UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN
SCIENZE DEGLI ALIMENTI

CICLO XXXIII

Preventing Fraud and Assessing Safety in the Fishery and Aquaculture Sectors through Rapid Analytical Strategies

Coordinatore:
Chiar.ma Prof. **Chiara Dall'Asta**

Tutore:
Chiar.mo Prof. **Sergio Ghidini**

Chiar.ma Prof. **Emanuela Zanardi**

Dottorando: **Maria Olga Varrà**

Anni Accademici 2017/2018 – 2019/2020

MARIA OLGA VARRÀ

Preventing Fraud and Assessing Safety in the Fishery and Aquaculture Sectors through Rapid Analytical Strategies

PHD THESIS

PHD THESIS SUPERVISORS:

Prof. Sergio Ghidini

Prof. Emanuela Zanardi

Research Group for Hygiene and Inspection of Food of Animal Origin

Department of Food and Drug, University of Parma

“The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.”

—— *William Lawrence Bragg* ——

INDEX

PREFACE

Food fraud affecting the fishery and aquaculture sectors has become a pressing problem not only for the significant economic losses, but also for the unavoidable repercussions on the safety of the whole supply chain.
Despite this, at a date, few steps have been taken to properly protect manufacturers, consumers, and the same fish and seafood products. Therefore, it is of utmost importance that the scientific research keeps becoming more oriented toward the development of analytical tools up with the times, rapid, cheap, and, above all, easy to understand and to be implemented.
The present doctoral thesis is focused on the evaluation of the potential of different fingerprinting and profiling analytical strategies as a means to solve many actual authenticity and safety challenges, in order to preserve the whole integrity of the fish and seafood production chain. A particular focus was put, throughout the dissertation, towards the mislabelling of the country of origin of fish and seafood products, which is one of the fastest-growing economically motivated fraud affecting the sector. The feasibility of using one single technique to address production methods and rearing systems authenticity problems at once was also tested, together with the investigation of the possibility to concomitantly assess safety issues relating to the presence of potentially toxic metals.

- The thesis opens with an *Introduction* part providing an overview of the current quality and authenticity problems which occur along the fish and seafood production chains and summarises the legislative and analytical tools currently available to detect and tackle fraud.

- The *Objective of the Thesis* part sums up the established goals of the research project developed during the PhD program.

The scientific contribution of this thesis was structured into two macro sections, each one subdivided into different chapters introduced by a state-of-the-art review.

- The *First Section* focuses on the potential of vibrational spectroscopy and, in particular, near infrared spectroscopy (NIR) to deter fish and seafood fraud and it is in turn divided into three chapters.

  Chapter one is a systematic review of the studies dealing with the use of qualitative vibrational spectroscopy combined with chemometrics for the resolution of the main authenticity issues affecting fish and seafood products.

  Chapter two presents an analytical integrated approach exploiting NIR spectroscopy and multivariate data analysis to rapidly and simultaneously investigate the authenticity of European sea bass according to production methods, rearing systems, and geographical provenances.

  Chapter three questions NIR spectroscopy-based fingerprints of lipid and protein degradation patterns of salt-ripened anchovies as a practical mean of ensuring the traceability of fishery products after industrial processing and at different stages of the production chain.

- The *Second Section* is divided into four chapters investigating the possibility of using the inorganic elemental composition of fish and seafood as an alternative strategy to monitor safety and traceability.

  Chapter four is a summary of the most recent and promising studies suggesting the inorganic composition of fish and, specifically, multielement composition, as a tool to reconstruct the history of fish, molluscs, and crustaceans.

  In *Chapter five* stable isotope ratios of carbon and nitrogen and rare earth elements analyses performed on wild and farmed sea bass tissues from different origins were evaluated about their ability to detect mislabelling concerning the provenance and the method of production.

  In *Chapter six* a method aimed to the protection and promotion of traditional local fishery products is presented. The chemometric discrimination among the mineral, trace- and ultra-trace element

profiles of Chioggia cuttlefish (traditional Italian products with a recognized quality mark) and common cuttlefish from other sources was studied by inductively-coupled mass spectrometry (ICP-MS) and chemometrics.

Chapter seven deals with the development of classification rules to easily distinguish among high-quality transformed anchovy products from the Cantabrian Sea and other competing but lower-quality products of different origin by exploiting information behind multielement patterns measured by ICP-MS. Decision trees built by using different machine learning algorithms were used for this purpose.

- Finally, the *General Conclusions and Future Outlook* part briefly summarises the achieved goals within the project, underlines the contribution of the results in the field of fish quality, safety, and authenticity, and discusses the ongoing practical applications of the developed methodologies.

INTRODUCTION

A general framework of fraud in the fish and seafood sector

The fishery and aquaculture sector has seen notable growth and development in recent years, to such an extent that 179 million tonnes of fish and seafood products were produced in 2018 and more than 204 million tonnes are expected to be produced by the year 2030 (FAO, 2020). This great expansion is the result of multiple and complex changes that have moved over time the commercial and distribution structure of almost all countries of the world towards an increasingly globalized environment.

Improvement of the logistics of goods transport, storage, free-trade agreements among countries, advances in aquaculture techniques or sea farming, as well as the growing general awareness of beneficial effects deriving from fish consumption have led to an increased demand for fish supply, to a point that consumption increased by 122% during the 1999-2018 period (FAO, 2020). Concurrent with this, the fishery and aquaculture production has become very diversified in terms of species and transformed products marketed and this inevitably contributed to the increase of the asymmetry in between all the players in the supply chain and the final consumer, as well as to a general atmosphere of confusion and uncertainty.

Because of all these reasons, fish and seafood products are particularly prone to fraudulent practices, being today the second-most likely category of food traded internationally at risk of fraud (European Parliament, 2013).

At present, no univocal and harmonized definition of *food fraud* has been established by the European Union (European Parliament, 2014). Anyway, the term *food fraud* globally refers to all kind of illegal practices (intentional or accidental), including substitution, addition, tampering or misrepresentation of foodstuffs, food ingredients or food packaging, as well as incorrect or misleading claims accompanying a food product usually motivated by illicit financial gains (Spink and Moyer, 2011).

In the context of fish and seafood products, fraud can range from species substitution and mislabelling involving the production method or the geographical provenance of fish to the undeclared or illegal use of food additives (Reilly, 2018). Specifically, mislabelling and species substitution have been reported as being the most frequent fraudulent activities,

affecting many biological species and occurring at every stage of the fish marketing chain, from primary production to catering (Pardo, Jiménez, & Pérez-Villarreal 2016; Oceana, 2016). For instance, the rate of mislabelling of white fish species in Europe was reported to be around 6% (European Commission, 2015), while in Italy it was found to reach about 44% for cephalopod products, 17% for crustaceans, and 14% for fish, especially when the products are imported from Vietnam, Thailand and China (Guardone et al., 2017).

In contrast, no occurrence data concerning the misrepresentation of the geographical origin are currently available, but the fact of the matter is that this practice keeps getting encouraged by the so-called country-of-origin effect, according to which the consumers increasingly tend to associate high quality fish products with specific production areas because of particular sensorial characteristics, ethical or ecological motivations. Considering that the overall perception of quality by consumers directly influences the global economic and market values of the product, it is important to underline that fish fraud is first of all responsible for significant worldwide economic losses, but, at the same time, it poses a problem for the sustainability of fish stocks and marine biodiversity since it potentially masks illegal, unreported, and unregulated fishing (Verrez-Bagnis, Sotelo, Mendes, Silva, Kappel, & Schröder, 2019).

Additionally, food mislabelling can have in certain cases some significant health consequences for consumers. It is the case, for instance, of the fraudulent substitution of fish with some naturally toxic fish species such as pufferfish or fish belonging to the Gempylidae family. Similarly, public health implications may sometimes derive from the substitution of wild fish with aquaculture fish from certain areas contaminated with heavy toxic metals, pesticides, or antibiotic residues (FAO, 2018). Therefore, given the possible health consequences, the concept of *food fraud* should be broadened to refer more properly to *food crime* (Sun, 2008).

Tackling fish fraud and monitoring fish authenticity and safety on several fronts

Regulatory and technical tools

During the last years, European Union has issued a comprehensive series of regulations ensuring the provision of labelling and traceability information along the production chain and to the final consumers, thus being the official instruments to prevent fraud and specifically protect the fishery and aquaculture sector. These regulations consist of the so-called Common Fisheries Policy (Council Regulation (EC) No. 1224/2009, Commission Implementing Regulation (EU) No. 404/2011, and Council Regulation (EU) No. 1379/2013). In addition, general rules for the provision of information to the final consumers included into Regulation 1169/2011 also apply to pre-packed fish products.

Traceability documentation, accompanying the products from catching or farming to market, is an effective tool to keep trace of origin and characteristics of fish since it requires biological species, provenances, date of fishing, batch, etc. to be recorded during the manufacturing chain. This way, traceability documentation allows to effectively recall products from the supply chain when some problems occur, thus being of utmost importance to prevent fraud and improve safety.

As for labelling mandatory requirements, the article 35 of Regulation (EU) No. 1379/2013 specifically established the following information to be provided to the final consumers within the labels of fresh and certain transformed fish and seafood product: i. the commercial designation plus the scientific name of the species; ii. the production method (caught or farmed fish); iii. the production area of the caught or farmed fish and the category of fishing gear for caught fish; iv. the possible defrosting; v. the date of minimum durability (where appropriate).

Lastly, the Regulation (EU) 2017/625 lays the groundwork for a risk-based control of the authenticity of foodstuffs by the implementation of standardized and validated methods to strengthen food labelling and traceability truthfulness.

In addition to official requirements, voluntary tools to be implemented throughout the food supply chain have been developed as prevention instruments, in order to help industries in better identifying the vulnerability of foodstuffs and mitigate the occurrence of fraud (Soon, Krzyzaniak, Shuttlewood, Smith, & Jack, 2019). For instance, the Vulnerability Assessment and Critical Control Points (VACCP) is a useful tool for the identification of those stages of the manufacturing process which are the most at risk of fraudulent activities, thus helping in reducing the probability of fraud (Spink, Ortega, Chen, & Wu, 2017). Other free-available food fraud vulnerability assessment tools such as the Safe Supply of Affordable Food Everywhere (SSAFE) (SSAFE, 2018) have been successfully implemented into certain food chains such as those of olive oil (Yan, Erasmus, Toro, Huang, & Van Ruth, 2020), milk (Yang et al., 2019), and spices (Silvis, Van Ruth, Van der Fels-Klerx, & Luning, 2017), and effectively helped industries in monitoring susceptibility of foodstuffs and reducing fraud.

Based on this scenario, it is clear that the authenticity issues concerning fish and seafood are of utmost importance for regulators, as well as for primary producers, food industry, and consumers, but in order to be effective, the tackle against fraud must be accompanied by the development of practical instruments providing certainty of accurate labelling with science-based evidence.

Analytical tools

Food authentication is known to be the process by which food is verified to be in accordance with its label description. Authentication analyses, therefore, are those analytical investigations aimed to verify the truthfulness of the claimed information regarding the product.

So far, the requirement for specific, valid, and accurate methods to ensure fish chain integrity is driven not only by food authorities, but also by the same manufacturer and processors.

The interest of food scientists towards the development of reliable methods for the resolution of several food authenticity and safety issues is well documented by the increasing number of scientific works which, albeit

through different methodologies, have attempted to address the same problems. At a date, trends in authenticity and traceability research are oriented towards the combination of different complementary techniques based on organic and/or inorganic composition of food, in order to comprehensively characterise foodstuffs and identify with great accuracy the discriminatory features to solve the authenticity issue under investigation.

It is clear, from the analysis of the latest literature, that several targeted and untargeted approaches laying down on the principles of mass spectrometry, molecular and nuclear spectroscopy, chromatographic separation, as well as bio-molecular and sensory techniques have been widely exploited to this purpose, demonstrating their exceptional multipurpose qualities for fish and seafood authenticity testing (Primrose, Woolfe, & Rollinson, 2010; Li, Boyd, & Sun, 2016; Mazzeo & Siciliano, 2016; Abbas et al., 2018; Asensio, González, García, & Martín, 2018; Esteki et al., 2018; Haynes, Jimenez, Pardo, & Helyar, 2019). Despite this, applications for the detection of geographical provenance has been less explored since it is more difficult to extract information from complex matrices such as fish to be modelled and univocally related to origin. Most of these techniques are known to share certain common disadvantages, such as long time needed for analysis, high costs of the equipment, the need of an elaborate and long-lasting sample preparation prior to analysis, destructiveness, and the demand for qualified personnel. On the other hand, as more consolidated within the research community, these techniques excel by their high accuracy, specificity, and sensitivity, to the point that many of them are considered gold standards for food official controls.

Nevertheless, considering that the assessment of the authenticity of fish and seafood products is complicated by the rapid deterioration of goods and the consequent need for very fast processing operations, what is currently most demanded by manufactures is the development of methods so rapid, easy to be applied, cheap, and, possibly non-destructive and eco-friendly, that they can certainly meet the needs of modern industry (Spink et al., 2016).

With advancing technology in computer science and data analytics, high-throughput applications based on *fingerprinting* and *profiling* strategies are

growing, as they allow to study food composition thoroughly and carefully, as well as any genetic or external environmental factor influencing the food identity, thus bypassing many obstacles related to the application of the conventional techniques (Esslinger, Riedl, & Fauhl-Hassek, 2014).

Vibrational fingerprinting spectroscopy and multielement profiling, when coupled with innovative chemometrics or machine learning techniques, can be regarded as forthcoming essential tools for the differentiation of fish samples in a comprehensive and detailed way, even if they are still under development and uncertainties remain.

Based on this background, the incoming chapters of this PhD Thesis, which is focused on the development of rapid and practical analytical strategies for fish authenticity and safety based on near infrared spectroscopy on the one hand and multielement profiling techniques on the other, are introduced by two review articles, where more detailed information about the principles behind the methods together with an overview of recent findings and applications have been provided.

References

Abbas, O., Zadravec, M., Baeten, V., Mikuš, T., Lešić, T., Vulić, A., Pleadin, J. (2018). Analytical methods used for the authentication of food of animal origin. *Food Chemistry*, *246*, 6-17.
https://doi.org/10.1016/j.foodchem.2017.11.007

Asensio, L., González, I., García, T., & Martín, R. (2018). Determination of food authenticity by enzyme-linked immunosorbent assay (ELISA). *Food Control, 19*, 1-8. https://doi:10.1016/j.foodcont.2007.02.010.

Commission Implementing Regulation (EU) No 404/2011 of 8 April 2011 laying down detailed rules for the implementation of Council Regulation (EC) No 1224/2009 establishing a Community control system for ensuring compliance with the rules of the Common Fisheries Policy. *Official Journal of the European Union*, *L112*, 1–30. Available online at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011R0404&from=EN

Council Regulation (EC) No 1224/2009 of 20 November 2009 establishing a Community control system for ensuring compliance with the rules of the common fisheries policy, amending Regulations (EC) No 847/96, (EC) No 2371/2002, (EC) No 811/2004, (EC) No 768/2005, (EC) No 2115/2005, (EC) No 2166/2005, (EC) No 388/2006, (EC) No 509/2007, (EC) No 676/2007, (EC) No 1098/2007, (EC) No 1300/2008, (EC) No 1342/2008 and repealing Regulations (EEC) No 2847/93, (EC) No 1627/94 and (EC) No 1966/2006. *Official Journal of the European Union, L343,* 1–50. Available online at: http://eur-lex.europa.eu/legal-ontent/EN/TXT/PDF/?uri=CELEX:32009R1224&
from=EN

Esslinger, S., Riedl, J., & Fauhl-Hassek, C. (2014). Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Research International*, *60*, 189-204. https://doi.org/10.1016/j.foodres.2013.10.015

Esteki, M., Simal-Gandara, J., Shahsavari, Z., Zandbaaf, S., Dashtaki, E., & Vander Heyden, Y. (2018). A review on the application of chromatographic methods, coupled to chemometrics, for food authentication (Chromatography-chemometrics in food authentication). *Food Control, 93*, 165-182. https://doi:10.1016/j.foodcont.2007.02.010.doi:10.1016/j.foodcont.2018.06.015.

European Parliament (2013). Report on the food crisis, fraud in the food chain and the control thereof. Available online at: https://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT +REPORT+A7-2013-0434+0+DOC+XML+V0//EN

European Parliament, (2014). European Parliament resolution of 14 January 2014 on the food crisis, fraud in the food chain and the control thereof (2013/2091(INI)). *Official Journal of the European Union, C482*, 22-30. Available online at: https://eur-lex.europa.eu/legal-ontent/EN/TXT/HTML/?uri=CELEX :52014IP0011(01)&from=IT

European Commission. 2015. Fish substitution (2015). Available online at: https://ec.europa.eu/food/safety/official_controls/food_fraud/fish_substitution_en

FAO (2020). The State of World Fisheries and Aquaculture 2020. Sustainability in

action. Rome. Available online at: https://doi.org/10.4060/ca9229en

Guardone, L., Tinacci, L., Costanzo, F., Azzarelli, D., D'Amico, P., Tasselli, G., Magni, A., Guidi, A., Nucera, D. & Armani, A. (2017). DNA barcoding as a tool for detecting mislabeling of fishery products imported from third countries: An official survey conducted at the border inspection post of Livorno-Pisa (Italy). *Food Control, 80*, 204–216. https://doi.org/10.1016/j.foodcont.2017.03.056.

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. https://doi:10.1016/j.foodcont.2007.02.010.doi:10.1016/j.foodcont.2019.02.010.

Li, L., Boyd, C. E., & Sun, Z. (2016). Authentication of fishery and aquaculture products by multi-element and stable isotope analysis. *Food Chemistry, 194*, 1238–1244. https://doi.org/10.1016/j.foodchem.2015.08.123

Mazzeo, M.F. & Siciliano, R.A. Proteomics for the authentication of fish species. Journal of proteomics, 147, 119-124. https://doi:10.1016/j.foodcont.2007.02.010.doi:10.1016/j.jprot.2016.03.007.

Oceana. 2016. Deceptive dishes: seafood swaps found worldwide. Available online at: http://usa.oceana.org/publications/reports/deceptive-dishes-seafood-swaps-found-worldwide

Pardo, M.Á., Jiménez, E., & Pérez-Villarreal, B. (2016) Misdescription incidents in seafood sector. *Food Control, 62,* 277–283. https://doi.org/10.1016/j.foodcont.2015.10.048

Primrose, S., Woolfe, M., & Rollinson, S. (2010). Food forensics: methods for determining the authenticity of foodstuffs. *Trends in Food Science & Technology, 21*, 582-590. https://doi.org/10.1016/j.tifs.2010.09.006.

Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004.

Regulation (EU) No 1379/2013 of the European Parliament and of the Council of 11 December 2013 on the common organisation of the markets in fishery and aquaculture products, amending Council Regulations (EC) No 1184/2006 and (EC) No 1224/2009 and repealing Council Regulation (EC) No 104/2000.

