography-(quadrupole-time of flight)-Mass Spectrometry Method and Chemometrics. *Journal of Food and Feed Science - Technology*.

Chen, D., Li, X., Tao, Y., Pan, Y., Wu, Q., Liu, Z., . . . Yuan, Z. (2013). *Journal of Chromatography B*, 45-50.

Chen, D., Li, X., Tao, Y., Pan, Y., Wu, Q., Liu, Z., . . . Yuan, Z. (2013). Development of a liquid chromatography-tandem mass spectrometry with ultrasound-assisted extraction method for the simultaneous determination of sudan dyes and their metabolites in the edible tissues and eggs of food-producing animals. *Journal of Chromatography B*, 45-50.

Ciarmiello, L. F., Mazzeo, M. F., Minasi, P., Peluso, A., De Luca, A., Piccirillo, P., . . . Carbone, V. (2014). Analysis of Different European Hazelnut (Corylus avellana L.) Cultivars: Authentication, Phenotypic Features, and Phenolic Profile. *Journal of Agricultural and Food Chemistry*, 6236-6246.

Cizkova, H., Rajchl, A., Snebergrova, J., & Voldrich, M. (2013). Filbertone as a marker for the assessment of hazelnut spread quality. *Czech Journal of Food Sciences*, 81-87.

Cordero, C., Liberto, E., Bicchi, C., Rubiolo, P., Schieberle, P., Reichenbach, S. E., & Tao, Q. (2010). Profiling food volatilesby comprehensive two-dimensional gas chromatography coupled with mass spectrometry: advanced fingerprinting approaches for comparative analysis of the volatile fraction of roasted hazelnuts (Corylus avellana L.) from different orig. *Journal of Chromatography A*, 5848-5858.

Cubero-Leon, E., De Rudder, O., & Maquet, A. (2018). Metabolomics for organic food authentication: Results from a long-term field study in carrots. *Food Chemistry*, 760-770.

Cuevas, F., Moreno-Rojas, J., & Ruiz-Moreno, M. (2017). Assessing a traceability technique in fresh oranges (Citrus sinensis L. Osbeck) with an HS-SPME-GC-MS method. Towards a volatile characterisation of organic oranges. *Food Chemistry*, 1930-1938.

Danezis, G. P., & Georgiou, C. A. (2022). Elemental metabolomics: food elemental assessment could reveal geographical origin. *Current Option in Food Science, 44*, 100812.

Dexter, J. E., & Matsuo, R. R. (1980). Relationship between Durum Wheat Protein Properties and Pasta Dough Rheology and Spaghetti Cooking Quality. *Journal of Agricultural and Food Chemistry*, 889-902.

Diaz, R., Pozo, O. J., Sancho, J. V., & Hernandez, F. (2014). Metabolomic approaches for orange origin discrimination by ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. *Food Chemistry*, 84-93.

Ding, Y., Jiang, G., Huang, L., Chen, C., Sun, J., & Zhu, C. (2020). DNA barcoding coupled with high-resolution melting analysis for nut species and walnut milk beverage authentication. *Journal of Science of Food and Agriculture*, 2372-2379.

Downs, M. L., Baumert, J. L., Taylor, S. L., & Mills, E. C. (2016). Mass Spectrometric Analysis of Allergens in Roasted Walnuts. *Journal of Proteomics*, 62-69.

D'Urso, G., d'Aquino, L., Pizza, C., & Montoro, P. (2015). Integrated Mass Spectrometric And Multivariate Data Analysis Approach For The Discrimination Of Organic And Conventional Strawberry (Fragaria x Ananassa Duch.) Crops. *Food Research International*, 264-272.

D'Urso, G., Maldini, M., Pintore, G., d'Aquino, L., Montoro, P., & Pizza, C. (2016). Characterisation of Fragaria vesca fruit from Italy following a metabolomics approach through integrated mass spectrometry techniques. *LWT - Food Science and Technology*, 387-395.

Elliott, C. (2019). *Elliott Review into the Integrity and Assurance of Food Supply Networks - Final Report.* HM Government.

Esteki, M., Farajmand, B., Amanifar, S., Berkhordari, R., Ahadiyan, Z., Dashtaki, E., . . . Vander Heyden, Y. (2017). Classification and authentication of Iranian walnuts according to their geographical origin based on gas chromatographic fatty acid fingerprint analysis using pattern recognition methods. *Chemometrics and Intelligent Laboratory Systems*, 251-258.

Eurofins. (s.d.). Tratto da https://www.eurofins.de/food-analysis/analytical-testing/food-authenticity/

European Commission. (s.d.). Tratto da https://food.ec.europa.eu/safety/agri-food-fraud_en

FDA, U. (2017, March 11). MPM: V-10. Nuts and Nut Products Methods. *MPM: V-10. Nuts and Nut Products Methods*.

Fiedler, K., McGrath, S., Callahan, J., & Ross, M. (2014). Characterization of Grain-Specific Peptide Markers for the Detection of Gluten by Mass Spectrometry. *Journal of Agricultural and Food Chemistry, 62*, 5835−5844.

Frank, N., Bessaire, T., Tarres, A., Goyon, A., & Delatour, T. (2017). Development of a Quantitative Multi-Compound Method for the Detection of 14 Nitrogen-Rich Adulterants by LC-MS/MS in Food Materials. *Food Additives & Contaminants: Part A*, 1842-1852.

Frank, N., Bessaire, T., Tasser, A., Goyion, A., & Delatour, T. (2017). Development of a quantitative multi-compound method for the detection of 14 nitrogen-rich adulterants by LC-MS/MS in food materials. *Food Additives & Contaminants: Part A, 34*(11), 1842-1852.

Galvin-King, P., Haughey, A., & Elliott, C. T. (2017). Herb and spice fraud; the drivers, challenges and detection. *Food Control*, 85-97.

Geng, P., Harnly, J., & Chen, P. (2016). Differentiation of Bread Made with Whole Grain and Refined Wheat (T. aestivum) Flour Using LC/MS-based chromatographic Fingerprinting and Chemometric Approaches. *Journal of Food Composition and Analysis*, 92-100.

Geng, P., Harnly, J., & Chen, P. (2016). Differentiation of bread made with whole grain and refined wheat (T. aestivum) flour using LC/MS-based chromatographic fingerprinting and chemometric approaches . *Journal of Food Composition and Analysis, 47*, 92-100.

Ghaste, M., Narduzzi, L., Carlin, S., Vrhovsek, U., Shulaev, V., & Mattivi, F. (2015). Chemical composition of volatile aroma metabolites and their glycosylated precursors that can uniquely differentiate individual grape cultivars. *Food Chemistry*, 309-319.

Ghisoni, S., Lucini, L., Rocchetti, G., Chiodelli, G., Farinelli, D., Tombesi, S., & Trevisan, M. (2020). Untargeted metabolomics with multivariate analysis to discriminate hazelnut (Corylus avellana L.) cultivars and their geographical origin. *Journal of the Science of Food and Agriculture*, 500-508.

Giannetti, V., Mariani, M. B., Mannino, P., & Marini, F. (2017). Volatile fraction analysis by HS-SPME/GC-MS and chemometric modeling for traceability of apples cultivated in the Northeast Italy. *Food Control*, 215-221.

Gil Solsona, R., Boix, C., Ibanez, M., & Sancho, J. (2017). The classification of almonds (Prunus dulcis) by country and variety using UHPLC-QTOF-based untargeted metabolomics. *Food Additives & Contaminants: Part A*, 395-403.

Gilbert-Lopez, B., Valdés, A., Acunha, T., Garcia-Canas, V., Simò, C., & Cifuentes, A. (2017). Foodomics: LC and LC-MS-based omics strategies in food science and nutrition. In S. Fanali, P. R. Haddad, C. Poole, & M. Riekkola, *Liquid Chromatography* (p. 838). Elsevier.

Hammann, S., Korf, A., Bull, I., Hayen, H., & Cramp, L. (2019). Lipid profiling and analytical discrimination of seven cereals using high temperature gas chromatography coupled to high resolution quadrupole time-of-flight mass spectrometry. *Food Chemistry, 282*, 27-35.

Haynes, E., Jimenez, E., Pardo, M. A., & Helyar, S. J. (2019). The future of NGS (Next Generation Sequencing) analysis in testing food authenticity. *Food Control, 101*, 134-143.

Hebling, C. M., McFarland, M. A., Callahan, J. H., & Ross, M. M. (2012). Global Proteomic Screening of Protein Allergens and Advanced Glycation Endproducts in Thermally Processed Peanuts. *Journal of Agricultural and Food Chemistry*, 5638-5648.

Hidalgo, A., Galbiati, D., Cavanna, D., & Suman, M. (2019). Evaluation of chemical indices for the identification of incubator-reject eggs in egg products. *Food Control*, 106767.

Jandric, Z., Islam, M., Singh, D. K., & Cannavan, A. (2015). Authentication of Indian citrus fruit/fruit juices by untargeted and targeted metabolomics. *Food Control*, 181-188.

Johnson, A. E., Sidwick, K. L., Pirgozliev, V. R., Edge, A., & Thompson, D. F. (2019). The Effect of Storage Temperature on the Metabolic Profiles Derived from Chicken Eggs. *Food Control*, 106930.

Johnson, A. E., Sidwick, K. L., Pirogzliev, V. R., Edge, A., & Thompson, D. F. (2019). The use of metabonomics to uncover differences between the small molecule profiles of eggs from cage and barn housing systems. *Food Control*, 165-170.

Johnson, A., Sidwick, K., Pirgozliev, V., Edge, A., & Thompson, D. (2018). The Metabonomic Profiling of Chicken Eggs During Storage Using High Performance Liquid Chromatography Quadrupole Time-of-Flight Mass Spectrometry. *Analytical Chemistry*, 7489-7494.

Kendirci, P., & Onogur, T. A. (2011). Investigation of Volatile Compounds and Characterization of Flavor Profiles of Fresh Pistachio Nuts (Pistacia vera L.). *International Journal of Food Properties*, 319-330.

Khakimov, B., Mongi, R. J., Sorensen, K. M., Ndabikunze, B. K., Chove, B. E., & Engelsen, S. B. (2016). A comprehensive and comparative GC-MS metabolomics study of non-volatiles in Tanzanian grown mango, pineapple, jackfruit, baobab and tamarind fruits. *Food Chemistry*, 691-699.

Klockmann, S., Reiner, E., Bachmann, R., Hackl, T., & Fischer, M. (2016). Food Fingerprinting: Metabolomic Approaches for Geographical Origin Discrimination of Hazelnuts (Corylus avellana) by UPLC-QTOF-MS. *Journal of Agricultural and Food Chemistry*, 9253-9262.

Klockmann, S., Reiner, E., Cain, N., & Fischer, M. (2017). Food Targeting: Geographical Origin Determination of Hazelnuts (Corylus avellana) by LC-QqQ-MS/MS-Based Targeted Metabolomic Application. *Journal of Agricultural and Food Chemistry*, 1456-1465.

Knödler, M., Most, M., Schieber, A., & Carle, R. (2010). A novel approach to authenticity control of whole grain durum wheat (Triticum durum Desf.) flour and pasta, based on analysis of alkylresorcinol composition. *Food Chemistry, 118*(1), 177-181.

Lamboni, Y., Frisvad, J. C., Hell, K., Linneman, A. L., Nout, R. M., Tamo, M., . . . Smid, E. J. (2016). Occurence of Aspergillus section Flavi and section Nigri and Aflatoxins in raw cashew kernels (Anacardiumoccidentale L.) from BeninL. *LWT - Food Science and Technology*, 71-77.

Laursen, K. H., Mihailova, A., Kelly, S. D., Epov, V. N., Berail, S., Schjoerring, J. K., . . . Husted, S. (2013). Is it really organic? – Multi-isotopic analysis as a tool to discriminate between organic and conventional plants. *Food Chemistry*, 2812-2820.

Lim, D., Long, N., Mo, C., Dong, Z., Cui, L., Kim, G., & Kwon, S. (2017). Combination of mass spectrometry-based targeted lipidomics and supervised machine learning algorithms in detecting adulterated admixtures of white rice. *Food Research International, 100*(1), 814-821.

Lin, H., Zhao, J., Sun, L., Chen, Q., & Zhou, F. (2011). Freshness measurement of eggs using near infrared (NIR) spectroscopy and multivariate data analysis. *Innovative Food Science & Emerging Technologies , 12*(2), 182-186.

Liu, R., Hei, W., He, P., & Li, Z. (2011). Simultaneous determination of fifteen illegal dyes in animal feeds and poultry products by ultra-high performance liquid chromatography tandem mass spectrometry. *Journal of Chromatography*, 2416-2422.

Maestri, E., Imperiale, D., & Marmiroli, N. (2018). Nuts, nut products and other seeds. In J. Morin, & M. Lees, *FoodIntegrity Handbook* (p. 127-136). Nantes, France: Eurofins Analytics France.

Mannino, G., Gentile, C., & Maffei, M. E. (2019). Chemical partitioning and DNA fingerprinting of some pistachio (Pistacia vera L.) varieties of different geographical origin. *Phytochemistry*, 40-47.

Matthews, S., Santra, M., Mensack, M., Wolfe, P., Byrne, P., & Thompson, H. (2012). Metabolite profiling of a diverse collection of wheat lines using ultraperformance liquid chromatography coupled with time-of-flight mass spectrometry. *Plos One*, e44179.

Matthews, S., Santra, M., Mensack, M., Wolfe, P., Byrne, P., & Thompson, H. (2012). Metabolite Profiling of a Diverse Collection of Wheat Lines Using Ultraperformance Liquid Chromatography Coupled with Time-of-Flight Mass Spectrometry. *PLoS ONE, 7*(8), e44179.

Miano, B., Righetti, L., Piro, R., Dall'Asta, C., Folloni, S., Galaverna, G., & Suman, M. (2018). Direct analysis real-time–high-resolution mass spectrometry for Triticum species authentication. *Food Additives & Contaminants: Part A*, 2291-2297.

Miao, H., Fan, S., Zhou, P. P., Zhang, L., Zhao, Y. F., & Wu, Y. N. (2010). Determination of melamine and its analogues in egg by gas chromatography-tandem mass spectrometry using an isotope dilution technique. *Food Additives & Contaminants*, 1497-1506.

Mie, A., Laursen, K. H., Aberg, K. M., Forshed, J., Lindahl, A., Thorup-Kristensen, K., . . . Husted, S. (2014). Discrimination of conventional and organic white cabbage from a long-term field trial study using untargeted LC-MS-based metabolomics. *Analytical and Bioanalytical Chemistry*, 2885-2897.

Mondal, A. R., Desmarchelier, A., Konings, E., Acheson-Shalom, R., & Delatour, T. (2010). Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) Method Extension To Quantify Simultaneously Melamine and Cyanuric Acid in Egg Powder and Soy Protein in Addition to Milk Products. *Journal of Agricultural and Food Chemistry*, 11574-11579.

Montero-Prado, P., Bentayeb, K., & Nerin, C. (2012). Pattern recognition of peach cultivars (Prunus persica L.) from their volatile components. *Food Chemistry*, 724-731.

Nascimento, C. F., Santos, P. M., Rodrigues Pereira-Filho, E., & Rocha, F. R. (2016). Recent advances on determination of milk adulterants. *Food Chemistry*, 1232-1244.

No 1169/2011 of the European Parliament and of the council of 25 October 2011 on the provision of food information to consumers (2011).

Papapetros, S., Louppis, A., Kosma, I., Kontakos, S., Badeka, A., & Kontominas, M. G. (2018). Characterization and differentiation of botanical and geographical origin of selected popular sweet cherry cultivars grown in Greece. *Journal of Food Composition and Analysis*, 48-56.

Pastor, K., Ačanski, M., Vujić, D., Bekavac, G., Milovac, S., & Kravić, S. (2016). Rapid Method for Small Grain and Corn Flour Authentication Using GC/EI–MS and Multivariate Analysis. *Food Analytical Methods , 9*, 443-450.

Pastor, K., Acanski, M., Vujic´, D., Jovanovic´, D., & Wienkoop, S. (2016). Authentication of Cereal Flours by Multivariate Analysis of GC–MS Data. *Chromatographia, 79*, 1387–1393.

Pastor, K., Ilić, M., Vujić, D., Jovanović, D., & Ačanski, M. (2019). Characterization of Fatty Acids in Cereals and Oilseeds from the Republic of Serbia by Gas Chromatography – Mass Spectrometry (GC/MS) with Chemometrics. *Analytical Letters*, 1177-1189 .

Pontes, M., Pererira, J., & Camara, J. S. (2012). Dynamic headspace solid-phase microextraction combined with one-dimensional gas chromatography–mass spectrometry as a powerful tool to differentiate banana cultivars based on their volatile metabolite profile. *Food Chemistry*, 2509-2520.

Ponzi, L. (2017). Turning Science into Solutions. In M. Monti, & L. Ponzi, *Cibo Criminale (Criminal Food)* (p. 87-88). Newton Compton Editori.

Porto-Figueira, P., Freitas, A., Cruz, C., Figueira, J., & Camara, J. (2015). Profiling of passion fruit volatiles: An effective tool to discriminate between species and varieties. *Food Research International*, 408-418.

Prandi , B., Bencivenni , M., Tedeschi, T., Marchelli , R., Dossena, A., Galaverna, G., & Sforza, S. (2012). Common wheat determination in durum wheat samples through LC/MS analysis of gluten peptides. *Analytical and Bioanalytical Chemistry, 403*, 2909-2914.

Prandi, B., Bencivenni, M., Tedeschi, T., Marchelli, R., Dossena, A., Galaverna, G., & Sforza, S. (2012). Common wheat determination in durum wheat samples through LC/MS analysis of gluten peptides. *Analytical and Bioanalytical Chemistry*, 2909-2914.

Prelle, A., Spadaro, D., Garibaldi, A., & Gullino, M. L. (2012). Aflatoxins monitoringin Italian hazelnut products by LC-MS. *Food Additives & Contaminants: Part B: Surveillance*, 279-285.

Rabadan, A., Pardo, J. E., Gomez, R., Alvarruiz, A., & Alvarez-Ortì, M. (2017). Usefulness of physical parameters for pistachio cultivar differentiation. *Scientia Horticulturae*, 7-11.

R-Biopharm AG. (2020, April). Tratto da R-Biopharm AG: https://food.r-biopharm.com/products/durotest-s/

Reg. CE 853/2004. (s.d.).

Righetti, L., Rubert, J., Galaverna, G., Folloni, S., Ranieri, R., Stranska-Zachariasova, M., . . . Dall'Asta, C. (2016). Characterization and Discrimination of Ancient Grains: A Metabolomics Approach. *International Journal of Molecular Sciences, 17*(1217), 1-14.

Righetti, L., Rubert, J., Galaverna, G., Hurkova, K., Dall'Asta, C., Hajslova, J., & Stranska-Zachariasova, M. (2018). A novel approach based on untargeted lipidomics reveals differences in the lipid pattern among durum and common wheat. *Food Chemistry*, 775-783.

Righetti, L., Rubert, J., Galaverna, G., Hurkova, K., Dall'Asta, C., Hajslova, J., & Stranska-Zachariasova, M. (2018). A novel approach based on untargeted lipidomics reveals differences in the lipid pattern among durum and common wheat. *Food Chemistry, 240*, 775-783.

Russo, R., Cusano, E., Perissi, A., Ferron, E., Severino, V., Parente, A., & Chambery, A. (2014). Ultra-high performance liquid chromatography tandem mass spectrometry for the detection of durum wheat contamination or adulteration. *Journal of Mass Spectrometry, 49*(12), 1239-1246.

Sataque Ono, E. Y., Hirooka, E. Y., Rossi, C. N., & Ono, M. A. (2011). Mycotoxins in Seeds and Nuts. In V. R. Preedy, R. R. Watson, & V. B. Patel, *Nuts and Seeds in Health and Disease Prevention* (p. 121-127). Elsevier.

Stachniuk, A., & Fornal, E. (2016). Liquid Chromatography-Mass Spectrometry in the Analysis of Pesticide Residues in Food. *Food Analytical Methods*, 1654-1665.

Steingass, C. B., Langen, J., Carle, R., & Schmarr, H. (2014). Authentication of pineapple (Ananas comosus [L.] Merr.) fruit maturity stages by quantitative analysis of c- and d-lactones using headspace solid-phase microextraction

and chirospecific gas chromatography–selected ion monitoring mass spectrometry (GC-MS). *Food Chemistry*, 496-503.

Strojnik, L., Stopar, M., Zlatic, E., Kokalj, D., Gril, M. N., Zenko, B., . . . Ogrinc, N. (2018). Authentication of key aroma compounds in apple using stable isotope approach. *Food Chemistry*, 766-773.

Taghizadeh, S. F., Goumenou, M., Rezaee, R., Alegakis, T., Kokaraki, V., Anesti, O., . . . Karimi, G. (2019). Cumulative risk assessment of pesticide residues in different Iranian pistachio cultivars: Applying the source specific HQs and adversity specific HIa approaches in Real Life Risk Simulations (RLRS). *Toxicology Letters*, 91-100.

UPOV. (1994, November 2004). GUIDELINES FOR THE CONDUCT OF TESTS FOR DISTINCTNESS, HOMOGENEITY AND STABILITY. *GUIDELINES FOR THE CONDUCT OF TESTS FOR DISTINCTNESS, HOMOGENEITY AND STABILITY*.

Vinas, P., Campillo, N., Ferez-Melgarejo, G., & Hernandez-Cordoba, M. (2012). Determination of Melamine and Derivatives in Foods by Liquid Chromatography Coupled to Atmospheric Pressure Chemical Ionization Mass Spectrometry and Diode Array Detection. *Analytical Letters, 45*, 2508–2518.

Wadood, S., Boli, G., Xiaowen, Z., Raza, A., & Yimin, W. (2019). Geographical discrimination of Chinese winter wheat using volatile compound analysis by HS-SPME/GC-MS coupled with multivariate statistical analysis. *Journal of Mass Spectrometry, 55*, e4453.

Wong, Y., & Mok, C. (2013). A single analytical procedure for the simultaneous and confirmatory determination of melamine and related compounds in various food matrices by isotope dilution gas chromatography-mass spectrometry (ID-GC-MS). *Analytical Methods, 5*, 2305.

Author Contributions

Conceptualization: Daniele Cavanna, Michele Suman; bibliographic study and writing-original draft preparation: Giuseppe Sammarco, Daniele Cavanna; writing -review and editing: all authors; supervision. Michele Suman, Cecilia Loffi

FRAUD	INDUSTRY FIELD	RESEARCH FIELD					
	Common approach	Analytical Technique	Analysis Type	Data Analysis	Short method description	Adulteration LOQ (Limit of Quantitation)	Reference
Adulteration of durum wheat with common wheat	Immunoassay tests NIR spectroscopy	Liquid Chromatography - Mass Spectrometry/Mass Spectrometry (LC-MS/MS)	Target	Linear Regression	Evaluation of C17:0/C21:0 alkylresorcinol homologue ratio in flour and pasta.	5%	(Knödler, Most, Schieber, & Carle, 2010)
	Monoclonal antibody assay (Durotest S - https://food.r-biopharm.com/products/durotest-s/)	LC-MS/MS	Target	Linear Regression	The area ratio of two specific peptides is used to quantify the percentage of common wheat in durum wheat samples.	3%	(Prandi , et al., 2012)
	Reverse phase-high pressure liquid chromatography (RP-HPLC) and free zone capillary electrophoresis (FZCE)	LC-High Resolution MS (LC-HRMS)	Non-Target	Orthogonal Partial Least Square – Discriminant Analysis (OPLS-DA)	A multivariate study that allows the identification of several markers, with digalactosyl diglyceride (DGDG) 36:4 as the most promising one.	3%	(Righetti, et al., 2018)
	Stable isotopes and compound-specific isotopes analysis	LC-HRMS	Non-Target	OPLS-DA	A multivariate study identified heterocyclic amines, polyketides and phospholipids as discriminant compounds.	N.T.	(Matthews, et al., 2012)
		LC-HRMS; LC-MS/MS	Non-Target	Linear Regression	A peptide of Puroindoline-a (Pin-a) that is present only in common wheat was detected and	1%	(Russo, et al., 2014)

					subsequently, a target method has been developed.		
Adulteration of bread		LC-HRMS LC-MS/MS	Non-Target Target	Soft Independent Modelling of Class Analogies (SIMCA) Linear Regression	A multivariate study identified Alk(en)ylresorcinols, Diglycerides and phosphatidylethanolamine as robust markers for the differentiation between whole wheat and refined wheat bread. Identification of specific peptide markers of wheat, rye and spelt and subsequent quantification of their presence in the bread	N.T. 5% for each species	(Geng, Harnly, & Chen, 2016) (Bonick, Huschek, & Rawel, 2017)
False geographical origin declaration		Gas Chromatography-Mass Spectrometry (GC-MS) LC-HRMS	Non-Target Non-Target	Linear Discriminant Analysis (LDA) OPLS-DA	Chinese durum wheat coming from 3 different geographical areas is discriminated with multivariate models created with the 32 main volatile compounds Selection, with multivariate models, of different chemical markers that allow the discrimination of durum wheat coming from Italy, Europe and Extra Europe	N.T. N.T.	(Wadood, Boli, Xiaowen, Raza, & Yimin, 2019) (Cavanna, Loffi, Dall'Asta, & Suman, 2020)
Addition of exogenous	HPLC -MS	LC-MS/MS	Target	Linear Regression	Quantitative analysis of 14 nitrogen-rich adulterants in Corn, Rye, Oat and Wheat	0.05–0.20 mg/kg based	(Frank, Bessaire,

nitrogen-rich adulterants		LC-MS	Target	Linear Regression	Quantitative analysis of Melamine, Ameline, Ammelide and Cyanuric Acid in cereal flour	on the molecules 0.125 mg/kg	Tasser, Goyion, & Delatour, 2017) (C) (Vinas, Campillo, Ferez-Melgarejo, & Hernandez-Cordoba, 2012)
		GC-MS	Target	Linear Regression	Quantitative analysis of Melamine, Ameline, Ammelide and Cyanuric Acid in cereals	0.09-0.25 mg/kg based on the molecules	(Wong & Mok, 2013)
False "gluten-free" declaration		LC-HRMS	Target	Linear Regression	Specific peptide markers are able to identify the presence of wheat flour in oat flour	1 ppm of wheat gluten in oats	(Fiedler, McGrath, Callahan, & Ross, 2014)
False cereal species declaration		LC-HRMS	Non-Target	OPLS-DA	A metabolomic study that allows the identification of specific markers responsible for the discrimination of three ancient grains varieties (emmer, einkorn and spelt)	N.T.	(Righetti, et al., 2016)

		GC-MS	Non-Target	Principal Coordinates Analysis (PCoA)	Authentication of Barley, oats, wheat, and corn species is executed by studying the intensities of fatty acids that were selected with multivariate studies	N.T.	(Pastor, Ilić, Vujić, Jovanović, & Ačanski, 2019)
		GC-MS	Non-Target	PCoA	Authentication of small grains and corn species is executed by studying the sugar and lipids profiles	N.T.	(Pastor, et al., 2016) (Pastor, Acanski, Vujic´, Jovanovic´, & Wienkoop, 2016)
		GC-HRMS	Non-Target	Principal Component Analysis (PCA)	A multivariate study on the cereals' lipid profile allows the discrimination of 7 different species (Barley, Wheat, Einkorn, Emmer, Oats, Rye, Spelt)	N.T.	(Hammann, Korf, Bull, Hayen, & Cramp, 2019)
		Direct Analysis in Real Time-HRMS (DART-HRMS)	Non-Target	PCA PLS-DA	Study on DART-MS as an unconventional untargeted approach for the classification of wheat species	N.T.	(Miano, et al., 2018)

White rice adulteration		LC-MS/MS	Target	RF Support Vector Machines (SVM)	The intensities of 17 lysoglycerophospholipids, coupled with statistical predictive models, are used to discriminate authentic Korean White Rice from mixtures with Chinese White Rice	5% of Chinese Rice addition	(Lim, et al., 2017)

Tab.1: Wheat and other cereals. Industrial and academic approach for frauds detection

MATRIX	**INDUSTRY FIELD**	**RESEARCH FIELD**						
	Common Approach	*Fraud*	*Analytical Technique*	*Analysis Type*	*Data Analysis*	*Short Method Description*	*Adulteration LOQ (Limit of Quantitation)*	*Reference*
Apple	Near Infrared Spectroscopy (NIR) Isotope Ratio Mass Spectrometry (IRMS) Nuclear Magnetic Resonance Spectroscopy (NMR) Polymerase Chain Reaction (PCR)	Geographical origin	Gas-Chromatography-Mass Spectrometry (GC-MS)	Non-target	(Partial Least Square-Discriminant Analysis) PLS-DA	Characterization of flavour composition of apple cultivars in the Northeast Italy obtained through different cultivation methods, by Headspace-Solid Phase Microextraction (HS-SPME)/GC-MS analysis of volatile fraction with chemometric tools for class modelling	N.T.	(Giannetti, Mariani, Mannino, & Marini, 2017)
		Authenticity	GC-IRMS	Non-target	/	Determination of apple aroma through extraction via HS-SPME of key volatile	N.T.	(Strojnik, et al., 2018)

						compounds analyzed then using GC-IRMS technique, for authenticity assessment		
Banana		Botanical origin	GC-MS	Non-target	(Principal Component Analysis)PCA SL-DA	Evaluation and determination of the volatile profile of five different banana varieties by dHS-SPME/GC-MS and multivariate analysis	N.T.	(Pontes, Pererira, & Camara, 2012)
Cherry		Origin	GC-MS	Non-target	Multivariate analysis of variance (MANOVA) Linear Discriminant Analysis (LDA)	Characterization and differentiation of four sweet cherry Greek cultivars, according to both botanical and geographical origin. One of the approaches employed is the identification and semi-quantification of volatile compounds through SPME/GC-MS	N.T.	(Papapetros, et al., 2018)
Citrus fruit		Authenticity	LC-MS	Non-target Target	PCA Soft Independent Modelling of Class Analogies (SIMCA) Orthogonal PLS-DA (OPLS-DA) T-test	Study to assess the authenticity of Indian Citrus fruit, detecting its principal markers, starting through an untargeted metabolomic approach and then focusing on these markers via targeted metabolomics (LC-MS)	N.T.	(Jandric, Islam, Singh, & Cannavan, 2015)

Mango Pineapple Jackfruit Baobab fruit Tamarind fruit		Authenticity	GC-MS	Non-target	PCA Hierarchical Cluster Analysis (HCA)	This study compares the non-volatile metabolite of five fruits grown in Tanzania, employing a broad-spectrum GC-MS metabolomics new methodology in fruits	N.T.	(Khakimov, et al., 2016)
Goji berry		Geographical origin	LC-MS	Target	T-test PCA Cluster Analysis (CA) Flexible Statistic-Data Analysis (FS-DA)	Characterization of carotenoid fraction in goji berries through High-Performance LC-Diode Array Detection-MS/MS (HPLC-DAD-MS), to define their geographical origin, distinguishing between Italian and Asian samples	N.T.	(Bertoldi, et al., 2018)
		Geographical origin	LC-MS	Non-target	PCA PLS-DA	Discrimination of phytochemical content between four different geographic origins of Goji berries by applying non-targeted LC–quadrupole Time of Flight (qToF)–MS metabolite profiling, coupled with multivariate analysis	N.T.	(Bodia-Pons, et al., 2014)
		Traceability	GC-MS	Target	PLS-DA	A targeted approach using HS-SPME-GC-MS was performed to	N.T.	(Cuevas, Moreno-Rojas, &

Orange						compare flavour compounds of 'Navelina' and 'Salustiana' orange cultivars from organic and conventional management systems		Ruiz-Moreno, 2017)
		Geographical origin	LC-MS	Non-target	PLS-DA OPLS-DA	Spanish orange authentication, different from foreign Argentinean, Brazilian and South African, using UHPLC-QTOF MS coupled with multivariate analysis.	N.T.	(Diaz, Pozo, Sancho, & Hernandez, 2014)
Passion fruit		Authenticity	GC-MS	Non-target	PCA PLS-DA	HS-SPME/GC-MS and multivariate analysis to discriminate between the volatile composition of nine passion fruits grown at Madeira Island (Portugal)	N.T.	(Porto-Figueira, Freitas, Cruz, Figueira, & Camara, 2015)
Peach		Authenticity	GC-MS	Non-target	PCA Canonical Correlation Analysis (CCA)	The volatile compounds of four peach cultivars (Prunus persica L.) were studied through HS-SPME/GC-MS, in order to avoid fraud and protect the European Designation of Origin ‘Melocotón de Calanda’	N.T.	(Montero-Prado, Bentayeb, & Nerin, 2012)
Pineapple		Authenticity	GC-MS	Target	Analysis of variance (ANOVA)	Quantitative determination of δ-lactones	0.001-0.115 mg/kg based	(Steingass, Langen, Carle,

					Kruskal-Wallis test	(d-C8, d-C10) and γ-lactones (c-C6, c-C8, c-C10) by HS-SPME/GC-MS in pineapples, at three progressing post-harvest stages of fully ripe air-freighted and green-ripe sea-freighted fruits, covering the relevant shelf-life of the fruits	on the molecule	& Schmarr, 2014)
Grape		Authenticity	GC-MS LC-MS	Non-target	CA Heatmap visualization	Molecular profiling of volatile aroma metabolites and their precursors in six *Vitis vinifera* cultivars two American species and two interspecific crosses. Chemical profiling was achieved through the combined use of two orthogonal techniques, GC-MS and LC-High Resolution MS (LC-HRMS), before and after enzymatic hydrolysis	N.T.	(Ghaste, et al., 2015)
Wild strawberry		Geographical origin	LC-MS	Non-target Target	PCA	The phytochemical content of *Fragaria vesca* L. grown in South Italy was investigated. Extracts were submitted	0.004-0.09 g/L based on each molecule	(D'Urso, et al., 2016)

						to untargeted LC-Electrospray-Orbitrap-MS (LC-ESI-Orbitrap-MS) and the selected known metabolites were quantitatively analysed by targeted LC-ESI-QTrap-MS/MS		

Tab.2: Fruits. Industrial and academic approach for frauds detection

MATRIX	**INDUSTRY FIELD**	**RESEARCH FIELD**						
	Common Approach	*Fraud*	*Analytical Technique*	*Analysis Type*	*Data Analysis*	*Short Method Description*	*Adulteration LOQ*	*Reference*
Almond	Visual and Organoleptic examination	Geographical Origin	LC-HRMS	Non-target	PLS-DA	Ultra HPLC-QTOF-based metabolomics approach aimed to classify almond variety and country	N.T.	(Gil Solsona, Boix, Ibanez, & Sancho, 2017)
Cashew nut	Enzyme-linked immunosorbent assay (ELISA) DNA markers analysis Lateral flow immunoassay	Contamination	LC-MS/MS	Target	ANOVA and T-test	Detection of fungal contamination by *Aspergillus* and aflatoxins through tandem mass spectrometry in raw cashew kernels (disinfected and non-disinfected samples) of two Beninese zones	0.2-1.1 µg/kg based on the molecule	(Lamboni, et al., 2016)

		Contamination	LC-MS/MS	Non-target	ANOVA	Profiling of fungal metabolites and mycotoxins of cashew nuts from Nigeria and South Africa through LC-MS/MS	0.03-73.00 µg/kg based on the molecule	(Adetunij, et al., 2019)
Chestnut		Contamination	LC-MS/MS	Target	ANOVA	Detection of *Aspergillus* and *Penicillium* toxins in fresh and dried chestnuts of northern Italy, extracted and purified through immunoaffinity or prepacked columns, and analysed using HPLC-MS/MS	0.05-2.50 µg/kg based on the molecule	(Bertuzzi, Rastelli, & Pietri, 2014)
		Geographical origin	GC x GC-MS	Non-target	Comprehensive Template Matching	Profiling of volatile fractions of roasted hazelnuts from nine different geographical origins, sampled by headspace-solid phase micro extraction and then analysed through GC x GC-MS	N.T.	(Cordero, et al., 2010)
		Contamination	LC-MS/MS	Target	Kruskal-Wallis Mann-Whitney	Analyses of hazelnut samples for aflatoxins are carried out, extracting these compounds with two different extraction mixtures and analysing them	0.20-0.44 µg/kg based on the molecule	(Prelle, Spadaro, Garibaldi, & Gullino, 2012)

Hazelnut						through tandem mass spectrometry. Samples are divided according to their geographical origin, the agricultural practice adopted and commercial typology		
		Authenticity	LC-MS	Non-target	HCA	Authentication of 29 different European hazelnut cultivars achieved by phenotypic features studies, phenolic profiling and mass spectrometric analyses of peptide/protein components extracted from kernels	N.T.	(Ciarmiello, et al., 2014)
		Geographical origin	LC-MS	Non-target	Support Vector Machines (SVM) SIMCA	Untargeted metabolomic approaches for geographical origin discrimination of hazelnuts from Georgia, Turkey, Italy, France and Germany, carried out using a UPLC-MS technique	N.T.	(Klockmann, Reiner, Bachmann, Hackl, & Fischer, 2016)
		Geographical origin	LC-MS/MS	Target	SVM	Targeted metabolomic approaches for geographical origin discrimination of hazelnuts from	<20-15%	(Klockmann, Reiner, Cain, & Fischer, 2017)

						Georgia, Turkey, Italy, Spain, France and Germany, analysing 20 non-polar key metabolites through the UPLC-MS/MS technique		
		Geographical origin	LC-MS	Non-target	OPLS-DA	The untargeted metabolomic approach aimed to discern 15 hazelnut cultivars and to discriminate the geographical origin of 6 cultivars, profiling phenolic and sterolic compounds through the LC-MS technique	N.T.	(Ghisoni, et al., 2020)
		Authenticity	GC-MS	Target	Linear Regression Analysis	Hazelnut authenticity assessment through the identification of Filbertone marker via GC-MS analysis	5 µg/kg	(Cizkova, Rajchl, Snebergrova, & Voldrich, 2013)
Peanut		Labelling	LC-MS/MS	Non-target	Mascot	Detection of protein allergens (Ara h 1, Ara h 2 and Ara h 3) and Advanced Glycation Endproducts (AGE) in raw and roasted hazelnuts through proteomics approach, using Sodium Dodecyl Sulphate-Polyacrylamide Gel	N.T.	(Hebling, McFarland, Callahan, & Ross, 2012)

						Electrophoresis (SDS-PAGE), western blot and LC-MS/MS techniques		
Pistachio nut		Geographical origin	LC-MS/MS	Non-target	ANOVA PCA	Discrimination of Pistachio nuts of different origins via detection of key metabolites through HPLC-DAD-MS/MS	N.T.	(Mannino, Gentile, & Maffei, 2019)
		Contamination	GC-MS	Non-target	Mann-Whitney Kruskal-Wallis	Detection of volatile compounds through GC-MS for cumulative risk assessment of pesticide residues in different Iranian pistachio cultivars	0.0003-0.049 mg/kg based on the molecule	(Taghizadeh, et al., 2019)
		Authenticity	GC-MS	Non-Target	ANOVA PCA	Volatile compounds investigation and flavour profiles characterization of fresh pistachio nuts of several Turkish varieties	N.T.	(Kendirci & Onogur, 2011)
		Adulteration	LC-MS	Non-target	PCA HCA	Detection of green pea adulteration in pistachio nut powder and related products through untargeted HPLC-QTOF MS	N.T.	(Cavus, Us, & Guzelsoy, 2018)
Walnut		Labelling	LC-MS/MS	Non-target	Mascot ANOVA	This study evaluates the changes in solubility and	N.T.	(Downs, Baumert,

						detectability of allergens from roasted walnuts, obtained employing different roasting temperatures and times, using tandem mass spectrometry methods		Taylor, & Mills, 2016)
		Geographical origin	GC	Non-Target	PCA-LDA	Gas chromatographic fatty acids fingerprints in combination with multivariate statistics to classify Iranian walnuts based on their geographical origin	N.T.	(Esteki, et al., 2017)

Tab.3: Nuts. Industrial and academic approach for frauds detection

MATRIX	**INDUSTRY FIELD**	**RESEARCH FIELD**						
	Common Approach	*Fraud*	*Analytical Technique*	*Analysis Type*	*Data Analysis*	*Short Method Description*	*Adulteration LOQ*	*Reference*
Egg	Dry matter testing for water addition	Adulteration	LC-MS/MS	Target	Linear Regression Analysis	Study on an accurate method for the simultaneous determination of 15 illegal dyes in eggs and	0.01-5.61 µg/Kg based on the molecule	(Liu, Hei, He, & Li, 2011)

	Lactic acid quantification (Reg. CE 853/2004)					other matrices, through UHPLC-MS/MS		
	Enzymatic assays with a specific molecular marker target (3-hydroxybutyric acid) (Reg. CE 853/2004). Chromatography-Mass Spectrometry	Adulteration	LC-MS/MS	Target	Linear Regression Analysis	Simultaneous determination of Sudan dyes and their metabolites in eggs and other 12 animal-derived foods through ultra-sound assisted extraction and LC-ESI-MS/MS analysis	0.2 µg/Kg	(Chen, et al., 2013)
		Freshness	LC-MS	Non-target Target	T-test PCA ANOVA Post-hoc Tukey	Metabonomic profiling through HPLC-QTOF-MS was carried out to unravel differences in the small molecule profiles of egg yolks. A follow-up study was then performed to better understand how choline was involved in egg aging	N.T.	(Johnson A. E., Sidwick, Pirgozliev, Edge, & Thompson, 2019)
		Freshness	LC-MS	Target	PCA OPLS-DA	Metabolomic approach on 31 selected compounds involved in freshness or no freshness of egg products, using UHPLC-HRMS technique and multivariate statistical analysis	N.T.	(Cavanna, Catellani, Dall'Asta, & Suman, 2018)
		Freshness	LC-MS	Non-target Target	ANOVA PCA T-test	Application of metabonomic techniques to uncover differences in	N.T.	(Johnson A. , Sidwick, Pirgozliev,

						the small molecule of chicken eggs to combat fraud regarding freshness within the egg industry. HPLC-QTOF-MS and multivariate statistical analysis were both employed for this study		Edge, & Thompson, 2018)
		Incubation	LC-MS	Target	ANOVA T-test	Study on the efficiency of European legislative indices and other compounds for incubator-reject eggs (IRE) detection in egg products, using the UHPLC-HRMS technique	N.T.	(Hidalgo, Galbiati, Cavanna, & Suman, 2019)
		Adulteration	LC-MS/MS	Target	Horwitz ratio	Demonstration of the ability of a clean-up-free LC-MS/MS method to quantify melamine and cyanuric acid in egg powder and soy protein	0.02-0.1 mg/Kg based on the molecule	(Mondal, Desmarchelier, Konings, Acheson-Shalom, & Delatour, 2010)
		Adulteration	GC-MS/MS	Target	Linear regression analysis	Development of a method for the simultaneous determination of melamine and its analogues in the egg using GC-MS/MS	N.T.	(Miao, et al., 2010)
		Authenticity	LC-MS	Non-target	PCA T-test F-test	Metabonomic techniques to prevent fraud regarding the	N.T.	(Johnson A. E., Sidwick, Pirogzliev,

						misrepresentation of egg farming systems, employing HPLC-QToF-MS and multivariate statistical analysis		Edge, & Thompson, 2019)

Tab.4: Egg and egg products. Industrial and academic approach for fraud detection.

OTHER ANALYTICAL TECHNIQUES FOR FOOD AUTHENTICITY ISSUES

The present thesis project not only focused on chromatographic-spectrometric techniques but also considered spectroscopic and innovative technologies, aimed at both rapid and confirmatory analysis. These approaches include NIR spectroscopy, GC coupled with the ion mobility system (IMS), NMR spectroscopy.

NIR Spectroscopy

NIR spectroscopy is in the region of 800-2500 nm, or 12500-4000 cm^{-1} (Siesler, Ozaki, Kawata, & Heise, 2002). It is a non-destructive, rapid, direct analysis method, for this reason, it has a wide field of applications (agriculture, food engineering, pharmaceutical industries, environmental science…) (Ozaki, Genkama, & Futami, 2017). NIR spectroscopy employs different spectral analysis methods, such as chemometrics, to extract chemical info from the spectra, traditional spectral analysis (based on spectra differences), for simpler spectra, two-dimensional correlation spectroscopy, which brings a simplification of the spectra visualization, quantum chemical calculations, with whom it is possible to calculate wavenumbers, and even absorption intensities of the XH overtones (X=C, O, and N) (Mark & Workman, 2018) (Noda & Ozaki, 2005) (Nishida, Shigeto, Yabumoto, & Hamaguchi, 2012). Anyway, original spectra often undergo pretreatments or data transformations before the analysis. There are four different categories of pretreatment methods: noise reduction, such as smoothing; baseline correction, which includes the second derivative, the standard normal variate (SNV), the multiplicative scatter correction (MSC); the centering and normalization; resolution enhancement (Ozaki, Genkama, & Futami, 2017) (Varmuza & Filzmoser, 2009). Both electronic and vibrational (overtones and combinations) transitions are present in the NIR region, the electronic ones are *d-d* transitions of charge transfer (CT) bands, and π-π^* transitions of conjugated systems (Figgis, 1966). In NIR spectroscopy, some bands overlap each other or many bands, because of the overtones and the combination modes. Therefore, the analysis of the NIR bands is not so easy in general. Due to its non-destructiveness and the possibility to perform *in situ* analysis, NIR spectroscopy is ideal to measure the absorption spectrum of an entire matrix, without any previous treatment (Ozaki, Genkama, & Futami, 2017) (Ozaki, Near-Infrared Spectroscopy—Its Versatility in Analytical Chemistry, 2012). In the 1960s, Karl Norris managed to exploit non-destructive analysis on agricultural products, since then, food engineering applications of NIR spectroscopy were the most representative strategy. Several NIR methods were considered time-consuming and cost-effective alternatives, including various applications on food matrices, such as grains, fruits, vegetables, meats, fish, and dairy (Oliveira & Franca, 2011). The main analytes of the agricultural products are sugar, starch, protein, oil… all of them contain -CH, -NH, -OH groups in their compounds, and they are

concentrated enough to be detected by NIR analyses. Nowadays, portable spectrometers are more and more common, as they permit measuring fruits' qualities and agricultural facilities, directly in fields (Ozaki, Genkama, & Futami, 2017).

GC-IMS Analysis

GC-IMS represents another rapid, direct, eventually non-destructive, high-throughput testing and screening of volatile organic compounds (VOCs) of food commodities. It has a quick response, it is user-friendly and cost-effective (Wang, Chen, & Sun, 2020). This approach combines the chromatographic separation of volatile compounds, based on their polarity, to the subsequent ion mobility separation, according to the compounds' dimension and shape. This innovative ion mobility strategy, coupled with the classic GC method, provides a versatile solution for sensitive and selective volatile fingerprinting. The GC-IMS process can be divided into five phases: sample introduction, such as headspace injection mode, compound separation by chromatographic effect, ion generation, usually a soft chemical ionization, initiated by a low-radiation tritium (^{3}H) source, ion separation and detection, through the drift tube and faraday plate, respectively. Since IMS separates ions according to their mobilities rather than mass, this technique allows us to selectively detect molecules having the same mass, but a different structure (Kanu & Hill Jr., 2008) (G.A.S. Dortmund, s.d.). In particular, the IMS workflow starts with the gaseous sample introduced into the ionization reaction region by a carrier gas (usually nitrogen), then several the sample undergoes an ionization reaction with the carrier gas molecules, under the effect of the ion source, leading to the formation of various product ions. These enter the drift region, colliding with a counter-flowing drift gas, that modifies their mobility rates, based on the shape and dimension, resulting in a fine separation (Gu, Zhang, Wang, Wang, & Du, 2021). Advances in this type of technology, such as the introduction of the capillary GC column combined with the IMS, enhanced its use in the field of food quality and safety, i.e. olive oil classification, strawberry *Botrytis cinerarius* detection, characteristic composition in wine detection… Furthermore, in recent years, relevant attention was given to this analytical approach in the analysis/microanalysis of numerous food volatiles, due to its advantages (Garrido-Delgado, et al., 2011) (Vandendriessche, Keulemans, Geeraerd, Nicolai, & Hertog, 2012).

NMR Spectroscopy

NMR is a phenomenon for the study of the physical, chemical, and biological properties of matter. It occurs when nuclei of certain atoms are in a static magnetic field and are then exposed to another oscillating magnetic field. Only some nuclei undergo this phenomenon, depending on the presence of a particular property, named spin. NMR finds applications in several scientific fields; it is

employed to study chemical structures by using simple one-dimensional approaches. Two-dimensional techniques are exploited to define more complicated molecular structures, whereas time domain (TD) NMR spectroscopic strategies are for probing molecular dynamics in solutions. Solid-state NMR is then performed to determine the molecular structure of solids. Therefore, the versatility of this technology makes it a relevant tool in the sciences (Hornak, 2017). NMR spectroscopy has become increasingly popular in food science, to evaluate and analyse various food commodities (i.e., beverages, oils and lipids, vegetables, meat, and dairy products). Initially, NMR use in the food field was limited to low-resolution techniques, whereas recently, high-resolution strategies were introduced to study solid and liquid matrices, for different aims, such as authentication and classification, quality control, sensorial analysis, structural characterization, compositional analysis, unraveling molecular mechanism and interactions of food components, and evaluation of nutritional approaches to health (Spyros & Dais, 2013) (Marcone, et al., 2013) (Malmendal, et al., 2011) (Fernandes, Bras, Mateus, & Freitas, 2015) (Ramakrishnan & Luthria, 2017). A major frequency of NMR applications on food matrices was mainly due to the development of powerful multinuclear/multidimensional and solvent suppression NMR techniques but also advances in hardware, such as cryoprobes, high-throughput approaches, and easy-to-use software. NMR spectroscopy represents a relevant tool for the emerging necessity of food companies and control agencies to satisfy consumers' standards for safe and high-quality foods. The non-destructive character, the elevated accuracy, the reproducibility, the possibility to perform the analyses without any separation and/or purification phases, makes the NMR technology a key player in analyzing multicomponent systems, complex matrices like foods. Furthermore, NMR is also a quantitative technique, since, under certain conditions, the signal area is directly proportional to the number of the nuclei that produce it, and various nuclei can be selected for analysis, according to the nature of the food and the information to obtain (Bharti & Roy, 2012). NMR spectroscopy is one of the few analytical solutions able to provide info about molecular regio/stereochemistry. It is possible to selectively study each atom, and the experiments can be carried out under physiological conditions, a fundamental aspect of biological studies. In addition, the employment of statistical analysis on NMR data led to a significant number of applications for food analysis. However, despite the growing interest in NMR spectroscopy for food studies, it still is underexploited in this field, mostly because of the high cost, the relatively low sensitivity, and the lack of familiarity of food scientists with this technology (Hatzakis, 2018). On the other hand, in the last few years, innovative, cost-effective technologies NMR-based were developed to encourage enterprises in investing in this type of analytical strategy. They were projected as rapid, direct and easy-to-use instrumentations, in order to

favour their introduction into the industrial environment, filling the gap caused by the complexity and the remarkable costs of the classic NMR facilities.

Chemometrics: Multivariate Statistical Analysis

Chemometrics is the discipline that exploits mathematical, statistical, and other formal logic methods to define and choose optimal measurement protocols and experiments, extracting relevant and useful chemical information from the raw data. Differently from the classical approach, which is devoted to understanding the effects, so which factors are influential or not, the chemometric approach underlines other goals, such as prediction, pattern recognition, classification, etc. It uses multivariate methods, considering all the model variables at the same time. In addition, these variables can also be correlated among themselves (Heberger, 2008). This advantage of handling multiple variables contextually can find a useful application in food science studies, as the companies carry out several measurements throughout the production chain. This brings a huge amount of data, and it is of great interest to consider all the features, gaining more relevant information from the collected data (Bro, et al., 2002). Multivariate techniques are mainly divided into two groups: unsupervised and supervised ones. Supervised techniques (STs) are applied to two groups of variables: independent, related to the data obtained from the analysis, and dependent, related to the samples' labels. STs interface independent and dependent variables, returning a model for sample classification. These techniques include partial least square (PLS), *k*-nearest neighbor (KNN), support vector machine (SVM), probabilistic neural network (PNN), linear discriminant analysis (LDA)… Unsupervised techniques (UTs) do not consider dependent variables for their models. Indeed, they provide models exploiting only independent variables, and sample classes are created according to the structure of the variables. These include clustering approaches, such as principal component analysis (PCA), hierarchical cluster analysis (HCA), *k*-means cluster analysis (KMCA), fuzzy C-means (FCM)… (Anzanello, Ortiz, Limberger, & Mariotti, 2014). In particular, PCA and PLS-DA are two of the techniques mostly employed for this thesis project, combining both unsupervised and supervised models. PCA is a mathematical algorithm aimed at reducing the data dimensionality, maintaining most of the data set variation at the same time. It achieves this reduction by defining directions, named principal components (PCs), which drive the maximal variation in the data. Specifically, the PCs of a high number of variables, represented as points, in a Cartesian space, are a sequence of $\boldsymbol{p}$-unit vectors (vectors of length). The $\boldsymbol{i}$-th vector has the direction that fits the data better than the others, and it is orthogonal to the first $\boldsymbol{i} - 1$ vector. Two unit and orthogonal vectors are also called orthonormal, and together they form an orthonormal basis. PCA is the process of computing PCs and using them for a change of basis, which is a technique aimed at redefining vectors according to a

different set of basis elements. This is achieved considering only a few first PCs (Jolliffe & Cadima, 2016). In data analysis, the first PC extrapolated from a set of ***p*** variables is formed by linearly combining the original variables that explain the most variance. The second PC is composed of the variables that explain the most variance once the effect of the first one is removed, and then it is possible to go through ***p*** iterations until the entire variance is explained (Brems, 2017). PCA is mainly employed for exploratory data analysis and prediction models. For these targets, PCs are represented as eigenvectors of the data's covariance matrix. Thus, they are nonzero vectors that belong to a square matrix giving the covariance between each pair of elements of the vector. PCA is the simplest of the eigenvector-based multivariate analysis (Abdi & Williams, 2010). The outputs of the PCA are reported as component scores (or factor scores), which represent the transformed variable values corresponding to a specific data point, and loadings, the weight by which each standardized original variable has to be multiplied to obtain the component score (Shaw, 2019).

PLS regression (PLS-R) is a statistical supervised method aimed at searching for the linear regression model by projecting predicted and observable variables to a new space, instead of looking for hyperplanes of maximum variance between response and independent variables. PLS-DA is a variant of PLS-R when the response variable is categorical and not numerical. This technique is successfully applied to models having many more predictors than observations or with multicollinearity issues, that is the high intercorrelation among explanatory variables in a multiple regression model and leads to incorrect outcomes (Kim, 2019) (Pérez-Enciso & Tenenhaus, 2003). In particular, this approach is exploited to find crucial relations between two data matrices, X and Y. A PLS model focuses on finding the multidimensional direction in the X space able to explain the maximum multidimensional variance direction in the Y space. The general model of PLS is:

$$X = TP^T + E$$
$$Y = UQ^T + F$$

Where X is a n x m matrix of predictors, Y is a n x p matrix of responses. T and U are n x l matrices that represent, respectively, the projection of X, named X scores, and the projection of Y, named Y scores. P and Q are, respectively, m x l and p x l orthogonal loading matrices. E and F figure the error terms. The decomposition of X and Y is performed in order to maximise the covariance between T and U. PLS-DA is then used for predictive and descriptive aims and discriminative variable selection. Thus, PLS-DA unites dimensionality reduction and discriminant analysis, ideal for an application on high-dimensional (HD) data. Moreover, PLS-DA does not assign the data to a particular distribution, and this makes it more flexible than other discriminant algorithms, like LDA. The most common

application of this supervised approach is to classify the observations (samples) in the defined classes (labels), with the largest predicted indicator variable (Lee, Liong, & Jemain, 2018) (Chevallier, Bertrand, Kohler, & Courcoux, 2006).

REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, 433-459.

Anzanello, M. J., Ortiz, R. S., Limberger, R., & Mariotti, K. (2014). Performance of some supervised and unsupervised multivariate techniques for grouping authentic and unauthentic Viagra and Cialis. *Egyptian Journal of Forensic Sciences*, 83-89.

Bharti, S. K., & Roy, R. (2012). Quantitative 1H NMR spectroscopy. *Trends in Analytical Chemistry*, 5-26.

Brems, M. (2017, April 17). *A One-Stop Shop for Principal Component Analysis.* Retrieved from Towards Data Science: https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c

Bro, R., van den Berg, F., Thybo, A., Andersen, C. M., Jorgensen, B. M., & Andersen, H. (2002). Multivariate data analysis as a tool in advanced quality monitoring in the food production chain. *Trends in Food Science & Technology*, 235-244.

Chevallier, S., Bertrand, D., Kohler, A., & Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics*, 221-229.

Fernandes, A., Bras, N. F., Mateus, N., & Freitas, V. D. (2015). A study of anthocyanin self-association by NMR spectroscopy. *New Journal of Chemistry*, 2602-2611.

Figgis, B. N. (1966). *Introduction to Ligand Fields.* Krieger Pub Co.

G.A.S. Dortmund. (n.d.). *IMS Working Principle*. Retrieved from GAS Dortmund: https://www.gas-dortmund.de/Technology/IMS-Working-Principle/1_388.html

Garrido-Delgado, R., Mercader-Trejo, F., Sielemann, S., De Bruyn, W., Arce, L., & Valcarcel, M. (2011). Direct classification of olive oils by using two types of ion mobility spectrometers. *Analytica Chimica Acta*, 108-115.

Gu, S., Zhang, J., Wang, J., Wang, X., & Du, D. (2021). Recent development of HS-GC-IMS technology in rapid and nondestructive. *Trends in Analytical Chemistry*, 116435.

Hatzakis, E. (2018). Nuclear Magnetic Resonance (NMR) Spectroscopy in Food Science: A Comprehensive Review. *Comprehensive Reviews in Food Science and Food Safety*, 189-220.

Heberger, K. (2008). Chapter 7 - Chemoinformatics—multivariate mathematical–statistical methods for data evaluation. In K. Vekey, A. Telekes, & A. Vertes, *Medical Applications of Mass Spectrometry* (pp. 141-169). Elsevier Science.

Hornak, J. P. (2017). *The Basic of NMR*. Retrieved from Rochester Institute of Technology: https://www.cis.rit.edu/htbooks/nmr/inside.htm

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A*, 20150202.

Kanu, A. B., & Hill Jr., H. H. (2008). Ion mobility spectrometry detection for gas chromatography. *Journal of Chromatography A*, 12-27.

Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean Journal of Anesthesiology*, 558-569.

Lee, L. C., Liong, C., & Jemain, A. A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst*, 3526-3539.

Malmendal, A., Amoresano, C., Trotta, R., Lauri, I., De Tito, S., Novellino, E., & Randazzo, A. (2011). NMR spectrometers as “magnetic tongues”: Prediction of sensory descriptors in canned tomatoes. *Journal of Agricultural and Food Chemistry*, 10831-10838.

Marcone, M. F., Wang, S., Albabish, W., Somnarain, D., Hill, A., & Nie, S. (2013). Diverse food-based applications of nuclear magnetic resonance (NMR) technology. *Food Research International*, 729-747.

Mark, H., & Workman, J. J. (2018). *Chemometrics in Spectroscopy.*

Nishida, J., Shigeto, S., Yabumoto, S., & Hamaguchi, H. (2012). Anharmonic coupling of the CH-stretch and CH-bend vibrations of chloroform as studied by near-infrared electroabsorption spectroscopy. *The Journal of Chemical Physics*, 234501.

Noda, I., & Ozaki, Y. (2005). *Two-Dimensional Correlation Spectroscopy: Applications in Vibrational and Optical Spectroscopy*.

Oliveira, L. S., & Franca, A. S. (2011). Applications of near infrared spectroscopy (NIRS) in food quality evaluation. In D. A. Medina, & A. M. Laine, *Food Quality: Control, Analysis and Consumer Concerns.* Nova Science Publishers, Inc.

Ozaki, Y. (2012). Near-Infrared Spectroscopy—Its Versatility in Analytical Chemistry. *Analytical Sciences*, 545-563.

Ozaki, Y., Genkama, T., & Futami, Y. (2017). Near-Infrared Spectroscopy. In J. C. Lindon, G. E. Tranter, & D. W. Koppenaal, *Encyclopedia of Spectroscopy and Spectrometry* (p. 3584).

Pérez-Enciso, M., & Tenenhaus, M. (2003). Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (pls-da) approach. *Human Genetics*, 581-592.

Ramakrishnan, V., & Luthria, D. L. (2017). Recent applications of NMR in food and dietary studies. *Journal of the Science of Food and Agriculture*, 33-42.

Shaw, P. J. (2019). *Introductory Multivariate Statistics for the Environmental Science 1st Edition.* Wiley.

Siesler, H. W., Ozaki, Y., Kawata, S., & Heise, H. M. (2002). Near-Infrared Spectroscopy. Principles, Instruments, Applications. *Journal of Chemometrics*, 636-638.

Spyros, A., & Dais, P. (2013). *NMR spectroscopy in food analysis.* Cambridge: Royal Society of Chemistry.

Vandendriessche, T., Keulemans, J., Geeraerd, A., Nicolai, B. M., & Hertog, M. L. (2012). Evaluation of fast volatile analysis for detection of Botrytis cinerea infections in strawberry. *Food Microbiology*, 406-414.

Varmuza, K., & Filzmoser, P. (2009). *Introduction to Multivariate Statistical Analysis in Chemometrics.* CRC Press.

Wang, S., Chen, H., & Sun, B. (2020). Recent progress in food flavor analysis using gas chromatography–ion mobility spectrometry (GC–IMS). *Food Chemistry*, 126158.

SENSOMICS: BASIS AND PRINCIPLE

Flavour perception is constituted by multiple sensory features, that are due to the complex combination of volatile molecules in the food commodity. For instance, the aroma plays a relevant role in contributing to the taste of wine or alcoholic beverages, whereas for meat and cheese, it is a mix of taste and smell that defines their flavoury characteristics (Chapman, et al., 2019).

Sensomics belongs to the wide family of 'omics' sciences applied to food, and its target is to describe the sensory characteristics of food commodities. Its point of strength is represented by the high capability of unraveling unknown information about sensory properties and the effects of several ingredients, as well as industrial processes on the sensory features of foods and beverages (Vrzal & Olsovska, 2019). This evaluation approach turns useful for companies and scientists by providing remarkable information on the sensory quality of food, which is then handled to assess the consumers' appreciation (Vivek, Subbarao, Routray, Kamini, & Dash, 2020).

Recently, papers regarding sensory studies on several food commodities are becoming frequent. A work about the potent odorant of medium-roasted Arabica coffee and its corresponding brew quantification was done by Mayer *et al.* (Mayer, Czerny, & Grosch, 2000). Another research evaluated the impact of monosodium glutamate (MSG) and its substitutes on chicken soup by correlating sensory parameters and consumers' acceptability (Wang, Zhang, & Adhikari, 2019). Sensory analysis can also be applied to evaluating the influence of specific processes, such as the emulsification of full-fat mayonnaise. Sensory and instrumental tools were employed to estimate the effect of the process (Olsson, Hakansson, Purhagen, & Wendin, 2018). The sensory analysis turns out to be effective also for sustainability studies, a work focused on the sensory properties and consumers' liking of almonds grown with a reduced quantity of water (Lipan, et al., 2019). Another application of sensory studies is on natural food, 20 ingredients were evaluated to define what is considered natural by the customers (Chambers V, Chambers IV, & Castro, 2018).

The present thesis work exploited the sensory analysis coupled with instrumental analysis to solve authenticity issues related to the matrices of interest, with the so-called sensomics approach. Interfacing flavour analysis and instrumental analytical strategies could be optimal for food companies, as the former aims to directly evaluate the potential rate of appreciation of the products by consumers, giving at the same time orthogonal indications for specific analytical solutions. For example, the study of the food aroma can be merged with the evaluation of the volatile organic compounds (VOCs), as well as the identification of molecules through high-resolution techniques could find an interesting match with the analysis of the food taste. Therefore, this approach can provide for innovative and robust applications aimed at solving authenticity and safety issues, as well as for the quality improvement of diverse food products.

REFERENCES

Chambers V, E., Chambers IV, E., & Castro, M. (2018). What Is "Natural"? Consumer Responses to Selected Ingredients. *Foods*, 65.

Chapman, J., Elbourne, A., Khanh Troung, V., Newman, L., Gangadoo, S., Rajapaksha Pathirannahalage, P., . . . Cozzolino, D. (2019). Sensomics - From conventional to functional NIR spectroscopy - Shining light over the aroma and taste of foods. *Trends in Food Science & Technology*, 274-281.

Lipan, L., Cano-Lamadrid, M., Corell, M., Sendra, E., Hernandez, F., Stan, L., . . . Carbonell-Barrachina, A. A. (2019). Sensory Profile and Acceptability of HydroSOStainable Almonds. *Foods*, 64.

Mayer, F., Czerny, M., & Grosch, W. (2000). Sensory study of the character impact aroma compounds of a coffee beverage. *European Food Research and Technology*, 272-276.

Olsson, V., Hakansson, A., Purhagen, J., & Wendin, K. (2018). The Effect of Emulsion Intensity on Selected Sensory and Instrumental Texture Properties of Full-Fat Mayonnaise. *Foods*, 9.

Vivek, K., Subbarao, K. V., Routray, W., Kamini, N. R., & Dash, K. K. (2020). Application of Fuzzy Logic in Sensory Evaluation of Food Products: a Comprehensive Study. *Food and Bioprocess Technology*, 1-29.

Vrzal, T., & Olsovska, J. (2019). Sensomics - basic principles and practice. *KVASNY PRUMYSL*, 166-173.

Wang, S., Zhang, S., & Adhikari, K. (2019). Influence of Monosodium Glutamate and Its Substitutes on Sensory Characteristics and Consumer Perceptions of Chicken Soup. *Foods*, 71.

FOOD COMMODITIES OF INTEREST

The focus of the abovementioned analytical technologies was on strategic food matrices. This strategy is due to the wide employment of this food by the companies, and the increasing rate of illegal activities on them, for economic purposes. The combination of these two aspects has pushed researchers to develop disparate techniques to fight/prevent fraud, as well as preserve food quality.

Apples

Regarding the apple chain, the dehydrated matrix was taken into account in the present study. Figure 1 reports the production workflow, from the harvesting to the packaging.

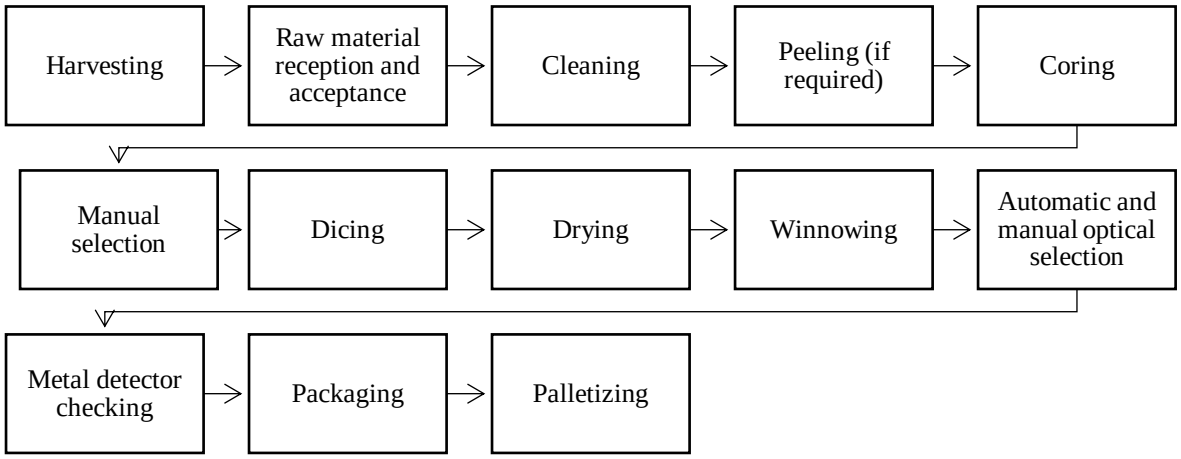


Fig. 1 *Apples dehydration workflow*

Once the fruits reach the industrial plant, two initial selections are done, according to the quality and the ripening rate. Afterward, there is the peeling step, which could be accomplished mechanically or chemically, by employing NaOH solution. Subsequently, the coring step is followed by blanching, a pre-cooking phase to inactivate enzymes responsible for the fruit browning. Another way to avoid this browning phenomenon is gaseous sulfurization, keeping the SO_2 concentration below 500 ppm, or 250 ppm if the apple is for direct consumption. The residual moisture rate could vary from 5 to 24 %, according to the form of the cut (cube, slice, powder, granulate…). The drying process is achieved by exploiting diverse methodologies, such as convective drying with hot air, suitable for cubes and slices, spray drying for powder and granulate, and drum drying for puree. At this point,

manual and optical checks, followed by metal detector control are done, before the packaging and palletizing final steps.

The scientific studies on apple matrix mainly considered flavours identification, chemical composition, taste, and variety, because of their unique aromatic characteristics (Elss, Preston, Appel, Heckel, & Schreier, 2006). Authenticity frauds are some of the most complex to deal with, and the scientific papers about them are relatively few. Illegal authenticity activities not only regard apples' geographical origin but also their botanical origin, usually related to the cultivation area. The geographical origin claim is then a fundamental feature for the apple quality assessment, for this reason, authenticity-oriented frauds find their target, which is strictly economical.

Among the techniques applied for apple authenticity studies, gas chromatography (GC) and high-pressure liquid chromatography (HPLC) were employed by Jihong *et al.*, making possible to discriminate among eight cultivars (Jihong, et al., 2007). HPLC was also used for flavonoid quantification, to distinguish among four different cultivars, as well as the antioxidant activity, evaluated by measuring the inhibition of lipid peroxidation in rat liver microsomes by the apple samples. Moreover, this study also considered samples from different harvesting campaigns, to assess the robustness of the methods over the yearly variability (van der Sluis, Dekker, De Jager, De Jager, & Jongen, 2001). Flavour characterisation is another valuable strategy to define apples cultivated in the North-Eastern area of Italy. Headspace-solid phase microextraction/GC-mass spectrometry (HS-SPME/GC-MS) was the technology exploited for the aroma analysis, and the raw data obtained were handled by a chemometrics approach (Giannetti, Boccacci Mariani, & Marini, Volatile fraction by HS-SPME/GC-MS and chemiometric modeling for traceability of apples cultivated in the Northeast Italy, 2017). The aromatic compounds were taken into account also with the GC-combustion-isotope ratio mass spectrometry (GC-C-IRMS), comparing commercial aromas with the ones extracted from the fruit, to verify their authenticity (Strojnik, et al., 2019). A bidimensional LC (LCxLC) approach was performed to identify diverse phenolic compounds in several apple varieties. Particularly, the analytical method combines the hydrophilic interaction chromatography (HILIC), in the first separation dimension, and the reverse phase (RP) LC in the second dimension, using the diode array detection (DAD) and the MS as detectors. This approach could be successfully applied also on untargeted and targeted metabolomics studies (Montero, Herrero, Ibanez, & Cifuentes, 2013).

The apple-based products analysed in the present project are the dehydrated apple cubes. The scientific literature, to the best of the author’s knowledge, does not provide papers about authenticity studies on this specific product. This aspect makes the current studies novel and original, as well as of great interest to food companies, that widely employ this commodity for several bakery products.

Hazelnuts

The hazelnut chain, reported in Figure 2, is represented, from the harvesting to the roasting and pasting processing.

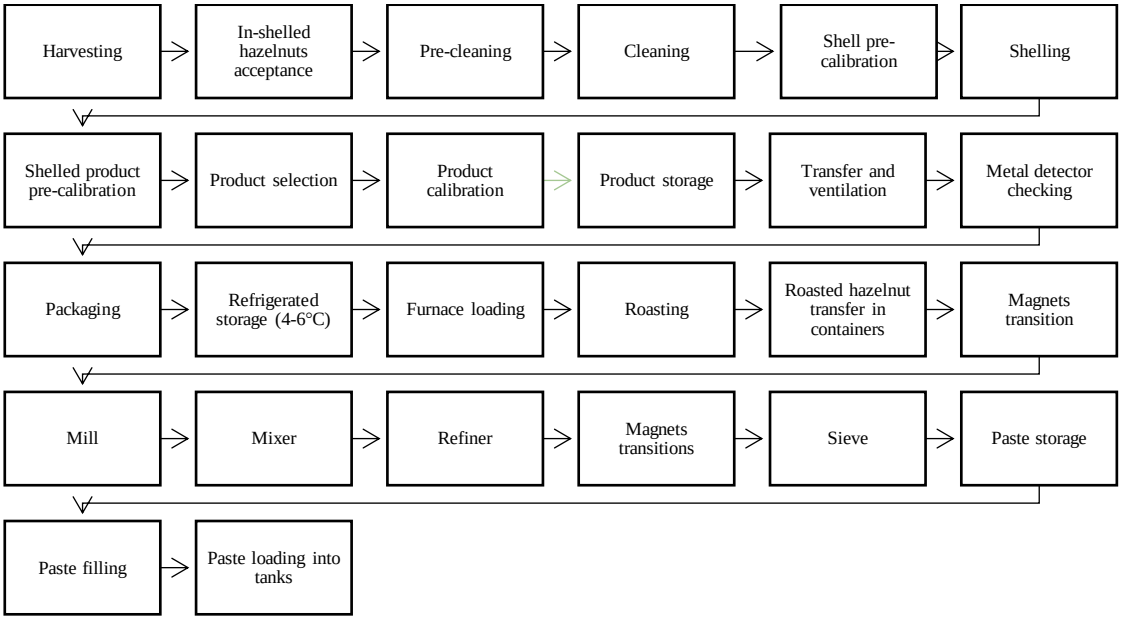


Fig. 2 *Hazelnut chain workflow*

After the harvesting and the acceptance, pre-cleaning and cleaning steps are done by initially eliminating wood, powder, sticks, and then heavier fragments. Subsequently, shell pre-calibration is accomplished, considering dimensions between 17 (small) and 19 (big) mm. The shelling, calibration, and metal detector check are followed by fresh hazelnut storage or loading in the furnace for the roasting step. Magnets, mill, mixer and refiner are employed to produce paste products. The roasting process occurs under hot air or infrared radiation, in a continuous cycle, at different temperatures (130-160°) and timing (20-40 minutes), according to the kernel diameter and degree of roasting required. The peeling is achieved during the cooling step. The hazelnut grain is obtained through a granulator equipped with rotatory blades, that could be regulated in relation to the size desired. Pastes are obtained following two processing phases, grinding and subsequent refining. The final result is a dense and homogenous product, with 65-70 % of oil.

The hazelnut chain is currently committed by authenticity frauds, as the nation of origin and the indication of the species is perhaps the most important characteristic for tree nuts concerning illegal activities. This information is typically required on the label and frequently these features are the basement for the development of protected denominations such as PGI and PDO certifications. In addition, fraudulent actions can derive from declarations regarding organic, Kosher, or Halal production too for this category of products.

DNA-based methods are commonly employed for hazelnut species or cultivar assessment, such as polymerase chain reaction (PCR) or immunological methods like enzyme-linked immunosorbent assay (ELISA), for instance (Rohman, et al., 2021). Other methods based on fatty acid analysis, chemical markers (i.e. filbertone, tocopherol), or volatile profiles, have been developed to verify the authenticity of the species or cultivar, even in highly processed products (Čížková, Rajchl, Šnebergrová, & Voldřich, 2013) (Castillo, Caballero, Blanch, & Herraiz, 2002). Furthermore, it is commonly proved that the chemical profile may distinguish nuts based on their geographic origin in some situations, although DNA markers are less effective in this kind of assessment (Esteki, et al., 2017). However, even if current methods can identify correctly species and cultivars, the provenance assessment is generally harder and more complex (Food Integrity Handbook, 2018). In addition, non-targeted techniques carried out with spectroscopic methodologies have been recently developed, to build authentic product fingerprints based on the chemical composition of different samples. This approach has been coupled with screening techniques based on portable instruments in the last few years (Manfredi, et al., 2018).

The entire hazelnut chain, from the fresh matrix to the roasted and the pasted ones, was considered for the project, due to the supplier's availability, and the big interest of the company in respecting the 100 % Italian claim, protecting the PDO and PGI certified Italian hazelnuts.

Final remarks

All the reported research regarding authenticity studies was mainly conducted by academic groups. This state-of-art of the two chains highlights a promising interest in looking for analytical solutions for facing/preventing fraud, but, contextually, it puts in evidence a scientific gap between university and enterprise. All the innovative and effective technologies above-mentioned are barely considered by food companies, only a few among the big ones invested in them and in the expertise required to work with them. Besides the costs, there is also a lack of communication between the academy and the industry. Many companies do not consider open innovation with the academic research group useful for several reasons, such as the confidentiality policy, the lack of knowledge about the innovative technologies, the different research attitudes adopted by the university…

The present thesis work was carried out mainly in an industrial environment. In spite of that, the project was drafted and planned together with an academic research group, with whom there was a close and fruitful collaboration. Therefore, each analytical solution considered during the studies was exploited with an academic approach to research, then further evaluations about its potential application in the enterprise environment were accomplished.

REFERENCES

Anzanello, M. J., Ortiz, R. S., Limberger, R., & Mariotti, K. (2014). Performance of some supervised and unsupervised multivariate techniques for grouping authentic and unauthentic Viagra and Cialis. *Egyptian Journal of Forensic Sciences*, 83-89.

Bharti, S. K., & Roy, R. (2012). Quantitative 1H NMR spectroscopy. *Trends in Analytical Chemistry*, 5-26.

Bro, R., van den Berg, F., Thybo, A., Andersen, C. M., Jorgensen, B. M., & Andersen, H. (2002). Multivariate data analysis as a tool in advanced quality monitoring in the food production chain. *Trends in Food Science & Technology*, 235-244.

Castillo, R. d., Caballero, G., Blanch, & Herraiz. (2002). Enantiomeric composition of filbertone in hazelnuts and hazelnut oils from different geographical origins. *J. Am. Oil Chem. Soc.*, 589-592. doi:10.1007/s11746-002-0527-1

Chevallier, S., Bertrand, D., Kohler, A., & Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. *Journal of Chemometrics*, 221-229.

Čížková, Rajchl, Šnebergrová, & Voldřich. (2013). Filbertone as a marker for the assessment of hazelnut spread quality. *Czech J. Food Sci.*, 81-87. doi:10.17221/493/2011-CJFS

Elss, S., Preston, C., Appel, M., Heckel, F., & Schreier, P. (2006). Influence of technological processing on apple aroma analysed by high resolution gas chromatography-mass spectrometry and on-line gas chromatography-combustion/pyrolysis-isotope ratio mass spectrometry. *Food Chemistry*, 269-276.

Esteki, B., F., S., A., R., B., Z., A., E., D., . . . Y., V. H. (2017). Classification and authentication of Iranian walnuts according to their geographical origin based on gas chromatographic fatty acid fingerprint analysis using pattern recognition methods. *Chemom. Intell. Lab. Syst.*, 251-258. doi:10.1016/j.chemolab.2017.10.014

Fernandes, A., Bras, N. F., Mateus, N., & Freitas, V. D. (2015). A study of anthocyanin self-association by NMR spectroscopy. *New Journal of Chemistry*, 2602-2611.

Figgis, B. N. (1966). *Introduction to Ligand Fields.* Krieger Pub Co.

Food Integrity Handbook. (2018). doi:https://doi.org/10.32741/fihb

G.A.S. Dortmund. (s.d.). *IMS Working Principle*. Tratto da GAS Dortmund: https://www.gas-dortmund.de/Technology/IMS-Working-Principle/1_388.html

Garrido-Delgado, R., Mercader-Trejo, F., Sielemann, S., De Bruyn, W., Arce, L., & Valcarcel, M. (2011). Direct classification of olive oils by using two types of ion mobility spectrometers. *Analytica Chimica Acta*, 108-115.

Giannetti, V., Boccacci Mariani, M., & Marini, F. (2017). Volatile fraction by HS-SPME/GC-MS and chemiometric modeling for traceability of apples cultivated in the Northeast Italy. *Food Control*, 215-221.

Gu, S., Zhang, J., Wang, J., Wang, X., & Du, D. (2021). Recent development of HS-GC-IMS technology in rapid and nondestructive. *Trends in Analytical Chemistry*, 116435.

Hatzakis, E. (2018). Nuclear Magnetic Resonance (NMR) Spectroscopy in Food Science: A Comprehensive Review. *Comprehensive Reviews in Food Science and Food Safety*, 189-220.

Heberger, K. (2008). Chapter 7 - Chemoinformatics—multivariate mathematical–statistical methods for data evaluation. In K. Vekey, A. Telekes, & A. Vertes, *Medical Applications of Mass Spectrometry* (p. 141-169). Elsevier Science.

Hornak, J. P. (2017). *The Basic of NMR*. Tratto da Rochester Institute of Technology: https://www.cis.rit.edu/htbooks/nmr/inside.htm

Jihong, W., Haijan, G., Lei, Z., Xiaojun, L., Fang, C., Zhenfu, W., & Xiaosong, H. (2007). Chemical compositional characterization of some apple cultivars. *Food Chemistry*, 88-93.

Kanu, A. B., & Hill Jr., H. H. (2008). Ion mobility spectrometry detection for gas chromatography. *Journal of Chromatography A*, 12-27.

Lee, L. C., Liong, C., & Jemain, A. A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst*, 3526-3539.

Malmendal, A., Amoresano, C., Trotta, R., Lauri, I., De Tito, S., Novellino, E., & Randazzo, A. (2011). NMR spectrometers as "magnetic tongues": Prediction of sensory descriptors in canned tomatoes. *Journal of Agricultural and Food Chemistry*, 10831-10838.

Manfredi, M., Robotti, E., Quasso, F., Mazzucco, E., Calabrese, G., & Marengo, E. (2018). Fast classification of hazelnut cultivars through portable infrared spectroscopy and chemometrics. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.*, 427-435.

Marcone, M. F., Wang, S., Albabish, W., Somnarain, D., Hill, A., & Nie, S. (2013). Diverse food-based applications of nuclear magnetic resonance (NMR) technology. *Food Research International*, 729-747.

Mark, H., & Workman, J. J. (2018). *Chemometrics in Spectroscopy.*

Montero, L., Herrero, M., Ibanez, E., & Cifuentes, A. (2013). Profiling phenolic compounds from different apple varieties using comprehensive two-dimensional liquid chromatography. *Journal of Chromatography*, 275-283.

Nishida, J., Shigeto, S., Yabumoto, S., & Hamaguchi, H. (2012). Anharmonic coupling of the CH-stretch and CH-bend vibrations of chloroform as studied by near-infrared electroabsorption spectroscopy. *The Journal of Chemical Physics*, 234501.

Noda, I., & Ozaki, Y. (2005). *Two-Dimensional Correlation Spectroscopy: Applications in Vibrational and Optical Spectroscopy*.

Oliveira, L. S., & Franca, A. S. (2011). Applications of near infrared spectroscopy (NIRS) in food quality evaluation. In D. A. Medina, & A. M. Laine, *Food Quality: Control, Analysis and Consumer Concerns.* Nova Science Publishers, Inc.

Ozaki, Y. (2012). Near-Infrared Spectroscopy—Its Versatility in Analytical Chemistry. *Analytical Sciences*, 545-563.

Ozaki, Y., Genkama, T., & Futami, Y. (2017). Near-Infrared Spectroscopy. In J. C. Lindon, G. E. Tranter, & D. W. Koppenaal, *Encyclopedia of Spectroscopy and Spectrometry* (p. 3584).

Ramakrishnan, V., & Luthria, D. L. (2017). Recent applications of NMR in food and dietary studies. *Journal of the Science of Food and Agriculture*, 33-42.

Ringner, M. (2008). What is principal component analysis? *Nature Biotechnology*, 303-304.

Rohman, A., Erwanto, Y., M., H. M., M., R., Aldawoud, M., T., & Galanakis, C. M. (2021). The application of DNA-based methods for authentication analysis: Examples in halal. *Food Authentication and Traceability*, 195-213.

Siesler, H. W., Ozaki, Y., Kawata, S., & Heise, H. M. (2002). Near-Infrared Spectroscopy. Principles, Instruments, Applications. *Journal of Chemometrics*, 636-638.

Spyros, A., & Dais, P. (2013). *NMR spectroscopy in food analysis.* Cambridge: Royal Society of Chemistry.

Strojnik, L., Stopar, M., Zlatic, E., Kokalj, D., Naglic Gril, M., Zenko, B., . . . Ogrinc, N. (2019). Authentication of key aroma compounds in apple using stable isotope approach. *Food Chemistry*, 766-773.

van der Sluis, A., Dekker, M., De Jager, A., De Jager, M. F., & Jongen, W. (2001). Activity and Cioncentration of Polyphenolic Antioxidants in Apple: Effect of Cultivar, Harvest, Year and Storage Conditions. *Journal of Agricultural and Food Chemistry*, 3606-3613.

Vandendriessche, T., Keulemans, J., Geeraerd, A., Nicolai, B. M., & Hertog, M. L. (2012). Evaluation of fast volatile analysis for detection of Botrytis cinerea infections in strawberry. *Food Microbiology*, 406-414.

Varmuza, K., & Filzmoser, P. (2009). *Introduction to Multivariate Statistical Analysis in Chemometrics.* CRC Press.

Wang, S., Chen, H., & Sun, B. (2020). Recent progress in food flavor analysis using gas chromatography–ion mobility spectrometry (GC–IMS). *Food Chemistry*, 126158.

Aim of the thesis

The general goal of the present thesis was the development and set up of innovative non-targeted methodologies, aimed at facing authenticity issues geographically origin-related, on specific raw materials, with both rapid and confirmatory strategies, merged with chemometrics tools, such as multivariate statistical analysis and data fusion approach. Simultaneously, a similar approach was also applied in the sensomics field, trying to define a practical method to reliably authenticate the commodities of interest, also identifying some active molecules responsible for consumer overall liking evaluation on specific food products.

The Ph.D. thesis is structured as follows:

- NIR spectroscopy application on hazelnut samples for the authentication of Italian products (Chapter 1).
- GC-IMS technology was applied to hazelnut and dehydrated apple samples. The study on hazelnuts provides data elaboration using two different platforms, as well as coupling with sensory analysis data. The study on the dehydrated apples provides the development and the validation of a method to finely discern between Italian and non-Italian samples (Chapter 2)
- IRMS technique for the geographical origin evaluation of hazelnut samples. A data fusion approach was exploited also in this section, as IRMS data were merged with ICP-MS and ICP-OES ones (Chapter 3).
- LC-HRMS instrumentation was employed for the confirmatory study about the geographical provenience of dehydrated apple samples (Chapter 4)
- Other two analytical methodologies, ASAP-MS and NMR spectroscopy, present preliminary and promising results about both hazelnut and apple samples' authenticity (Chapter 5).

CHAPTER 1

NIR SPECTROSCOPY

NIR Spectroscopy

General Overview

NIR spectroscopy is a common analytical technique, widely employed at the academic level, and it is also promising as a tool for food companies. Its growing popularity is due to the low costs, limited or even inexistent sample preparation, non-destructiveness, and rapidity. All these characteristics make the technology useful for the online control made by the quality control departments. Further, it can be defined as an environmental-friendly technique, as it does not require toxic solvents or any particular pollutant consumables (Porep, Kammerer, & Carle, 2015). NIR spectroscopy is a potential process analyser, that could be successfully integrated into the process analytical technology (PAT) approach, due to the capability of fingerprinting materials and studying diverse physicochemical occurrences throughout the processing chain. Among the advantages that NIR spectroscopy can provide to the food industries, there are the evaluation of raw materials, time optimisation in the process cycle, the substitution of slow, expensive, and non-green lab procedures, as well as the possibility to deeply study the process, considering eventual product innovation (Grassi & Casiraghi, 2022). The NIR spectroscopy technique is extensively adopted for performing quantitative analyses of chemical constituents. However, it presents a growing number of applications as qualitative analysis tools, such as sample definition, authenticity studies, sensory evaluation, rheological and technological features, and physical characteristics. For industrial applications, several design constructions of NIR instruments have been developed, for hyperspectral imaging, fibre optical, portable devices, direct contact and tube-integrated probes, and automated sample cell loading (Porep, Kammerer, & Carle, 2015). Despite the recent progress in bringing the NIR technology to the industrial level, other implementations seem to be necessary for transferring knowledge about NIR spectroscopy and the multivariate analysis approach, to deal with its data, to the next food technologists (Grassi & Alamprese, Advances in NIR spectroscopy applied to process analytical technology in food industries, 2018).

The goal of the work

In the present chapter, an application of the NIR spectroscopy was evaluated for the authentication of Italian hazelnut samples, following the production chain from the fresh material to the roasted and the pasted ones.

This chapter was submitted to "*Vibrational Spectroscopy*". For additional details see the section "Author".

References

Grassi, S., & Alamprese, C. (2018). Advances in NIR spectroscopy applied to process analytical technology in food industries. *Current Opinion in Food Science*, 17-21.

Grassi, S., & Casiraghi, E. (2022). Advances in NIR Spectroscopy Analytical Technology in Food Industries. *Foods*, 1250.

Porep, J. U., Kammerer, D. R., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science & Technology*, 211-230.

Near Infrared Spectroscopy and Multivariate Statistical Analysis as Rapid Tools for the Geographical Origin Assessment of Italian Hazelnuts

Giuseppe Sammarco[1,3], Chiara Dall'Asta[3], Michele Suman[1,2]

[1] Advanced Laboratory Research, Barilla G. e R. Fratelli S.p.A., Parma, Italy
[2] Department for Sustainable Food Process, Catholic University Sacred Heart, Piacenza, Italy
[3] Department of Food and Drug, University of Parma, Parma, Italy

Highlights

- The geographical origin could represent a guarantee of quality for some food matrices
- Food Authenticity studies could prevent or fight false origin declaration
- Italian hazelnut variates are sensible to the authenticity fraud
- Near Infrared spectroscopy could be a rapid and effective solution for the origin assessment

Abstract

The geographical origin assessment of Italian hazelnuts is nowadays a relevant topic, aimed at the protection of provenience certificates. Near Infrared (NIR) spectroscopy could be a functional candidate for preventing and fighting illegal activities related to this matrix. The present study focuses on the exploitability of the NIR technique on the 'hazelnut chain' (fresh, roasted and paste), against the false origin declaration frauds, mainly concerning some of the best Italian varieties ('Nocciola Piemonte', 'Tonda Gentile Romana', 'Mortarella'). 216 spectra were recorded, for a total of n=144 for the training set, and n=72 for the validation set, considering fresh (n=57), roasted (n=107), and paste (n=52) hazelnuts as different matrices. The training set sample selection was made according to a Design of Experiment (DoE), that considered diverse factors, such as harvesting year, storage shelf life, and presence of peel. The validation set was composed of blended samples generated by mixing Italian and non-Italian ones, and real samples bought from local markets. Multivariate Statistical Analysis was employed for data handling and elaboration, both unsupervised and supervised models, Principal Component Analysis, and Partial Least Square-Discriminant Analysis were built to simplify, observe, and classify the samples. A variables selection was performed by filtering the most important ones considering the Variable Importance in Projection (VIP) scores. The predictive ability of the technology was evaluated by applying Classification List and Confusion Matrix approaches to a prediction set, providing a fit of the observations of this set into the selected

supervised model. The outcomes highlight valuable discrimination between authentic samples (related to two different harvesting year campaigns) with classification accuracy rates between 89 and 100%. Promising results about the application on blended and real samples were also obtained, especially as regards fresh and roasted hazelnuts, which presented classification accuracy rates of 81 and 91%. Therefore, this analytical technique could play a strategic role in the geographical origin assessment considering it is a rapid, direct, non-destructive, and cost-effective approach.

Keywords

NIR Spectroscopy, Hazelnut Chain, Multivariate Statistical Analysis, Geographical Origin.

INTRODUCTION

Hazelnut (*Corylus avellana L.*) is nowadays a valuable commodity for both bakery/chocolate industries and as a packed snack. This dried fruit presents health benefits, like the risk reduction of cancer and cardiovascular disease development, thanks to the content of *omega-3* fatty acids, manganese superoxide dismutase, vitamin E, and proanthocyanidins [1]. Various scientific studies focused on the geographical origin of hazelnuts, to preserve the Protected Geographical Indication (PGI) and the Protected Designation of Origin (PDO) [2] [3] [4] [5]. Indeed, a growing request for precise geographical provenience of several food matrices is occurring, as food authentication represents a fundamental aspect for farmers and producers, whereas, for consumers, the origin assures quality, as well as organoleptic and nutritional characteristics [6]. A false declaration of origin could seriously affect regional and national economies; thus, food authentication turns relevant in preventing illegal activities [7]. Italian hazelnuts represent the excellence of the territory, and their quality is confirmed by the abovementioned PDO and PGI certification. In particular, 'Tonda Gentile delle Langhe' in the Piedmont region gained the PGI in 1996, under the name of 'Nocciola Piemonte' [2] as well as 'Tonda di Giffoni' from Campania, while 'Tonda Gentile Romana', from Latium, has the PDO. Unfortunately, these products are often subject to authenticity fraud, due to their costs and the amount available. In 2020, more than 62% of hazelnuts were produced in Turkey, whereas Italy was the second country, with ca. 13% of world production. Other high-ranked countries were the United States (6%), Azerbaijan (4.6%), and Georgia (3%) [8]. The abundance of Turkish products and the lower prices of Azerbaijani and Georgian push dishonest producers/suppliers to perpetrate a false origin declaration, by adding hazelnuts from these locations, or even by substituting Italian ones with them [9]. Different analytical solutions have been approached to fight/prevent this illegal activity: ^{1}H Nuclear Magnetic Resonance (NMR), a robust and high-res technique, allowed to gain relevant accuracy of a discrimination model for the analyses of polar metabolites, considering five

countries and four harvesting years [10]. An interesting number of papers is about untargeted and targeted metabolomic strategies for hazelnuts' authenticity, Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-ToF) Mass Spectrometry (MS) analysis of kernels peptides/proteins was exploited for assessment of diverse cultivars, together with the evaluation of agronomical and morphological features, and kernel polyphenols characterization [11]. Cerulli *et al.* provided a detailed investigation of the phenolic content of both fresh and roasted hazelnuts from Giffoni ('Tonda di Giffoni', Italy), using Liquid Chromatography-Electrospray Ionization Linear Ion-Trap-Orbitrap Tandem MS (LC-ESI/LTQ-Orbitrap/MS/MS^n) [3]. Another work assessed the geographical origin and cultivars of hazelnuts through Ultra High-Pressure LC coupled to Quadrupole ToF MS (UHPLC-ESI/QToF-MS) and multivariate statistical analysis [12]. The Automatic untargeted Metabolic Profiling Analysis with Chemometrics (AuMPAC) approach was also useful to screen the Total Ion Chromatogram (TIC) peaks and revealing potential differences between hazelnut samples from precise locations [13]. Klockmann, in two diverse studies, worked on both untargeted and targeted solutions for geographical evaluation. Initially, via UHPLC-QToF-MS, 196 authentic samples were distinguished, they were from five countries and harvested in two years. A training set classification accuracy of 99.5% was reached by combining two classification models, Soft Independent Modeling of Class Analogy (SIMCA) and Support Vector Machine (SVM) [14]. Afterward, with an LC-ESI-Triple Quadrupole-MS (LC-ESI-QqQ-MS) it was possible to identify and quantify 20 non-polar key metabolites for the geographical discrimination [15].

Near Infrared (NIR) spectroscopy is one of the rapid techniques able to prevent/combat food frauds, as NIR radiation can superficially penetrate the food materials (ca. 2-3 mm). Thus, food samples having small dimensions, or previously milled, can be directly analysed, with limited or even inexistent preparation [16] [17]. The analytical strategy has been already tested for authenticity studies on hazelnuts. In particular, a work focused on the authenticity of the Italian PDO hazelnut 'Nocciola Romana' highlights how NIR spectroscopy, merged with two different classification approaches, SIMCA and Partial Least Squares-Discriminant Analysis (PLS-DA), led to high prediction capability of the examined samples (ca. 92% of classification accuracy). The instrument used was an FT-NIR, supplied with a halogen-tungsten lamp, and an indium gallium arsenide detector, working in the 4000-10000 cm^{-1} spectral range, with a nominal resolution of 4 cm^{-1}. The analyses were performed on the whole fruit (unshelled nut) [4]. The same Italian PDO hazelnut was assessed through NIR spectroscopy, using the SVM-DA routine. In particular, an Acousto-Optic Tunable Filter-NIR (AOTF-NIR) was employed, equipped with a reflectance post-dispersive optical disposition, and an indium gallium arsenide detector. The spectral range was from 1100 to 2300 nm, with a resolution of 2 nm. The analyses were carried out on the entire unshelled samples, acquiring

two spectra on both the kernel's sides, then averaged. The specificities, sensitivities, and accuracies obtained were about 96.0, 95.0, and 95.5%, respectively [5].

These valuable results demonstrate the effectiveness of the abovementioned strategy, even though they did not consider other Italian varieties, such as 'Nocciola Piemonte' or 'Tonda di Giffoni', as well as they only approached raw/fresh hazelnuts, without extending the attention to the other matrices of the chain (roasted hazelnuts, paste hazelnuts…). This study aimed at exploiting NIR spectroscopy, coupled with both unsupervised and supervised statistical models, for evaluating the geographical origin of the main Italian cultivars, from the field (fresh material) to the consumer (hazelnut-derived products).

MATERIALS AND METHODS

Sampling

Raw, roasted, peeled roasted, and paste hazelnuts were sampled from different Italian and not Italian regions, considering both the 2020 and the 2021 harvesting campaigns, and different storage shelf-lives, short and long for 2020. Short shelf-life samples were processed and analysed right after the harvesting, whereas the long shelf-life ones were stored for ca. 6 months (4-8°C, controlled atmosphere storage) before being processed and analysed. This could have influenced their chemical profile, for this reason the storage time was considered as a sampling factor. The Italian samples include PGI "Tonda Gentile delle Langhe" from Piedmont, PDO "Nocciola Romana" from Lazio, and "Mortarella" from Campania. For each matrix, these three varieties were mixed to have 'Italian samples' (N=36, 9 raw, 9 roasted, 9 peeled roasted, and 9 paste of hazelnuts) The same number of lots, considering the same factors, were from Turkey, Azerbaijan, and Georgia, for a total of 108 not Italian samples. The authenticity of both Italian and non-Italian samples was guaranteed by a trusted Barilla Company supplier. All the samples were stored in a cold room, with a controlled temperature of 4-6 °C. These samples were employed to create mixes, for each product the Italian hazelnuts were randomly mixed with one of the non-Italian, at different percentages of adulteration (10, 20, 50, 70, 90%). Fresh samples were composed of Italian and Georgian hazelnuts, the roasted ones presented Italian and Azerbaijani matrices, while the pastes were prepared by mixing Italian and Turkish samples. Different geographical origins were picked to create the blended samples to confirm the consistency of the model and the technique, which were not related to specific differences between a precise area and Italy. These mixes, together with other samples purchased from the local markets, were analysed as a validation set. The products, bought at the markets, were named 'real samples' since they were from a real environment, directly taken from the shelves and analysed. Tables S1 and

S2 (Supplementary Materials) list, respectively, all the training and validation set samples employed for the present study.

NIR Spectroscopy

The analyses were carried out through the NIRS^{TM} DS2500 (FOSS Analytics, Hilleroed, Denmark). Fresh and roasted samples were previously minced with Grindomix GM 200 (Retsch GmbH, Haan, Germany), at 7000 rpm for 15 seconds, whereas paste hazelnuts were directly analysed without any pre-treatment. The milling step was adopted to homogenise the sample, in order to avoid significative deviations in the measurements. Approximately 50 g of each sample was placed in a small (fresh and roasted hazelnuts) or slurry (paste of hazelnuts) quartz cup for the analyses. NIR spectra were acquired in the 400-2500 nm wavelength range, in reflectance mode with a spectral resolution of 0.5 nm, at room temperature. The instrument was equipped with two detectors: a Silicon detector, for wavelengths from 400 to 1100 nm, and a Lead Sulfide one, for wavelengths from 1100 to 2500 nm. The number of data points for each spectrum is 4200, but the elaborated data matrix contained 280 data points since the WinISI software (FOSS Analytics, Hilleroed, Denmark) averaged the NIR data (number of columns averaged: 15), in order to reduce the number of columns to export as a .csv file to being processed. Further, the reduction of the data points found more compatibility with the number of samples per class (9 for fresh and paste, 18 for roasted hazelnuts). Each sample was analysed in duplicate, and each analysis provided 7 sub-samples, so the products were analysed at 7 different points, in order to avoid the sampling effect.

Sample preparation & analysis

The spectra of 216 samples were acquired, n=144 samples were used as the training set, to create a statistical model, whereas n=72 samples, mixed and real ones, were selected as the validation set to challenge the model previously built. Each product (fresh, roasted, and paste) was considered and studied individually. The absorbance values at different wavelengths were exported in an XLSX matrix using WinISI software (FOSS Analytics, Hilleroed, Denmark). The spectra were normalized by Standard Normal Variate (SNV), to eliminate slope variation and correct the scattering effects [18]. 2nd derivatives emphasize small spectral variations, favouring the differentiation among samples, and providing a better data interpretation, whilst the Savitzky-Golay filter reduces the high frequency of the noise, and it is fundamental to avoid interferences from the background, as a smoothing method [19]. After this data pre-treatment, multivariate statistical models were built to assess the differences among the classes (geographical origins). The statistical approach was carried out with SIMCA software (version 16.0, Umetrics, Umea, Sweden), starting with an unsupervised

Principal Component Analysis (PCA) to look for preliminary class separation. Afterward, a supervised Partial Least Square-Discriminant Analysis (PLS-DA) was used, giving class information to the model, and enhancing the geographical differentiation of the samples. Both training and validation sets were exploited to create a prediction set, the Classification List and Confusion Matrix displayed the fit of the observations of this set into the selected supervised model; the fit was assessed from the YPredPS value, which is the probability of fitting the models in question. Therefore, this classification model showed the study's robustness and the predictive ability of the NIR technique regarding the geographical authenticity of Italian hazelnuts.

RESULTS AND DISCUSSION

144 samples were analysed as a training set, considering the different products from the chain (fresh, roasted, and paste hazelnuts). Figure 1 shows an example of averaged NIR spectra of short shelf life, 2020 harvesting campaign, and fresh hazelnut samples labeled according to their origin.

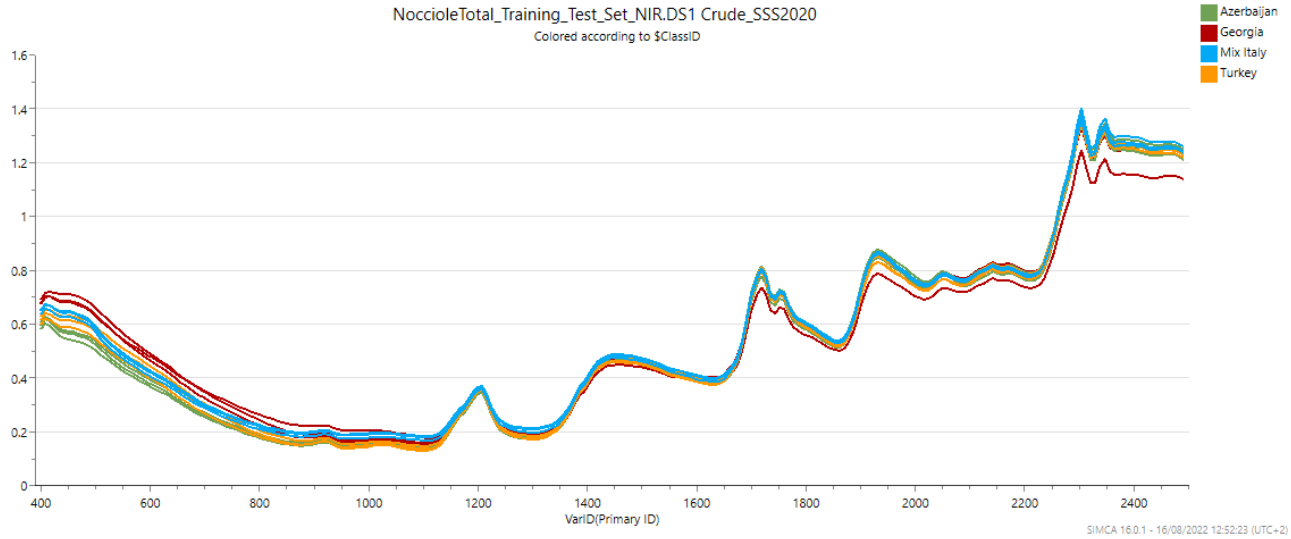


Figure 1. Averaged NIR spectra of short shelf-life samples harvested during the 2020 campaign. (Green spectrum: Azerbaijan, red spectrum: Georgia, turquoise spectrum: Mix Italy, yellow spectrum: Turkey)

A PCA model was built to preliminary summarise and visualise the data, considering each matrix individually. Both fresh and paste hazelnuts presented 36 samples, 9 from Georgia, 9 from Azerbaijan, 9 from Turkey, and 9 from Italy; roasted hazelnuts encompassed 72 samples, 36 with peeling and 36 without it, 18 from each abovementioned origin. Figures 2A-2B-2C show how the unsupervised PCA drove the clustering of mixed Italian samples.

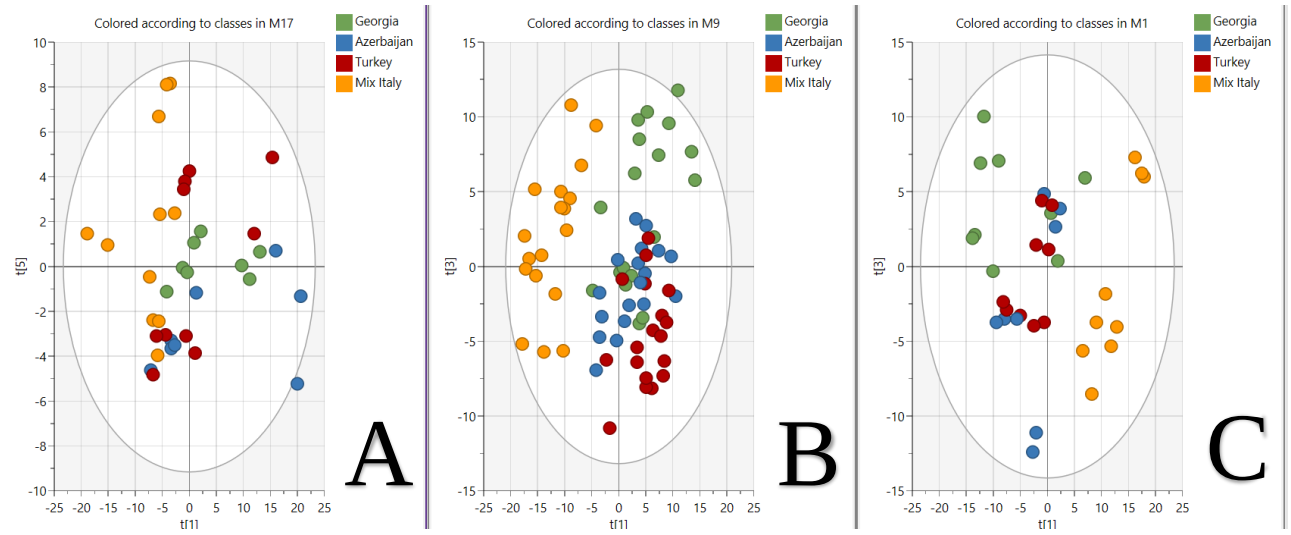


Figure 2. A) PCA score plot of hazelnut pastes NIR data. B) PCA score plot of roasted hazelnuts NIR data. C) PCA score plot of fresh hazelnuts NIR data. (Green dots: Georgia, blue dots: Azerbaijan, red dots: Turkey, orange dots: Italy)

All the score plots highlight a clear separation between Italian and not Italian samples, but the clustering among the latter ones is not ideal. This could be partially due to the similar harvesting area of the Georgian, Azerbaijani, and Turkish samples, but also the number of samples related to the classes. Indeed, the separation of not Italian roasted samples (B) is considerably better, as this matrix contains more samples per class than the others. However, these preliminary results can be defined as promising, as an unsupervised model. These outcomes were confirmed by the supervised PLS-DA model, as a classification approach. Figures 3A-3B-3C illustrate the score plots, with a clearer class clustering.

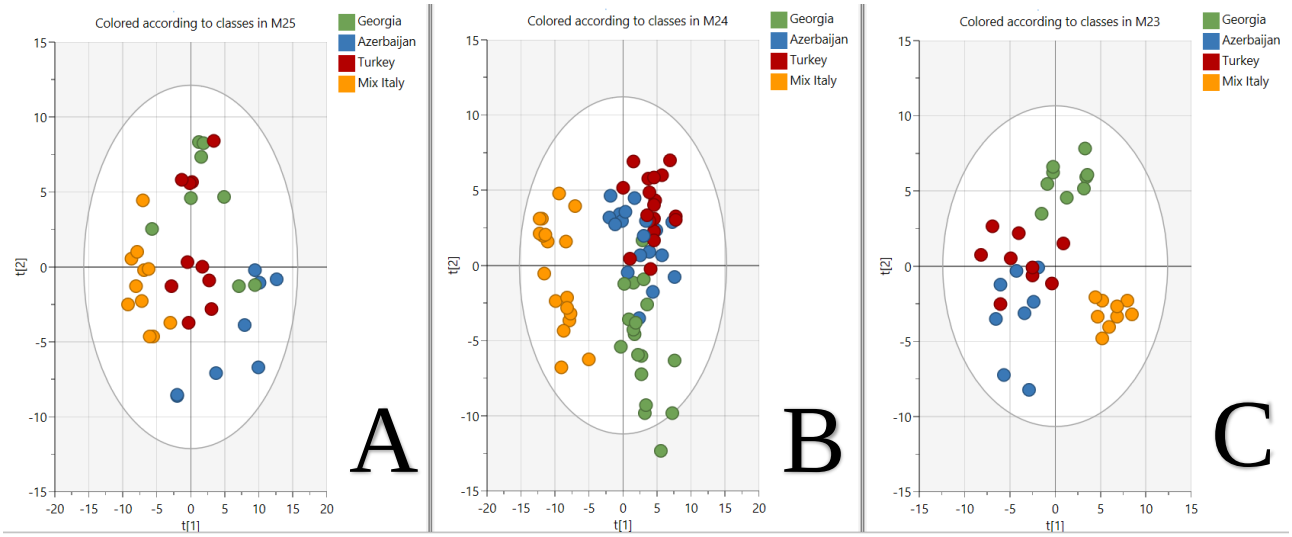


Figure 3. A) PLS-DA score plot of hazelnut pastes NIR data. B) PLS-DA score plot of roasted hazelnuts NIR data. C) PLS-DA score plot of fresh hazelnuts NIR data. (Green dots: Georgia, blue dots: Azerbaijan, red dots: Turkey, orange dots: Italy)

In order to filter among the most important variables that allowed class discrimination, the Variable Importance in Projection (VIP) scores were evaluated. They estimate the importance of each variable

used in a PLS model and they are often employed for variable selection [20]. In this study, a variable having a VIP score greater than 1 was considered important for the model, whereas the others were excluded. The most relevant absorbance regions, that define the Italian samples cluster, in the spectra presented absorption bands around 757 nm, 765.5 nm, 774 nm, 782.5 nm, 791 nm, 1020.5 nm, 1029 nm, 1037.5 nm, 1148 nm, 1360.5 nm, 1658 nm, 1666.5 nm, and 1675 nm. The bands between 750 and 800 nm were associated with the third overtone region, and they could underline the presence of the water because of the relation of this region with the O-H stretching, explaining so the moisture content. The region between 1000 and 1100 nm is at the beginning of the second overtone region, and it could be linked to the aromatic compound's presence, related to the polyphenols. The band at 1360.5 nm is in the same overtone region, and it is associated with -CH, CH_2, and $-CH_3$ groups, inherent in the lipid component. The 1650-1670 nm area has the same relation, and it is at the interface between the second and the first overtone region.

To verify the reliability of the statistical models built employing the training set, ca. 1/3 of the samples from this set were picked to generate test sets. Thus, 12 samples were considered for fresh and paste hazelnut samples, and 18 for the roasted ones. Confusion matrices were created to estimate the models' performances, they are listed as Table 1A-B-C. Excellent classification rate was achieved as regards the fresh hazelnut samples, while the models' performances get worse with the processed products. Concerning the roasted samples, as well as in the PLS-DA score plot, there is a misclassification between Azerbaijani and Turkish samples, and this influenced the multi-class confusion matrix, leading to 74% as correctness rate. Nonetheless, all the Italian samples were well-classified, and the one-class confusion matrix presented 100% of correctness score. About the hazelnut paste, both multi-class and one-class confusion matrices reported bad classification outcomes for the non-Italian samples, according to the PLS-DA score plot.

To confirm the training models robustness and the predictive ability of the technique, the validation set was also analysed. It contains mixtures of Italian and non-Italian samples, at different percentages of the latter ones, and real samples bought at the local markets, with precise origin declarations. For the blends, the harvesting year and storage shelf life were 'mixed', as all the DoE factors were considered during the preparation, in order to be as adherent as possible to the training set. Both training and validation sets were used to build a prediction set, that was aimed at estimating the predictive ability of the supervised model regarding authentic and non-authentic samples. For each type of matrix present in the prediction set, a new PLS-DA model was created. The Classification List for Discriminant Analysis models displays the observations, the original dummy variable in YVarPS (1 or 0), and the predicted dummy variable in YPredPS. Depending on a range of values, from the YPredPS is possible to define to which class a sample belongs:

- <0.35 the samples do not belong to the class
- Between 0.35 and 0.65 the samples are borderline
- >0.65 the samples belong to the class

These intervals were defined by the SIMCA Software employed for the data elaboration [21]. Tables S3A-B-C show how this approach led to the classification of the prediction set, according to the different matrices, highlighting the predictive ability of the NIR spectroscopy on this food authenticity application. The Classification List referred to the fresh hazelnuts worked on a prediction set that contained mixed Italian-Georgian samples, and real Turkish and Italian samples, as validation. Thus, in this set, also authentic Italian, Georgian and Turkish samples were included, from the training set. Authentic samples were correctly classified, except for a Georgian sample, which was borderline between the Georgian and the Turkish classes. The blended ones presented the highest YPredPS values for the Turkish class, even though none of these values were major than 0.65 (borderline). Only 90% of Georgian samples showed relevant values for the Georgian class. On the other side, in this circumstance real samples were weirdly classified, as the Italians had the highest values for the Turkish class. In particular, the Brand n.1 samples were borderline (YPredPS < 0.65), whereas the Brand n.2 ones were classified as Turkish. That could be explained by the eventual difference in the Italian variety analysed. Indeed, the authentic Italian samples constituted a mixture, in equal amounts, of Piedmont, Latium, and Campania hazelnuts. However, other varieties could have been used, since the label generically indicated 'Italy' as provenience. The valuable point is that all the mixtures were defined as "anomalous", no blended samples were indeed assigned to the Italian cluster, even the ones having 90% of Italian hazelnuts. The Confusion matrices, reported as Table 2A-B-C, visually summarise the models' performance. All the mixes not classified as Italian samples were considered correctly classified, as the main target is to detect Italian frauded products. As expected, remarkable classification accuracy rates were achieved for authentic samples (100-89%) from the training set, while interesting outcomes went out from the validation set samples (81%). A similar trend was observed for roasted hazelnut samples, where Italian and Azerbaijani mixtures and classes were evaluated. Only one authentic Italian sample was misclassified, and only one blended sample, with 90% of Italian peeled roasted hazelnuts, had a YPredPS value > 0.65, so Italian-classified. Concerning the real samples, none of them were assigned to the Italian group, as no one had the same composition as the authentic samples. One Piedmont sample presented an interesting value of the predicted variable, it could be explained by the fact that an aliquot of Italian samples contains hazelnuts from that region. The Confusion Matrix returned valuable results for both authentic and non-authentic sample evaluations, with accuracy rates of, respectively, 100-94% and 91%. Different outcomes were derived from the hazelnut paste samples prediction set, as blended samples having from 10 to 50%

Turkish aliquots were Italian classified, whereas 70 and 90% ones were defined as anomalous. This could be due to the sample heterogeneity, despite several sampling points during the analysis. This output led to a bad validation set classification (accuracy rate of 33%), despite the excellent accuracy of the training set (100%). After the indication of potential fraud risk from the present NIR approach regarding the Italian-labelled real samples, the Brands n.1 (fresh hazelnuts) and n.4 (hazelnut paste) have been subjected to other investigations with different analytical techniques, confirming the suspects (results not published yet).

CONCLUSION

This study particularly focused on the geographical authenticity of hazelnuts from the fresh product to the roasted (peeled or not), and hazelnut paste, very employed by the food companies for the production of chocolate, ice cream, candy, pastries, cakes, and the preparation of spreadable creams. NIR spectroscopy provided valuable discrimination of the authentic samples, related to two different harvesting year campaigns and with diverse storage shelf lives. It demonstrated to be also highly promising in a reliable evaluation of blended and real samples, especially concerning fresh and roasted products, whereas hazelnut pastes showed poor outputs, which underlined a low sensitivity of the technique on this specific matrix. Therefore, the results of this study highlight the possibility of a functional approach of NIR technology to the hazelnut chain food authenticity area. The ideal workflow for applying the methodology provides the construction of a big database, containing a remarkable number of authentic samples data, finely selected according to the varieties and the geographical origins of interest, in order to subsequently create a robust predictive model, that will be able to rapidly reveal whether an independent sample is authentic or not.

REFERENCES

[1] E. Alphan, M. Pala, F. Açkurt and T. Yilmaz, “NUTRITIONAL COMPOSITION OF HAZELNUTS AND ITS EFFECTS ON GLUCOSE AND LIPID METABOLISM,” *Acta Hortic,* pp. 305-310, 1997.

[2] M. Oddone, M. Aceto, M. Baldizzone, D. Musso and D. Osella, “Authentication and Traceability Study of Hazelnuts from Piedmont, Italy,” *Journal of Agricultural and Food Chemistry,* p. 3404–3408, 2009.

[3] A. Cerulli, A. Napolitano, M. Masullo, C. Pizza and S. Piacente, “LC-ESI/LTQOrbitrap/MS/MSn Analysis Reveals Diarylheptanoids and Flavonol O-glycosides in Fresh and Roasted Hazelnut (Corylus avellana cultivar “Tonda di Giffoni”),” *Natural Product Communications,* vol. 13, no. 9, pp. 1123-1126, 2018.

[4] A. Biancolillo, S. De Luca, S. Bassi, L. Roudier, R. Bucci, A. Magrì and F. Marini, “Authentication of an Italian PDO hazelnut ("Nocciola Romana") by NIR spectroscopy,” *Environmental Science and Pollution Research,* p. 28780–28786, 2018.

[5] R. Moscetti, E. Radicetti, D. Monarca, M. Cecchini and R. Massantini, “Near infrared spectroscopy is suitable for the classification of hazelnuts according to Protected Designation of Origin,” *Journal of the Science of Food and Agriculture,* p. 2619–2625, 2015.

[6] D. Bandoniene, T. Meisel, A. Rachetti and C. Walkner, “A tool to assure the geographical origin of local food products (glasshouse tomatoes) using labeling with rare earth elements,” *Journal of the Science of Food and Agriculture,* p. 4769–4777, 2018.

[7] G. Lo Feudo, A. Naccarato, G. Sindona and A. Tagarelli, “Investigating the origin of tomatoes and triple concentrated tomato pastes through multielement determination by inductively coupled plasma mass spectrometry and statistical analysis,” *Journal of Agricultural and Food Chemistry,* p. 3801–3807, 2010.

[8] Tridge, “Hazelnut Kernel,” 2020. [Online]. Available: https://www.tridge.com/intelligences/hazelnut-in-shell/production.

[9] C. Lang, N. Weber, M. Moller, L. Schramm, S. Schelm, O. Kohlbacher and M. Fischer, “Genetic authentication: Differentiation of hazelnut cultivars using polymorphic sites of the chloroplast genome,” *Food Control,* p. 108344, 2021.

[10] R. Bachmann, S. Klockmann, J. Haerdter, M. Fischer and T. Hackl, “1H NMR Spectroscopy for Determination of the Geographical Origin of Hazelnuts,” *Journal of Agricultural and Food Chemistry,* pp. 11873-11879, 2018.

[11] L. F. Ciarmiello, M. F. Mazzeo, P. Minasi, A. Peluso, A. De Luca, P. Piccirillo, R. A. Siciliano and V. Carbone, “Analysis of Different European Hazelnut (Corylus avellana L.) Cultivars: Authentication, Phenotypic Features, and Phenolic Profiles,” *Journal of Agricultural and Food Chemistry,* pp. 6236-6246, 2014.

[12] S. Ghisoni, L. Lucini, G. Rocchetti, G. Chiodelli, D. Farinelli, S. Tombesi and M. Trevisan, “Untargeted metabolomics with multivariate analysis to discriminate hazelnut (Corylus avellana L.) cultivars and their geographical origin,” *Journal of the Science of Food and Agriculture,* pp. 500-508, 2020.

[13] L. Han, Y. Zhang, J. Song, M. Fan, Y. Yu, P. Liu, Q. Zheng, Q. Chen, C. Bai, T. Sun and Y. She, “Automatic untargeted metabolic profiling analysis coupled withChemometrics for improving metabolite identification quality toenhance geographical origin discrimination capability,” *Journal of Chromatography A,* pp. 12-20, 2018.

[14] S. Klockmann, E. Reiner, R. Bachmann, T. Hackl and M. Fischer, “Food Fingerprinting: Metabolomic Approaches for Geographical Origin Discrimination of Hazelnuts (Corylus avellana) by UPLC-QTOF-MS,” *Journal of Agricultural and Food Chemistry,* pp. 9253-9262, 2016.

[15] S. Klockmann, E. Reiner, N. Cain and M. Fischer, “Food Targeting: Geographical Origin Determination of Hazelnuts (Corylus avellana) by LC-QqQ-MS/MS-Based Targeted Metabolomics Application,” *Journal of Agricultural and Food Chemistry,* pp. 1456-1465, 2017.

[16] P. Pallav, G. G. Diamond, D. A. Hutchins, R. J. Green and T. H. Gan, “A Near-Infrared (NIR) Technique for Imaging Food Materials,” *Journal of Food Science,* vol. 74, no. 1, pp. E23-E33, 2009.

[17] Metrohm NIRSystems, “A guide to near-infrared spectroscopic analysis of industrial manufacturing processes - Preface,” in *A guide to near-infrared spectroscopic analysis of industrial manufacturing processes*, Herisau, Switzerland, Metrohm AG, 2013, p. 5.

[18] A. Zaid, N. Abu-Khalaf, S. Mudalal and M. Petracci, “Differentiation between Normal and White Striped Turkey Breasts by Visible/Near Infrared Spectroscopy and Multivariate Data Analysis,” *Food Science of Animal Resources,* vol. 40, no. 1, pp. 96-105, 2019.

[19] G. Qiu, E. Lu, H. Lu, S. Xu, F. Zeng and Q. Shui, “Single-Kernel FT-NIR Spectroscopy for Detecting Supersweet Corn (Zea mays L. Saccharata Sturt) Seed Viability with Multivariate Data Analysis,” *Sensors,* vol. 18, no. 1010, p. 16, 2018.

[20] I. Chong and C. Jun, “Performance of some variable selection methods when multicollinearity is present,” *Chemometrics and Intelligent Laboratory Systems,* vol. 78, pp. 103-112, 2005.

[21] MKS Umetrics, “Dynacentrix,” 26 January 2015. [Online]. Available: https://www.dynacentrix.com/telecharg/SimcaP/SIMCA14_User_Guide.pdf.

TABLES

A)

	Georgia	Azerbaijan	Turkey	Italy	Outlier	Total
Georgia	3					100%
Azerbaijan		3				100%
Turkey			3			100%
Italy				3		100%
Total						100%

	Not Italy	Italy	Outlier	Total
Not Italy	9			100%
Italy		3		100%
Total				100%

B)

	Georgia	Azerbaijan	Turkey	Italy	Outlier	Total
Georgia	6					100%
Azerbaijan		0	6			0%
Turkey			6			100%
Italy				5		100%
Total						74%

	Not Italy	Italy	Outlier	Total
Not Italy	18			100%
Italy		5		100%
Total				100%

C)

	Georgia	Azerbaijan	Turkey	Italy	Outlier	Total
Georgia	1	1		1		33.33%
Azerbaijan		1	1	1		33.33%
Turkey	1	1	1			33.33%
Italy				3		100%
Total						50%

	Not Italy	Italy	Outlier	Total
Not Italy	4		5	44.44%
Italy		3		100%
Total				72%

Table 1. A) Fresh hazelnut samples prediction set (test set) multi-class (above) and one-class (below) confusion matrices. B) Roasted hazelnut samples prediction set (test set) multi-class (above) and one-class (below) confusion matrices. C) Hazelnut paste samples prediction set (test set) multi-class (above) and one-class (below) confusion matrices.

A)

	Italy	Georgia	Turkey	Validation	Outlier	Total
Italy	9					100%
Georgia		8			1	89%
Turkey			9			100%
Validation			2	18	2	81%

B)

	Italy	Azerbaijan	Validation	Outlier	Total
Italy	17			1	94%
Azerbaijan	18				100%
Validation	1	1	32	1	91%

C)

	Italy	Turkey	Validation	Outlier	Total
Italy	11				100%
Turkey		10			100%
Validation	6	1	4	1	33%

Table 2. A) Fresh hazelnut samples prediction set (training-validation sets) Confusion Matrix. B) Roasted hazelnut samples prediction set (training-validation sets) Confusion Matrix. C) Hazelnut paste samples prediction set (training-validation sets) Confusion Matrix.

SUPPLEMENTARY MATERIALS

ORIGIN	MATRIX	PEELING	HARVESTING YEAR	SHELF LIFE
GEORGIA	Fresh	/	2020	Short
GEORGIA	Fresh	/	2020	Short
GEORGIA	Fresh	/	2020	Short
AZERBAIJAN	Fresh	/	2020	Short
AZERBAIJAN	Fresh	/	2020	Short
AZERBAIJAN	Fresh	/	2020	Short
TURKEY	Fresh	/	2020	Short
TURKEY	Fresh	/	2020	Short
TURKEY	Fresh	/	2020	Short
ITALY	Fresh	/	2020	Short
ITALY	Fresh	/	2020	Short
ITALY	Fresh	/	2020	Short
GEORGIA	Roasted	Yes	2020	Short
GEORGIA	Roasted	Yes	2020	Short
GEORGIA	Roasted	Yes	2020	Short
AZERBAIJAN	Roasted	Yes	2020	Short
AZERBAIJAN	Roasted	Yes	2020	Short
AZERBAIJAN	Roasted	Yes	2020	Short
TURKEY	Roasted	Yes	2020	Short
TURKEY	Roasted	Yes	2020	Short
TURKEY	Roasted	Yes	2020	Short
ITALY	Roasted	Yes	2020	Short
ITALY	Roasted	Yes	2020	Short
ITALY	Roasted	Yes	2020	Short
GEORGIA	Roasted	No	2020	Short
GEORGIA	Roasted	No	2020	Short
GEORGIA	Roasted	No	2020	Short
AZERBAIJAN	Roasted	No	2020	Short
AZERBAIJAN	Roasted	No	2020	Short
AZERBAIJAN	Roasted	No	2020	Short
TURKEY	Roasted	No	2020	Short

TURKEY	Roasted	No	2020	Short
TURKEY	Roasted	No	2020	Short
ITALY	Roasted	No	2020	Short
ITALY	Roasted	No	2020	Short
ITALY	Roasted	No	2020	Short
GEORGIA	Paste	/	2020	Short
GEORGIA	Paste	/	2020	Short
GEORGIA	Paste	/	2020	Short
AZERBAIJAN	Paste	/	2020	Short
AZERBAIJAN	Paste	/	2020	Short
AZERBAIJAN	Paste	/	2020	Short
TURKEY	Paste	/	2020	Short
TURKEY	Paste	/	2020	Short
TURKEY	Paste	/	2020	Short
ITALY	Paste	/	2020	Short
ITALY	Paste	/	2020	Short
ITALY	Paste	/	2020	Short
GEORGIA	Fresh	/	2020	Long
GEORGIA	Fresh	/	2020	Long
GEORGIA	Fresh	/	2020	Long
AZERBAIJAN	Fresh	/	2020	Long
AZERBAIJAN	Fresh	/	2020	Long
AZERBAIJAN	Fresh	/	2020	Long
TURKEY	Fresh	/	2020	Long
TURKEY	Fresh	/	2020	Long
TURKEY	Fresh	/	2020	Long
ITALY	Fresh	/	2020	Long
ITALY	Fresh	/	2020	Long
ITALY	Fresh	/	2020	Long
GEORGIA	Roasted	Yes	2020	Long
GEORGIA	Roasted	Yes	2020	Long
GEORGIA	Roasted	Yes	2020	Long
AZERBAIJAN	Roasted	Yes	2020	Long
AZERBAIJAN	Roasted	Yes	2020	Long

AZERBAIJAN	Roasted	Yes	2020	Long
TURKEY	Roasted	Yes	2020	Long
TURKEY	Roasted	Yes	2020	Long
TURKEY	Roasted	Yes	2020	Long
ITALY	Roasted	Yes	2020	Long
ITALY	Roasted	Yes	2020	Long
ITALY	Roasted	Yes	2020	Long
GEORGIA	Roasted	No	2020	Long
GEORGIA	Roasted	No	2020	Long
GEORGIA	Roasted	No	2020	Long
AZERBAIJAN	Roasted	No	2020	Long
AZERBAIJAN	Roasted	No	2020	Long
AZERBAIJAN	Roasted	No	2020	Long
TURKEY	Roasted	No	2020	Long
TURKEY	Roasted	No	2020	Long
TURKEY	Roasted	No	2020	Long
ITALY	Roasted	No	2020	Long
ITALY	Roasted	No	2020	Long
ITALY	Roasted	No	2020	Long
GEORGIA	Paste	/	2020	Long
GEORGIA	Paste	/	2020	Long
GEORGIA	Paste	/	2020	Long
AZERBAIJAN	Paste	/	2020	Long
AZERBAIJAN	Paste	/	2020	Long
AZERBAIJAN	Paste	/	2020	Long
TURKEY	Paste	/	2020	Long
TURKEY	Paste	/	2020	Long
TURKEY	Paste	/	2020	Long
ITALY	Paste	/	2020	Long
ITALY	Paste	/	2020	Long
ITALY	Paste	/	2020	Long
GEORGIA	Fresh	/	2021	Short
GEORGIA	Fresh	/	2021	Short
GEORGIA	Fresh	/	2021	Short

AZERBAIJAN	Fresh	/	2021	Short
AZERBAIJAN	Fresh	/	2021	Short
AZERBAIJAN	Fresh	/	2021	Short
TURKEY	Fresh	/	2021	Short
TURKEY	Fresh	/	2021	Short
TURKEY	Fresh	/	2021	Short
ITALY	Fresh	/	2021	Short
ITALY	Fresh	/	2021	Short
ITALY	Fresh	/	2021	Short
GEORGIA	Roasted	Yes	2021	Short
GEORGIA	Roasted	Yes	2021	Short
GEORGIA	Roasted	Yes	2021	Short
AZERBAIJAN	Roasted	Yes	2021	Short
AZERBAIJAN	Roasted	Yes	2021	Short
AZERBAIJAN	Roasted	Yes	2021	Short
TURKEY	Roasted	Yes	2021	Short
TURKEY	Roasted	Yes	2021	Short
TURKEY	Roasted	Yes	2021	Short
ITALY	Roasted	Yes	2021	Short
ITALY	Roasted	Yes	2021	Short
ITALY	Roasted	Yes	2021	Short
GEORGIA	Roasted	No	2021	Short
GEORGIA	Roasted	No	2021	Short
GEORGIA	Roasted	No	2021	Short
AZERBAIJAN	Roasted	No	2021	Short
AZERBAIJAN	Roasted	No	2021	Short
AZERBAIJAN	Roasted	No	2021	Short
TURKEY	Roasted	No	2021	Short
TURKEY	Roasted	No	2021	Short
TURKEY	Roasted	No	2021	Short
ITALY	Roasted	No	2021	Short
ITALY	Roasted	No	2021	Short
ITALY	Roasted	No	2021	Short
GEORGIA	Paste	/	2021	Short

GEORGIA	Paste	/	2021	Short
GEORGIA	Paste	/	2021	Short
AZERBAIJAN	Paste	/	2021	Short
AZERBAIJAN	Paste	/	2021	Short
AZERBAIJAN	Paste	/	2021	Short
TURKEY	Paste	/	2021	Short
TURKEY	Paste	/	2021	Short
TURKEY	Paste	/	2021	Short
ITALY	Paste	/	2021	Short
ITALY	Paste	/	2021	Short
ITALY	Paste	/	2021	Short

Table S1. Training set samples list

ORIGIN	MATRIX	PEELING	HARVESTING YEAR	SHELF LIFE
90% ITA-10% GEO	Fresh	/	Mixed	Mixed
90% ITA-10% GEO	Fresh	/	Mixed	Mixed
90% ITA-10% GEO	Fresh	/	Mixed	Mixed
80% ITA-20% GEO	Fresh	/	Mixed	Mixed
80% ITA-20% GEO	Fresh	/	Mixed	Mixed
80% ITA-20% GEO	Fresh	/	Mixed	Mixed
50% ITA-50% GEO	Fresh	/	Mixed	Mixed
50% ITA-50% GEO	Fresh	/	Mixed	Mixed
50% ITA-50% GEO	Fresh	/	Mixed	Mixed
30% ITA-70% GEO	Fresh	/	Mixed	Mixed
30% ITA-70% GEO	Fresh	/	Mixed	Mixed
30% ITA-70% GEO	Fresh	/	Mixed	Mixed
10% ITA-90% GEO	Fresh	/	Mixed	Mixed
10% ITA-90% GEO	Fresh	/	Mixed	Mixed
10% ITA-90% GEO	Fresh	/	Mixed	Mixed
GERMANY (BRAND 1)	Fresh	/	Not Specified	Not Specified
GERMANY (BRAND 1)	Fresh	/	Not Specified	Not Specified
ITALY (BRAND 2)	Fresh	/	Not Specified	Not Specified
ITALY (BRAND 2)	Fresh	/	Not Specified	Not Specified

TURKEY (BRAND 3)	Fresh	/	Not Specified	Not Specified
ITALY (BRAND 4)	Fresh	/	Not Specified	Not Specified
90% ITA-10% AZE	Roasted	Yes	Mixed	Mixed
90% ITA-10% AZE	Roasted	Yes	Mixed	Mixed
90% ITA-10% AZE	Roasted	Yes	Mixed	Mixed
80% ITA-20% AZE	Roasted	Yes	Mixed	Mixed
80% ITA-20% AZE	Roasted	Yes	Mixed	Mixed
80% ITA-20% AZE	Roasted	Yes	Mixed	Mixed
50% ITA-50% AZE	Roasted	Yes	Mixed	Mixed
50% ITA-50% AZE	Roasted	Yes	Mixed	Mixed
50% ITA-50% AZE	Roasted	Yes	Mixed	Mixed
30% ITA-70% AZE	Roasted	Yes	Mixed	Mixed
30% ITA-70% AZE	Roasted	Yes	Mixed	Mixed
30% ITA-70% AZE	Roasted	Yes	Mixed	Mixed
10% ITA-90% AZE	Roasted	Yes	Mixed	Mixed
10% ITA-90% AZE	Roasted	Yes	Mixed	Mixed
10% ITA-90% AZE	Roasted	Yes	Mixed	Mixed
90% ITA-10% AZE	Roasted	No	Mixed	Mixed
90% ITA-10% AZE	Roasted	No	Mixed	Mixed
90% ITA-10% AZE	Roasted	No	Mixed	Mixed
80% ITA-20% AZE	Roasted	No	Mixed	Mixed
80% ITA-20% AZE	Roasted	No	Mixed	Mixed
80% ITA-20% AZE	Roasted	No	Mixed	Mixed
50% ITA-50% AZE	Roasted	No	Mixed	Mixed
50% ITA-50% AZE	Roasted	No	Mixed	Mixed
50% ITA-50% AZE	Roasted	No	Mixed	Mixed
30% ITA-70% AZE	Roasted	No	Mixed	Mixed
30% ITA-70% AZE	Roasted	No	Mixed	Mixed
30% ITA-70% AZE	Roasted	No	Mixed	Mixed
10% ITA-90% AZE	Roasted	No	Mixed	Mixed
10% ITA-90% AZE	Roasted	No	Mixed	Mixed
10% ITA-90% AZE	Roasted	No	Mixed	Mixed
PIEDMONT-ITALY (BRAND 1)	Roasted	No	2020	Not Specified

PIEDMONT-ITALY (BRAND 1)	Roasted	No	2020	Not Specified
NOT SPECIFIED (BRAND 2)	Roasted	No	Not Specified	Not Specified
NOT SPECIFIED (BRAND 2)	Roasted	No	Not Specified	Not Specified
TURKEY (BRAND 3)	Roasted	No	Not Specified	Not Specified
90% ITA-10% TUR	Paste	/	Mixed	Mixed
90% ITA-10% TUR	Paste	/	Mixed	Mixed
90% ITA-10% TUR	Paste	/	Mixed	Mixed
80% ITA-20% TUR	Paste	/	Mixed	Mixed
80% ITA-20% TUR	Paste	/	Mixed	Mixed
80% ITA-20% TUR	Paste	/	Mixed	Mixed
50% ITA-50% TUR	Paste	/	Mixed	Mixed
50% ITA-50% TUR	Paste	/	Mixed	Mixed
50% ITA-50% TUR	Paste	/	Mixed	Mixed
30% ITA-70% TUR	Paste	/	Mixed	Mixed
30% ITA-70% TUR	Paste	/	Mixed	Mixed
30% ITA-70% TUR	Paste	/	Mixed	Mixed
10% ITA-90% TUR	Paste	/	Mixed	Mixed
10% ITA-90% TUR	Paste	/	Mixed	Mixed
10% ITA-90% TUR	Paste	/	Mixed	Mixed
ITALY (BRAND 1)	Paste	/	Not Specified	Not Specified

Table S2. Validation set samples list (GEO: Georgia, AZE: Azerbaijan, TUR: Turkey, ITA: Italy)

A)

PRIMARY ID	CLASSID	YVARPS (GEO)	YPREDPS (GEO)	YVARPS (TUR)	YPREDPS (TUR)	YVARPS (ITA)	YPREDPS (ITA)
126457 1	Mix Italy	0	0.103024	0	-0.220251	1	1.11723
126457 2	Mix Italy	0	0.141049	0	-0.191527	1	1.05048
126457 3	Mix Italy	0	-0.0522517	0	0.0454415	1	1.00681
156558 1	Mix Italy	0	0.152813	0	0.129595	1	0.717592
156558 2	Mix Italy	0	0.0676003	0	0.0818914	1	0.850508
156558 3	Mix Italy	0	-0.0881445	0	0.276094	1	0.812051
162337 1	Mix Italy	0	0.0555349	0	0.198459	1	0.746006
162337 2	Mix Italy	0	0.0827885	0	0.0035476	1	0.913664
162337 3	Mix Italy	0	-0.041297	0	0.198789	1	0.842508

125592 1	Georgia	1	0.651518	0	0.118009	0	0.230473
125592 2	Georgia	1	0.786136	0	-0.283087	0	0.496951
125592 3	Georgia	1	0.346715	0	0.370123	0	0.283162
156552 1	Georgia	1	1.09346	0	-0.0119724	0	-0.0814849
156552 2	Georgia	1	0.788581	0	0.206965	0	0.00445423
156552 3	Georgia	1	0.849925	0	0.358205	0	-0.20813
162331 1	Georgia	1	1.03219	0	0.161971	0	-0.194162
162331 2	Georgia	1	1.29426	0	-0.148282	0	-0.145973
162331 3	Georgia	1	0.98892	0	-0.0247449	0	0.0358245
125594 1	Turkey	0	-0.123576	1	1.09433	0	0.0292502
125594 2	Turkey	0	0.253148	1	0.507297	0	0.239555
125594 3	Turkey	0	-0.0312145	1	0.705516	0	0.325698
156554 1	Turkey	0	-0.206616	1	1.14214	0	0.0644712
156554 2	Turkey	0	0.160973	1	0.87367	0	-0.0346422
156554 3	Turkey	0	0.265316	1	0.757887	0	-0.0232033
162333 1	Turkey	0	0.255654	1	0.88234	0	-0.137993
162333 2	Turkey	0	0.209908	1	0.686136	0	0.103956
162333 3	Turkey	0	-0.0364053	1	1.08146	0	-0.0450507
10%GEO_A	Validation	0	0.336636	0	0.277404	0	0.385961
10%GEO_B	Validation	0	0.371335	0	0.280265	0	0.3484
10%GEO_C	Validation	0	0.336352	0	0.18587	0	0.477778
20%GEO_A	Validation	0	0.2334	0	0.522004	0	0.244596
20%GEO_B	Validation	0	0.261098	0	0.563111	0	0.175791
20%GEO_C	Validation	0	0.319641	0	0.506696	0	0.173663
50%GEO_A	Validation	0	0.381518	0	0.234594	0	0.383888
50%GEO_B	Validation	0	0.406656	0	0.468975	0	0.124369
50%GEO_C	Validation	0	0.427705	0	0.463563	0	0.108733
70%GEO_A	Validation	0	0.597565	0	0.47613	0	-0.0736949
70%GEO_B	Validation	0	0.496327	0	0.561745	0	-0.0580713
70%GEO_C	Validation	0	0.512618	0	0.567852	0	-0.0804704
90%GEO_A	Validation	0	0.668371	0	0.437721	0	-0.106092
90%GEO_B	Validation	0	0.599183	0	0.353958	0	0.0468592
90%GEO_C	Validation	0	0.704085	0	0.451214	0	-0.155298
GERMANY (BRAND 1)	Validation	0	-0.0970129	0	1.15502	0	-0.0580056
GERMANY (BRAND 1)	Validation	0	-0.184469	0	1.19857	0	-0.0141041
ITALY (BRAND 2)	Validation	0	0.152209	0	0.538396	0	0.309395
ITALY (BRAND 2)	Validation	0	0.205402	0	0.48907	0	0.305528
TURKEY (BRAND 3)	Validation	0	-0.173891	0	1.56458	0	-0.390687
ITALY (BRAND 4)	Validation	0	-0.391625	0	0.99283	0	0.398796
ITALY (BRAND 4)	Validation	0	-0.471111	0	1.11393	0	0.357184

B)

PRIMARY ID	CLASSID	YVARPS (AZE)	YPREDPS (AZE)	YVARPS (ITA)	YPREDPS (ITA)
126456 1	Mix Italy	0	-0.141404	1	1.1414
126456 2	Mix Italy	0	-0.183445	1	1.18344
126456 3	Mix Italy	0	0.177759	1	0.82224
126458 1	Mix Italy	0	0.000317007	1	0.999683
126458 2	Mix Italy	0	0.0509401	1	0.94906
126458 3	Mix Italy	0	0.0588798	1	0.94112
156565 1	Mix Italy	0	-0.00971645	1	1.00972
156565 2	Mix Italy	0	0.0231152	1	0.976885
156565 3	Mix Italy	0	0.0822551	1	0.917745
156572 1	Mix Italy	0	-0.112216	1	1.11222
156572 2	Mix Italy	0	-0.00376785	1	1.00377
156572 3	Mix Italy	0	-0.0112921	1	1.01129
162344 1	Mix Italy	0	0.0772374	1	0.922763
162344 2	Mix Italy	0	-0.00598621	1	1.00599
162344 3	Mix Italy	0	-0.0547754	1	1.05478
162351 1	Mix Italy	0	0.18767	1	0.81233
162351 2	Mix Italy	0	0.400086	1	0.599914
162351 3	Mix Italy	0	0.232718	1	0.767282
125599 1	Azerbaijan	1	1.08915	0	-0.08915
125599 2	Azerbaijan	1	0.70413	0	0.29587
125599 3	Azerbaijan	1	0.776734	0	0.223266
125605 1	Azerbaijan	1	0.823975	0	0.176025
125605 2	Azerbaijan	1	0.990022	0	0.00997758
125605 3	Azerbaijan	1	1.25296	0	-0.252959
156560 1	Azerbaijan	1	0.931748	0	0.0682524
156560 2	Azerbaijan	1	1.18838	0	-0.188382
156560 3	Azerbaijan	1	0.960782	0	0.0392181
156567 1	Azerbaijan	1	1.01091	0	-0.0109149
156567 2	Azerbaijan	1	1.06559	0	-0.065589
156567 3	Azerbaijan	1	1.00373	0	-0.00372696
162339 1	Azerbaijan	1	0.927163	0	0.0728369
162339 2	Azerbaijan	1	0.840961	0	0.159039
162339 3	Azerbaijan	1	1.02088	0	-0.020885
162346 1	Azerbaijan	1	0.806184	0	0.193816
162346 2	Azerbaijan	1	0.796085	0	0.203915
162346 3	Azerbaijan	1	1.04223	0	-0.0422344
10%AZE_A	Validation	0	0.583626	0	0.416374
10%AZE_B	Validation	0	0.541796	0	0.458204
10%AZE_C	Validation	0	0.458374	0	0.541626

20%AZE_A	Validation	0	0.662646	0	0.337354
20%AZE_B	Validation	0	0.686096	0	0.313904
20%AZE_C	Validation	0	0.498101	0	0.501899
50%AZE_A	Validation	0	0.69312	0	0.30688
50%AZE_B	Validation	0	0.562933	0	0.437067
50%AZE_C	Validation	0	0.582447	0	0.417553
70%AZE_A	Validation	0	0.878421	0	0.121579
70%AZE_B	Validation	0	0.728532	0	0.271468
70%AZE_C	Validation	0	0.595359	0	0.404641
90%AZE_A	Validation	0	0.85835	0	0.14165
90%AZE_B	Validation	0	0.794844	0	0.205156
90%AZE_C	Validation	0	0.64518	0	0.35482
PEELED_10%AZE_A	Validation	0	0.586851	0	0.413149
PEELED_10%AZE_B	Validation	0	0.411006	0	0.588994
PEELED_10%AZE_C	Validation	0	0.331082	0	0.668918
PEELED_20%AZE_A	Validation	0	0.582958	0	0.417042
PEELED_20%AZE_B	Validation	0	0.566964	0	0.433036
PEELED_20%AZE_C	Validation	0	0.498702	0	0.501298
PEELED_50%AZE_A	Validation	0	0.77965	0	0.22035
PEELED_50%AZE_B	Validation	0	0.921628	0	0.0783723
PEELED_50%AZE_C	Validation	0	0.582757	0	0.417243
PEELED_70%AZE_A	Validation	0	0.583322	0	0.416678
PEELED_70%AZE_B	Validation	0	0.805897	0	0.194103
PEELED_70%AZE_C	Validation	0	0.610351	0	0.389649
PEELED_90%AZE_A	Validation	0	0.398627	0	0.601373
PEELED_90%AZE_B	Validation	0	0.727767	0	0.272233
PEELED_90%AZE_C	Validation	0	0.429862	0	0.570138
PIEDMONT-ITALY (BRAND 1)	Validation	0	0.991534	0	0.00846633
PIEDMONT-ITALY (BRAND 1) / (BRAND 2)	Validation	0	0.607288	0	0.392712
	Validation	0	1.12066	0	-0.120659
TURKEY (BRAND 3)	Validation	0	1.17226	0	-0.172263

C)

PRIMARY ID	CLASSID	YVARPS (TUR)	YPREDPS (TUR)	YVARPS (ITA)	YPREDPS (ITA)
126459 1	Mix Italy	0	0.0943932	1	0.905607
126459 2	Mix Italy	0	0.0448929	1	0.955107
126459 3	Mix Italy	0	-0.0501643	1	1.05016

126459 4	Mix Italy	0	-0.0461408	1	1.04614
126459 5	Mix Italy	0	0.0331995	1	0.9668
156579 1	Mix Italy	0	-0.0301674	1	1.03017
156579 2	Mix Italy	0	-0.00736547	1	1.00737
156579 3	Mix Italy	0	0.0236831	1	0.976317
162358 1	Mix Italy	0	-0.000957817	1	1.00096
162358 2	Mix Italy	0	0.147697	1	0.852303
162358 3	Mix Italy	0	0.032282	1	0.967718
125612 1	Turkey	1	0.802459	0	0.197541
125612 2	Turkey	1	1.1854	0	-0.185402
125612 3	Turkey	1	0.988	0	0.0119998
125612 4	Turkey	1	0.948284	0	0.0517159
125612 5	Turkey	1	0.837845	0	0.162155
156575 2	Turkey	1	1.1308	0	-0.130802
156575 3	Turkey	1	1.16438	0	-0.164375
162354 1	Turkey	1	0.944865	0	0.0551346
162354 2	Turkey	1	0.980779	0	0.0192211
162354 3	Turkey	1	0.775836	0	0.224164
10%TUR_A	Validation	0	0.0192276	0	0.980772
10%TUR_B	Validation	0	-0.0550626	0	1.05506
20%TUR_A	Validation	0	0.0724492	0	0.927551
20%TUR_B	Validation	0	0.0534062	0	0.946594
50%TUR_A	Validation	0	0.321127	0	0.678873
50%TUR_B	Validation	0	0.30465	0	0.69535
70%TUR_A	Validation	0	0.564006	0	0.435994
70%TUR_B	Validation	0	0.5517	0	0.4483
90%TUR_A	Validation	0	0.546389	0	0.453611
90%TUR_B	Validation	0	0.446266	0	0.553734
ITALY (BRAND 1)	Validation	0	0.500045	0	0.499955
ITALY (BRAND 1)	Validation	0	0.74299	0	0.25701

Table S3. A) Fresh hazelnut samples prediction set (training-validation sets) Classification List. B) Roasted hazelnut samples prediction set (training-validation sets) Classification List. C) Hazelnut paste samples prediction set (training-validation sets) Classification List. The percentages indicate the amount of not Italian samples in the mixture, whereas letters and numbers after the samples code are referred to as different batches, not replicates. (Green YPredPS value: sample belongs to the class, yellow YPredPS value: sample is borderline, white YPredPS value: sample does not belong to the class) (Primary ID: observation name; Class ID: original observation class; YVarPS: original dummy variable prediction set; YPredPS: predicted dummy variable prediction set; GEO: Georgia; AZE: Azerbaijan; TUR: Turkey; ITA: Italy)

CHAPTER 2

GC-IMS ANALYSIS

GC-IMS Analysis

General Overview

GC-IMS technology figures an innovative approach, that is gaining increasing demand among food scientists from both academic and industrial sides. The capability of the technique to accurately characterise and detect volatile organic compounds (VOCs) in agri-food products, coupled with the need to follow their volatile profile change during the storage, processing and distribution, is leading this tool to an industrial interest. Food quality and safety could be ensured by the GC-IMS system, by classifying and detecting adulterants, monitoring flavour change, and detecting microbial contamination in agri-food products (Gu, Zhang, Wang, Wang, & Du, Recent development of HS-GC-IMS technology in rapid and non-destructive detection of quality and contamination in agri-food products, 2021). Among the other advantages of this technique, fast detection, simple operation and instrumental portability make it a convenient strategy for the chemical analysis of food flavour. Compared to the classic gas chromatography-mass spectrometry (GC-MS) approach, GC-IMS provides a high separation efficiency, rapidity in obtaining qualitative results employing a chromatographic library, and the atmospheric pressure as a working condition, instead of the vacuum, that facilitates further development of the instrument (portability, low costs, easy operation…) (He, et al., 2020). On the other hand, the GC-IMS system presents limitations regarding the qualitative analysis accuracy, the IMS response is non-linear, hence the concentration measurements, at ppbv and pptv, are difficult to achieve. Moreover, GC-IMS databases are not completed yet, so it could also be tough to accomplish a putative molecular identification (He, et al., 2020). Therefore, further efforts are needed to overcome the limitations of the technique, and this could favour major investments by enterprises in the technology.