Regulation (EU) 2017/625 of the European Parliament and of the Council of 15 March 2017 on official controls and other official activities performed to ensure the application of food and feed law, rules on animal health and welfare, plant health and plant protection products, amending Regulations (EC) No 999/2001, (EC) No 396/2005, (EC) No 1069/2009, (EC) No 1107/2009, (EU) No 1151/2012, (EU) No 652/2014, (EU) 2016/429 and (EU) 2016/2031 of the European Parliament and of the Council, Council Regulations (EC) No 1/2005 and (EC)

No 1099/2009 and Council Directives 98/58/EC, 1999/74/EC, 2007/43/EC, 2008/119/EC and 2008/120/EC, and repealing Regulations (EC) No 854/2004 and (EC) No 882/2004 of the European Parliament and of the Council, Council Directives 89/608/EEC, 89/662/EEC, 90/425/EEC, 91/496/EEC, 96/23/EC, 96/93/EC and 97/78/EC and Council Decision 92/438/EEC (Official Controls Regulation).

Reilly, A. (2018). Overview of food fraud in the fisheries sector. *FAO Fisheries and Aquaculture Circular* (C1165), I-21. Available online at: http://www.fao.org/3/i8791en/I8791EN

Silvis, I. C. J., Van Ruth, S. M., Van der Fels-Klerx, H. J., & Luning, P. A. (2017). Assessment of food fraud vulnerability in the spices chain: An explorative study. *Food Control, 81*, 80-87. https://doi.org/10.1016/j.foodcont.2017.05.019

Soon, J. M., Krzyzaniak, S. C., Shuttlewood, Z., Smith, M., & Jack, L. (2019). Food fraud vulnerability assessment tools used in food industry. *Food Control, 101*, 225-232. https://doi.org/10.1016/j.foodcont.2019.03.002

Spink, J., & Moyer, D. C. (2011). Defining the public health threat of food fraud. *Journal of food science, 76*, 9, R157-R163.
https://doi.org/10.1111/j.1750-3841.2011.02417.x

Spink, J., Fortin, N. D., Moyer, D. C., Miao, H., & Wu, Y. (2016). Food fraud prevention: policy, strategy, and decision-making–implementation steps for a government agency or industry. *CHIMIA International Journal for Chemistry*, *70*(5), 320-328. https://doi.org/10.2533/chimia.2016.320

Spink, J., Ortega, D. L., Chen, C., & Wu, F. (2017). Food fraud prevention shifts the food risk focus to vulnerability. *Trends in Food Science & Technology, 62*, 215-220. https://doi.org/10.1016/j.tifs.2017.02.012

SSAFE food fraud vulnerability assessment tool (2018). Available online at: httpS://www.ssafe-food.org/

Sun, D. W. (Ed.) (2008). Modern techniques for food authentication. Elsevier/Academic Press, London, UK.

Verrez-Bagnis, V., Sotelo, C. G., Mendes, R., Silva, H., Kappel, K., & Schröder, U. (2019). Methods for Seafood Authenticity Testing in Europe. *Reference Series in Phytochemistry*, 2063–2117. https://doi:10.1007/978-3-319-78030-6_69

Yan, J., Erasmus, S. W., Toro, M. A., Huang, H., & van Ruth, S. M. (2020). Food fraud: Assessing fraud vulnerability in the extra virgin olive oil supply chain. *Food Control, 111*, 107081. https://doi.org/10.1016/j.foodcont.2019.107081

Yang, Y., Huisman, W., Hettinga, K. A., Liu, N., Heck, J., Schrijver, G. H., & van Ruth, S. M. (2019). Fraud vulnerability in the Dutch milk supply chain: Assessments of farmers, processors and retailers. *Food Control, 95*, 308-317.
https://doi.org/10.1016/j.foodcont.2018.08.019

AIMS AND OBJECTIVES OF THE THESIS

The present PhD Thesis was aimed at investigating the possibility of developing rapid analytical methods based on NIR spectroscopic fingerprinting or elemental profiling strategies merged with multivariate data analysis or machine learning techniques to uncover and prevent fraudulent commercial practices targeting the fishery and aquaculture sectors and assessing the safety of fish and seafood products.

Below are the specific purposes and objectives set for the PhD research activity.

1. Address multiple authenticity issues of fish and seafood, intended both as raw and transformed products, through the development of chemometric-based discriminant models using the whole near infrared (NIR) spectrum and chemometrics. In particular, an attempt to answer the following questions was made:

   a. Is it possible to use one single NIR spectrum to concurrently discriminate the production method (wild vs. farmed), the farming system (extensively-, semi-intensively-, and intensively-reared fish), and the geographical origin (central, eastern, and western Mediterranean Sea) of sea bass?

   b. Do the NIR spectra of salt-ripened anchovies continue to hold sufficient and useful information related to the provenance of the products even after heavy processing? Do the lipid and protein degradation patterns show valuable traces of the original composition of the raw fish for traceability purposes?

   The related answers are reported in the *First Section* of this PhD Thesis, *Chapters 1, 2, and 3.*

2. Verify whether the inorganic composition of specific marine fishing areas is reflected in the inorganic composition of fish and seafood to a point that it can be effectively used to trace back to the origin of the product. Although going against the flow, this objective was focused on the possibility of developing classification methods starting from a *whole element profiling approach* and progressively moving towards a *targeted* and *parsimonious approach* by identifying the lowest but the

most significant number of inorganic markers to be used to achieve satisfying classification goals. Specifically, the following questions arose:

a. Is the merging of the outputs from stable isotope ratio analysis of carbon and nitrogen and rare earth elements analysis an accurate analytical alternative to evaluate the truthfulness of information reported on sea bass label concerning its production and geographical history?

b. Can the geographical imprint of Italian traditional cuttlefish from Chioggia be revealed through the multivariate analysis of key elements and used to promote local fishery products? Is there the possibility to take advantages of the quite stable composition of seawater of Venice Lagoon (where the cephalopod is produced) which is characterised by a limited replacement of fresh water?

c. Taking into consideration that presence of minerals and trace- and ultra-trace elements in fish is strongly dependent not only on the sea composition but also on the manufacturing environment, is it feasible to select those elements less influenced by the transformation process? Since Cantabrian anchovy is particularly prone to fraudulent substitution with other products due to its overall reputation and high commercial value, is there the possibility of generating simplified classification rules using selected elements to trace Cantabrian anchovy to its origin before and after packaging and sale?

The related answers are reported in the *Second Section* of this PhD Thesis, *Chapters 4, 5, 6, and 7.*

FIRST SECTION

CHAPTER 1

Review Article

Approaching Authenticity Issues in Fish and Seafood Products by Qualitative Spectroscopy and Chemometrics

Sergio Ghidini, Maria Olga Varrà* and Emanuela Zanardi

Department of Food and Drug, University of Parma, Strada del Taglio 10, 43126 Parma, Italy.

*Correspondence: mariaolga.varra@studenti.unipr.it; Tel.: +39-0521-902-761

Adapted from

Ghidini, S., Varrà, M. O., & Zanardi, E. (2019). Approaching authenticity issues in fish and seafood products by qualitative spectroscopy and chemometrics. Molecules, 24(9), 1812.

Abstract: The intrinsically complex nature of fish and seafood, as well as the complicated organisation of the international fish supply and market, make today the struggle against counterfeiting and falsification of fish and seafood products very difficult. The development of fast and reliable omics strategies based on spectroscopy in conjunction with multivariate data analysis has been attracting great interest from food scientists, so that the studies linked to fish and seafood authenticity have increased considerably in recent years. The present work has been designed to review the most promising studies dealing with the use of qualitative spectroscopy and chemometrics for the resolution of the key authenticity issues of fish and seafood products, with a focus on species substitution, geographical origin falsification, production method or farming system misrepresentation, and fresh for frozen/thawed product substitution. Within this framework, the potential vibrational spectroscopy combined with both unsupervised and supervised chemometric techniques have been highlighted, each time pointing out the trends in using one or another analytical approach and the performances achieved.

Abbreviations: ANN = artificial neural networks; BBN = Bayesian belief network; FIR = far-infrared; FDA = factorial discriminant analysis; FFT = fast Fourier transform; FT = Fourier transform; HCA = hierarchical cluster analysis; HSI = hyperspectral imaging; IR = infrared; k-NN = k-nearest neighbours; LDA = linear discriminant analysis; LW-NIR = long-wave near infrared; MIR = mid-infrared; NMR = nuclear magnetic resonance; MSC = multiplicative scatter correction; NIR = near-infrared; OPLS-DA = orthogonal partial least square-discriminant analysis; PCA = principal component analysis; PLS-DA = partial least square-discriminant analysis; PNN = probabilistic neural network; QDA = quadratic factorial analysis; SERS = surface-enhanced Raman spectroscopy; SG = Savitzky-Golay smoothing; SIMCA = soft independent modelling of class analogy; SNV = standard normal variate; SVM = support vector machine; SW-NIR = short-wave near infrared; UV = ultraviolet; Vis = visible.

1. Introduction

The demand for fish and seafood products has increased notably during the last years, mostly as a consequence of the new special attention paid by consumers towards healthier food. The technological development that has invested the whole fisheries sector has additionally contributed to overcome the well-known obstacles to export fish and seafood worldwide, deriving from the high vulnerability of the products, to the point that today more than 35% of all caught and cultured fish is traded across national boundaries (FAO, 2018).

The growing competitiveness of the sector and diversification in fish supply chain have, in turn, led to the presence of a huge variety of look-alike products on the international market, whose global quality features are, however, quite different. More than 700 different species of fish, 100 of molluscan and 100 of crustacean are, in fact, used as food for humans (Rehbein & Oehlenschläger, 2009).

In this scenario, what is remarkable is that consumers demand not only for more fish, but for even safer and higher-quality fish, whilst the deliberate or accidental lack of transparency about the identity of products and fraudulent or negligent activities just keep on growing.

Based on what has been recently reported by the Food and Agriculture Organization, fish and related products have become among the most vulnerable to fraud category of food. Nevertheless, the effective monitoring of illicit practices in the fisheries sector is hampered by the increasing spread of highly processed fish products, in which the presence of different types of fraud can be hidden with ease (FAO, 2018).

The voluntary substitution of commercially valuable fish species with lower quality ones, represents the most recurrent form of fish fraud, although substitution can also take place accidentally when species look so similar that they are mistaken for each other.

The geographical provenance and the production process are other current authenticity topics concerning fish and seafood products, whose falsification, hard to bring to light, has a negative economic impact. Despite being economically motivated, mislabelling concerning these issues may

occasionally represent a risk to public health. The illegal commercialisation of poisonous fish species (Tetraodontidae, Molidae, Diodontidae, and Canthigasteridae families) or the replacement of certain kind of raw fish fillets with gastro-intestinal toxic fish (i.e. those belonging to the Gempylidae family) are just some of many examples. Likewise, occurrence of some harmful marine biotoxins may be linked to the geographical distribution of the producing organisms (Van Dolah, 2000), while the presence of higher levels of heavy metals or residues of antibiotic and pesticides can be found in farmed products than in wild ones (Fallah, Saei-Dehkordi, Nematollahi, & Jafari, 2011; Okocha, Olatoye, & Adedeji, 2018; Kelly, Ikonomou, Higgs, Oakes, & Dubetz, 2011).

Ensuring a clear discrimination of authenticity of fish and seafood is today of special concern not only for consumers, but also for producers, traders, and industries. Traceability throughout the whole production chain and at all stages of the market, covered by Regulations 178/2002/EC (Council regulation (EC) No 178/2002), 1005/2008/EC (Council regulation (EC) No 1005/2008), 1224/2009/EC (Council regulation (EC) No 1224/2009), is considered to be the starting point for the assurance of a high level of safety and quality of food and ingredients, as it represents the basic instrument not only for preventing illegal activities, but also for protecting consumers through the opportunity to access information about fish exact nature and characteristics.

Specific regulations for the provision of information to consumers (Regulation (EU) No 1169/2011), and for the requirement to uniquely identify fish and seafood on the label (Regulation (EU) No 1379/2013) play also an essential role in providing more transparency regarding the nature of the products, as they allow consumers to make informed choices and further contribute to the implementation of seafood traceability.

As a matter of fact, labels of all unprocessed and some processed fishery and aquaculture products must include information on both the commercial and the scientific names of the species, whether the fish has been caught or farmed, the catch or the production area, the fishing gear used, whether the product has been defrosted, and the date of minimum durability (where appropriate).

Many other voluntary claims can also be reported on the label, including the date of catch/harvest for wild/aquaculture products, information about the production techniques and practices, environmental and ethical information (Regulation (EU) No 1379/2013).

All the claimed declaration appearing on the label must always be checked to verify whether they are truthful. Therefore, in spite of the utility of the traceability system, the fisheries sector needs effective methods to address the problem of fish authenticity and ensure product quality. Innovative analytical approaches based on the evaluation of total spectral properties, are rapidly gaining ground at all levels of current food authenticity research, thanks to their ability to simultaneously provide lots of information related to physical and chemical characteristics of the food matrix.

Recent advances in chemometrics, moreover, have represented the major turning point in the dissemination of 'fingerprinting strategies', as they allow to study all the genetic, environmental and other external factors influencing the food identity, and to bypass many obstacles related to the application of the conventional techniques (Esslinger et al., 2014). This way, chemometrics can be now considered an essential tool for differentiation of similar samples according to the authentication issues of interest.

Till now, several spectroscopic techniques in conjunction with chemometrics have been used as rapid, simple, and cheap tools for fish quality and authenticity testing. Among these, vibrational (near-infrared [NIR], mid-infrared [MIR], Raman), fluorescence or absorption ultraviolet-visible (UV-Vis), and nuclear magnetic resonance (NMR) spectroscopies, together with hyperspectral imaging (HSI) spectroscopy, represent the most used ones, even if they are still being developed.

Based on this background, the present review article has been designed to highlight uses and developments of fast and reliable omics strategies based on UV-Vis, NIR, MIR, Raman, and HSI spectroscopies, with the attempt to address the key authenticity challenges within the fish and seafood sector.

To this end, a brief discussion concerning basilar concepts underlying these techniques has been provided, and it has been accompanied by a short overview about the implementation of several chemometric tools, in order to

highlight the potential benefits in extracting relevant information from spectral data.

The main body of this review focuses specifically on the application, over the years, of spectroscopy and chemometrics to distinguish products in accordance with the species, production method (wild or farmed), the farming system (conventional or organic; intensive, semi-intensive or extensive), the geographical provenance (different FAO areas and countries of origin), and the processing technique (fresh or fresh/thawed), which, at present, correspond to the key authenticity concerns for which there must be ongoing and effective monitoring.

2. A conceptual framework of spectroscopy and chemometrics

Spectroscopy is the study of electromagnetic radiation interacting with matter, which can be absorbed, transmitted, or scattered on the basis of both the specific frequency of the radiation and the physical/chemical nature of the matter.

When absorbed, radiation leads to a change in energy states of the atoms, nuclei, molecules, or crystals that make up matter, inducing an electronic, vibrational, or rotational transition, depending on the energy of the incident radiation (Picò, 2012). When the radiation, at a specific frequency, is scattered by molecules (as in Raman spectroscopy), some changes can occur in the energy of the incident photon, which transfer parts of its energy to the matter. In any case, the result of these interactions is a spectrum enclosing many features of the matter analysed, which, when properly interpreted with the help of chemometrics, can be used in a great number of different applications.

In choosing the most appropriate spectroscopic method to be used, consideration should be given to some factors, which go beyond the purely analytical purposes: the physical state and the chemical composition of the sample, sensitivity, specificity, and overall accuracy of the technique, scale of operation, time of analysis, and cost/availability of the instrumentation (Shrieber, 2008). For sake of conciseness, the main features related to spectroscopic techniques used mostly in the food authentication field are summarised in Table 1.

Table 1. Comparison of different spectroscopic techniques used for food authentication purposes: summary of the main characteristics

Spectroscopic technique		Wavelength range (nm)	Interaction light-matter	Basic principle	Sensitive compounds	Information obtained	Applications	Possible limitations
UV-Vis:	**UV**	2–4 x 10^2	Absorption/ emission	Electronic transitions	Double-conjugated bonds; isolated double, triple, peptide bonds; aromatic and carbonyl groups	Molecular structure	Qualitative/ quantitative	Need of sample preparation pH and temperature interferences
	Vis	4–7.5 x 10^2						
IR[1]:	**NIR**	7.5–25 x 10^2	Absorption	Vibrations/ Rotations of molecular bonds (changes in dipole moments)	Polar bonds (N-H, C-H, O-H, S-H, C-O)	Chemical bonds and physical structure	Qualitative/ quantitative	Water interferences Overlapping of spectral peaks
	MIR	2.5–25 x 10^3						
Raman		2.5 x 10^3–10 x 10^5	Scattering	Vibrations of molecular bonds (changes in polarizability)	Non polar double or triple bonds (C=C, C≡C)	Chemical bonds and physical structure	Qualitative/ quantitative	Fluorescence and photodecomposition interferences Low-Intensity Peaks
HSI		Varying by spectroscopic modules	Absorption/ emission/ scattering	Varying by vibrational spectroscopic modules	Varying by vibrational spectroscopic modules	Varying by vibrational spectroscopic modules	Qualitative/ Quantitative/ Spatial	Varying by vibrational spectroscopic modules

[1] IR electromagnetic region taken into consideration do not include FIR range (2.5–10 x 10^4 nm) since it is not commonly used in food authentication studies. [2] H-1, C-13 and P-31 are the most frequently investigated nuclei in food science-related NMR applications

.1 UV-Vis absorption and fluorescence emission spectroscopy

UV-Vis spectroscopy involves the electronic excitation of molecules containing specific chromophore groups, which results from the absorption of photons at two wavelength regions of the electromagnetic spectrum. In the absorption mode, the amount of light retained by sample is measured, while in the fluorescence mode the amount of light emitted after absorption is taken into consideration (Shrieber, 2008). Typically, the UV-Vis spectrum is characterised by broad absorption or emission peaks which reflect the molecular composition of the matrix: by exploiting the unicity absorption or emission patterns of the entire spectrum, or by measuring the absorbance or fluorescence intensity of the analyte at one wavelength, this spectrum can be used for many food analytical qualitative and quantitative applications, respectively (Penner, 2017; Strasburg, 1995).

2.2. IR spectroscopy

Infrared spectroscopy involves three different sub-regions of the electromagnetic spectrum, namely NIR, MIR and FIR, whose absorption by samples results in vibrations of atoms in molecular bonds (Xu, Riccioli, & Sun, 2015). These vibrations give out a great amount of information related not only to chemical bonding, but also to the general molecular conformation, the structure, and the intermolecular interactions within the sample (Rodriguez-Saona & Allendorf, 2011). This way, IR spectra enclose total sample composition, whose pattern of peaks distribution represents a unique signature profile and whose intensity of bands is linked to the concentration of specific compounds (Rodriguez-Saona, Giusti, & Shotts, 2016; (Lohumi, Lee, Lee, & Cho, 2015). The NIR spectrum of food samples results from absorption by molecular bonds containing prevalently light atoms and it is characterised by the presence of broad and overlapping overtone and combination bands (Blanco & Villarroya, 2002; Cen & Yong, 2007). By contrast, spectral signature in the MIR region is characterised by the presence of more intense and delineated bands, whose position and intensity are more informative of molecule's concentration in the sample (Cheng et al., 2013; Cozzolino & Murray, 2012). Here too, the spectral profile is complex and data mining is very difficult without the use of multivariate

data analysis. Finally, with reference to FIR spectroscopy, it is noted that no applications to food authentication are currently available since it relates to molecules containing halogen atoms, organometallic compounds, and inorganic compounds, whose interest for is more limited within the context of food research (Stuart, 2004).