The goal of the work

The following chapter discusses the GC-IMS applications on both hazelnut and apple commodities. Two rapid, qualitative, non-targeted approaches have been exploited for the geographical origin assessment of Italian products. The first study was submitted to "*Food Research International*", whereas the second was submitted to "*Food Analytical Methods*". For additional details see the section "Author".

References

Gu, S., Zhang, J., Wang, J., Wang, X., & Du, D. (2021). Recent development of HS-GC-IMS technology in rapid and non-destructive detection of quality and contamination in agri-food products. *TrAC Trends in Analytical Chemistry*, 116435.

He, W., Ren, F., Wang, Y., Gao, X., Wang, X., Dai, X., & Song, J. (2020). Application of GC-IMS in Detection of Food Flavor Substances. *2020 International Conference on Green Chemical and Environmental Science* (p. 012030). Seattle, Washington, US: IOP Publishing.

A Geographical Origin assessment of Italian Hazelnuts: Gas Chromatography-Ion mobility spectrometry coupled with Multivariate Statistical Analysis and Data Fusion approach

Giuseppe Sammarco[1,3], Daniele Bardin[1], Federica Quaini[1], Chiara Dall'Asta[3], Joscha Christmann[4], Philipp Weller[4], Michele Suman[1,2]

[1] Sensory and Analytical Food Science, Barilla G. e R. Fratelli S.p.A., Parma, Italy
[2] Department for Sustainable Food Process, Catholic University Sacred Heart, Piacenza, Italy
[3] Department of Food and Drug, University of Parma, Parma, Italy
[4] Institute of Analytics and Bioanalytics, Faculty of Biotechnology, Mannheim University of Applied Sciences, Mannheim, Germany

Abstract

Hazelnut is a commodity that has gained interest in the food science community concerning its authenticity. The quality of the Italian hazelnuts is guaranteed by Protected Designation of Origin and Protected Geographical Indication certificates. However, due to their modest availability and the high price, fraudulent producers/suppliers blend, or even substitute, Italian hazelnuts with others from different countries, having a lower price, and often a lower quality. To contrast or prevent these illegal activities, the present work investigated the application of the Gas Chromatography-Ion mobility spectrometry (GC-IMS) technique on the hazelnut chain (fresh, roasted, and paste of hazelnuts). The raw data obtained were handled and elaborated using two different ways, software for statistical analysis, and a programming language. In both cases, Principal Component Analysis and Partial Least Squares-Discriminant Analysis models were exploited, to study how the Volatile Organic Profiles of Italian, Turkish, Georgian, and Azerbaijani products differ. A prediction set was extrapolated from the training set, for a preliminary models' evaluation, then an external validation set, containing blended samples, was analysed. Both approaches highlighted an interesting class separation and good model parameters (accuracy, precision, sensitivity, specificity, F1-score). Moreover, a data fusion approach with a complementary methodology, sensory analysis, was achieved, to estimate the performance enhancement of the statistical models, considering more discriminant variables and integrating the same time further information correlated to quality aspects. GC-IMS could be a key player as a rapid, direct, cost-effective strategy to face authenticity issues regarding the hazelnut chain.

Keywords

Hazelnut (*Corylus avellana*), Gas Chromatography-Ion Mobility System, Food Authenticity, Multivariate Statistical Analysis, Sensory Analysis, Data Fusion

INTRODUCTION

Authentic food can be defined as "a product where there is a match between its actual characteristics and the corresponding food product claims" (Food Integrity Handbook, 2018). According to the Food and Agriculture Organization of the United Nations (FAO) and the World Health Organization (WHO), "food authenticity is the quality of a food to be genuine and undisputed in its nature, origin, identity, and claims, and to meet expected properties" (Food and Agriculture Organization of United Nations & World Health Organization, 2018). Italy, as well as other Southern Europe countries (Greece, Spain, Portugal…), is increasingly interested in food authenticity research since it produces a relevant number of foods with Protected Designation of Origin (PDO), Protected Geographical Indication (PGI), and Traditional Speciality Guaranteed (TSG) certifications. In spite of that, nations such as China and the United States are significantly increasing their commitment to the food authenticity field in the last decade (Danezis, Tsagkaris, Camin, Brusic, & Georgiou, 2016). Hazelnuts are nuts belonging to the class *Corylus avellana* and *C. maxima*, the main global country producer is Turkey (ca. 70 % of the global production), followed by Italy, which is one of the principal importers (Maestri, Imperiale, & Marmiroli, 2020). Hazelnuts are commercialized with or without shells and consumed fresh and roasted, or they can be employed as an ingredient since hazelnut paste is a popular commodity in confectionery. The nation of origin and the indication of the species could be one of the most important characteristics for tree nuts concerning food fraud, as this information is typically required on the label, and frequently these features are the basement for the development of protected denominations. In particular, Italian hazelnuts represent a high-quality food, certified by the abovementioned PDO and PGI brands. This aspect could have determined a bigger price than the other producing/importing countries. The vast availability of hazelnuts from Turkey, and the lower selling price of other big producers, such as Azerbaijan and Georgia, favoured illegal activities of false origin declaration by fraudulent suppliers, "contaminating" Italian batches with cheaper and/or more abundant products or even entirely substituting them (Lang, et al., 2021). According to the Food Integrity Handbook (Morin & Lees, 2018), several analytical approaches were exploited to solve hazelnut-related food authenticity issues. Ruiz del Castillo *et al.* evaluated the potential of enantiomeric analysis of (*E*)-5-methyl-hept-2-en-4-one (filbertone) for authenticity control of hazelnut and hazelnut oil. A pre-separation of the hazelnut extract with high-pressure liquid chromatography (HPLC) was followed by the enantioselective gas chromatography (GC) analysis of

the resulting fraction (Ruiz del Castillo, Gomez Caballero, Blanch, & Herraiz, 2002). Filbertone was also analysed for the authenticity assessment of hazelnut-based products. The compound was determined through headspace solid phase micro extraction-GC-mass spectrometry (HS-SPME-GC-MS), and it was considered a proper marker to estimate the quality of commercial hazelnut spreads. According to these analyses, the spreads were classified into three groups, samples with minimal hazelnuts content (less than 1 % - filbertone concentration lower than 4 µg/kg), with middle hazelnuts content (1-10 % - filbertone 4-45 µg/kg), and with high hazelnuts content (> 10 %, filbertone > 45 µg/kg) (Cizkova, Rajchl, Snebergrova, & Voldrich, Filbertone as a marker for the assessment of hazelnut spread quality, 2013). Fourier-transform infrared (FTIR) spectroscopy was also employed for the discrimination of hazelnuts from different origins/cultivars, according to their IR signal intensities. The multivariate statistical approach comprehended the unsupervised principal component analysis (PCA), followed by linear discriminant analysis (LDA) and partial least square-discriminant analysis (PLS-DA), with or without variable selection, and it permitted valuable discrimination among the groups (Manfredi, et al., Fast classification of hazelnut cultivars through portable infrared spectroscopy and chemometrics, 2018). polymerase chain reaction (PCR), and immunological methods, like enzyme-linked immunosorbent assay (ELISA) are often performed for hazelnut species/cultivars/origins characterization (Rohman, et al., 2021). However, in the evaluation of the geographical origin, DNA-based markers are less effective than other chemical compounds (Esteki, et al., 2017). Regarding other spectroscopic techniques exploited for hazelnuts' authenticity issues, near-infrared (NIR) and nuclear magnetic resonance (NMR) were the most relevant in the scientific literature. NIR analyses were performed by Biancolillo *et al.*, to authenticate the Italian PDO hazelnut "Nocciola Romana". Two different classification approaches were taken into account, PLS-DA and soft independent modelling of class analogies (SIMCA), and both of them showed a high predictive ability, being applied in external validation on a test set (Biancolillo, et al., Authentication of an Italian PDO hazelnut ("Nocciola Romana") by NIR spectroscopy, 2018). The same technology was tested also by Moscetti *et al.*, for the assessment of the same PDO hazelnut. In this study, an algorithm for the selection of the best pre-treatment was carried out, and, besides SIMCA and PLS-DA, also *k*-Nearest Neighbour (KNN) and Support Vector Machine (SVM) were considered as discriminant routines. SVM has been revealed to be the optimal approach, with the best classification performance rate (Moscetti, Radicetti, Monarca, Cecchini, & Massantini, 2014). ^{1}H-NMR was used for geographical origin determination by Bachmann *et al.*, considering 262 authentic samples from five different countries, over four harvesting years Both non-polar and polar extraction protocols were achieved, but the polar fraction was the most suitable for the work. It was used for the data analysis, monitoring of the sample preparation, and measurement through PCA. LDA was then

found to be the best machine learning algorithm for the model classification, providing a 91 % cross-validation accuracy on the training set, and 96 % on the test set (Bachmann, Klockmann, Haerdter, Fischer, & Hackl, 1H NMR Spectroscopy for Determination of the Geographical Origin of Hazelnuts, 2018). Despite an interesting application of spectroscopic strategies for the authentication of hazelnuts, the researchers' focus was mainly on spectrometric techniques, especially for the geographical origin assessment. Rosso *et al.* focused on the volatile metabolome of high-quality hazelnuts, "Ordu" from Turkey and "Tonda Romana" from Italy, to evaluate their evolution along the production chain, from the harvest to the storage and the roasting phase, employing HS-SPME-GCxGC-MS. Selected pattern recognition was used to mine the outputs, whereas PCA, Fisher ratio, HCA, and analysis of variance (ANOVA) permitted to find 'decisional markers' among the most important features (Rosso, et al., 2018). Comprehensive bi-dimensional GC hyphenated with MS technology was applied for an advanced fingerprinting approach aimed at the comparative analysis of the volatile fraction of roasted hazelnuts from different origins. An HS-SPME-GCxGC-qMS solution was exploited, gaining patterns processed with a 'chromatographic fingerprinting' and a 'comprehensive template matching', and the resulting markers had a distribution that can be linked to the sensory properties, the geographical origin, and the effect of thermal treatment on the compound classes (Cordero, et al., Profiling food volatiles by comprehensive two-dimensional gas chromatography coupled with mass spectrometry: Advanced fingerprinting approaches for comparative analysis of the volatile fraction of roasted hazelnuts from different origins, 2010). Mono-dimensional GC, coupled with MS was approached by Han *et al.* for an automatic metabolic profiling analysis with chemometrics (AuMPAC), for the geographical origin estimation of hazelnuts from six Chinese regions. AuMPAC consented to monitor the Total Ion Chromatographic (TIC) peaks showing evident differences among the abovementioned regions. Afterward, a chemometric peak resolution method was applied for the screened peaks, and the recovered features were then analysed by ANOVA. The peaks with significant differences were employed for an origin discrimination model, based on two-way encoding PLS, obtaining an accuracy of up to 98 % (Han, et al., Automatic untargeted metabolic profiling analysis coupled with Chemometrics for improving metabolite identification quality to enhance geographical origin discrimination capability, 2018). Hazelnut cultivars from different areas were extensively studied to differentiate according to their geographical origins. Agronomical and pomological characterisations were initially achieved, then a polyphenolic extract was analysed through HPLC/UV analysis, and the phenolic fraction from the HPLC separation was also analysed by electrospray ionization-multistage ion trap MS (ESI-ITMS^n). The protein fraction was also extracted with a different protocol, and it was analysed via matrix-assisted laser desorption ionization-time of flight-MS (MALDI-ToF-MS) (Ciarmiello L. F., et al., Analysis of

Different European Hazelnut (Corylus avellana L.) Cultivars: Authentication, Phenotypic Features, and Phenolic Profiles, 2014). Klockmann, in two diverse papers, perpetrated both food fingerprinting, by discerning hazelnuts from different locations with an untargeted metabolomics approach, using ultra-high performance LC-quadrupole ToF (UPLC-QToF) and food targeting, determining the hazelnuts' geographical origin through a targeted metabolomics approach, exploiting LC-triple quadrupole-MS/MS (LC-QqQ-MS/MS) for the quantitation of the proper 'origin markers', identified by the previous strategy (Klockmann, Reiner, Bachmann, Hackl, & Fischer, Food Fingerprinting: Metabolomic Approaches for Geographical Origin Discrimination of Hazelnuts (Corylus avellana) by UPLC-QTOF-MS, 2016) (Klockmann, Reiner, Cain, & Fischer, Food Targeting: Geographical Origin Determination of Hazelnuts (Corylus avellana) by LC-QqQ-MS/MS-Based Targeted Metabolomics Application, 2017). An innovative technique that could be potentially helpful in origin investigation is gas chromatography-Ion mobility spectrometry (GC-IMS). This analytical solution interfaces the comprehensive separation by gas chromatography and ion mobility systems allowing fingerprinting of the volatile fraction from solid and liquid samples, with limited or even inexistent sample preparation. In addition, this technology enhances the analysis dimensionality, by combining the analytical power of the high-res chromatographic separation with the analytical selectivity of IMS, which has a Limit of Detection (LOD) range from 0.2 µg/m3 to 2 mg/m3 (Ruszkiewicz, et al., 2022). IMS is a technology for the detection of separated gaseous compounds in a mixture of analytes. It includes an ionization source and a drift tube, which is then constituted of an ionization zone and a migration zone (Yin, et al., 2021). The beta-emitting tritium (^{3}H) source is very solid and does not require an additional power supply. The Reactant Ion Peak (RIP), in the case of the radioactive source employed, represents $H_3O^+(H_2O)^n$ ions (hydronium ions), that are fundamental for the ionization of the Volatile Organic Compounds (VOCs), through charge-transfer reaction (Jurado-Campos, Martin-Gomez, Saavedra, & Arce, 2021). Subsequently, the ions pass across the drift tube, a fixed distance tube, under an electric field applied. They are divided passing into the drift tube, at a time that depends on their shape and charge distribution. Therefore, the IMS tool permits the separation of isomeric compounds. When hyphenated with a GC column, everything that is eluted is ionized and subsequently separated in the IMS cell, the electrical signal generated is recorded at the end of the tube through a Faraday plate, and the ions discrimination is favoured by a counter-flowing inert gas in the tube. A 3D graph is obtained as result, each compound is characterised by a retention time, a drift time, and an intensity value (Eiceman, Karpas, & Hill, 2016) (Garrido-Delgado, Dobao-Prieto, Arce L., & Valcarcel, 2015). This technology was also performed for food authenticity issues, particularly on the geographical origin assessment. Honey samples with different botanical origins were studied, merging HS-GC-IMS with multivariate statistical models, PCA, LDA, and kNN. It was

demonstrated, by comparing the PCA-LDA models, the complementarity of the technique with the NMR-based profiling of honey samples (Gerhardt, Birkenmeier, Schwolow, Rohn, & Weller, 2018). Another vulnerable commodity whose geographical origin was assessed by GC-IMS is olive oil. Gerhardt *et al.* compared the technology to the conventional isothermal capillary column (CC)-IMS system, in order to differentiate between extra virgin olive oil (EVOO) from Italy and Spain. GC-IMS highlighted a valuable resolving power for the untargeted profiling of VOCs in a complex matrix such as EVOO (Gerhardt, Birkenmeier, Sanders, Rohn, & Weller, 2017). Afterward, GC-IMS data were also merged with Fourier-transform mid-infrared (FT-MIR) data, for the authentication of olive oil and honey samples. Datasets were combined with a low-level data fusion approach, and a multivariate classification was performed, by PCA-LDA or PLS-DA. Data fusion is an effective tool for better classification performance. The present study focuses on the geographical assessment of the Italian hazelnut chain, from the fresh matrix to the roasted and the pasted one. This chain does not present robust scientific literature concerning authenticity frauds, and GC-IMS could represent an innovative, functional, and cost-effective strategy to deal with them. Furthermore, to evaluate how and whether the data fusion approach could improve the statistical analysis, the GC-IMS data matrix was merged with a matrix derived from sensory analysis, a complementary methodology that can provide more discriminant variables to the model for the geographical origin assessment.

MATERIALS AND METHODS

Sampling

Fresh, roasted, and paste hazelnut batches were sampled from different Italian and non-Italian regions, considering both the 2020 and the 2021 harvesting campaigns, and different storage shelf life, short and long (roughly 6 months after harvest) for 2020. The Italian samples include PGI "Tonda Gentile delle Langhe" from Piedmont, PDO "Nocciola Romana" from Lazio, and "Mortarella" from Campania. For each matrix, these three varieties were equally blended to have 'Italian samples' (N=36, 9 raw, 9 roasted, 9 peeled roasted, and 9 paste of hazelnuts) The same number of lots, considering the same sampling factors, were from Turkey, Azerbaijan, and Georgia, for a total of 108 non-Italian samples. All the samples were stored in a cold room, with a controlled temperature of 4-6 °C. For the validation set, these samples were used to create mixed ones; for each product, the Italian hazelnuts were mixed with one of the non-Italian, at different percentages of adulteration (10, 20, 50, 70, 90 %). Fresh samples were constituted of Italian and Georgian hazelnuts, the roasted ones had Italian and Azerbaijani matrices inside, whereas the pastes were made by blending Italian and Turkish samples. These mixes were employed as a test set and analysed as real samples. Tables 1 and 2 show the training and the test sets, respectively.

ORIGIN	MATRIX	PEELING	HARVESTING YEAR	SHELF LIFE
GEORGIA	Fresh	/	2020	Short
GEORGIA	Fresh	/	2020	Short
GEORGIA	Fresh	/	2020	Short
AZERBAIJAN	Fresh	/	2020	Short
AZERBAIJAN	Fresh	/	2020	Short
AZERBAIJAN	Fresh	/	2020	Short
TURKEY	Fresh	/	2020	Short
TURKEY	Fresh	/	2020	Short
TURKEY	Fresh	/	2020	Short
ITALY	Fresh	/	2020	Short
ITALY	Fresh	/	2020	Short
ITALY	Fresh	/	2020	Short
GEORGIA	Roasted	Yes	2020	Short
GEORGIA	Roasted	Yes	2020	Short
GEORGIA	Roasted	Yes	2020	Short
AZERBAIJAN	Roasted	Yes	2020	Short
AZERBAIJAN	Roasted	Yes	2020	Short
AZERBAIJAN	Roasted	Yes	2020	Short
TURKEY	Roasted	Yes	2020	Short
TURKEY	Roasted	Yes	2020	Short
TURKEY	Roasted	Yes	2020	Short
ITALY	Roasted	Yes	2020	Short
ITALY	Roasted	Yes	2020	Short
ITALY	Roasted	Yes	2020	Short
GEORGIA	Roasted	No	2020	Short
GEORGIA	Roasted	No	2020	Short
GEORGIA	Roasted	No	2020	Short
AZERBAIJAN	Roasted	No	2020	Short
AZERBAIJAN	Roasted	No	2020	Short
AZERBAIJAN	Roasted	No	2020	Short
TURKEY	Roasted	No	2020	Short
TURKEY	Roasted	No	2020	Short

TURKEY	Roasted	No	2020	Short
ITALY	Roasted	No	2020	Short
ITALY	Roasted	No	2020	Short
ITALY	Roasted	No	2020	Short
GEORGIA	Paste	/	2020	Short
GEORGIA	Paste	/	2020	Short
GEORGIA	Paste	/	2020	Short
AZERBAIJAN	Paste	/	2020	Short
AZERBAIJAN	Paste	/	2020	Short
AZERBAIJAN	Paste	/	2020	Short
TURKEY	Paste	/	2020	Short
TURKEY	Paste	/	2020	Short
TURKEY	Paste	/	2020	Short
ITALY	Paste	/	2020	Short
ITALY	Paste	/	2020	Short
ITALY	Paste	/	2020	Short
GEORGIA	Fresh	/	2020	Long
GEORGIA	Fresh	/	2020	Long
GEORGIA	Fresh	/	2020	Long
AZERBAIJAN	Fresh	/	2020	Long
AZERBAIJAN	Fresh	/	2020	Long
AZERBAIJAN	Fresh	/	2020	Long
TURKEY	Fresh	/	2020	Long
TURKEY	Fresh	/	2020	Long
TURKEY	Fresh	/	2020	Long
ITALY	Fresh	/	2020	Long
ITALY	Fresh	/	2020	Long
ITALY	Fresh	/	2020	Long
GEORGIA	Roasted	Yes	2020	Long
GEORGIA	Roasted	Yes	2020	Long
GEORGIA	Roasted	Yes	2020	Long
AZERBAIJAN	Roasted	Yes	2020	Long
AZERBAIJAN	Roasted	Yes	2020	Long
AZERBAIJAN	Roasted	Yes	2020	Long

TURKEY	Roasted	Yes	2020	Long
TURKEY	Roasted	Yes	2020	Long
TURKEY	Roasted	Yes	2020	Long
ITALY	Roasted	Yes	2020	Long
ITALY	Roasted	Yes	2020	Long
ITALY	Roasted	Yes	2020	Long
GEORGIA	Roasted	No	2020	Long
GEORGIA	Roasted	No	2020	Long
GEORGIA	Roasted	No	2020	Long
AZERBAIJAN	Roasted	No	2020	Long
AZERBAIJAN	Roasted	No	2020	Long
AZERBAIJAN	Roasted	No	2020	Long
TURKEY	Roasted	No	2020	Long
TURKEY	Roasted	No	2020	Long
TURKEY	Roasted	No	2020	Long
ITALY	Roasted	No	2020	Long
ITALY	Roasted	No	2020	Long
ITALY	Roasted	No	2020	Long
GEORGIA	Paste	/	2020	Long
GEORGIA	Paste	/	2020	Long
GEORGIA	Paste	/	2020	Long
AZERBAIJAN	Paste	/	2020	Long
AZERBAIJAN	Paste	/	2020	Long
AZERBAIJAN	Paste	/	2020	Long
TURKEY	Paste	/	2020	Long
TURKEY	Paste	/	2020	Long
TURKEY	Paste	/	2020	Long
ITALY	Paste	/	2020	Long
ITALY	Paste	/	2020	Long
ITALY	Paste	/	2020	Long
GEORGIA	Fresh	/	2021	Short
GEORGIA	Fresh	/	2021	Short
GEORGIA	Fresh	/	2021	Short
AZERBAIJAN	Fresh	/	2021	Short

AZERBAIJAN	Fresh	/	2021	Short
AZERBAIJAN	Fresh	/	2021	Short
TURKEY	Fresh	/	2021	Short
TURKEY	Fresh	/	2021	Short
TURKEY	Fresh	/	2021	Short
ITALY	Fresh	/	2021	Short
ITALY	Fresh	/	2021	Short
ITALY	Fresh	/	2021	Short
GEORGIA	Roasted	Yes	2021	Short
GEORGIA	Roasted	Yes	2021	Short
GEORGIA	Roasted	Yes	2021	Short
AZERBAIJAN	Roasted	Yes	2021	Short
AZERBAIJAN	Roasted	Yes	2021	Short
AZERBAIJAN	Roasted	Yes	2021	Short
TURKEY	Roasted	Yes	2021	Short
TURKEY	Roasted	Yes	2021	Short
TURKEY	Roasted	Yes	2021	Short
ITALY	Roasted	Yes	2021	Short
ITALY	Roasted	Yes	2021	Short
ITALY	Roasted	Yes	2021	Short
GEORGIA	Roasted	No	2021	Short
GEORGIA	Roasted	No	2021	Short
GEORGIA	Roasted	No	2021	Short
AZERBAIJAN	Roasted	No	2021	Short
AZERBAIJAN	Roasted	No	2021	Short
AZERBAIJAN	Roasted	No	2021	Short
TURKEY	Roasted	No	2021	Short
TURKEY	Roasted	No	2021	Short
TURKEY	Roasted	No	2021	Short
ITALY	Roasted	No	2021	Short
ITALY	Roasted	No	2021	Short
ITALY	Roasted	No	2021	Short
GEORGIA	Paste	/	2021	Short
GEORGIA	Paste	/	2021	Short

GEORGIA	Paste	/	2021	Short
AZERBAIJAN	Paste	/	2021	Short
AZERBAIJAN	Paste	/	2021	Short
AZERBAIJAN	Paste	/	2021	Short
TURKEY	Paste	/	2021	Short
TURKEY	Paste	/	2021	Short
TURKEY	Paste	/	2021	Short
ITALY	Paste	/	2021	Short
ITALY	Paste	/	2021	Short
ITALY	Paste	/	2021	Short

Table 1. *Training Set sample list*

ORIGIN	MATRIX	PEELING	HARVESTING YEAR	SHELF LIFE
90% ITA-10% GEO	Fresh	/	Mixed	Mixed
90% ITA-10% GEO	Fresh	/	Mixed	Mixed
90% ITA-10% GEO	Fresh	/	Mixed	Mixed
80% ITA-20% GEO	Fresh	/	Mixed	Mixed
80% ITA-20% GEO	Fresh	/	Mixed	Mixed
80% ITA-20% GEO	Fresh	/	Mixed	Mixed
50% ITA-50% GEO	Fresh	/	Mixed	Mixed
50% ITA-50% GEO	Fresh	/	Mixed	Mixed
50% ITA-50% GEO	Fresh	/	Mixed	Mixed
30% ITA-70% GEO	Fresh	/	Mixed	Mixed
30% ITA-70% GEO	Fresh	/	Mixed	Mixed
30% ITA-70% GEO	Fresh	/	Mixed	Mixed
10% ITA-90% GEO	Fresh	/	Mixed	Mixed
10% ITA-90% GEO	Fresh	/	Mixed	Mixed
10% ITA-90% GEO	Fresh	/	Mixed	Mixed
90% ITA-10% AZE	Roasted	Yes	Mixed	Mixed
90% ITA-10% AZE	Roasted	Yes	Mixed	Mixed
90% ITA-10% AZE	Roasted	Yes	Mixed	Mixed
80% ITA-20% AZE	Roasted	Yes	Mixed	Mixed
80% ITA-20% AZE	Roasted	Yes	Mixed	Mixed
80% ITA-20% AZE	Roasted	Yes	Mixed	Mixed

50% ITA-50% AZE	Roasted	Yes	Mixed	Mixed
50% ITA-50% AZE	Roasted	Yes	Mixed	Mixed
50% ITA-50% AZE	Roasted	Yes	Mixed	Mixed
30% ITA-70% AZE	Roasted	Yes	Mixed	Mixed
30% ITA-70% AZE	Roasted	Yes	Mixed	Mixed
30% ITA-70% AZE	Roasted	Yes	Mixed	Mixed
10% ITA-90% AZE	Roasted	Yes	Mixed	Mixed
10% ITA-90% AZE	Roasted	Yes	Mixed	Mixed
10% ITA-90% AZE	Roasted	Yes	Mixed	Mixed
90% ITA-10% AZE	Roasted	No	Mixed	Mixed
90% ITA-10% AZE	Roasted	No	Mixed	Mixed
90% ITA-10% AZE	Roasted	No	Mixed	Mixed
80% ITA-20% AZE	Roasted	No	Mixed	Mixed
80% ITA-20% AZE	Roasted	No	Mixed	Mixed
80% ITA-20% AZE	Roasted	No	Mixed	Mixed
50% ITA-50% AZE	Roasted	No	Mixed	Mixed
50% ITA-50% AZE	Roasted	No	Mixed	Mixed
50% ITA-50% AZE	Roasted	No	Mixed	Mixed
30% ITA-70% AZE	Roasted	No	Mixed	Mixed
30% ITA-70% AZE	Roasted	No	Mixed	Mixed
30% ITA-70% AZE	Roasted	No	Mixed	Mixed
10% ITA-90% AZE	Roasted	No	Mixed	Mixed
10% ITA-90% AZE	Roasted	No	Mixed	Mixed
10% ITA-90% AZE	Roasted	No	Mixed	Mixed
90% ITA-10% TUR	Paste	/	Mixed	Mixed
90% ITA-10% TUR	Paste	/	Mixed	Mixed
90% ITA-10% TUR	Paste	/	Mixed	Mixed
80% ITA-20% TUR	Paste	/	Mixed	Mixed
80% ITA-20% TUR	Paste	/	Mixed	Mixed
80% ITA-20% TUR	Paste	/	Mixed	Mixed
50% ITA-50% TUR	Paste	/	Mixed	Mixed
50% ITA-50% TUR	Paste	/	Mixed	Mixed
50% ITA-50% TUR	Paste	/	Mixed	Mixed
30% ITA-70% TUR	Paste	/	Mixed	Mixed

30% ITA-70% TUR	Paste	/	Mixed	Mixed
30% ITA-70% TUR	Paste	/	Mixed	Mixed
10% ITA-90% TUR	Paste	/	Mixed	Mixed
10% ITA-90% TUR	Paste	/	Mixed	Mixed
10% ITA-90% TUR	Paste	/	Mixed	Mixed

Table 2. *Test set sample list*

Sample preparation

Ca. 10 g of both fresh and roasted hazelnuts were initially minced with the knife mill Grindomix GM 200 (Retsch, Haan-Gruiten, Germany). 0.5 g were weighed into a 20 mL headspace vial, incubated then 60 °C for 5 mins, under agitation (500 rpm). Hazelnut pastes were directly weighed, incubated, and analysed as they were, without any mincing step. To evaluate the method's repeatability, each sample was double-prepared and injected.

Instrumental parameters

The GC-IMS instrument (FlavourSpec®, G.A.S. Dortmund, Dortmund, Germany) was equipped with a syringe and the autosampler PAL3-RSI Series II (CTC Analytics AG, Zwingen, Switzerland) for the headspace injection mode. The injection volume was set to 0.5 mL, and both the syringe and the injector port temperatures were at 80 °C. The chromatographic separation step was carried out with an FS-SE-54-CB-0.5 GC column (30 m length, internal diameter 0.32 mm, film thickness 0.5 µm), at the constant temperature of 40 °C. Nitrogen was the carrier gas. The separation was done without a thermal ramp, only a flow ramp was exploited: the program started at 2 mL/min for 5 mins, then the flow was brought to 31 mL/min in 4 mins, then to 100 mL/min in 20 sec, keeping it at this value for the last 2 mins, for a total GC runtime of 11 mins. The elute was subjected to the drift tube, for the ion mobility separation step. Both drift tube flow and temperature were kept constant, 150 mL/min and 45 °C, respectively. The carrier gas was nitrogen, the tube length was 9.8 cm, and the drift voltage and time were 5 kV and 30 ms, operated in positive ionization mode.

Data elaboration

Fresh, roasted, and paste hazelnut samples were separately considered, so data analyses were individually performed on each matrix sample set. This was due to the strong impact that the product processing could have on the volatile profiles of the different hazelnut-based commodities. The output of a GC-IMS analysis was a 3D graph, with the y-axis as GC retention time, the x-axis as the IM drift time, and the z-axis showing the detector response or signal intensity. To favour the graphical

visualisation, the heat map 2D fingerprint representation was ideal, the 3D-2D conversion was achieved by transforming the z-axis into a colour signature for each spot. Thus, the more intense the signal, the more coloured the spot. (Vera, Companioni, Meacham, & Gygax, 2016) An example of a fresh hazelnut sample GC-IMS 2D heat map is shown in Figure 1.

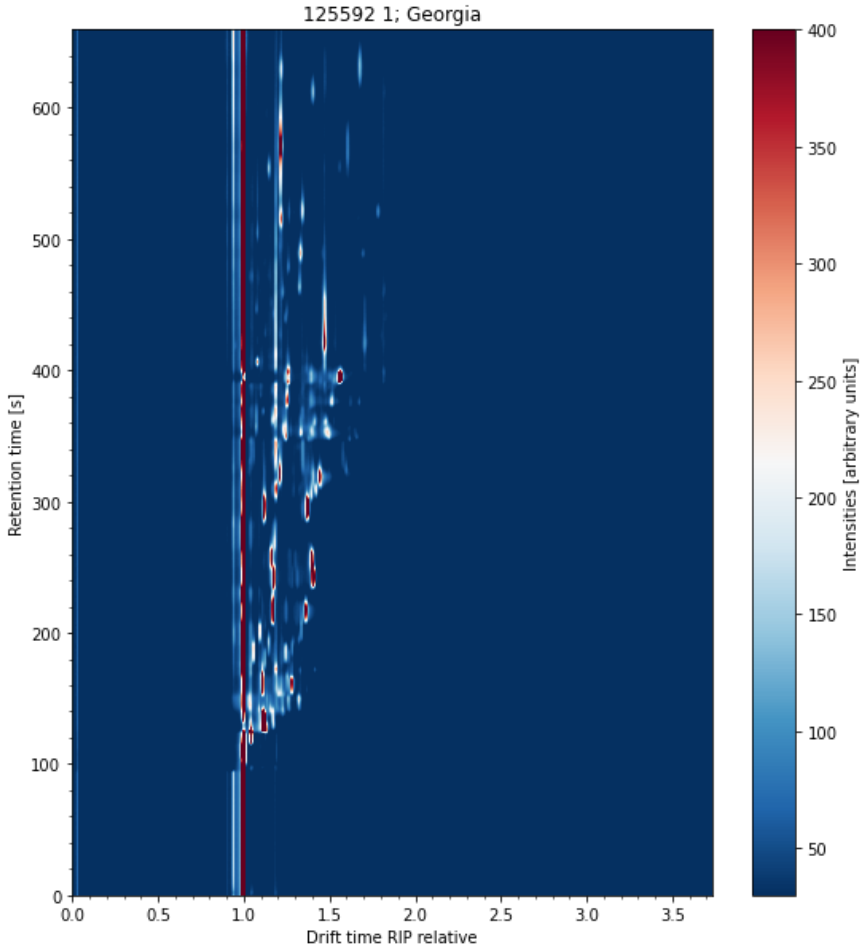


Fig. 1 *2D GC-IMS chromatogram of a fresh hazelnut sample*

GC-IMS raw data were elaborated in two different ways: by manually selecting the most intense spots, or by working on the whole spectrum. The former is achieved by creating a manual area set, which was done through the VOCal software (version 0.1.3 – G.A.S. Dortmund, Dortmund, Germany), selecting all the visible spots on the heat map of the entire samples set, considering both training and test sets. The list of spots/areas can be better visualised using a VOCal software module, called "Galerie". The module permits observation of all the areas picked for the project, showing them in all the samples analysed. From this plot, the signal intensities were exported in an Excel spreadsheet, and this obtained matrix was processed by SIMCA software (Version 16.0.1, Umetrics, Umea, Sweden). In this case, also to assess the method repeatability, both replicates were included in the processing and modelling steps. For the statistical analysis and the model setup, the spectra were normalised according to the reactant ion peak (RIP) position, the red line in Figure 1 at 1.0 ms drift time. The data elaboration of the full spectrum was achieved using the programming language Python

(Python, version 3.8.12) via Visual Studio Code (Microsoft & Electron Framework, version 1.74.1), through the *gc-ims-tools* package (version 0.1.2) for GC-IMS data handling and analysis (Christmann, Rohn, & Weller, 2022). In this other case, the package allowed us to create a mean value of the two replicates, returning score plots with a minor data dispersion. Starting from the initial spectra, all of them were normalised according to RIP, set at 1.0 ms drift time. Subsequently, it was cut, in order to better visualise the graph, and to define the area where all the spots were, that will be considered for the elaboration. Data were auto-scaled, so the mean centering was combined with weighting by the inverse standard deviation for each variable, prior to being processed with multivariate statistical models (Christmann, Rohn, & Weller, 2022). These two different approaches were evaluated to compare the data elaboration performed only on the spots of interest, using dedicated software, with defined, or eventually limited, statistical models to be applied, and the elaboration was done via coding with Python, which consents to work on the whole graph, including the risk of picking background signals, or others due to instrumental variation. However, this last approach is more flexible, as every statistical model can be applied with the proper code.

Sensory analysis

Sensory evaluation is defined as the discipline used to evoke, measure, analyze and interpret reactions to those characteristics of foods and materials as they are perceived by the sense of sight, smell taste, touch and hearing (Stone & Sidel, 1985).

The methodology used for this study is the profile attribute analysis (PAA) which is a sensory analysis methodology carried by a panel composed of expert tasters to objectively describe and measure the sensory characteristics of a product. Data obtained with the PAA can be entered efficiently into automated processing systems and are perfectly suitable for statistical analysis and aggregations.

Tasting booths following ISO 8589:2007 (*General guidance for the design of test rooms Sensory evaluations*) were used to perform the sensory evaluation. The panel was composed of eight (8) experienced assessors. The sample set tasted by the assessors was composed of one batch for each sample, for a total of 48 samples (12 fresh hazelnut lots, 24 roasted ones, unpeeled and peeled, 12 paste ones). A first tasting session was run to describe the sensory characteristics of all the samples to select the attributes to be used for the quantitative evaluation. In the subsequent session, the intensity of the sensory attributes was measured for each hazelnut sample, using a 12 points scale.

The results were then analysed by running the ANOVA test and the PCA. The main aim of ANOVA is to identify and quantify which factors are responsible for the variability of the response, while PCA maps out how products perform in terms of sensory characteristics and provide a visual snapshot of the sensory space that the samples sit in. Both ANOVA and PCA allow to analyse sensory data for

detecting and quantifying differences between products. Figure 2 displays an example of a PCA score plot of the fresh hazelnut sample set.

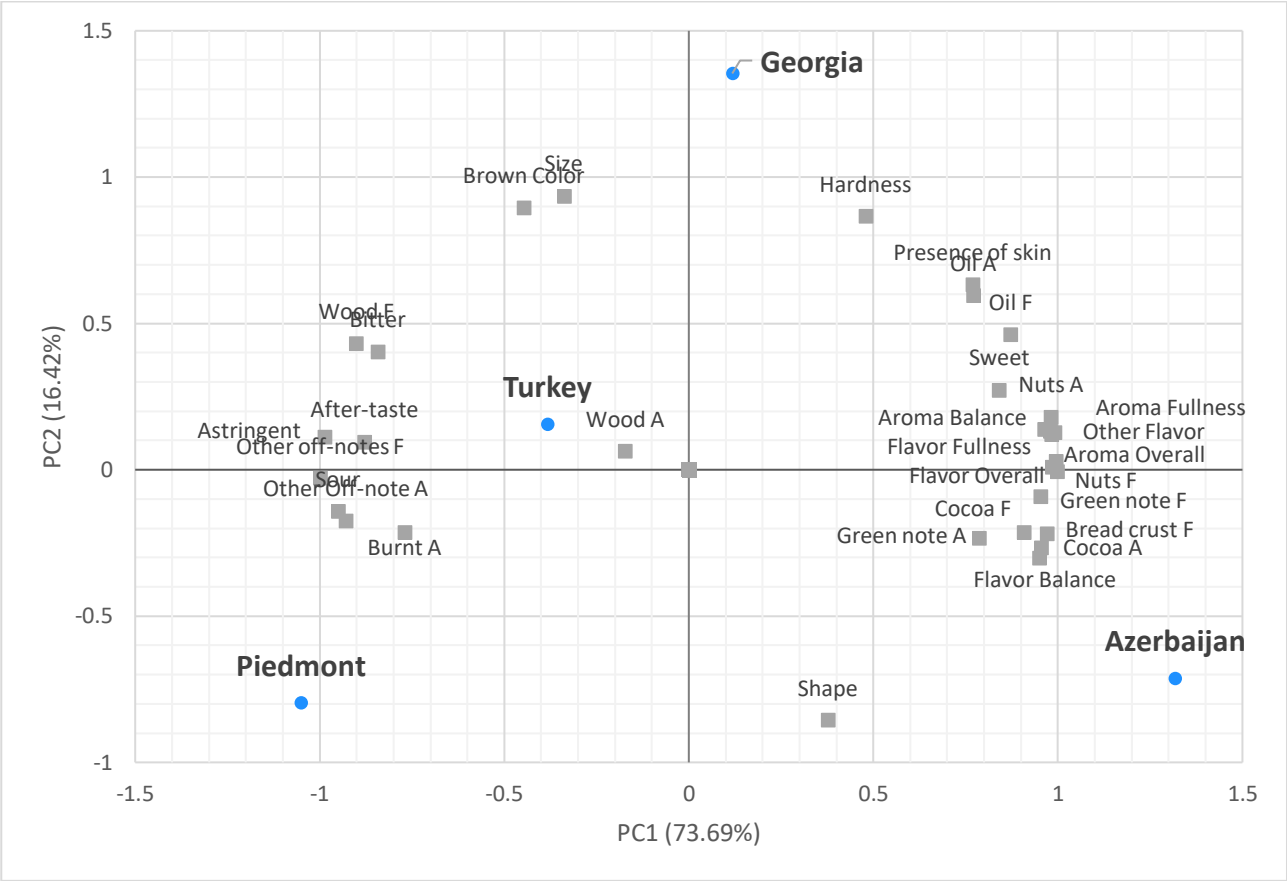


Fig. 2) *Example of a PCA score plot of the fresh hazelnut sample set.*

Data Fusion

Data from GC-IMS and sensory analyses were merged by low-level data fusion with the original data. The GC-IMS data matrix obtained from the manual area set presented a suitable dimensionality for the fusion with the sensory analysis matrix. Thus, the original data were merged without pre-treatment with a SIMCA software functionality, and then the fused block was auto-scaled before the processing. On the other hand, the entire spectrum handled with Python code returned a huge data matrix. This was due to the unfolding of three-way GC-IMS data, that generated a matrix with a lower number of rows (number of samples) and a larger number of columns (retention x drift times). After this unfolding, the matrix could be merged with the sensory one by concatenation (Schwolow, Rohn, Gerhardt, & Weller, 2019). However, the enormous difference in features number between the two datasets (ca. 3 million from GC-IMS vs 40 from sensory), was difficult to handle, even with a pre-treatment, which included “within-” and “between-block” scaling. The dominance of the GC-IMS dataset strongly impacted the model, and the data fusion did not significantly improve the models. Therefore, this specific approach was employed only using the matrix from the manual area set,

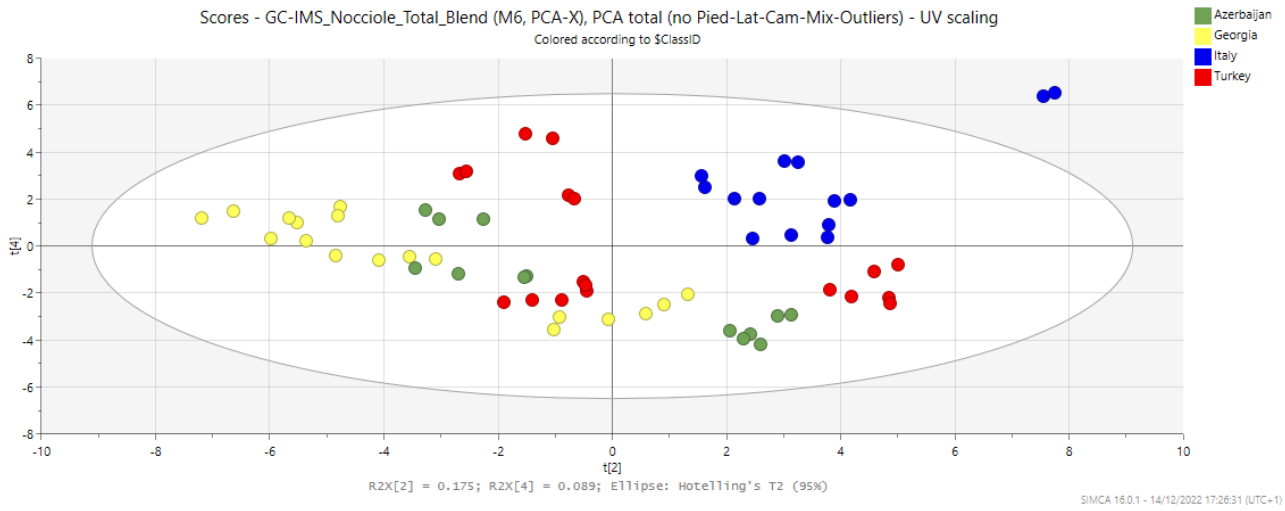
concatenating it with the sensory analysis matrix by using SIMCA software, to evaluate how the merging could potentially work on the models.

RESULTS AND DISCUSSION

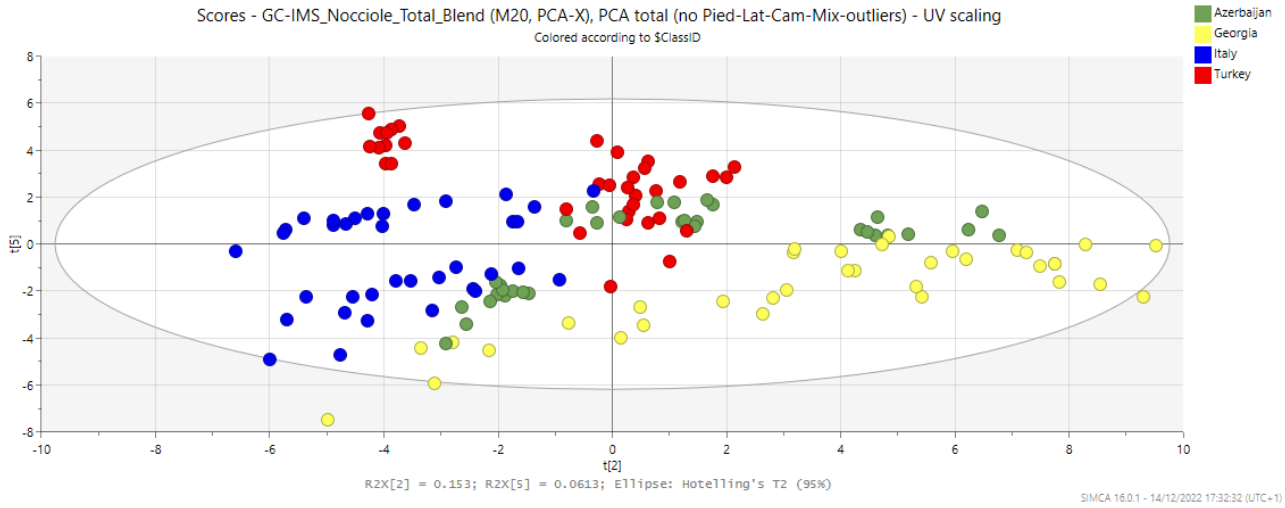
SIMCA software data elaboration

The data matrix obtained from the GC-IMS analysis by manually selecting the visible spots had between 70 and 100 variables, a manageable dimension by SIMCA software, and the multivariate statistical analysis was fundamental to deal with all of these variables at the same time, finding a correlation among them. The workflow adopted started with PCA for a preliminary visualisation of the class clustering. Figures 3A)-B)-C) show the PCA score plots of fresh, roasted, and hazelnut paste, respectively.

A)



B)



C)

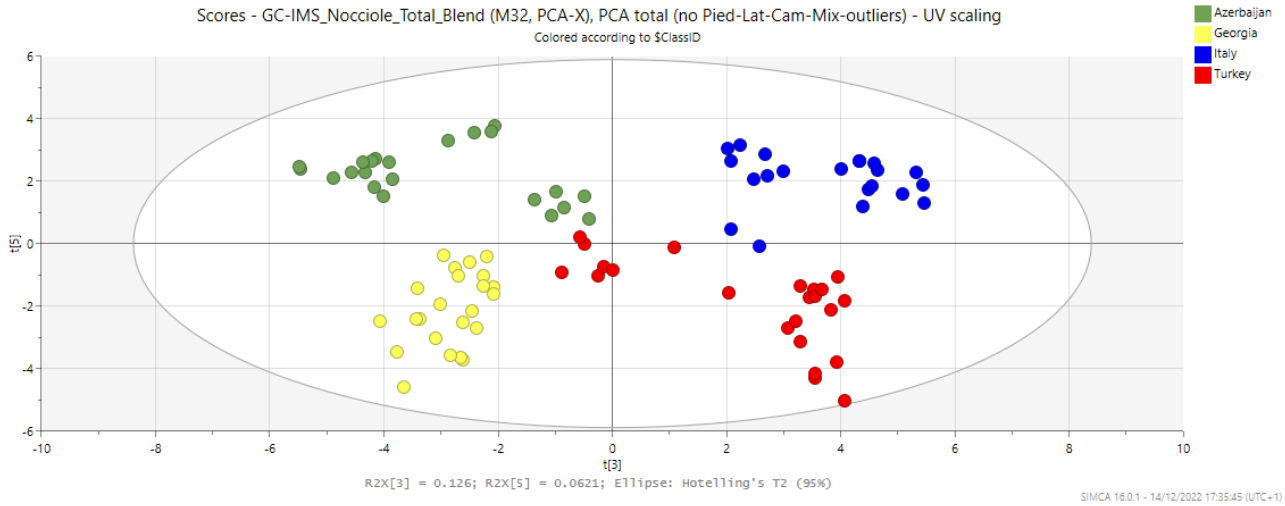
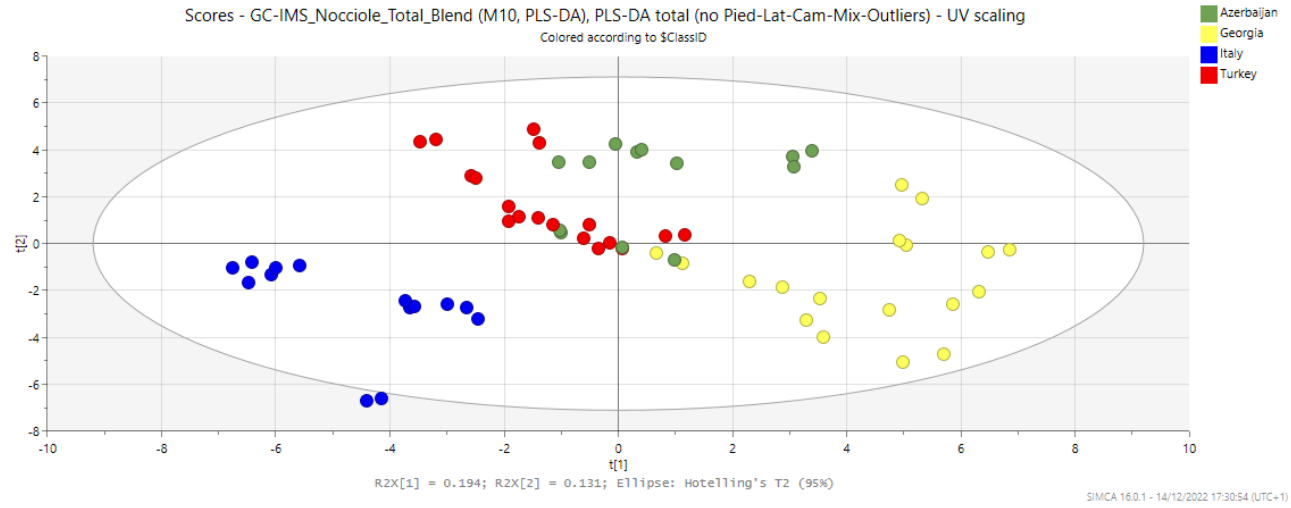


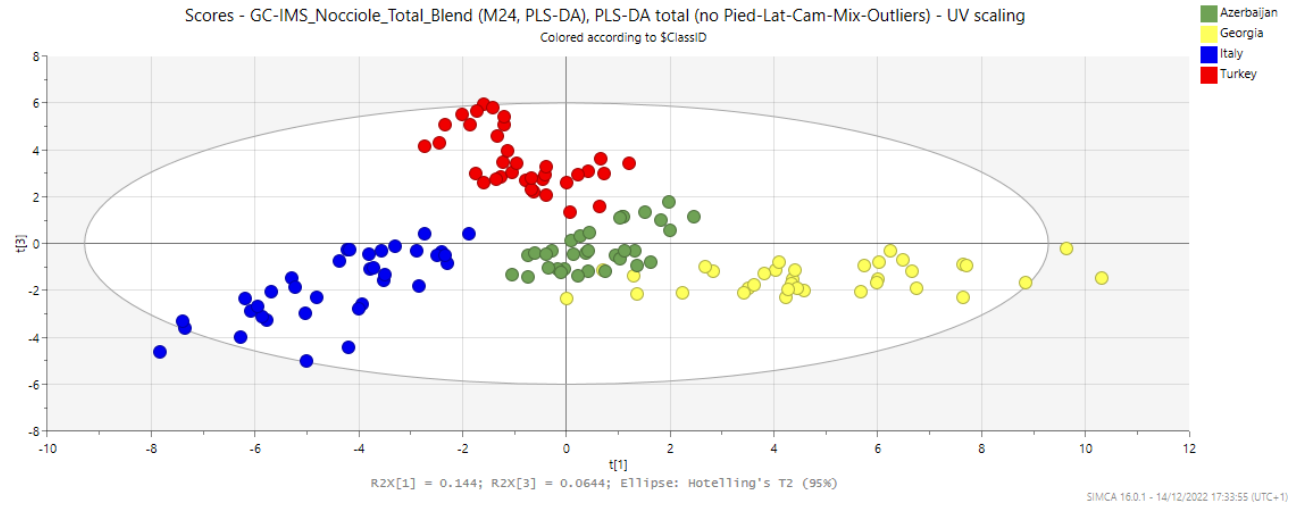
Fig. 3A) *PCA score plot of the fresh hazelnut sample set.* 3B) *PCA score plot of the roasted hazelnut sample set.* 3C) *PCA score plot of the hazelnut paste sample set. (Green dots: Azerbaijan, yellow dots: Georgia, blue dots: Italy, red dots: Turkey)*

The plots highlight a discrete separation between Italian and non-Italian samples, considering different components from the 1st since the first factor that drove the separation, with the biggest variance explained is the harvesting year. By working with the other components, it is possible to achieve a geographical origin-based clustering. However, using the other PCs led to a low explained variance, which is a statistical indication of how much variation in a dataset is attributed to each of the principal components created by the PCA (Kumar, 2022). In this case, the overall explained variance was between 20 and 30 %, this pointed out a data dispersion, mainly due to the relevant number of qualitative variables / DoE factors (storage shelf-life, harvesting year, presence of peel) related to the number of samples. The main target of the PCA is the data dimensionality reduction, in order to be able to visualise a 2D plot and extract the features. It is an unsupervised approach since it does not label the classes, hence the observations are positioned into the plot only depending on the variables that drive them (Ghojogh & Crowley, 2022). Supervised analysis, such as PLS-DA, by labelling the groups, permits to reach of a better clustering, decreasing the data dispersion. Figures 4A)-B)-C) show the PLS-DA score plots of fresh, roasted, and paste hazelnut sample sets, underlining a better class separation, with a clear division between Italian and non-Italian products, even with a discrete grouping inside the non-Italian cluster, in some cases.

A)



B)



C)

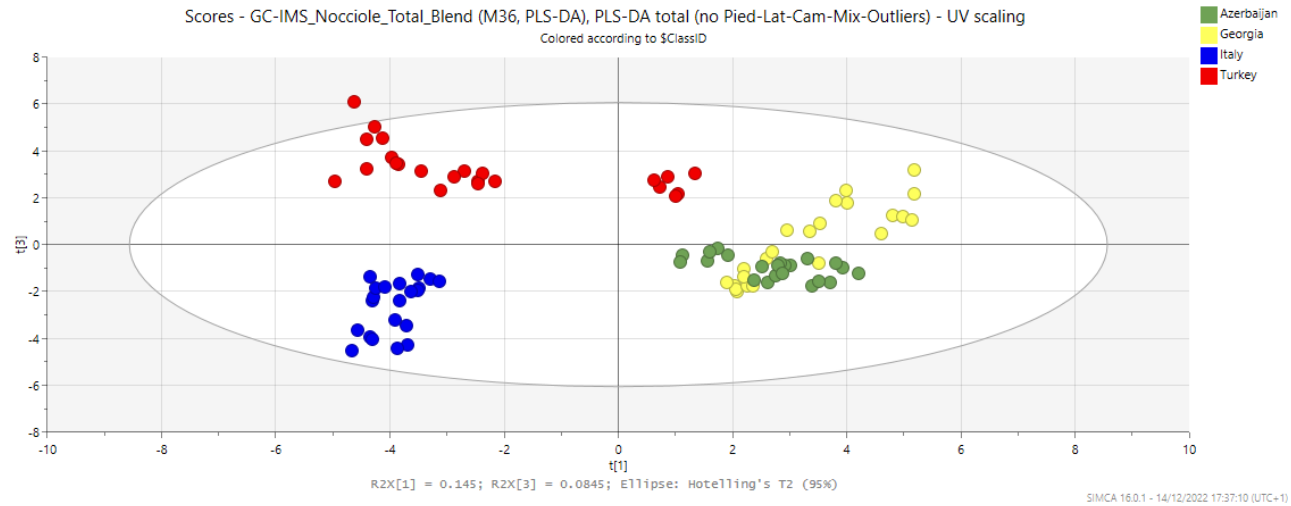


Fig. 4A) *PLS-DA score plot of the fresh hazelnut sample set.* 4B) *PLS-DA score plot of the roasted hazelnut sample set.* 4C) *PLS-DA score plot of the hazelnut paste sample set. (Green dots: Azerbaijan, yellow dots: Georgia, blue dots: Italy, red dots: Turkey)*

A misclassification table was created as well. It is a tabular way to estimate the performance of the prediction model. Each item in the misclassification table designates the number of predictions by

the model, where the classification is done correctly or incorrectly (Mohajon, 2020). In the present study, 26 samples for fresh hazelnuts, 47 for roasted hazelnuts, and 34 for hazelnut pastes were picked from the training set and used as a prediction set. An accuracy score of 100 % for all the supervised models was obtained, confirming their statistical robustness (Tables 3A-B-C).

A)

	Members	Correct	Georgia	Azerbaijan	Turkey	Italy	No class (YPred <= 0)
Georgia	4	100%	4	0	0	0	0
Azerbaijan	4	100%	0	4	0	0	0
Turkey	6	100%	0	0	6	0	0
Italy	12	100%	0	0	0	12	0
No class	0		0	0	0	0	0
Total	26	100%	4	4	6	12	0

B)

	Members	Correct	Georgia	Azerbaijan	Turkey	Italy	No class (YPred <= 0)
Georgia	9	100%	9	0	0	0	0
Azerbaijan	11	100%	0	11	0	0	0
Turkey	11	100%	0	0	11	0	0
Italy	16	100%	0	0	0	16	0
No class	0		0	0	0	0	0
Total	47	100%	9	11	11	16	0

C)

	Members	Correct	Georgia	Azerbaijan	Turkey	Italy	No class (YPred <= 0)
Georgia	7	100%	7	0	0	0	0
Azerbaijan	8	100%	0	8	0	0	0
Turkey	7	100%	0	0	7	0	0
Italy	12	100%	0	0	0	12	0
No class	0		0	0	0	0	0
Total	34	100%	7	8	7	12	0

Table 3A) *Misclassification table of the selected fresh hazelnut prediction set (n=26).* 3B) *Misclassification table of the selected roasted hazelnut prediction set (n=47).* 3C) *Misclassification table of the selected hazelnut paste prediction set (n=34). (Green cell: samples correctly classified, yellow cell: samples misclassified)*

To further assess the robustness of the predictive models, and also their feasibility in a real environment, the test set was analysed. The selected workflow was the same one used for the training set, as well as the manual area set and the matrix export. Thus, the test set was used as a prediction set, considering only the Italian class, and the other one was used for the mixed samples. This was suitable to simulate an industrial approach, where a sample is analysed, and if its volatile profile matches the Italian samples, it is considered authentic, otherwise, it is discarded or, eventually, analysed with a confirmatory methodology. A classification list was performed on the test set: it

displayed the observation (sample IDs), the original dummy variables as YVarPS, which can range from 1 to 0, and the predicted dummy variables as YPredPS. From this last value, it is possible to define the sample class:

- <0.35 the samples do not belong to the class
- Between 0.35 and 0.65 the samples are borderline
- >0.65 the samples belong to the class

(MKS Umetrics, 2015). In this work, only samples with a YPredPS bigger than 0.65 were considered Italian. Tables 4A)-B)-C) report the classification list outcomes from the fresh, roasted, paste of hazelnut test sets.