2.3. Raman spectroscopy

Raman spectroscopy is a molecular vibration technique based on the inelastic Raman scattering, a physical effect that comes with molecular vibrations and triggers a change in the polarizability of the molecule (Boyaci et al., 2015). In particular, this kind of spectroscopy focuses on the measurement of those small fraction of the radiation which is scattered by specific categories of compounds at higher or lower frequencies than incident photons. The typical Raman spectrum, showing intensities of the scattered light versus the wavelengths of Raman shift, is characterised by sharp and well-resolved bands, which provide information about molecular structure and composition of the matter analysed.

For a long time after its discovery, Raman spectroscopy has been poorly exploited in food applications, by reason of several analytical disadvantages and interference (see Table 1). These drawbacks have now been overcome thanks to the overall technological improvement of Raman equipment: by way of example, surface-enhanced Raman spectroscopy (SERS), has recently made it possible to surmount hurdles related to faint scattering signals (Zheng & He, 2014).

2.4. Hyperspectral imaging

HSI is a technique cobbling together spectroscopy and computer vision to give several useful information concerning physico-chemical characteristics of samples in relation to their specific spatial distribution. Briefly, HSI systems provide several hyperspectral images of the tested sample, corresponding to three-dimensional data containers, of which each sub-image is a map showing spatial distribution of the sample constituents in relation to each single wavelength (Feng & Sun, 2012; Wu & Sun, 2013).

Over the recent years, the steady usage growth of HIS technology in the field of food research has been mainly driven by the availability of different instrumental configurations that allow to exploit fluorescence, absorbance, or light scattering phenomena. On the other side, application of spectral imaging technologies is not at all widespread in the food industry, due to variety of factors ranging from high costs and low availability of instrumentations, to the computation speed and necessity of expertise by users (Roberts, Power, Chapman, Chandra, & Cozzolino, 2018).

2.5. Qualitative chemometric methods

Raw spectra resulting from spectroscopic analyses are usually characterised by broad and unresolved bands containing too much information, some of which are certainly useful and need to be retained, but some of which hamper the correct data interpretation and need to be removed. Recent advances in chemometrics have marked an important milestone in spectra analysis, since they have simplified the identification of hidden interrelations between variables providing the key for discrimination and classification of samples Rodriguez-Saona, Giusti, & Shotts, 2016; Oliveri & Simonetti, 2016). In other words, qualitative chemometrics methods help to recognise similarities and dissimilarities within spectral data, which can be used to confirm the authenticity or detect adulteration of food samples (Manley & Baeten, 2008). Based on the explorative or predictive nature of the methodology, qualitative chemometric techniques are usually classified into unsupervised and supervised techniques. While unsupervised techniques are independent of prior knowledge of class membership of samples to perform classification, supervised techniques call for such knowledge. Brief descriptions of the principles behind the chemometric techniques which are being used to a greater extent are provided below.

2.5.1. Spectral pre-treatments

Pre-treatment of spectral data is recognized as being fully integrated into the chemometric set-up itself. Prior to develop chemometric models, raw spectroscopic data are suggested to be pre-processed by applying some corrections, aimed to enhance spectral properties and minimize those

fraction of systematic variation which do not contain relevant information to the discrimination of samples. One such systematic variation is the sum of different physical effects which arise during instrumental acquisition of spectra (e.g. light scattering or background fluorescence phenomena), which are responsible for the appearance, especially in solids samples, of multiplicative, additive, non-linearity effects (e.g. overlapping bands, baseline shifts/drifts, random noise) (Rinnan, Van der Berg, & Engelsen, 2009). Thus, pre-processing algorithms are usually classified into signal correction methods (e.g. multiplicative scatter correction, MSC; standard normal variate, SNV), differentiation methods (first, second, or third order derivation), and filtering-based methods (e.g. orthogonal signal correction, OSC; orthogonal wavelet correction, OWAVEC) (Pereira, Reis, Saraiva, & Marques, 2011).

While signal correction and filtering-based methods are conceived to retain only the spectral information mainly by suppressing the light-scattering effects, derivative-based methods also help to reduce the spectral complexity through the separation of the broad overlapping bands.

A more detailed description of spectral pre-processing techniques can be widely found in literature (Rinnan et al., 2009, Rinnan, 2014; Engel et al., 2013).

Either way, it is essential to point out that spectral filters are most often concatenated to exploit the effects of each one, but this concatenation might increase model complexity and background noise, resulting in an inaccurate chemometric modelling of data and, thus, in wrong predictions. For this reason, it is recommended to customize the selection of the pre-treatments prior to perform chemometric analysis according to the spectroscopic technique used and the sample characteristics, trying to restrict, whenever possible, their number.

2.5.2. Unsupervised methods

Unsupervised methods look at the study of variability among samples for the purpose of identifying their natural characteristics and possible similarities among them, without the need to provide any information about the class to which samples belong.

Between the various available techniques, principal component analysis (PCA) is the most used one. PCA is a quite basic projection method able to reduce the original correlated variables into a smaller number of new uncorrelated latent variables (known as principal components), containing as much systematic variation as possible of the original data (Ziegel, 2004). Score plots outputs deriving from PCA applications, show in a simple and intuitive graphical way the hidden structures among samples, the interrelations among variables and between samples and variables, the probable presence of any outliers, and possible groupings or dispersion of sample according to specific class membership.

Hierarchical cluster analysis (HCA) in another frequently employed unsupervised methods, based on the splitting of samples into different clusters. This splitting is based on the degree of analogy among samples and it is generally performed by evaluating the Mahalanobis or Euclidean distance between the same samples. The hierarchical approach followed is thus aimed at constructing a ladder, in which the most closely related samples are first classified into small groups, and then progressively assembled into bigger groups including less similar samples (Oliveri et al., 2016). Results of HCA are graphically expressed by tree diagrams (dendrograms), showing relationships among clusters; nevertheless, despite being easily computable, dendrograms are often misunderstood, since the number of clusters to be considered is arbitrary, making the interpretation of results more subjective than objective.

2.5.3. Supervised methods

Supervised techniques require the previous knowledge of the class membership of the samples tested, which can be used to develop predictive models able to discriminate and classify future unidentified samples. There are several different chemometric techniques belonging to the category of the supervised methods, most of which require a training set (to found classification rules for sample), and a test set (to assess the predictability of the model developed) (Voncina, 2009).

Linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) are variance-based methods which use Euclidean distance to find

those combinations of the original variables determining maximum separation among the different groups of samples (Rodriguez-Saona, et. al, 2016). Both techniques presume that the measurements within each class are normally distributed, but while LDA supposes that dispersion (covariance) is identical for all the classes, QDA, on the contrary, allows the possibility of different dispersion to be present within different classes (Oliveri et. al, 2016). Although QDA is considered an extension of LDA, there are some common limitations, for instance the risks of overfitting and failing in classification, especially when the samples size for each class in unbalanced.

K-nearest neighbours (k-NN) clustering is one of the simplest method to discriminate samples on the basis of the distance among them. Basically, after choosing the adequate number of k-neighbour samples, the algorithm identifies the k-nearest samples of known class membership to select classification of unknown samples. This method, unlike LDA and QDA, do not require any prior assumption and its success is independent of the homogeneity of samples number in each tested class (Berrueta, Alonso-Salces, & Héberger, 2007). Among supervised machine learning approaches, support vector machine (SVM) is particularly advantageous when samples classification is complicated by the non-linearity and the high dimensional space. The core of the method is the use of specific functions for pattern analysis (kernel algorithms), through which the margin of separation between classes is maximised and complex classification problems that are not linear in the initial dimension (but may be at high dimensional spaces) are resolved (Rodriguez-Saona et al., 2016).

Similarly, artificial neural network (ANN) is a machine learning method characterised by the ability of adapt to the data, providing classification also in presence of non-linearity input-output relationships. Since structured and organized in a less complex way than SVM, ANN usually generate a more rapid response at a lower computational cost; these efforts, however, are counterbalanced by reduction in accuracy (Rodriguez-Saona, 2016; Ahmad, Khalid, &Yosuf, 2002). Nevertheless, ANN suffers from poor data generalisation and, by consequence, it is inclined to return model's overfitting errors. This tendency to overfitting is the main reason why

accurate ANN computation analyses call for a very high number of samples to be considered, and, at the same time, require strict internal and external validations to be performed, where the training set and the test set should enclose as much similar variability as possible (Berrueta et al., 2007).

Soft independent modelling of class analogy (SIMCA) is alternative pattern recognition method which first perform individual PCA on the samples for each of the class they must be assigned to, in order to compress original variables into a smaller number of new principal components. Principal components and critical distances computed are then used to delineate a confidence limit for each class. Unknown samples are then assigned to the class to which they get close by projection into the resulting multidimensional space (Manley et al., 2008). SIMCA is particularly useful when samples belong to several different classes, but, since maximum class-separation is not covered by the method, the interpretation of the outcomes may be difficult if not impossible (Rodriguez-Saona et. al, 2016).

Regression-based supervised discriminant analyses exploit specific classification algorithms to model the interrelations existing between measured variables (i.e. spectra) and qualitative parameters (i.e. class membership), such that maximum separation between the different groups of samples is achieved. Partial least square-discriminant analysis (PLS-DA) and orthogonal partial least square-discriminant analysis (OPLS-DA) belong to this category of techniques. PLS-DA involves a standard PLS regression to find interrelations between the X-matrix (containing measured variables) and Y-matrix (containing categorical variables) by building new variables (latent variables). These interrelations allow not only to classify new samples into one of the Y-groups based on measured spectrum, but also to identify variables that mostly contribute to the classification. Despite PLS-DA has the advantage of modelling noisy and highly collinear data efficiently, the technique is often unsuccessful when the non-related (orthogonal) variability in the X-matrix is substantial, since hinder the correct interpretation of the results (Rodriguez-Saona et al., 2016). This drawback can be overcome by the application of OPLS-DA, through which the orthogonal variability within the X matrix is separated from the related (predicted) variability and then modelled apart. Consequently, if samples

cannot be discriminated along the predictive direction, the orthogonal variability may be handled to increase the effectiveness of discrimination among classes (Bylesjö et al., 2007).

3. Authenticating fish and seafood through the application of qualitative spectroscopy and chemometrics

Spectroscopic and chemometric analyses have been used over the years for many applications in fishery research, those in the authentication field being among the most promising ones. Some of the works concerning the flexibility of spectroscopy in fish and seafood analysis have been already reviewed by different authors (Uddin & Okazaki, 2006; Cozzolino & Murray, 2012; Cheng et al., 2013; He, Wu, & Sun, 2015; Fiorino et al., 2018), but they have been mainly centred on illustration of the advances of the available techniques for quality attributes assessment, as well as on the advantages and limitations of the single type of technique over traditional methods.

Therefore, in the following section, more attention has been paid to the resolution, on a case-by-case basis, of the weightiest authentication issues in the fish and seafood sector, namely species substitution, geographical origin falsification, production method or farming system misrepresentation, and fresh for frozen/thawed product substitution, each time pointing out the trends in using one or another method as well as the discrimination performances achieved, which are considered to be the most intuitive parameters used for chemometric models diagnostics. An overview of the most frequently investigated authentication issues in the fishery sector and the trend of using each spectroscopic technique over the years by the scientific community are plotted in Figure 1 and Figure 2, respectively.

3.1. Species substitution

Substitution or counterfeit of high-values fish species with low-values ones has many quality and safety implications. Therefore, the confirmation of scientific and commercial names declared on the label through the use of rapid and low-cost methods is increasingly popular in food research.

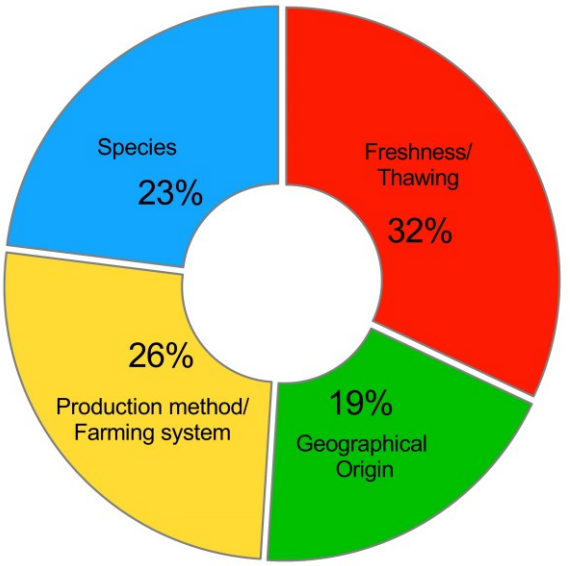


Figure 1. Percentage distribution of the authenticity issues covered by the scientific literature reviewed in the present work. Data were collected in February 2019 from the web search engine Google Scholar (search criteria: time period: "any time", and keywords: ''fish and/or seafood"; "authenticity" "spectroscopy"; "chemometrics".

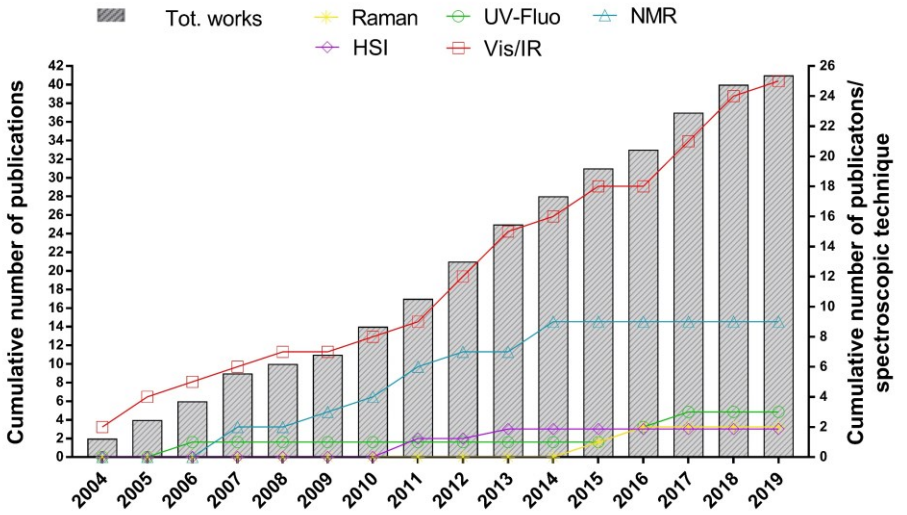


Figure 2. Combined bars and lines graph, where bars (plotted against the left Y-axis) show the cumulative number of scientific works concerning the use of spectroscopy and chemometrics for fish authentication purposes, and lines (plotted against the right Y-axis) show the cumulative number of works using each spectroscopic technique. Data were collected in February 2019 from the web search engine Google Scholar (search criteria: time period: "any time", and keywords: ''fish and/or seafood"; "authenticity" "spectroscopy"; "chemometrics".

3.1.1. Application of vibrational spectroscopy

An early study explored the Vis-NIR spectroscopy as a tool to detect the counterfeit of Atlantic blue crabmeat (Callinectes sapidus) with blue swimmer crabmeat (*Portunus pelagicus*) in 10% increments, taking into consideration their different commercial values (Gayo, Hale, & Blanchard, 2006). Qualitative chemometric analysis was performed on 400–2498 nm Vis/NIR spectra (previously subjected to different pretreatments to evaluate the effects on model performance), by means of a full-spectrum PCA and of a sequential-spectrum PCA. As a result, both the first derivative-pretreated full spectra and second derivative-pretreated sequential spectra, highlighted a trend of samples towards moving from the left part to the right part of the PCA score plot with increased adulteration levels, but authors identified the sequential approach using the 400–1700 nm second derivative spectra as being the most informative and, thus, the most suitable approach (Gayo et al., 2006).

Based on the fact that the past several years have seen a sharp rise in the interest towards the portability of the instruments, which may provide great flexibility especially in on-line, in-line, and at-line routine quality control, a study performed by O'Brien et al. (2013), explored the ability of a hand-held NIR spectrometer to give positive results of discrimination between high-value and low-value whole fish and fish fillets species (O'Brien, Hulse, Pfeifer, & Siesler, 2013). In particular, the objective was to discriminate between two different species of mullet (red mullet from mullet), cod (winter cod from cod), and trout (samlet from salmon trout). NIR spectra (906–1648 nm) obtained from skin (whole fish) and meat (fish fillets), were first pre-processed and then elaborated by PCA and SIMCA analysis. Successful PCA results were achieved only in separating the whole mullet samples, but the discrimination performances improved significantly also for mullet fillets after the application of the SIMCA analysis. PCA failed to discriminate both whole cod and cod fillets, but, here too, SIMCA predictions provided a correct assignment of the tested fish samples. Similar outcomes for samlet from salmon trout were achieved (Gayo et al., 2006). Thus, despite PCA investigation failed, SIMCA supervised analysis clearly

outlined the possibility to authenticate high quality fish species which are potentially substitutable with lower-quality alternatives. Still in the context of the use of hand-held and compact NIR devices, a broader attempt to distinguish fillets and patties of Atlantic cod (*Gadus morhua*) from those of haddock (*Melanogrammus aeglefinus*) was recently made (Grassi, Casiraghi, & Alamprese, 2018). Raw fillets and patties of the two fish species were scanned at 950–1650 nm (by the portable instrument) or at 800–2222 nm (by a benchtop instrument) and, after being pre-treated with SNV, MSC, or Savitzky-Golay smoothing (SG) coupled with first or second derivative, they were elaborated by means of supervised LDA and SIMCA analysis. Regardless of instrumentation used, the best LDA models were computed on MSC spectra of both fillets and patties, since the correct classification rate in external validation step reached 100% [52]. SIMCA class-modelling strategy allowed to obtain 100% correctly classified SNV, SG-first derivative, or SG-second derivative fillets spectra acquired by benchtop NIR, and 100% correctly classified MSC fillets spectra acquired with portable NIR (Grassi et al., 2018). As for patties, samples acquired by benchtop NIR and portable NIR were 100% correctly classified when spectra were subjected to SG-first derivative or SG-second derivative, and SNV or MSC, respectively. Basically, the worst SIMCA outcomes in prediction for patties and fillets were obtained for SG-second derivative spectra acquired with the portable instrument. Despite these results, no significant differences in the performances of the two instruments tested were found, thus confirming equivalent discrimination powers also in processed product.

Different species of freshwater fish of the Cyprinidae family, namely black carp (*Mylopharyngodon piceus*), grass carp (*Ctenopharyngodon idellus*), silver carp (*Hypophthalmichthys molitrix*), bighead carp (*Aristichthys nobilis*), common carp (*Cyprinus carpio*), crucian (*Carassius auratus*), and bream (*Parabramis pekinensis*), were also investigated by NIR spectroscopy (Lv, Xu, You, & Xiong, 2017). Fish samples were scanned in the 1000–1799 nm region, MSC pre-treated, and pre-reduced in dimensionality by different methods, including PCA, PLS, and fast Fourier transform (FFT). In this case, LDA models were built by using only nine pre-selected spectra wavelengths from the entire spectrum and results obtained showed a good prediction ability of

the adopted strategy: PCA-LDA and FFT-LDA models, in fact, showed 100% accuracy, specificity, sensitivity, and precision, even if most of the information was not taken into account by calculation (Lv et al., 2017).

Zhang et al. (2017) attempted to classify marine fish surimi by 1100–2500 nm NIR spectroscopy, according to the species by which products were composed, namely white croaker (*Argyrosomus argentatus*), hairtail (*Trichiurus haumela*), and red coat (*Nemipterus virgatus*) (Zhang et al., 2017). According to results obtained from PCA of the pre-processed spectra, the presence of a well-defined and separated cluster associated with red coat surimi species was observed, but the separation of the other two species of surimi samples was not clear (Zhang et al., 2017). However, as regards LDA results, 100% correct classification rate for external validation datasets after MSC pre-treatment was achieved, demonstrating once again the greater effectiveness of supervised analyses compared to unsupervised ones.

Species authenticity was also studied by comparing FT-NIR and FT-MIR spectra of red mullet and plaice fillets (higher-value species) to those of Atlantic mullet and flounder fillets (lower-value species) (Alamprese & Casiraghi, 2015). LDA and SIMCA analysis applied to differently pre-treated NIR and MIR spectra (800–2500 nm and 2500–14300 nm spectral ranges, respectively), allowed to clearly discriminate Atlantic mullet fillets from those of the more valuable red mullet. While LDA gave a 100% correct classification percentage in prediction (irrespective the spectroscopic technique considered), sensitivity and specificity higher than 70% and 100%, respectively, were calculated for FT-NIR spectra subjected to SIMCA analysis (Alamprese & Casiraghi, 2015). Poorer but acceptable results were obtained for flounder and plaice fillets discrimination: in this case, FT-IR spectroscopy showed the best discrimination power, with a prediction ability higher than 83% and a specificity of 100%.

The usefulness of NIR spectroscopy was explored to identify different fish species used to make fishmeal under industrial conditions. The 1100-2500 nm raw or second derivative NIR spectra of samples containing salmon, blue whiting, and other (i.e. mackerel or herring) fish species were elaborated by PCA, LDA, and DPLS (PLS-DA). Models developed allow to correctly classify, on average, more than 80% of the fish meal samples into

the three groups assigned according to the fish species (Cozzolino, Chree, Scaife, & Murray, 2005).

In contrast to the multiple applications of NIR spectroscopy, only one study explored the discrimination abilities of MIR spectroscopy (Sousa, Moreira, Saraiva, & de Almeida, 2018). This study coupled SG- and SNV-pre-treated MIR spectra (2500–20000 nm) with chemometrics (PCA) to specifically detect adulteration of Atlantic salmon (*Salmo salar*) mini-burgers with different percentage (from 0 to 100%, in steps of 10%) of Salmon trout (*Onconrhynchus mykiss*). The resulting 11 formulations of salmon burgers were grouped into 11 distinct clusters, even when the samples were stored for different periods of time before acquisition (Sousa et al., 2018).

Only two applications of Raman spectroscopy concerning fish species authentication are available. The aim of the first study was to discriminate 12 different fish fillets of different species by using pre-treated Raman spectra in the range 300–3400 cm−1 (about 3940–33333 nm) recorded by a Raman spectrometer equipped with a 532 nm laser exciting source (Rašković, Heinke, Rösch, & Popp, 2016). HCA analysis applied to the Raman spectra revealed the presence of three major clusters, one corresponding to fish from Salmonidae family (rainbow trout and chum salmon), one corresponding to various freshwater fish (zander, Nile perch, pangasius, and European seabass), and one corresponding to various saltwater fish (Atlantic herring, Atlantic pollock, Alaska pollock, Atlantic cod, blue grenadier, and yellowfin tuna). Within these large clusters, spectra were also grouped according to their species in sub-clusters, with a high degree of accuracy of the spectral classification on species level (95.8%) (Rašković et al., 2016). Similarly, PCA analysis performed on 5000–50000 nm Raman spectra (acquired by using a 785 nm laser exciting source) allowed to discriminate among horse mackerel (*Trachurus trachurus*), European anchovy (*Engraulis encrasicolus*), Bluefish (*Pomatamus saltatrix*), Atlantic salmon (*Salmo salar*), and flying gurnard (*Trigla lucerna*) samples. In this case, however, the study was less rapid and more elaborate since the spectral acquisition was performed on the previously extracted lipid fraction of fish (Velioğlu, Temiz, & Boyaci, 2015).

3.2. Production method and farming system misrepresentation

The differentiation of the production method of fish and seafood is another relevant aspect in certifying authenticity and traceability. During the last few years, the wild fish catches have been decreasing compared to the aquaculture production, thus supply of the market in farmed products has been growing very fast. From a compositional and organoleptic point of view, wild fish is quite different from aquaculture one, and this diversity is inevitably reflected on the different economic value of the two types of products (Bagr & Gab-Alla, 2007; Grigorakis, 2007; Grikorasis, Taylor, & Alexis, 2003).

By way of example, wild fish is usually characterised by higher levels of muscle protein, saturated and polyunsaturated fatty acids, while farmed fish by a higher content of total lipid and monounsaturated fatty acids (Fuentes, Fernández-Segovia, Serra, & Barat, 2010; Lenas, Chatziantoniou, Nathanailides, & Triantafillou, 2011). Consequently, illegal substitution of higher-value wild fish with lower-value farmed fish is not an uncommon occurrence. Additionally, aquaculture fish consist of a number of high-variable products (i.e. extensively, semi-intensively, or intensively farmed fish, as well as organic or conventional farmed fish), whose final characteristics, since influenced by the husbandry environment and, above all, by the diet, are slight and very difficult to identify. This the reason is why the authentication of the production method (wild or farmed, organic or conventional), but also of the farming system of the aquaculture products is of extreme importance from the standpoint of fraud prevention and transparency towards consumers.

3.2.1. Application of vibrational spectroscopy

Among various vibrational spectroscopic methods applied to differentiate production processes and farming systems of fish, NIR is once again the most widely used one. No application of UV or Raman spectroscopy, to the best of our knowledge, are currently available.

Ottavian et al. (2012) proposed a comparison between the classification performances of wild and farmed European sea bass obtained by three different NIR spectroscopic/chemometric approaches, and the classification

performances obtained using the only chemical and morphometric features (Ottavian et al., 2012). The use of 1100–2500 nm raw spectra, WPTER-pre-treated spectra (wavelet packet transform for efficient pattern recognition) or of some parameters predicted by building a regression-based model, were found to be equivalent in terms of predictability assessed by PLS-DA and no differences between classification obtained by these models and classification obtained by using only chemical and morphometric data was observed. Moreover, authors identified (by using the variable influence of projection indexes, VIP) the wavelengths related to the absorbance of fat, fatty acids, and water as most influential in differentiating the production process of the fish tested.

More recently, the systems behind the production of European sea bass, was also investigated by applying unsupervised PCA and supervised OPLS-DA to 1100–2500 nm NIR spectra (Ghidini et al., 2019). PCA built to SNV-SG-second derivative spectral data did not returned a clear separation of groups, mainly as a consequence of the fact that the intraclass variability among samples was higher than the among-class variability between samples. A correct classification rate of 100% for both wild and farmed sea bass was instead achieved by OPLS-DA, and, in this case, authors found VIP indexes related to proteins exerting a greater contribution to the variance between the two types of fish. A deeper insight into the different farming systems of aquaculture samples, moreover, showed the ability of NIR and OPLS-DA to authenticate 67%, 80%, 100% of extensively, semi-intensively, and intensively-reared subjects, respectively, thanks above all to the spectral bands associated with protein absorption (Ghidini et al., 2019). Concrete tank-cultured sea bass were also successfully discriminated from sea cage-cultured sea bass during storage, by means of Vis-NIR spectroscopy coupled with PLS-DA (C. Costa et al., 2011). The best performances (87% of correct classification) were observed for spectral measurements performed at 48 hours post-mortem (C. Costa et al., 2011). However, the greater contributions of the wavelengths to the PLS discrimination of samples analysed at 48 hours post mortem were different from those of samples analysed at 96 hours post, thus classification by farming system may have

been affected also by other unrelated factors, such as the well-known compositional changes occurring during shelf life.

Authentication by NIR and SIMCA analysis of European sea bass raised into extensive ponds, semi-intensive ponds, intensive tanks, and intensive sea-cages, was also performed both on fresh fillets and freeze-dried fillets (Xiccato, Trocino, Tulli, & Tibaldi, 2004). Authors found the freeze-drying of the samples giving the best classification outcomes. The same results were obtained when classifying fresh minced fillets and freeze-dried fillets of farmed European sea bass according to the semi-intensive conventional or the organic production system (Trocino et al., 2012). SIMCA classification based on second-derivative spectra (1100–2500 nm) of samples, in fact, generated good results when fitted on the freeze-dried fillet (65-75% of correct classification), and worse results when performed on fresh fillets (20-25% of correct classification) (Trocino et al., 2012).

All these results are particularly informative about problems posed by water when analysing high-moisture foods like fish. One of the main drawbacks of NIR spectroscopy is, in fact, the difficulty in separating relevant from useless information from spectra, in which peaks of water are predominant. These peaks, when included in chemometric calculations may hinder reliable features related to functional groups of molecules of interest and, thus, produce misleading result, especially when samples only slightly differ, such as in the case of fish reared under different conditions.

Following these principles, NIR spectroscopy was also used to directly authenticate freeze-dried rainbow trout fillets by rearing farm and, at the same time, to check whether NIR discriminating capability changed between raw and cooked freeze-dried fillets (Dalle Zotte et al., 2014). Rainbow trout samples came from three different aquaculture systems, varying in average well water temperatures, of which one consisted in indoor rearing at 11-14 °C, one in outdoor rearing at 9-11 °C, and one in outdoor rearing at 3-14 °C. Results for classification by farm (using SNV and second derivative 1100–2500 nm spectra of raw samples) showed approximately 97-100% of accuracy, with k-NN analysis giving the best overall statistical performances and PLS-DA the worst ones. As for cooked freeze-dried samples discrimination, the accuracy was approximately the

same as those obtained for raw samples (90-100% for LDA, QDA, k-NN and 80% for PLS-DA), highlighting that the cooking process did not alter the capabilities of the technique to discriminate sample by rearing farm (Dalle Zotte et al., 2014).

3.3. Geographical origin falsification

Proving the geographical origin authenticity of fish and seafood often involves the use of multi-disciplinary and cross-disciplinary approaches which take account of the environmental and genetic backgrounds affecting fish final characteristics (Abbas et al., 2018). Several published scientific research concerning the use of spectroscopic methods pointed out the usefulness in classification of fish and seafood according to country or FAO area of origin.

3.3.1. Application of vibrational spectroscopy

Unlike the other authentication issues discussed above, NIR spectroscopy has been less explored for fish geographical origin identification. The reason, probably, is the great difficulty experienced in modelling total variability of NIR spectra and uniquely steering it to provenance, since provenance is the sum of a huge amount of different intrinsic or extrinsic factors (genetic, growth pattern, feeding regime, muscular activity, water temperature and salinity, etc.).

A traceability model able to predict the geographical origin of Chinese tilapia fillets coming from four different Chinese provinces, was developed by NIR spectroscopy (Y. Liu et al., 2015). SIMCA analysis, performed on 1000–2500 nm spectra of the minced samples, allowed more than 80% of fillets from Guangdong, Hainan, and Fujian provinces and 75% of fillets from the Fujian province to be correctly and exclusively assigned to the corresponding area of origin. Several locations in northern China Sea and East China Sea, from which sea cucumber (*Apostichopus japonicus*) came from, were also identified by using NIR spectroscopy (Guo et al., 2018). In this case, authors found pre-treated (SNV or MSC, and second derivative) 1000–1800 nm spectra to give the best performance in PCA, since 100% correct classification rate were obtained both in the internal calibration

model and in the external validation model. Similarly, 100% of sea cucumber analysed by means of diffuse reflectance MIR spectroscopy (fingerprint 5800–16600 nm region) combined with SIMCA, were discriminated by the Chinese geographical region of provenance (Wu et al., 2010).

The last available application of NIR spectroscopy concerned the authentication of European sea bass according to Western, Central, or Eastern Mediterranean Sea provenances, by using OPLS-DA as classification technique (Ghidini et al., 2019). Results showed an overall discrimination performance of 89% according to these geographical origins, with 100% of Eastern, 88% of Central and 85% of Western Mediterranean Sea samples being correctly classified. The VIP index analysis, moreover, allowed to identify lipid-associated bands as the most influential variables on samples geographic discrimination.

3.4. Discrimination between fresh and frozen/thawed fish and seafood

Fish is commonly processed by freezing in order to be preserved from deterioration. Frozen fish, however, is usually characterised by much lower quality and commercial value compared to fresh fish. Therefore, fraudulent practices consisting in the substitution of fresh with frozen/thawed products are not uncommon events (Verrez-Bagnis et al., 2018). Considering that labelling of fish must state if the fish is fresh, frozen, or previously frozen (or refreshed), discriminating fresh from frozen/thawed products is one of the most important authenticity issue. Anyway, the differentiation between fresh and frozen/thawed products is hampered by difficulties in detecting those tiny physical and chemical variations occurring during freeze storage, which, moreover, do not cause any perceptible organoleptic change (Verrez-Bagnis et al., 2018; Tokur, Ozkütük, Atici, Ozyurt, & Ozyurt, 2006). Therefore, the rapid confirmation of fish freshness by spectroscopy has been widely studied during the last few years and several published research is currently available.

3.4.1. Application of fluorescence and vibrational spectroscopy

Front-face fluorescence spectroscopy is one of the earliest spectroscopic technique historically applied to differentiate fresh from frozen/thawed fish.

It has been demonstrated that typical changes in fluorescence spectra of aromatic amino acids, nucleic acids, and nicotinamide adenine dinucleotide (NADH) occur during storage, as a consequence of several reactions involving free amino acids and carbonyl compounds of reducing sugars, formaldehyde (produced from trimethylamine oxide), and malondialdehyde (produced from oxidation of fish lipids during storage). Therefore, changes in fluorescence of fish samples may be considered as fingerprints for fresh and aged fish fillets identification (Karoui, Thomas, & Dufour, 2006). The fluorescence emission spectra of tryptophan (305–400 nm) recorded directly on whiting fillets and elaborated by factorial discriminant analysis (FDA) led to correct classification rates of 62.5% and 70.8% in the calibration and validation set, respectively. NADH fluorescence spectra (360–570 nm), indeed, were found to have a higher potential to differentiate fresh from frozen/thawed products as they allowed to achieve 100% of correct discrimination for both calibration and validation set. More recently, the same authors confirmed the success of a similar methodology in authenticating freshness of sea bass samples (Karoui, Hassoun, & Ethuin, 2017). Fluorescence emission spectra at 340 and 380 nm, elaborated by FDA, led to 94.87% of total correct classification rate. Additionally, the elaboration of NADH fluorescence spectra by Fisher's linear discriminant analysis, was stated as a reliable method to rapidly discriminate fresh and frozen/thawed large yellow croaker fillets, since 100% of total correct classification rate was achieved (Gao et al., 2016).

More applications of IR spectroscopy are reported in the published literature. Uddin and Okazaki (2004) used NIR reflectance spectroscopy on dry extract of horse mackerel specimens to evaluate freshness (Uddin & Okazaki, 2004). Both PCA (using 1100–2500 nm spectra) and SIMCA analysis (using only three selected wavelengths which were strongly related to protein content) successfully discriminated 100% of fresh and frozen/thawed samples. Thereafter, the same authors performed further investigations on fresh and frozen/thawed red sea bream by using Vis/NIR spectroscopy in the 400–1100 nm region (Uddin et al., 2009). In this case, raw spectra were used to build an LDA model, by which 100% classification accuracy in prediction was reached. PLS-DA of SG-smoothed spectra (670–

1100 nm) of shrimps subjected to different treatments (including ice, water, and brine at various salt concentrations), also led to 100% of fresh and frozen/thawed samples to be authenticated (Zhang & Cheng, 2013).

Another study was directed to compare classification ability of Vis-NIR (380–1080 nm) and NIR (1100–2500) spectroscopy in authenticating fresh and frozen/thawed swordfish and, through the application of PLS-DA, it was found that, in this case, Vis-NIR spectra gave better results in the external validation (≥ 96.7% of correctly classified samples) (Fasolato et al., 2012). Despite worse outcomes were obtained by using the only NIR region, the technique, combined with SVMs, also allowed to authenticate 93% of fresh and 83% of frozen/thawed sole (*Solea vulgaris*) samples (Fasolato et. al, 2008). Again, high accuracy (90%) and sensitivity (80%) in prediction were observed for the discrimination of fresh and frozen/thawed tuna sample by Vis-NIR spectral analysis (350–2500 nm) combined with PLS-DA (Reis et al., 2017), while better and more homogenous SIMCA prediction results were obtained when using MIR (2500–14300 nm) instead of NIR (800–2500 nm) regions for the discrimination between fresh and previously frozen Atlantic mullet fillets (Ottavian, Fasolato, Facco, & Barolo, 2013). Ottavian et *al*. (2013) proposed an interesting three-step-approach based only on NIR spectra and latent variable modelling techniques to develop a species-independent classifier able to simultaneously discriminate between fresh and frozen/thawed fish and, remarkably, overall classification accuracy of the method ranged between 80% and 91%, based on the strategy adopted and the instrument used (Ottavian et al., 2013). By contrast, the only MIR region was found to be useful for determining whether whiting fish fillets have been frozen/thawed: when FDA was applied to the 3300–3570 nm MIR subregion (usually related to fatty acids absorption), 87.5% of sample spectra in the validation set was correctly identified (Karoui et al., 2007).

Finally, one single application of Raman spectroscopy to the authentication of fresh fish is up to now available (Velioğlu et al., 2015). Lipid fraction of fish from several species (horse mackerel, European anchovy, bluefish, Atlantic salmon, red mullet, and flying gurnard) was extracted from three samples batches (fresh samples, once frozen/towed samples, and twice frozen/thawed samples), and then collected by a Raman spectrometer along

the 5000–50000 nm spectral range and using a 785 nm laser exciting source. Chemometric analysis, performed by PCA, allowed to identify three different cluster in the score plot, each corresponding to one of the three batches of fish investigated (Velioğlu et al., 2015).

3.4.2. Application of hyperspectral imaging spectroscopy

Discrimination between fresh and frozen/thawed cod fillet was studied by Vis-NIR/HSI, using both a handheld interactance probe and an imaging spectrometer (for automatic online analysis at typical industrial speeds) (Sivertsen, Kimiya, & Heia, 2011). Spectra resulting from the two instruments were pre-treated (SNV and second derivative) and statistically analysed by applying the Rosenblatt's perceptron linear classifier to the first and third principal component of the imaging data. Results showed that fresh cod fillets can be completely separated from fresh/thawed cod fillets using only few wavelengths in the Vis region, mainly related to the oxidation of haemoglobin and myoglobin which occur during freezing/thawing (Zhu, Zhang, He, Liu, & Sun, 2013). Similarly, hyperspectral data from Vis-NIR/HSI (380–1030 nm) combined with least square-SVMs, returned an average correct classification rate of 91.67% for fresh and frozen/thawed halibut fillets (Zhu et al., 2013).

4. Conclusions and limitations to overcome

Recent increases of complexity and competitiveness of the fishery and seafood sectors, have resulted in the presence, on the international market, of a huge variety of fresh and processed products, but, at the same time, have meant that the risk of fraud deriving from substitution among look-alike products is now exponentially higher than it was even a few years ago. Thus, ensuring the truthfulness of fish and seafood claims concerning their quality and origin, has become an exceptionally important topic, firstly with a view to enable consumers to make informed decisions.