A)

Primary ID	YVarPS (Georgia)	YPredPS (Georgia)	YVarPS (Italy)	YPredPS (Italy)
10%GEO_A1	0	0.61578	0	0.38422
10%GEO_A2	0	0.422563	0	0.577437
10%GEO_B1	0	0.624495	0	0.375505
10%GEO_B2	0	0.352342	0	0.647658
10%GEO_C1	0	0.567206	0	0.432794
10%GEO_C2	0	0.525967	0	0.474033
20%GEO_A1	0	0.69405	0	0.30595
20%GEO_A2	0	0.615975	0	0.384025
20%GEO_B1	0	0.632878	0	0.367122
20%GEO_B2	0	0.610177	0	0.389823
20%GEO_C1	0	0.605968	0	0.394032
20%GEO_C2	0	0.586561	0	0.413439
50%GEO_A1	0	0.83216	0	0.16784
50%GEO_A2	0	0.796909	0	0.203091
50%GEO_B1	0	0.756463	0	0.243537
50%GEO_B2	0	0.701625	0	0.298376
50%GEO_C1	0	0.677982	0	0.322018
50%GEO_C2	0	0.711677	0	0.288323
70%GEO_A1	0	0.747834	0	0.252166
70%GEO_A2	0	0.655781	0	0.344219
70%GEO_B1	0	0.579242	0	0.420758
70%GEO_B2	0	0.575767	0	0.424233
70%GEO_C1	0	0.575282	0	0.424718
70%GEO_C2	0	0.524255	0	0.475745
90%GEO_A1	0	0.662705	0	0.337295
90%GEO_A2	0	0.660799	0	0.339201
90%GEO_B1	0	0.545649	0	0.454351

90%GEO_B2	0	0.582536	0	0.417464
90%GEO_C1	0	0.496819	0	0.503181
90%GEO_C2	0	0.469684	0	0.530316

B)

Primary ID	YVarPS (Azerbaijan)	YPredPS (Azerbaijan)	YVarPS (Italy)	YPredPS (Italy)
10%AZE_A1	0	-0.0164122	0	1.01641
10%AZE_A2	0	0.132099	0	0.867901
10%AZE_B1	0	0.256292	0	0.743708
10%AZE_B2	0	0.252218	0	0.747782
10%AZE_C1	0	0.207088	0	0.792912
10%AZE_C2	0	0.212533	0	0.787467
20%AZE_A1	0	0.0976739	0	0.902326
20%AZE_A2	0	0.202276	0	0.797724
20%AZE_B1	0	0.0180982	0	0.981902
20%AZE_B2	0	-0.0249361	0	1.02494
20%AZE_C1	0	0.202015	0	0.797985
20%AZE_C2	0	0.318547	0	0.681453
50%AZE_A1	0	0.310914	0	0.689086
50%AZE_A2	0	0.207079	0	0.792921
50%AZE_B1	0	0.443647	0	0.556353
50%AZE_B2	0	0.462625	0	0.537375
50%AZE_C1	0	0.470671	0	0.529329
50%AZE_C2	0	0.435949	0	0.564051
70%AZE_A1	0	0.384327	0	0.615673
70%AZE_A2	0	0.340855	0	0.659145
70%AZE_B1	0	0.518991	0	0.48101
70%AZE_B2	0	0.505171	0	0.494829
70%AZE_C1	0	0.376595	0	0.623405
70%AZE_C2	0	0.441973	0	0.558027
90%AZE_A1	0	0.404706	0	0.595294
90%AZE_A2	0	0.412488	0	0.587512
90%AZE_B1	0	0.560054	0	0.439946
90%AZE_B2	0	0.551305	0	0.448695
90%AZE_C1	0	0.646187	0	0.353813
90%AZE_C2	0	0.585984	0	0.414016
PEELED_10%AZE_A1	0	-0.0218881	0	1.02189
PEELED_10%AZE_A2	0	0.204611	0	0.795389

PEELED_10%AZE_B1	0	0.00203001	0	0.99797
PEELED_10%AZE_B2	0	0.018167	0	0.981833
PEELED_10%AZE_C1	0	0.10027	0	0.89973
PEELED_10%AZE_C2	0	0.164626	0	0.835374
PEELED_20%AZE_A1	0	0.313332	0	0.686668
PEELED_20%AZE_A2	0	0.321357	0	0.678643
PEELED_20%AZE_B1	0	0.352264	0	0.647736
PEELED_20%AZE_B2	0	0.351118	0	0.648882
PEELED_20%AZE_C1	0	0.206639	0	0.793361
PEELED_20%AZE_C2	0	0.259315	0	0.740685
PEELED_50%AZE_A1	0	0.377949	0	0.622051
PEELED_50%AZE_A2	0	0.440446	0	0.559554
PEELED_50%AZE_B1	0	0.381423	0	0.618577
PEELED_50%AZE_B2	0	0.347268	0	0.652732
PEELED_50%AZE_C1	0	0.345953	0	0.654047
PEELED_50%AZE_C2	0	0.445335	0	0.554665
PEELED_70%AZE_A1	0	0.537875	0	0.462125
PEELED_70%AZE_A2	0	0.483036	0	0.516964
PEELED_70%AZE_B1	0	0.351321	0	0.648679
PEELED_70%AZE_B2	0	0.360874	0	0.639126
PEELED_70%AZE_C1	0	0.500878	0	0.499122
PEELED_70%AZE_C2	0	0.486643	0	0.513357
PEELED_90%AZE_A1	0	0.568573	0	0.431427

PEELED_90%AZE_A2	0	0.502258	0	0.497742
PEELED_90%AZE_B1	0	0.524738	0	0.475262
PEELED_90%AZE_B2	0	0.525537	0	0.474463
PEELED_90%AZE_C1	0	0.586571	0	0.413429
PEELED_90%AZE_C2	0	0.61657	0	0.38343

C)

Primary ID	YVarPS (Turkey)	YPredPS (Turkey)	YVarPS (Italy)	YPredPS (Italy)
90%TUR_A	0	1.32139	0	-0.321394
90%TUR_B	0	1.33313	0	-0.333127
70%TUR_A	0	1.17264	0	-0.172644
70%TUR_B	0	1.26649	0	-0.266494
50%TUR_A	0	0.892638	0	0.107362
50%TUR_B	0	1.13559	0	-0.135585
20%TUR_A	0	0.691167	0	0.308833
20%TUR_B	0	0.746397	0	0.253603
10%TUR_A	0	0.856022	0	0.143978
10%TUR_B	0	0.782204	0	0.217796

Table 4A) *Fresh hazelnut test set Classification List.* 4B) *Roasted hazelnut test set Classification List.* 4C) *Hazelnut paste test set Classification List. The numbers in the Sample ID indicate different batches, whereas the letters indicate the replicate. (GEO: Georgia, AZE: Azerbaijan, TUR: Turkey, ITA: Italy) (Green YPredPS value: sample belongs to the class, yellow YPredPS value: sample is borderline, white YPredPS value: sample does not belong to the class)*

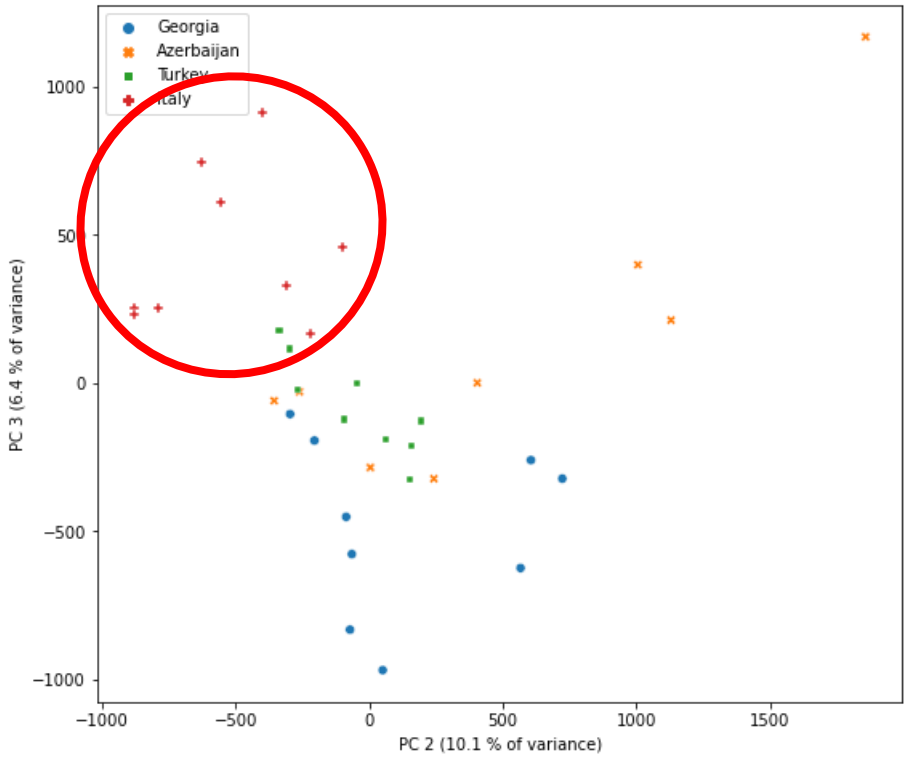
The Classification list related to the fresh hazelnut samples gave valuable outcomes since none of the mixes were classified as Italian. This confirmed the good sensitivity and the robustness of the technique on this matrix, as even the samples having only the 10 % of the non-Italian products were not wrongly classified. Concerning the roasted hazelnut samples, the performance was worse than the one referred to as the fresh ones. Indeed, all the samples, both peeled and unpeeled roasted hazelnuts, with 90 % Italian products were classified as Italian; also, many samples having 80 and 50 % Italian hazelnuts were misclassified. Thus, with the roasted matrix, the analytical tool is less robust, and only samples having more than 50 % of non-Italian hazelnuts were grouped as anomalous. Hazelnut paste samples were all classified as non-Italian, also with high values of YPredPS. This points out a good performance of the methodology to tackle the authenticity issues of Italian materials. However, samples with a big percentage of Italian products presented relevant YPredPS values for the Turkish class; this does not denote an ideal performance by the model. Likely, a higher

number of samples should be needed to better understand the robustness of the methodology on the paste matrix.

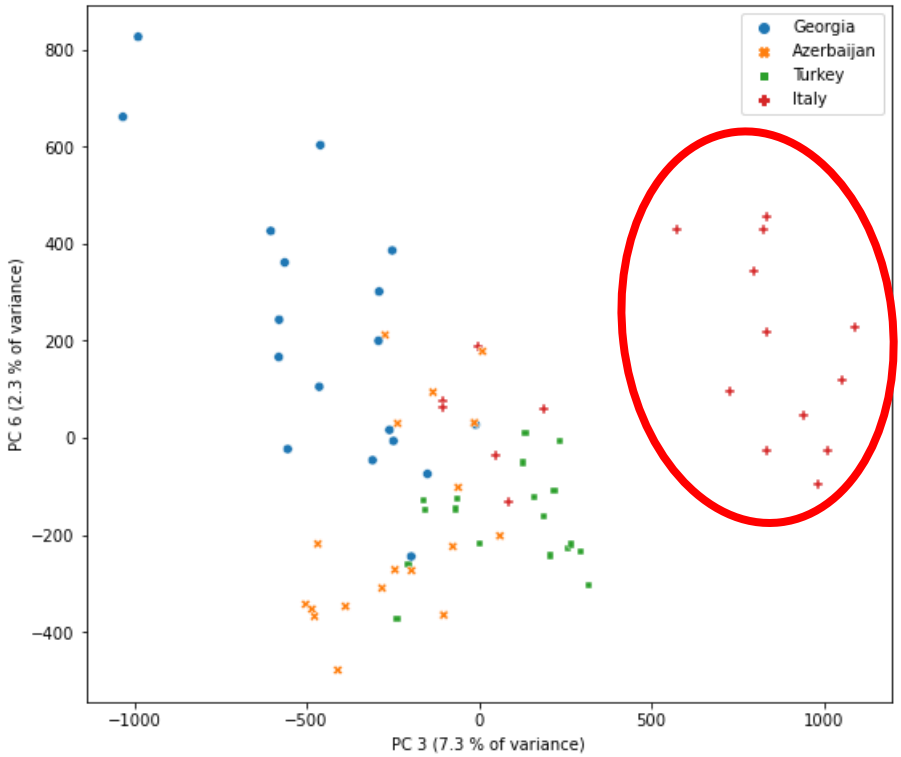
Python data elaboration with gc-ims-tools package

The workflow selected for the data elaboration with the "gc-ims-tools" package in Python coding was analogue to the one adopted with the SIMCA software, so PCA as an unsupervised model, then PLS-DA as supervised, and other tests, on the prediction set, taken from the training set, and on the test set, to assess the model's robustness. Figures 5A-B-C display the PCA score plots of fresh, roasted, and paste hazelnut samples, respectively.

A)



B)



C)

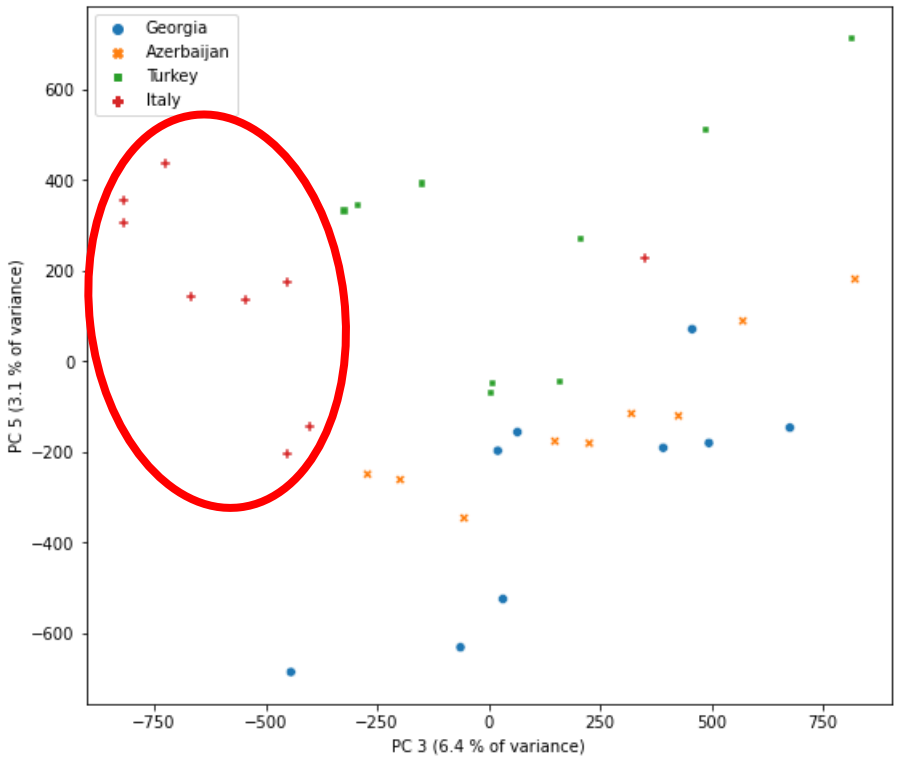
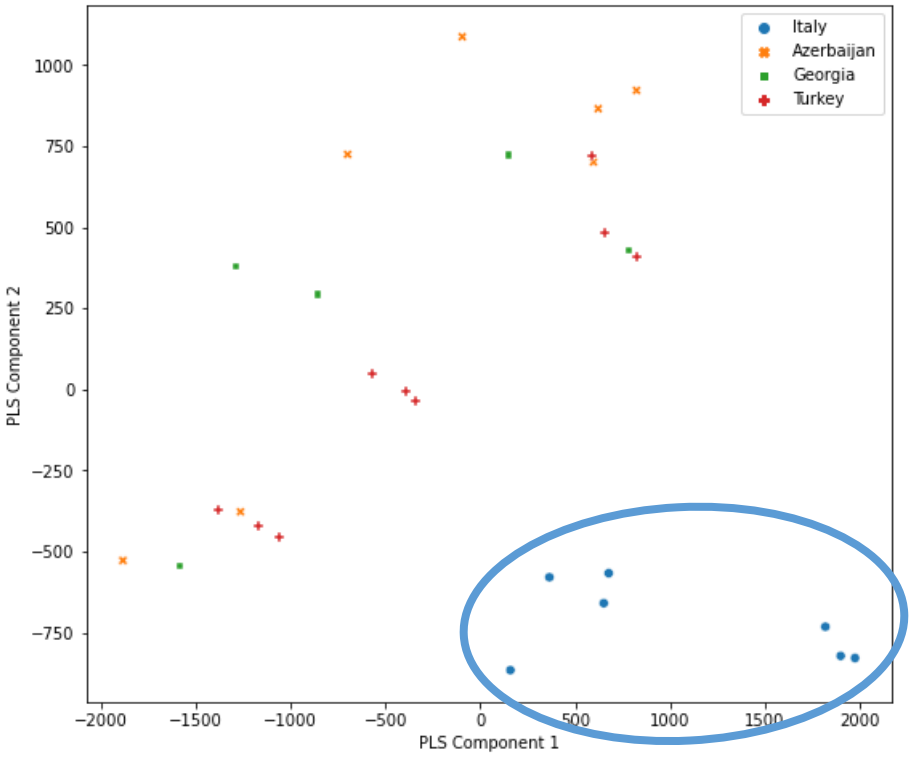


Fig. 5A) *PCA score plot of the fresh hazelnut sample set.* 5B) *PCA score plot of the roasted hazelnut sample set.* 5C) *PCA score plot of the hazelnut paste sample set. (Blue dots: Georgia, orange crosses: Azerbaijan, green squares: Turkey, red crosses: Italy, red ellipse: Italian cluster)*

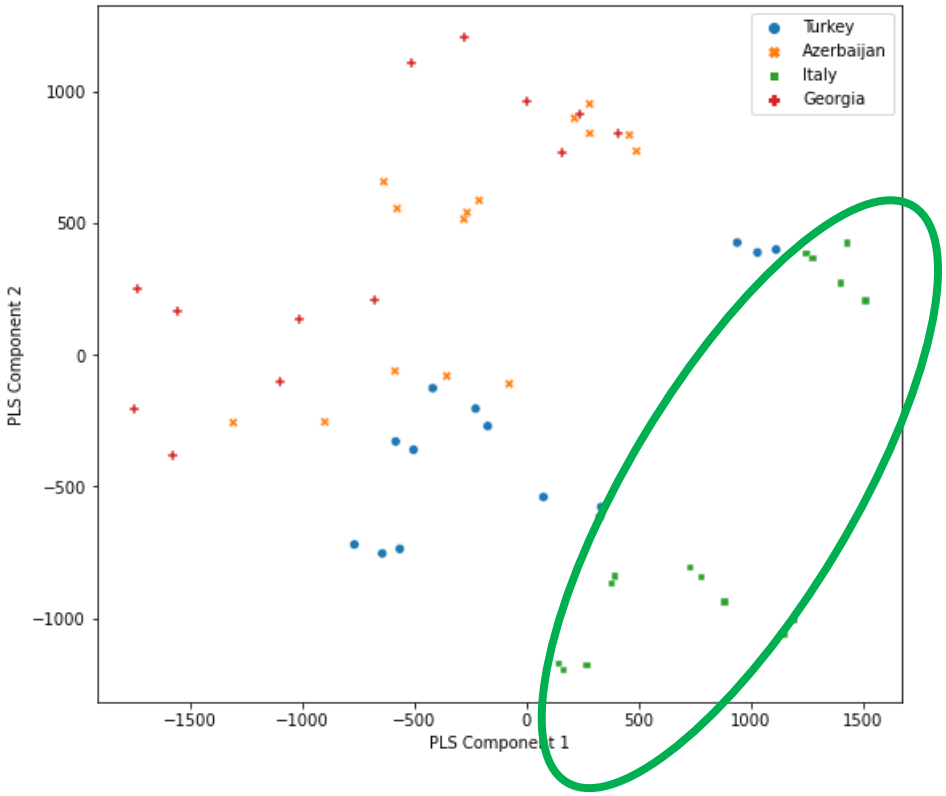
As expected, a preliminary geographical origin-based separation was achieved by considering PCs different from the first one, since, on the whole spectrum, the harvesting year and the storage shelf-

life factors were more relevant than in the model made from the manual area set. It was evident in the roasted hazelnut sample set PCA, where, out of the Italian cluster, there were six Italian samples, from the 2020 harvesting campaign, with short storage shelf-life. Nevertheless, this trend was not so evident also in the models related to the fresh and paste of hazelnuts, where the Italian clusters were quite well-defined, considering the unsupervised method. Moving to the supervised models, figures 6A-B-C show PLS-DA score plots of fresh, roasted, and paste hazelnut sample sets, respectively.

A)



B)



C)

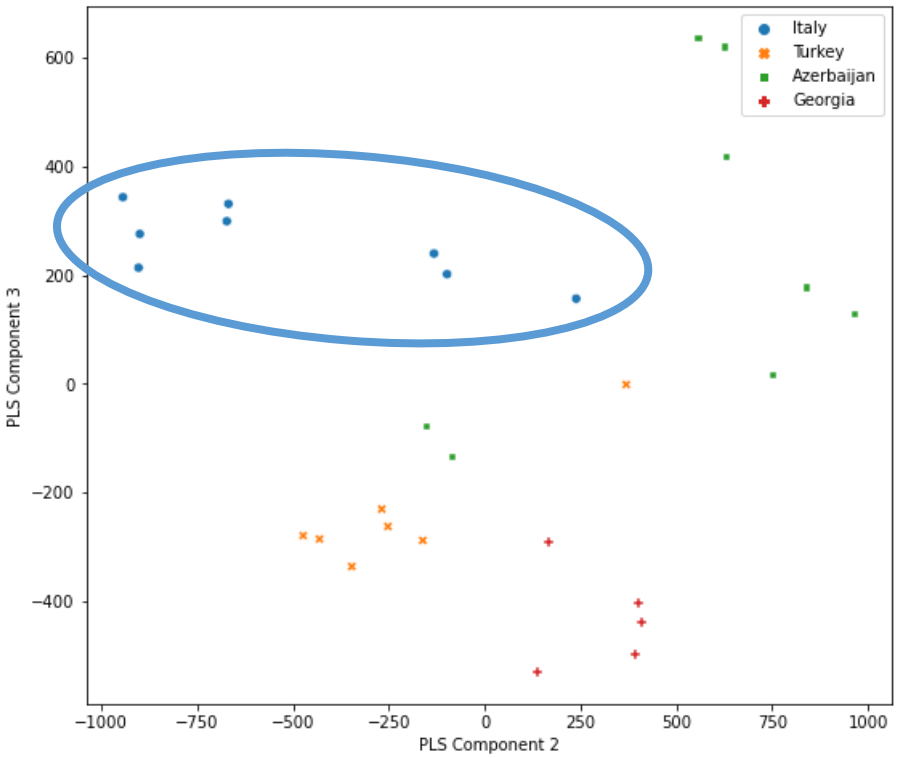


Fig. 6A) *PLS-DA score plot of the fresh hazelnut sample set. (Blue dots: Italy, orange crosses: Azerbaijan, green squares: Georgia, red crosses: Turkey, blue ellipse: Italian cluster)* 6B) *PLS-DA score plot of the roasted hazelnut sample set. (Blue dots: Turkey, orange crosses: Azerbaijan, green squares: Italy, red crosses: Georgia, green ellipse: Italian cluster)* 6C) *PLS-DA score plot of the hazelnut paste sample set. (Blue dots: Italy, orange crosses: Turkey, green squares: Azerbaijan, red crosses: Georgia, blue ellipse: Italian cluster)*

Predictably, the clustering improved, and it was possible to define a precise Italian group. In the roasted hazelnut score plot, it is still evident a clear division between samples having a different harvesting year and storage shelf-life. It is interesting how the Italian samples were well-separated from the other classes; this could be explained by taking into account the areas of cultivation in the non-Italian countries in the present study. Turkey, Azerbaijan, and Georgia are all around the "Black Sea Region", hence the climatic and soil conditions are similar. This could be reflected in products having analogue chemical characteristics, and so the VOCs. To initially estimate the robustness of the supervised model, the train-test split was done, extracting 20 % of samples from the training set to generate the test set. On this latter set, the PLS-DA algorithm was applied, in order to obtain the model accuracy score. Fresh, roasted, and paste hazelnut models reported an accuracy score of 86 %, 100 %, and 88 %, respectively. These values underlined a good model performance; however, to better assess it, and to simulate the application of the approach in a real situation, the validation set, which included both blended samples and others picked from the training set, was employed for a Random Forest Classification. It is a so-called "ensemble learning" machine learning method, used to create prediction models. Random Forests are a collection of classification and regression trees, elementary models using binary split on the variable to define predictions. Decision trees provide a perceptive method for predicting output, that divides "high" vs. "low" values of a variable linked to it. Nonetheless, this methodology often returns low accuracy in the case of intricate datasets. Random Forest builds classification and regression trees exploiting random training sets and subsets variables for modelling outputs. Outcomes from each tree are cumulated, returning a prediction for each observation. This leads to higher accuracy than a single decision tree model (Breiman, Random Forests, 2001) (Breiman, Friedman, Stone, & Olshen, 1984) (Speiser, Durkalski, & Lee, Random forest classification of etiologies for an orphan disease, 2015) (Speiser, Miller, Tooze, & Ip, 2019). Random Forest Classifier was then applied to the validation set after the model was trained using the training set. Prediction variables were generated and compared with the ones from the validation set, and, from this comparison, the following parameters were evaluated:

- Accuracy: the ratio of the correctly classified samples to the whole pool of samples.
- Precision: the ratio of the correctly positive classified by the model to all positive samples (true and false positive).
- Sensitivity: the ratio of the correctly positive classified by the model to all that are positive (true positive and false negative).
- F1 score: the harmonic mean of the precision and sensitivity.
- Specificity: the correct negative classified by the model to all that are negative (true negative and false positive).

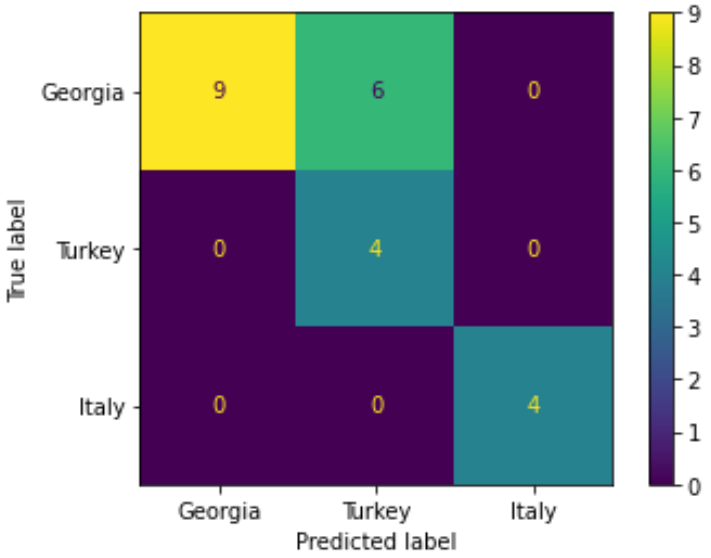
(Ghoneim, 2019). Table 6 reports the abovementioned parameters after the Random Forest Classification of fresh, roasted, and paste hazelnut validation sets.

MATRIX	ACCURACY	PRECISION	SENSITIVITY	F1-SCORE	SPECIFICITY
FRESH HAZELNUTS	74 %	90 %	74 %	76 %	74 %
ROASTED HAZELNUTS	100 %	100 %	100 %	100 %	100 %
HAZELNUT PASTES	92 %	92 %	92 %	90 %	92 %

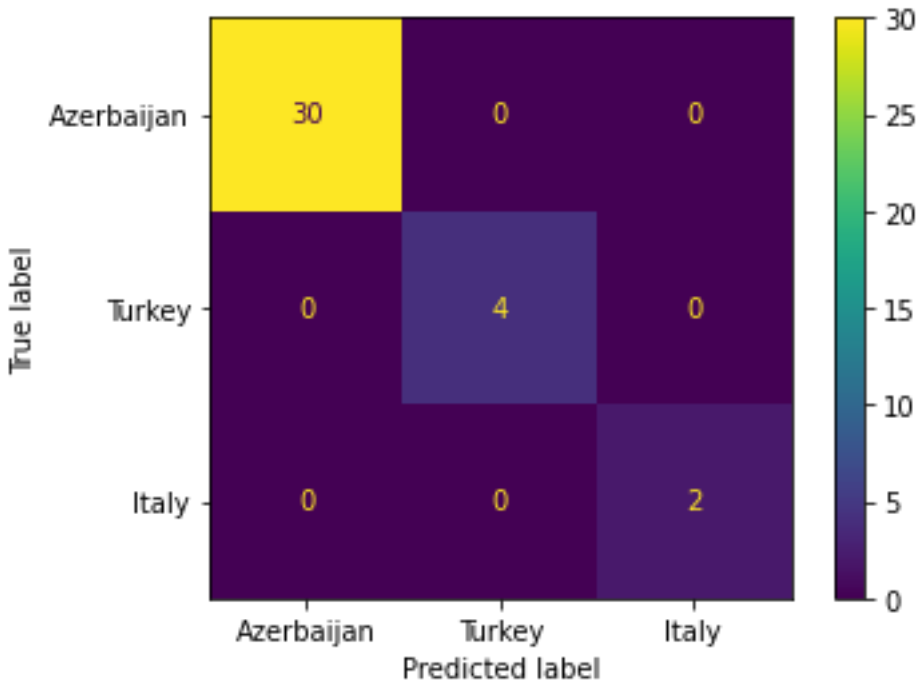
Table 6 *Accuracy, Precision, Sensitivity, F1-Score, and Specificity values of fresh, roasted, paste hazelnut validation sets.*

To conclude the assessment and to better visualise the model performance on the validation set, a confusion matrix was built. It is a table where each row represents the number of samples in the actual class, whereas each column is about the number of samples in the predicted class (Powers, 2011). Figures 7A-B-C represent the confusion matrix related to, respectively, fresh, roasted, and paste hazelnut validation sets.

A)



B)



C)

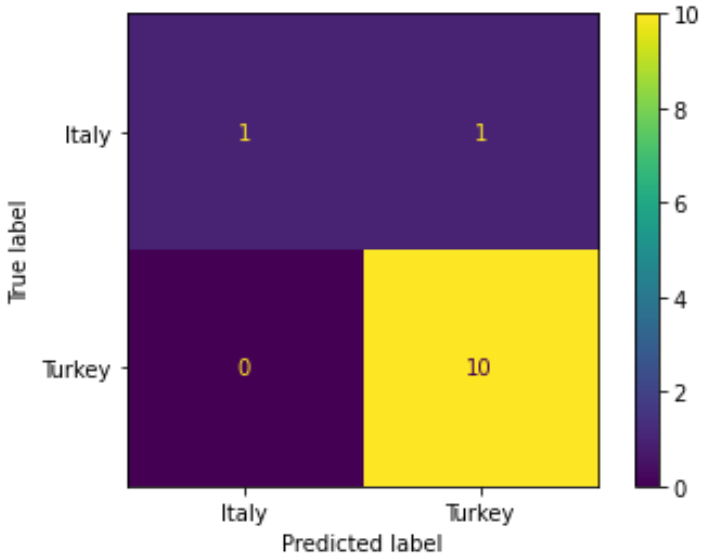


Table 7A) *Confusion matrix of the fresh hazelnut validation set.* 7B) *Confusion matrix of the roasted hazelnut validation set.* 7C) *Confusion matrix of hazelnut paste validation set.*

All the mixed samples (Italian – non-Italian at different percentages) were placed into the non-Italian class before the statistical analysis, and they were correctly classified as non-Italian. This approach also explains the non-ideal parameter values for the fresh hazelnut model. Some of the mixed samples, initially labelled as Georgian, were classified as Turkish, the class related to the other samples picked from the training set since the mix between Italian and Georgian products was evaluated, in some cases, by the model as Turkish class. This led to lower accuracy, precision, sensitivity, specificity, and F1-score. However, the model results were effective, because all the not

100 % authentic Italian samples were classified as non-Italian, making it possible, also on external sets, to detect anomalous lots, even with a percentage of Italian products.

Data fusion

Sensory analysis encompasses several sciences to better comprehend the sensory properties of commodities and the consumer's response to these properties. (Chambers & McGuire, 2003) Some parameters, according to the matrix analysed, are initially set. To assign values to these parameters, category scales are employed, for objective assessment and subjective response. (McEwan & Lyon, 2003) Tables 8A-B-C display all the sensory parameters of the hazelnut analysis, and their relative values, obtained by doing the mean of the assessors' evaluation of each sample.

A)

Piedmont 2021 A	Piedmont 2020 B	Piedmont 2020 A	Georgia 2021 B	Georgia 2021 A	Georgia 2020 B	Georgia 2020 A	Azerbaijan 2021 B	Azerbaijan 2021 A	Azerbaijan 2020 B	Azerbaijan 2020 A	product
7.00	7.22	7.00	9.25	9.50	9.00	8.78	5.13	5.13	7.33	7.44	Brown Color
8.00	8.39	8.72	5.38	5.00	5.11	5.00	9.00	9.00	9.11	8.00	Shape
6.25	7.17	7.06	8.63	8.88	6.11	6.22	4.88	4.63	7.67	7.11	Size
8.13	8.11	8.72	10.25	10.50	9.33	10.33	9.75	10.25	11.00	10.89	Presence of
3.25	4.83	5.28	4.63	4.75	2.67	3.44	5.75	5.63	6.78	6.22	Aroma
3.88	4.94	5.50	4.88	5.00	2.67	3.56	5.75	5.50	6.89	6.67	Aroma
3.88	4.94	5.50	4.88	5.00	2.67	3.56	5.75	5.50	6.89	6.78	Aroma Overall
2.88	4.11	4.50	4.25	3.38	2.56	3.00	4.50	4.13	6.56	6.22	Nuts A
2.00	1.72	1.89	1.75	1.75	1.44	1.44	1.75	2.00	1.78	1.89	Coffee A
1.63	1.89	2.11	2.00	1.88	1.33	1.33	2.75	3.13	2.11	2.78	Cocoa A
1.88	3.06	2.94	2.00	2.00	1.44	2.11	2.13	2.00	3.89	3.00	Bread crust A
5.13	4.56	4.83	5.25	5.13	4.89	4.11	4.50	5.13	4.11	4.67	Wood A
2.88	3.78	3.67	3.75	4.00	1.67	3.33	3.75	3.63	4.11	4.00	Oil A
3.88	3.78	3.72	4.00	4.00	2.89	3.11	4.63	4.88	4.22	3.78	Green note A
1.88	1.61	1.28	1.38	1.38	1.44	1.44	1.13	1.13	1.00	1.11	Burnt A
7.00	4.28	3.61	4.88	5.00	8.00	5.00	3.50	3.88	2.11	2.22	Other Off-note
6.63	6.22	6.56	7.38	7.88	5.33	6.22	7.00	7.38	7.11	7.22	Hardness
5.38	5.50	4.72	5.25	5.25	5.00	4.89	5.38	5.00	5.00	4.22	Patina
5.38	5.17	5.33	5.13	5.00	5.44	5.22	5.00	5.13	5.11	4.78	Patina type
7.00	6.89	6.33	6.88	6.88	7.44	7.22	6.25	6.38	6.22	6.11	Astringent
3.75	4.72	5.61	4.25	3.88	2.56	3.33	5.75	5.75	6.89	6.56	Flavor
3.88	4.89	5.61	5.00	4.75	2.56	3.33	5.88	5.88	7.00	6.89	Flavor
4.00	4.89	5.56	5.00	4.75	2.56	3.44	6.00	6.00	7.00	7.00	Flavor Overall
3.50	4.50	5.33	4.88	3.75	2.56	3.11	5.25	5.13	7.22	7.11	Nuts F
1.75	2.06	2.56	2.25	2.00	1.44	1.78	3.13	2.88	2.11	2.22	Cocoa F
1.63	1.72	2.22	1.75	1.88	1.67	2.00	1.88	2.00	1.56	1.78	Coffee F
1.63	2.61	2.94	2.00	1.88	1.67	2.00	3.00	2.75	4.00	4.22	Bread crust F
5.25	5.06	4.61	4.75	5.25	5.33	5.11	3.88	4.13	4.56	4.67	Wood F
3.13	4.17	4.78	3.88	4.38	1.89	3.33	4.13	4.13	5.00	4.67	Oil F
3.13	3.78	4.61	3.50	4.00	3.11	3.00	4.75	5.13	4.22	4.00	Green note F
1.75	2.61	2.11	3.00	2.88	1.89	1.56	3.63	3.25	4.11	3.11	Other Flavor
3.75	3.50	3.94	4.25	4.00	2.56	2.78	4.13	4.25	4.22	4.00	Sweet
5.25	4.78	4.11	4.63	4.63	5.33	4.89	4.38	4.38	3.56	3.67	Sour
5.88	5.33	4.61	5.75	5.88	7.89	6.78	4.13	4.00	4.00	3.56	Bitter
1.50	1.61	1.61	1.75	1.50	2.67	3.11	1.13	1.50	1.00	1.11	Burnt F
6.88	5.00	3.78	4.88	5.38	9.33	8.00	3.50	3.13	2.00	2.78	Other off-
9.00	8.89	8.44	8.50	8.75	9.89	9.33	8.13	8.25	9.00	8.78	After-taste

Peeled Georgia 2020 A	Peeled Azerbaijan 2021 B	Peeled Azerbaijan 2021 A	Peeled Azerbaijan 2020 B	Peeled Azerbaijan 2020 A	product
7.89	5.50	6.00	6.00	6.89	Brown Color
6.00	8.50	9.25	9.44	9.22	Shape
8.00	5.00	5.00	7.78	7.00	Size
5.89	6.88	7.00	4.22	5.78	Presence of
4.00	5.00	4.88	6.11	5.11	Aroma
4.78	5.13	5.13	6.67	5.11	Aroma
4.78	5.13	5.13	6.78	5.11	Aroma Overall
5.22	5.13	5.25	6.78	5.11	Nuts A
4.22	2.50	3.38	3.22	2.78	Coffee A
1.89	3.00	2.88	3.00	2.00	Cocoa A
3.22	2.88	3.63	4.33	4.22	Bread crust A
5.00	4.38	4.13	3.22	3.89	Wood A
3.78	3.63	3.75	4.22	3.89	Oil A
1.89	3.38	3.00	2.89	2.78	Green note A
5.00	2.75	3.25	2.89	2.00	Burnt A
5.89	5.00	4.50	3.11	2.89	Other Off-note
6.11	6.13	6.13	6.67	7.11	Hardness
5.89	6.00	5.88	5.67	5.89	Patina
5.11	5.63	6.13	5.89	5.22	Patina type
7.67	6.88	6.38	6.11	7.11	Astringent
3.78	4.88	4.88	5.89	5.22	Flavor
4.00	5.25	5.00	6.11	6.00	Flavor
3.89	5.13	5.13	6.11	6.11	Flavor Overall
4.78	5.88	5.88	6.22	5.22	Nuts F
1.78	3.00	3.25	3.11	2.78	Cocoa F
3.33	2.88	3.13	3.22	3.11	Coffee F
2.00	3.63	4.00	4.44	4.11	Bread crust F
5.67	4.50	4.75	3.33	3.89	Wood F
4.00	4.13	4.38	4.33	5.00	Oil F
2.67	3.75	3.88	3.22	3.67	Green note F
3.11	3.25	3.50	3.33	3.00	Other Flavor
2.78	4.00	3.88	3.67	3.67	Sweet
5.78	5.00	4.75	4.22	4.44	Sour
7.22	5.25	4.88	4.89	4.89	Bitter
6.11	2.75	3.25	3.00	2.78	Burnt F
7.00	5.25	4.75	3.00	4.22	Other off-
9.67	8.75	8.63	8.78	9.33	After-taste

B)

Turkey 2021 B	Turkey 2021 A	Turkey 2020 B	Turkey 2020 A	Piedmont 2021 B
7.63	8.13	8.78	7.78	6.75
5.38	5.38	6.00	5.67	7.75
6.75	6.88	6.22	6.11	5.88
9.50	9.13	9.78	10.00	7.63
4.13	4.13	5.78	6.00	3.13
4.13	4.13	5.78	6.11	3.63
4.13	4.25	5.89	6.22	3.75
3.50	3.00	5.33	5.11	2.75
1.88	1.50	1.44	2.00	1.75
1.88	2.13	1.89	2.67	1.63
1.75	1.88	3.00	3.00	1.50
4.63	4.00	4.33	4.78	5.38
3.25	3.13	4.00	3.89	3.00
3.13	3.13	4.22	3.89	3.63
1.25	1.13	1.22	1.11	1.75
5.00	4.88	2.78	2.67	6.88
7.38	7.13	7.00	7.33	6.75
5.88	5.50	5.00	4.89	5.88
4.75	5.00	5.11	5.67	4.75
7.25	7.00	6.33	6.22	7.50
4.13	4.13	5.33	6.11	3.50
4.25	4.13	6.00	6.44	3.63
4.75	4.75	6.00	6.44	3.63
4.00	4.13	6.00	6.22	3.50
1.50	1.50	2.33	3.11	1.75
1.75	1.50	1.78	2.11	1.63
2.00	1.88	2.89	2.22	1.50
5.13	4.88	5.11	4.33	4.88
3.38	3.38	4.00	4.33	2.88
3.63	4.25	3.67	4.11	2.88
2.13	2.25	2.89	2.89	2.25
3.38	3.75	3.44	3.44	3.50
5.13	5.00	4.67	4.22	5.00
5.50	4.75	5.00	4.89	6.00
1.63	1.50	1.67	1.67	1.50
6.00	5.75	3.00	3.00	7.25
8.50	8.38	9.00	9.00	8.75

Unpeeled Azerbaijan 2020 A	Peeled Turkey 2021 B	Peeled Turkey 2021 A	Peeled Turkey 2020 B	Peeled Turkey 2020 A	Peeled Piedmont 2021 B	Peeled Piedmont 2021 A	Peeled Piedmont 2020 B	Peeled Piedmont 2020 A	Peeled Georgia 2021 B	Peeled Georgia 2021 A	Peeled Georgia 2020 B
7.89	5.88	5.75	7.78	6.33	6.88	7.13	5.44	5.82	5.25	4.63	7.78
9.33	6.63	5.88	6.00	7.00	8.13	8.13	9.06	9.29	6.00	6.13	5.44
7.78	5.88	6.38	5.78	5.00	5.38	5.13	7.83	7.65	9.00	8.75	7.11
8.11	4.00	4.25	4.67	5.11	6.38	6.88	4.06	3.71	3.38	3.00	5.67
5.89	5.13	5.00	5.11	5.33	5.75	6.00	5.28	5.41	5.88	6.13	3.89
6.33	5.38	5.00	5.33	6.00	5.88	6.38	5.61	6.06	6.13	6.13	4.00
6.33	5.38	5.13	5.44	6.00	6.50	7.00	5.67	6.12	6.25	6.50	4.00
6.89	6.25	5.75	5.89	6.22	6.50	7.00	6.06	6.12	6.50	6.63	4.11
3.11	3.00	3.13	3.11	2.22	3.75	3.50	3.33	3.47	4.00	3.88	3.22
2.78	2.88	3.25	2.00	2.89	3.50	3.88	2.67	3.24	3.00	3.88	2.11
4.11	4.00	3.75	3.89	4.22	4.00	3.88	3.83	3.94	4.25	4.63	2.11
4.33	4.13	4.13	4.67	3.89	4.13	4.13	3.33	4.47	3.88	4.00	5.11
3.78	4.00	3.88	3.44	4.11	3.88	4.00	4.22	4.06	4.13	4.13	3.33
2.33	3.25	3.38	2.78	3.00	3.13	2.50	2.50	2.35	3.13	3.13	2.78
2.89	2.88	2.63	4.11	3.67	3.13	3.13	2.83	3.59	2.13	2.13	5.00
3.11	4.00	4.13	4.56	2.89	3.88	3.75	3.28	3.76	3.38	2.88	7.00
6.78	7.00	6.88	6.78	6.22	6.25	6.38	6.67	6.12	5.88	6.00	6.56
5.11	5.88	5.88	5.11	5.11	6.00	5.75	5.83	5.94	6.00	5.63	5.56
5.44	5.63	5.63	5.22	5.11	5.75	5.75	5.72	5.65	5.88	6.13	4.78
6.11	6.25	6.50	7.44	7.00	6.63	6.38	7.11	7.35	6.25	6.13	7.44
6.22	4.75	5.00	4.44	4.78	4.63	5.38	5.06	5.12	5.50	5.50	2.89
6.67	5.25	5.25	5.22	5.00	5.38	6.00	5.11	5.41	5.88	6.13	3.11
6.78	5.25	5.25	5.22	5.00	5.50	6.13	5.17	5.53	6.00	6.25	3.22
6.89	6.13	5.88	5.33	5.78	6.63	7.13	5.94	6.06	6.50	6.88	3.67
3.11	3.38	3.25	2.22	3.00	3.50	3.88	3.00	3.12	3.25	3.50	1.89
3.33	3.13	3.38	3.11	3.33	3.25	3.25	3.17	3.41	4.00	3.75	2.11
4.89	3.25	3.25	3.33	2.67	3.63	4.00	3.33	3.88	3.88	4.38	2.00
3.89	4.75	5.13	4.33	4.33	5.00	4.75	4.28	4.24	4.13	4.13	5.00
4.56	4.13	4.00	4.00	3.67	3.75	4.00	4.50	4.71	4.75	4.63	3.33
3.11	3.38	3.38	2.00	3.00	3.13	2.88	2.83	3.00	3.50	3.25	2.11
3.11	2.88	2.88	2.89	2.89	2.75	3.13	2.94	2.88	3.75	3.63	2.22
4.11	3.88	4.25	3.11	3.22	3.88	3.63	3.50	4.00	4.38	4.00	3.11
4.11	4.63	5.00	5.00	4.89	4.88	4.75	5.06	5.00	4.75	4.88	6.22
4.89	5.63	5.63	6.11	6.67	5.50	5.13	6.33	5.47	4.88	5.13	7.78
3.00	3.63	2.88	4.89	4.78	3.50	3.00	4.44	3.94	2.25	2.00	6.67
3.00	4.88	4.88	4.78	4.89	5.00	4.88	4.56	4.53	3.88	3.88	8.11
9.00	8.75	8.63	9.22	9.33	8.75	8.75	9.28	9.18	8.75	8.63	9.78

Unpeeled Turkey 2021 A	Unpeeled Turkey 2020 B	Unpeeled Turkey 2020 A	Unpeeled Piedmont 2021 B	Unpeeled Piedmont 2021 A	Unpeeled Piedmont 2020 B	Unpeeled Piedmont 2020 A	Unpeeled Georgia 2020 B	Unpeeled Georgia 2020 A	Unpeeled Azerbaijan 2021 B	Unpeeled Azerbaijan 2021 A	Unpeeled Azerbaijan 2020 B
8.25	6.67	7.11	7.88	7.75	7.94	7.53	8.11	7.89	6.63	7.00	7.67
6.38	7.22	6.11	7.50	7.75	9.06	9.12	5.22	7.00	8.25	8.67	9.11
6.25	5.78	5.22	5.50	5.88	7.71	7.71	6.11	7.11	5.25	5.00	8.11
8.13	4.78	4.67	7.88	8.00	6.12	6.53	7.11	6.67	9.88	10.00	6.78
4.13	5.78	4.22	5.00	5.00	5.53	6.18	3.89	5.00	3.88	4.83	5.78
4.75	5.89	4.78	5.00	5.13	5.88	6.29	4.00	5.44	4.63	5.33	5.89
4.75	5.22	4.78	5.13	5.13	6.00	6.24	4.00	5.44	4.50	5.33	6.00
4.88	5.78	5.00	5.50	5.25	6.29	7.24	4.22	5.44	4.88	5.50	6.67
3.13	2.89	2.33	3.50	3.00	3.35	3.18	2.89	3.22	2.75	3.17	3.89
2.88	2.78	2.89	3.25	3.00	2.71	3.88	2.00	2.89	2.63	3.00	2.67
3.13	3.89	3.22	3.50	2.88	4.06	3.88	3.78	3.89	3.13	3.67	4.33
4.63	4.56	4.78	4.00	4.25	3.88	3.53	4.78	4.11	4.75	4.33	4.00
3.88	4.22	3.56	3.88	3.75	4.24	4.41	3.22	3.67	3.50	4.17	4.00
2.25	3.00	2.22	2.38	2.13	2.41	2.82	2.56	2.11	2.88	3.17	2.78
2.88	3.67	3.67	3.75	4.50	4.06	2.24	5.11	3.89	3.25	3.00	4.00
5.00	3.89	4.22	4.63	5.00	3.82	2.94	5.67	5.00	5.25	4.67	3.22
6.75	6.78	7.11	7.00	6.88	6.59	6.41	6.33	6.56	6.00	6.00	6.89
6.25	5.56	5.00	6.00	6.13	5.24	5.47	5.67	5.78	6.25	5.83	6.22
5.75	5.78	4.89	6.00	5.63	5.41	5.41	6.00	6.22	5.88	5.83	5.44
7.25	6.89	7.11	7.13	7.88	7.18	7.29	6.78	7.56	7.13	6.83	7.11
5.00	5.22	3.89	4.13	3.63	5.12	5.24	3.89	4.11	4.00	4.50	5.78
5.13	5.56	4.22	4.25	3.88	5.18	5.53	4.00	4.22	4.50	5.00	6.00
5.13	5.56	4.22	4.13	3.88	5.29	5.76	4.00	4.22	4.50	5.00	6.11
5.13	5.56	4.33	5.25	4.88	5.71	6.71	4.11	4.22	4.63	5.17	6.11
3.38	2.67	2.33	2.88	2.75	3.12	3.47	2.89	2.11	2.88	2.67	2.89
3.13	3.11	4.11	3.13	2.88	3.06	3.76	3.22	3.11	3.00	3.00	3.89
2.88	3.44	3.11	3.00	2.75	2.76	3.06	3.00	3.11	2.63	3.50	3.11
4.75	4.11	4.67	5.00	4.88	4.24	4.00	5.11	5.22	4.63	4.50	4.33
4.13	4.22	3.44	3.63	4.00	4.12	4.65	3.89	3.67	3.63	4.33	4.11
2.38	3.78	3.00	2.38	2.13	2.71	2.88	3.00	2.67	2.88	3.33	3.00
3.00	3.00	2.33	2.75	2.13	2.71	3.47	3.00	2.89	3.00	3.33	3.22
3.88	3.22	3.11	3.88	3.88	3.82	3.76	2.78	3.56	3.75	3.33	3.67
5.00	5.11	5.89	5.25	5.00	5.41	4.82	5.22	5.56	5.25	5.17	5.00
5.25	6.11	6.78	6.13	7.00	6.35	6.12	7.11	7.11	5.38	5.67	6.00
3.75	4.78	5.33	4.38	5.25	5.12	4.06	6.00	4.22	4.25	4.33	4.11
4.75	4.11	6.11	5.63	6.00	5.12	4.12	6.22	6.00	5.13	4.67	4.00
8.88	9.56	9.11	9.13	9.00	9.35	9.18	9.33	9.33	9.00	8.83	9.22

Unpeeled Turkey 2021 B
8.13
6.25
6.00
8.13
4.25
4.88
4.88
5.13
3.38
3.00
3.25
5.13
4.38
3.00
3.88
4.88
6.75
6.25
5.63
7.00
4.75
5.00
5.00
5.25
3.13
2.88
3.38
4.88
4.00
3.13
2.75
4.25
5.13
5.38
3.88
4.75
8.88

C)

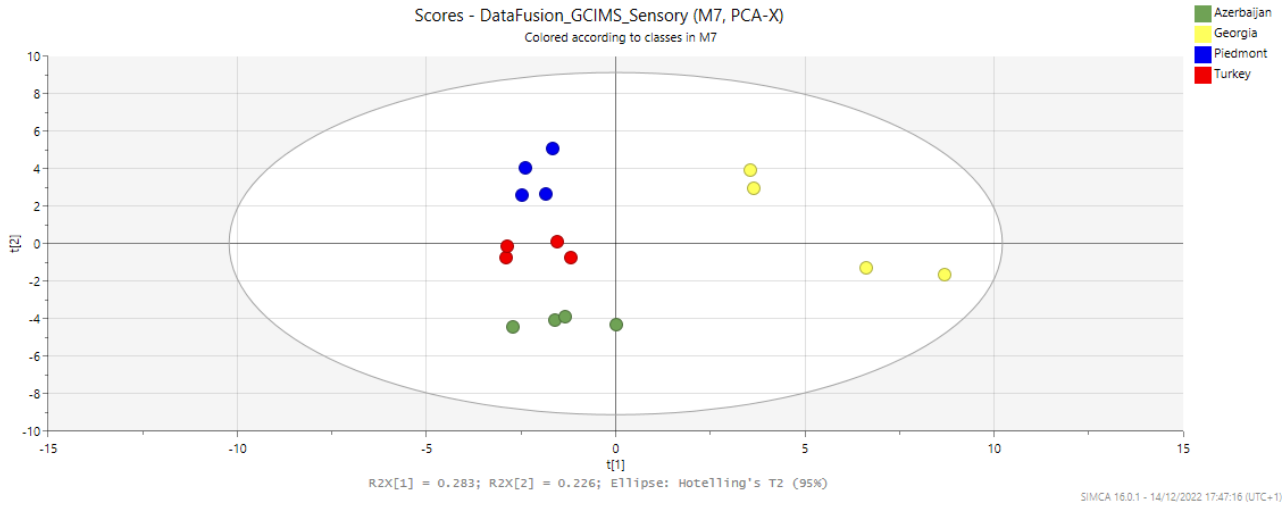
Piedmont 2021 B	Piedmont 2021 A	Piedmont 2020 B	Piedmont 2020 A	Georgia 2021 B	Georgia 2021 A	Georgia 2020 B	Georgia 2020 A	Azerbaijan 2021 B	Azerbaijan 2021 A	Azerbaijan 2020 B	Azerbaijan 2020 A	product
8.13	8.13	7.06	7.06	4.88	5.20	8.33	8.67	6.38	7.00	5.00	4.89	Brown Color
7.00	6.88	5.06	4.89	4.50	4.00	5.78	4.89	6.25	5.75	4.22	4.11	Presence of
4.63	5.13	6.17	6.06	5.13	5.80	5.11	6.00	5.50	4.75	6.33	6.78	Aroma
5.50	5.88	6.56	6.50	5.38	6.20	5.56	6.11	5.50	6.00	6.67	6.44	Aroma
5.63	5.88	6.72	6.50	5.38	6.40	5.67	6.33	5.88	5.88	6.67	6.33	Aroma
5.75	6.00	7.28	6.94	5.63	6.00	6.22	6.44	5.88	6.00	6.33	5.89	Nuts A
3.88	4.00	3.72	3.50	2.88	3.40	3.00	2.78	3.25	3.13	2.11	2.78	Coffee A
3.63	4.00	2.94	3.33	2.88	3.20	2.22	2.78	3.13	3.00	1.78	2.00	Cocoa A
3.88	4.00	3.72	3.44	3.63	4.00	3.11	2.89	3.25	3.38	4.00	3.78	Bread crust A
4.38	3.88	3.83	3.78	3.75	3.60	3.22	3.22	3.75	4.50	2.89	2.00	Wood A
5.00	4.88	4.67	4.78	5.00	5.20	4.33	4.33	4.63	4.63	4.67	4.67	Oil A
3.25	2.75	2.50	2.33	3.50	3.40	2.00	2.11	3.38	2.88	2.67	2.89	Green note A
4.25	3.00	2.61	2.61	1.75	2.00	2.78	2.67	2.88	2.75	1.33	1.44	Burnt A
4.50	4.38	2.83	2.78	3.88	3.40	3.33	3.11	4.00	4.00	1.78	1.44	Other Off-
6.75	7.13	5.22	5.39	4.88	4.80	5.44	5.67	6.13	6.25	4.89	5.00	Hardness
8.00	7.88	7.56	7.83	6.75	7.00	7.11	7.11	7.50	8.00	6.67	7.00	Patina
7.63	7.75	7.22	7.44	6.75	6.80	7.78	6.56	7.50	7.88	6.78	7.33	Patina type
8.13	8.00	7.78	7.22	6.88	7.00	7.67	7.11	6.88	7.25	6.00	6.11	Astringent
3.63	4.13	6.06	6.11	4.88	5.80	4.89	5.22	5.25	5.38	6.67	6.78	Flavor
3.75	4.25	6.61	6.67	5.38	6.00	5.67	5.89	5.88	6.00	6.67	6.89	Flavor
4.50	4.88	6.83	6.72	5.38	6.40	5.89	6.22	5.88	6.13	6.67	7.00	Flavor
4.88	5.50	7.06	7.06	5.38	6.20	6.22	6.89	6.13	6.13	6.56	7.22	Nuts F
3.25	3.75	4.22	4.33	3.25	3.60	4.00	3.22	3.88	3.13	2.22	1.78	Cocoa F
4.00	3.75	3.44	3.56	3.38	3.20	3.11	2.89	3.38	3.13	1.89	2.00	Coffee F
3.00	3.25	3.61	3.33	3.63	4.40	3.00	3.22	3.63	3.88	3.89	4.11	Bread crust F
5.38	5.25	4.22	3.94	4.00	3.60	4.89	4.67	3.88	4.38	2.22	2.67	Wood F
4.75	5.00	6.00	6.11	5.00	5.40	6.00	5.89	5.00	5.13	6.11	5.89	Oil F
2.75	2.75	3.17	3.22	3.50	3.40	2.78	2.78	3.38	3.13	3.11	4.22	Green note F
4.13	3.75	3.78	3.33	3.75	3.80	4.11	3.78	3.75	3.63	3.89	3.00	Other Flavor
4.00	3.88	4.44	4.39	5.13	4.60	3.78	4.11	5.13	4.88	4.89	5.33	Sweet
6.00	6.13	5.50	5.56	5.00	4.40	6.33	6.00	5.25	5.25	3.56	3.78	Sour
8.38	7.88	4.94	4.94	5.88	4.80	5.11	5.22	5.75	6.00	4.22	4.22	Bitter
6.25	5.88	3.28	2.67	2.88	2.20	4.78	3.67	3.25	3.25	1.33	1.44	Burnt F
6.88	5.75	2.89	3.06	4.38	3.40	3.22	4.00	4.00	4.00	1.78	2.11	Other off-
10.25	9.75	9.67	9.28	9.63	9.40	9.33	9.22	9.38	9.13	9.33	9.11	After-taste

Turkey 2020 A	Turkey 2020 B	Turkey 2021 A	Turkey 2021 B
6.11	6.00	6.13	6.63
6.22	5.22	6.13	5.50
6.56	7.22	6.00	5.75
6.78	7.33	5.75	6.13
6.89	7.33	5.75	6.00
6.89	7.33	5.88	6.00
2.78	2.78	3.25	3.25
2.67	3.00	3.13	3.38
3.00	4.11	2.88	3.50
2.89	2.89	4.00	4.38
4.89	4.78	4.13	4.75
3.00	2.11	3.00	3.13
1.56	1.33	2.88	3.13
1.67	1.67	4.13	4.13
5.89	5.78	6.00	5.88
7.33	7.44	7.25	7.13
7.33	8.00	7.00	7.25
6.89	6.78	7.13	7.13
6.33	6.89	5.38	5.25
6.89	7.44	5.88	6.13
7.11	7.44	6.00	6.38
7.33	7.44	6.13	6.50
2.89	3.00	3.75	3.88
3.33	3.11	3.13	3.75
4.11	4.00	3.75	3.75
3.67	2.78	4.13	4.38
6.00	5.78	5.13	5.00
3.89	2.89	3.38	3.25
3.89	3.89	3.88	3.88
5.00	4.67	5.00	5.13
5.11	4.78	5.13	5.25
4.78	4.56	5.63	6.25
2.33	2.00	3.63	4.13
2.78	2.00	4.00	4.13
9.33	9.44	9.25	9.38

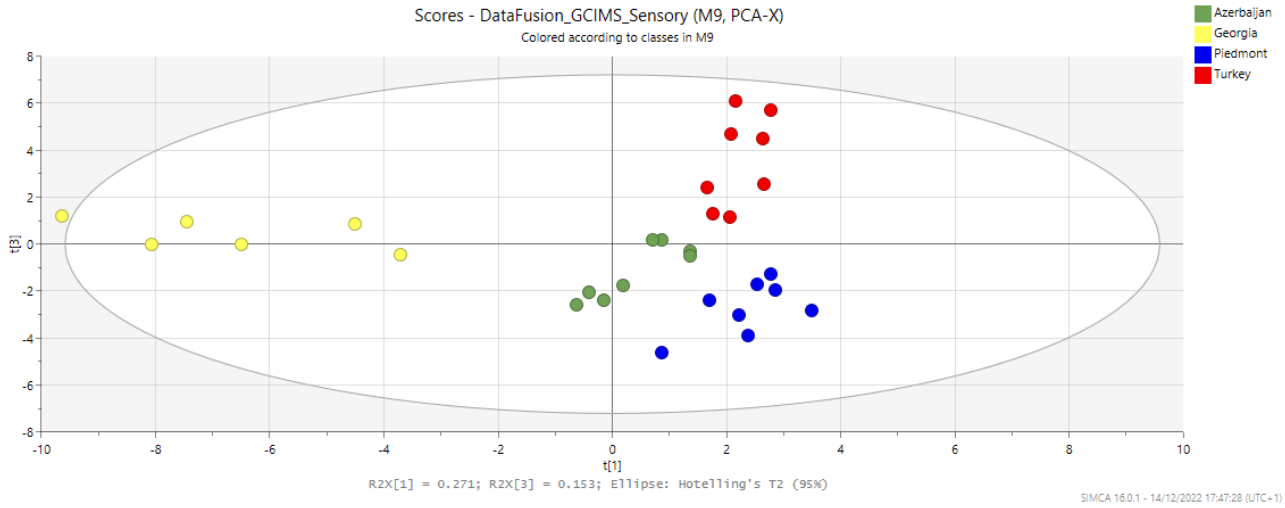
Table 8A) *Mean values of sensory parameters from fresh hazelnut analysis.* 8B) *Mean values of sensory parameters from roasted hazelnut analysis.* 8C) *Mean values of sensory parameters from fresh hazelnut analysis. (Letters after the sample ID indicate technical replicates, different products taken from the same batch; A = Aroma, F = Flavour)*

These features were concatenated with the ones from the GC-IMS analysis, and extracted with the manual area set, employing the SIMCA software functionality. Since the number of features from the two data sets was similar, no pre-treatment was achieved prior to merging them. Low-level data fusion was then performed, and, afterward, the entire data matrix obtained was auto-scaled, to make all the variables comparable. One batch, in duplicate, per type of sample (considering origin, harvesting year, and storage shelf-life) was used for the sensory analysis, hence only the matching data from the GC-IMS analysis were merged, to better investigate the efficacy of the approach, avoiding the dominance of one technique on the other. Thus, considering the limited number of samples employed, this could be defined as a 'proof-of-concept' of the analytical-sensory data fusion approach. The same protocol for the statistical analysis of the GC-IMS outputs was carried out, starting from the unsupervised PCA to the supervised PLS-DA. Figures 7A-B-C show the PCA score plots of the fresh, roasted, and paste hazelnut samples after the data fusion.

A)



B)



C)

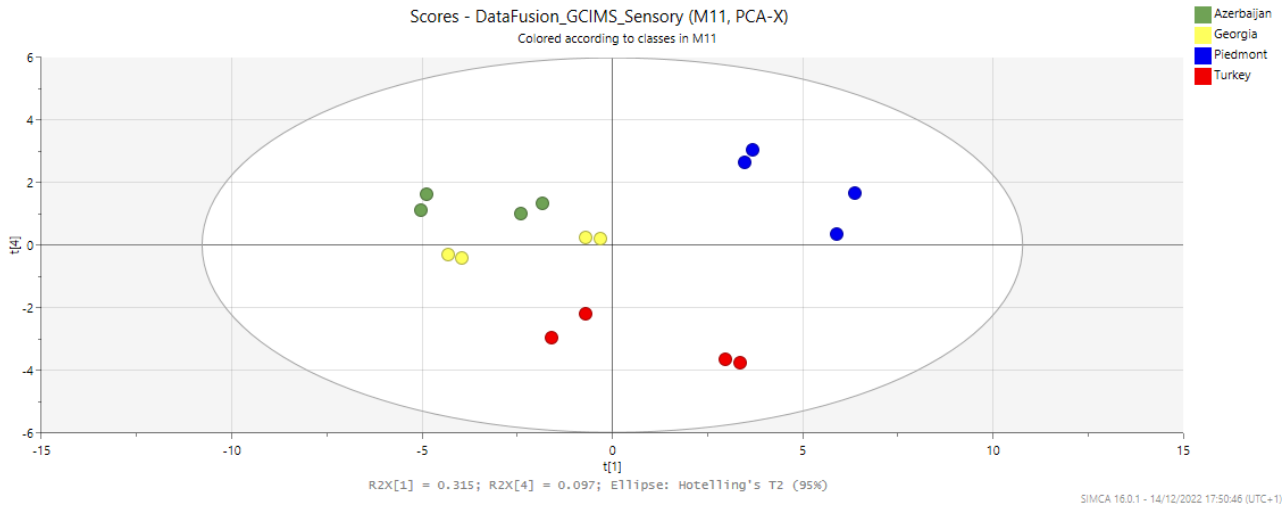
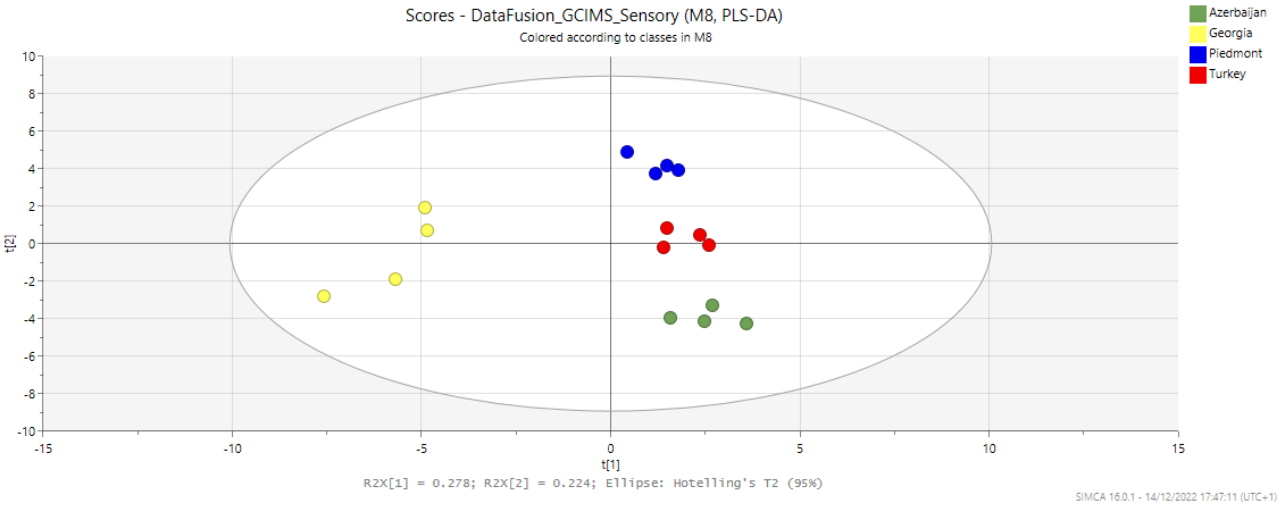


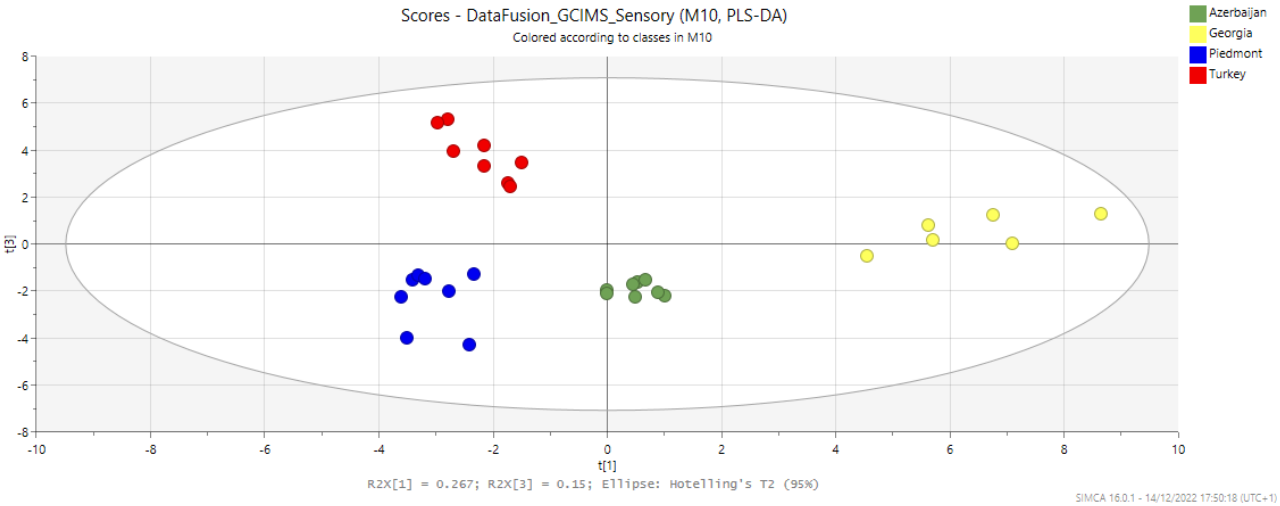
Figure 7A) *PCA score plot of fresh hazelnut samples after data fusion.* 7B) *PCA score plot of roasted hazelnut samples after data fusion.* 7C) *PCA score plot of hazelnut paste samples after data fusion. (Green dots: Azerbaijan, yellow dots: Georgia, blue dots: Piedmont, red dots: Turkey)*

The Italian mix could not be reproduced for the sensory analysis, hence only samples from one Italian region, Piedmont, were considered (PDO “Tonda Gentile Trilobata”). As it is possible to see, the group clustering seems to be improved, the classes were clearly divided. In addition, the first PC already can efficiently separate the samples according to their geographical origin, so the explained variance was increased, with a lower data dispersion and harvesting year effect. It is curious to notice how the fusion of the data sets brought a distinct division of the Georgian cluster from the others in the fresh and roasted score plots, and this grouping is then inverted in the hazelnut paste score plot, where Piedmont is the class finely separated, similarly to the GC-IMS outcome. Figures 8A-B-C indicate the PLS-DA score plots of the fresh, roasted, and paste hazelnut samples after the data fusion.