The overview presented in this review clearly highlights the effective support provided by analytical approaches based on spectroscopy and multivariate data analysis to the evaluation and monitoring of fish and seafood products authenticity. Fluorescence, vibrational, NMR and HSI

spectroscopic applications have been discussed, with an accent on the trends toward their use for several authentication purposes. In this connection, IR spectroscopy has been the most exploited technique, especially in studies concerning species and fresh for frozen/thawed products substitutions. NMR, instead, showed many applications in the field on the production method, farming system, and geographical origin identification. By contrast, Raman and HIS have provided very encouraging results in some fish authentication fields, but their overall potential has so far been largely ignored.

In addition, the attractiveness of spectroscopy and chemometrics is evidenced by not only by the large literature provided in the present review, but also by several other applications covering a wide range of food and foodstuffs: fruits and vegetables, honey, wine, edible oils and fats, cereal and cereal-based products, milk and dairy products (Yang, Irudayaraj, & Paradkar, 2005; Dos Santos, Páscoa, & Lopes, 2017; Maione, Barbosa, & Barbosa, 2019; Cozzolino, 2014; Kamal & Karoui, 2015) have been successfully investigated and authenticated by means of spectroscopy.

Nevertheless, some critical reflections should be made about the problems related to the use of spectroscopy and chemometrics, which still have not been overcome. In accordance to what has been already reported and to our opinion, the research papers analysed were found to be highly variable to each other in terms of analytical set-up, e.g. sample pre-processing, spectral ranges, spectra pre-treatments, resolutions, number of samples tested, and statistical elaboration. This variability, as easily understood from Section 3, is further worsened by the fact that only a few of the works analysed reported in-depth statistical outputs and, where present, they were not comparable to each other.

A critical and objective evaluation of these works is severely hampered also by lack, in certain cases, of comprehensive data with regards to the validation of the results. Alongside with the internal cross-validation, the external validation of the qualitative chemometric model is, in our opinion, a crucial points in assessing the overall goodness of the classifiers and avoid misleading interpretations. The last aspect which should be emphasised is that a detailed description of the characteristics of sample dataset was not

often reported and the lack of standardisation of external factors (e.g. storage times and conditions), may have interfered with spectral analysis, possibly affecting the robustness of the model. In this scenario, a recommendation for the future works is to consider the intrinsically natural variability of the fish products (as well as those of all other foodstuffs), and to organise the sampling in such a way that as much of the expected variability of samples is collected during the calibration stage. That way, the of robustness of the models could make the way to the spread of applications also in the industrial sector.

As final remark, no technique should be universally regarded as the optimal solution. However, the possibility of using UV, IR, Raman, and NMR spectroscopies with no distinction for food authentication purposes is still an obstacle to overcome, and therefore, in accordance to our experience, untargeted NIR spectroscopy represent the most versatile option thanks to its high sensitivity to organic molecules of food, cost-effectiveness and ease of use. Additionally, the use of NIR spectroscopy with supervised chemometric method able to separate relevant from non-relevant spectral variation, like OPLS-DA, should be encouraged since the interpretability of results is enhanced.

.

References

Abbas, O., Zadravec, M., Baeten, V., Mikuš, T., Lešić, T., Vulić, A., Prpić, J., Jemeršić, L., & Pleadin, J. (2018). Analytical methods used for the authentication of food of animal origin. *Food Chemistry, 246,* 6-17. https://doi.org/10.1016/j.foodchem.2017.11.007

Adewale, P., Mba, O., Dumont, M. J., Ngadi, M., & Cocciardi, R. (2014). Determination of the iodine value and the free fatty acid content of waste animal fat blends using FT-NIR. *Vibrational Spectroscopy*, 72, 72–78. https://doi.org/10.1016/j.vibspec.2014.02.016

Ahmad, A. R. (2002). Machine Learning Using Support Vector Machines 2 . Artificial Neural Network. *Artificial Intelligence*, 1–8.

Alamprese, C., & Casiraghi, E. (2015). Application of FT-NIR and FT-IR spectroscopy to fish fillet authentication. *LWT - Food Science and Technology*, *63*(1), 720–725. https://doi.org/10.1016/j.lwt.2015.03.021

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007, July 27). Supervised pattern recognition in food analysis. *Journal of Chromatography A*. https://doi.org/10.1016/j.chroma.2007.05.024

Blanco, M., & Villarroya, I. (2002). NIR spectroscopy: a rapid-response analytical tool. *TrAC Trends in Analytical Chemistry, 21* (4). https://doi.org/10.1016/S0165-9936(02)00404-1

Boyaci, I. H., Temiz, H. T., Geniş, H. E., Soykut, E. A., Yazgan, N. N., Güven, B., Uysal, R.S., Bozkurt, A.G., Ilaslan, K., Torun, O., & Şeker, F.C.D. (2015). Dispersive and FT-Raman spectroscopic methods in food analysis. *RSC Advances*, *5*, *56606*. https://doi.org/10.1039/c4ra12463d

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., & Trygg, J. (2007). OPLS Discriminant Analysis, Combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics, 20*, 341-351. https://doi:10.1002/cem.1006

Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science and Technology*, *18*(2), 72–83. https://doi.org/10.1016/j.tifs.2006.09.003

Cheng, J. H., Dai, Q., Sun, D. W., Zeng, X. A., Liu, D., & Pu, H.B. (2013). Applications of non-destructive spectroscopic techniques for fish quality and safety evaluation and inspection. *Trends in Food Science and Technology, 34,* 18-31. https://doi.org/10.1016/j.tifs.2013.08.005

Costa, C., D'Andrea, S., Russo, R., Antonucci, F., Pallottino, F., & Menesatti, P. (2011). Application of non-invasive techniques to differentiate sea bass (Dicentrarchus labrax, L. 1758) quality cultured under different conditions. *Aquaculture International*, *19*(4), 765–778. https://doi.org/10.1007/s10499-010-9393-9

Council regulation (EC) No 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. *Official Journal of the European Communities. 2002, 31* (01/02/2002), 1-24.

Council regulation (EC) No 1005/2008 of 29 September 2008 establishing a Community system to prevent, deter and eliminate illegal, unreported and unregulated fishing, amending Regulations (EEC) No 2847/93, (EC) No 1936/2001 and (EC) No 601/2004 and repealing Regulations (EC) No 1093/94 and (EC) No 1447/1999. *Official Journal of the European Union. L. 2008, 286*, 1-32.

Council regulation (EC) No 1224/2009 of 20 November 2009 establishing a Community control system for ensuring compliance with the rules of the common fisheries policy, amending Regulations (EC) No 847/96, (EC) No 2371/2002, (EC) No 811/2004, (EC) No 768/2005, (EC) No 2115/2005, (EC) No 2166/2005, (EC) No 388/2006, (EC) No 509/2007, (EC) No 676/2007, (EC) No 1098/2007, (EC) No 1300/2008, (EC) No 1342/2008 and repealing Regulations (EEC) No 2847/93, (EC) No 1627/94 and (EC) No 1966/2006. *Official Journal of the European Union. L. 2009, 343*, 1-50.

Cozzolino, D., Chree, A., Scaife, J. R., & Murray, I. (2005). Usefulness of Near-infrared reflectance (NIR) spectroscopy and chemometrics to discriminate fishmeal batches made with different fish species. *Journal of Agricultural and Food Chemistry, 53,* 4459-4463. https://doi.org/10.1021/jf050303i

Cozzolino, D., & Murray, I. (2012). A review on the application of infrared technologies to determine and monitor composition and other quality characteristics in raw fish, fish products, and seafood. *Applied Spectroscopy Reviews, 47*(3), 207-218. https://doi.org/10.1080/05704928.2011.639106

Cozzolino, D. (2014). An overview of the use of infrared spectroscopy and chemometrics in authenticity and traceability of cereals. *Food Research International, 60,* 262-265. . https://doi: 10.1016/j.foodres.2013.08.034.

Dalle Zotte, A., Ottavian, M., Concollato, A., Serva, L., Martelli, R., & Parisi, G. (2014). Authentication of raw and cooked freeze-dried rainbow trout (Oncorhynchus mykiss) by means of near infrared spectroscopy and data fusion. *Food Research International*, *60*, 180–188. https://doi.org/10.1016/j.foodres.2013.10.033

Dolah, F. M. Van. (2000). Marine Algal Toxins: Origins, Health Effects, and Their Increased Occurrence. *Environmental Health Perspectives*, *108* (October 1999), 133. https://doi.org/10.2307/3454638

Dos Santos, C. A. T., Páscoa, R. N. M. J., & Lopes, J. A. (2017). A review on the application of vibrational spectroscopy in the wine industry: From soil to bottle. *TrAC - Trends in Analytical Chemistry, 88,* 100-118. https://doi.org/10.1016/j.trac.2016.12.012

Engel, J., Gerretzen, J., Szymańska E., Jansen, J. J., Downey, G., Blanchet, L., & Buydens, L.M.C. (2013). Breaking with trends in pre-processing? *TrAC Trends in Analytical Chemistry, 50,* 96–106. . https://doi:10.1016/j.trac.2013.04.015.

Esslinger, S., Riedl, J., & Fauhl-Hassek, C. (2014). Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Research International, 60,* 189-204. https://doi.org/10.1016/j.foodres.2013.10.015

Fallah, A. A., Saei-Dehkordi, S. S., Nematollahi, A., & Jafari, T. (2011). Comparative study of heavy metal and trace element accumulation in edible tissues of farmed and wild rainbow trout (Oncorhynchus mykiss) using ICP-OES

technique. *Microchemical Journal*, *98*(2), 275–279. https://doi.org/10.1016/j.microc.2011.02.007

FAO (2018). The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals, Volume 35, Rome, Italy. ISBN 978-92-5106-029-2.

FAO, (2018). Overview of food fraud in the fisheries sector. Fisheries and Aquaculture Circular No. 1165, Rome, Italy. ISBN 978-92-5-130402-0.

Fasolato, L., Manfrin, A., Corrain, C., Perezzani, A., Arcangeli, G., Rosteghin, M., & Serva, L. (2008). Assessment of quality-parameters and authentication in sole (Solea vulgaris) by NIRS (Near infrared reflectance spectroscopy). *Industrie Alimentari*, *47*, 355-361.

Fasolato, L., Balzan, S., Riovanto, R., Berzaghi, P., Mirisola, M., Ferlito, J. C., … Novelli, E. (2012). Comparison of visible and near-infrared reflectance spectroscopy to authenticate fresh and frozen-thawed swordfish (xiphias gladius L). *Journal of Aquatic Food Product Technology*, *21*(5), 493–507. https://doi.org/10.1080/10498850.2011.615103

Feng, Y. Z. & Sun, D. W (2012). Application of hyperspectral imaging in food safety inspection and control: a review. *Critical reviews in food science and nutrition*, *52*, 1039-1058. https://doi.org/10.1080/10408398.2011.651542

Fiorino, G. M., Garino, C., Arlorio, M., Logrieco, A. F., Losito, I., & Monaci, L. (2018). Overview on Untargeted Methods to Combat Food Frauds: A Focus on Fishery Products. *Journal of Food Quality*. https://doi.org/10.1155/2018/1581746

Fuentes, A., Fernández-Segovia, I., Serra, J.A., & Barat, J.M. (2010). Comparison of wild and cultured sea bass (Dicentrarchus labrax) quality. *Food Chemistry, 119,* 1514-1518. https://doi:10.1016/j.foodchem.2009.09.036.

Gabr, H.R., & Gab-Alla, A.A.F.A. (2007). Comparison of biochemical composition and organoleptic properties between wild and cultured finfish. *Journal of Fisheries and Aquatic Science, 2*, 77-81. https://doi:10.3923/jfas.2007.77.81.

Gayo, J., Hale, S. A., & Blanchard, S. M. (2006). Quantitative analysis and detection of adulteration in crab meat using visible and near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, *54*(4), 1130–1136. https://doi.org/10.1021/jf051636i

Ghidini, S., Varrà, M. O., Dall'Asta, C., Badiani, A., Ianieri, A., & Zanardi, E. (2019). Rapid authentication of European sea bass (Dicentrarchus labrax L.) according to production method, farming system, and geographical origin by near infrared spectroscopy coupled with chemometrics. *Food Chemistry*, *280*, 321-327. https://doi.org/10.1016/j.foodchem.2018.12.075

Grassi, S., Casiraghi, E., & Alamprese, C. (2018). Handheld NIR device: A non-targeted approach to assess authenticity of fish fillets and patties. *Food Chemistry, 243,* 382-388. https://doi.org/10.1016/j.foodchem.2017.09.145

Grigorakis, K., Taylor, K.D.A., & Alexis, M.N. (2003). Organoleptic and volatile aroma compounds comparison of wild and cultured gilthead sea bream (Sparus aurata): sensory differences and possible chemical basis. *Aquaculture, 225*, 109-119. https://doi.org/10.1016/S0044-8486(03)00283-7

Grigorakis, K. (2007). Compositional and organoleptic quality of farmed and wild gilthead sea bream (Sparus aurata) and sea bass (Dicentrarchus labrax) and factors affecting it: A review. *Aquaculture, 272*, 55-75. https://doi.org/10.1016/j.aquaculture.2007.04.062.

Gao, Y., Tang, H., Ou, C., Li, Y., Wu, C., & Cao, J. (2016). Differentiation between fresh and frozen-thawed large yellow croaker based on front-face fluorescence spectroscopy technique. *Transactions of the Chinese Society of Agricultural Engineering, 32*, 279–285. https://doi.org/10.11975/j.issn.1002-6819.2016.16.038.

Guo, X., Cai, R., Wang, S., Tang, B., Li, Y., & Zhao, W. (2018). Non-destructive geographical traceability of sea cucumber (Apostichopus japonicus) using near infrared spectroscopy combined with chemometric methods. *Royal Society Open Science*. https://doi.org/10.1098/rsos.170714

Hatzakis, E. (2019). Nuclear Magnetic Resonance (NMR) Spectroscopy in Food Science: A Comprehensive Review. *Comprehensive Reviews in Food Science and Food Safety*, *18*, 189-220. https://doi.org/10.1111/1541-4337.12408.

He, H. J., Wu, D., & Sun, D. W. (2015). Nondestructive Spectroscopic and Imaging Techniques for Quality Evaluation and Assessment of Fish and Fish Products. *Critical Reviews in Food Science and Nutrition*. https://doi.org/10.1080/10408398.2012.746638

Kamal, M. & Karoui, R. (2015). Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review. *Trends in Food Science & Technology, 46*, 27-48. https://doi.org/10.1016/j.tifs.2015.07.007.

Karoui, R., Hassoun, A., & Ethuin, P. (2017). Front face fluorescence spectroscopy enables rapid differentiation of fresh and frozen-thawed sea bass (Dicentrarchus labrax) fillets. *Journal of Food Engineering*, *202*, 89–98. https://doi.org/10.1016/j.jfoodeng.2017.01.018

Karoui, R., Thomas, E., & Dufour, E. (2006). Utilisation of a rapid technique based on front-face fluorescence spectroscopy for differentiating between fresh and frozen–thawed fish fillets. *Food Research International, 39*, 349-355. https://doi.org/10.1016/j.foodres.2005.08.007.

Karoui, R., Lefur, B., Grondin, C., Thomas, E., Demeulemester, C., De Baerdemaeker, J., & Guillard, A. S. (2007). Mid-infrared spectroscopy as a new tool for the evaluation of fish freshness. *International Journal of Food Science and Technology*, *42*(1), 57–64. https://doi.org/10.1111/j.1365-2621.2006.01208.x

Kelly, B. C., Ikonomou, M. G., Higgs, D. A., Oakes, J., & Dubetz, C. (2011). Flesh residue concentrations of organochlorine pesticides in farmed and wild salmon from British Columbia, Canada. *Environmental Toxicology and Chemistry*, *30*(11), 2456–2464. https://doi.org/10.1002/etc.662

Lenas, D., Chatziantoniou, S., Nathanailides, C., & Triantafillou, D. (2011). Comparison of wild and farmed sea bass (Dicentrarchus labrax L) lipid quality. Procedia Food Science, 1, 1139–1145. https://doi.org/10.1016/j.profoo.2011.09.170

Li, L., Boyd, C. E., & Sun, Z. (2016). Authentication of fishery and aquaculture products by multi-element and stable isotope analysis. *Food Chemistry*, *194*,

1238–1244. https://doi.org/10.1016/j.foodchem.2015.08.123

Liu, Y., Ma, D. hong, Wang, X. chang, Liu, L. ping, Fan, Y. xia, & Cao, J. xuan. (2015). Prediction of chemical composition and geographical origin traceability of Chinese export tilapia fillets products by near infrared reflectance spectroscopy. *LWT - Food Science and Technology*, *60*(2), 1214–1218. https://doi.org/10.1016/j.lwt.2014.09.009

Lohumi, S., Lee, S., Lee, H., & Cho, B. K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science and Technology, 46,* 85-98. https://doi.org/10.1016/j.tifs.2015.08.003

Lv, H., Xu, W., You, J., & Xiong, S. (2017). Classification of freshwater fish species by linear discriminant analysis based on near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy, 25*(1), 54-62. https://doi.org/10.1177/0967033516678801

Maione, C., Barbosa Jr. F., & Barbosa, R.M. (2019). Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review. *Computers and Electronics in Agriculture*, *157*, 436-446. https://doi:10.1016/j.compag.2019.01.020.

Manley, M., & Baeten, V. (2008). Spectroscopic technique: near infrared (NIR) spectroscopy. Modern techniques for food authentication. Sun, D. W, Eds.; Academic Press, San Francisco, CA, USA, 51-102. https://doi:10.1016/B978-0-12-814264-6.00003-7.

O'Brien, N., Hulse, C. A., Pfeifer, F., & Siesler, H. W. (2013). Near infrared spectroscopic authentication of seafood. *Journal of Near Infrared Spectroscopy*. https://doi.org/10.1255/jnirs.1063

Oliveri, P., & Simonetti, R. (2016). Chemometrics for food authenticity applications. Advances in food authenticity testing. Woodhead Publishing, 2016, 701–728. https://doi:10.1016/B978-0-08-100220-9.00025-4.

Okocha, R. C., Olatoye, I. O., & Adedeji, O. B. (2018, August 8). Food safety impacts of antimicrobial use and their residues in aquaculture. *Public Health Reviews*. BioMed Central Ltd. https://doi.org/10.1186/s40985-018-0099-2

Ottavian, M., Facco, P., Fasolato, L., Novelli, E., Mirisola, M., Perini, M., & Barolo, M. (2012). Use of near-infrared spectroscopy for fast fraud detection in seafood: Application to the authentication of wild European sea bass (Dicentrarchus labrax). *Journal of Agricultural and Food Chemistry,* 60, 639-648. https://doi.org/10.1021/jf203385e

Ottavian, M., Fasolato, L., Facco, P., & Barolo, M. (2013). Foodstuff authentication from spectral data: Toward a species-independent discrimination between fresh and frozen-thawed fish samples. *Journal of Food Engineering*, *119*(4), 765–775. https://doi.org/10.1016/j.jfoodeng.2013.07.005

Pereira, A.C., Reis, M.S., Saraiva, P.M., & Marques, J.C. (2011). Madeira wine ageing prediction based on different analytical techniques: UV–vis, GC-MS, HPLC-DAD. *Chemometrics and Intelligent Laboratory Systems, 105*, 43-55. https://doi:10.1016/j.chemolab.2010.10.009

Penner, M. H. (2017). Basic Principles of Spectroscopy. *Food analysis*, Food Science Text Series, 2nd ed.; Nielsen S.S., Eds., Springer, Cham. 79-88; ISBN 978-3-319-

45776-5

Picò, Y. (2012). Chemical analysis of food: Techniques and applications. Picò, Y., Eds; Academic Press, San Francisco, CA, USA. ISBN 978-0-1238-4862-8.