A)



B)



C)

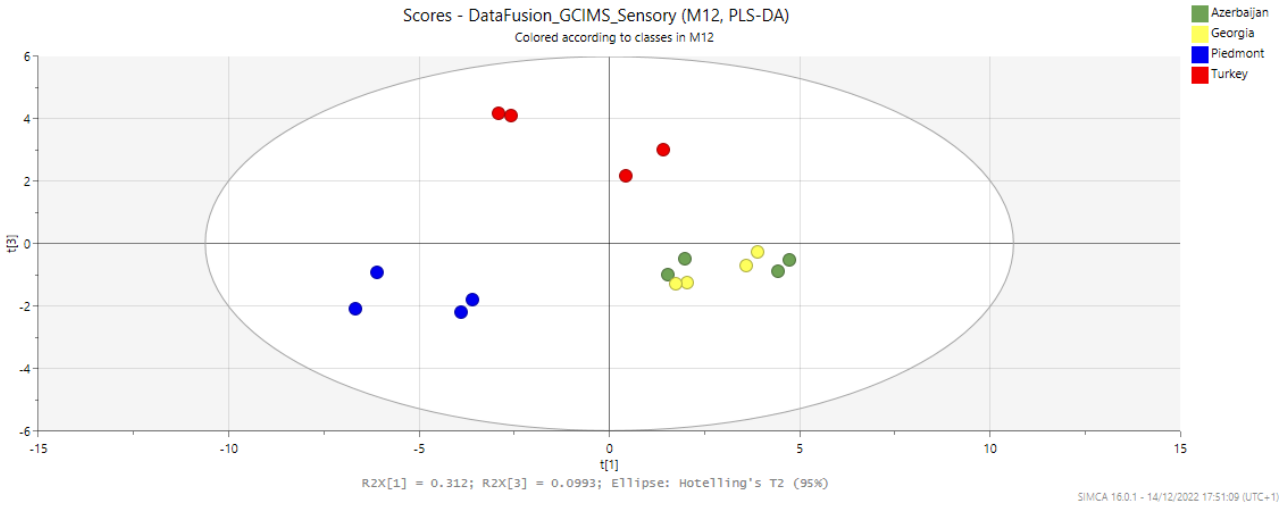


Figure 8A) *PLS-DA score plot of fresh hazelnut samples after data fusion.* 8B) *PLS-DA score plot of roasted hazelnut samples after data fusion.* 8C) *PLS-DA score plot of hazelnut paste samples after data fusion. (Green dots: Azerbaijan, yellow dots: Georgia, blue dots: Piedmont, red dots: Turkey)*

As expected, the supervised models improved the preliminary grouping, highlighting the inter-class differences. Therefore, this 'proof-of-concept' study reported interesting outcomes about the data fusion strategy, that revealed to be a tool for statistical model improvement when complementary/orthogonal techniques were considered. However, considering the limited number of samples per class, and the also the moderate number of variables included, further studies are necessary to robustly define the effectiveness of the approach.

CONCLUSIONS

Hazelnuts and their processed products are very common in bakery companies, for their versatility in different preparations. The growing demand for high-quality ingredients has led producers/suppliers to search for the best varieties, that, in the case of the hazelnut commodity, are strictly related to the geographical area of cultivation. Therefore, the authentication of this matrix is a relevant aspect from both a quality and safety point of view. The GC-IMS technology could be a functional candidate for facing authenticity issues linked to the hazelnut chain. The instrumental parameters can be set and optimised for gaining the best sample fingerprinting, as well as the sample preparation for the ideal extraction of the VOCs. This opens up the possibility for research on this technology, which can be performed at both academic and industrial levels. Simultaneously, the speed, sensitivity, and cost-effectiveness of the tool are also exploitable for routine analysis in the company's quality control department. In addition, the coupling with multivariate statistical analysis permits dealing with the big number of features of the 3D data matrix obtained, extrapolating useful chemical info. The models created employing a relevant number of samples can be applied on real lots in an automatised way.

Furthermore, it also demonstrated preliminary efficacy in merging data from this technique with a complementary one, such as sensory analysis, concerning the statistical modelling.

REFERENCES

Bachmann, R., Klockmann, S., Haerdter, J., Fischer, M., & Hackl, T. (2018). 1H NMR Spectroscopy for Determination of the Geographical Origin of Hazelnuts. *Journal of Agricultural and Food Chemistry*, 11873-11879.

Biancolillo, A., De Luca, S., Bassi, S., Roudier, L., Bucci, R., Magrì, A. D., & Marini, F. (2018). Authentication of an Italian PDO hazelnut ("Nocciola Romana") by NIR spectroscopy. *Environmental Science and Pollution Research*, 28780–28786.

Breiman, L. (2001). *Random Forests.* Berkeley, CA.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees.* New York: Taylor & Francis.

Chambers, E., & McGuire, B. (2003). SENSORY EVALUATION | Practical Considerations. In B. Caballero, *Encyclopedia of Food Sciences and Nutrition (Second Edition)* (p. 5136-5141). Elsevier Science Ltd.

Christmann, J., Rohn, S., & Weller, P. (2022). gc-ims-tools – A new Python package for chemometric analysis of GC-IMS data. *Food Chemistry*, 133476.

Ciarmiello, L. F., Mazzeo, M. F., Minasi, P., Peluso, A., De Luca, A., Piccirillo, P., . . . Carbone, V. (2014). Analysis of Different European Hazelnut (Corylus avellana L.) Cultivars: Authentication, Phenotypic Features, and Phenolic Profiles. *Journal of Agricultural and Food Chemistry*, 6236-6246.

Cizkova, H., Rajchl, A., Snebergrova, J., & Voldrich, M. (2013). Filbertone as a marker for the assessment of hazelnut spread quality. *Czech Journal of Food Science*, 81-87.

Cordero, C., Liberto, E., Bicchi, C., Rubiolo, P., Schieberle, P., Reichenbach, S. E., & Tao, Q. (2010). Profiling food volatiles by comprehensive two-dimensional gas chromatography coupled with mass spectrometry: Advanced fingerprinting approaches for comparative analysis of the volatile fraction of roasted hazelnuts from different origins. *Journal of Chromatography A*, 5848-5858.

Danezis, G. P., Tsagkaris, A. S., Camin, F., Brusic, V., & Georgiou, C. A. (2016). Food authentication: Techniques, trends & emerging approaches. *Trends in Analytical Chemistry, 85*, 123-132. doi:https://doi.org/10.1016/j.trac.2016.02.026

Eiceman, G., Karpas, Z., & Hill, H. J. (2016). *Ion Mobility Spectrometry, Third Edition.* Boca Raton: Taylor & Francis Group.

Esteki, B., F., S., A., R., B., Z., A., E., D., . . . Y., V. H. (2017). Classification and authentication of Iranian walnuts according to their geographical origin based on gas chromatographic fatty acid fingerprint analysis using pattern recognition methods. *Chemom. Intell. Lab. Syst.*, 251-258. doi:10.1016/j.chemolab.2017.10.014

Food and Agriculture Organization of United Nations & World Health Organization. (2018). CODEX COMMITTEE ON FOOD IMPORT AND EXPORT INSPECTION AND CERTIFICATION SYSTEMS. *DISCUSSION PAPER ON FOOD INTEGRITY AND FOOD AUTHENTICITY*, (p. 1-7). Brisbane, Australia.

Food Integrity Handbook. (2018). doi:https://doi.org/10.32741/fihb

Garrido-Delgado, R., Dobao-Prieto, M., Arce L., & Valcarcel, M. (2015). Determination of volatile compounds by GC-IMS to assign the quality of virgin olive oil. *Food Chemistry*, 572-579.

Gerhardt, N., Birkenmeier, M., Schwolow, S., Rohn, S., & Weller, P. (2018). Volatile-Compound Fingerprinting by Headspace-Gas-Chromatography Ion-Mobility Spectrometry (HS-GC-IMS) as a Benchtop Alternative to 1H NMR Profiling for Assessment of the Authenticity of Honey. *Analytical Chemistry*, 1777–1785.

Ghojogh, B., & Crowley, M. (2022). Unsupervised and Supervised Principal Component Analysis: Tutorial. *arXiv*.

Ghoneim, S. (2019, April 2). *Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on?* Tratto da Towards Data Science: https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124

Han, L., Zhang, Y., Song, J., Fan, M., Yu, Y., Liu, P., . . . She, Y. (2018). Automatic untargeted metabolic profiling analysis coupled with Chemometrics for improving metabolite identification quality to enhance geographical origin discrimination capability. *Journal of Chromatography A*, 12-20.

Jurado-Campos, N., Martin-Gomez, A., Saavedra, D., & Arce, L. (2021). Usage considerations for headspace-gas chromatography-ion mobility spectrometry as a suitable technique for qualitative analysis in a routine lab. *Journal of Chromatography A*, 461937.

Klockmann, S., Reiner, E., Bachmann, R., Hackl, T., & Fischer, M. (2016). Food Fingerprinting: Metabolomic Approaches for Geographical Origin Discrimination of Hazelnuts (Corylus avellana) by UPLC-QTOF-MS. *Journal of Agricultural and Food Chemistry*, 9253-9262.

Klockmann, S., Reiner, E., Cain, N., & Fischer, M. (2017). Food Targeting: Geographical Origin Determination of Hazelnuts (Corylus avellana) by LC-QqQ-MS/MS-Based Targeted Metabolomics Application. *Journal of Agricultural and Food Chemistry*, 1456-1465.

Kumar, A. (2022, August 11). *PCA Explained Variance Concepts with Python Example*. Tratto da VitalFlux: https://vitalflux.com/pca-explained-variance-concept-python-example/#:~:text=Conclusion-,What%20is%20Explained%20Variance%3F,component%20analysis%20(PCA)%20method.

Lang, C., Weber, N., Moller, M., Schramm, L., Schelm, S., Kohlbacher, O., & Fischer, M. (2021). Genetic authentication: Differentiation of hazelnut cultivars using polymorphic sites of the chloroplast genome. *Food Control*, 108344.

Maestri, E., Imperiale, D., & Marmiroli, N. (2020). Nuts, nut products and other seeds. In *Food Integrity Handbook.* doi:https://doi.org/10.32741/fihb

Manfredi, M., Robotti, E., Quasso, F., Mazzucco, E., Calabrese, G., & Marengo, E. (2018). Fast classification of hazelnut cultivars through portable infrared spectroscopy and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 427-435.

McEwan, J. A., & Lyon, D. H. (2003). SENSORY EVALUATION | Sensory Rating and Scoring Methods. In B. Caballero, *Encyclopedia of Food Sciences and Nutrition (Second Edition)* (p. 5148-5152). Elsevier Science Ltd.

MKS Umetrics. (2015, January 26). *Dynacentrix*. Tratto da https://www.dynacentrix.com/telecharg/SimcaP/SIMCA14_User_Guide.pdf

Mohajon, J. (2020, May 9). *Confusion Matrix for Your Multi-Class Machine Learning Model.* Tratto da Towards Data Science: https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826

Morin, J., & Lees, M. (2018). *FoodIntegrity Handbook.* Nantes: Eurofins Analytics France.

Moscetti, R., Radicetti, E., Monarca, D., Cecchini, M., & Massantini, R. (2014). Near infrared spectroscopy is suitable for the classification of hazelnuts according to Protected Designation of Origin. *Journal of the Science of Food Agriculture*, 2619-2625.

Powers, D. M. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 37-63.

Rohman, A., Erwanto, Y., M., H. M., M., R., Aldawoud, M., T., & Galanakis, C. M. (2021). The application of DNA-based methods for authentication analysis: Examples in halal. *Food Authentication and Traceability*, 195-213.

Rosso, M. C., Liberto, E., Spigolon, N., Fontana, M., Somenzi, M., Bicchi, C., & Cordero, C. (2018). Evolution of potent odorants within the volatile metabolome of high-quality hazelnuts (Corylus avellana L.): evaluation by comprehensive two-dimensional gas chromatography coupled with mass spectrometry. *Analytical and Bioanalytical Chemistry*, 3491-3506.

Ruiz del Castillo, M. L., Gomez Caballero, E., Blanch, G. P., & Herraiz, M. (2002). Enantiomeric composition of filbertone in hazelnuts and hazelnut oils from different geographical origins. *Journal of the American Oil Chemists' Society*, 589-592.

Ruszkiewicz, D., Myers, R., Henderson, B., Yusof, H., Meister, A., Moreno, S., . . . Thomas, C. (2022). Peppermint protocol: first results for gas chromatography-ion mobility spectrometry. *Journal of Breath Research*, 036004.

Schwolow, S., Rohn, S., Gerhardt, N., & Weller, P. (2019). Data fusion of GC-IMS data and FT-MIR spectra for the authentication of olive oils and honeys—is it worth to go the extra mile? *Analytical and Bioanalytical Chemistry*, 6005-6019.

Speiser, J. L., Durkalski, V. L., & Lee, W. M. (2015). Random forest classification of etiologies for an orphan disease. *Statistics in Medicine*, 887-899.

Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, 93-101.

Stone, H., & Sidel, J. L. (1985). *Sensory Evaluation Practices.* London, UK: Academic Press.

Vera, L., Companioni, E., Meacham, A., & Gygax, H. (2016). Real Time Monitoring of VOC and Odours Based on GC-IMS at Wastewater Treatment Plants. *Chemical Engineering Transactions*, 79-84.

Yin, J., Wu, M., Lin, R., Li, X., Ding, H., Han, L., . . . Li, Z. (2021). Application and development trends of gas chromatography–ion mobility spectrometry for traditional Chinese medicine, clinical, food and environmental analysis. *Microchemical Journal*, 3745.

Untargeted Gas Chromatography-Ion Mobility Spectrometry approach for the geographical origin evaluation of dehydrated apples

Giuseppe Sammarco[1,3], Chiara Dall'Asta[3], Michele Suman[1,2]

[1] Advanced Laboratory Research, Barilla G. e R. Fratelli S.p.A., Parma, Italy
[2] Department for Sustainable Food Process, Catholic University Sacred Heart, Piacenza, Italy
[3] Department of Food and Drug, University of Parma, Parma, Italy

Abstract

Gas Chromatography-Ion Mobility Spectrometry (GC-IMS) is an interesting candidate to face the geographical origin declaration fraud on dehydrated apple samples, as it represents a rapid, cost-effective, and sensitive solution for food authenticity issues. A design of experiment (DoE) led to a robust sampling, taking into account different factors, such as harvesting year, presence of peel, variety. The sample preparation was limited: GC-IMS analytical method permitted to obtain of a 3D graph in 11 minutes, and the multivariate statistical analysis returned a clear separation between Italian and non-Italian samples, considering both unsupervised and supervised approaches. The statistical model, created employing a training set, was applied on a further test set, with a good overall performance. Thus, GC-IMS could play a relevant role as a tool to prevent/fight false origin declaration frauds, but also, potentially, for other kinds of food authenticity and safety frauds.

Keywords

Dehydrated apples, Gas Chromatography-Ion Mobility Spectrometry, Food Authenticity, Multivariate statistical analysis

INTRODUCTION

Food quality and safety represent nowadays important aspects for food companies. The quality is directly related to the variety and the geographical origin of the commodities. Each cultivar is indeed characterised by a composition profile that is affected by the climate, soil, and agricultural practices, and so at the by the geographical origin. The variety and the origin could be impactful on the commercial price of the food (Feng, Wu, Zhu, He, & Zhang, 2021). In the European Union (EU), the apple (*Malus x domestica Borkh.*) is one of the most important fruit as a harvested product. In 2021, Italy ranked 6th as an apple producer country and 2nd as an apple exporter (Atlas Big, 2021) (World's Top Exports, 2021). Besides the huge amount, the quality of Italian apples is guaranteed by a set of EU geographical indications (GIs). In 2003, apples from Val di Non gained the Protected Designation

of Origin (PDO) label, while the ones from South Tyrol and Valtellina received the Protected Geographical Indication (PGI) label, in 2005 and 2010, respectively (Aguzzoni, et al., 2020). This contributed to a higher price of Italian products, compared to the other big producer, such as China, Chile, France, Hungary, Poland… In fact, in 2022, Italian apples cost 3.09 US dollars (USD) per kg, more than European and extra-European countries having important apple production (China 2.98 USD kg^{-1}, Chile 2.24 USD kg^{-1}, Poland 1.82 USD kg^{-1}, Hungary 1.50 USD kg^{-1}, France 0.93 USD kg^{-1}) (Global Product Prices, 2022). This could lead the suppliers to an economically-driven authenticity fraud, by blending apples from different areas and selling them as 'Made in Italy' food. As regards the geographical origin assessment of the fresh matrix, Li et al. employed near-infrared (NIR) spectroscopy, coupled with principal component analysis (PCA) as a multivariate statistical tool and successive projection algorithm (SPA) for the variable selection. Backpropagation neural network (BPNN), support vector machine (SVM), and extreme learning machine (ELM) were the three pattern recognition methods used. The SPA-ELM model was the most robust, leading to an identification accuracy of 98.33 % on the calibration set, and 96.67 % on the prediction set (Li, et al., 2018). The NIR spectroscopy technique was also adopted by Schmutzler & Huck, which employed a novel automated surface scanning technique, thatallowed the authors to lower the prediction errors with the Partial Least Squares (PLS) regression, and a successful multivariate clustering was achieved with the PCA, identifying 160 Golden Delicious apple samples from South Tyrol towards 235 products harvested in 20 countries (Schmutzler & Huck, 2014). Another relevant analytical strategy performed for the authenticity studies is the multi-isotopic ratio. The multi-element and multi-isotope approach aimed at the characterisation of the PDO and PGI apples grown in northern Italy. This allowed the authors to classify samples based on their cultivation area, applying a linear discriminant analysis (LDA), whose outputs demonstrated a successful sample classification, with a balanced accuracy of > 96 %. Even a regional classification was achieved, due to a variable selection (Aguzzoni, et al., 2020). A comprehensive study on the geographical origin of Slovenian apples took into account different approaches, multi-element analysis, multi-isotopic ratios (Carbon, nitrogen, oxygen, hydrogen), and selected chemical and physical properties (fruit mass, antioxidant activity, ascorbic acid content, and total phenols). A good separation was reached thanks to the $\delta\ ^{18}O$ and δ D values in water and the concentration of Rb and S in fruit juice (Bat, et al., 2012). Some works focused on apple-based products' geographical origin, mainly apple juice. Guo et al. exploited the headspace solid-phase microextraction coupled with gas chromatography (HS-SPME-GC) to develop a geographical discrimination model according to the volatile profile of the apple juice samples. Stepwise linear discriminant analysis (SLDA) was carried out as a classification model, and it highlighted successful discriminations of samples considering the variety and geographical

provenience, with a 100 % and 89.8 % success rate, respectively. Then, the compounds selected by this model were identified through gas chromatography-mass spectrometry (GC-MS) (Guo, Yue, & Yuan, 2012). Another apple-based commodity commonly employed by food companies for several products is the dehydrated apple cube/slice, but, to the best of the authors' knowledge, there is a lack of scientific works about food authenticity and/or geographical origin assessment of this matrix. One innovative, rapid, direct, and cost-effective technique that could be potentially useful for provenience discrimination is gas chromatography-ion mobility spectrometry (GC-IMS). This analytical strategy combines the double separation by the gas chromatography and the ion mobility systems, permitting the detection of the volatile fingerprinting of solid and liquid samples, with a limited, or even inexistent sample pre-treatment. Further, this technology improves the analysis dimensionality, by interfacing the analytical selectivity from the high-res chromatographic separation with the analytical selectivity of IMS, which has a limit of detection (LOD) range from 0.2 µg m^{-3} to 2 mg m^{-3} (Ruszkiewicz, et al., Peppermint protocol: first results for gas chromatography-ion mobility spectrometry, 2022). Besides other applications, this approach was also employed for food authenticity studies, focusing on the geographical origin assessment. Gerhardt et al. evaluated the botanical origin of honey samples, combining resolution-optimised HS-GC-IMS with chemometric analysis, PCA, LDA, and k nearest neighbor (kNN), also demonstrating, by comparing the PCA-LDA models, the complementarity of the technique with the NMR-based profiling of honey samples (Gerhardt, Birkenmeier, Schwolow, Rohn, & Weller, 2018). Olive oil is another matrix whose geographical origin was assessed by GC-IMS. The technology was compared to the conventional isothermal capillary column (CC)-IMS system in the geographical differentiation of extra virgin olive oil (EVOO) from Italy and Spain. GC-IMS provided superior resolving power for non-targeted profiling of VOC fractions in a complex matrix like EVOO (Gerhardt, Birkenmeier, Sanders, Rohn, & Weller, 2017). Subsequently, GC-IMS data were also fused with fourier-transform mid-infrared (FT-MIR) data for the authentication of olive oil and honey samples. Datasets were merged with a low-level data fusion approach, and a multivariate classification was carried out, by PCA-LDA or PLS-DA. Data fusion has turned out to be an effective strategy for improving classification performance (Schwolow, Gerhardt, Rohn, & Weller, 2019). The technology was also recently employed for the provenience evaluation of less common commodities, such as the Molixiang grape, and Sichuan pepper (Huajiao) (Feng, et al., 2022) (Feng, Wang, Wang, Huang, & Kan, 2022). The GC-IMS technology could be a relevant interface between the academic and industrial contexts. It represents a fast, cost-effective, and easy-to-use strategy, preserving a remarkable sensitivity. Therefore, companies are motivated in investing in this type of technology, as they can rapidly screen a good number of samples, without excessive costs related to high-res instrument, consumables, or

particular expertise. In this scenario, this study about the geographical assessment of Italian dehydration apple samples, reports a concrete application of the GC-IMS technique in an industrial environment.

MATERIALS AND METHODS

Sampling

A Design of Experiment (DoE) was conducted to have a robust dehydrated apple sampling. Different factors were considered, such as dehydration rate, presence of peeling, harvesting year, and variety. The regions of interest selected were France, China, Hungary, Poland, and Italy as the main geographical target, since the aim is to discriminate between Italian and non-Italian samples. A training set (n=59) was employed to build a statistical model, that was subsequently applied to a test set (n=12). For this set, samples from the 2022 harvesting campaign were picked, as well as samples from another location (Chile), and brought to a local Italian market. Table S1-S2 – Supplementary Materials listed the training and test sets, respectively, and their respective factors. The 2020 and 2021 samples were analysed in different periods and stored at 2-8 °C.

Sample preparation

Ca. 10 g dehydrated apple cubes were initially minced with the knife mill Grindomix GM 200 (Retsch, Haan-Gruiten, Germany). 0.5 g were weighed in a 20 mL headspace vial, incubated then 60 °C for 10 mins. To evaluate the method's repeatability, each sample was double-prepared and injected.

Instrumental parameters

The GC-IMS instrument (FlavourSpec®, G.A.S. Dortmund, Dortmund, Germany) was equipped with a syringe and the autosampler PAL3-RSI Series II (CTC Analytics AG, Zwingen, Switzerland) for the headspace injection mode. The injection volume was 0.5 mL, and both the syringe and the injector port were at 80 °C. The chromatographic separation step was achieved with an FS-SE-54-CB-0.5 GC column (30 m length, internal diameter 0.32 mm, film thickness 0.5 μm), that was kept at 40 °C. Nitrogen was employed as carrier gas. The separation was done without a thermal ramp (isothermal conditions), while a flow ramp was adopted: the program started at 2 mL min^{-1} for 5 mins, then the flow was brought to 31 mL min^{-1} in 4 mins, then to 100 mL min^{-1} in 20 s, keeping it at this value for the last 2 mins, for a total GC runtime of 11 mins. The elute was then conveyed to the drift tube, for the ion mobility separation step. Both drift tube flow and temperature were kept constant, 150 mL

min^{-1} and 45 °C, respectively. The carrier gas was nitrogen, the tube length was 9.8 cm, and the drift voltage and time were 5 kV and 30 ms, with a positive ionisation mode.

Data elaboration

The output of a GC-IMS analysis was a 3D chromatogram, the y-axis represented the GC retention time, the x-axis was for the IM drift time, and the z-axis retuned the detector response, so the signal intensity. To facilitate the graphical visualisation, the heat map 2D fingerprint representation was ideal, the 3D-2D conversion was carried out by transforming the z-axis into a colour signature for each spot. Therefore, the higher the response, the more coloured was the spot (Vera, Companioni, Meacham, & Gygax, 2016). An example of a dehydrated apple sample GC-IMS 2D heat map was shown in Figure 1.

A manual area set was created through the VOCal software (version 0.1.3 – G.A.S. Dortmund, Dortmund, Germany), picking all the visible spots on the heat map, and considering all the samples of the study, from both training and test sets. Table S3 reports all the area-set integration parameters. This list of spots/areas can be visualised using a VOCal software module, named "Galerie". The module allows us to observe all the areas selected for the project, as also shown in Figure 2.

From this plot, the spectra were exported in an Excel spreadsheet, and this generated matrix was processed by SIMCA software (Version 16.0.1, Umetrics, Umea, Sweden). For the statistical analysis and the model setup, the spectra were aligned according to the Reactant ion Peak (RIP) position, the red line in Figure 1 at 1.0 ms drift time.

RESULTS AND DISCUSSION

The data matrix obtained from the GC-IMS analysis contained a large number of variables, hence the multivariate statistical analysis was a valuable tool to handle and elaborate it. The workflow adopted provided for a PCA to preliminary visualise the class clustering and evaluate the data dispersion. Figure 3 shows the PCA score plot, which highlights a good separation between Italian and non-Italian samples, considering as principal components the 2^{nd} and the 4^{th}. The 1^{st} and the 3^{rd} mainly divided the samples according to the harvesting year rather than the origin, for this reason the others were exploited (data not shown). The explained variance is a statistical indication of how much variation in a dataset is attributed to each of the principal components created by the PCA (Kumar, 2022). In this case, the overall explained variance is around 30 %, this pointed out a data dispersion, mainly due to the relevant number of qualitative variables / DoE factors (dehydration rate, harvesting year, presence of peel, variety) related to the number of samples (59).

R^2X is a coefficient that measures the goodness of fit of the model to the real data, its value is from 0 to 1 (Casella & Berger, 2001). In this model, the coefficient was 0.779, highlighting valuable goodness of fit, despite the abovementioned data dispersion. The main aim of the PCA is to reduce the data dimensionality, in order to make the visualisation possible on a 2D plot and to extract the features. It is an unsupervised method, as it does not label the classes, so the observations are placed into the plot only according to the variables that drive their position (Ghojogh & Crowley, 2022). Supervised analysis, such as PLS-DA, by labelling the groups, allows the achievement of a good clusterisation, decreasing the data dispersion. Figure 4 shows the PLS-DA score plot, underlining a better class separation, even though it is possible to discern only between Italian vs non-Italian groups. For supervised models another parameter had to be considered, the Q^2, related to the goodness of prediction or predicted variation (Nepomuceno, Cruz Junho, Carneiro-Ramos, & da Silva Martinho, 2021). It is the R^2 when the PLS built on a training set is applied on a test set. In this model, R^2 was 0.716 and Q^2 was 0.566. The last model applied was the Orthogonal PLS-DA (OPLS-DA), which combines the Orthogonal Signal Correction (OSC), and the PLS-DA. OSC removes the info from the X-block (independent variables), orthogonal to the Y-block (dependent variables) (Beckwith-Hall, et al., 2002). Fig. 5 shows the OPLS-DA score plot, where it is possible to see a multi-class separation. Therefore, with the PCA there was a preliminary visualisation of how the observations started to get clustered, then the separation became finer with the supervised PLS-DA, and the OPLS-DA permitted to discriminate even among all the geographical locations considered. In this last model, R^2 was 0.764, while Q^2 was 0.593. Since these models, with the class labelling and the OSC, could have forced the separation, their goodness was evaluated through the permutation test. It allows for permutation (randomly assigning) the data to the classes, and the model has then performed again. The permutation could bring to a wrong class assignment, hence R^2(cum) and Q^2(cum) should be lower than the original model values (Triba, et al., 2015). Figure 6 reports the permutation plot obtained from the test to assess the OPLS-DA model goodness.

R^2 and Q^2 values were, respectively, (0.0, 0.201) and (0.0, -0.418). This result confirms the validity of the supervised model, as the parameters from the permutation, were considerably lower than the ones from the original model. Finally, a misclassification table was created as well. It consists in applying the model, built on the training set, on a prediction set, to assess its accuracy. In this study, 17 samples were extracted from the training set and used as a prediction set. Only one sample was misclassified between French and Hungarian classes, and an accuracy score of 94.12 % for the supervised model was obtained. (Table 1)

To further assess the robustness of the predictive models, and also its applicability in a real situation, a test set was analysed. The workflow adopted was the same for the training set, as well as the manual

area set and the matrix export (the samples were analysed in duplicate). Besides Italian samples from the 2022 campaign, some other Hungarian samples from the 2021 crop were included, and samples from a different region were not considered in the training set, Chile. Thus, the test set was used as a prediction set but considering only the Italian and non-Italian classes. This was useful to simulate a real industrial approach, where a sample is analysed, and if it belongs to the Italian class, it is evaluated as authentic, otherwise, it is discarded or, eventually, analysed with a confirmatory/high-res technique. A Classification list was performed on the test set, it displayed the observation (sample IDs), the original dummy variables as YVarPS, which can range from 1 to 0, and the predicted dummy variables as YPredPS. From this value, it is possible to define the sample class:

- <0.35 the samples do not belong to the class
- Between 0.35 and 0.65 the samples are borderline
- >0.65 the samples belong to the class

(MKS Umetrics, 2015). In this study, only samples having a YPredPS bigger than 0.65 were considered Italian. Table 2 reports the Classification List outcomes, highlighting an interesting predictive ability of GC-IMS on this specific application.

All the Italian samples were correctly classified, whereas one replicate of a Chilean sample and two replicates of the same Hungarian sample were misclassified as Italian. A Misclassification table was also created to evaluate the model accuracy score, which was around 78 %. (Table 3)

6 Chilean samples were assigned to the “No class” group. This could be due to the absence of Chilean samples in the training set; hence, it was reasonable to have borderline YPredPS values, without an authentic sample as a reference. This led to the “No class” group classification. However, only 3 measurements were wrongly assigned, confirming the GC-IMS technology as a reliable analytical strategy for this type of matrix.

CONCLUSIONS

The present work describes the application of untargeted GC-IMS fingerprinting, a recent analytical approach, for the geographical origin evaluation of dehydrated apple samples. The developed method allowed us to obtain a relevant number of features, with good repeatability. The chemometric elaboration brought valuable results, Italian and non-Italian samples were well separated, employing both unsupervised and supervised models. The predictive ability was assessed on the training and the test set as well, showing acceptable accuracy scores. This strategy could be exploited at both industrial and academic levels, since it could be used for research purposes, but also as a rapid and direct technique in routine laboratories, or, potentially, online, directly on the production chain, to discard anomalous samples, or to go deeper with other high-res approaches.

DECLARATIONS

Ethics approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Competing interest

The authors declare no competing interests.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Giuseppe Sammarco. The first draft of the manuscript was written by Giuseppe Sammarco and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

No funding was received to assist with the preparation of this manuscript.

Availability of data and materials

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

Aguzzoni, A., Bassi, M., Pignotti, E., Robatscher, P., Scandellari, F., Tirler, W., & Tagliavini, M. (2020). Sr isotope composition of Golden Delicious apples in Northern Italy reflects the soil 87Sr/86Sr ratio of the cultivation area. *Journal of The Science of Food and Agriculture*, 3666-3674.

Atlas Big. (2021). *World Apple Production by Country*. Tratto da Atlas Big: https://www.atlasbig.com/en-cn/countries-by-apple-production#:~:text=China%20is%20the%20largest%20apple,50%20%25%20of%20world's%20total%20apple.

Bat, K., Vidrih, R., Necemer, N., Vodopivec, B., Mulic, I., Kump, P., & Ogrinc, N. (2012). Characterization of Slovenian Apples with Respect to Their Botanical and Geographical Origin and Agricultural Production Practice. *Food Technology and Biotechnology*, 107-116.

Beckwith-Hall, B. M., Brindle, J. T., Barton, R. H., Coen, M., Holmes, E., Nicholson, J. K., & Antti, H. (2002). Application of orthogonal signal correction to minimise the effects of physical and biological variation in high resolution 1H NMR spectra of biofluids. *The Analyst*, 1283-1288.

Casella, G., & Berger, R. L. (2001). *Statistical Inference (second edition)*. Pacific Groove, California, US: Duxbury/Thomson Learning.

Feng, L., Wu, B., Zhu, S., He, Y., & Zhang, C. (2021). Application of Visible/Infrared Spectroscopy and Hyperspectral Imaging With Machine Learning Techniques for Identifying Food Varieties and Geographical Origins. *Nutrition and Food Science Technology*, 680357.

Feng, T., Sun, J., Song, S., Wang, H., Yao, L., Sun, M., . . . Chen, D. (2022). Geographical differentiation of Molixiang table grapes grown in China based on volatile compounds analysis by HS-GC-IMS coupled with PCA and sensory evaluation of the grapes. *Food Chemistry: X*, 100423.

Feng, X., Wang, H., Wang, Z., Huang, P., & Kan, J. (2022). Discrimination and characterization of the volatile organic compounds in eight kinds of huajiao with geographical indication of China using electronic nose, HS-GC-IMS and HS-SPME-GC–MS. *Food Chemistry*, 131671.

Gerhardt, N., Birkenmeier, M., Sanders, D., Rohn, S., & Weller, P. (2017). Resolution-optimized headspace gas chromatography-ion mobility spectrometry (HS-GC-IMS) for non-targeted olive oil profiling. *Analytical and Bioanalytical Chemistry*, 3933-3942.

Gerhardt, N., Birkenmeier, M., Schwolow, S., Rohn, S., & Weller, P. (2018). Volatile-Compound Fingerprinting by Headspace-Gas-Chromatography Ion-Mobility Spectrometry (HS-GC-IMS) as a Benchtop Alternative to 1H NMR Profiling for Assessment of the Authenticity of Honey. *Analytical Chemistry*, 1777–1785.

Ghojogh, B., & Crowley, M. (2022). Unsupervised and Supervised Principal Component Analysis: Tutorial. *arXiv*.

Global Product Prices. (2022). *Apples - prices by country, around the world, June 2022*. Tratto da Global Product Prices: https://www.globalproductprices.com/rankings/apple_prices/

Guo, J., Yue, T., & Yuan, Y. (2012). Feature selection and recognition from nonspecific volatile profiles for discrimination of apple juices according to variety and geographical origin. *Journal of Food Science*, 1090-1096.

Kumar, A. (2022, August 11). *PCA Explained Variance Concepts with Python Example*. Tratto da VitalFlux: https://vitalflux.com/pca-explained-variance-concept-python-example/#:~:text=Conclusion-,What%20is%20Explained%20Variance%3F,component%20analysis%20(PCA)%20method.

Li, C., Li., L., Wu, Y., Lu, M., Yang, Y., & Li, L. (2018). Apple Variety Identification Using Near-Infrared Spectroscopy. *Journal of Spectroscopy*.

MKS Umetrics. (2015, January 26). *Dynacentrix*. Tratto da https://www.dynacentrix.com/telecharg/SimcaP/SIMCA14_User_Guide.pdf

Nepomuceno, G., Cruz Junho, C. V., Carneiro-Ramos, M. S., & da Silva Martinho, H. (2021). Partial least squares—discriminant analysis (PLS-DA). *Scientific Reports*.

Ruszkiewicz, D., Myers, R., Henderson, B., Yusof, H., Meister, A., Moreno, S., . . . Thomas, C. (2022). Peppermint protocol: first results for gas chromatography-ion mobility spectrometry. *Journal of Breath Research*, 036004.

Schmutzler, M., & Huck, C. (2014). Automatic sample rotation for simultaneous determination ofgeographical origin and quality characteristics of apples based on nearinfrared spectroscopy (NIRS). *Vibrational Spectroscopy*, 97-104.

Schwolow, S., Gerhardt, N., Rohn, S., & Weller, P. (2019). Data fusion of GC-IMS data and FT-MIR spectra for the authentication of olive oils and honeys—is it worth to go the extra mile? *Analytical and Bioanalytical Chemistry*, 6005-6019.

Triba, M. N., Le Moyec, L., Amanthieu, R., Goossens, C., Bouchemal, N., Nahon, P., . . . Savarin, P. (2015). PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. *Molecular BioSystems*, 13-19.

Vera, L., Companioni, E., Meacham, A., & Gygax, H. (2016). Real Time Monitoring of VOC and Odours Based on GC-IMS at Wastewater Treatment Plants. *Chemical Engineering Transactions*, 79-84.

World's Top Exports. (2021). *Apples Exports by Country Plus Average Prices*. Tratto da World's Top Exports: https://www.worldstopexports.com/apples-exports-by-country/

TABLES

	Members	Correct	Italy	China	France	Hungary	Poland	No class
Italy	4	100%	4	0	0	0	0	0
China	3	100%	0	3	0	0	0	0
France	5	100%	0	0	5	0	0	0
Hungary	3	66.67%	0	0	1	2	0	0
Poland	2	100%	0	0	0	0	2	0
No class	0		0	0	0	0	0	0
Total	17	94.12%	4	3	6	2	2	0

Table 1 *Misclassification table of a selected prediction set (n=17). (Green cell: samples correctly classified, yellow cell: samples misclassified)*

Sample ID	Class ID	YVarPS (ITA)	YPredPS (ITA)	YVarPS (NOT ITA)	YPredPS (NOT ITA)
Chile 1 A	Not Italy	0	0.465	1	0.535
Chile 1 B	Not Italy	0	0.334	1	0.666
Chile 2 A	Not Italy	0	0.351	1	0.649
Chile 2 B	Not Italy	0	0.561	1	0.439
Chile 3 A	Not Italy	0	0.598	1	0.402
Chile 3 B	Not Italy	0	0.742	1	0.258
Chile 4 A	Not Italy	0	0.524	1	0.476
Chile 4 B	Not Italy	0	0.607	1	0.393
Italy 1 A	Italy	1	1.107	0	-0.107
Italy 1 B	Italy	1	1.080	0	-0.080
Italy 2 A	Italy	1	1.306	0	-0.306
Italy 2 B	Italy	1	1.229	0	-0.229
Italy 3 A	Italy	1	1.180	0	-0.180
Italy 3 B	Italy	1	1.258	0	-0.258
Italy 4 A	Italy	1	1.317	0	-0.317
Italy 4 B	Italy	1	1.379	0	-0.379
Hungary 1 A	Not Italy	0	1.163	1	-0.163
Hungary 1 B	Not Italy	0	0.993	1	0.007
Hungary 2 A	Not Italy	0	0.004	1	0.996
Hungary 2 B	Not Italy	0	-0.048	1	1.048
Hungary 3 A	Not Italy	0	0.102	1	0.898
Hungary 3 B	Not Italy	0	0.118	1	0.882
Hungary 4 A	Not Italy	0	0.168	1	0.832
Hungary 4 B	Not Italy	0	0.142	1	0.858

Table 2 *Dehydrated apple test set Classification List. The numbers in the Sample ID indicate different batches, whereas the letters indicate the replicate. (Green YPredPS value: sample belongs to the class, yellow YPredPS value: sample is borderline, white YPredPS value: sample does not belong to the class)*

	Members	**Correct**	**Italy**	**Not Italy**	**No class**
Italy	8	100%	8	0	0
Not Italy	16	56.25%	3	7	6
No class	0		0	0	0
Total	17	78.12%	11	7	6

Table 3 *Dehydrated apple test set Misclassification Table*.

FIGURES

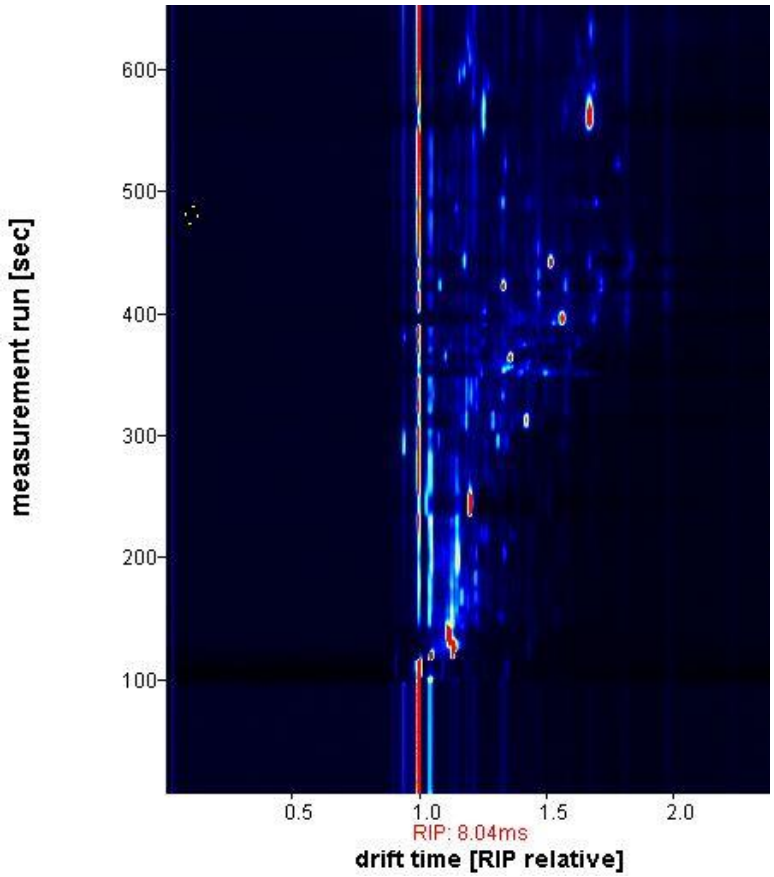


Fig. 1 *2D GC-IMS chromatogram of dehydrated apple samples*

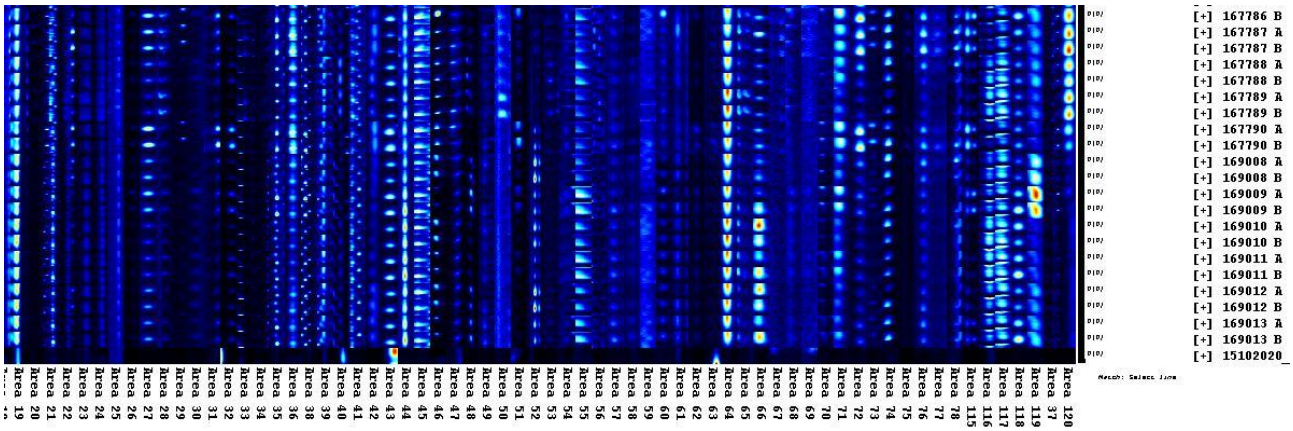


Fig. 2 *Part of the Galerie plot window, showing all the observation areas in all the samples of the project.*

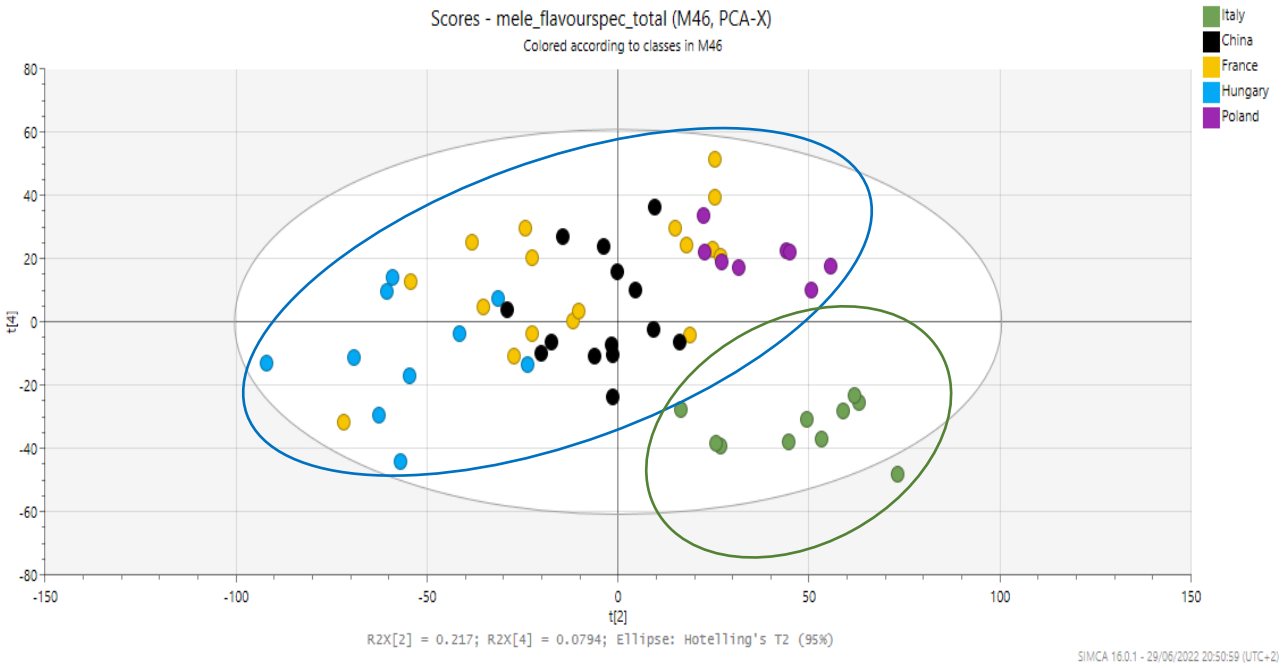


Fig. 3 *PCA score plot of the dehydrated apple sample set. (Green dots: Italy, black dots: China, yellow dots: France, turquoise dots: Hungary, purple dots: Poland. Green group: Italian samples, turquoise group: non-Italian samples)*

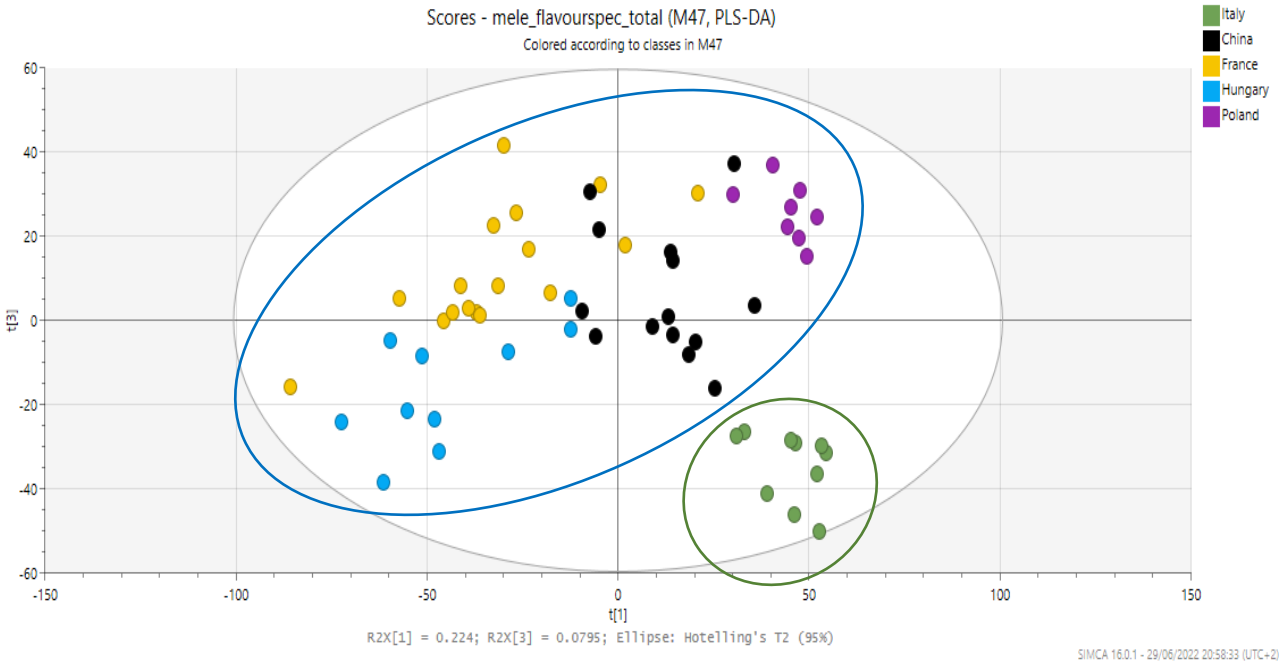


Fig. 4 *PLS-DA score plot of the dehydrated apple sample set. (Green dots: Italy, black dots: China, yellow dots: France, turquoise dots: Hungary, purple dots: Poland. Green group: Italian samples, turquoise group: non-Italian samples)*

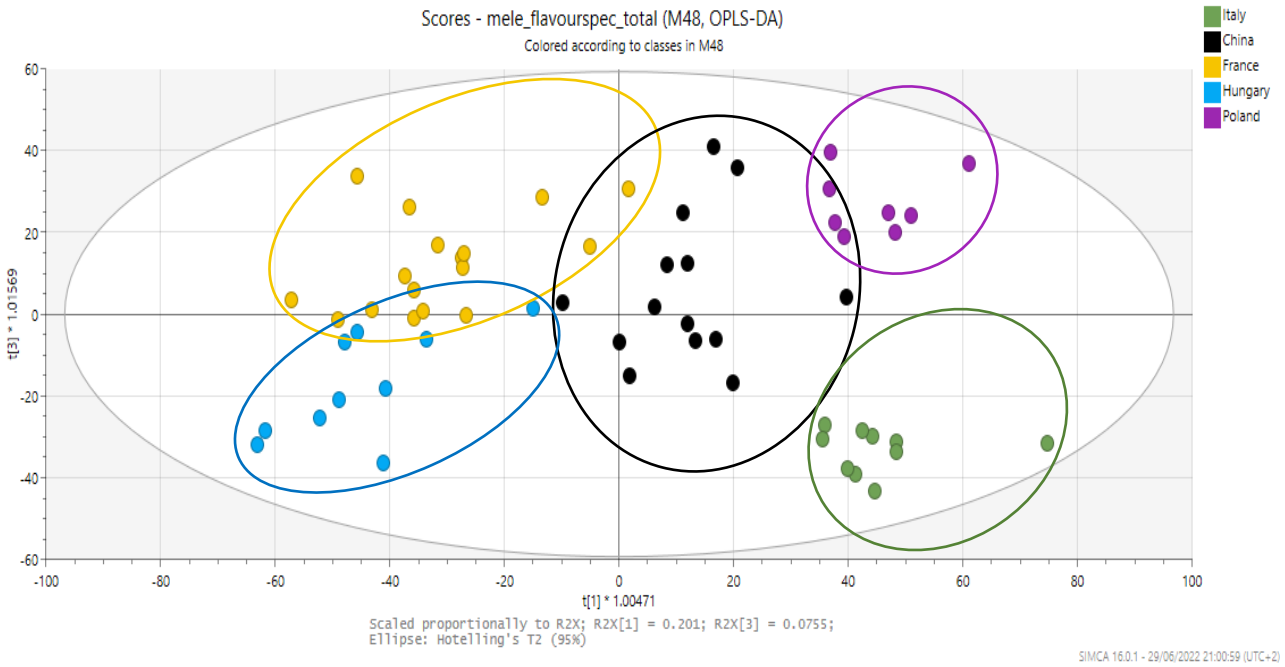


Fig. 5 *OPLS-DA score plot of the dehydrated apple sample set. (Green dots: Italy, black dots: China, yellow dots: France, turquoise dots: Hungary, purple dots: Poland. Green group: Italian samples, black group: Chinese samples, yellow group: French samples, turquoise group: Hungarian samples, purple group: Polish samples)*

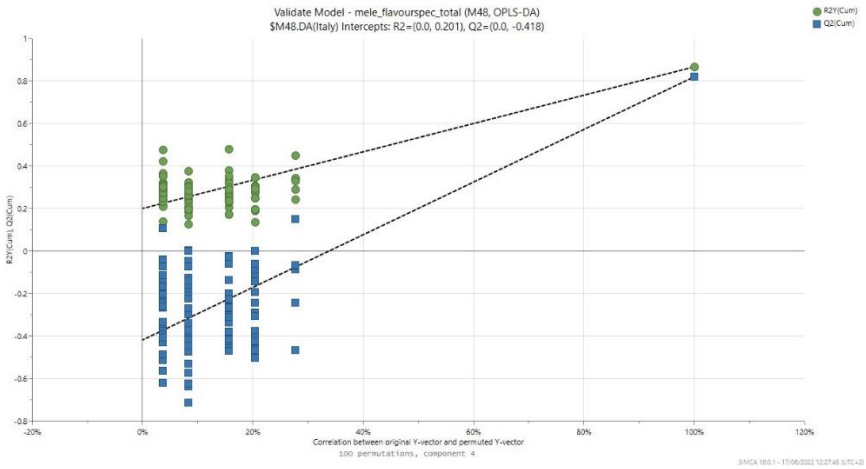


Fig. 6 *Permutation plot for the OPLS-DA model evaluation. Number of permutations=100 (Green dots: $R^2Y(cum)$, blue squares: $Q^2(cum)$)*

SUPPLEMENTARY MATERIALS

Sample ID	Year	Variety	Dehydration rate	Presence of peel	Origin
122931	2020	Pink Lady	Total	Yes	Italy
122932	2020	Pink Lady	Partial	No	Italy
122933	2020	Pink Lady	Partial	No	China
122934	2020	Pink Lady	Total	Yes	China
122935	2020	Pink Lady	Partial	Yes	China
122936	2020	Pink Lady	Total	No	China
123006	2020	Fuji	Total	No	Italy
123007	2020	Granny Smith	Partial	No	Italy
123010	2020	Fuji	Total	Yes	France
123015	2020	Granny Smith	Total	Yes	France
123017	2020	Fuji	Partial	No	France
123019	2020	Granny Smith	Partial	No	France
123021	2020	Fuji	Partial	Yes	France
123022	2020	Granny Smith	Total	No	France
123023	2020	Fuji	Total	No	France
123025	2020	Granny Smith	Partial	Yes	France
166460	2021	Pink Lady	Total	No	France
166466	2021	Golden Delicious	Total	No	France
166467	2021	Granny Smith	Partial	Yes	France
166468	2021	Fuji	Total	No	France
166469	2021	Golden Delicious	Partial	No	France
166471	2021	Pink Lady	Partial	No	France
166472	2021	Pink Lady	Partial	Yes	France
166473	2021	Pink Lady	Total	Yes	France
166474	2021	Golden Delicious	Total	Yes	France
166476	2021	Pink Lady	Partial	No	Hungary
166479	2021	Fuji	Total	No	Hungary
166480	2021	Golden Delicious	Total	Yes	Hungary
166481	2021	Pink Lady	Total	Yes	Hungary
166482	2021	Pink Lady	Total	No	Hungary
166483	2021	Granny Smith	Total	Yes	Hungary
166484	2021	Golden Delicious	Partial	No	Hungary
166486	2021	Fuji	Partial	Yes	Hungary
166487	2021	Granny Smith	Total	Yes	Hungary
166490	2021	Fuji	Total	Yes	Hungary
167047	2021	Fuji	Total	Yes	China
167048	2021	Fuji	Partial	Yes	China

167052	2021	Golden Delicious	Total	Yes	China
167053	2021	Fuji	Total	No	China
167054	2021	Granny Smith	Total	No	China
167057	2021	Golden Delicious	Partial	Yes	China
167059	2021	Golden Delicious	Total	No	China
167060	2021	Golden Delicious	Partial	No	China
167061	2021	Granny Smith	Partial	No	China
167062	2021	Fuji	Partial	No	China
167063	2021	Golden Delicious	Partial	No	Poland
167064	2021	Golden Delicious	Total	Yes	Poland
167065	2021	Fuji	Partial	No	Poland
167066	2021	Fuji	Total	Yes	Poland
167067	2021	Fuji	Total	No	Poland
167068	2021	Fuji	Partial		Poland
167069	2021	Golden Delicious	Total	No	Poland
167070	2021	Golden Delicious	Partial	Yes	Poland
167785	2021	Golden Delicious	Partial	No	Italy
167788	2021	Golden Delicious	Total	No	Italy
167790	2021	Granny Smith	Total	No	Italy
169008	2021	Granny Smith	Total	Yes	Italy
169011	2021	Fuji	Partial	Yes	Italy
169013	2021	Pink Lady	Partial	Yes	Italy

Table S1 *Dehydrated apple samples training set, according to the DoE*

Sample ID	Year	Variety	Dehydration rate	Presence of peel	Origin
Chile 1	2022	Pink Lady	Total	Yes	Chile
Chile 2	2022	Pink Lady	Partial	Yes	Chile
Chile 3	2022	Pink Lady	Total	No	Chile
Chile 4	2022	Pink Lady	Partial	No	Chile
Italy 1	2022	Pink Lady	Total	Yes	Italy
Italy 2	2022	Pink Lady	Total	No	Italy
Italy 3	2022	Granny Smith	Total	Yes	Italy
Italy 4	2022	Granny Smith	Total	No	Italy
Hungary 1	2021	Granny Smith	Total	Yes	Hungary
Hungary 2	2021	Granny Smith	Partial	Yes	Hungary
Hungary 3	2021	Granny Smith	Total	No	Hungary
Hungary 4	2021	Granny Smith	Partial	Yes	Hungary

Table S2 *Dehydrated apple samples test set, according to the DoE*

Area label	IMS center [Dt/a.u.]	IMS width [Dt/a.u.]	Spectra Start [#n]	Spectra End [#n]
Area 1	1.148	0.106	484.000	874.000
Area 2	1.128	0.053	834.000	1022.000
Area 3	1.140	0.108	983.000	1938.000
Area 4	1.280	0.053	914.000	1143.000
Area 5	1.422	0.066	847.000	1089.000
Area 6	1.348	0.056	753.000	982.000
Area 7	1.518	0.040	2462.000	2758.000
Area 8	1.571	0.060	1857.000	2058.000
Area 9	1.596	0.050	2085.000	2355.000
Area 10	1.051	0.056	632.000	2570.000
Area 11	1.136	0.050	1897.000	2597.000
Area 12	1.054	0.056	2570.000	5814.000
Area 13	1.295	0.089	5464.000	6096.000
Area 14	1.323	0.073	2422.000	2920.000
Area 15	1.227	0.053	4225.000	4589.000
Area 16	1.743	0.053	5585.000	5975.000
Area 17	1.687	0.060	5370.000	5841.000
Area 18	1.217	0.046	1789.000	2139.000
Area 19	1.317	0.053	1749.000	2112.000
Area 20	1.366	0.040	1708.000	2099.000
Area 21	1.130	0.063	2597.000	5706.000
Area 22	1.247	0.099	497.000	1103.000
Area 23	1.221	0.086	5612.000	6218.000
Area 24	1.226	0.089	4602.000	5343.000
Area 25	1.214	0.060	3418.000	4185.000
Area 26	1.621	0.073	3458.000	3983.000
Area 27	1.051	0.050	5814.000	6204.000
Area 28	1.613	0.056	5423.000	5841.000
Area 29	1.303	0.060	4091.000	4629.000
Area 30	1.219	0.063	3028.000	3391.000
Area 31	1.689	0.050	3876.000	4306.000
Area 32	1.224	0.053	2462.000	2987.000
Area 33	1.506	0.052	1780.000	2195.000
Area 34	1.601	0.073	1587.000	1803.000
Area 35	1.621	0.079	2906.000	3431.000
Area 36	1.573	0.076	1332.000	1668.000
Area 37	1.138	0.082	514.000	726.000
Area 38	1.324	0.042	654.000	766.000
Area 39	1.342	0.040	888.000	1039.000
Area 40	1.258	0.059	984.000	1123.000
Area 41	1.210	0.047	900.000	1028.000
Area 42	1.203	0.042	721.000	855.000
Area 43	1.262	0.040	1190.000	1336.000

Area 44	1.427	0.038	1414.000	1554.000
Area 45	1.305	0.059	1397.000	1554.000
Area 46	1.499	0.063	1615.000	1744.000
Area 47	1.562	0.054	1788.000	1967.000
Area 48	1.342	0.077	1923.000	2101.000
Area 49	1.338	0.042	2269.000	2398.000
Area 50	1.182	0.052	2783.000	2996.000
Area 51	1.088	0.038	1951.000	2129.000
Area 52	1.186	0.038	2062.000	2174.000
Area 53	1.260	0.035	1855.000	1990.000
Area 54	1.212	0.047	1520.000	1677.000
Area 55	1.193	0.047	1436.000	1531.000
Area 56	1.130	0.047	1503.000	1671.000
Area 57	1.046	0.056	710.000	1693.000
Area 58	1.049	0.049	1710.000	1883.000
Area 59	1.046	0.033	1906.000	2006.000
Area 60	1.046	0.042	2336.000	2716.000
Area 61	1.042	0.054	2057.000	2331.000
Area 62	1.520	0.040	2051.000	2152.000
Area 63	1.429	0.026	2029.000	2185.000
Area 64	1.418	0.080	1604.000	1755.000
Area 65	1.437	0.047	2286.000	2392.000
Area 66	1.701	0.033	2275.000	2370.000
Area 67	1.563	0.056	2599.000	2789.000
Area 68	1.666	0.047	2593.000	2767.000
Area 69	1.131	0.045	687.000	995.000
Area 70	1.130	0.061	1000.000	1470.000
Area 71	1.127	0.059	2124.000	2347.000
Area 72	1.086	0.028	2331.000	2688.000
Area 73	1.123	0.066	1710.000	1889.000
Area 74	1.480	0.068	1822.000	1973.000
Area 75	1.189	0.047	1202.000	1364.000
Area 76	1.252	0.056	1649.000	1850.000
Area 77	1.334	0.038	2169.000	2252.000
Area 78	1.258	0.026	2593.000	2878.000
Area 79	1.346	0.056	2437.000	2577.000
Area 80	1.409	0.047	1811.000	2006.000
Area 81	1.051	0.042	508.000	620.000
Area 82	1.048	0.033	453.000	503.000
Area 83	1.402	0.080	1157.000	1319.000
Area 84	1.588	0.054	2006.000	2180.000
Area 85	1.555	0.068	2191.000	2308.000
Area 86	1.206	0.035	1106.000	1185.000
Area 87	1.210	0.084	2286.000	2431.000
Area 88	1.650	0.056	2152.000	2291.000
Area 89	1.268	0.042	676.000	771.000
Area 90	1.237	0.016	637.000	788.000
Area 91	1.272	0.063	816.000	956.000

Area 92	1.636	0.042	1744.000	1855.000
Area 93	1.677	0.035	2057.000	2135.000
Area 94	1.594	0.033	1688.000	1783.000
Area 95	1.408	0.045	2823.000	2973.000
Area 96	1.485	0.054	2873.000	3024.000
Area 97	1.692	0.052	2744.000	2856.000
Area 98	1.626	0.047	2845.000	3013.000
Area 99	1.684	0.045	2929.000	3024.000
Area 100	1.651	0.073	1632.000	1716.000
Area 101	1.588	0.030	1481.000	1604.000
Area 102	1.550	0.040	1039.000	1168.000
Area 103	1.405	0.040	939.000	1039.000
Area 104	1.438	0.040	1045.000	1146.000
Area 105	1.378	0.038	1369.000	1475.000
Area 106	1.351	0.026	1341.000	1447.000
Area 107	1.115	0.026	2353.000	2482.000
Area 108	1.655	0.033	1951.000	2034.000
Area 109	1.094	0.030	453.000	525.000
Area 110	1.479	0.056	2644.000	2789.000
Area 111	1.127	0.035	2649.000	2828.000
Area 112	1.437	0.033	2644.000	2800.000
Area 113	1.334	0.052	453.000	531.000
Area 114	1.378	0.048	2127.000	2219.000
Area 115	1.468	0.031	1970.000	2209.000
Area 116	1.724	0.030	1949.000	2140.000
Area 117	1.791	0.036	2428.000	2564.000
Area 118	1.318	0.023	1140.000	1232.000
Area 119	1.331	0.089	1626.000	1855.000
Area 120	1.333	0.032	1544.000	1623.000

Table S3 *GC-IMS area set integration parameters.*

CHAPTER 3

IRMS ANALYSIS

IRMS Analysis

General Overview

IRMS technology is nowadays one of the most suitable analytical strategies to assess food and beverage authenticity, starting from the accurate determination of stable isotope ratios inside the analysed products (Kelly, Brodie, & Hilkert, 2018). Thus, this technique shows interesting applications in the food field, as it could be adopted for analysing a wide range of products, generating crucial information about several food components, especially when hyphenated with separation technologies, such as gas chromatography. Among possible examples of the use of IRMS data, there are the determination of synthetic or natural ingredients and fertilizers, the substitution of one or more ingredient(s), and the origin of foods and their ingredients. All these aspects make the IRMS a very powerful strategy to detect fraudulent and unsafe procedures, giving a remarkable contribution to authenticating products of known origin (van Leeuwen, Prenzler, Ryan, & Camin, 2014). Further, IRMS allows the scientists to carry out bulk analysis, converting all the biomass, optimising the time and the costs of the sample preparation, and following a green approach, as no extraction or similar treatments have to be done (Forstel, 2007).

Despite these points of strength and the versatility of the technology, the scientific literature does not report a relevant number of publications about the IRMS technique, as many studies are performed by private companies, that are not interested in sharing their results (Forstel, 2007). In addition, the lack of extended literature is due to the high instrumentation costs, the routine maintenance, and the expertise required, not only for performing the analyses but also to statistically interpret the raw data obtained.

The goal of the work

This section is aimed at evaluating the geographical origin of Italian hazelnut samples mainly through IRMS technology. Data fusion with orthogonal strategies (ICP-MS and ICP-OES) was also achieved, allowing the authors to create a multivariate statistical model. The following work was submitted to "The *Journal of the Science of Food and Agriculture*". For additional details see the section "Author".

References

Forstel, H. (2007). The natural fingerprint of stable isotopes—use of IRMS to test food authenticity. *Analytical and Bioanalytical Chemistry*, 541-544.

Kelly, S., Brodie, C., & Hilkert, A. (2018). Isotopic-Spectroscopic Technique: Stable Isotope-Ratio Mass Spectrometry (IRMS). In D. Sun, *Modern Techniques for Food Authentication (Second Edition)* (pp. 349-413). Academic Press.

van Leeuwen, K. A., Prenzler, P. D., Ryan, D., & Camin, F. (2014). Gas Chromatography-Combustion-Isotope Ratio Mass Spectrometry for Traceability and Authenticity in Foods and Beverages. *Comprehensive Reviews in Food Science and Food Safety*, 814-837.

Hazelnut products traceability through combined Isotope Ratio Mass Spectrometry and Multi-elemental Analysis

Giuseppe Sammarco[1,3], Mattia Rossi[4], Michele Suman[1,2], Daniele Cavanna[5], Laura Viotto[6], Piero Pettenà[6], Chiara Dall'Asta[3], Paola Iacumin[4]*

[1] Advanced Laboratory Research, Barilla G. e R. Fratelli S.p.A., Parma, Italy
[2] Department for Sustainable Food Process, Catholic University Sacred Heart, Piacenza, Italy
[3] Department of Food and Drug, University of Parma, Parma, Italy
[4] Department of Chemical, Life and Environmental Sustainability Sciences, University of Parma, Parma, Italy
[5] LabGo S.r.L. Università di Parma, Parco Area delle Scienze 157/A, 43124 Parma, Italy
[6] Mérieux NutriSciences Italia - Chelab S.r.l., Via Fratta 25, 31023 Resana (TV), Italy

*Corresponding Author: Prof. Dr. Michele Suman, Ph.D
Barilla G. R. F.lli SpA, Advanced Laboratory Research - via Mantova 166, Parma (Italy)
Catholic University Sacred Heart, Department for Sustainable Food Process - Piacenza (Italy)
Phone: +39 0521 262332
E-mail: michele.suman@barilla.com

Highlights

- False origin declaration represents a common fraud for the Italian hazelnuts
- IRMS is successfully applied for the geographical assessment of different matrices
- $\delta\ ^{13}C$ and $\delta\ ^{18}O$ differ in samples, according to the soil and climatic conditions
- IRMS data are positively merged with ICP-OES and ICP-MS ones

Abstract

Isotope Ratio Mass Spectrometry (IRMS) could play a key role in origin discrimination. The present study aims to assess the provenience of Italian hazelnuts, by analysing relative isotopic ratios of carbon and oxygen. Method development is performed by evaluating samples' repeatability, reproducibility, and robustness. The results are reproducible and robust, having acceptable standard deviations. One-way ANOVA demonstrates the significant statistical difference between Italian and non-Italian samples. Furthermore, a data fusion approach, with Inductively Coupled Plasma–Optical Emission Spectroscopy (ICP-OES) and Inductively Coupled Plasma-Mass Spectrometry (ICP-MS), permitted to build of multivariate statistical models to confirm the differences of geographical provenience. A Design of Experiment (DoE) is created to sample correctly, considering factors such as variety, processing, and peel percentage. N=96 hazelnut lots, from Italy, Turkey, Georgia, and Azerbaijan are analysed for the geographical assessment: this strategy demonstrates promising

potentialities, as food isotopic abundances reflect ground and climate-related features, typical of precise locations.

Keywords

EA-IRMS, TC-IRMS, geographical origin, hazelnuts, multi-elemental analysis, data fusion

INTRODUCTION

Hazelnut (*Corylus avellana* L.) represents one of the most economically relevant matrices for the bakery and chocolate industries. Ca. 90 % of the global crop is indeed exploited by the food industry. (Moscetti, Radicetti, Monarca, Cecchini, & Massantini, Near infrared spectroscopy is suitable for the classification of hazelnuts according to Protected Designation of Origin, 2015) According to the statistics database of the International Nut and Dried Fruit Council, in the 2019/2020 season, Turkey is the leading hazelnut producer in the world, ca. 75 % of the global production, followed by Italy (7 %), Azerbaijan (3 %), the USA (3 %), Georgia (2 %), and others. (https://www.nutfruit.org/industry/statistics) The authentication of cultivars from different locations is fundamental for the determination of hazelnut provenance. For instance, high-quality Italian products, such as "Tonda di Giffoni and "Tonda Gentile delle Langhe", received the Protected Geographical Indication, (PGI) while "Nocciola Romana" is a Protected Designation of Origin. (PDO) (Ciarmiello, et al., 2014). High quality and certifications enhance the market price: in 2014, the highest selling price was for Italian hazelnuts at 5.207 USD/t, Turkish products were sold at an 18 % lower price, followed by the United States (24 %), Georgia (31 %) and Azerbaijan. (49 %). However, the geographical origin has a relative impact as a market parameter, since varieties usually are not indicated in the final product. On the other hand, market fluctuations could be potentially due to the location of harvesting, i.e. crop failure, as well as climatic changes, which may rapidly upset prices. Therefore, these aspects could boost demand for the geographical authentication of hazelnuts and hazelnut products in quality control (Bachmann, Klockmann, Haerdter, Fischer, & Hackl, 1H NMR Spectroscopy for Determination of the Geographical Origin of Hazelnuts, 2018). Several techniques have been applied for the geographical assessment of hazelnuts and hazelnut products. ^{1}H Nuclear Magnetic Resonance (NMR) spectroscopy studies on hazelnuts from different European countries and harvest years yield a location discrimination accuracy of 91 % for the training set and 96 % for the test set (Bachmann, Klockmann, Haerdter, Fischer, & Hackl, 1H NMR Spectroscopy for Determination of the Geographical Origin of Hazelnuts, 2018). Various papers about hazelnut traceability describe a metabolomics approach through Liquid Chromatography-Mass Spectrometry. (LC-MS) Klockmann *et al.* employed Ultraperformance LC-Electrospray Ionisation-Quadrupole

Time of Flight-MS, (UPLC-ESI-QToF-MS) merged with the multivariate statistics, as an untargeted tool to gain hazelnuts fingerprints (Klockmann, Reiner, Bachmann, Hackl, & Fischer, 2016). The same research group moved from an untargeted to targeted metabolomics approach, to finely identify metabolites responsible for the geographical differentiation, by exploiting LC-ESI-triple quadrupole (QqQ)-MS, (LC-ESI-QqQ-MS) multivariate statistics, and a prediction model using support vector machine (SVM) classification (Klockmann, Reiner, Cain, & Fischer, 2017). Volatile compounds were considered for a geographical classification as well. Cordero *et al.* worked on hazelnuts from nine different locations, comparably roasted for texture and flavour requirements, sampled by headspace-solid phase microextraction (HS-SPME), and then analysed by Gas Chromatography x Gas Chromatography-quantitative Mass Spectrometry (GCxGC-qMS). Chromatographic fingerprinting and Comprehensive Template Matching were used for data processing, highlighting the usefulness of fingerprinting for a sample comparison, and the comprehensive approach for specificity and quality assessment of the products (Cordero, et al., 2010). Rapid techniques, such as Near Infrared (NIR) spectroscopy, were also employed. The classification of hazelnuts according to PDO certification is feasible through NIR spectroscopy. In particular, a study focused on the Italian "Nocciola Romana", showed classification performance rates with 96 % specificities, 95 % sensitivities, and 95.5 % accuracies (Moscetti, Radicetti, Monarca, Cecchini, & Massantini, Near infrared spectroscopy is suitable for the classification of hazelnuts according to Protected Designation of Origin, 2015). The same tool was exploited, testing two different classification approaches, the discriminant Partial Least Squares-Discriminant Analysis (PLS-DA), and the class-modeling Soft Independent Modeling of Class Analogies, (SIMCA) both having high values of accuracy, sensitivity, and specificity (Biancolillo, et al., Authentication of an Italian PDO hazelnut ("Nocciola Romana") by NIR spectroscopy, 2018). Isotope Ratio Mass Spectrometry (IRMS) is a technique used to provide information on the geographic, chemical, and biological origins of commodities. The ability to gain this information is due to the relative isotopic abundances of the elements which comprise the matrix (Muccio & Jackson, 2009). The $\delta\,^{2}$H- and $\delta\,^{18}$O-values (‰) of plant and plant-derived food samples, for instance, give insights into their geographical origin, since water is the exclusive source of hydrogen for all organic compounds in nature. The $\delta\,^{15}$N- and $\delta\,^{13}$C-values (‰) also could provide relevant information regarding soil composition and growing conditions, useful for the isotopic fingerprint of samples from different geographical regions. (Krauß, Vieweg, & Vetter, 2020) In this paper, we use IRMS techniques (stable isotopes of oxygen and carbon), Inductively Coupled Plasma-Optical Emission Spectroscopy (ICP-OES) (Ca, K), and Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) (Fe, Mn) to assess the Italian geographical origin of raw, roasted, and paste

hazelnuts. This is done by comparing the data obtained on Italian, Turkish, Georgian, and Azerbaijani samples.

MATERIALS AND METHODS

Sampling

Raw, roasted, peeled roasted, and paste hazelnuts were sampled from different Italian regions, PGI "Tonda Gentile delle Langhe" from Piedmont, PDO "Nocciola Romana" from Lazio, and "Mortarella" from Campania. For each product, these three varieties were mixed to have Italian samples (N=24, 6 raw, 6 roasted, 6 peeled roasted and 6 paste of hazelnuts) The same number of lots were from Turkey, Azerbaijan, and Georgia, for a total of 72 not Italian samples, considering both the 2020 and the 2021 harvesting campaigns, and for 2020 also different storage shelf life, the short one, in which the fresh hazelnuts were immediately processed after the crop, and the long one, in which the fresh hazelnuts were stored for ca. 6 months, in controlled atmosphere environment, before being processed. The same lots were used for both $^{13}C/^{12}C$ and $^{18}O/^{16}O$ measurements, and each lot of roasted and paste hazelnuts was obtained by processing the same lots of raw hazelnuts employed for the analyses.

Isotopic and chemical analysis

IRMS-Isotope Ratio Mass Spectrometry

Stable isotope ratios are reported in terms of delta (δ) according to IUPAC (International Union of Pure and Applied Chemistry):

$$\delta\ {}^{a}\mathrm{E} = \frac{{}^{a/b}\mathrm{R}}{{}^{a/b}\mathrm{R_{RM}}} - 1 = 10^{3}\left(\frac{{}^{a/b}\mathrm{R}}{{}^{a/b}\mathrm{R_{RM}}} - 1\right)‰$$

where E is the element of interest (oxygen and carbon in our case), R is the ratio between the abundances (number n of nuclides) of the isotopes ${}^{a}E$ and ${}^{b}E$ in the sample and the international reference material RM, respectively (${}^{a/b}\mathrm{R} = n^{18}\mathrm{O}/\,n^{16}\mathrm{O}$ or $n^{13}\mathrm{C}/\,n^{12}\mathrm{C}$; RM = V-SMOW for oxygen and V-PDB for carbon). The following international standards were used to calibrate the mass spectrometer: mineral oil NBS-22, (IAEA, International Atomic Energy Agency, Vienna, Austria) sucrose IAEA-CH-6, (IAEA), and graphite USGS24 (IAEA) for $\delta\ ^{13}C$ measurements; caffeine IAEA-600, (IAEA) benzoic acid IAEA-601, (IAEA) and sucrose IAEA-CH-6 (IAEA) for $\delta\ ^{18}O$ measurements. Polyethylene IAEA-CH-7, (IAEA) and benzoic acid IAEA-602 (IAEA) were used as Quality Control samples for $\delta\ ^{13}C$ measurements and $\delta\ ^{18}O$ measurements, respectively. Tin (for carbon) and silver (for oxygen) capsules were employed for sample and reference weighting. (Säntis analytical AG, Teufen, Switzerland)

The analyses were carried out using IRMS instrument Delta V Advantage (Thermo Scientific, Bremen, Germany), hyphenated with the Elemental Analyser Flash EA 2000 (Thermo Scientific) in Dynamic Flash Combustion (DFC) mode for $^{13}C/^{12}C$ measurements, and in High-Temperature Conversion (HTC) for $^{18}O/^{16}O$ measurements. Both the modes provide for gas separation with the Separation column (Thermo Scientific), employed at 45°C for $^{13}C/^{12}C$ analysis, and 80°C for $^{18}O/^{16}O$ analysis. Moreover, the device is equipped with the autosampler MAS 200R for solids, (Thermo Scientific) and it is interfaced with the MS via the ConFlo IV, (Thermo Scientific) able to address sample and reference gases.

Raw and roasted hazelnuts were initially minced with Grindomix GM 200 (Retsch GmbH, Haan, Germany) and then homogenised using Precellys Evolution Homogenizer (Bertin Instruments, Montigny-le-Bretonneux, France). About 0.080-0.100 mg of paste hazelnuts were weighed in a tin capsule and directly analysed without any pre-treatment sample. Fifteen $^{13}C/^{12}C$ analyses of a raw hazelnut sample were performed during the same run to evaluate repeatability expressed as standard deviation, SD (SD ≅ 0.12‰); For $^{18}O/^{16}O$ bulk analyses, 0.18-0.20 mg of a raw hazelnut sample were weighed in silver capsules, then they were left for one week in the SuperDry (Totech EU, Netherlands) to avoid moisture interferences during the analyses Afterward, the sample was analysed 10 times obtaining SD ≅ 0.38 ‰.

ICP-OES-Inductively Coupled Plasma-Optical Emission Spectroscopy

ICP-OES analyses, for the detection and quantitation of Ca and K elements, were performed by the iCAP 6300 Radial ICP-OES spectrometer (Thermo Scientific, Bremen, Germany). A grating spectrometer allowed to disperse of the emission spectra of specific elements, whereas a photosensitive medium monitored the intensity of the emission lines. Through a calibration curve is then possible to quantify the content of the elements of interest.

The samples underwent acidic digestion using the microwave MARS6 (CEM Corporation, Charlotte, North Carolina, US). The obtained solutions were nebulised, and the aerosol was conveyed to the plasma through the torch. For the solid matrices, the Limit of Detection (LOD) was 5 mg/kg, while the Limit of Quantification (LOQ) was 25 mg/kg.

ICP-MS-Inductively Coupled Plasma-Mass Spectrometry

ICP-MS analyses, for the detection and quantitation of Mn and Fe elements, were provided by the ICP-MS Agilent 7900 spectrometer, using the MassHunter 4.2 software (Agilent Technologies, Santa Clara, California, US). The sample preparation is analogously done to the one for the ICP-OES analyses, so the sample underwent acidic digestion with the microwave, the obtained solution was

nebulised, and the aerosol was addressed to the plasma through the torch. For the solid matrices, the LOD varies according to the element, for Mn it was 0.0083 mg/kg, and for Fe, it was 0.33 mg/kg. The LOQ values are also element-based, it was 0.02 mg/kg for Mn, and 1 mg/kg for Fe.

RESULTS AND DISCUSSION

IRMS data

Samples were analysed on the same day, for repeatability assessment, and again on another day, for reproducibility evaluation. Some of them underwent a robustness test as well, by varying a parameter not strictly related to the isotopic ratio, such as weight. Table 1 reports the number of data, the $\delta\,^{13}C$ average values, and the standard deviation for each single hazelnut dataset. It is noteworthy, however, that the distribution of the data from the different localities is not normal; thus, the values represent only approximately the distribution around the central values.

Provenance	Number of analyses	10^3 x $^{13}\delta$ Average	10^3 x $^{13}\delta$ Standard deviation
Italy	51	-27.59	0.55
Georgia	51	-28.77	0.46
Azerbaijan	50	-28.79	0.25
Turkey	51	-28.70	0.60

Table 1. *The standard deviation for $\delta^{13}C$ for bulk raw, roasted, peeled roasted, and paste hazelnut from Italy, Georgia, Azerbaijan, and Turkey.*

A one-way ANOVA test was conducted to verify the null hypothesis H_o: $\mu_{Italy} = \mu_{Georgia} = \mu_{Azerbaijan} = \mu_{Turkey}$. The test (Table 2) indicates the null hypothesis is rejected. The tests H_o: $\mu_{Italy} = \mu_{Georgia}$, H_o: $\mu_{Italy} = \mu_{Azerbaijan}$, H_o: $\mu_{Italy} = \mu_{Turkey}$ were also performed: the difference between Italian and not Italian samples was confirmed.

	Sum of squares	Degree of freedom	Mean square	F	p(same)
Between groups	51.69	3	17.23	73.27	$5.75x10^{-32}$
Into groups	46.80	199	0.24		
Total	98.49	202			

Table 2. *ANOVA for the four groups considered.*

The standard deviation values were slightly high. This could be due to the diverse shelf-life storage, as well as two harvesting campaigns could have also impacted this aspect. Table 3 reports all the $^{13}C/^{12}C$, similar values in the samples having the same shelf life, the same treatment (fresh, roasted, paste), harvested in the same campaign, explained the deviations among samples with different 'experimental factors'.

		ITALY	GEORGIA	AZERBAIJAN	TURKEY
Raw hazelnuts	SSS 2020	-27.52	-27.93	-28.72	-28.33
		-27.61	-27.91	-28.75	-28.38
		-27.56	-28.00	-28.79	-28.29
		-27.74	-27.96	-28.69	-28.40
				-28.80	
				-28.78	
				-28.61	
	LSS 2020	-27.95	-28.75	-28.82	-28.83
		-27.95	-28.77	-28.70	-28.07
		-27.28	-28.57	-28.48	-28.00
		-27.40	-28.84	-28.28	-27.69
	2021	-27.03	-29.32	-29.06	-28.98
		-26.71	-29.36	-29.02	-27.93
		-27.11	-28.87	-28.42	-28.54
		-26.82	-29.00	-28.33	-28.19
Roasted hazelnuts	SSS 2020	-27.78	-28.04	-28.79	-28.95
		-27.79	-28.05	-28.78	-28.99
		-27.70	-28.19	-28.92	-28.88
		-27.74	-27.98	-28.82	-29.00
		-27.69			
		-27.73			
		-27.85			
	LSS 2020	-26.99	-28.61	-28.90	-28.80
		-27.10	-28.71	-28.87	-29.07
		-28.23	-28.63	-29.08	-28.48
		-27.95	-28.79	-28.92	-28.58
	2021	-26.17	-29.51	-28.57	-27.91
		-26.80	-29.35	-28.78	-27.91

		-28.34	-28.79	-28.86	-28.96
		-28.17	-28.92	-28.82	-29.05
Peeled roasted hazelnuts	**SSS 2020**	-27.85	-28.55	-29.14	-29.09
		-27.87	-28.54	-29.02	-29.09
		-27.64	-28.52	-28.88	-28.93
		-27.62	-28.55	-29.00	-28.93
			-28.56		
			-28.41		
			-29.95		
	LSS 2020	-28.02	-29.15	-28.79	-28.05
		-28.07	-28.98	-28.31	-28.15
		-27.05	-28.93	-28.33	-27.74
		-27.42	-28.78	-28.85	-27.70
	2021	-26.97	-29.18	-29.06	-28.99
		-27.01	-29.47	-28.28	-29.42
		-27.02	-29.08	-28.34	-28.76
		-27.17	-28.92	-28.50	-28.88
Hazelnut paste	**SSS 2020**	-27.94	-28.63	-29.04	-29.33
		-28.03	-28.73	-29.02	-29.33
		-28.38	-28.61	-29.05	-29.39
		-28.01	-28.68	-29.05	-29.43
					-29.62
					-29.71
					-30.63
	LSS 2020	-27.31	-28.89	-28.98	-28.92
		-27.20	-28.67	-29.04	-28.82
		-29.00	-29.12	-29.21	-28.42
		-29.02	-28.86	-28.83	-28.38
	2021	-27,99	-29,59	-28,89	-28,00
		-27,28	-29,37	-28,45	-27,85
		-27,40	-28,81	-28,86	-28,97
		-27,21	-28,93		-29,19

Table 3. *Measured $10^3 x \delta^{13}C$ values of bulk raw, roasted, peeled roasted, and paste hazelnuts from Italy, Georgia, Azerbaijan, and Turkey (SSS: Short Shelf-life Storage, LSS: Long Shelf-life Storage)*

Italian hazelnuts always have a $\delta^{13}C$ higher than the others, whereas Turkish, Azerbaijani, and Georgian ones have values very close to each other. This could be explained by the areas of harvesting in these countries, as all of them are around the "Black Sea Region" (Figure S1 – Supplementary Materials). Carbon in the hazelnuts mainly derives from the atmospheric CO_2, absorbed by the plant through photosynthesis. A relevant factor that regulates the opening of the stomata is the water stress that the plant undergoes. This mostly depends on the type of soil and the climatic conditions. Low temperatures and high humidity permit the stomata to open, whereas opposite conditions favour their closure. The opening and the closure brought to the carbon isotopic fractionation of CO_2 absorbed by the plant. The closure allows the heaviest ^{13}C, while the opening is for the lightest one (^{12}C). (Lavergne, Sandoval, Hare, Graven, & Prentice, 2020) Therefore, it is reasonable to gain similar $\delta^{13}C$ ‰ in products from these regions.

Different hazelnut matrices present small differences in $\delta\ ^{13}C$ values, as the processing probably impacts the isotopic fractionation since roasting and the paste preparation modify the chemical profile of the matrix. However, besides the deviations of the δ values due to the experimental factors considered, the trend of the values in the class considered (Italy, Georgia, Azerbaijan, Turkey) is the same, justifying the reliability of the technique. The constant δ values difference from Italian and non-Italian samples, independent from the storage shelf life, the harvesting year, and the type of matrix analysed could suggest a relation of the bulk material isotopic ratio with parameters strictly related to each location, and therefore, implicated in the δ differences. Besides the mineralogy and chemistry of the soil, altitude could have relevance. Actually, the $\delta^{13}C$ values frequently increase with altitude. However, Chen *et al.* (2017), who investigated these variations, concluded that not always altitude favours ^{13}C enrichment. Enrichment may depend on other several factors, like soil moisture, air temperature, atmospheric CO_2 concentrations, and leaf morphological and physiological traits. For instance, aridity and decreasing atmospheric pressure have a negative impact on ^{13}C, at high altitudes. (Chen, Wang, & Jia, 2017). Turkish hazelnuts are cultivated in a mountainous area close to the Black Sea, at altitudes from 1500 to 2400 m (an et al., 2020); hazelnuts from Azerbaijan grow mostly in the Zagatala Reserve area, where the altitude ranges from 630 to 3368 m (Enpi East Flag project: www.enpi-fled.org) Georgian products grow at 450-650 m above sea level. (Mirotadze, 2005). The Italian samples are cultivated at a mean altitude of 500 m. The isotopic ratio-altitude comparison was evaluated considering each harvesting year individually since the altitude and the soil conditions could change in every crop. Fig. S2 demonstrates how the samples from the 2020 harvesting campaign present a sort of relation between $^{13}C/^{12}C$ and the altitude, with a lower ratio at high elevations.

Despite there being no linear relation between δ ^{13}C and altitude parameters, it is clear that Italian samples, cultivated at the lowest altitude, present the highest isotopic ratio, whereas Azerbaijani and Turkish samples, harvested at the highest altitudes, have the lowest ratio. Georgian samples did not properly fit this relation, as their δ ^{13}C values are close to the Turkish and Azerbaijani ones, but the altitude of harvesting is more similar to the Italian ones. Hazelnuts from the 2021 crop underwent the same workflow, and the outcomes did not confirm the relation, as in this case, Georgian samples present the lowest isotopic ratio, despite they were cultivated at altitudes not so different from the Italian products. (Fig. S3)

Considering these results, the $^{13}C/^{12}C$ values could not be strictly related to the altitude, while soil conditions might be a potential factor, as the highest values were always found in the Italian hazelnuts, in both the 2020 and 2021 campaigns, and lower values were for the other countries, very similar in both harvesting years, reflecting the 'neighboring' position of these areas, and so an analogue organic composition of the soil.

For δ ^{18}O ‰, the same hazelnut batches were used, and the same asset employed for ^{13}C analyses was performed, (repeatability, reproducibility, and robustness tests) considering higher standard deviations limit for precision assessment. This is due to the different "elemental molecules" considered to gain information about the isotopic ratio. For the $^{13}C/^{12}C$ measurement, CO_2 is the molecule generated by a redox reaction, whereas, for the $^{18}O/^{16}O$ measurement, CO is the molecule obtained by a pyrolysis reaction. These different reactions, to generate the elemental molecules, impact the measurement, as the pyrolysis could not have good CO yields when the reactor is not well conditioned. (Bilke & Mosandl, 2002). Therefore, it is acceptable to have higher standard deviation values. (From 0.70 to 1.20 ‰) Table 4 shows the same summary of values, as analogously done for carbon.

A)

Provenance	Number of analyses	10^3 x $^{18}\delta$ Average	10^3 x $^{18}\delta$ Standard deviation
Italy	45	23.30	1.20
Georgia	48	20.03	0.96
Azerbaijan	47	21.20	1.04
Turkey	50	19.41	0.71

B)

	ITALY	**GEORGIA**	**AZERBAIJAN**	**TURKEY**
R a w S S c	23.56	20.09	22.86	19.23

		24.90	20.19	22.96	19.03
		24.69	20.95	23.05	19.97
			20.09	23.32	19.84
		22.01	19.17	21.13	20.28
	LSS2020	22.50	19.85	21.43	19.55
		22.30	19.29	20.16	18.54
		21.85	19.37	19.88	19.19
		22.04	18.89	20.49	18.45
	2021	21.69	18.95	20.83	19.22
		21.94	18.54	19.81	18.88
			18.92		19.08
		25.14	21.13	23.39	20.32
		24.37	21.77	22.62	20.43
	SSS2020	24.99	21.69	22.79	20.20
		24.82	21.91	22.32	20.21
		24.81			
Roasted hazelnuts		25.55			
		22.54	19.23	21.10	18.42
	LSS2020	22.21	20.80	20.81	18.57
		22.36	21.04	20.77	19.40
		22.45	20.31	20.83	18.47
		22.18	19.31	20.89	18.37
	2021	22.08	19.75	20.12	18.47
		22.02	18.89	20.50	18.37
		21.71	18.50		
Peeled roasted hazelnuts	**SSS2020**	24.85	20.77	21.92	19.78
		24.47	21.13	22.11	19.52
		24.34	21.27	21.42	20.06
		24.60	20.88	21.37	19.63
				21.76	19.84
					20.62
					19.77
	LSS2020	22.32	20.80	21.92	18.83
		22.41	21.04	21.26	19.28

Hazelnut paste		21.83	20.31	20.84	18.43
		22.11	20.26	20.40	19.16
	2021	22.99	19.40	20.01	19.38
		23.95	19.08	20.01	18.94
			19.06	19.03	17.91
			18.40		18.38
	SSS2020	24.72	21.26	22.02	19.75
		24.09	20.67	21.30	19.36
		23.73	20.60	21.99	20.10
		24.47	20.54	21.56	19.83
		24.04		21.13	20.28
		24.27		21.43	19.55
	LSS2020	23.29	21.12	20.83	20.59
		22.83	19.85	20.74	20.67
		21.99	20.16	20.14	
			19.52		
	2021	23.45	19.68	20.46	20.05
		22.92	19.31	19.99	19.87
			19.09	20.11	19.55
			18.76	20.45	19.11

Table 4 *A) Mean, standard deviation, variance, and standard error values of δ^{18} O (bulk raw, roasted, peeled roasted, and paste hazelnuts) for Italy, Georgia, Azerbaijan, and Turkey. B) Measured δ^{18}O values of bulk raw, roasted, peeled roasted, and paste hazelnuts from Italy, Georgia, Azerbaijan, and Turkey.*

The number of analyses for this isotopic ratio does not correspond to one of the carbon analyses, this was due to the presence of some outlier values. These values were removed from the model. A one-way ANOVA test was performed also for these analyses and in this case, the δ differences were more relevant than the ones obtained from the $^{13}C/^{12}C$ measurements. (Table 5)

	Sum of squares	Degree of freedom	Mean square	F	p(same)
Between groups	408.98	3	136.33	139.9	$1.86x10^{-47}$
Into groups	181.22	186	0.97		
Total	590.19	189			

Table 5 *ANOVA results for the oxygen isotope data f the four groups considered.*

As expected, standard deviation values increased for this isotopic ratio. Also, for the oxygen isotopes, the Italian samples exhibit the highest delta values, while the Turkish ones have the lowest values. The trend could be confirmed by relating O isotopic composition with parameters that might impact it, like relative humidity, temperature, amount of precipitation, and distance from the sea and/or other evaporation sources. (Taous, et al., 2020) Some climatic parameters (see https://climateknowledgeportal.worldbank.org/) are related to the delta values, a comprehensive portal that provides historical and future climate, vulnerabilities, and impacts. Italy and Azerbaijan exhibit the highest annual temperature (13.85 and 13.76 °C in 2020, and 13.50 and 14.29 °C in 2021, respectively) and the highest $^{18}O/^{16}O$ as well. Moreover, Italy has annual precipitation (769.9 mm in 2020, and 730.6 mm in 2021) and delta values higher than Azerbaijan (445 mm in 2020, and 473.1 mm in 2021) demonstrating that both temperature and precipitation influence the oxygen isotopic ratio. Azerbaijani mean value is slightly higher than the Georgian one, despite an important annual precipitation amount of the latter. (97.59 cm in 2020, 105.01 cm in 2021) This could be explained by the "mitigation" effect of the temperature since the Azerbaijan area registered 3.84 °C more than the Georgian area in 2020, and 4.23 °C in 2021. Fig. S4A-S4B show δ ^{18}O ‰-annual precipitation amount-annual temperature histograms, highlighting differences between Italian and not Italian samples, driven by these two climatic parameters.

ICP-OES / ICP-MS / IRMS data

ICP-OES and ICP-MS analyses were focused on the main elements that highlighted differences between Italian and non-Italian samples. Calcium (Ca), Iron (Fe), Manganese (Mn), and Potassium (K) were the selected elements, whose values are listed in Table 5.

		ITALY	GEORGIA	AZERBAIJAN	TURKEY
Raw hazelnuts (Ca)	**SSS2020**	1533.33	1680	1800	1810
	LSS2020	1340	1059	1560	1620
	2021	1666.7	1124	984	1680
Raw hazelnuts (Fe)	**SSS2020**	38.2	35.7	34.5	36.3
	LSS2020	44.5	29.2	37.5	33.9
	2021	47.4	37.7	31.7	55
Raw hazelnuts (Mn)	**SSS2020**	38	88	39.3	90
	LSS2020	49	83	36.8	80
	2021	38.1	70	33.6	89
	SSS2020	6786.7	8570	5880	5580

Raw hazelnuts (K)	**LSS2020**	7503.3	6570	5670	6170
	2021	8656.7	6960	5950	7510
Roasted hazelnuts (Ca)	**SSS2020**	1563.3	1570	1620	2130
	LSS2020	1299.7	1094	1560	1680
	2021	1483.3	1175	1540	1690
Roasted hazelnuts (Fe)	**SSS2020**	34.5	35.2	35.4	30.2
	LSS2020	41.3	32.5	41.7	42.8
	2021	48.1	36.6	33.7	53
Roasted hazelnuts (Mn)	**SSS2020**	42.9	71	53	116
	LSS2020	43.2	89	43.5	80
	2021	45.6	69	34.2	122
Roasted hazelnuts (K)	**SSS2020**	8003.3	8790	5550	6910
	LSS2020	7426.7	6710	5730	6470
	2021	8526.7	7700	6950	7510
Peel. Roasted hazelnuts (Ca)	**SSS2020**	1413.3	1620	1540	1710
	LSS2020	1463.7	1240	1250	1380
	2021	1433.3	1180	1530	1520
Peel. Roasted hazelnuts (Fe)	**SSS2020**	34.2	34.3	33.8	30.9
	LSS2020	33.6	38.6	43.7	47.2
	2021	39.2	36.7	38.9	39.8
Peel. Roasted hazelnuts (Mn)	**SSS2020**	35.9	81	46	86
	LSS2020	42.2	106	57	92
	2021	42.2	67	44	101
Peel. Roasted hazelnuts (K)	**SSS2020**	8630	8710	5630	7060
	LSS2020	7816.7	7080	5920	5880
	2021	8356.7	7400	6860	7050
Hazelnut Paste (Ca)	**SSS2020**	1333.3	1360	1370	1540
	LSS2020	1186.7	1186	1310	1500
	2021	1245	/	1240	1400

Hazelnut	**SSS2020**	31.8	37.7	31.7	22.3
Paste	**LSS2020**	33.5	37.8	33.3	36.6
(Fe)	**2021**	35.4	/	30.7	39.3
Hazelnut	**SSS2020**	31.8	70	31.7	22.3
Paste	**LSS2020**	39.7	90	40	71
(Mn)	**2021**	35.5	/	40.3	73
Hazelnut	**SSS2020**	7656.7	8510	5740	7000
Paste	**LSS2020**	6446.7	7180	5590	6450
(K)	**2021**	7053.3	/	5410	7330

Table 6. *ICP-OES (Ca, K), and ICP-MS data (Mn, Fe) (mg/Kg) for bulk raw, roasted, peeled roasted, and paste hazelnuts from Italy, Georgia, Azerbaijan, and Turkey.*

The different shelf life and the harvesting year impacted the hazelnut minor chemical component distribution. These components come from the soil and their assimilation is probably influenced by the climatic condition. These ICP-OES, ICP-MS, and IRMS data have been pulled together and statistically treated with multivariate analysis (SIMCA software, version 16.0, Umetrics, Umea, Sweden). This scaling method is the most objective, as it puts all the variables on the same 'importance level' in the data analysis. It is useful when variables are of different kinds and not directly comparable numerically (Eriksson, Byrne, Johansson, Trygg, & Vikstrom, 2013). Initially, the data elaboration workflow considered an unsupervised model, the Principal Component Analysis (PCA), aimed at enhancing the interpretability of a dataset with a high number of features per observation, while preserving as much information as possible and providing a multidimensional data visualisation. (Jolliffe & Cadima, 2016) On the other hand, this model does not present any indications regarding the classes of the samples, the data are basically placed in a bi-dimensional space, according to the process of computing the principal components. Fig. S5 shows the PCA model built considering both ICP-OES and IRMS datasets.

The score plot already highlights a discrete differentiation. In particular, the first Principal Component (PC) allows the separation of the Italian and the non-Italian groups. Considering these acceptable outcomes from the unsupervised model, the following natural step was the supervised model, a Partial Least Square – Discriminant Analysis, a versatile algorithm useful for descriptive modelling as well as for discriminative variable selection (Lee, Liong, & Jemain, 2018). Thus, the samples are basically 'labelled' depending on their class, and this led to a 'driven' clustering. Fig. S6 demonstrates how the separation between Italian and not Italian clusters has improved, as expected, by moving to the supervised model.

The loading plot (Fig. S7) indicates how the model variables relate to each other and the group, underlining which ones are the most important for the class separation.

As expected, δ ^{13}C and δ ^{18}O ‰ from both repeatability and reproducibility analysis are higher in the Italian samples than the others, whereas Calcium and Manganese concentrations are predominant in Georgia and Turkish samples. Azerbaijani samples present the lowest values of Potassium in all the matrices analysed.

CONCLUSION

δ ^{13}C and δ ^{18}O IRMS analysis, coupled with a multi-elemental approach, represents a key tool for discrimination of the geographical origin of hazelnuts and hazelnut-based products. In the future, $^{2}H/^{1}H$ and $^{15}N/^{14}N$ could also be considered for the geographical origin assessment, as well as the chemical compounds extracted from the matrix, instead of the bulk analysis data. In addition, the discriminative ability of the technique was also assessed by a statistical approach based on the pooled isotopic and chemical data. An interesting and functional exploitation of this research, from both academic and industrial points of view, could be the creation every year of a reference database, containing data of samples from the geographical areas of interest, used as reference. According to the isotopic ratio values obtained, their average and standard deviation, it could be possible to set a range within whom the sample is considered authentic. In this way, the database could be employed for the detection of false declaration frauds.

REFERENCES

An, N., Turp, M., Turkes, M., & Kurnaz, M. (2020). Mid-Term Impact of Climate Change on Hazelnut Yield. *Agriculture, X*, 159.

Bachmann, R., Klockmann, S., Haerdter, J., Fischer, M., & Hackl, T. (2018). 1H NMR Spectroscopy for Determination of the Geographical Origin of Hazelnuts. *Journal of Agricultural and Food Chemistry*(66), 11873-11879.

Biancolillo, A., De Luca, S., Bassi, S., Roudier, L., Bucci, R., Magrì, A., & Marini, F. (2018). Authentication of an Italian PDO hazelnut ("Nocciola Romana") by NIR spectroscopy. *Environmental Science and Pollution Research*(25), 28780–28786.

Bilke, S., & Mosandl, A. (2002). Measurements by gas chromatography/pyrolysis/mass spectrometry: Fundamental conditions in 2H/1H isotope ratio analysis. *Rapid Communications in Mass Spectrometry*, *16*(5), 468-472.

Camin, F., Larcher, R., Perini, M., Bontempo, L., Bertoldi, D., Gagliano, G., . . . Versini, G. (2010). Characterisation of authentic Italian extra-virgin olive oils by stable isotope ratios of C, O and H and mineral composition. *Food Chemistry*(118), 901-909.

Camin, F., Pavone, A., Bontempo, L., Wehrens, R., Paolini, M., Faberi, A., . . . Mannina, L. (2016). The use of IRMS, 1H NMR and chemical analysis to characterise Italian chemical analysis to characterise Italian. *Food Chemistry*(196), 98-105.

Chen, Z., Wang, G., & Jia, Y. (2017). Foliar δ13C Showed No Altitudinal Trend in an Arid Region and Atmospheric Pressure Exerted a Negative Effect. *Frontiers in Plant Science, 8*, 1070.

Ciarmiello, L., Mazzeo, M., Minasi, P., Peluso, A., De Luca, A., Piccirillo, P., . . . Carbone, V. (2014). Analysis of Different European Hazelnut (Corylus avellana L.) Cultivars: Authentication, Phenotypic Features, and Phenolic Profiles. *Journal of Agricultural and Food Chemistry*(62), 6236-6246.

Cordero, C., Liberto, E., Bicchi, C., Rubiolo, P., Schieberle, P., Reichenbach, S., & Tao, Q. (2010). Profiling food volatiles by comprehensive two-dimensional gas chromatography coupled with mass spectrometry: Advanced fingerprinting approaches for comparative analysis of the volatile fraction of roasted hazelnuts (Corylus avellana L.) from different ori. *Journal of Chromatography A*(1217), 5848-5858.

Eriksson, L., Byrne, T., Johansson, E., Trygg, J., & Vikstrom, C. (2013). 13 Centering and Scaling. In M. U. AB (Ed.), *Multi- and Megavariate Data Analysis - Basic Principles and Applications* (p. 490). Malmo, Sweden: Umetrics Academy.

Jolliffe, I., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of The Royal Society A: Mathematical, Physical, and Engineering Sciences, 374*(2065).

Kelly, S. (2003). Using stable isotope ratio mass spectrometry (IRMS) in food authentication and traceability. In M. Lees (Ed.), *Food Authenticity and Traceability* (pp. 156-183). Cambridge, the UK: Woodhead Publishing Ltd.

Kelly, S., Baxter, M., Chapman, S., Rhodes, C., Dennis, J., & Brereton, P. (2002). The application of isotopic and elemental analysis to determine the geographical origin of premium long grain rice. *European Food Research and Technology*(214), 72-78.

Klockmann, S., Reiner, E., Bachmann, R., Hackl, T., & Fischer, M. (2016). Food Fingerprinting: Metabolomic Approaches for Geographical Origin Discrimination of Hazelnuts (Corylus avellana) by UPLC-QTOF-MS. *Journal of Agricultural and Food Chemistry*(64), 9253-9262.

Klockmann, S., Reiner, E., Cain, N., & Fischer, M. (2017). Food Targeting: Geographical Origin Determination of Hazelnuts (Corylus avellana) by LC-QqQ-MS/MS-Based Targeted Metabolomics Application. *Journal of Agricultural and Food Chemistry*(65), 1456-1465.

Krauß, S., Vieweg, A., & Vetter, W. (2020). Stable isotope signatures ($\boldsymbol{\delta}$2H-, $\boldsymbol{\delta}$13C-, $\boldsymbol{\delta}$15N-values) of walnuts (Juglans regia L.) from different regions in Germany. *Journal of the Science of Food and Agriculture, IV*(100), 1625-1534.

Lavergne, A., Sandoval, D., Hare, V. J., Graven, H., & Prentice, I. C. (2020). Impacts of soil water stress on the acclimated stomatal limitation of photosynthesis: Insights from stable carbon isotope data. *Global Change Biology, 26*(12), 7158-7172.

Lee, L., Liong, C., & Jemain, A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analysy, 143*(15), 3526-3539.

Manca, G., Camin, F., Coloru, G., Del Caro, A., Depentori, D., Franco, M., & Versini, G. (2001). Characterization of the Geographical Origin of Pecorino Sardo Cheese by Casein Stable Isotope (13C/12C and 15N/14N) Ratios and Free Amino Acid Ratios. *Journal of Agricultural and Food Chemistry, III*(49), 1404-1409.

Mirotadze, N. (2005). Hazelnut in Georgia. *Acta Hortic, I*(686), 29-34.

Moscetti, R., Radicetti, E., Monarca, D., Cecchini, M., & Massantini, R. (2015). Near infrared spectroscopy is suitable for the classification of hazelnuts according to Protected Designation of Origin. *Journal of the Science of Food and Agriculture*(95), 2619-2625.

Muccio, Z., & Jackson, G. (2009). Isotope ratio mass spectrometry. *Analyst*(134), 213-222.

Ogrinc, N., Kosir, I., Kocjancic, M., & Kidric, J. (2001). Determination of Authenticy, Regional Origin, and Vintage of Slovenian Wines Using a Combination of IRMS and SNIF-NMR Analyses. *Journal of Agricultural and Food Chemistry*(49), 1432-140.

Rossmann, A., Haberhauer, G., Holzl, S., Horn, P., Pichlmayer, F., & Voerkelius, S. (2000). The potential of multielement stable isotope analysis for regional origin assignment of butter. *European Food Research and Technology*(211), 32-40.

Sah, S., & Brumme, R. (2003). Altitudinal gradients of natural abundance of stable isotopes of nitrogen and carbon in the needles and soil of a pine forest in Nepal. *Journal of Forest Science, I*(49), 19-26.

Taous, F., Amenzou, N., Marah, H., Maia, R., Maguas, C., Bahmad, L., & Kelly, S. (2020). Stable isotope ratio analysis as a new tool to trace the geographical origin of Argan oils in Morocco. *Forensic Chemistry, 17*, 100198.

SUPPLEMENTARY MATERIALS

Fig. S1 *Black Sea Region map (www.agefotostock.com)*

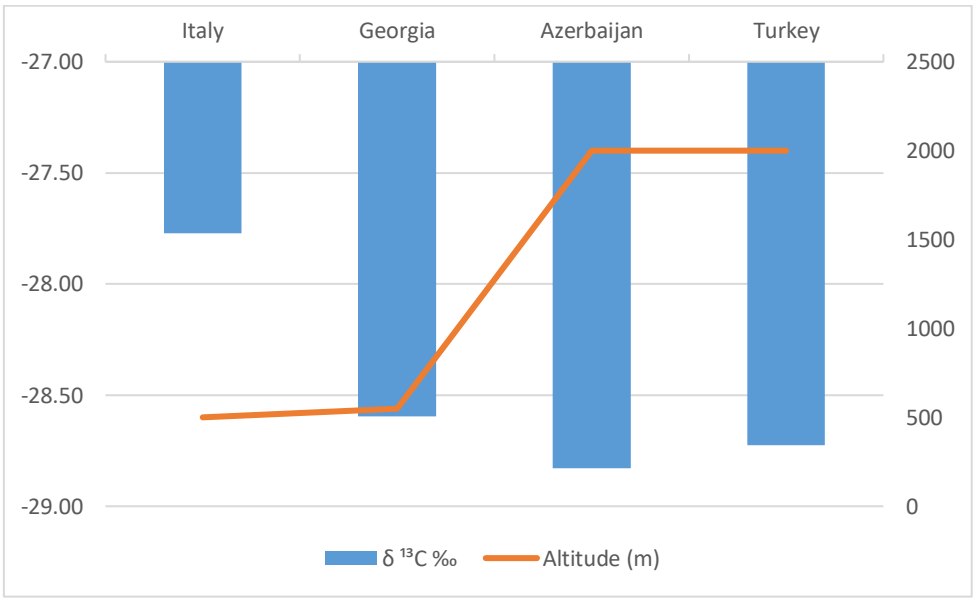


Fig S2. *Mean δ^{13}C-altitude values histogram plot of the 2020 campaign hazelnut sample set*

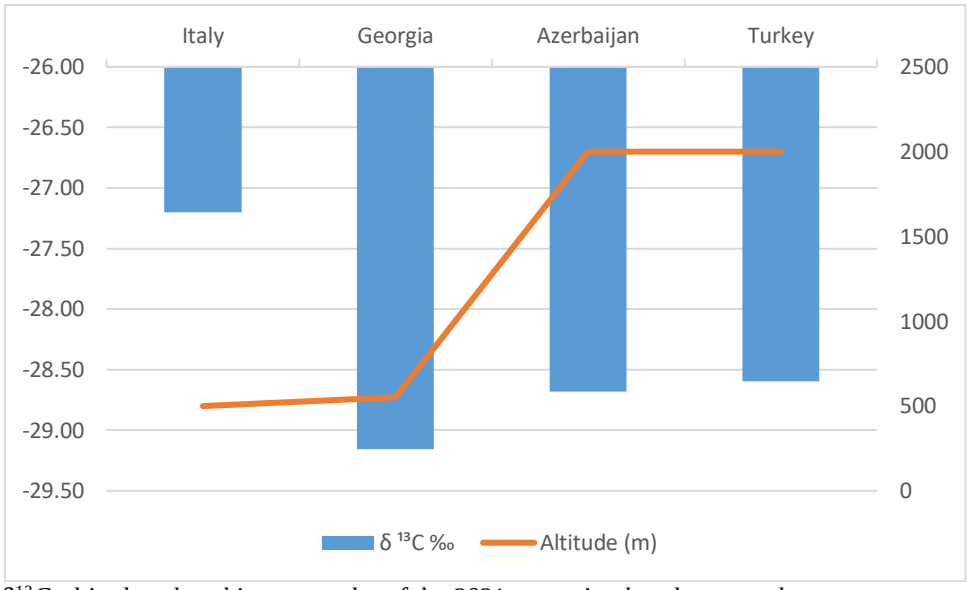
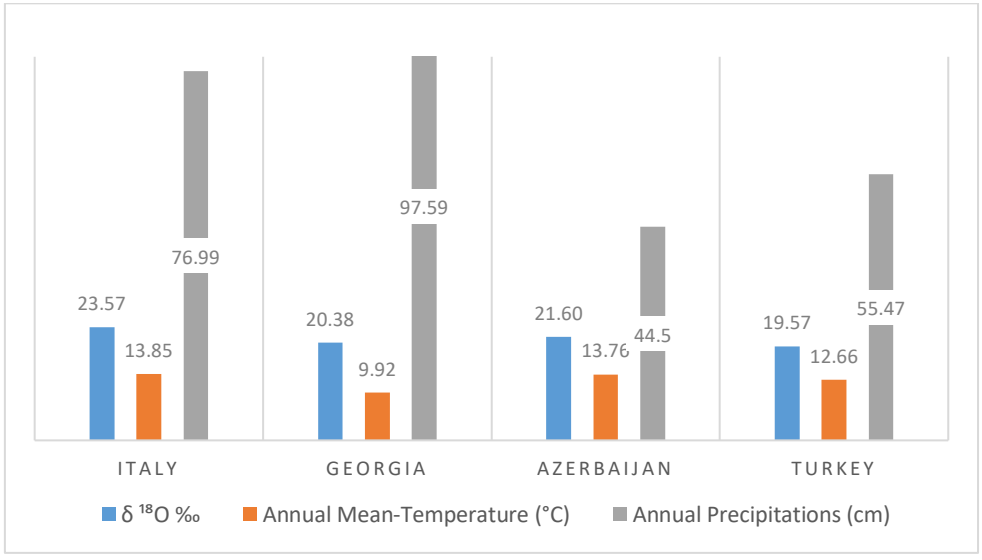


Fig S3. *Mean δ^{13}C-altitude values histogram plot of the 2021 campaign hazelnut sample set*

A)



B)

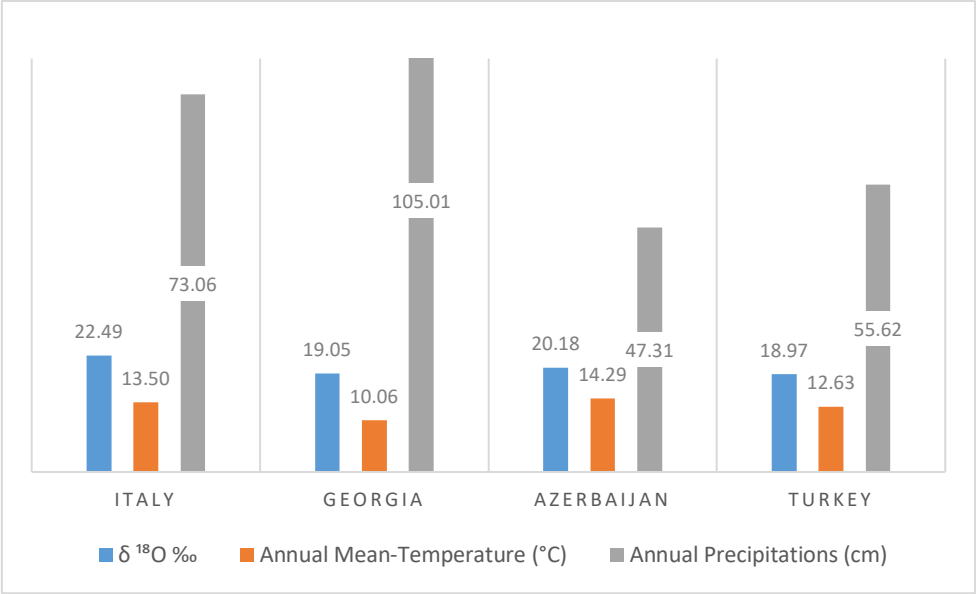


Fig. S4A) *Mean δ^{18}O-annual precipitation amount-annual temperature values histograms for 2020.* S4B) *Mean δ^{18}O-annual precipitation amount-annual temperature values histograms for 2021.*

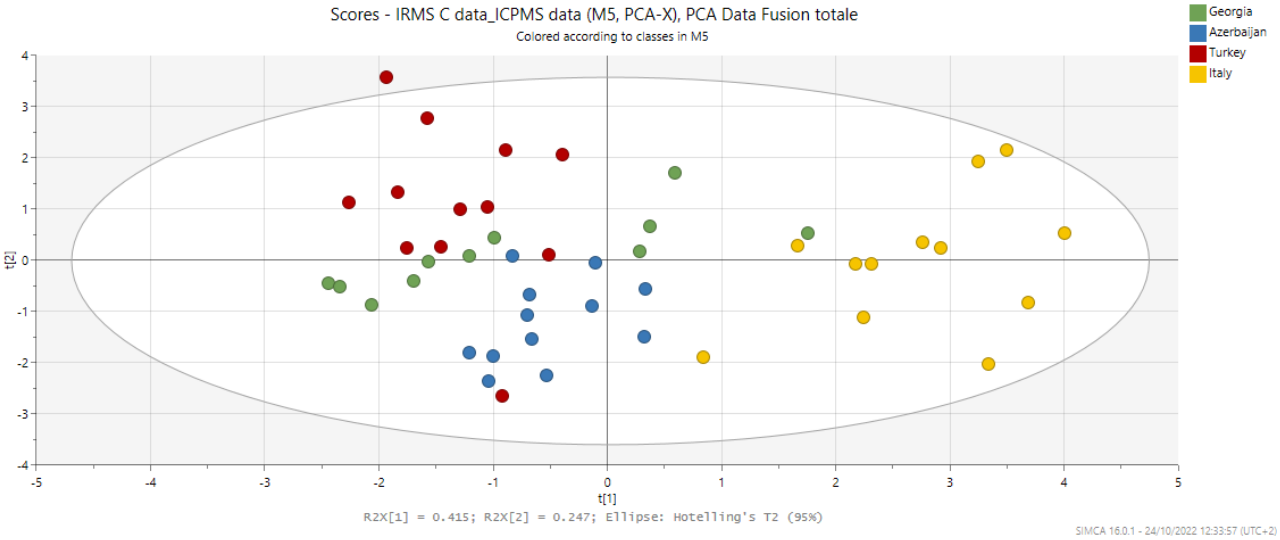


Fig. S5 *PCA score plot of hazelnut samples obtained using ICP-OES, ICP-MS, and IRMS data. (Green dots: Georgia, blue dots: Azerbaijan, red dots: Turkey, yellow dots: Italy)*

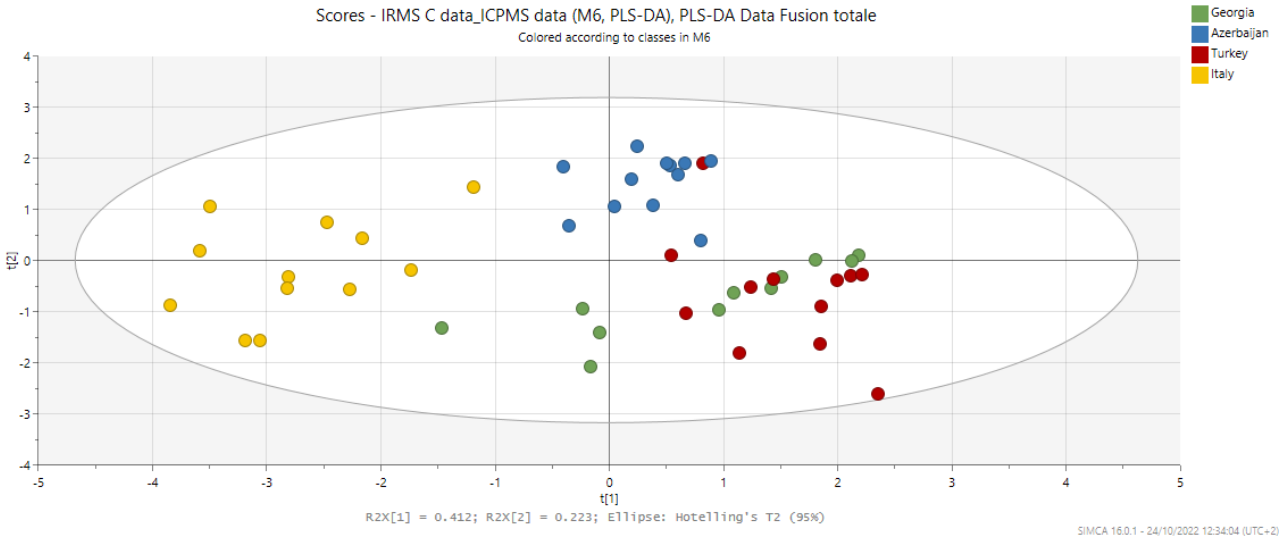


Fig. S6 *PLS-DA score plot of hazelnut ICP-OES, ICP-MS, and IRMS data. (Green dots: Georgia, blue dots: Azerbaijan, red dots: Turkey, yellow dots: Italy)*

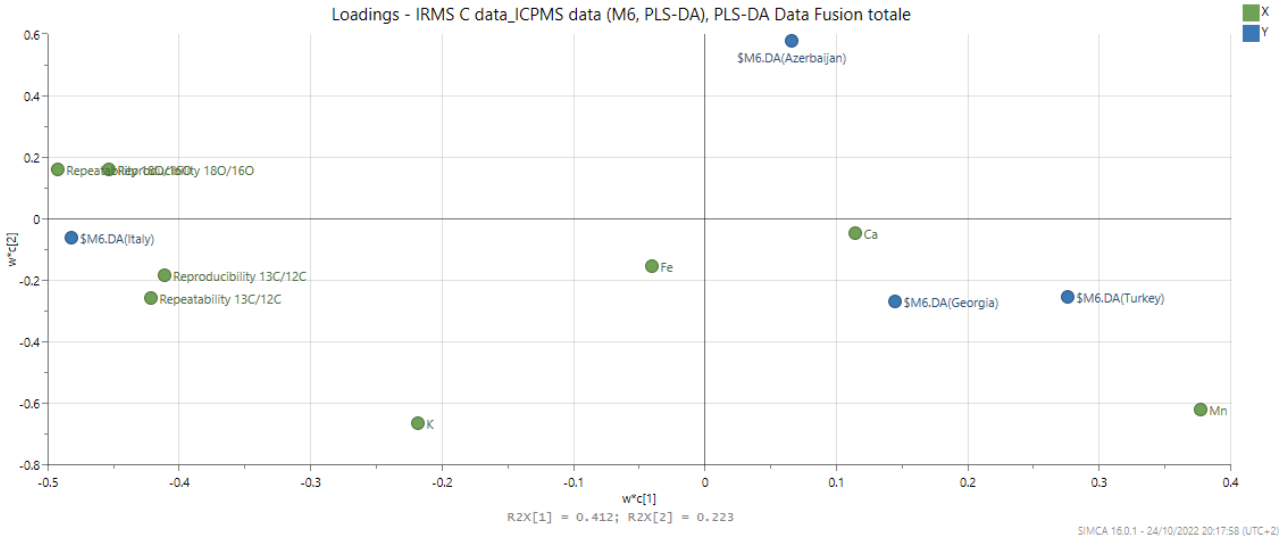


Fig. S7 *PLS-DA loading plot of hazelnut data obtained using ICP-OES, ICP-MS, and IRMS data. (Green dots: variables, blue dots: groups)*

CHAPTER 4

LC-HRMS ANALYSIS

LC-HRMS Analysis

General Overview

LC-HRMS is a remarkable approach to food analysis, it was exploited to study a wide range of food products. The technological progress led to enhanced performance facilities and better data processing tools, which make LC-HRMS a key player in several food-related studies (Donato, Cacciola, Beccaria, Dugo, & Mondello, 2012).

Both targeted and untargeted approaches are adopted for authenticity issues, the targeted ones are commonly used for monitoring bioactive substances, whose content and distribution in food commodities could depend on climate conditions, water resources, cultivation practices… All of these factors are related to geographical origin, so this targeted analysis can potentially contribute to food authentication studies. Non-targeted approaches, such as metabolomic fingerprinting, gained popularity in recent years since it is possible to evaluate and correlate a huge number of spectral detected variables. From these, it is possible to extrapolate numerous information via chemometric methods (Campmajo, Nunez, & Nunez, 2019).

Besides the feasibility of this technique for authenticity frauds, it represents an important tool for the detection and quantification of contaminants in food matrices. For example, food companies developed analytical methods for measuring mycotoxins in wheat- and oat-based foods accurately and sensitively (Suman & Catellani, 2008).

The high sensitivity and resolution, the capability to obtain chemical structural information, even molecular identification, and the robustness of this technology make it one of the most used for food analysis. However, there is a wide gap between academic and industrial research as regards LC-HRMS. This is prevalently due to the very expensive costs of the diverse spectrometers in commerce, as well as the expertise required for both data acquisition and elaboration. Small-medium food enterprises are not encouraged in investing on this instrumentation. Nevertheless, a functional solution could be represented by an open innovation approach, where universities are available in developing an analytical method and performing the analyses in their labs, according to the projects and targets of the companies, that will be the financers.

The goal of the work

The present chapter explains an untargeted metabolomic LC-HRMS approach for the geographical origin evaluation of Italian dehydrated apple samples. The study was submitted to "*Journal of Mass Spectrometry*". For additional details see the section "Author".

References

Campmajo, G., Nunez, N., & Nunez, O. (2019). The Role of Liquid Chromatography-Mass Spectrometry in Food Integrity and Authenticity. In G. Kamble, *Mass Spectrometry - Future Perceptions and Applications* (p. 134). intechOpen.

Donato, P., Cacciola, F., Beccaria, M., Dugo, P., & Mondello, L. (2012). Advances in LC-MS for Food Analysis. *LCGC Supplements*, 15-24.

Suman, M., & Catellani, D. (2008). Ion trap lC-MS/MS as a valid multi-method to determine trichothecenes and zearalenone in the food industry. *World Mycotoxin Journal*, 255-262.

Untargeted Metabolomics Liquid Chromatography-High Resolution Mass Spectrometry approach for the Geographical Origin Assessment of Italian Dehydrated Apples

Giuseppe Sammarco[1,3], Chiara Dall'Asta[3], Michele Suman[1,2]

[1] Advanced Laboratory Research, Barilla G. e R. Fratelli S.p.A., Parma, Italy
[2] Department for Sustainable Food Process, Catholic University Sacred Heart, Piacenza, Italy
[3] Department of Food and Drug, University of Parma, Parma, Italy

Abstract

Geographical provenience is nowadays a relevant aspect of the authenticity and the quality of many food commodities. Dehydrated apple cubes/slices represent an ingredient commonly used by food companies for bakery products. However, this apple-based matrix is not so known and studied from an analytical point of view. In the present work, seven compounds were identified as key molecules to distinguish between Italian and non-Italian samples, through an untargeted ultrahigh-pressure liquid chromatography-high-resolution mass spectrometry (UHPLC-HRMS) approach. This methodology was merged with multivariate statistical analysis, and the principal features were studied and identified considering several identification steps. Samples from 2020 and 2021 harvesting campaigns, with partial and total dehydration rates, with or without peel, and from different apple varieties were considered for the study, for a total of 91 samples. Afterward, the same analysis protocol was applied to an external set (n=12 samples), included in the statistical models, searching for the key compounds identified in the training set. Interesting and significant results underlined the potentiality of the UHPLC-HRMS technology as a confirmatory strategy for the geographical origin assessment of dehydrated apple commodities.

Keywords

Dehydrated Apples, Geographical Origin, UHPLC-HRMS, Multivariate Statistical Analysis, Compounds Identification

INTRODUCTION

Liquid chromatography-high resolution mass spectrometry (LC-HRMS) technique represents a useful strategy to finely analyse food commodities, in order to obtain structural information about the main components of foods (Aydogan, 2020) or to prevent/face safety and authenticity issues, such as contaminants detection (Kunzelmann, Winter, Aberg, Hellenas, & Rosén, 2018). In particular, food authenticity is nowadays a relevant aspect, related to both the quality and safety of food commodities (Selamat, Rozani, & Murugesu, 2021). Geographical origin declaration is one of the peculiarities as regards food authenticity since it could be linked to the quality of the products (Katerinopoulou, Kontogeorgos, Salmas, Patakas, & Ladavos, 2020). Apple (*Malus x domestica Borkh.*) is a very common fruit cultivated in Europe. Italy is one of the best apple producers (6^{th}) and exporters (2^{nd}) in Europe (Atlas Big, 2021) (World's Top Exports, 2021). Furthermore, the Italian apple quality, certified by European geographical indications (GIs), makes them renowned as well as more expensive than the ones from the main producer countries, according to Global Product Prices (Global Product Prices, 2022). This pushed Italian apple producers/suppliers/retailers to invest in technologies to assess apple authenticity, in terms of geographical provenience. Thus, the first steps were taken to identify false declaration of origin fraud, where Italian products are mixed or even substituted by non-Italian matrices and sold as 'Made in Italy'.

Several analytical methodologies were considered to fight the above-mentioned type of fraud. Multiple elements analysis was performed to differentiate apple samples from five Chinese regions. Fourteen elements were analysed using graphite furnace atomic absorption spectrometry (GFAAS) and inductively coupled plasma atomic emission spectroscopy (ICP-AES). Principal component analysis (PCA), linear discriminant analysis (LDA), and back-propagation artificial neuron network analysis (BP-ANN) were employed for the modelling (Zhang, et al., 2019). Both electronic nose (EN) and electronic tongue (ET) were exploited to study the variety and origin of 126 apple samples from seven Chinese regions. PCA, LDA, support vector machine (SVM), and Partial Least Square-Discriminant Analysis (PLS-DA) were performed for classification modelling (Wu, Yue, & Yuan, 2018). Fourier-transform near-infrared (FT-NIR) spectroscopy was another technology carried out to define the soluble solids content of 'Fuji' apple samples from three origins. NIR wavelengths were selected by competitive adaptive reweighted sampling algorithm (CARS), and PLS-DA was adopted for the classification (Li, et al., 2018). FT-NIR spectral feature extraction was combined with model search strategies, wavelength selection, and deep learning with multivariate regression analysis to authenticate Chinese 'Fuji' samples (Bai, Xiong, Huang, Zhou, & Zhang, 2019). Genetic differences

were found in apple samples from three harvestings in northeastern Spain through 16 simple sequence repeats. A further set of cultivars with Spanish and international products was used as a reference. Genetic analyses were carried out by Bayesian model-based clustering, revealing remarkable differences between the two groups (Urrestarazu, Miranda, Santestaban, & Royo, 2012).

An apple-based commodity often employed by food companies for diverse finished products is the dehydrated apple cube/slice, but, to the best of the authors' knowledge, there is a gap in scientific works about food authenticity and/or geographical origin assessment of this apple-based product. In this scenario, LC-HRMS could represent a solid analytical strategy to face authenticity issues related to the dehydrated apple commodity. This technique was, indeed, largely adopted for geographical origin assessment in the food field. Popping *et al.* developed an analytical method for the identification of the geographical origin of Parmigiano Reggiano, a protected designation of origin (PDO) Italian cheese, employing non-targeted MS and chemometrics. A database with authentic samples was created, and then 32 grated cheese samples were investigated, gaining an accuracy rate of 87.5 % (Popping, De Dominicis, Dante, & Nocetti, 2017). Another authenticity study, regarding the differences in the lipid pattern between durum and common wheat, involved the MS. A first exploratory step was carried out by analysing 52 samples from two common and durum wheat varieties. Subsequently, it was extended to an independent sample set (173 samples) to better assess the coherence of the models built (Righetti, et al., A novel approach based on untargeted lipidomics reveals differences in the lipid pattern among durum and common wheat, 2018). Olive oil is another commodity largely studied about its geographical origin. Two different tools, LC-electrospray ionization-time of flight MS (LC-ESI-ToF MS) and LC-ESI-ion trap MS (LC-ESI-IT MS) were applied to characterise, and subsequently quantify, phenolic compounds in the north Moroccan virgin olive oil. 25 key phenolic compounds, from 156 samples harvested in 7 different Moroccan areas, were identified and quantified. Multivariate statistical analysis was also performed, to test the capability of phenolic profiles to define the geographical origin of the examined oil samples (Bajoub, Carrasco-Pancorbo, Ajal, Ouazzani, & Fernandez-Gutierrez, 2015).

Therefore, the robustness of a confirmatory technique such as LC-HRMS, and its wide application in the food field, represented the main factors for employing this analytical strategy in the present study, with an untargeted approach, aiming at discriminating between Italian and non-Italian dehydrated apples, and also contributing to fill the scientific lack of knowledge about this apple-based product.

MATERIALS AND METHODS

Chemicals and samples

Acetonitrile (ACN), ammonium formate (AF), formic acid (FA), iso-propanol (i-Prop), methanol (MeOH), and n-hexane were all purchased from VWR International Ltd (Poole, England, UK). Analytical standard chloramphenicol was purchased from Sigma-Aldrich (St. Louis, Missouri, US). Water was bi-distilled utilizing a Milli-Q system (Millipore, Bedford, Massachusetts, US).

A design experiment (DoE) was conducted to have a functional dehydrated apple sampling. Several factors were considered, such as dehydration rate, presence of peeling, harvesting year, and variety. The chosen areas of interest were France, China, Chile, Hungary, Poland, and Italy as the main geographical target. A training set (n=91) was employed to build a statistical model, that was subsequently applied to a test set (n=12). For this set, samples from the 2022 harvesting campaign were picked as well. Table 1-2 listed the training and test sets, respectively, and their corresponding factors.

Sample ID	Year	Variety	Dehydration rate	Presence of peel	Origin
122931	2020	Pink Lady	Total	Yes	Italy
122932	2020	Pink Lady	Partial	No	Italy
122933	2020	Pink Lady	Partial	No	China
122934	2020	Pink Lady	Total	Yes	China
122935	2020	Pink Lady	Partial	Yes	China
122936	2020	Pink Lady	Total	No	China
123004	2020	Fuji	Partial	No	Italy
123005	2020	Fuji	Partial	Yes	Italy
123006	2020	Fuji	Total	No	Italy
123007	2020	Granny Smith	Partial	No	Italy
123009	2020	Granny Smith	Total	Yes	Italy
123010	2020	Fuji	Total	Yes	France
123011	2020	Fuji	Total	No	China
123012	2020	Granny Smith	Partial	Yes	China
123013	2020	Fuji	Partial	No	China
123014	2020	Granny Smith	Total	Yes	China
123015	2020	Granny Smith	Total	Yes	France
123016	2020	Granny Smith	Total	No	China
123017	2020	Fuji	Partial	No	France
123019	2020	Granny Smith	Partial	No	France
123021	2020	Fuji	Partial	Yes	France
123022	2020	Granny Smith	Total	No	France
123023	2020	Fuji	Total	No	France
123025	2020	Granny Smith	Partial	Yes	France
166460	2021	Pink Lady	Total	No	France
166461	2021	Granny Smith	Partial	Yes	France
166462	2021	Fuji	Partial	No	France
166463	2021	Granny Smith	Total	Yes	France

166464	2021	Granny Smith	Total	No	France
166465	2021	Fuji	Partial	Yes	France
166466	2021	Golden Delicious	Total	No	France
166467	2021	Granny Smith	Partial	Yes	France
166468	2021	Fuji	Total	No	France
166469	2021	Golden Delicious	Partial	No	France
166471	2021	Pink Lady	Partial	No	France
166472	2021	Pink Lady	Partial	Yes	France
166473	2021	Pink Lady	Total	Yes	France
166474	2021	Golden Delicious	Total	Yes	France
166475	2021				
166476	2021	Pink Lady	Partial	No	Hungary
166477	2021	Golden Delicious	Partial	Yes	Hungary
166478	2021	Granny Smith	Partial	Yes	Hungary
166479	2021	Fuji	Total	No	Hungary
166480	2021	Golden Delicious	Total	Yes	Hungary
166481	2021	Pink Lady	Total	Yes	Hungary
166482	2021	Pink Lady	Total	No	Hungary
166483	2021	Granny Smith	Total	Yes	Hungary
166484	2021	Golden Delicious	Partial	No	Hungary
166485	2021	Granny Smith	Partial	No	Hungary
166486	2021	Fuji	Partial	Yes	Hungary
166487	2021	Granny Smith	Total	Yes	Hungary
166488	2021	Fuji	Partial	No	Hungary
166489	2021	Pink Lady	Partial	Yes	Hungary
166490	2021	Fuji	Total	Yes	Hungary
167047	2021	Fuji	Total	Yes	China
167048	2021	Fuji	Partial	Yes	China
167049	2021	Pink Lady	Partial	Yes	China
167050	2021	Granny Smith	Partial	Yes	China
157051	2021	Pink Lady	Total	No	China
167052	2021	Golden Delicious	Total	Yes	China
167053	2021	Fuji	Total	No	China
167054	2021	Granny Smith	Total	No	China
167055	2021	Pink Lady	Partial	No	China
167056	2021	Pink Lady	Total	Yes	China
167057	2021	Golden Delicious	Partial	Yes	China
167058	2021	Granny Smith	Total	Yes	China
167059	2021	Golden Delicious	Total	No	China

167060	2021	Golden Delicious	Partial	No	China
167061	2021	Granny Smith	Partial	No	China
167062	2021	Fuji	Partial	No	China
167063	2021	Golden Delicious	Partial	No	Poland
167064	2021	Golden Delicious	Total	Yes	Poland
167065	2021	Fuji	Partial	No	Poland
167066	2021	Fuji	Total	Yes	Poland
167067	2021	Fuji	Total	No	Poland
167068	2021	Fuji	Partial		Poland
167069	2021	Golden Delicious	Total	No	Poland
167070	2021	Golden Delicious	Partial	Yes	Poland
167783	2021	Granny Smith	Partial	No	Italy
167784	2021	Fuji	Partial	No	Italy
167785	2021	Golden Delicious	Partial	No	Italy
167786	2021	Pink Lady	Partial	No	Italy
167787	2021	Pink Lady	Total	No	Italy
167788	2021	Golden Delicious	Total	No	Italy
167790	2021	Granny Smith	Total	No	Italy
169008	2021	Granny Smith	Total	Yes	Italy
169009	2021	Granny Smith	Partial	Yes	Italy
169010	2021	Fuji	Total	Yes	Italy
169011	2021	Fuji	Partial	Yes	Italy
169012	2021	Pink Lady	Total	Yes	Italy
169013	2021	Pink Lady	Partial	Yes	Italy

Table 1 *Dehydrated apple samples training set, according to the DoE*

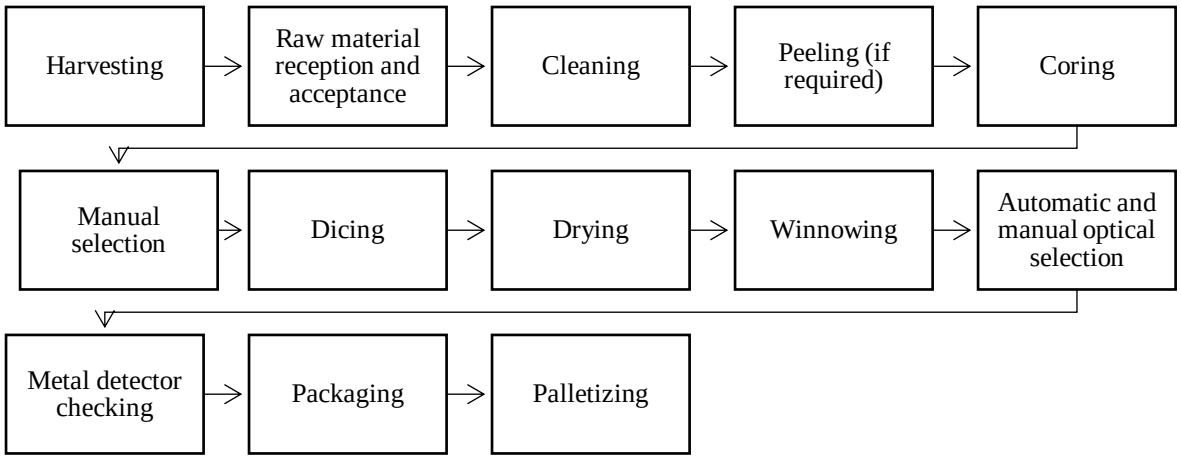
Sample ID	Year	Variety	Dehydration rate	Presence of peel	Origin
Chile 1	2022	Pink Lady	Total	Yes	Chile
Chile 2	2022	Pink Lady	Partial	Yes	Chile
Chile 3	2022	Pink Lady	Total	No	Chile
Chile 4	2022	Pink Lady	Partial	No	Chile
Italy 1	2022	Pink Lady	Total	Yes	Italy
Italy 2	2022	Pink Lady	Total	No	Italy
Italy 3	2022	Granny Smith	Total	Yes	Italy
Italy 4	2022	Granny Smith	Total	No	Italy
Hungary 1	2021	Granny Smith	Total	Yes	Hungary
Hungary 2	2021	Granny Smith	Partial	Yes	Hungary
Hungary 3	2021	Granny Smith	Total	No	Hungary
Hungary 4	2021	Granny Smith	Partial	Yes	Hungary

Table 2 *Dehydrated apple samples test set, according to the DoE*

The 2020 and 2021 samples were analysed in different periods and stored at 2-8 °C.

Dehydration process

According to the supplier indications, the dehydrated samples collected underwent the following treatment



To avoid the fruit browning before the dehydration phase, a pre-cooking step was carried out, in order to inactivate the enzymes that are responsible for the browning phenomenon (blanching). The moisture rate went from 5 (low) to 10 (normal) %, and the drying process was achieved by employing a hot air flow on the fruit. It is the most suitable to dry cubes or slices.