Rašković, B., Heinke, R., Rösch, P., & Popp, J. (2016). The Potential of Raman Spectroscopy for the Classification of Fish Fillets. *Food Analytical Methods*. https://doi.org/10.1007/s12161-015-0312-6

Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC. *Official Journal of European Union Legislation, L304*, 18-63.

Regulation (EU) No 1379/2013 of the European Parliament and of the Council of 11 December 2013 on the common organisation of the markets in fishery and aquaculture products, amending Council Regulations (EC) No 1184/2006 and (EC) No 1224/2009 and repealing Council Regulation (EC) No 104/2000. *Official Journal of European Union* Lesislation*, L354*, 12-14.

Rehbein, H., & Oehlenschläger, J. (2009). Fishery products: quality, safety and authenticity. Rehbein, H., Oehlenschläger, J., Eds., John Wiley & Sons, Chichester, West Sussex, UK. ISBN 978-1-4051-4162-8.

Reis, M. M., Martínez, E., Saitua, E., Rodríguez, R., Pérez, I., & Olabarrieta, I. (2017). Non-invasive differentiation between fresh and frozen/thawed tuna fillets using near infrared spectroscopy (Vis-NIRS). *LWT - Food Science and Technology*, *78*, 129–137. https://doi.org/10.1016/j.lwt.2016.12.014

Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends in Analytical Chemistry*. https://doi.org/10.1016/j.trac.2009.07.007

Rinnan, Å. (2014). Pre-processing in vibrational spectroscopy–when, why and how. *Analytical Methods, 6*, 7124-7129. https://doi:10.1039/C3AY42270D.

Roberts, J., Power, A., Chapman, J., Chandra, S., & Cozzolino, D. (2018). A short update on the advantages, applications and limitations of hyperspectral and chemical imaging in food authentication. *Applied Sciences, 8,* 505. https://doi:10.3390/app8040505.

Rodriguez-Saona, L. E., Giusti, M. M., & Shotts, M (2016). Advances in infrared spectroscopy for food authenticity testing. Downey, G., Eds., Woodhead Publishing, Duxford, UK, 2016. ISBN: 978-0-08-100220-9.

Sacchi, R., & Paolillo, L. (2007). NMR for Food Quality and Traceability. In *Advances in Food Diagnostics*, 101–117. Blackwell Publishing. https://doi.org/10.1002/9780470277805.ch5

Schrieber, A. (2008). Introduction to Food Authentication. *Modern Techniques for Food Authentication*. Sun, D. W, Eds., Academic Press, San Francisco, CA, USA. https://doi:10.1016/B978-0-12-814264-6.00001-3

Sivertsen, A. H., Kimiya, T., & Heia, K. (2011). Automatic freshness assessment of cod (Gadus morhua) fillets by Vis/Nir spectroscopy. *Journal of Food Engineering*, *103*(3), 317–323. https://doi.org/10.1016/j.jfoodeng.2010.10.030

Sobolev, A.P., Circi, S., & Mannina, L. (2016). Advances in Nuclear Magnetic Resonance Spectroscopy for Food Authenticity Testing. *Advances in Food Authenticity Testing*, Woodhead Publishing, 147-170. https://doi:10.1016/B978-0-08-100220-9.00006-0.

Sousa, N., Moreira, M., Saraiva, C., & de Almeida, J. (2018). Applying Fourier Transform Mid Infrared Spectroscopy to Detect the Adulteration of Salmo salar with Oncorhynchus mykiss. *Foods, 7*(4), 55. https://doi.org/10.3390/foods7040055

Strasburg, G. M. & Ludescher, R. D. (1995). Theory and applications of fluorescence spectroscopy in food research. *Trends in Food Science and Technology*, *6,* 69-75. https://doi:10.1016/S0924-2244(00)88966-9

Stuart, B. H. (2004). Infrared Spectroscopy: Fundamentals and Applications. Stuart, B. H., Eds., Jhon Wiley & Sons Ltd, Chichester, WS, UK. ISBN 0-470-85427-8.

Tokur, B., Ozkütük, S., Atici, E., Ozyurt, G., & Ozyurt, C. E. (2006). Chemical and sensory quality changes of fish fingers, made from mirror carp (Cyprinus carpio L., 1758), during frozen storage (− 18 C*). Food Chemistry, 99*, 335-341. https://doi:10.1016/j.foodchem.2005.07.044.

Trocino, A., Xiccato, G., Majolini, D., Tazzoli, M., Bertotto, D., Pascoli, F., & Palazzi, R. (2012). Assessing the quality of organic and conventionally-farmed European sea bass (Dicentrarchus labrax). *Food Chemistry, 131,* 427-433. https://doi.org/10.1016/j.foodchem.2011.08.082

Uddin, M. & Okazaki, E. (2004). Classification of fresh and frozen-thawed fish by near-infrared spectroscopy. *Journal of Food Science, 69*, C665-C668, https://doi.org/10.1111/j.1750-3841.2004.tb18015.x.

Uddin, M., & Okazaki, E. (2006). Applications of Vibrational Spectroscopy to the Analysis of Fish and Other Aquatic Food Products. Handbook of Vibrational Spectroscopy. John Wiley & Sons Ltd, Chichester. https://doi.org/10.1002/9780470027325.s8955

Uddin, M., Okazaki, E., Turza, S., Yumiko, Y., Tanaka, M., & Fukuda, Y. (2009). Non-destructive Visible/NIR Spectroscopy for Differentiation of Fresh and Frozen-thawed Fish. *Journal of Food Science*, *70*(8), c506–c510. https://doi.org/10.1111/j.1365-2621.2005.tb11509.x

Velioglu, H. M., Temiz, H. T., & Boyaci, I. H. (2015). Differentiation of fresh and frozen-thawed fish samples using Raman spectroscopy coupled with chemometric analysis. *Food Chemistry, 172, 283-290*. https://doi.org/10.1016/j.foodchem.2014.09.073

Verrez-Bagnis, V., Sotelo, C. G., Mendes, R., Silva, H., Kappel, K., & Schröder, U. (n.d.). Methods for Seafood Authenticity Testing in Europe. https://doi.org/10.1007/978-3-319-54528-8_69-1

Voncina, D. B. (2009). Chemometrics in analytical chemistry. *Nova Biotechnologica, 9*, 211-216, https://doi.org/10.5740/jaoacint.

Wu, Z., Tao, L., Zhang, P., Li, P., Zhu, Q., Tian, Y., & Yang, T. (2010). Diffuse reflectance mid-infrared Fourier transform spectroscopy (DRIFTS) for rapid identification of dried sea cucumber products from different geographical areas. *Vibrational Spectroscopy, 53*, 222-226. https://doi.org/10.1016/j.vibspec.2010.03.008.

Wu, D. & Sun, D. W (2013). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part I: Fundamentals. *Innovative Food Science & Emerging Technologies, 19*, 1-14. https://doi:10.1016/j.ifset.2013.04.014

Xiccato, G., Trocino, A., Tulli, F., & Tibaldi, E. (2004). Prediction of chemical composition and origin identification of european sea bass (Dicentrarchus labrax L.) by near infrared reflectance spectroscopy (NIRS). *Food Chemistry, 86*, 275-281. https://doi.org/10.1016/j.foodchem.2003.09.026

Xu, J. L., Riccioli, C., & Sun, D. W. (2015). An Overview on Nondestructive Spectroscopic Techniques for Lipid and Lipid Oxidation Analysis in Fish and Fish Products. *Comprehensive Reviews in Food Science and Food Safety, 14*. https://doi.org/10.1111/1541-4337.12138

Yang, H., Irudayaraj, J., & Paradkar, M.M. (2005). Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. *Food Chemistry, 93*, 25-32. https://doi:10.1016/j.foodchem.2004.08.039.

Ziegel, E. R. (2014). A User-Friendly Guide to Multivariate Calibration and Classification. *Technometrics, 46*, 108-111. https://doi.org/10.1198/004017004000000167.

Zhang, A., & Cheng, F. (2013). Identification of fresh shrimp and frozen-thawed shrimp by Vis/NIR spectroscopy. Proceedings of the 2nd International Conference on Nutrition and Food Sciences IPCBEE, Singapore, 53, IACSIT Press https://doi.org/10.7763/IPCBEE.

Zhang, X. Y., Hu, W., Teng, J., Peng, H. H., Gan, J. H., Wang, X. C. & Liu, Y. (2017). Rapid recognition of marine fish surimi by one-step discriminant analysis based on near-infrared diffuse reflectance spectroscopy. *International Journal of Food Properties, 20*(12), 2932-2943. https://doi.org/10.1080/10942912.2016.1261153

Zheng, J., & He, L. (2014). Surface-Enhanced Raman Spectroscopy for the Chemical Analysis of Food. *Comprehensive Reviews in Food Science and Food Safety*, *13*(3), 317–328. https://doi.org/10.1111/1541-4337.12062

Zhu, F., Zhang, D., He, Y., Liu, F., & Sun, D. W. (2013). Application of Visible and Near Infrared Hyperspectral Imaging to Differentiate Between Fresh and Frozen-Thawed Fish Fillets. *Food and Bioprocess Technology*, *6*(10), 2931–2937. https://doi.org/10.1007/s11947-012-0825-6

CHAPTER 2

Research Article

Rapid Authentication of European Sea Bass (Dicentrarchus labrax L.) according to Production Method, Farming System, and Geographical Origin by Near Infrared Spectroscopy Coupled with Chemometrics

Sergio Ghidini[a], Maria Olga Varrà[a]*, Chiara Dall'Asta[a], Anna Badiani[b], Adriana Ianieri[a], and Emanuela Zanardi[a]

aDepartment of Food and Drug, University of Parma, Strada del Taglio 10, 43126 Parma, Italy.

[b]Department of Veterinary Medical Science, University of Bologna, Via Tolara di Sopra 50, 40064 - Ozzano dell'Emilia, Bologna, Italy

*Correspondence: mariaolga.varra@studenti.unipr.it; Tel.: +39-0521-902-761

Adapted from

Ghidini, S., Varrà, M. O., Dall'Asta, C., Badiani, A., Ianieri, A., & Zanardi, E. (2019). Rapid authentication of European sea bass (Dicentrarchus labrax L.) according to production method, farming system, and geographical origin by near infrared spectroscopy coupled with chemometrics. Food chemistry, 280, 321-327.

Abstract: Chemometric analysis of near-infrared spectroscopy (NIRS) data was applied to investigate the possibility to rapidly authenticate European sea bass (*Dicentrarchus labrax* L.) according to production method (wild or farmed), rearing system (extensive, semi-intensive or intensive), and geographical origin (Western, Central or Eastern Mediterranean Sea). NIR spectra from 1100 to 2500 nm were subjected to an exploratory principal component analysis (PCA) followed by orthogonal partial last square-discriminant analysis (OPLS-DA) to develop classifiers able to distinguish samples according to the various conditions under study. Models provided a correct classification rate of 100% for both wild and farmed sea bass, and of 67%, 80%, 100% for extensively, semi-intensively, and intensively-reared subjects, respectively. As for geographical provenance, 100% of Eastern, 88% of Central and 85% of Western Mediterranean Sea samples were correctly discriminated. The successful results obtained confirmed suitability of chemometric analysis applied to NIRS data for fast authentication of European sea bass origin.

Abbreviations: Area under the Receiver Operating Characteristic Curves, AUROC; Central Mediterranean Sea, CM; cross-validation, CV; analysis of variance testing of cross-validation predictive residuals, CV-ANOVA; Eastern Mediterranean Sea, EM; extensive system, Ext; farmed, F; intensive system, Int; near infrared spectroscopy, NIRS; orthogonal partial last square-discriminant analysis, OPLS-DA; principal component analysis, PCA; principal component, PC; root mean square error from cross-validation, RMSECV; root mean square error of estimation, RMSEE; root mean square error of prediction, RMSEP; second derivative, SD; Savitzky-Golay smoothing, SG; semi-intensive system, SI; standard normal variate, SNV; variable influence on projection, VIP; wild, W; ; Western Mediterranean Sea, WM.

1. Introduction

The expansion and globalization of the fisheries and aquaculture sector, together with the greater public awareness regarding food quality, have led to a growing interest in several issues related to fish authenticity and compliance with food legislation. According to European Regulation (EU) n. 1379/2013, fishery and aquaculture products must be labelled with the commercial designation, proper scientific name of the species, production method (e.g. caught, farmed), fishing gear (e.g. hook, trap, trawl), and catch or production area. Errors in label information about fish origin and production process are increasing in frequency to such an extent that today fish is the second category of food most vulnerable to fraud (FAO, 2018).

European sea bass (*Dicentrarchus labrax* L.) is one of the most economically important fish species in the whole Mediterranean area: Turkey and Greece have recently become the largest producing countries, while Italy, Spain and France continue to remain the world's leading importers (EUMORFA, 2017). Although wild fishing is well-established in Europe, the vast bulk of sea bass production comes from aquaculture systems, where fish are bred at different stocking densities and feed inputs (FAO 2005-2018, 2005). In extensive culture system, proper lagoon sites are generally involved, and no supplemental nutritional input is provided. In semi-intensive culture system, sea bass are usually farmed in small ponds or tank (allowing higher stocking density) and the natural feeding is artificially supplemented. Intensive systems, instead, are based on external complete diet nutrient input and frequently involve floating or submersible cages placed in coastal or open sea waters where fish are raised at high stocking densities (Moretti et al., 1999; FAO 2005-2018, 2005). By allowing a higher production yield to be achieved, intensive systems represent today the most frequent form of sea bass farming in the Mediterranean basin (EUMORFA, 2017).

Both geographical provenance and production method can influence overall characteristics of fish, resulting in a large variability among look-alike products of different origins, whose discriminating properties are often difficult to measure. Several analytical techniques are traditionally used to assess fish authenticity and traceability, but, although being well-

established, the need for faster, easier, and cheaper methods is growing. Untargeted fingerprinting approaches based on near infrared spectroscopy (NIRS) meet all these characteristics, as they provide multiple chemical and physical information to qualitatively/quantitatively characterise complex food matrix (Uddin & Okazaki, 2010). The shape of NIR spectra obtained from fish samples is the result of several interactions between NIR radiation and water, organic molecules like protein, carbohydrate and fat, and low-concentration constituents such as vitamins and minerals (Cozzolino, 2015). As matter of fact, NIR absorption bands in the wavelength range 780–2500 nm are associated with multiple overtones and combinations of fundamental vibrations of chemical bonds between light atoms, especially C-H, N-H, O-H, C=O, and S-H (Blanco & Villarroya, 2002), which are widely present in fish matrices.

Low sensitivity related to the high signal-to-noise ratio, and a high spectral complexity due to the overlapping peaks, are the most prominent disadvantages of NIRS applied to compositionally complex samples such as fish (Abbas et al., 2018). Multivariate data analysis such as principal component analysis (PCA) or orthogonal partial last square-discriminant analysis (OPLS-DA) techniques help to overcome these obstacles by separating useful information from noise, uncover hidden correlations, improve spectral features and interpretability, and provide a visual approach for data analysis (Granato et al., 2018). The subsequent outcomes are the identification of patterns within the results and their classification based on the relationship between the data (McGrath et al., 2018).

Several authors have reported that appropriate statistical treatments of NIR spectra allow to successfully discriminate between different fish species in fishmeal (Cozzolino et al., 2005), fresh from frozen-thawed cod fillet and Atlantic salmon (Sivertsen et al.), and tilapia fillets according to their geographical origin (Liu et al., 2015). Research about sea bass authentication by using NIR spectroscopy has focused prevalently on discrimination of wild from farmed specimens (Ottavian et al., 2012) according to different intensity of farming system used (Xiccato et al., 2004; Majolini et al., 2010), and samples according to production techniques and practices used (organically- vs. conventionally-produced sea bass, Trocino et al., 2012;

concrete tanks- vs. sea cage-cultured sea bass, Costa et al., 2011). Anyway, analysis of the literature shows that the technique has not been explored to classify sea bass samples according to geographical area of provenance.

Therefore, the aim of the present work was to explore the possibility of using NIRS combined with chemometric analysis as a rapid and non-destructive tool to discriminate and classify European sea bass according to production method (wild vs. farmed), farming system (intensive vs. semi-intensive vs. extensive), and geographical origin (Western vs. Central vs. Eastern Mediterranean Sea).

2. Materials and Methods

2.1. Set of sea bass samples

A total of 144 European sea bass specimens (*Dicentrarchus labrax* L.) collected during 2 periods in 2012 (spring-summer, n = 77; autumn-winter, n = 77) were considered in this study. The dataset included wild (W; n = 34) and farmed samples (F; n = 110) respectively caught in fishing areas or bred in fish farms located in the Mediterranean basin. Fifty of them came from Western Mediterranean Sea (WM; FAO fishing subareas 37.1.2 and 37.1.3), sixty-four were from Central Mediterranean Sea (CM; FAO fishing subareas 37.2.1 and 37.2.2) and thirty were from Eastern Mediterranean Sea (EM; FAO fishing subareas 37.3.1). Based on information about the fish breeding system (in terms of stocking density) declared by farmers, aquaculture sea bass was identified as intensively, semi-intensively and extensively reared samples. Intensively-reared sea bass (Int; n = 80) were raised in submersible or floating cages, located in various open sea areas, at a stocking density up to 30 kg/m3; semi-intensively reared ones (SI; n = 20) were reared in earthen tank at a stocking density up to 1 kg/m3; extensively reared ones (Ext; n = 10) were reared in coastal lagoons (valliculture) at a stocking density up to 0.0025 kg/m3.

2.2. Samples preparation and NIR analysis

After removing the skin and the viscera, right and left fillets were separated from each sea bass. The two fillets of the same sample were first ground and homogenized by using a blender (Multiquark System ZK 100, Braun,

Kronbergim Taunus, Germany) and then divided in representative sub-samples that were individually packed and stored at -20°C up to the time of analysis. Before NIRS measurement, each frozen sample was thawed overnight at 4 °C and all the spectra acquired the following day. Samples were stabilized at room temperature for 30 min prior to the collection of each spectrum.

NIR analysis was performed by using a NIRFlex® N-500 (Büchi Labortechnik AG, Flawil, Switzerland) at a wavelength range of 1100 to 2500 nm with a spectral resolution of 1 nm, by placing an aliquot of sample inside a 35 mm diameter round quartz cuvette. Spectral data were recorded in reflectance (R) units and then converted in absorbance (A) units through the logarithm of reciprocal of R (1/R). Each spectrum acquisition was the result of 32 single scans and 4 spectra were acquired by rotating 90° the cuvette consecutively; these operations were repeated twice, so the final number of spectra for each sample was 8. A single mean spectrum was obtained by averaging the 8 individual spectra of each sample. The final data matrix used in subsequent calculations, consisted of 1401 variables (i.e. wavelengths) and 144 observations (i.e. sea bass samples).

2.3. Chemometric analysis and validation of the results

Raw data matrix was imported into SIMCA-P v.14.1 software (Umetrics, Umeå, Sweden) to perform multivariate analysis. Spectra were scaled, mean-centered and mathematically pre-treated by standard normal variate (SNV) to correct light scattering effects, followed by second derivation (SD) and a 15-points Savitzky-Golay smoothing (SG), to reduce baseline shift and improve the spectral properties.