Sample preparation

Ca. 10 g dehydrated apple cubes were initially minced with the knife mill Grindomix GM 200 (Retsch, Haan-Gruiten, Germany). 1.0 g were weighed in a 50 mL falcon tube, then 10 mL n-hexane was added to extract and eliminate the fats from the samples, which could interfere with the LC-MS analysis by suppressing the metabolites' signals. The fatty extraction was favoured through an ultrasonic bath, for 15 minutes. The fatty fraction was so thrown away, and the samples were dried using a nitrogen flow. Subsequently, the metabolites extraction was performed by adding 10 mL of Milli-Q purified water, with 70 % of methanol, and 1 µg/mL of chloramphenicol as standard to evaluate the extraction efficiency. The extraction was again favoured by the ultrasonic bath, for 15 mins. Preliminary centrifugation was achieved by using a Rotina 380R (Hettic Lab Technology, Tuttlingen, Germany), at 5000 rpm for 1 min, keeping its temperature at 4°C. The supernatant was transferred in a 2 mL Eppendorf tube and centrifuged at 14000 rpm for 10 mins, always keeping the

temperature at 4 °C. The supernatant was finally filtered into an HPLC vial using a 0.22 µm polytetrafluoroethylene (PTFE) syringe filter (Phenomenex, Torrance, California, US) and stored at -20°C before the analysis. For each session, the same protocol was applied to empty tubes, following all the steps, except for the addition of the dehydrated apple powders. These samples were indicated as "Extraction blanks". To estimate the method's repeatability, 20 % of the samples were double-prepared. Moreover, a quality control (QC) sample was obtained by mixing 10 µL from each sample, extraction blanks excluded.

LC-HRMS method

HPLC analyses were carried out through a Dionex UltiMate 3000 UHPLC instrument (Thermo Fisher Scientific, Inc, Waltham, Massachusetts, US). Two different methods were applied, exploiting two HPLC columns. Kinetex® Biphenyl 100 x 2.1 mm, 2.6 µm particle size analytical column (Phenomenex, Torrance, California, US) was used for aromatic and organic components separation, in order to focus on the (poly)phenolic fraction, present in the apple matrix, as also reported in the literature (Oleszek, Lee, Jaworski, & Price, 1988). The analytical column was equipped with a precolumn security guard ultra C18, and both of them were kept at 40°C. Gradient elution was executed having FA as mobile phase modifiers, with a stable flow rate of 0.500 mL/min. The gradient conditions were: in the first minute, the system moved from 95 % to 80 % of phase A (water 0.1 % FA) and 20 % of phase B (ACN 0.1 % FA), it remained constant for 1 minute, then in another minute, it reached 95 % of phase B. It stayed at this level for half a minute, and in another half a minute, it returned to 95 % of phase A, and it was kept at this percentage for the remaining 2 minutes, before column re-equilibration, for a total runtime of 6 minutes. The autosampler temperature was maintained at 5°C, and the injection volume was 4 µL. The other chromatographic method involved the Acquity UPLC® BEH HILIC 100 x 2.1 mm, 1.7 µm particle size analytical column (Waters Corporation, Milford, Massachusetts, US). It was used for the sugars and the very polar components separation. The analytical column was equipped with a precolumn security guard ultra C18 as well, and both of them were kept at 40°C. Gradient elution was executed having FA and AF as mobile phase modifiers, with a stable flow rate of 0.500 mL/min. The gradient conditions were: for the initial 1.4 minutes, the system stayed from 90 % of phase A (ACN:i-Prop 70:30 0.1 % FA) and 10 % of phase B (MeOH 5 mM AF 0.1 % FA), in 0.1 minutes it moved from 10 to 60 % of phase B, and it was kept constant for half a minute. Afterward, in another half a minute, it returned at 10 % of phase B, and it was maintained at this percentage for 2.5 minutes, before column re-equilibration, for a total runtime of 5 minutes. The autosampler temperature was maintained at 5°C, and the injection volume was 4 µL.

Mass spectrometry analysis was carried out through a benchtop Q Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, Massachusetts) equipped with a heated (H) ESI interface (Thermo Fisher Scientific, Waltham, Massachusetts). Two analytical sequences were launched, one with a negative ionization mode, applying the LC method that used the Biphenyl column, and the other one with a positive ionization mode, applying the LC method that used the HILIC column, both of them performing a "Full Scan-data dependent fragmentation" experiment. Ionization source conditions were: sheath and auxiliary gas flow rates of 40 and 12 arbitrary units, respectively; heater temperature of 290°C, with a spray voltage of 3.2 kV (ESI pos) and −3.0 kV (ESI neg). The capillary was kept at 270°C and the S-lens RF level was set at 55 AU for both the acquisition modes. The full-scan accurate mass spectra, from 75 to 1000 m/z for both the ionization modes, were obtained with a resolution of 70 000 full width at half maximum (FWHM) (m/z 200), automatic gain control (ACG) target 1e6, and maximum injection time of 200 milliseconds. In all the experiments, the data-dependent tandem mass spectrometry (MS/MS) acquisition was executed with a resolution of 17 500 FWHM (m/z 200) and intensity threshold 2e4. The quadrupole isolation window was kept at 1.0 m/z with a TopN value of 5. The scan range was from 50 to the fragmented mass m/z (m/z +25), automatic gain control target 2e5, maximum injection time of 50 milliseconds, and normalized collision energy (NCE) of 20, 40, 100 % for positive ionization mode, and 20, 50, 100 % for negative ionization mode. Samples and "extraction blanks" were randomly injected to avoid systematic bias. The QC samples were injected at the beginning of the sequences and every 6 sample injections.

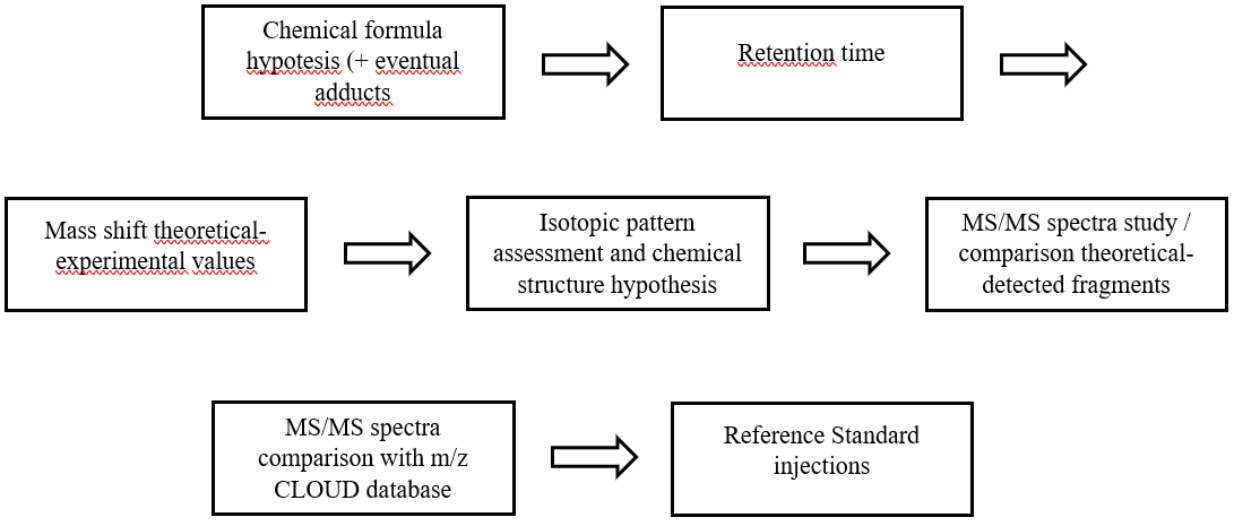
Data elaboration

UPLC-HRMS raw data were recorded via Xcalibur software (version 4.1, Thermo Fisher Scientific, Waltham, Massachusetts); peaks alignment, "Extraction blanks" subtraction, and features extraction were achieved through Compound Discoverer software (version 3.3, Thermo Fisher Scientific, Waltham, Massachusetts); the mass range inspected was between 75 *m/z* and 1000 *m/z* from 1 to 6/5 minutes of the chromatographic runs. The values of the principal parameters for the feature extractions are the following ones: precursor ion deviation of 5 ppm; maximum retention time shift of 1 minute; minimum peak intensity for a peak to be selected 100 000 AU; relative intensity tolerance for isotope search 30%. Structures prediction was also performed using "ChemSpider" databases, setting a maximum mass shift of 3 ppm. For the "*m/z CLOUD*" MS/MS library search, the precursor mass tolerance used was 0.4 Da while the fragments' mass tolerance was 10 ppm. The resulting data matrixes for both positive and negative ionization modes, containing the area values provided by Compound Discoverer for all the features, were exported and elaborated with SIMCA software

(version 16.0, Umetrics, Umea, Sweden) for multivariate statistical analysis. Data were Pareto scaled, then a preliminary PCA was carried out, to initially check the samples clusterisation, and the QCs positions in the scores plot. Afterward, features were filtered, selecting only the ones that had a coefficient of variation (CV) % lower than 40 % in the QC samples. A new PCA was performed to evaluate the expected improvement in sample separation and the QCs positioning. Subsequently, supervised PLS-DA models were built, and the samples are labelled, hence the data dispersion was lowered, and the clustering was improved. This workflow was initially adopted for the training set statistical analysis; to assess the robustness of it, also samples from the test set were included in the models. Thus, PCA and PLS-DA were re-performed, and to finally estimate their consistency, 43 samples, from both training and test sets, were picked from the overall set to create a prediction set, then a classification list and misclassification table were generated to visually assess the classification accuracy rate of the models.

Molecular identification

Molecular identification was accomplished according to these steps:



According to "The Standard Initiative in Metabolomics" (Sumner, et al., 2007) (Schymanski, et al., 2014), and Cavanna *et al.* (Cavanna, Zanardi, Dall'Asta, & Suman, 2018).

RESULTS AND DISCUSSION

Extraction efficiency evaluation

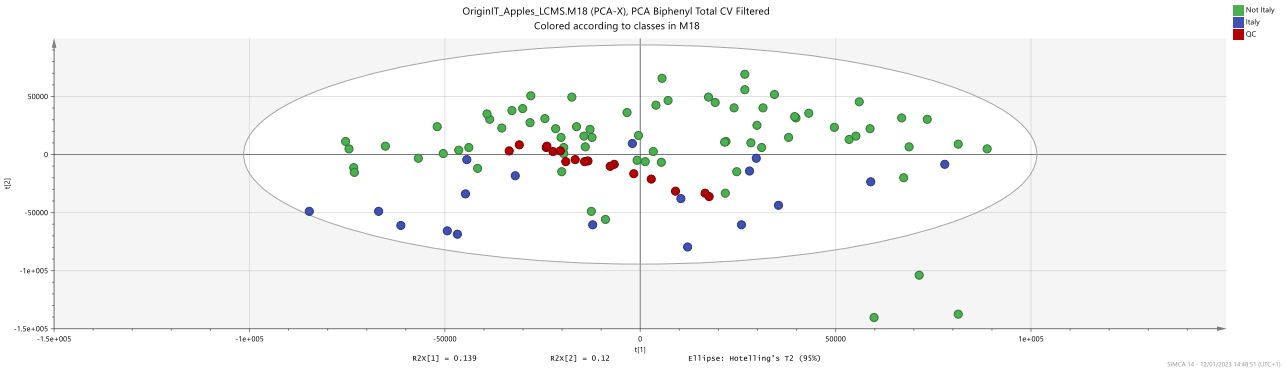
To estimate the goodness of the extraction procedure, an evaluation of the internal standard outcomes was done. Chloramphenicol was the standard added to each sample, in order to monitor the extraction protocol. Its peak intensity was assessed for all the sample injections. Hence, relevant information concerning the analytical sequence was gained. The CV % of the peak area values of the standard

were exploited to appraise the extraction repeatability/reproducibility and to exclude any mistakes that occurred during the sample preparation. The chloramphenicol was detectable mainly using a negative ionization mode, and since the extraction protocol was the same for both the chromatographic methods, only the outcomes from the analyses in negative ionization mode were considered. The areas CV was 36 %, and this result highlighted an acceptable extraction procedure efficiency; the retention times did not significantly shift throughout the sequence, a further indication of its goodness of it (Cavanna, Catellani, Dall'Asta, & Suman, 2018).

Multivariate Statistical Analysis

The PCA resulting from both whole data set elaborations (positive and negative ionization mode) highlighted a relevant data dispersion, also confirmed by the low explained variance described by the PCs (ca. 25-28 %), a slight separation between Italian and non-Italian samples, and a discrete clustering of the QC samples at the centre of the score plots (Figures S1A-B – Supplementary Materials). The dispersion of the data could be due to the important number of qualitative variables / DoE factors (dehydration rate, harvesting year, presence of peel, variety) related to the number of samples. After the filtering, with the selection of the features having area CV % values lower than 40 % in the QC samples, the models' performance slightly improved (Figures 1A-B). Especially for the positive ionization mode score plot, the QC cluster was tighter at the centre, and also the explained variance increased (ca. 36 %).

A)



B)

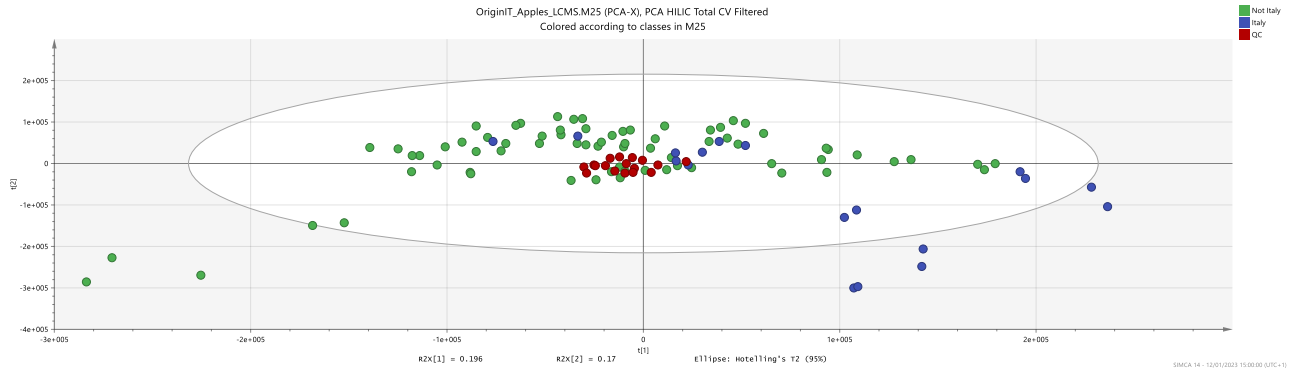
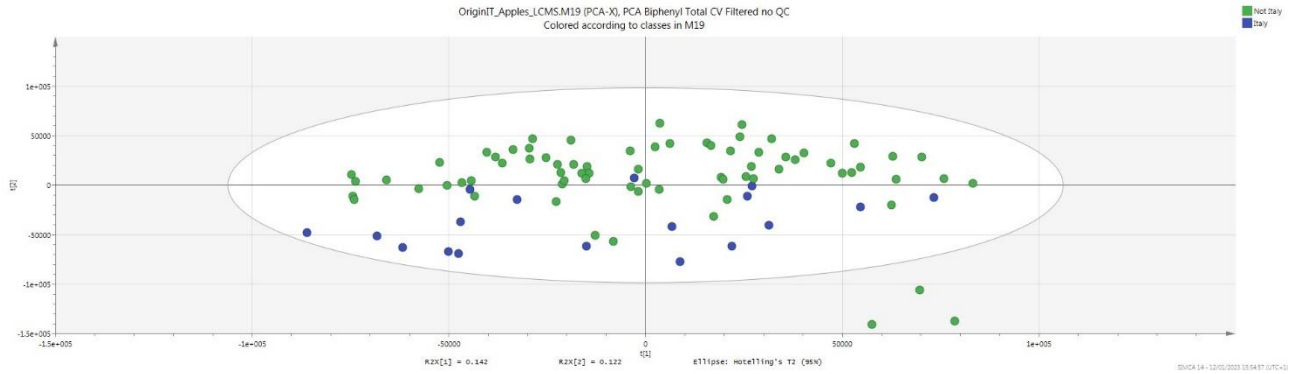


Figure 1A) *PCA score plot of the dehydrated apple sample set (n= 108) in negative ionization mode, after features filtering.* 1B) *PCA score plot of the dehydrated apple sample set (n= 108) in positive ionization mode, after features filtering. (Green dots: Not Italy, blue dots: Italy, red dots: QC samples)*

The QC samples' positions on the score plots confirmed the consistency of the analytical protocol, and it assured that the grouping was accomplished thanks to real variability. Another indication concerning the goodness of the analytical procedure was the proximity of the replicates in the plots. The PCA model built without the QC samples pointed out a better separation between Italian and non-Italian classes (Figures 2A-B).

A)



B)

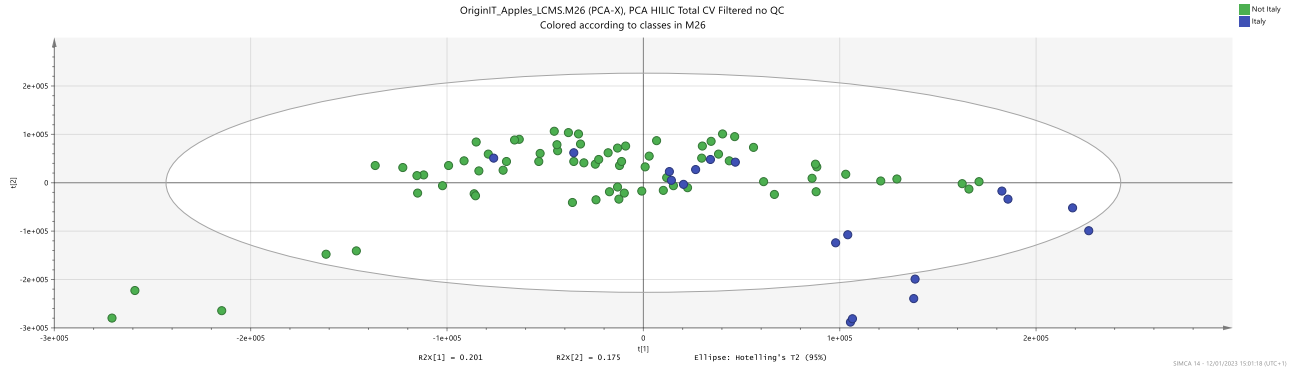
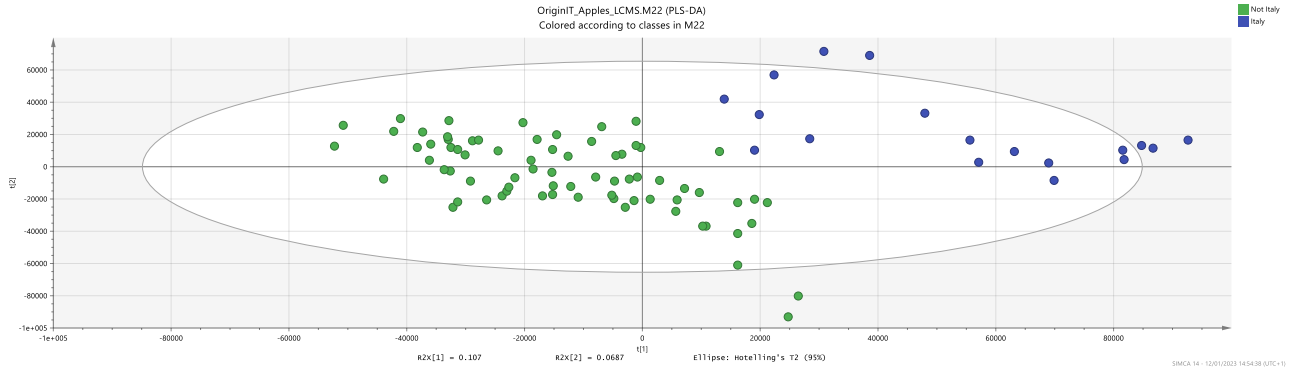


Figure 2A) *PCA score plot of dehydrated apple sample set without QC samples (n=91) in negative ionization mode, after features filtering.* 2B) *PCA score plot of dehydrated apple sample set without QC samples (n=91) in positive ionization mode, after features filtering. (Green dots: Not Italy, blue dots: Italy)*

The supervised PLS-DA models for both methods employed clearly increased the clustering between Italian and non-Italian (Figures 3A-B).

A)



B)

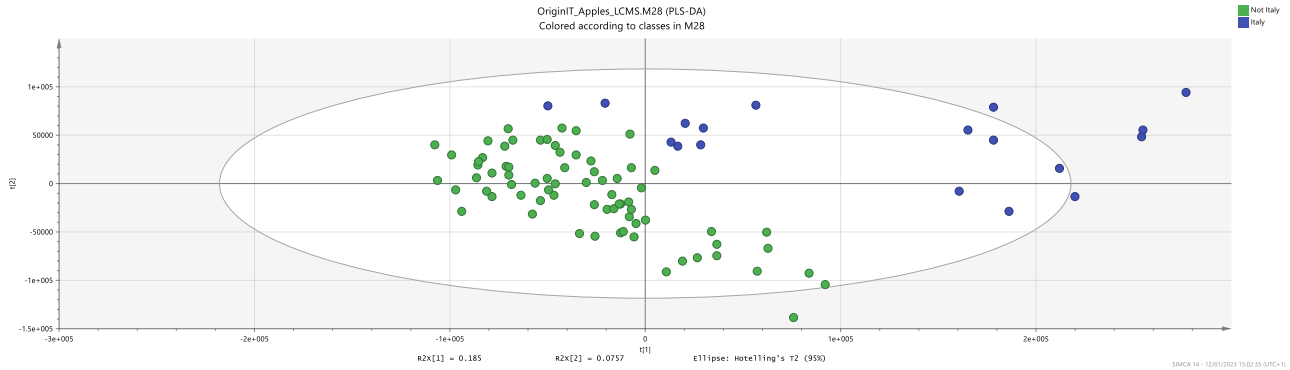


Figure 3A) *PLS-DA score plot of dehydrated apple sample set without QC samples (n=91) in negative ionization mode, after features filtering.* 3B) *PLS-DA score plot of dehydrated apple sample set without QC samples (n=91) in positive ionization mode, after features filtering. (Green dots: Not Italy, blue dots: Italy)*

Tables 3A-B explain the models' performances, by summarising R^2X (cum), R^2Y (cum), and Q^2 of all the models built, for both negative and positive ionization modes, respectively. R^2 is a coefficient that indicates the goodness of fit of the model to the actual data, its value ranges from 0 to 1 (Casella & Berger, 2001). For supervised models, another parameter had to be taken into account, the Q^2, related to the goodness of prediction (Nepomuceno, Cruz Junho, Carneiro-Ramos, & da Silva Martinho, 2021).

A)

Model	PCA (No Filter)	PCA (Filtered)	PLS-DA (No Filter)	PLS-DA (Filtered)
R^2X (cum)	0.7	0.715	0.341	0.344
R^2Y (cum)	/	/	0.930	0.909
Q^2	0.341	0.378	0.688	0.702

B)

Model	PCA (No Filter)	PCA (Filtered)	PLS-DA (No Filter)	PLS-DA (Filtered)

R^2X (cum)	0.769	0.798	0.494	0.516
R^2Y (cum)	/	/	0.933	0.921
Q^2	0.460	0.530	0.758	0.727

Table 3A) *Value of R^2X (cum), R^2Y (cum), and Q^2 parameters of all the models for negative ionization mode.* Table 3B) *Value of R^2X (cum), R^2Y (cum), and Q^2 parameters of all the models for positive ionization mode.*

To assess the robustness of the classification models, a prediction set was generated, by combining some samples picked from the training set, and the test set. A classification list and misclassification table were created to visually investigate the consistency of the models. The misclassification table, or confusion matrix, is useful for machine learning classification questions, in which the outcome is two or more groups. It is a matrix with different combinations, depending on the class number, of actual and predicted values (Narkhede, 2018). The classification List reports the sample IDs, their class IDs, the original dummy variables as YVarPS, which can range from 1 to 0, and the predicted dummy variables as YPredPS. From this value, it is possible to delineate the sample group:

- <0.35 the samples do not belong to the class
- Between 0.35 and 0.65 the samples are borderline
- >0.65 the samples belong to the class

(MKS Umetrics, 2015). Tables 4A-B show misclassification tables of the prediction sets after the feature filtering.

A)

	Members	**Correct**	**Not Italy**	**Italy**	**No class (YPred <= 0)**
Not Italy	32	100%	32	0	0
Italy	11	90.91%	1	10	0
No class	0		0	0	0
Total	43	97.67%	33	10	0

B)

	Members	**Correct**	**Not Italy**	**Italy**	**No class (YPred <= 0)**
Not Italy	32	100%	32	0	0
Italy	11	81.82%	2	9	0
No class	0		0	0	0
Total	43	95.35%	34	9	0

Table 4A) *Misclassification table of dehydrated apple prediction set (n=43 samples) after features filtering, in negative ionization mode.* 4B) *Misclassification table of dehydrated apple prediction set (n=43 samples) after features filtering, in positive ionization mode. (Green cell: samples correctly classified, yellow cell: samples misclassified)*

Tables S1A-B display the classification lists related to the prediction set after feature filtering, always considering the two different ionization modes, and, consequently, the diverse chromatographic

methods employed. Both prediction tables put in evidence a valuable models' performance, with a correctness score of 97.67 and 95.35 % for negative and positive ionization modes prediction sets, respectively.

Molecular identification

The most important features, selected from the multivariate statistical study and the interpretation of the results carried out through Compound Discoverer 3.3 software, were putatively identified, following the steps reported in the "Molecular identification" paragraph of the "Materials and Methods" section. The level of identification reached for both Italian and non-Italian markers was the second and the third, no molecules were identified at the first level due to the unavailability of the reference standards to inject. Table 6 reports the compounds selected and their related information; p-value and adjusted p-value (with Benjamini-Hochberg correction, to lower the false-discovery rate) parameters were also listed, to further prove the statistically significant difference of the features between Italian and non-Italian classes.

Name	Formula	Pseudomolecolar ion	Detected m/z	RT (min)	Mass error (ppm)	CV in QCs (%)	Group areas Italy	Group areas Not Italy	P-value	Adj. P-value	ID type
2-(acetylamino)-3-(1h-indol-3-yl)propanoic acid	C13 H14 N2 O3	[M-H]-1	245.09309	2.337	-0.29	5.09	1.55e8	9.60e7	3.36E-03	2.05E-02	2
Benzoyl-β-D-glucopyranose	C13 H16 O7	[M+FA-H]-1	329.08796	1.775	0.42	9.88	1.53e7	9.45e5	6.69E-09	1.14E-06	3
2-phenylethyl 6-O-(6-deoxy-l-mannopyranosyl)-D-glcopyranoside	C20 H30 O10	[M-H]-1	429.17693	2.806	0.67	8.09	2.95e7	1.85e7	2.53E-04	3.02E-03	3
L-glutamine	C5 H10 N2 O3	[M+H]+1	147.07638	2.334	-0.28	17.85	3.51e7	1.07e8	9.50E-03	4.69E-02	3
L-iditol	C6 H14 O6	[M+H]+1	183.08631	0.859	-0.05	4.09	3.10e8	9.2e7	5.86E-07	3.89E-05	3
Dethiobiotin	C10 H18 N2 O3	[M+H]+1	215.13902	3.056	0.01	10.51	4.80e7	2.89e7	2.22E-05	6.76E-04	3

5-methyl-5'-thioadenosine	C11 H15 N5 O3 S	[M+H]+1	298.09676	0.765	-0.26	7.20	5.51e7	8.89e6	2.40E-09	5.66E-07	2

Table 6 *Summary of identified molecules*

Figures S2-S3 report examples of extracted ion chromatograms (XIC) of a compound detected with the positive ionization mode (dethiobiotin) and another detected with the negative ionization mode (2-(acetylamino)-3-(1h-indol-3-yl) propanoic acid), together with their related spectra. Italian sample XICs, in both positive and negative modes, present peaks more intense than the others of non-Italian samples.

The compounds displayed in table 6 were not previously identified in terms of Italian or non-Italian markers. Some of them were reportedly present in the apple commodity, according to the Human Metabolome Database (HMDB), such as dethiobiotin and L-iditol. Glutamine is an amino acid highly present in various apple cultivars (Ma, et al., 2018). The others could be formed because of the processing phases (blanching for enzymatic inactivation and drying). 2-(acetylamino)-3-(1h-indol-3-yl) propanoic acid presents a chemical structure similar to the tryptophan amino acid, whose amount in a medium-size apple is ca. 2 mg (WebMD Editorial Contributors, 2022). Benzoyl-β-D-glucopyranose and 2-phenylethyl 6-O-(6-deoxy-l-mannopyranosyl)-D-glucopyranoside are aromatic compounds of the apple fruit according to the FooDB database and Søltoft-Jensen and Hansen (Søltoft-Jensen & Hansen, 2005), present as O-glycosidic adducts in the analysed samples. 5-methyl-5'-thioadenosine (MTA) nucleosidase is an enzyme linked to fruit ripening, that acts on the 5-methylthioadenosine for catalysing the formation of methylthioribose (MTR) (Kushad, Richardson, & Ferro, 1985). In the processing of the fruit, hypothetically, the enzymatic inactivation contributed to the degradation of the MTA nucleosidase, favouring the presence of MTA in the dehydrated products. Therefore, putative markers linked to both fresh and processed apples were identified, underlining the originality of the work, and considering the scientific gap concerning the study of the authenticity of the dehydrated matrix.

CONCLUSIONS

In the present work, seven apple compounds were identified as markers for the geographical origin assessment of Italian dehydrated apple cubes. An untargeted metabolomic approach was employed, with two chromatographic methods to separate and study both aromatic/organic and polar components. This methodology coupled with multivariate statistical analysis has led to interesting results, permitting the clustering of Italian samples with a classification model, also confirmed by applying it to a prediction set. No reference standards were available in the research laboratory at the

end of the training and test sets analyses, so the identification was only putative. The next steps could involve standard injections to reach the first level of identification of the seven molecules putatively identified. Further, a targeted approach may consent to quantify these compounds. Once the features are quali-quantitatively analysed, the quality control department of the companies can focus only on these compounds, exploiting less expensive technologies, for facing/preventing authenticity issues.

REFERENCES

Atlas Big. (2021). *World Apple Production by Country*. Retrieved from Atlas Big: https://www.atlasbig.com/en-cn/countries-by-apple-production#:~:text=China%20is%20the%20largest%20apple,50%20%25%20of%20world's%20total%20apple.

Aydogan, C. (2020). Recent advances and applications in LC-HRMS for food and plant natural products: a critical review. *Analytical and Bioanalytical Chemistry*, 1973-1991.

Bai, Y., Xiong, Y., Huang, J., Zhou, J., & Zhang, B. (2019). Accurate prediction of soluble solid content of apples from multiple geographical regions by combining deep learning with spectral fingerprint features. *Postharvest Biology and Technology*, 110943.

Bajoub, A., Carrasco-Pancorbo, A., Ajal, E., Ouazzani, N., & Fernandez-Gutierrez, A. (2015). Potential of LC–MS phenolic profiling combined with multivariate analysis as an approach for the determination of the geographical origin of north Moroccan virgin olive oils. *Food Chemistry*, 292-300.

Casella, G., & Berger, R. L. (2001). *Statistical Inference (second edition)*. Pacific Groove, California, US: Duxbury/Thomson Learning.

Cavanna, D., Catellani, D., Dall'Asta, C., & Suman, M. (2018). Egg product freshness evaluation: A metabolomic approach. *Journal of Mass Spectrometry*, 1-13.

Cavanna, D., Zanardi, S., Dall'Asta, C., & Suman, M. (2018). Ion Mobility Spectrometry coupled to Gas Chromatography: a rapid tool to assess eggs freshness. *Food Chemistry*.

Global Product Prices. (2022). *Apples - prices by country, around the world, June 2022*. Retrieved from Global Product Prices: https://www.globalproductprices.com/rankings/apple_prices/

Katerinopoulou, K., Kontogeorgos, A., Salmas, C. E., Patakas, A., & Ladavos, A. (2020). Geographical Origin Authentication of Agri-Food Products: A Review. *Foods*, 489.

Kunzelmann, M., Winter, M., Aberg, M., Hellenas, K., & Rosén, J. (2018). Non-targeted analysis of unexpected food contaminants using LC-HRMS. *Analytical and Bioanalytical Chemistry*, 5593-5602.

Kushad, M. M., Richardson, D. G., & Ferro, A. J. (1985). 5′-Methylthioadenosine Nucleosidase and 5-Methylthioribose Kinase Activities and Ethylene Production during Tomato Fruit Development and Ripening. *Plant Physiology*, 525–529.

Li, X., Huang, J., Xiong, Y., Zhou, J., Tan, X., & Zhang, B. (2018). Determination of soluble solid content in multi-origin 'Fuji' apples by using FT-NIR spectroscopy and an origin discriminant strategy. *Computers and Electronics in Agriculture*, 23-31.

Ma, S., Neilson, A. P., Lahne, J., Peck, G. M., O'Keefe, S. F., & Stewart, A. C. (2018). Free amino acid composition of apple juices with potential for cider making as determined by UPLC-PDA. *Journal of the Institue of Brewing*, 467-476.

MKS Umetrics. (2015, January 26). *Dynacentrix*. Retrieved from https://www.dynacentrix.com/telecharg/SimcaP/SIMCA14_User_Guide.pdf

Narkhede, S. (2018, May 9). *Understanding Confusion Matrix*. Retrieved from Towards Data Science: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Nepomuceno, G., Cruz Junho, C. V., Carneiro-Ramos, M. S., & da Silva Martinho, H. (2021). Partial least squares—discriminant analysis (PLS-DA). *Scientific Reports*.

Oleszek, W., Lee, C. Y., Jaworski, A. W., & Price, K. R. (1988). Identification of Some Phenolic Compounds in Apples. *Journal of Agricultural and Food Chemistry*, 430-432.

Popping, B., De Dominicis, E., Dante, M., & Nocetti, M. (2017). Identification of the Geographic Origin of Parmigiano Reggiano (P.D.O.) Cheeses Deploying Non-Targeted Mass Spectrometry and Chemometrics. *Foods*.

Righetti, L., Rubert, J., Galaverna, G., Hurkova, K., Dall'Asta, C., Hajslova, J., & Stranska-Zachariasova, M. (2018). A novel approach based on untargeted lipidomics reveals differences in the lipid pattern among durum and common wheat. *Food Chemistry*, 775-783.

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environmental Science & Technology*, 2097-2098.

Selamat, J., Rozani, N. A., & Murugesu, S. (2021). Application of the Metabolomics Approach in Food Authentication. *Molecules*, 7565.

Søltoft-Jensen, J., & Hansen, F. (2005). 15 - New Chemical and Biochemical Hurdles. In D. Sun, *Emerging Technologies for Food Processing* (pp. 387-416). Academic Press.

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., . . . Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, 211-221.

Urrestarazu, J., Miranda, C., Santestaban, L. G., & Royo, J. B. (2012). Genetic diversity and structure of local apple cultivars from Northeastern Spain assessed by microsatellite markers. *Tree Genetics & Genomes*, 1163-1180.

WebMD Editorial Contributors. (2022, November 22). *Top Foods High in Tryptophan.* Retrieved from Nourish by WebMD: https://www.webmd.com/diet/foods-high-in-tryptophan#:~:text=Additionally%2C%20a%20medium%2Dsized%20apple,milligrams%20of%20this%20amino%20acid.

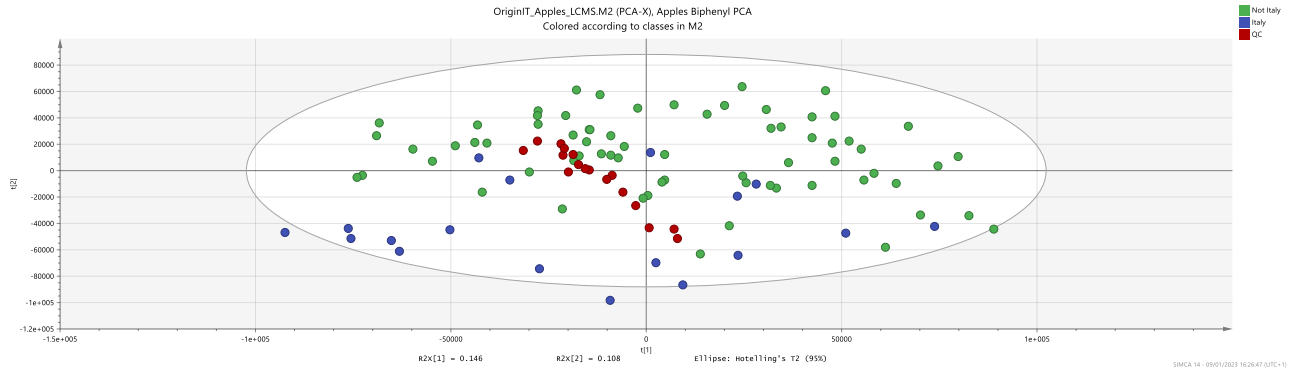
World's Top Exports. (2021). *Apples Exports by Country Plus Average Prices*. Retrieved from World's Top Exports: https://www.worldstopexports.com/apples-exports-by-country/

Wu, H., Yue, T., & Yuan, Y. (2018). Authenticity Tracing of Apples According to Variety and Geographical Origin Based on Electronic Nose and Electronic Tongue. *Food Analytical Methods*, 522-532.

Zhang, J., Nie, J., Kuang, L., Shen, Y., Zheng, H., Zhang, H., . . . Asim, S. (2019). Geographical origin of Chinese apples based on multiple element analysis. *Journal of the Science of Food and Agriculture*, 6182-6190.

SUPPLEMENTARY MATERIALS

A)



B)

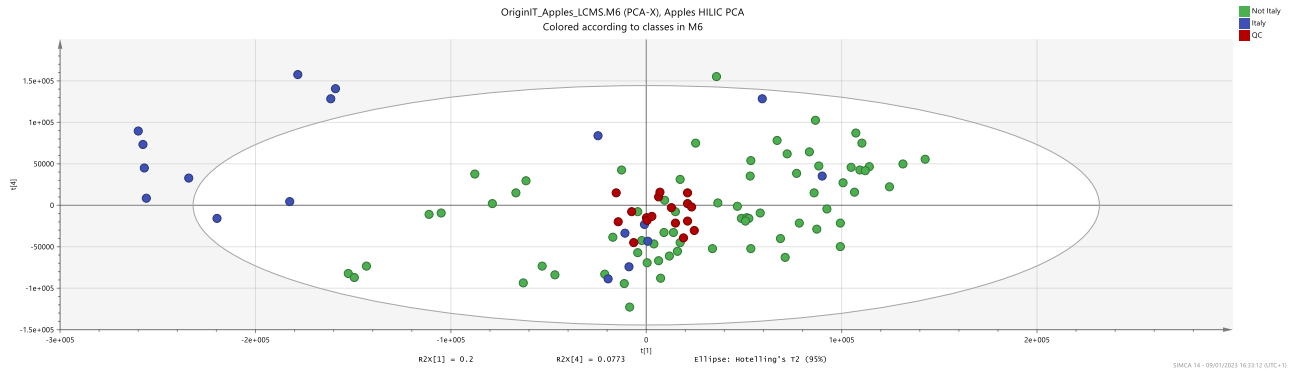


Figure S1A) *PCA score plot of the dehydrated apple sample set (n= 108) in negative ionization mode.* S1B) *PCA score plot of the dehydrated apple sample set (n= 108) in positive ionization mode. (Green dots: Not Italy, blue dots: Italy, red dots: QC samples)*

A)

Obs ID	ClassID	YVarPS (Not Italy)	YPredPS (Not Italy)	YVarPS (Italy)	YPredPS (Italy)
167783_A	Italy	0	-0.184852	1	1.18485
167783_B	Italy	0	-0.13514	1	1.13514
167787_A	Italy	0	0.256506	1	0.743494
167787_B	Italy	0	0.160604	1	0.839396
169008_A	Italy	0	-0.0431086	1	1.04311
169008_B	Italy	0	-0.0177169	1	1.01772
169011_A	Italy	0	0.255948	1	0.744052
169011_B	Italy	0	0.115115	1	0.884885
169013_B	Italy	0	0.278567	1	0.721433
166468_A	Not Italy	1	1.17823	0	-0.178228
166468_B	Not Italy	1	1.06468	0	-0.0646837
166469_A	Not Italy	1	0.924153	0	0.0758473
166469_B	Not Italy	1	0.870226	0	0.129774
166470_A	Not Italy	1	1.04317	0	-0.0431725
166470_B	Not Italy	1	0.967628	0	0.0323715
166474_A	Not Italy	1	1.01125	0	-0.0112524
166474_B	Not Italy	1	0.927905	0	0.0720953
166477_A	Not Italy	1	0.936796	0	0.0632045
166477_B	Not Italy	1	1.02038	0	-0.0203755
166483_A	Not Italy	1	1.08873	0	-0.0887262
166483_B	Not Italy	1	1.07165	0	-0.0716519
166489_A	Not Italy	1	0.911674	0	0.0883261
166489_B	Not Italy	1	0.891794	0	0.108206
167044_A	Not Italy	1	1.30393	0	-0.303931
167044_B	Not Italy	1	1.1328	0	-0.132804
167048_A	Not Italy	1	0.877042	0	0.122958
167048_B	Not Italy	1	0.86998	0	0.13002
167056_A	Not Italy	1	0.779747	0	0.220253
167056_B	Not Italy	1	0.830066	0	0.169934
167060_A	Not Italy	1	1.04965	0	-0.0496474
167060_B	Not Italy	1	0.980697	0	0.0193034
167061_A	Not Italy	1	0.832421	0	0.167579
167061_B	Not Italy	1	0.869751	0	0.130249
167067_A	Not Italy	1	1.04294	0	-0.042942
167067_B	Not Italy	1	1.05457	0	-0.0545718

167068_A	Not Italy	1	0.93451	0	0.0654898
167068_B	Not Italy	1	0.912171	0	0.0878295
Chile_1A_LM_cb_PL_2022	Not Italy	1	1.02139	0	-0.021387
Chile_1B_LM_cb_PL_2022	Not Italy	1	1.05368	0	-0.0536801
Hungary_1_NM_cb_GS_2021	Not Italy	1	1.10493	0	-0.104933
Hungary_2_LM_GS_2021	Not Italy	1	0.968348	0	0.0316519
Italy_1_LM_cb_PL_2022	Italy	0	0.501883	1	0.498117
Italy_2A_LM_PL_2022	Italy	0	0.451294	1	0.548706

B)

Obs ID	ClassID	YVarPS (Not Italy)	YPredPS (Not Italy)	YVarPS (Italy)	YPredPS (Italy)
166468_A_2_1	Not Italy	1	0.961921	0	0.0380791
166468_B_2_1	Not Italy	1	0.824016	0	0.175984
166469_A_2_1	Not Italy	1	0.863542	0	0.136458
166469_B_2_1	Not Italy	1	0.693409	0	0.306591
166470_A_2_1	Not Italy	1	1.11762	0	-0.117621
166470_B_2_1	Not Italy	1	0.95849	0	0.0415103
166474_A_2_1	Not Italy	1	0.863923	0	0.136077
166474_B_2_1	Not Italy	1	0.905372	0	0.0946279
166477_A_2_1	Not Italy	1	1.00451	0	-0.00450592
166477_B_2_1	Not Italy	1	0.932374	0	0.0676258
166483_A_2_1	Not Italy	1	1.02593	0	-0.0259312
166483_B_2_1	Not Italy	1	0.967502	0	0.0324984
166489_A_2_1	Not Italy	1	1.05354	0	-0.0535423
166489_B_2_1	Not Italy	1	1.00351	0	-0.00351363
167044_A_2_1	Not Italy	1	0.671337	0	0.328663
167044_B_2_1	Not Italy	1	0.777235	0	0.222765
167048_A_2_1	Not Italy	1	0.989391	0	0.0106094
167048_B_2_1	Not Italy	1	1.0132	0	-0.0131963
167056_A_2_1	Not Italy	1	0.789286	0	0.210714

167056_B_2_1	Not Italy	1	0.967817	0	0.0321827
167060_A_2_1	Not Italy	1	0.993479	0	0.00652093
167060_B_2_1	Not Italy	1	1.00081	0	-0.000805363
167061_A_2_1	Not Italy	1	0.922706	0	0.0772945
167061_B_2_1	Not Italy	1	0.943395	0	0.0566046
167067_A_2_1	Not Italy	1	1.03514	0	-0.0351386
167067_B_2_1	Not Italy	1	0.851433	0	0.148567
167068_A_2_1	Not Italy	1	1.01126	0	-0.0112625
167068_B_2_1	Not Italy	1	0.972333	0	0.027667
167783_A_2_1	Italy	0	0.0674541	1	0.932546
167783_B_2_1	Italy	0	-0.025593	1	1.02559
167787_A_2_1	Italy	0	-0.0747906	1	1.07479
167787_B_2_1	Italy	0	0.0818274	1	0.918173
169008_A_2_1	Italy	0	-0.0522812	1	1.05228
169008_B_2_1	Italy	0	-0.0277551	1	1.02776
169011_A_2_1	Italy	0	0.156497	1	0.843503
169011_B_2_1	Italy	0	0.0433602	1	0.95664
169013_B_2_1	Italy	0	0.360829	1	0.639171
Chile_1A_LM_cb_PL_2022_2_1	Not Italy	1	0.946381	0	0.0536186
Chile_1B_LM_cb_PL_2022_2_1	Not Italy	1	0.909414	0	0.0905861
Hungary_1_NM_cb_GS_2021_2_1	Not Italy	1	0.950754	0	0.0492455
Hungary_2_LM_GS_2021_2_1	Not Italy	1	0.963823	0	0.0361767
Italy_1_LM_cb_PL_2022_2_1	Italy	0	0.784413	1	0.215587
Italy_2A_LM_PL_2022_2_1	Italy	0	0.603577	1	0.396423

Table S1A) *Classification List of dehydrated apple prediction set after features filtering, in negative ionization mode.* S1B) *Classification List of dehydrated apple prediction set after features filtering, in positive ionization mode. The numbers in the Sample ID indicate different batches, whereas the letters indicate the replicate. (NM: normal moisture, LM: low moisture, PL: Pink Lady, GS: Granny Smith) (Green YPredPS value: sample belongs to the class, orange YPredPS value: sample is borderline, white YPredPS value: sample does not belong to the class)*

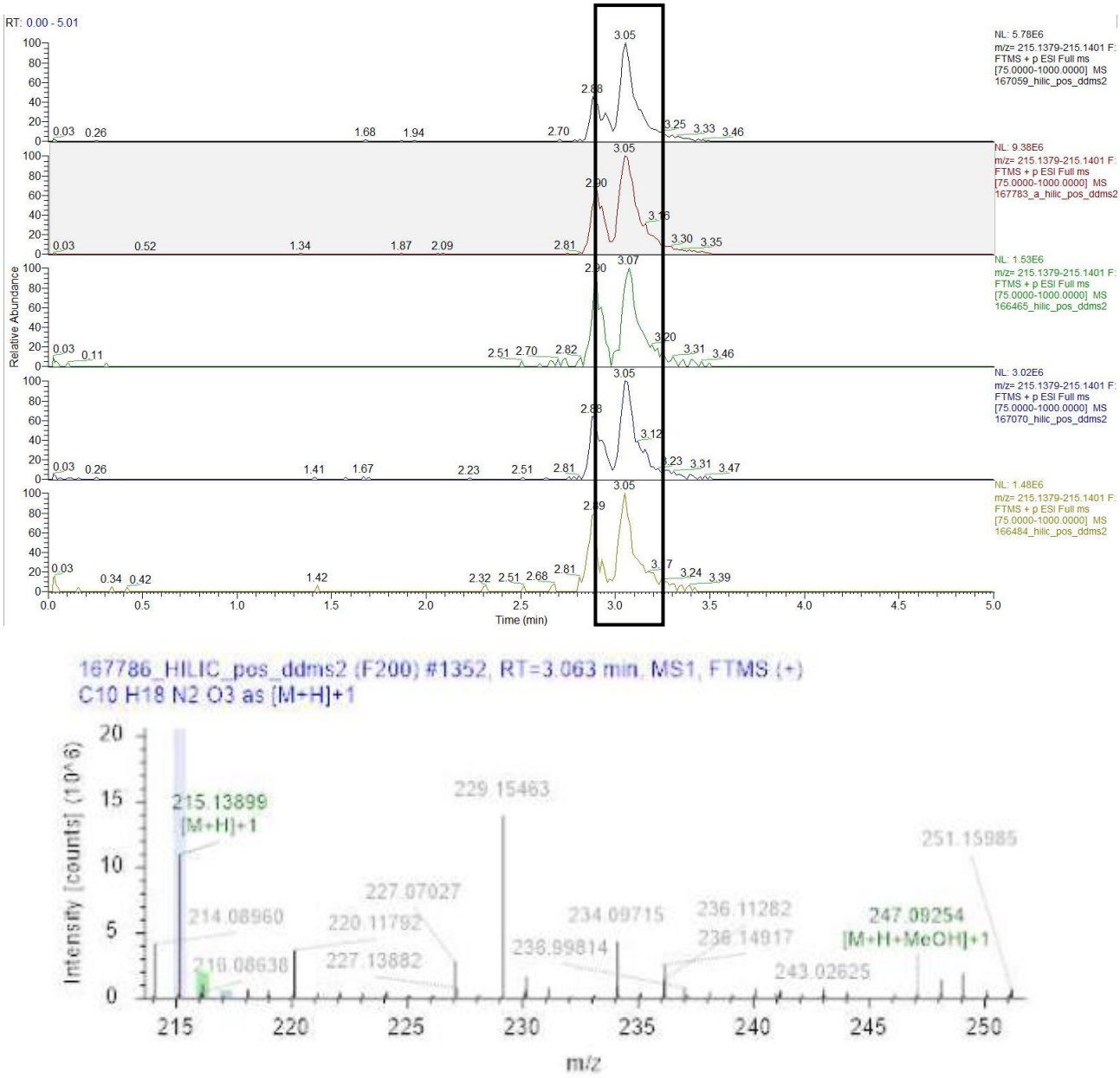


Fig. S2 *Above: extracted ion chromatograms (XIC) of putatively identified dethiobiotin molecule in samples from different countries (Positive ionization mode) (Black XIC: China, NL: 5.78E6, red XIC: Italy, NL: 9.38E6, green XIC: France, NL: 1.53E6, blue XIC: Poland, NL: 3.02E6, yellow XIC: Hungary, NL: 1.48E6). Below: MS spectrum is taken from the XIC peak.*

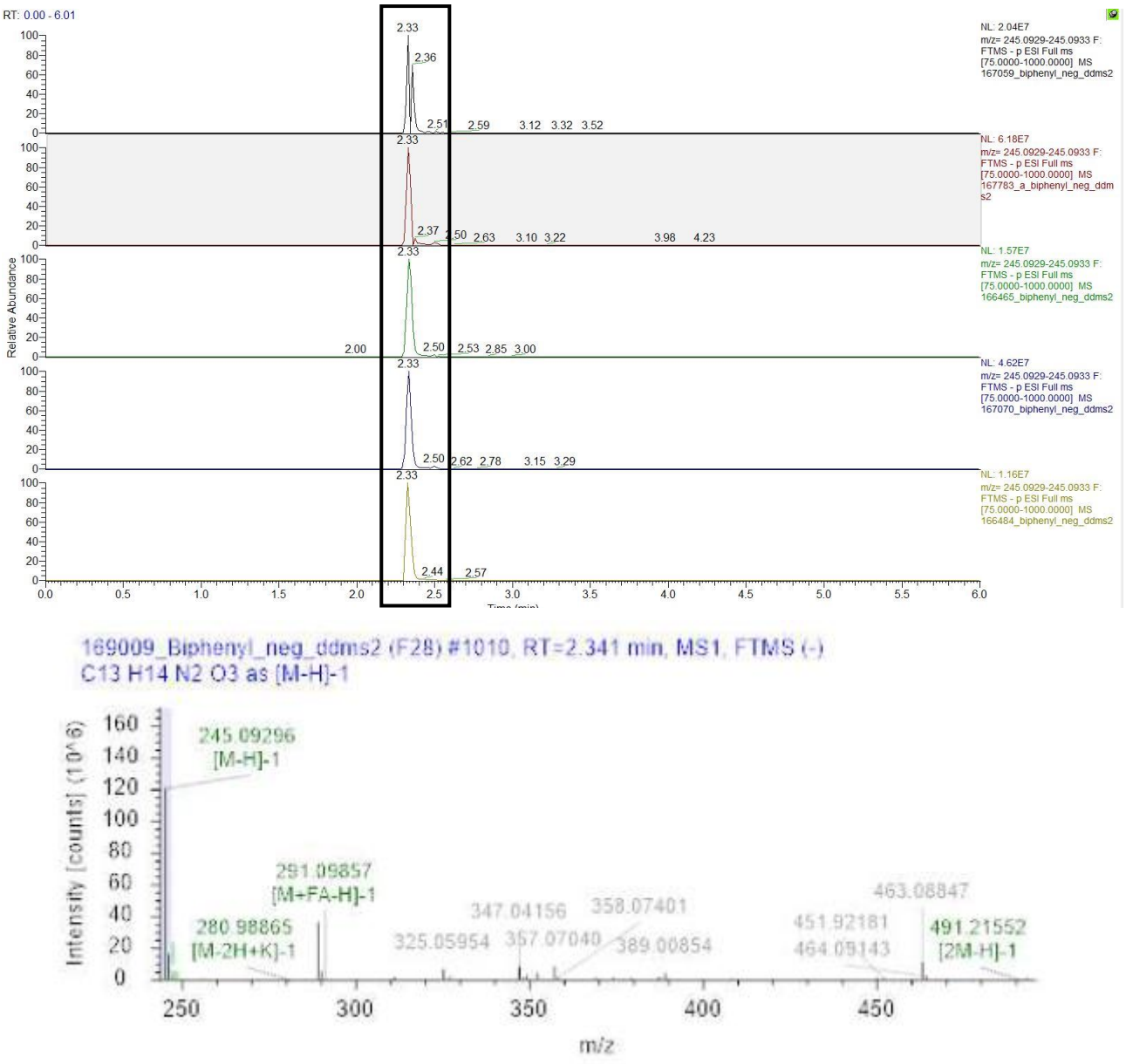


Fig. S3 *Above: extracted ion chromatograms (XIC) of putatively identified 2-(acetylamino)-3-(1h-indol-3-yl) propanoic acid molecule in samples from different countries (Negative ionization mode) (Black XIC: China, NL: 2.04E7, red XIC: Italy, NL: 6.18E7, green XIC: France, NL: 1.57E7, blue XIC: Poland, NL: 4.62E7, yellow XIC: Hungary, NL: 1.16E7). Below: MS spectrum is taken from the XIC peak.*

CHAPTER 5

OTHER ANALYTICAL STRATEGIES: CASE STUDIES

OTHER ANALYTICAL APPROACHES

In the present thesis work, other two analytical strategies have been employed, atmospheric solid analysis probe-mass spectrometry (ASAP-MS) and nuclear magnetic resonance (NMR) spectroscopy. A bigger amount of data is needed, as well as the analytical methods developed could be optimised. In spite of that, the results obtained look promising, hence the preliminary outcomes will be presented.

NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY FOR FOOD AUTHENTICITY ISSUES

Giuseppe Sammarco[1,3], Birk Schütz[4], Claudia Napoli[4], Chiara Dall'Asta[3], Michele Suman[1,2]

[1] Advanced Laboratory Research, Barilla G. e R. Fratelli S.p.A., Parma, Italy
[2] Department for Sustainable Food Process, Catholic University Sacred Heart, Piacenza, Italy
[3] Department of Food and Drug, University of Parma, Parma, Italy
[4] Bruker BioSpin GmbH, Rudolf-Plank-Straße 23, 76275 Ettlingen, Germany

INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is a robust analytical strategy, used systematically in food analysis and authenticity for years. Its first application in the food field was achieved in 1957 to measure moisture in foods through low-resolution NMR (Alberti, Belton, & Gil, 2002). The technique was consistently employed from the 1980s, to deal with the complexity of the food matrices. From then on, NMR applications in food science were commonly reported in the scientific literature, journals and books.

Foods represent very heterogeneous systems, encompassing many chemical compounds, that could present diverse compositions, depending on the harvesting, processing, storing, and maturation conditions. In this context, one-dimensional liquid or solid-state high-resolution NMR spectroscopy returns, within a single experiment, relevant structural and quantitative information as NMR parameters, such as chemical shifts, coupling constants, and signal intensities. Several nuclei can be chosen among ^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{31}P… according to the information needed. The above-mentioned experiments do not require any separation or particular sample preparation. In the case of the strong complexity of the food matrix, there is an important number of NMR-based methodologies able to solve these issues, from the homo- and hetero-nuclear multidimensional NMR to the coupling with separation techniques, like liquid chromatography (LC-NMR). Two-dimensional NMR experiments,

such as COSY, TOCSY, NOESY, HSQC… provide information related to the "communication" between nuclei (spin-spin or dipolar coupling), unveiling their connections, and facilitating the molecule characterization in the food product.

NMR data can be handled and elaborated through multivariate statistical approaches, in order to extract as much information as possible. This strategy, named metabonomics, does not strictly ask for the identification of the NMR signals, as in quantitative NMR, but it searches for features that can identify unique metabolites or key biomarkers. Both unsupervised and supervised multivariate statistical models are exploited for data visualization, plotting the spectral fingerprints, and used for studying metabolic changes and discriminating among sample classes (Spyros & Dais, Introduction, 2015). In particular, principal component analysis (PCA) is the unsupervised model aimed at using a reduced number of variables to explain most of the variation. Partial least squares-discriminant analysis (PLS-DA) is the supervised model that evaluates the labelled samples, in order to classify them, and assess the robustness and the accuracy of the technology; linear discriminant analysis (LDA), artificial neural networks (ANNs), and independent component analysis (ICA) could also have a key role as statistical methods in food assessment (Trygg, Holmes, & Lundstedt, 2007).

Sobolev *et al.* described in a review three starting points to summarise the interests of industries, suppliers, consumers and researchers. The first point concerns the food to be analysed, as the common interest is the deep knowledge of the food product. This could help producers in boosting agricultural practices, and companies in the industrial processes. Another point is about food frauds, to face food safety and authenticity issues, such as adulteration, geographical and botanical origin assessment. The third point regards the versatility of the NMR analytical strategy, as a single methodology could be applied to a wide range of food-related challenges (Sobolev, Ingallina, Spano, Di Matteo, & Mannina, 2022).

Olive oil, wine, and honey are among the most popular commodities known for authenticity frauds, such as the false declaration of origin, that have been faced through the NMR approach. Extra virgin olive oils from different geographical areas were studied via ^{1}H NMR spectroscopy. PCA and orthogonal PLS-DA (OPLS-DA) highlighted differences between Italian and Tunisian olive oils; further, the metabolic map NMR-related was able not only to evaluate the geographical provenience of the olive oils but also find a correlation with the climatic conditions (Rongai, et al., 2017). Another study focused on the geographical origin of olive oil samples from five Turkish regions. ^{1}H magnetic resonance imaging (MRI) technology, merged with the analysis of variance (ANOVA) test, was exploited, and it underlined diverse compositions of linoleoyl groups in the analysed samples. The fatty acyl content was studied as well, and this permitted us to distinguish between three Turkish regions, that differ from each other for the ^{1}H signal of the oleoyl content (Ok, 2014) (Un & Ok,

2018). The phenolic fraction was found to be relevant for the origin assessment of olive oil. ^{1}H and ^{31}P NMR were employed to discriminate among samples from four areas of Greece, and the phenolic compounds were differently distributed in samples having different varieties, from different cultivation zones (Laincer, et al., 2016) (Agiomyrgianaki, Petrakis, & Dais, 2012).

Concerning the wine commodity, minor compounds could play a key role in solving authenticity issues. However, water and ethanol signals have to be suppressed in the ^{1}H spectrum, in order to not interfere with the others (Esslinger, Fauhl-Hassek, & Wittkowski, 2015). Studies about Czech and German wines were conducted. Proton spectrum and multivariate statistics could be effectively combined to predict wine type and variety with valuable accuracy, whereas another work highlighted the spectral region between 5.1 and 9.8 ppm as the most suitable for the geographical provenience and variety evaluation of red wine samples (Mascellani, et al., 2021) (Magdas, Pirnau, Feher, Guyon, & Cozar, 2019). Bordeaux and Cabernet Sauvignon red wines were characterised by an NMR-based metabolomics approach, by quantifying metabolites from diverse classes. Soil and cultivation played a crucial part as well. The outcomes underlined the capability of this method to discern wines depending on the soil, the harvesting year, and their variety (Xu, et al., 2021) (Gougeon, da Costa, Guyon, & Richard, 2019).

Honey's chemical profile changes according to climatic and environmental conditions, and various molecules were identified as markers for geographical and botanical origins. An NMR metabolomic method permitted to analyse organic extracts from Acacia honey samples, with the aim to define their geographical origin. A differentiation was excellently accomplished between Italian and Eastern European samples, with a correct classification rate of 100 % (Schievano, Stocchero, Zuccato, Conti, & Piana, 2019). Sugar, free amino acids, organic acids, and sugar/water ratios were exploited for geographical evaluations, applying a method based on NMR and high-pressure liquid chromatography (HPLC). In particular, two discriminant functions were able to discern honey samples from different Greek regions, with a prediction accuracy of around 80 % (Karabagias, et al., 2018). OPLS-DA model was applied for picking spectral regions in a ^{1}H NMR experiment to define the geographical origin of honey samples obtained from three diverse sources. Samples from China, Hungary, Italy, and South America were classified with a surprising variance (Consonni, Cagliani, & Cogliati, 2013).

The present thesis project reports a case study concerning the geographical origin assessment, aimed at correctly classifying Italian hazelnut and hazelnut-based product samples (fresh, roasted, and paste hazelnuts) via untargeted ^{1}H NMR analysis, exploiting both high-resolution instrument (400 MHz) and low-resolution one (80 MHz).

HAZELNUT CHAIN CASE STUDY: AN NMR SPECTROSCOPY PRACTICAL APPROACH

Sampling and Instrumental parameters

Fresh, roasted, and paste hazelnut authentic samples were sampled from Georgia, Azerbaijan, Turkey, and Italy as the main geographical target. A design of experiment (DoE) was created, considering several factors, such as the harvesting campaign (2020 and 2021), the presence of peeling, and the industrial process. The Italian samples include "Tonda Gentile delle Langhe" from the Piedmont region, "Nocciola Romana" from the Latium region, and "Mortarella" from the Campania region. For each matrix, these three varieties were equally mixed to have 'Italian samples' (N=21, 9 fresh, 9 roasted, and 3 paste hazelnuts) Other samples were collected from Turkey, Azerbaijan, and Georgia, for a total of 59 non-Italian samples (19 Georgian, 20 Azerbaijani, 20 Turkish). All the samples were stored in a cold room, with a controlled temperature of 4-6 °C. All of them were selected for the 400 MHz analysis, whereas only roasted hazelnut samples (n=48) were considered for a proof-of-concept study with the low-field instrument (80 MHz).

Regarding the instrumental conditions, the ^{1}H NMR spectra were acquired at 300 K on a FoodScreener® NMR spectrometer (Bruker Biospin, Ettlingen, Germany), operating at a frequency of 400 MHz, using Bruker automatic sample changer (B-ACS 60). For each spectrum, 32 scans were acquired, the data were collected into 131K points, then reduced to 65K points, with a spectral width of 20 ppm, and an acquisition time of 3.895 s. The water and methanol signals were suppressed by performing the one-dimensional nuclear overhauser enhancement spectroscopy (NOESY) pulse sequence. The acquisitions were automatically achieved through the ICON-NMR and SampleTrack software (Bruker Biospin, Ettlingen, Germany), in ca. 12 min per sample.

The analyses with low-resolution instrumentation were carried out employing the Fourier 80® Benchtop NMR spectrometer (Bruker Biospin, Ettlingen, Germany), operating at a frequency of 80 MHz, and a temperature of 300 K, using the PAL RSI autosampler (CTC Analytics AG, Zwingen, Switzerland). For each spectrum, 96 scans were acquired, the data were collected into 32K points, then reduced to 16K points, with a spectral width of 20 ppm, and an acquisition time of 5.079 s. The water and methanol signals were suppressed by performing the one-dimensional NOESY pulse sequence. The acquisitions were automatically achieved through the ICON-NMR and SampleTrack software (Bruker Biospin, Ettlingen, Germany).

Sample preparation

10 g of each fresh and roasted hazelnut sample were milled at 10 000 rpm for 2 minutes with the knife mill GrindoMix GM200 (Retsch, Hann, Germany). 300 ± 5 mg of each sample (both milled fresh and roasted hazelnuts and hazelnut paste) were weighed in a 15 mL falcon tube. 2 mL of methanol (Merck Millipore, Burlington, Massachusetts, US) was added to the samples, then they were vortexed with the vortex shaker Intelli-Mixer RM-2 Rotator (Cole-Parmer, Vernon Hills, Illinois, US). The solutions were centrifuged with an EBA 200 S centrifuge (Hettich, Tuttlingen, Germany), at 6000 rpm for 15 minutes. 900 µL of supernatant were pipetted into a 1.8 mL cryovial. 100 µL of methanol-d_4 containing 0.03 % tetramethylsilane (TMS) (Sigma-Aldrich, St. Louis, Missouri, US). Each sample was then vortexed for 30 seconds with an IKA® Vortex 2 (IKA, Staufen, Germany) and 600 µL of the extract was pipetted into a 5 mm NMR tube (Sigma-Aldrich, St. Louis, Missouri, US). Each NMR tube was barcoded with labels generated with SampleTrack software, and then they were degassed for 20 seconds in an ultrasonic bath, before the NMR analysis.

Multivariate statistical analysis

Two different chemometrics strategies were approached, employing two platforms. The analysis of the spectra was carried out via MatLab programming language (The Mathworks, Version R2022b). The part of spectra considered ranged from 0.7 to 6.0 ppm, binned in 300 buckets, and excluding 4.5-5.2 and 2.8-3.6 ppm sections. The classification workflow started with the unsupervised PCA for the dimensionality reduction, so the model preserved a smaller number of components that explained the most variance of the original data in a lower dimensional space. The explained variance chosen for the dimension reduction was 99.93 %. The actual classification was accomplished through LDA with a nearest-class-mean classifier, and the model was then validated by exploiting the Monte-Carlo (10 runs), and the cross-validation (8-set) methods, building 8x10 models.

Python programming language (Python Software Foundation, Version 3.8.12) was also used for the multivariate statistical analysis. All the features of the spectra were extracted, then unsupervised PCA for the dimensionality reduction was carried out. The classification was achieved with the supervised PLS-DA, and the validation was done with the PLS regression, after creating dummy variables to represent the categories of the categorical independent variables.

Results & Discussion

^{1}H NMR analyses at 400 MHz

The high-resolution analyses of the methanolic extracts of fresh, roasted, and paste hazelnut samples returned spectra with the density of signals between 0 and 6 ppm. Considering the spectrum obtained

(Figure 1), and the approximative proton chemical shift of the common functional groups' table (Liu, 2021), it was possible to hypothesize the class compounds present in the samples.

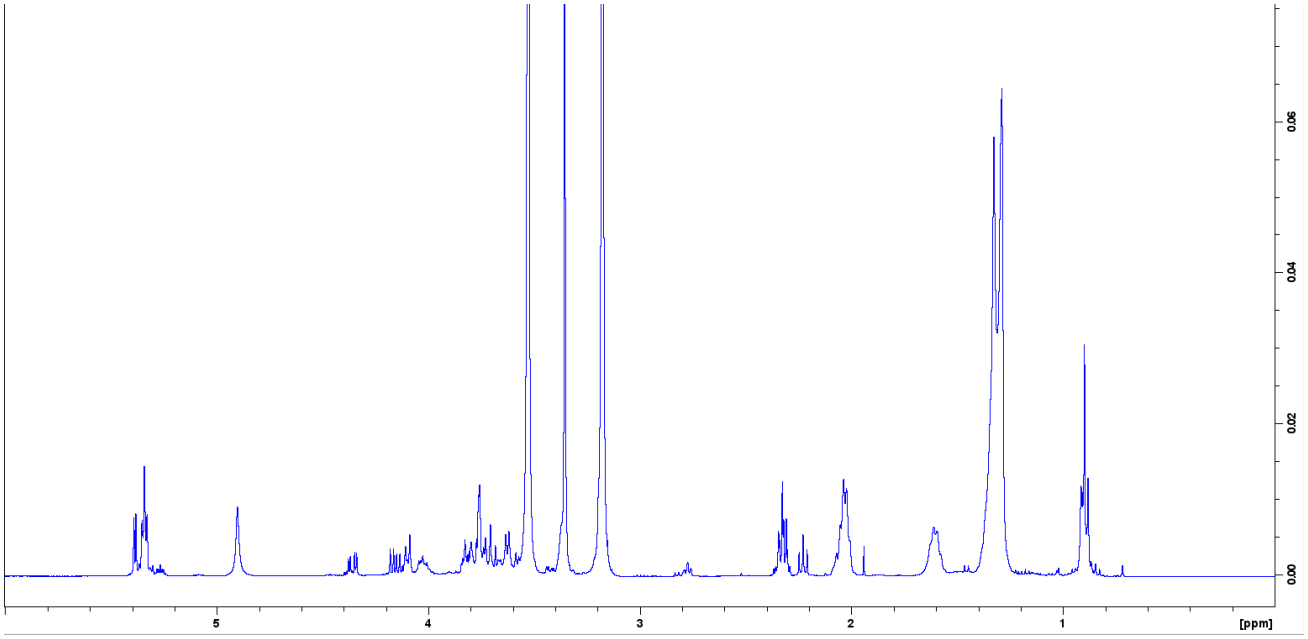


Fig. 1 *1H NMR spectrum of fresh hazelnut sample, acquired at 400 MHz. (From 0 to 6 ppm)*

The signals at the beginning of the spectrum (1-2 ppm) put in evidence the presence of hydrogen in the C-H bond, without any other groups in proximity, that could be related to the hydrocarbon chains of the fatty acids. The signals between 1.5 and 2.5 ppm could be linked to ketones, alkenes, and aromatic compounds, as they are typical of hydrogen in C-H bonds beside double bonds, such as C=O or C=C. The most intense signals from 3 to 4 ppm could be due to the bond of the hydrogen atom with a carbon connected to an electronegative atom, such as oxygen. Thus, these signals could be generated by the presence of the glycerol ethers (glycosylglycerides, phosphoglycerides, phosphatidylglycerols…) or esters, like triacyl-glycerols as one of the main components of the analysed matrix (Cialiè Rosso, et al., 2021). The other signals from 4 to 6 ppm could be related to the presence of phenolic compounds, present in both kernels and peels, and amines, that are mostly formed during the roasting process (Tas, 2017).

The TopSpin software (Version 4.1.4, Bruker Biospin, Ettlingen, Germany) permitted to display different spectra at the same time, so a preliminary visual comparison between them was done. Figure 2 shows an example of spectra comparison, considering 4 fresh hazelnut samples from diverse geographical areas.