Two different chemometric approaches were followed: an unsupervised PCA and a supervised OPLS-DA.

PCA is a projection method able to reduce the correlated variables of a matrix of independent variables (X) into a smaller number of new uncorrelated latent variables, known as principal components (PCs). PCs contain as much systematic variation as possible of the original and most of the variation is explained by the first two PCs variables (Naes, Isakson, Fearn, & Davies, 2002). PCA was preliminarily performed both on raw and

pre-processed spectra of the whole dataset, to explore their characteristics and detect clustering or trends among samples. The presence of outliers was also checked during this operation, by evaluating Hotelling's T2 range values (5% level of significance).

Subsequently, OPLS-DA was employed with the aim to build discrimination models able to distinguish samples according to their production method, stocking density and geographical origin. OPLS-DA is a discriminant and classification method based on OPLS regression that separates all systematic variation in an X-matrix into a related (predictive) and a non-related (orthogonal) part to a set of dependent dummy binary variables (Y) that describe the class membership of each observation in the X matrix. (Trygg & Wold, 2002).

Prior to developing OPLS classifiers, we divided the total number of samples in the ratio of 75:25 to create a training set (n = 108) and a test set (n = 36) respectively; the same proportion was respected to split spectra into the test set on the basis of each class membership, in order to ensure uniformity and a large experimental variation. While the training set was employed to build calibration models, the independent test set was reserved to externally validate them. All the full-spectrum PCA and OPLS-DA models computed were internally validated by a 7-fold cross-validation (CV) and their quality assessed by the statistical parameters R^2X_{cum} which represented the sum of predictive plus orthogonal variation in X matrix explained (goodness of fit), and Q^2_{cum} (goodness of prediction estimated by CV). For OPLS-DA models, R^2Y_{cum} was also evaluated (total sum of variation explained in Y matrix). The most influential absorption bands in the OPLS classification were identified by means of VIP (Variable Influence on Projection) parameter for predictive components.

The reliability of the classifiers was further evaluated using CV-ANOVA (analysis of variance testing of cross-validation predictive residuals), taking a p-value < 0.05 as an indication of a good model (Eriksson, Trygg, & Wold, 2008), RMSECV (Root Mean Square Error from cross-validation), and RMSEE (Root Mean Square Error of Estimation). An external-test set validation was finally performed. The resulting percentage of correctly classified observations and the RMSEP (Root Mean Square Error of

Prediction) were used to assess overall classification performances. Multi-class ROC (Receiver Operating Characteristic) analysis was further adopted to calculate the values of Area under the ROC Curves (AUROC). ROC curves display the classifier's true positive rate (sensitivity) versus the false positive rate (1 – specificity), as a function of the threshold value. AUROC vary from 1 for an ideal predictor to 0.5 for a random predictor (Fawcett, 2006).

3. Results

3.1. Spectral features interpretation

Raw NIR spectral data of whole sea bass dataset under investigation were characterised by considerable baseline shifts due to light-scattering effects, and broad overlapping absorption bands, which hindered spectral analysis. SNV, SD and SG treatments were therefore performed to correct baseline shifts and improve the separation of the peaks (Figure 1). The positions of the negative peaks in the second–derivative spectra match the positions of peaks in the original spectrum (Rinnan, Berg, & Engelsen, 2009). According to literature, NIR absorptions bands in SNV-SD-SG spectra have been assigned to various functional groups in water, proteins, and lipids, which represent the main constituents of fish flesh. NIR region with wavelengths from 1100 to 1300 is associated with the second overtone of the C-H stretching vibration of different chemical groups (–CH2, –CH3, –CH=CH–), while the combination of C-H stretching, and C-H deformation vibrations falls within the range 1300-1420 nm (weak NIR peaks around 1360 and 1395 nm). The range 1420-1600 nm is related to N-H stretching (first overtone) and O-H stretching (first overtone), where absorption at 1435 nm refers to O–H bonds in water (Osborne, 2000; Khodabux et al., 2007). The intensive bands at wavelengths between 1600 and 1800 nm depend on C–H and CH2 vibrations related to fatty acids content; the prominent peak at 1710 nm is due to the first overtone of C-H stretch (Aenugu et al., 2011). Wavelength region from 1800 to 2200 nm is characterised by O-H and N-H bonds combinations. Peaks around 2058 and 2174 nm are related to the absorption of the amide group (amide I and II) and have high correlation with proteins, whose content is also revealed at 1990 and 2180 nm (Osborne, 2000;

Cozzolino et al., 2005; Aenugu et al., 2011). The NIR region 2200-2500 nm is related to C-H combination vibrations of fatty acids: peaks at 2280, 2335 and 2352, correspond to C-H stretch/C-H deformation combination (Osborne, 2000; Liu, Zeng, & Sun, 2013). As it can be observed, all spectra show approximately the same profile; nevertheless, some clear variations of the peak intensity are evident in the NIR regions related to fat and protein. These differences could be explained by the influence exerted by environment, feeding regime, water quality, growth pattern, competition, and muscular activity, on fish flesh composition (Lenas et al., 2011; Trocino et al., 2012; Arechavala-Lopez et al., 2013). Wild sea bass, in fact, exhibit higher moisture, muscle protein content, and saturated and polyunsaturated fatty acids. By contrast, farmed specimens are characterised by higher contents of total lipid and monounsaturated fatty acids (Fuentes et al., 2010; Lenas et al., 2011). A more-in depth knowledge about spectral differences among the samples was achieved by means of multivariate data analysis.

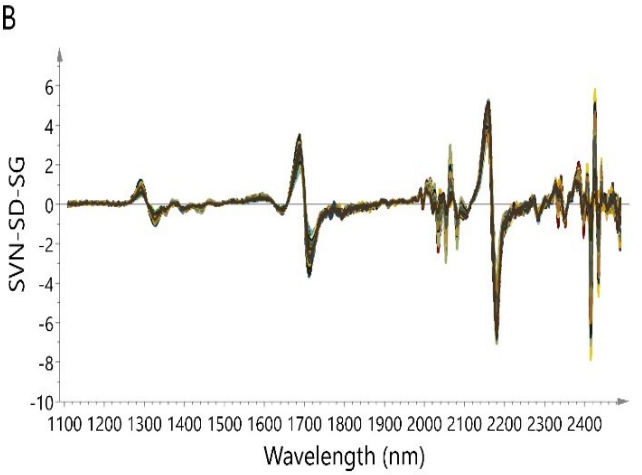


Figure 1. Pre-processed (SNV + SD) NIR spectra of European sea bass samples.

3.2. Preliminary PCA to explore natural clustering of the sea bass samples

Two PCA models were initially built with the raw and SNV-SD-SG spectral data of the 144 sea bass samples, with the aim to look at the distribution and detect any anomalies among samples. A total of 4 and 17 PCs were extracted

from raw data (R^2X = 0.998, Q^2 = 0.998) and pre-processed data- based PCA models (R^2X = 0.971, Q^2 = 0.936), respectively.

Score scatter plots of the first two PCs (PC1 and PC2) derived from the pre-processed spectra are reported in Figure 2, where in Figure 2A the observations were highlighted by production method, in Figure 2B by farming system, and in Figure 2C by provenance. In the three plot display modes, no well-defined groups were identified. As it can be observed, a slight separation between W and F samples can be observed, even if some observations overlapped (Figure 2A). Regarding farming system, almost all W/Ext reared samples had positive scores on the PC1; W ones were mainly represented by negative score on the PC2, while Ext reared ones by positive scores on the same component. By contrast, SI/Int reared sea bass were mainly located in the left part of the score plot, corresponding to the negative side of the PC1, and both the positive and negative side of the PC2. No differentiation was detected within the SI and Int systems (Figure 2B). Particular grouping behaviours did not emerge even according to the provenance (Figure 2C). Samples originating from WM, CM, and EM Sea were strongly overlapping each other, indicating that a high within-group variability existed in them. Only EM samples clustered as negative scores on the PC1. Some samples from each model were outside the 95% confidence ellipse, but they were not strong outliers according to Hotelling's T^2 test.

Loading plot reported in Figure 3, highlighted the contribution of each wavelength to PC1 and PC2 derived from PCA analysis of the SNV-SD-SG spectra. The most influential loadings on PC1 were observed around 2020-2100 nm and 2380-2440 nm, corresponding to the absorption of proteins and fatty acids, respectively (see *Section 3.1.*). Regarding PC2, the most important loadings were found around 1660-1730 nm (absorption of fat and fatty acids) and 2140-2220 (absorption of proteins). Since PC1 and PC2, together, accounted for 86% of the total variability in the data, it can be said that proteins and fatty acids contributed most to this faint separation of samples, even if overlapping samples complicated the clear identification of the contribution of the individual variables to each class.

The 17 PCs extracted from SNV-SD-SG spectral data were not found to be informative enough to achieve a clear group separation, mainly as a

consequence of higher intraclass variability greater than the among-class variability (Barker & Rayens, 2003) Since unsupervised classification analysis failed to produce satisfactory results, the ability of supervised OPLS-DA to discriminate sea bass was investigated.

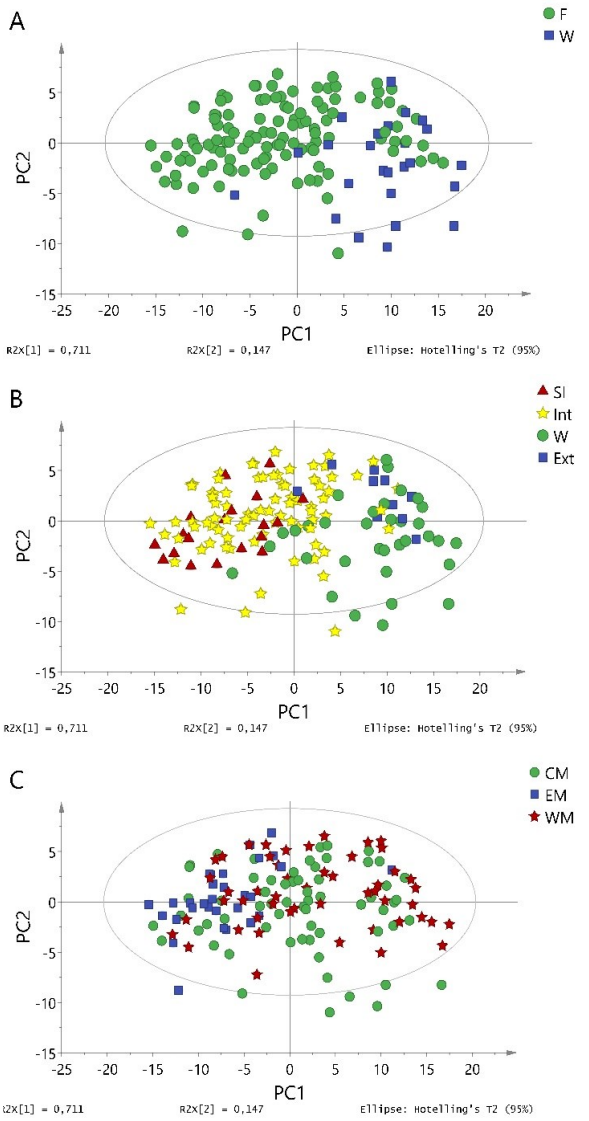


Figure 2. PCA score plot of the first two principal components (PC1 and PC2), showing sea bass clustering by production method (**A**), farming system (**B**), and geographical provenance (**C**). The ellipse identifies the 95% confidence interval for Hotelling's T^2.

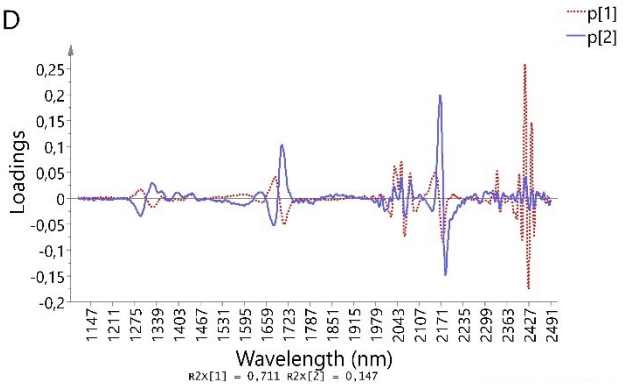


Figure 3. Loading plot of the first two loading vectors (p1 and p2), showing the influence of NIR wavelengths on the PCA model (p1, dotted line; p2, solid line).

3.3. Supervised OPLS-DA for discrimination and classification of the sea bass samples

Three different supervised OPLS-DA models were built on SNV-SD-SG training set (n=108) in order to pursue the following objectives: (1) classification of samples by production method (W vs. F); (2) classification of samples by farming system (W vs. Ext vs. SI vs. I); (3) classification of samples by geographical provenance (WM vs. CM vs. EM).

A good interclass variability can be observed along the t[1], i.e. the first predictive component, of the score plot for W vs. F calibration model (Figure 4A), where the negative scores corresponded to F samples and the positive scores to W samples. No tight clusterisation along the to[1] (first orthogonal component) was achieved, thus indicating a high intraclass variability. Four distinct clusters were also identified along the two first predictive components t[1] and t[2] of the score plot for samples modelled according to stocking density (Figure 4B). W, Ext, and SI/I subjects were well discriminated along the t[1]. Variability between SI and I subjects was collected by the t[2]. Score plot for geographical origins (Figure 4C) showed a grouping of EM samples that distributed prevalently as negative scores on the t[1]. The best separation between WM and CM samples was provided by the t[2]. Hotelling's T^2 test, applied to the samples from each model outside the 95% confidence ellipse, indicated the absence of strong outliers.

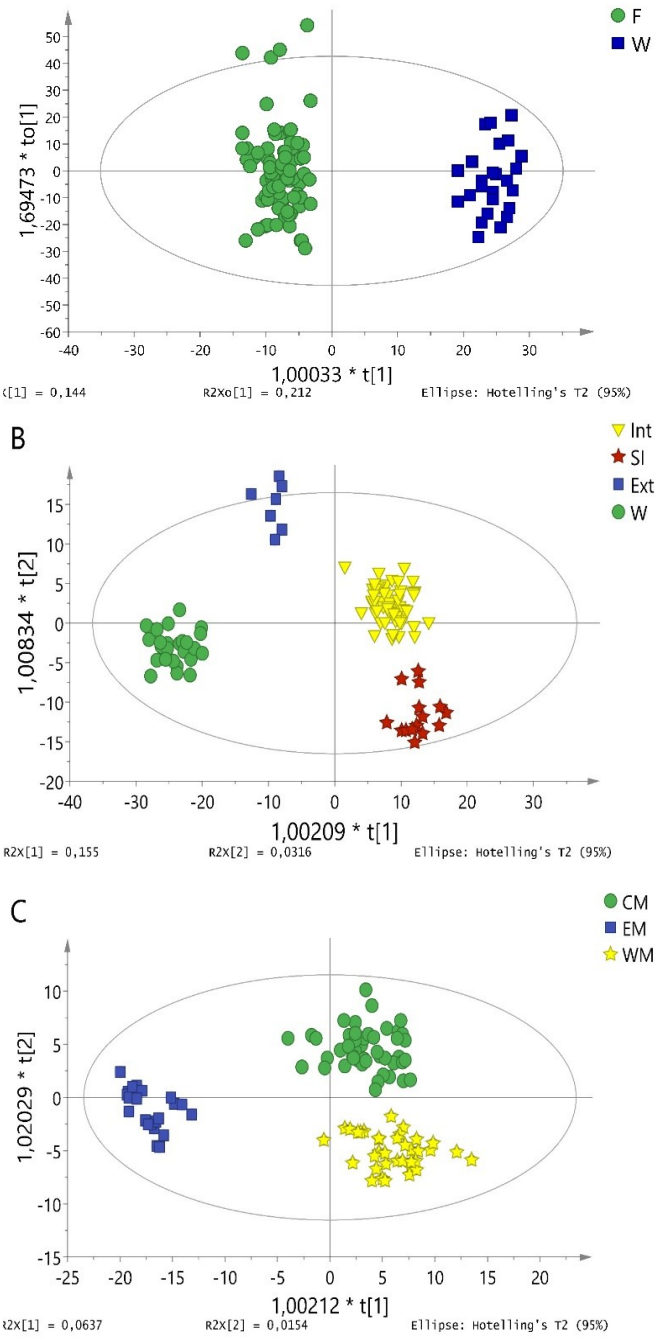


Figure 4. OPLS-DA score plot for sea bass discrimination based on production method (**A**), farming system (**B**), and geographical provenance (**C**). The ellipse identifies the 95% confidence interval for Hotelling's T^2.

The VIP scores for predictive components were then employed to visualize the spectral regions that mostly contributed to discriminate samples in each calibration. In general, a VIP threshold value greater than one, is considered to be relevant (Galindo-Prieto, B., Eriksson, L., & Trygg, J, 2014). In the present work, too many variables where characterised by VIP values > 1, so we established a cut-off value of 1.5 for significant absorption bands models (see Figure S1, Supplementary material).

As a result, we found an influence of 1630-1800 nm, 1980-2200 nm, and 2320-2450 nm wavelength regions on the W vs. F models related to fatty acids and peptides absorption. In particular, the main absorption peaks found at 1990 nm (VIP = 1.90) and 2058 (VIP = 1.96), have a strong correlation with the amide group (see *Section 3.1.*), indicating a great contribution of the proteins to the variance between the wild and farmed sea bass. Similarly, the most relevant variables influencing discrimination by rearing system were found to be in the region corresponding to N-H and O-H absorption (1190-1230 nm). The peaks around 1199 nm (VIP = 1.90) and 1223 nm (VIP = 1.83) showed the highest contributions, but many other variables also exhibited a minor impact on the model. These outcomes are not strongly supported by what is reported by the available literature. According to what has been suggested by most authors (Alasalvar et al., 2002; Orban et al., 2003; Fuentes et al., 2010; Ottavian et al., 2012), the total lipid amount, as well as those of saturated and mono-unsaturated fatty acids, are higher in farmed than wild fish, mainly as a consequence of both high stocking density and intensive feeding from fish feeds rich in terrestrial vegetable oils (Lenas et al., 2011). Flesh protein content, on the contrary, is less influenced by external feeding since it seems to be more related to intrinsic factors such as the fish species, variety, and size (Periago et al., 2005).

However, alongside this, a great contribution to fish flesh composition is also made by the muscular activity. According to Bell et al. (2007), the lower nitrogen content in farmed bass compared with wild specimens, cannot be simply explained by the dilution effect of higher lipid content, but also reflects the higher protein content of wild fish due to greater muscle mass. This final assumption may explain the results we obtained, thus suggesting

that production method and farming system probably affected protein content more than lipid content of our samples.

With regard to the geographical location of sea bass, it seemed to have a major influence on the lipid composition of fish; absorption of N-H and O-H bonds (1190-1230 nm) slightly influenced the group separation by provenance, but the lipid-associated bands around 1620-1720 nm were predominant (VIP > 2.0). The intrinsic variability of fatty acid composition of fish, in fact, depends on fishing ground, being strongly influenced by environmental conditions and geographical effects, such as water temperature, salinity and habitat (Saito, Ishihara, & Murase, 1997).