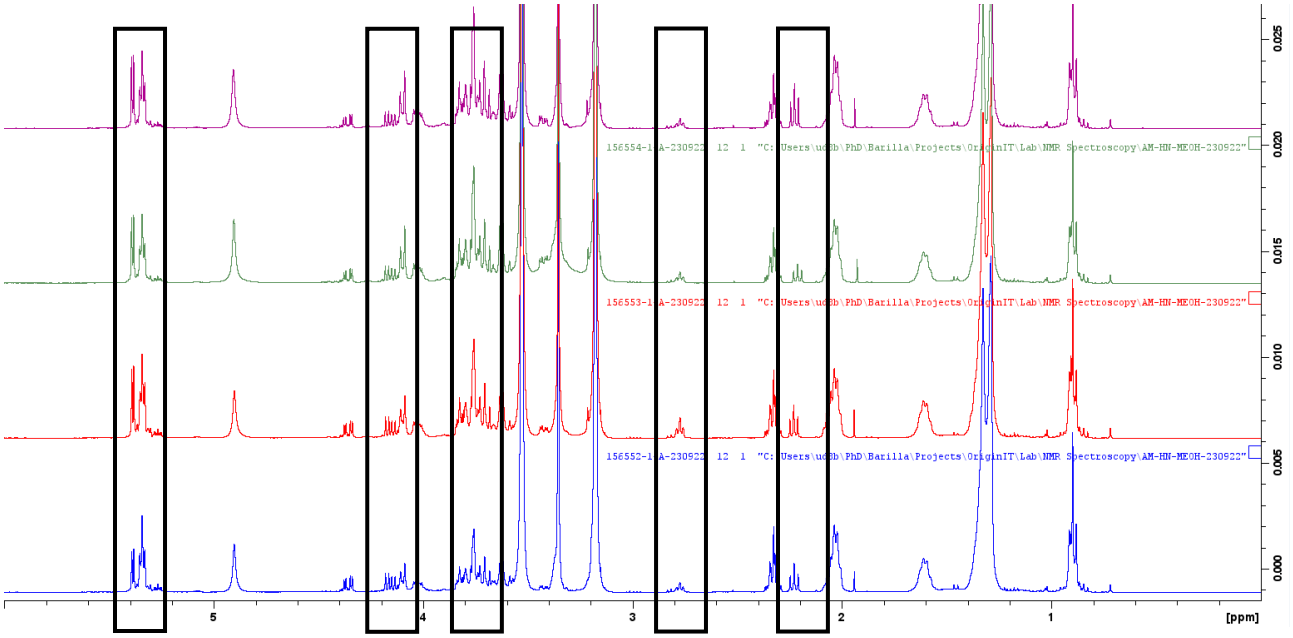


Fig. 2 *Fresh hazelnut sample ^{1}H NMR 400 MHz spectra comparison. (Blue spectrum: Georgia, red spectrum: Azerbaijan, green spectrum: Turkey, purple spectrum: Italy)*

The black rectangles in the figure highlight the macroscopic visual differences in NMR signals among the samples from the geographical origin selected. To render statistically significant these hypothetic differences, a multivariate statistical analysis was performed. The data elaboration via MatLab software exploited the LDA as a classification model, coupled with the nearest centroid classifier for the final classification. A train-test split was also done, by picking 20 % of the samples from the training set to generate the test set, and the model validation was realized with the Monte-Carlo and the cross-validation methods. Figure 3 presents the LDA plot of the entire sample set analysed, and the confusion matrix that summarises the model performance and its correctness percentage score.

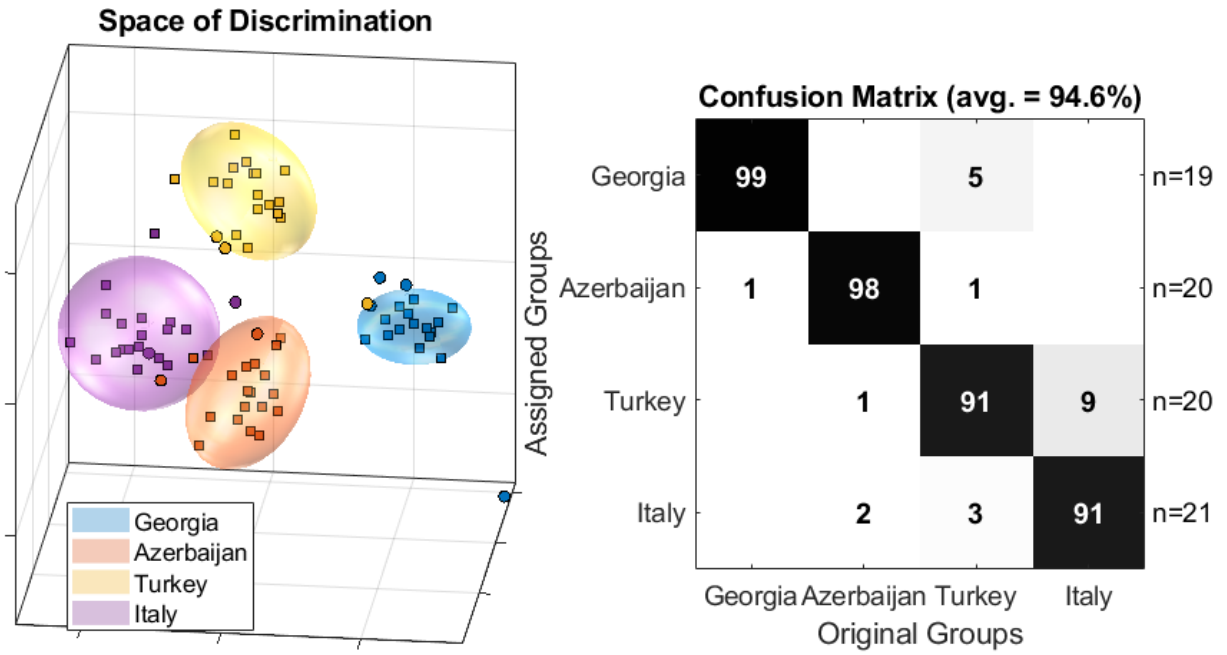


Fig. 3 *LDA plot and confusion matrix of the entire hazelnut sample set. (Squares: training set samples, circles: test set samples, blue: Georgia, orange: Azerbaijan, yellow: Turkey, purple: Italy)*

In this image is possible to evaluate the goodness of the model built, as the classes were properly separated, according to the geographical origin, and the test set samples (circles) were well-classified. The correctness score of the model had, indeed, a relevant value of 94.6 %.

The same sample set was elaborated with Python software, following a different classification protocol. The supervised model applied was PLS-DA, and the PLS was used to make the prediction on the test set (20 % of the samples picked from the training set). Figure 4 shows the PLS-DA plot of the entire sample set analysed.

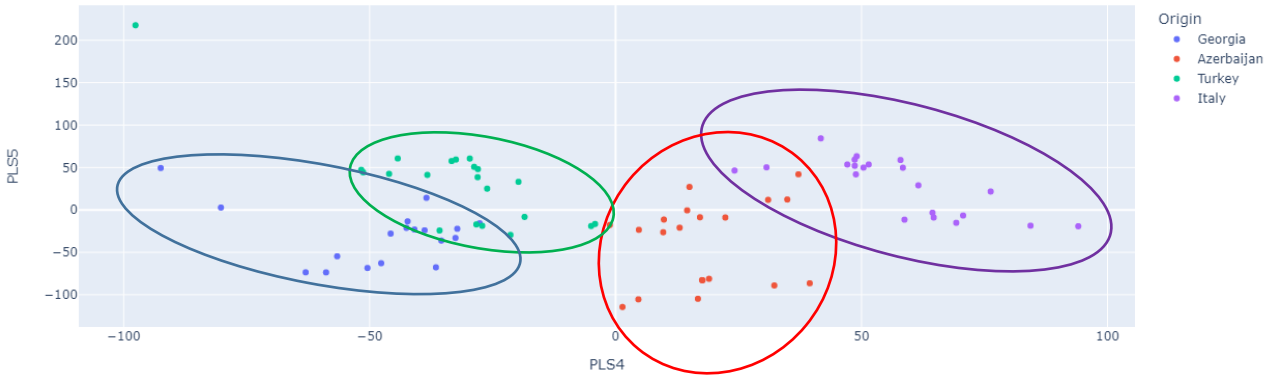


Figure 4 *PLS-DA plot of the entire hazelnut sample set. (Blue dots and cluster: Georgia, red dots and cluster: Azerbaijan, green dots and cluster: Turkey, purple dots and cluster: Italy)*

As visible, the samples are separated quite well depending on the geographical provenience, and the model consistency was assessed by the PLS regression model on the test set. Table 1 resumes averaged precision, recall, f1-score, and the accuracy percentage score of the model.

ACCURACY	AVG. PRECISION	AVG. RECALL	F1-SCORE
93.75 %	95 %	93.75 %	93.75 %

Table 1 *Summary of model performance parameters: accuracy, averaged precision, averaged recall, and averaged f1-score.*

Both data elaboration workflows reported interesting outcomes, with good class clustering, and valuable performance parameters, studying the same sample set on two different platforms.

^{1}H NMR analyses at 80 MHz

The low-resolution analyses, as expected, provided less resolved spectra, with a higher signal-to-noise ratio, which made the preliminary visual interpretation, as well as the statistical elaboration, more complicated. Figures 5 and 6 display, respectively, an example of a roasted hazelnut sample spectrum and the spectra comparison of samples from the selected countries.

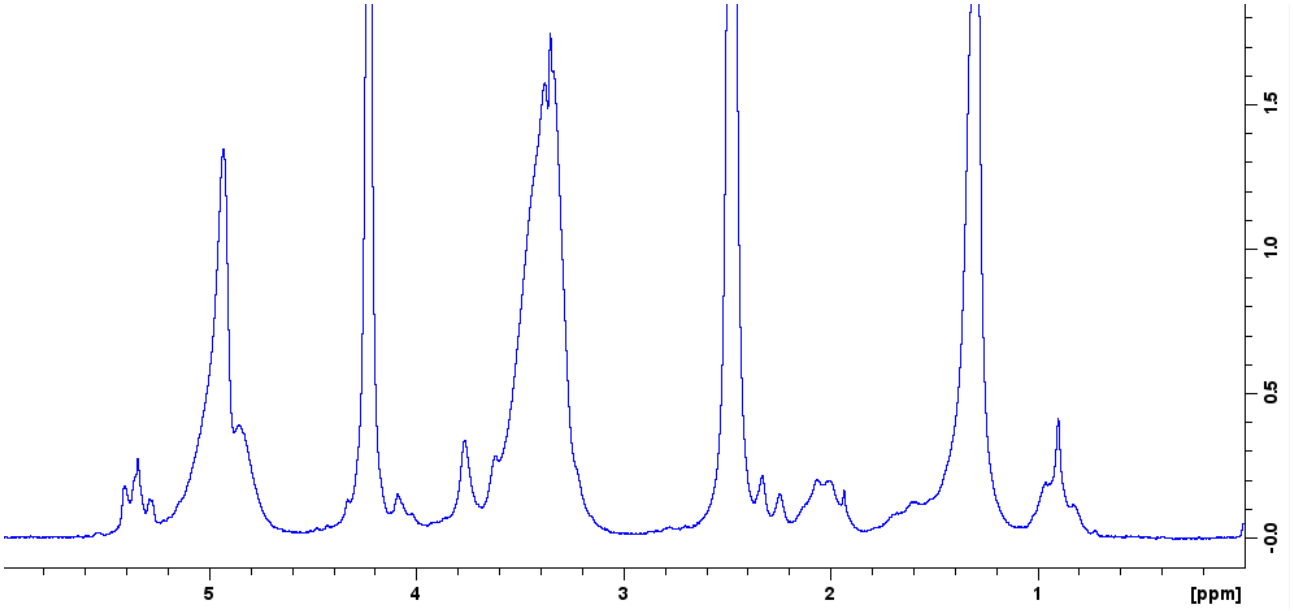


Fig. 5 *^{1}H NMR spectrum of roasted hazelnut sample, acquired at 80 MHz. (From 0 to 6 ppm)*

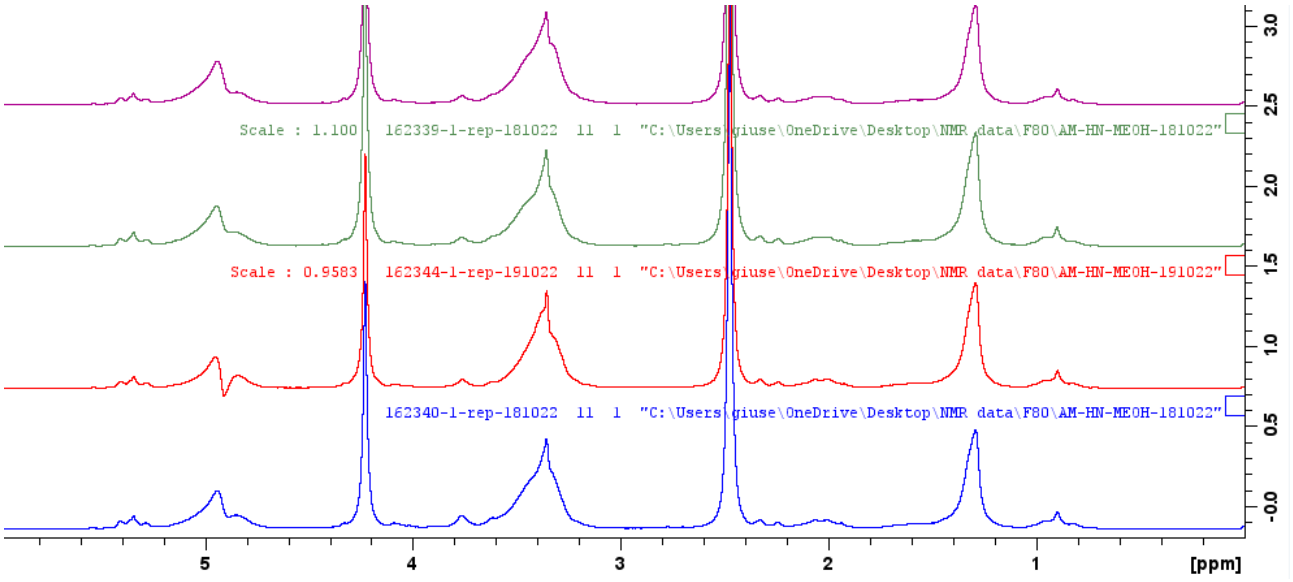


Fig. 6 *Roasted hazelnut sample ^{1}H NMR 80 MHz spectra comparison. (Blue spectrum: Turkey, red spectrum: Italy, green spectrum: Azerbaijan, purple spectrum: Georgia)*

A class compounds interpretation was harsh to carry out, with the non-resolved peaks, as well as it was difficult to search for macroscopic signal differences.

The multivariate statistical analysis followed the same protocols applied for the study of the high-resolution spectra. Figure 7 presents both the LDA plot and confusion matrix as outcomes of the classification model and its validation.

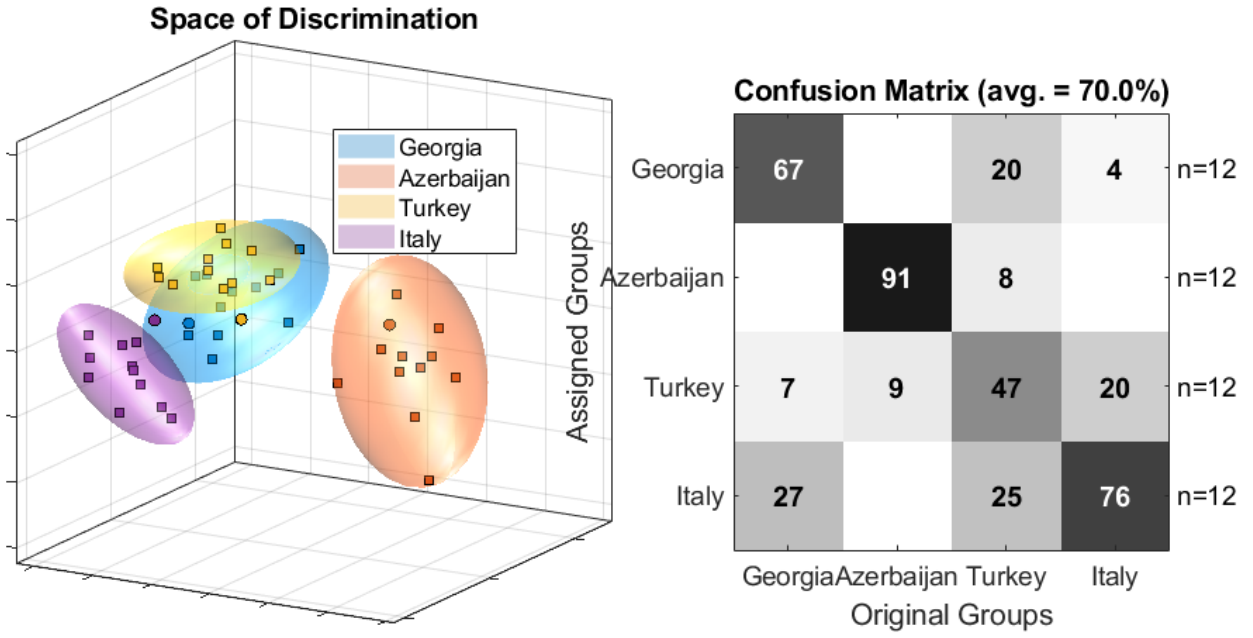


Fig. 7 *LDA plot and confusion matrix of the roasted hazelnut sample set. (Squares: training set samples, circles: test set samples, blue: Georgia, orange: Azerbaijan, yellow: Turkey, purple: Italy)*

The correctness percentage rate lowered to 70 %, and the clustering was discrete. Besides the low resolution giving less information to the model, the limited number of samples (n=48, 12 per class) also influenced the classification and the model performance in general.

The workflow adopted with the Python platform was applied to these measurements as well, figure 8 shows the PLS-DA score plot of the roasted hazelnut sample set analysed at 80 MHz.

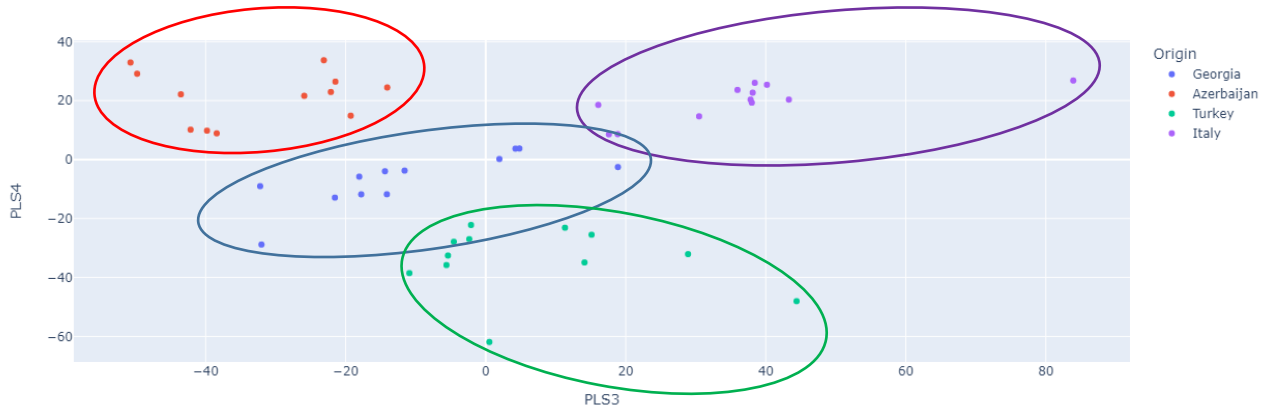


Fig. 8 *PLS-DA plot of the roasted hazelnut sample set. (Blue dots and cluster: Georgia, red dots and cluster: Azerbaijan, green dots and cluster: Turkey, purple dots and cluster: Italy)*

The model reported an acceptable clustering, and its robustness was estimated by the PLS regression on the test set (20 % of the samples from the training set). Table 2 displays the model performance parameters (averaged precision, recall, f1-score, and accuracy percentage score).

ACCURACY	AVG. PRECISION	AVG. RECALL	F1-SCORE
60 %	70.75 %	58.5 %	61 %

Table 2 *Summary of model performance parameters: accuracy, averaged precision, averaged recall, and averaged f1-score.*

Also in this case, both data elaboration approaches returned similar outcomes, putting in evidence a revisable model performance, related to the minor amount of information extractable from a low-res spectrum, and also the restricted number of samples employed for the specific study.

Conclusion & Perspectives

^{1}H NMR spectroscopy played a relevant role in the assessment of the geographical origin of food commodities. The present thesis work emphasized this aspect by applying this technique to the hazelnut chain, considering several sampling factors, such as the industrial process, the harvesting year, and the presence of peel. Further, the application aimed at evaluating outputs from the classic high-resolution technology and the innovative 80 MHz low-resolution one. The former returned very interesting results, with a fine multiclass separation and classification, and valuable model performance parameters. The low-res technique did not show optimal results in terms of model robustness, even though a discrete multiclass separation was anyway accomplished. An intriguing

next step could include a bigger sample set, especially as regard the 80 MHz measurements, as well as a molecular identification, to be done through the 400 MHz instrumentation. In this way, it will be possible to unravel authenticity markers, in order to give the quality control department or control agencies/governments a remarkable tool to specifically search for these molecules, without expensive techniques or expertise needed. To conclude, this work could be a starting point to mind the gap between the academic and industrial environments regarding the NMR spectroscopy approaches. The introduction of a rapid, portable, automatable, and relatively cheap instrumentation, represented by the Fourier 80, could attract food companies in investing in this technology, and to find a connection with the universities to optimise analytical methods and compare results, also from the high-res facilities.

REFERENCES

Agiomyrgianaki, A., Petrakis, P. V., & Dais, P. (2012). Influence of harvest year, cultivar and geographical origin on Greek extra virgin olive oils composition: A study by NMR spectroscopy and biometric analysis. *Food Chemistry*, 2561–2568.

Alberti, E., Belton, P. S., & Gil, A. M. (2002). Applications of NMR to food science. *Annual Reports on NMR Spectroscopy*, 109-148.

Cialiè Rosso, M., Stilo, F., Mascrez, S., Bicchi, C., Purcaro, G., & Cordero, C. (2021). Shelf-Life Evolution of the Fatty Acid Fingerprint in High-Quality Hazelnuts (Corylus avellana L.) Harvested in Different Geographical Regions. *Foods*, 685.

Consonni, R., Cagliani, L. R., & Cogliati, C. (2013). Geographical discrimination of honeys by saccharides analysis. *Food Control*, 543-548.

Esslinger, S., Fauhl-Hassek, C., & Wittkowski, R. (2015). Authentication of Wine by 1H-NMR Spectroscopy: Opportunities and Challenges. In S. B. Ebler, G. Sacks, S. Vidal, & P. Winterhalter, *Advances in Wine Research* (p. 85-108). Washington, DC: American Chemical Society.

Gougeon, L., da Costa, G., Guyon, F., & Richard, T. (2019). 1H NMR metabolomics applied to Bordeaux red wines. *Food Chemistry*, 125257.

Karabagias, I. K., Vlasiou, M., Kontakos, S., Drouza, C., Kontominas, M. G., & Keramidas, A. D. (2018). Geographical discrimination of pine and fir honeys using multivariate analyses of major and minor honey components identified by 1H NMR and HPLC along with physicochemical data. *European Food Research and Technology*, 1249-1259.

Laincer, F., Iaccarino, N., Amato, J., Pagano, B., Tenore, G., Tamendjari, A., . . . Randazzo, A. (2016). Characterization of monovarietal extra virgin olive oils from the province of Béjaïa (Algeria). *Food Research International*, 1123-1133.

Liu, X. (2021). 6.6: ¹H NMR Spectra and Interpretation (Part I). In X. Liu, *ORGANIC CHEMISTRY I* (p. 139-144). Kwantlen Polytechnic University.

Magdas, D. A., Pirnau, A., Feher, I., Guyon, F., & Cozar, B. I. (2019). Alternative approach of applying 1H NMR in conjunction with chemometrics for wine classification. *LWT*, 422-428.

Mascellani, A., Hoca, G., Babisz, M., Krska, P., Kloucek, P., & Havlik, J. (2021). 1H NMR chemometric models for classification of Czech wine type and variety. *Food Chemistry*, 127852.

Ok, S. (2014). Fast screening of Turkish olive oil by NMR spectroscopy for geographical determination and discrimination purposes. *Grasas y Aceites*, e024.

Rongai, D., Sabatini, N., Del Coco, L., Perri, E., Del Re, P., Simone, N., . . . Fanizzi, F. P. (2017). 1H NMR and Multivariate Analysis for Geographic Characterization of Commercial Extra Virgin Olive Oil: A Possible Correlation with Climate Data. *Foods*, 96.

Schievano, E., Stocchero, M., Zuccato, V., Conti, I., & Piana, L. (2019). NMR assessment of European acacia honey origin and composition of EU-blend based on geographical floral markers. *Food Chemistry*, 96-101.

Sobolev, A. P., Ingallina, C., Spano, M., Di Matteo, G., & Mannina, L. (2022). NMR-Based Approaches in the Study of Foods. *Molecules*, 7906.

Spyros, A., & Dais, P. (2015). Introduction. In A. Spyros, & P. Dais, *NMR Spectroscopy in Food Analysis. 1st ed.* Royal Society of Chemistry.

Tas, N. (2017). *INVESTIGATION OF CHEMICAL REACTIONS IN HAZELNUT INDUCED BY ROASTING.* Ankara, Turkey.

Trygg, J., Holmes, E., & Lundstedt, T. (2007). Chemometrics in Metabonomics. *Journal of Proteome Research*, 469-479.

Un, I., & Ok, S. (2018). Analysis of olive oil for authentication and shelf life determination. *Journal of Food Science and Technology*, 2476-2487.

Xu, S., Zhu, J., Zhao, Q., Gao, J., Zhang, H., & Hu, B. (2021). Quality evaluation of Cabernet Sauvignon wines in different vintages by 1H nuclear magnetic resonance-based metabolomics. *Open Chemistry*, 385-399.

AMBIENT MASS SPECTROMETRY FOR FOOD AUTHENTICITY ISSUES

Giuseppe Sammarco[1,3], Nicola Dreolin[2], Chiara Dall'Asta[3], Michele Suman[1,4]

[1] Advanced Laboratory Research, Barilla G. e R. Fratelli S.p.A., Parma, Italy
[2] Waters Corporation, Wilmslow SK9 4AX, United Kingdom
[3] Department of Food and Drug, University of Parma, Parma, Italy
[4] Department for Sustainable Food Process, Catholic University Sacred Heart, Piacenza, Italy

INTRODUCTION

Another technique taken into account in the present thesis work was the atmospheric solid analysis probe-mass spectrometry (ASAP-MS), an ambient mass spectrometry (AMS) technology. ASAP is an ionization source that permits fast analysis with limited or even inexistent sample preparation. The sample is physically introduced into the source through a glass capillary probe, then it is volatilised via thermal desorption, under heated nitrogen gas flow. The analytes ionization occurs by exploiting the corona discharge effect (Tose, Murgu, Vaz, & Romao, 2017) (Tan, et al., 2023) (Xiao, Miller, Parchert, Hayes, & Hochrein, 2016). This effect manifests when a direct or alternating current, generated between two electrodes, led to a high potential and is divided by a neutral fluid, such as air or, in the case of ASAP source, nitrogen, by ionization of this fluid. A plasma is obtained, and the electric charges circulate, going from the ions to the molecules of the surrounding neutral gas (Photonis, 2022). Thus, the inert atmosphere favours the formation of molecular ions through charge transfer from radical nitrogen.

This methodology could potentially find interesting applications for food authenticity issues, and it could also be at the interface between the academy and industry. Indeed, it is a rapid, direct, cost-effective technique, that enables the fingerprinting of the samples analysed. Thus, this facility can be employed by the quality control department of the food companies for rapidly screening lots from the production plant. On the other hand, academic research groups can create a connection with the industries by developing analytical methods, creating databases and data elaboration protocols.

Concerning the applications of the ASAP-MS on food authenticity, Tan *et al.* exploited the technology to authenticate Chinese oolong teas, firstly by discriminating among three varieties. principal component analysis (PCA) and k-nearest neighbour (kNN) gave the best results as statistical models. Secondly, possible adulterations of the Anxi Tieguanyin variety were detected, to preserve the protected geographical indication (PGI). One-class modelling through data-driven soft independent modelling of class analogies (DD-SIMCA) was adopted as a discriminant model, and it yielded remarkable sensitivity and specificity values (Tan, Chan, Lee, Xu, & Zhou, 2022). Two

different AMS techniques, direct analysis in real time (DART) and ASAP-MS were used for the authenticity study of oregano (*Origanum vulgare* L.). The data analysis was performed on the entire mass spectra, after proper pre-processing. SIMCA classification algorithm was exploited, then the model consistency was proven by adding new authentic oregano samples and in-house blended samples with other herbs. ASAP-MS reported optimal outcomes, and both training and validation sets were integrated with other samples and new adulterants. The predictive accuracy gained was 93 %, with a detection capability in the 5-20 % range, except for marjoram-adulterated samples (Damiani, Dreolin, Stead, & Dall'Asta, 2021).

This thesis study presents a concrete application of ASAP-MS methodology for the geographical origin evaluation of hazelnut and hazelnut-derived, and dehydrated apple samples. The analytical strategy was coupled with multivariate statistical models for data handling and elaboration.

CASE STUDY: ASAP-MS TECHNIQUE FOR AN AUTHENTICITY STUDY OF HAZELNUT AND APPLE COMMODITIES

Chemicals

Dichloromethane, methanol, and formic acid solvents were all purchased from VWR International Ltd (Poole, England, UK). Water was bi-distilled via the Milli-Q system (Millipore, Bedford, Massachusetts, US).

Instrumental parameters

ASAP ion source (Waters Corp., Wilmslow, England, UK) was hyphenated with the Waters QDa analyser (Waters Corp., Wilmslow, England, UK), a single-quadrupole MS detector. Sealed glass capillaries (1.9 mm diameter) were used as probes, their tips were dipped into the solution to analyse for roughly 10 seconds, then the capillary was immediately inserted into the source. Before this step, the capillary was previously put under heated N_2 flow for 90 seconds into the ASAP source, to remove potential impurities from its surface. This bake-out phase favours a relevant reduction of the spectral background. The parameters set for the analyses were positive ionization mode, gas flow 3.0 L/min, gas temperature 500°C °C, constant corona current 3 µA, and cone voltage 15 V. Two technical replicates were analysed per sample, within the range 100-1000 *m/z*, with a sampling frequency of 1 Hz.

Hazelnuts

Sampling

Initially, a limited number of samples was selected to perform a feasibility study, in order to develop and optimise the extraction protocol and the analytical method. Subsequently, a design of experiment (DoE) was created to achieve a robust sampling. Fresh, roasted, and paste samples were sampled from different countries, considering Italy as the main geographical target. The factors contemplated for the DoE included the harvesting year, different storage shelf life, the presence of peel, and the industrial process. The Italian samples provided by the supplier were PGI "Tonda Gentile delle Langhe" from Piedmont, PDO "Nocciola Romana" from Latium, and "Mortarella" from Campania. For all the hazelnut-type samples, these three varieties were mixed to have 'Italian samples' (N=36, 9 raw, 9 roasted, 9 peeled roasted, and 9 paste of hazelnuts) The same number of lots were purchased from Turkey, Azerbaijan, and Georgia, for a total of 108 non-Italian samples. All of them were kept refrigerated in a cold room, with a controlled temperature of 4-6 °C.

Sample preparation

Ca. 10 g of fresh and roasted hazelnuts were milled at 7000 rpm for 15 seconds, with the knife mill GrindoMix GM200 (Retsch, Hann, Germany). 0.5 g per sample (milled fresh and roasted, and paste) were weighed in falcon tubes, then 5 mL of dichloromethane and 5 mL of methanol were added. The mixtures were shaken for 10 minutes at 1300 rpm, with the rotatory shaker Multi Bio RS-24 (BioSan, Riga, Latvia). Fast centrifugation was then carried out at 4000 rpm for 1 minute, using the Rotina 380R centrifuge (Hettic Lab Technology, Tuttlingen, Germany). The supernatant was withdrawn and 1:10 diluted, 40 µL was taken and 360 µL of methanol was added in 1 mL amber glass vials. These were stored at -20°C before the analyses.

Multivariate statistical analysis

Data elaboration was accomplished by means of LiveID software (Version 2.0, Waters Corp, Wilmslow, England, UK). Raw data were collected into the spectral library of the software, and the threshold intensity of the total ion current (TIC) peak was set to 70 %, to both exclude background signals and gain good repeatability between replicates measurements. In the spectral library, the samples were grouped into the classes of interest, this is important for the validation test to perform after the statistical models building. Regarding the hazelnuts, spectra were assigned to the groups, and the model parameters included multi-class PCA coupled with linear discriminant analysis (LDA) for classification as model type, outlier detection by the standard deviation (threshold=5), binning

resolution 1.00, a mass range considered from 200 to 900 *m/z*. Leave-group-out validation method was applied to assess the consistency of the model.

Results & Discussion

ASAP-MS technology is an innovative tool in the analytical chemistry field, hence there is not much knowledge about extraction protocols and methods of analysis. Thus, different potential solutions were exploited, always trying to preserve the rapidity of the sample prep/analysis phases, which represents the main advantage of the technique. Regarding the hazelnut samples, several solutions were tried since they present both polar and non-polar components, so the aim was to extract as many compounds as possible, to extrapolate the highest amount of information from them. Methanol, methanol-dichloromethane (50:50 % w/w), and double-phase water-dichloromethane (50:50 % w/w) solutions were screened during the feasibility study, and the second one was revealed to be the most suitable and rapid one. Despite the toxicity of a halogen compound such as dichloromethane, this solvent, combined with methanol, was found the be very effective in lipids extraction, which represents a relevant component of the samples of interest (Cequier-Sanchez, Rodriguez, Ravelo, & Zarate, 2008). Once the extractions were done, the dilution step represented an important point: since this technology does not operate with a chromatographic separation before the mass spectrometric analysis when a highly concentrated extract is analysed, it is not consequential to have more intense signals. Therefore, the sample dilution becomes a remarkable factor to reach the proper concentration to have a good signals distribution, without ion suppressions, charge-competition phenomena, or deep contaminations of the ASAP source (Damiani, Dreolin, Stead, & Dall'Asta, 2021) (Kumbhani, Wingen, Perraud, & Finlayson-Pitts, 2017) . Spectra from hazelnut samples analysis required a 1:10 dilution to improve their quality, contextually keeping good signal intensities. Also, about the sample introduction on the probe, there were two possible ways to do it, by dipping the capillary into the solution or pipetting the extract on its tip. These two methods were approached, and the outcomes were analogue, so the dipping mode was employed because it requires no consumables (pipet tips) or pipetting skills. Capillary dipping time was standardised to 10 seconds per measurement, and this was more user-friendly and reproducible. To conclude as regards the analytical method optimisation, different trials were done for the heated gas temperature, in order to obtain an effective desorption of the solvent from the probe. The most suitable temperatures were 500 °C for hazelnut and apple extracts desorption. Both positive and negative ionization mode experiments were carried out for all the samples. Hazelnut samples spectra obtained from the negative ionization mode did not show additive information and the signals were less intense than the ones from the positive mode, in the

feasibility study. Therefore, only the positive ionization mode experiments were performed with this matrix.

Figure 1 shows the TIC peaks related to the two technical replicates measured during the same analysis, of a paste hazelnut sample. For each measurement, a single peak is generated, which starts with the probe introduction into the source, and gradually ends together with the desorption of the sample from the capillary.

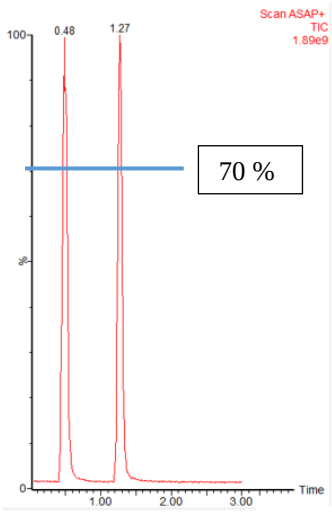


Fig. 1 *Classical TIC peaks obtained from ASAP-MS analysis of paste hazelnut sample. Blue lines indicate the threshold above whom the average spectra were taken.*

The mass spectra change throughout the TIC peak, as in the first part the lower-mass molecules are immediately desorbed, so their signal intensities decrease, while heavier molecules take more time to get completely desorbed, so their signal intensities gradually grow up. This is the reason for setting the relative intensity threshold for the peaks at 70 %. In this way, the data considered are more consistent and this favours good reproducibility. AMS techniques could present a relevant background, due to the eventual probe contaminations, or the opening of the source for the capillary introduction into it. Figure 2 presents the comparison between mass spectra related to the TIC peak (intensity > 70 %) and the one related to the background, of a paste hazelnut sample.

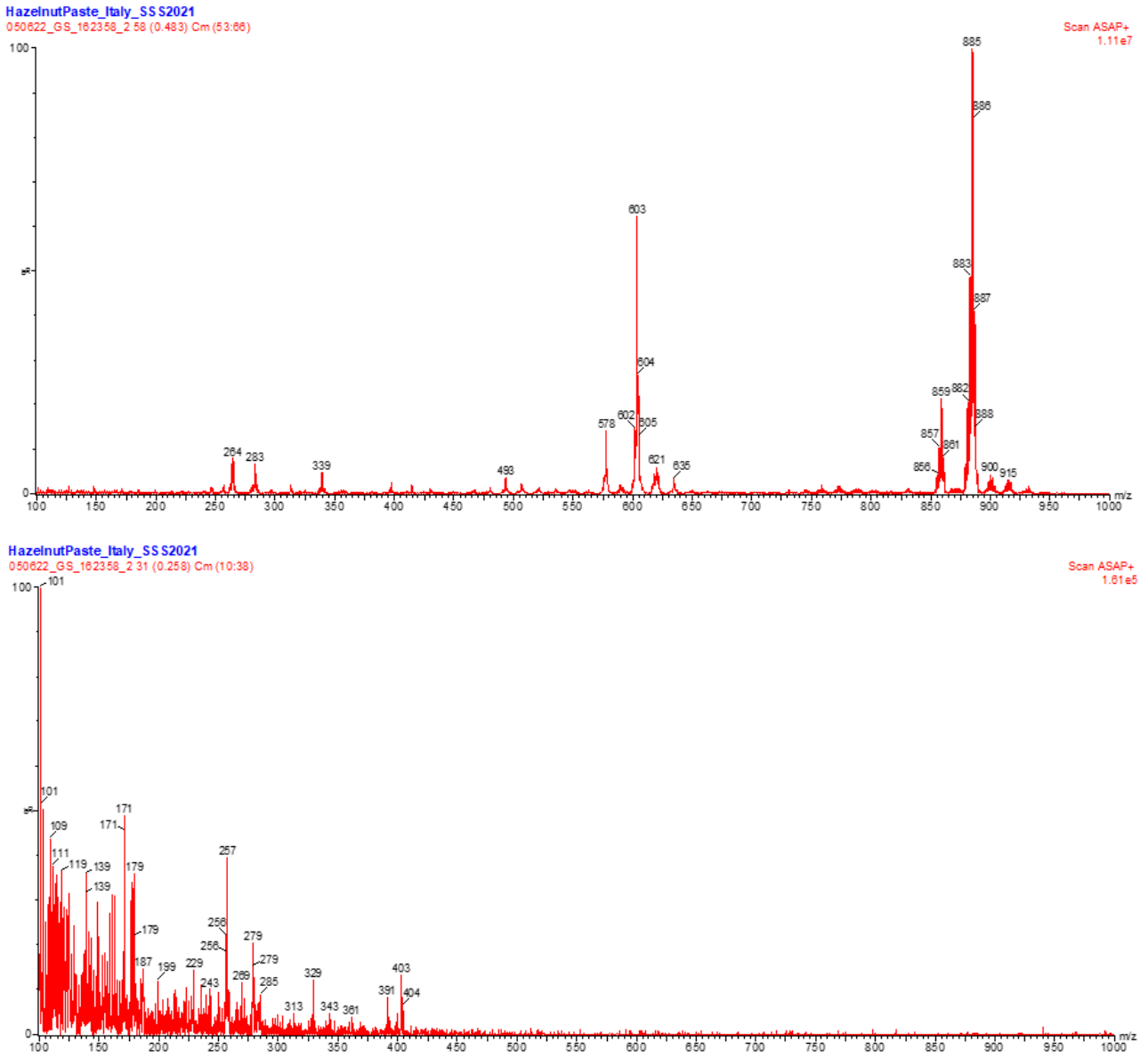


Fig. 2 *ASAP-MS raw spectrum from paste hazelnut sample acquisition. Below, the background's recorded spectrum of the same sample, taken before the TIC peak.*

It is evident that the main features derived from the sample measurement were not present also in the background mass spectra; further, the signal intensities were way lower than in one of the sample analyses. The enclosed structure of the ion source and the bake-out step of the capillary contributed to this "clean" background.

To extract useful information from these spectral fingerprints, and to give statistical significance to the possible differences between Italian and non-Italian products, a multivariate statistical model was built for both the matrices analysed. PCA-LDA model was employed to reduce the data dimensionality and classify the samples. PCA was also exploited to select the proper mass range features to include in the model, by considering the most important ones through the loadings plot. Figure 3 displays the tri-dimensional PCA-LDA score plot of the entire hazelnut sample set.

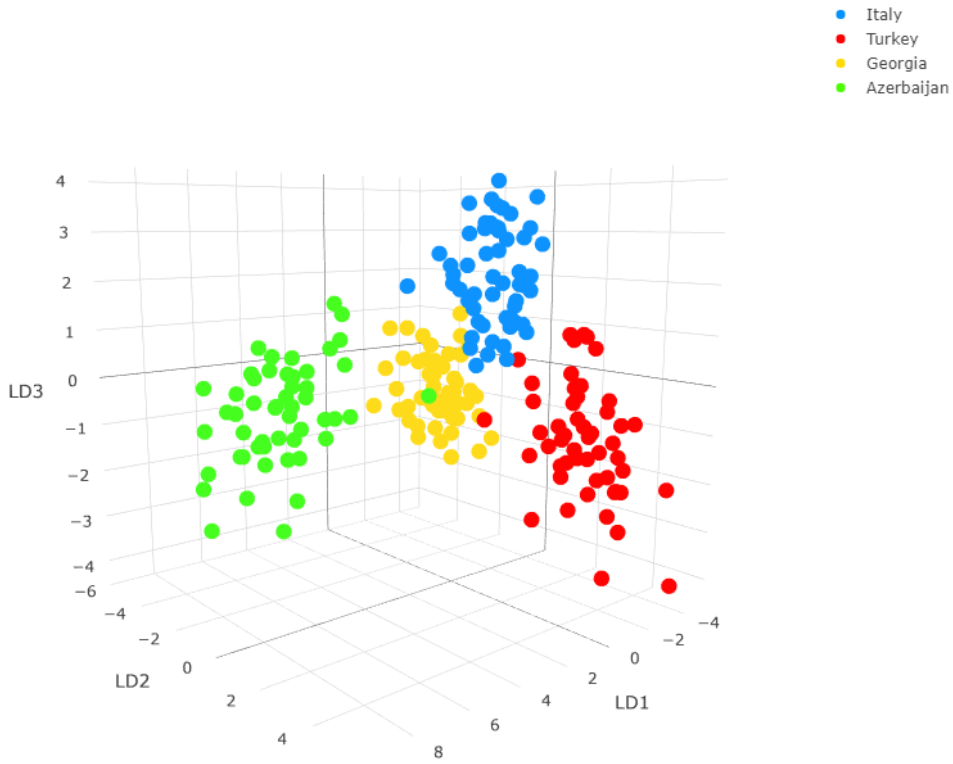


Fig. 3 *Tri-dimensional score plot of PCA-LDA (LD1 vs LD2 vs LD3) of the hazelnut sample set.*

The sample clusterization is evident, for the hazelnuts a multi-class classification was achieved, with a fine separation according to the different countries considered. To monitor the model robustness, some cross-validation methods were approached for both matrices. The basis of these methods is to take out from the training set part of the data, build the model, and make a prediction of the samples left out (Bro, Kjeldahl, Smilde, & Kiers, 2008). Leave-one-out cross-validation (LOOCV) is exploited to evaluate the machine learning algorithm performance in case of predictions on data extrapolated from the training set. It consists in generating several datasets, each of them by leaving out one sample, that will be then used as a test set. It is quite tough to carry out, but it provides an unbiased assessment of the model's performance. Moreover, it is easy to use, even though it is not suitable for large datasets or computationally complex models (Brownlee, 2020). In these cases, leave-group-out cross-validation (LGOCV) is a good solution to manage several prediction tasks. It consists in taking out of the sample set a group of data manually defined. Thus, it is less complicated to carry out the method (Liu & Rue, 2022). In the present study, the LGOCV cross-validation approach was used, Table 1 displays the results of it the hazelnut model, represented as a confusion matrix.

	Georgia	Azerbaijan	Turkey	Italy	Outlier	Total
Georgia	75 %		8 %	13 %	4 %	
Azerbaijan		87 %	3 %		10 %	
Turkey		2 %	80 %	12 %	6 %	
Italy	5 %		12 %	81 %	2 %	
Total						80.75 %

Table 1 *Confusion matrix from hazelnut statistical model's evaluation through LGOCV method. Total values report the model's correctness score.*

As it is possible to see, the correctness scores were not optimal, even if reasonable especially considering them from a high throughput screening purpose perspective. This could be due to several factors, such as the data dispersion, generated by many factors considered for the DoE, and the spectral differences throughout the sample measurement, caused by the diverse 'desorption rate' of the analytes on the probe. Another limitation is the data elaboration software, programmed to be user-friendly, rapid, and automatable, but not as flexible as *ad hoc* statistical elaboration software, or programming language scripts. However, these results could represent a relevant starting point for further confirmation of the applicability of the ASAP-MS methodology in the food authenticity field.

Dehydrated Apples

Sampling

The same sampling approach was employed for the dehydrated apple commodity, so an initial feasibility study was performed, using a limited number of samples, then a DoE was generated to achieve a consistent sampling. Diverse factors were taken into account, such as dehydration rate, presence of peel, harvesting campaign, and variety. The countries selected were France, China, Chile, Hungary, Poland, and Italy as the main geographical target, for a total of 91 samples.

Sample preparation

As for the hazelnut samples, ca. 10 g was initially milled. 0.5 g were weighed, and 10 mL of water with 70 % methanol and 0.1 % formic acid was added. Agitation and centrifugation steps were the same performed for the hazelnuts, while the dilution was avoided, so 400 µL of supernatants were directly pipetted into the 1 mL amber glass vials. These were equally stored at -20°C before the analyses.

Multivariate statistical analysis

The dehydrated apple samples statistical model was similar to the hazelnut samples, the only differences were about the model type, it was a bi-class PCA-LDA, and the mass range considered, from 100 to 600 *m/z*. As validation approaches, leave-20%-out was found to return discrete outcomes.

Results & Discussion

For the dehydrated apple samples, the extraction solution was "exported" from the protocol already employed for the liquid chromatography-mass spectrometry methodology (See "LC-HRMS Analysis" chapter). Interesting results from the feasibility study were found using water with 70 % methanol with 0.1 % formic acid, so it was selected for the extraction of the entire sample set. The preliminary study of the spectra during the feasibility study did not show issues related to the high concentration of the samples throughout the analytical sequence, hence no dilution was applied. Also for dehydrated apples, both positive and negative ionization mode acquisitions were performed. The spectra presented interesting features to study in both types of experiments. However, from the multivariate statistical evaluation of the negative mode data, no useful results were reached. Thus, only the outcomes from the positive mode experiment will be shown. Figure 4 reports the TIC peaks derived from the ASAP-MS acquisition, with the same intensity threshold of 70 % set for the hazelnut samples.

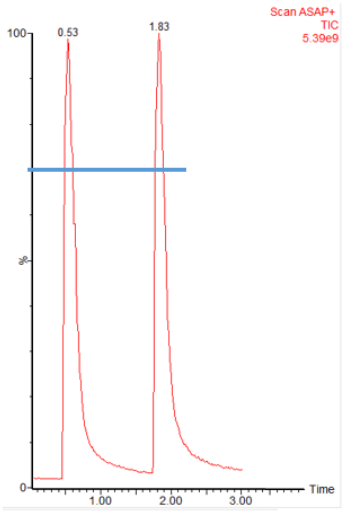


Fig. 4 *Classical TIC peaks obtained from ASAP-MS analysis of dehydrated apple sample. Blue lines indicate the threshold above whom the average spectra are taken.*

The comparison between the spectrum derived from the sample peak (> 70 %) and the one recorded from the background was carried out. Figure 5 displays them.

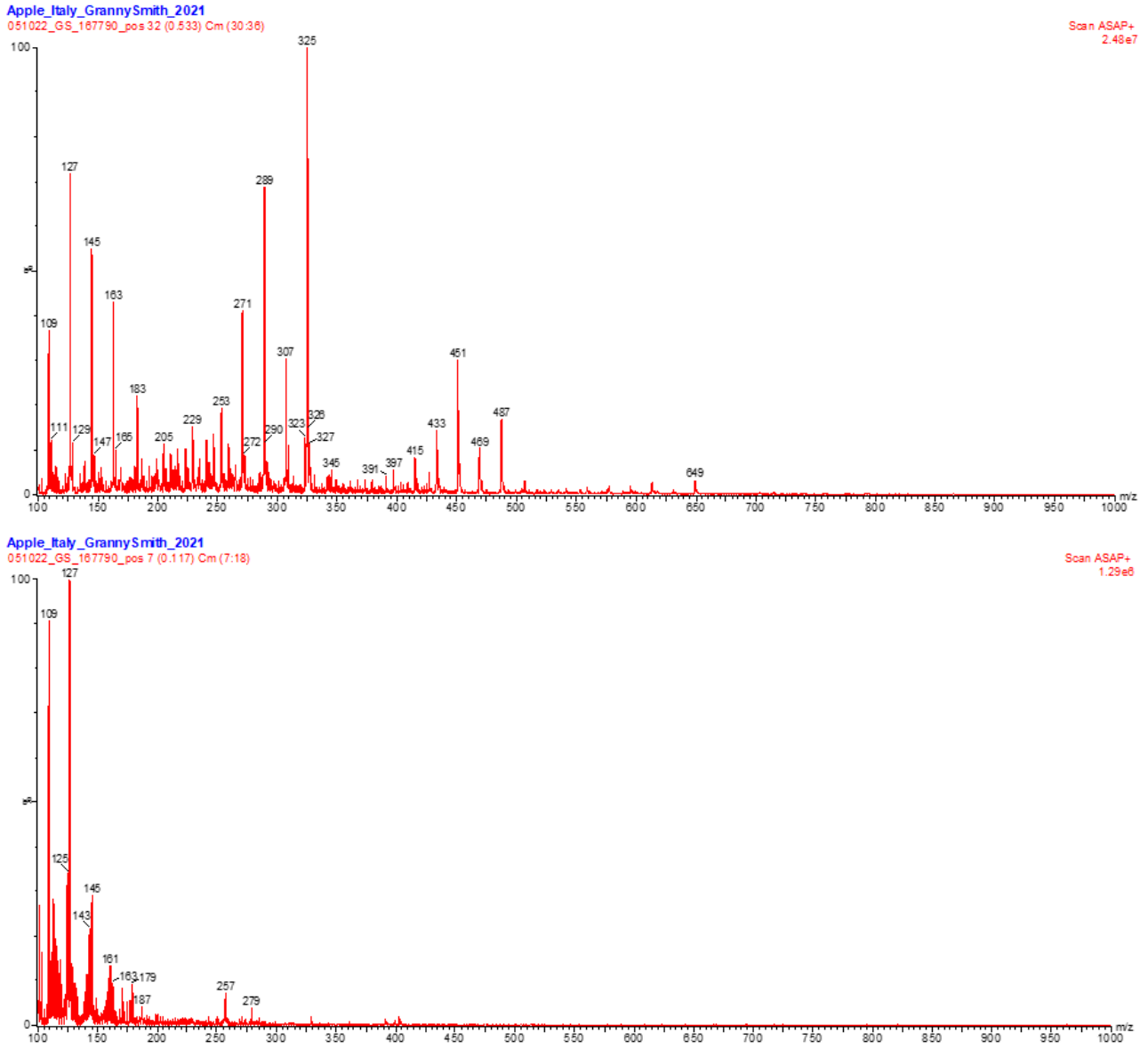


Fig. 5 *ASAP-MS raw spectrum from dehydrated apple sample acquisition. Below, the background's recorded spectrum of the same sample, taken before the TIC peak.*

These spectra clearly demonstrate that the main features of the TIC peak spectrum were not present in the background spectrum as well. Further, the base peak of the background spectrum is 20 times less intense than the one from the TIC.

Useful chemical information and data visualisation were achieved by performing the PCA-LDA model, as done for the hazelnut sample set. The spectrum interpretation and the loadings plot from the PCA led to the application of the model on the mass range of 100-600 *m/z*. In addition, the high data dispersion, due to the numerous DoE factors related to the number of samples, favoured a bi-class model, instead of a multi-class. Thus, samples from China and Chile were grouped as Extra European Union (EU), samples from France, Hungary, Poland were labelled as EU, and Italy was considered the main target. Figure 6 represents the tri-dimensional PCA-LDA based on the entire dehydrated apple sample set.

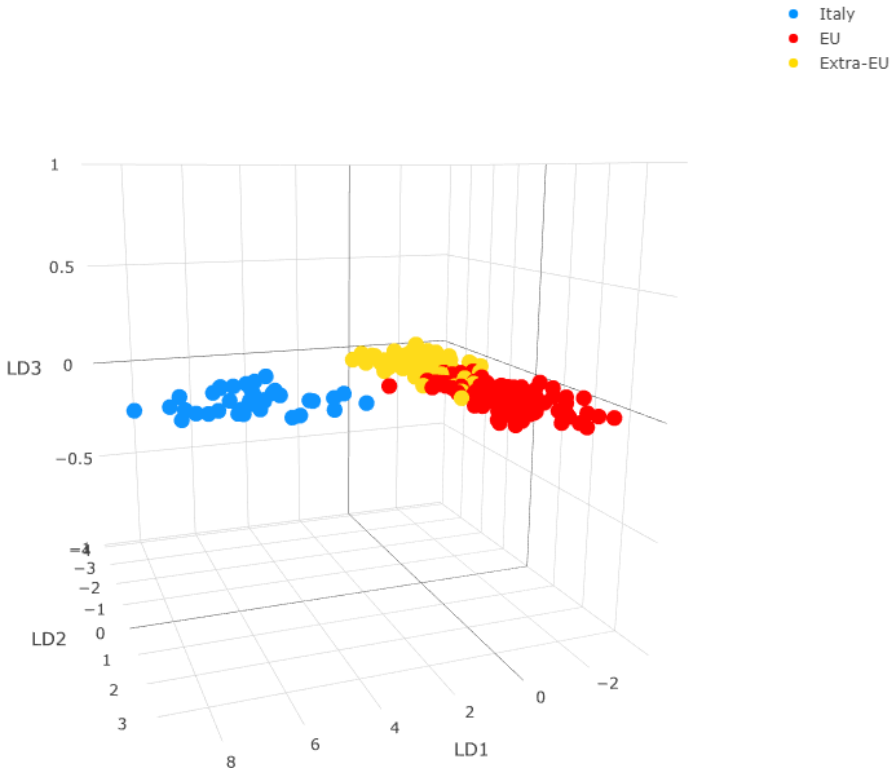


Fig. 3 *Tri-dimensional score plot of PCA-LDA (LD1 vs LD2 vs LD3) of the dehydrated apple sample set.*

An interesting class separation was reached, with the Italian samples evidently divided by the rest of them through the 1st dimension of the score plot. To finally estimate the model performance, leave-20%-out cross-validation, a variant of the LGOCV method, was carried out, with discrete outcomes, reported in a confusion matrix (Table 2).

	EU	Non-EU	Italy	Outlier	Total
EU	77 %	23 %			
Non-EU	22 %	78 %			
Italy	5 %	12 %	78 %	5 %	
Total					77.67 %

Table 2 *Confusion matrix from dehydrated apple statistical model's evaluation through LGOCV method. Total values report the model's correctness score.*

The results were analogue to the hazelnut statistical model, always taking into account the limitation of the study, technique, and data elaboration software. However, these outputs should not discourage the researchers from applying this analytical strategy, but they could serve as a promising achievement to start with.

Conclusions & Perspective

ASAP-MS represents an innovative AMS technique, able to overcome some critical issues of analogue technologies, such as the open ion source or the noisy background. In spite of that, it ensures rapidity in obtaining a spectral fingerprint of the sample analysed, which requires limited or even inexistent preparation. Furthermore, the cost is much lower compared to that of a classic mass spectrometer, and it did not necessitate any chromatographic separation. Coupling ASAP-MS with the LiveID software permits quick data interpretation and visualization with limited experience. The present thesis work, about the geographical origin assessment of hazelnut and apple products, highlighted the pros and cons of the adopted methodology. The groups' clustering, according to the areas of provenience, and the cross-validation methods showed the feasibility of the technique applied to this type of issue, even though the final results cannot be considered as robust enough for a highly reliable classification/authentication of the matrices of interest. A finer data handling and elaboration are required to develop and automatize the proper statistical analysis method. Once this will be accomplished, an external sample set is required to challenge and validate the model. In addition, the analysis of blended samples (different percentages of Italian and non-Italian products), should be performed to estimate the sensitivity of the technique. This conclusive approach could confirm the feasibility of the methodology to rapidly prevent and/or fight food authenticity issues.

REFERENCES

Bro, R., Kjeldahl, K., Smilde, A. K., & Kiers, H. A. (2008). Cross-validation of component models: A critical look at current methods. *Analytical and Bioanalytical Chemistry*, 1241-1251.

Brownlee, J. (2020, July 27). *LOOCV for Evaluating Machine Learning Algorithms.* Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/loocv-for-evaluating-machine-learning-algorithms/

Cequier-Sanchez, E., Rodriguez, C., Ravelo, A. G., & Zarate, R. (2008). Dichloromethane as a Solvent for Lipid Extraction and Assessment of Lipid Classes and Fatty Acids from Samples of Different Natures. *Journal of Agricultural and Food Chemistry*, 4297-4303.

Damiani, T., Dreolin, N., Stead, S., & Dall'Asta, C. (2021). Critical evaluation of ambient mass spectrometry coupled with chemometrics for the early detection of adulteration scenarios in Origanum vulgare L. *Talanta*, 122116.

Kumbhani, S. R., Wingen, L. M., Perraud, V., & Finlayson-Pitts, B. J. (2017). A cautionary note on the effects of laboratory air contaminants on ambient ionization mass spectrometry measurements. *Rapid Communications in Mass Spectrometry*, 1659-1668.

Liu, z., & Rue, H. (2022). Leave-group-out cross-validation for latent Gaussian models. *ArXiv*.

Photonis. (2022, August 5). *What's corona discharge?* Retrieved from Photonis: https://www.photonis.com/news/whats-corona-discharge#:~:text=What's%20corona%20discharge%3F,-Friday%2C%2005.08.2022&text=The%20corona%20effect%20is%20a,formation%20of%20an%20electric%20arc.

Tan, H. R., Can, L. Y., Ong, A., Xu, Y., Zhang, X., & Zhou, W. (2023). Atmospheric solids analysis probe-mass spectrometry (ASAP-MS) as a rapid fingerprinting technique to differentiate the harvest seasons of Tieguanyin oolong teas. *Food Chemistry*, 135135.

Tan, H. R., Chan, L. Y., Lee, H. H., Xu, Y., & Zhou, W. (2022). Rapid authentication of Chinese oolong teas using atmospheric solids analysis probe-mass spectrometry (ASAP-MS) combined with supervised pattern recognition models. *Food Control*, 108736.

Tose, L. V., Murgu, M., Vaz, B. G., & Romao, W. (2017). Application of Atmospheric Solids Analysis Probe Mass Spectrometry (ASAP-MS) in Petroleomics: Analysis of Condensed Aromatics Standards, Crude Oil, and Paraffinic Fraction. *Journal of The American Society for Mass Spectrometry*, 2401-2407.

Xiao, X., Miller, L. L., Parchert, K. J., Hayes, D., & Hochrein, J. M. (2016). *Atmospheric Solids Analysis Probe Mass Spectrometry (ASAP-MS) for the Rapid Identification of Pollens through Flavonoid Fingerprints.* Albuquerque, New Mexico, US: Sandia National Laboratories.

Conclusions and future perspectives

In the present thesis work, authenticity issues of hazelnut and apple samples have been studied. Particularly, several analytical solutions, both rapid and confirmatory, were applied to the two food commodities, to evaluate the geographical origin of Italian products. This was achieved to avoid a false declaration of origin fraud from the suppliers, but also for protecting the food excellence of the Italian territory, guaranteeing high-quality Barilla products.

Fresh, roasted and paste hazelnut samples were analysed through diverse solutions, each of them returned interesting results. NIR spectroscopy and GC-IMS spectrometry were found to be two optimal rapid, direct, cost-effective methods to assess the origin of Italian hazelnuts. NIR spectroscopy is the most suitable for on-line control of the materials used for industrial products, also with a relevant sensitivity and repeatability. However, it cannot allow the user to get information for molecular identification, or to study volatile organic compounds (VOCs). GC-IMS system permits to obtain a fingerprint of the volatile profile of the sample. In addition, a sensomics approach was performed, by merging GC-IMS data with the ones gained from a sensory analysis, with intriguing outputs. ASAP-MS was another fast methodology preliminary employed. It returned mass spectra in a few minutes, and from the spectral information, it was possible to estimate the geographical origin. Anyway, the low resolution does not consent an actual molecular identification, and the rapidity of the analysis, together with the absence of a separation technique before, led to not ideal outcomes from the validation tests. As a confirmatory approach, a preliminary study with NMR spectroscopy was carried out, considering both high- and low-resolution experiments. Different statistical elaboration workflows were used, with promising results from the 400 MHz analyses, whereas the 80 MHz instrument, which could be suitable for the industrial environment, did not bring exalting outputs. IRMS technology was another remarkable approach for food authentication and permitted the authors to discriminate between Italian and non-Italian samples without the influence of all the factors taken into account for the DoE. It was also found a logical relation between climatic/soil conditions and the stable isotopic ratios measured.

As regards dehydrated apple samples, GC-IMS and ASAP-MS were exploited. The former one returned impressive results, the volatile profile information was handled and elaborated with a multivariate statistical protocol able to discriminate all the samples according to their origin. ASAP-MS, analogously to the hazelnut samples analyses, reported promising results, even if further studies are needed, about both analytical method optimisation and chemometric elaboration. LC-HRMS was

the robust solution to confirm the geographical origin assessment. Seven molecules were putatively identified as provenience markers, and the statistical models built were able to discriminate and classify Italian and non-Italian samples with high correctness scores. Table 1 summarizes the advantages, the drawbacks, and industrial applications of all the analytical techniques considered.

ANALYTICAL TECHNIQUE	ADVANTAGES	DRAWBACKS	INDUSTRIAL APPLICATION
NIR Spectroscopy	Easy-to-use, rapid, non-destructive, cost-effective, green, reproducible, portable	Low sensitivity, indirect, low penetration depth, no molecular identification	Fast on-line/in-field sample control
GC-IMS	Easy-to-use, direct, rapid, cost-effective, high sensitivity, green	Destructive, only for volatile compounds, contamination by atmospheric vapours	Fast detection of anomalous samples (off-line)
ASAP-MS	Easy-to-use, rapid, direct, limited or no sample prep, cost-effective, portable (potentially)	Destructive, low sensitivity, low resolution	Fast control of anomalous sample (off-line), potential on-line and in-field analyses
IRMS	Rapid, high accuracy, green, reproducible	Destructive, expensive, (relative) low sensitivity, no molecular identification	Confirmatory analysis for food authentication
LC-HRMS	High sensitivity, high accuracy, high resolution, molecular identification	Destructive, expensive, time consuming, complex sample prep and data elaboration	Confirmatory analysis for food frauds detection, trace analysis (contaminants, pesticides, mycotoxins…)

NMR Spectroscopy	High accuracy, high resolution, (possible) limited sample prep	Destructive, expensive, time consuming, not applicable to every nuclei, (relative) low sensitivity	Confirmatory analysis for food frauds detection

Table 1. *Advantages, drawbacks, industrial applications of each technique. (NIR: near infrared; GC-IMS: gas chromatography-ion mobility spectrometry; ASAP-MS: atmospheric solid analysis prob-mass spectrometry; IRMS: isotope ratio mass spectrometry; LC-HRMS: liquid chromatography-high-resolution mass spectrometry; NMR: nuclear magnetic resonance)*

Many of the techniques illustrated here were used at the industrial level, in the Barilla research group laboratories. The open collaboration with the University of Parma led to a fruitful project, with the development of rapid analytical strategies that could be implanted as tools for routinary analyses, as well as the creation and optimization of confirmatory methods. Despite these last methods are not applicable for routinary controls, because of the sophisticated technology that is not suitable for the on-field analyses, they find valuable importance for deeply evaluating anomalous samples preliminary detected with the rapid tools. Furthermore, these instrumentations could be useful for fine research about the quality of the ingredients and their correlated products, together with the sensory analyses and the sensomics approach in general.

Author

Giuseppe Sammarco

Born in Salerno (SA), on July 9th, 1994

e-mail: giuseppe.sammarco@barilla.com / giuseppe.sammarco1@unipr.it

Studies

Giuseppe obtained a master's degree in Pharmaceutical Chemistry and Technologies in 2018, with maximum grades (110/110) at the "Università degli Studi di Salerno".

He had an experimental thesis project at the Bio-organic chemistry lab, at the Department of Pharmacy of the "Università degli Studi di Salerno", based on functional chemical proteomics.

In 2019, he had an Erasmus traineeship scholarship for a project at the University of Utrecht (The Netherlands), where he joined the Biomolecular Mass Spectrometry and Proteomics group.

In November 2019, he started his Ph.D. project in Food Science at the "Università degli Studi di Parma", under the supervision of Prof. Michele Suman and Prof. Chiara Dall'Asta. The goal of the project was to develop analytical strategies able to discriminate between Italian and non-Italian food products, contextually employing multivariate statistical analysis, data fusion, and sensomics approaches.

During the Ph.D. program, he was visiting student at the Waters Corporation (Wilmslow, England, UK), the University of Applied Science of Mannheim (Mannheim, Germany), and the Bruker Biospin (Ettlingen, Germany).

Scientific Activities

Papers published

- **Practical approach to develop a multi-group screening method for detection of mycotoxins, pesticides and veterinary drugs in food**
  Simone Moretti, Daniele Cavanna, Francesca Lambertini, Dante Catellani, **Giuseppe Sammarco**, Carolina Barola, Fabiola Paoletti, Giorgio Saluti, Roberta Galarini, Michele Suman
  J mass spectrom. 2020; 55:e4618. https://doi.org/10.1002/jms.4618

- **Fighting food frauds exploiting chromatography-mass spectrometry technologies: scenario comparison between solutions in scientific literature and real approaches in place in industrial facilities**
  Michele Suman, Daniele Cavanna, **Giuseppe Sammarco**, Francesca Lambertini, Cecilia Loffi
  TrAC Trends in Analytical Chemistry, vol. 142, 2021, 116305, ISSN 0165-9936, https://doi.org/10.1016/j.trac.2021.116305

- **Non-targeted high-resolution mass spectrometry study for evaluation of milk freshness**
  Cecilia Loffi, Daniele Cavanna, **Giuseppe Sammarco**, Dante Catellani, Chiara Dall'asta, Michele Suman
  Journal Of Dairy Science, vol. 104, issue 12, 2021, 12286-12294, ISSN 0022-0302, https://doi.org/10.3168/jds.2021-20285

- **Non-targeted authentication of black pepper using a local web platform: development, validation and post-analytical challenges of a combined nir spectroscopy and lasso method**
  Andrea Massaro, Marco Bragolusi, Alessandra Tata, Carmela Zacometti, Stephane Lefevre, Aline Frégière-Salomon, Jean-Louis Lafeuille, **Giuseppe Sammarco**, Ingrid Fiordaliso Candalino, Michele Suman, Roberto Piro
  Food Control, vol. 145, 2022, 109477, ISSN 0956 7135, https://doi.org/10.1016/j.foodcont.2022.109477

Papers submitted

- G. Sammarco, C. Dall'Asta, M. Suman - **Near Infrared Spectroscopy and Multivariate Statistical Analysis as Rapid Tools for the Geographical Origin Assessment of Italian Hazelnuts Chain** – Vibrational Spectroscopy (submitted)

- G. Sammarco, D. Bardin, F. Quaini, C. Dall'Asta, J. Christmann, P. Weller, M. Suman - **A Geographical Origin assessment of Italian Hazelnuts: Gas Chromatography-Ion mobility spectrometry coupled with Multivariate Statistical Analysis and Data Fusion approach** – Food Research International (submitted)

- G. Sammarco, C. Dall'Asta, M. Suman - **Untargeted Gas Chromatography-Ion Mobility Spectrometry approach for the geographical origin evaluation of dehydrated apples** – Analytica Chimica Acta (submitted)

- G. Sammarco, M. Rossi, M. Suman, D. Cavanna, L. Viotto, P. Pettenà, C. Dall'Asta, P. Iacumin - **Hazelnut products traceability through combined Isotope Ratio Mass Spectrometry and Multi-elemental Analysis** – Journal of the Science of Food and Agriculture (submitted)

- G. Sammarco, C. Dall'Asta, M. Suman - **Untargeted Metabolomics Liquid Chromatography-High Resolution Mass Spectrometry approach for the Geographical Origin Assessment of Italian Dehydrated Apples** – Journal of Mass Spectrometry (submitted)

Oral Communications

- Virtual European Geosciences Union General Assembly (19-30 April 2021)
  Hazelnut products traceability through Isotope Ratio Mass Spectrometry approach
  Giuseppe Sammarco, Mattia Rossi, Michele Suman, Daniele Cavanna, Chiara Dall'Asta, Paola Iacumin
  DOI: 10.5194/egusphere-egu21-4554

- Workshop PhD Food 2021 (Palermo, Italy, 14-15 September 2021)
  Innovative analytical strategies coupled with Multivariate Analysis to deal with complex issues in Food Integrity, Food Authenticity and Sensomics
  Giuseppe Sammarco, Michele Suman, Chiara Dall'Asta

- FRUTIC 2022: 14th International Symposium (Valencia, Spain, 29 June-1 July 2022)
  Dehydrated Apples Geographical Assessment through Gas Chromatography-Ion Mobility System Analysis
  Giuseppe Sammarco, Michele Suman, Chiara Dall'Asta

- Recent Advances in Food Analysis (RAFA) 2022 (Prague, Czech Republic, 6-9 September 2022)
  Hazelnut products traceability through Isotope Ratio Mass Spectrometry approach
  Giuseppe Sammarco, Mattia Rossi, Michele Suman, Daniele Cavanna, Chiara Dall'Asta, Paola Iacumin