3.3.1. Internal Validation of the OPLS-DA models

One of the main drawbacks related to the use of OPLS is its tendency to overfit data, mainly because of the discrepancy between the large number of variables and the low number of observations. This means that the classifiers may often not be able to predict the correct class membership of new samples, despite an excellent discrimination observed in the training set-related scores (Brereton, 2006; Westerhuis et al., 2008). To avoid misleading results, we performed an internal 7-fold CV on SNV-SD-SG training set used to compute calibration OPLS-DA models. Resulting statistical metrics are presented in Table 1. The calibration models were fitted with 1 to 3 predictive components and 5 to 7 orthogonal components (A), that captured 60-63% of the total variation in the X-matrix (R^2X_{cum}). In particular, OPLS-DA for different farming systems, led to the higher value of $R2X_{cum}$, even if about 46% of the variation was orthogonal in X (o-R^2X_{cum}), and thus indicative of a high within-class variance among samples. Values related to the predictive variation of X (p-R^2X_{cum}), were rather low for all the models, varying from 0.079 to 0.144; anyway, the predictive variation contained in the spectra described 88-97% of the class membership information (R^2Y_{cum}), thus indicating a good class separation in each model. The predictability values, given by cross-validated Q^2_{cum} parameter, ranged from 0.416 to 0.793, among which the model based on the production method showed the best performance and the model based on the geographical provenance the worst one. For W vs. F model, the lowest accuracy-associated errors, RMSECV and

RMSEE, were calculated, thus highlighting a better performance of this model compared to the others. The statistical significance of the overall OPLS-DA classifiers was confirmed at 95% confidence level (p CV-ANOVA < 0.05).

Table 1. Summary of OPLS-DA calibration statistics calculated by CV.

Parameter	OPLS-DA model		
	W vs. F	**W vs. Ext. vs. SI vs. Int**	**WM vs. CM vs. EM**
A	1 + 5	3 + 7	2 + 6
R^2X_{cum}	0.601	0.634	0.608
p-R^2X_{cum}	0.144	0.196	0.079
o-R^2X_{cum}	0.457	0.438	0.529
R^2Y_{cum}	0.967	0.907	0.885
Q^2_{cum}	0.793	0.562	0.416
RMSECV	0.194	0.239	0.352
RMSEE	0.080	0.112	0.162
p CV-ANOVA	2.65e-27	2.14e-15	1.79e-08

A = number of extracted predictive and orthogonal components. R^2X_{cum} = cumulative variation of the X block. p-R^2X_{cum} = cumulative predictive fraction of the variation of the X block. o-R^2X_{cum} = cumulative orthogonal fraction of the variation of the X block. R^2Y_{cum} = cumulative variation of the Y block explained. Q^2_{cum} = cumulative variation of the Y block predicted. RMSECV = root mean square error of cross-validation. RMSEE = root mean square error of estimation. p CV-ANOVA = p-value of cross validation-analysis of variance (significant level of 0.05).

3.3.2. External Validation of the OPLS-DA models

The test set (n = 36) was further used to perform a strict external validation, complementary to internal CV, with the aim to confirm the accuracy of the OPLS-DA classifiers to practically predict class labels of new sea bass samples.

An overview of the classification outcomes is presented in Table 2. As it can be observed, none of the W samples were misclassified in the F group (and vice versa). Globally, this model showed the best overall classification rate (100%), low RMSEP values of 0.246, and excellent AUROC values of 1. Correct class predictions for 100%, 67%, 80% and 100% of W, Ext, SI, and Int samples, were respectively obtained. Notably, the rearing systems antipodal to each other (W and Int) were perfectly allocated in their own class, while

the intermediate ones (Ext and SI) suffered from some misclassifications, due to their similarity. Ext class showed the lowest RMSEP value in the model and an AUROC value of 1, despite presenting the poorest classification rate (nearly of 67%). By contrast, Int class presented a 100% classification rate, but the lowest RMSEP and AUROC values. Geographical classification led to 100% of EM samples to be assigned to the correct class, while two samples of the CM group and two samples of the WM group were misclassified. Model's overall precision of nearly 89% was considered satisfactory, even if the CM group showed the weakest statistical performances compared to the other groups.

Table 2. External validation metrics for the independent sea bass test set.

		Single class			Overall model	
OPLS-DA model	**Class**	**Classification rate**	**RMSEP**	**AUROC**	**Total classification rate**	***p*-value**
Production method	W	100.00% (8/8)	0.246	1	100%	5.5e-06
	F	100.00% (28/28)		1		
Farming system	W	100.00% (8/8)	0.244	1	94.44%	1.3e-12
	Ext	66.67% (2/3)		1		
	SI	80.00% (4/5)		0.991		
	Int	100.00% (20/20)		0.947		
Provenance	WM	84.62% (11/13)	0.308	0.993	88.89%	9.5e-11
	CM	87.50% (14/16)		0.875		
	EM	100.00% (7/7)		1		

RMSEP = root mean square error of prediction; AUROC: area under Receiver Operator Characteristic curve; *p* value = assessed by Fisher's Exact Test (significant level of 0.05).

4. Conclusions

This work demonstrates the feasibility of a simple and rapid authentication of European sea bass by using NIRS combined with chemometric analysis. Supervised and unsupervised classification models were built to recognize the different origins of the samples, according to some of the European Regulation (EU) n. 1379/2013 requirements. Even if a preliminary classification performed by PCA analysis was not satisfying enough, results

of OPLS-DA showed a clear discrimination of 100% sample by production method, 94% by farming system, and 89% by geographical provenance. To interpret discriminant information, the VIP index was additionally used. Spectral bands associated with protein absorption were found to be significant towards a sample discrimination based on production method/farming system, while those associated to lipid absorption had a major contribution to sample geographical discrimination.

Models were internally and externally validated by a 7-fold CV and an independent test set, respectively, and statistical outputs confirmed the practical ability of prediction and the absence of overfitting. Overall, this study has shown the potential approach to enable the authenticity assessment of sea bass; it should be mentioned, however, that fingerprints database must be continuously expanded to obtain robust classification model.

The main advantages of the proposed analytical strategy over the traditional methodologies of food analysis are the rapidity and the ease of use in routine operations on a large-scale, suitable to implement efficient control systems.

SUPPLEMENTARY MATERIAL

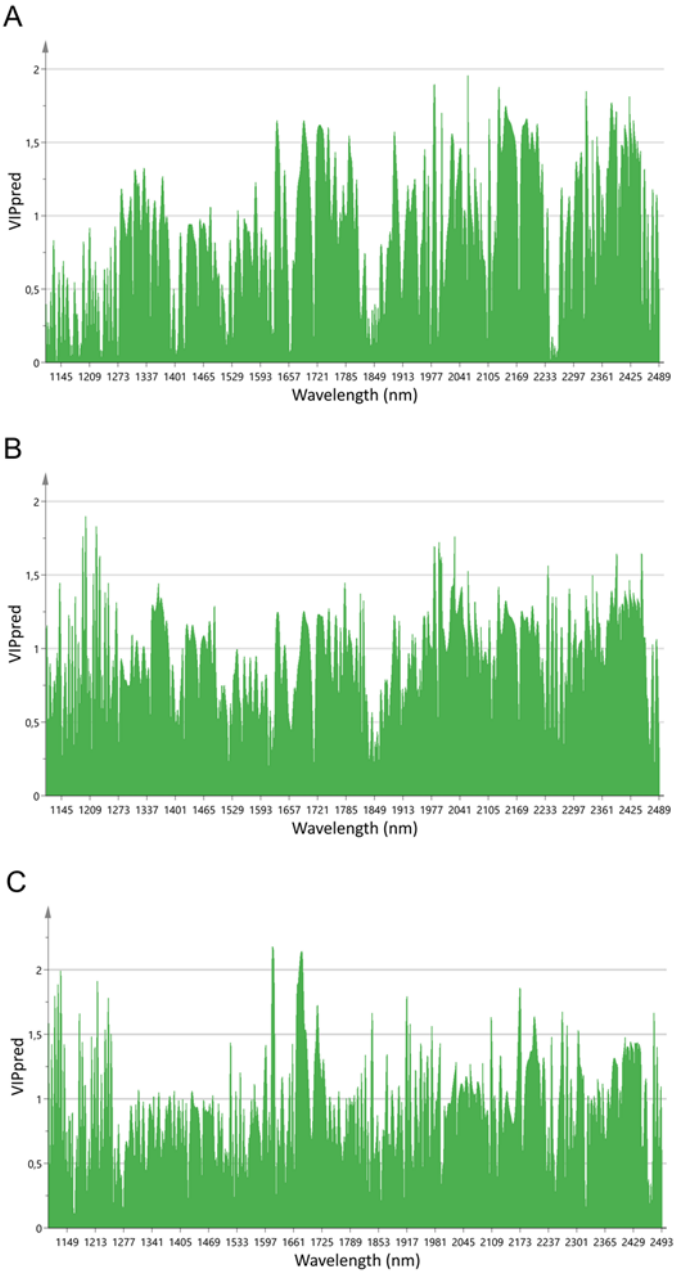


Figure S1. Variable importance in projection (VIP) plot for OPLS-DA predictive components, summarizing the contribution the NIR wavelengths make to discriminate samples according to production method (**A**), farming system (**B**), and geographical provenance (**C**).

References

Abbas, O., Zadravec, M., Baeten, V., Mikuš, T., Lešić, T., Vulić, A., Pleadin, J. (2018). Analytical methods used for the authentication of food of animal origin. *Food Chemistry*, *246*, 6–17. https://doi.org/10.1016/j.foodchem.2017.11.007

Aenugu, H. P. R., Sathis Kumar, D., Srisudharson, Parthiban, N., Ghosh, S. S., & Banji, D. (2011). Near infra red spectroscopy- An overview. *International Journal of ChemTech Research*, *3*(2), 825–836. https://doi.org/10.5772/24208

Alasalvar, C., Taylor, K. D. A., Zubcov, E., Shahidi, F., & Alexis, M. (2002). Differentiation of cultured and wild sea bass (Dicentrarchus labrax): Total lipid content, fatty acid and trace mineral composition. *Food Chemistry*, 79(2), 145–150. https://doi.org/10.1016/S0308-8146(02)00122-X

Arechavala-Lopez, P., Fernandez-Jover, D., Black, K. D., Ladoukakis, E., Bayle-Sempere, J. T., Sanchez-Jerez, P., & Dempster, T. (2013). Differentiating the wild or farmed origin of Mediterranean fish: A review of tools for sea bream and sea bass. *Reviews in Aquaculture*, *5*(3), 137–157.
https://doi.org/10.1111/raq.12006

Bell, J. G., Preston, T., Henderson, R. J., Strachan, F., Bron, J. E. Cooper, K., & Morrison, D.J. (2007). Discrimination of wild and cultured european sea bass (Dicentrarchus labrax) using chemical and isotopic analyses. *Journal of Agricultural and Food Chemistry*, *55*(15), 5934-5941.
http://dx.doi.org/10.1021/jf0704561

Blanco, M., & Villarroya, I. (2002). NIR spectroscopy: A rapid-response analytical tool. *TrAC - Trends in Analytical Chemistry*, *21*(4), 240–250.
https://doi.org/10.1016/S0165-9936(02)00404-1

Brereton, R.G. (2006). Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data. *TrAC - Trends in Analytical Chemistry*, *25*(11), 1103–1111.
https://doi.org/10.1016/j.trac.2006.10.005

Costa, C., D'Andrea, S., Russo, R., Antonucci, F., Pallottino, F., & Menesatti, P. (2011). Application of non-invasive techniques to differentiate sea bass (Dicentrarchus labrax, L. 1758) quality cultured under different conditions. *Aquaculture International*, *19*(4), 765–778. https://doi.org/10.1007/s10499-010-9393-9

Cozzolino, D. (2015). Infrared Spectroscopy as a Versatile Analytical Tool for the Quantitative Determination of Antioxidants in Agricultural Products, Foods and Plants. *Antioxidants*, *4*(3), 482–497. https://doi.org/10.3390/antiox4030482

Cozzolino, D., Chree, A., Scaife, J. R., & Murray, I. (2005). Usefulness of Near-infrared reflectance (NIR) spectroscopy and chemometrics to discriminate fishmeal batches made with different fish species. *Journal of Agricultural and Food Chemistry*, *53*(11), 4459–4463. https://doi.org/10.1021/jf050303i.

Eriksson, L., Trygg, J., & Wold, S. (2008). CV-ANOVA for significance testing of PLS and OPLS® models. *Journal of Chemometrics*, *22*(11–12), 594–600.
https://doi.org/10.1002/cem.1187

EUMOFA (European Market Observatory for Fisheries and Aquaculture products) (2017)**.** Highlights the EU in the World EU market supply consumption trade

EU landings aquaculture production. The EU fish Market, 2017 Edition.
http://www.eumofa.eu/documents/20178/108446/The+EU+fish+market+2017.pdf

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

FAO (2018). Overview of food fraud in the fisheries sector. Text by Reilly, A. *Fisheries and Aquaculture Circular No. 1165*. Rome, Italy.

FAO 2005-2018 (2005). Cultured Aquatic Species Information Programme. Dicentrarchus labrax. Cultured Aquatic Species Information Programme. Text by Bagni, M. *FAO Fisheries and Aquaculture Department* [online]. Rome. Updated 18 February 2005.
http://www.fao.org/fishery/culturedspecies/Dicentrarchus_labrax/en

Fuentes, A., Fernández-Segovia, I., Serra, J. A., & Barat, J. M. (2010). Comparison of wild and cultured sea bass (Dicentrarchus labrax) quality. *Food Chemistry*, *119*(4), 1514–1518. https://doi.org/10.1016/j.foodchem.2009.09.036

Galindo-Prieto, B., Eriksson, L., & Trygg, L. (2014). Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). *Journal of Chemometrics*, 28, 623-632. https://dx.doi.org/10.1002/cem.2627

Granato, D., Putnik, P., Kovačević, D. B., Santos, J. S., Calado, V., Rocha, R. S., Pomerantsev, A. (2018). Trends in Chemometrics: Food Authentication, Microbiology, and Effects of Processing. *Comprehensive Reviews in Food Science and Food Safety*, *17*(3), 663–677. https://doi.org/10.1111/1541-4337.12341

Khodabux, K., L'Omelette, M. S. S., Jhaumeer-Laulloo, S., Ramasami, P., & Rondeau, P. (2007). Chemical and near-infrared determination of moisture, fat and protein in tuna fishes. *Food Chemistry*, *102*(3), 669–675.
https://doi.org/10.1016/j.foodchem.2006.05.057

Lenas, D., Chatziantoniou, S., Nathanailides, C., & Triantafillou, D. (2011). Comparison of wild and farmed sea bass (Dicentrarchus labrax L) lipid quality. *Procedia Food Science*, *1*, 1139–1145.
https://doi.org/10.1016/j.profoo.2011.09.170

Liu, D., Zeng, X.-A., & Sun, D.-W. (2013). NIR Spectroscopy and Imaging Techniques for Evaluation of Fish Quality—A Review. *Applied Spectroscopy Reviews*, *48*(8), 609–628. https://doi.org/10.1080/05704928.2013.775579

Liu, Y., Ma, D.-H, Wang, X.-C, Liu, L.-P, Fan, Y.-X, & Cao, J.-X (2015). Prediction of chemical composition and geographical origin traceability of Chinese export tilapia fillets products by near infrared reflectance spectroscopy. *LWT - Food Science and Technology*, *60*(2), 1214–1218.
https://doi.org/10.1016/j.lwt.2014.09.009

Majolini, D., Trocino, A., Xiccato, G., & Santulli, A. (2010). Near infrared reflectance spectroscopy (NIRS) characterization of European sea bass (Dicentrarchus labrax) from different rearing systems. *Italian Journal of Animal Science*, *8*(2s), 860. https://doi.org/10.4081/ijas.2009.s2.860

McGrath, T. F., Haughey, S. A., Patterson, J., Fauhl-Hassek, C., Donarski, J., Alewijn, M., Elliott, C. T. (2018). What are the scientific challenges in moving from targeted to non-targeted methods for food fraud testing and how can they be addressed? – Spectroscopy case study. *Trends in Food Sceince and Technology,*

76(December 2017), 38–55. https://doi.org/10.1016/j.tifs.2018.04.001

Moretti, A., Fernandez-Criado, M. P., Cittolin, G., & Ruggero Guidastri. (1999). Manual on hatchery production of seabass and gilthead seabream, Volume 1. FAO, Rome.

Naes, T., Isakson, T., Fearn, T., &.Davies, T. (2002). A user-friendly guide to multivariate calibration and classification. *NIR Publications*, Chichester, UK.

Orban, E., Nevigato, T., Di Lena, G., Casini, I., & Marzetti, A. (2003). Differentiation in the lipid quality of wild and farmed seabass (Dicentrarchus labrax) and gilthead sea bream (Sparus aurata). *Journal of Food Science*, 68(1), 128–132.
https://doi.org/10.1111/j.1365-2621.2003.tb14127.x

Osborne, B. G. (2000). Near-Infrared Spectroscopy in Food Analysis. *Encyclopedia of Analytical Chemistry*, 1–14. https://doi.org/10.1002/9780470027318.a1018

Ottavian, M., Facco, P., Fasolato, L., Novelli, E., Mirisola, M., Perini, M., & Barolo, M. (2012). Use of near-infrared spectroscopy for fast fraud detection in seafood: Application to the authentication of wild European sea bass (Dicentrarchus labrax). *Journal of Agricultural and Food Chemistry*, *60*(2), 639–648. https://doi.org/10.1021/jf203385e

Periago, M. J., Ayala, M. D., López-Albors, O., Abdel, I., Martínez, C., García-Alcázar, A., Gil, F. (2005). Muscle cellularity and flesh quality of wild and farmed sea bass, Dicentrarchus labrax L. *Aquaculture*, 249(1–4), 175–188.
https://doi.org/10.1016/j.aquaculture.2005.02.047

Regulation (EU) No 1379/2013 of the European Parliament and of the Council of 11 December 2013 on the common organisation of the markets in fishery and aquaculture products, amending Council Regulations (EC) No 1184/2006 and (EC) No 1224/2009 and repealing Council Regulation (EC) No 104/2000. *Official Journal of the European Union*, L354, 1–21.

Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, *28*(10), 1201–1222.
https://doi.org/10.1016/j.trac.2009.07.007

Saito, H., Ishihara, K., & Murase, T. (1997). The Fatty Acid Composition in Tuna (Bonito, Euthynnus pelamis) Caught at Three Different Localities from Tropics to Temperate. *Journal of the Science of Food and Agriculture, 73*(1), 53-59.
https://doi.org/10.1002/(SICI)1097-0010(199701)73:1<53::AID-JSFA707>3.0.CO

Sivertsen, A. H., Kimiya, T., & Heia, K. (2011). Automatic freshness assessment of cod (Gadus morhua) fillets by Vis/Nir spectroscopy. *Journal of Food Engineering*, *103*(3), 317–323. https://doi.org/10.1016/j.jfoodeng.2010.10.030

Trocino, A., Xiccato, G., Majolini, D., Tazzoli, M., Bertotto, D., Pascoli, F., & Palazzi, R. (2012). Assessing the quality of organic and conventionally-farmed European sea bass (Dicentrarchus labrax). *Food Chemistry*, *131*(2), 427–433.
https://doi.org/10.1016/j.foodchem.2011.08.082

Trygg, J., & Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*, *16*(3), 119–128. https://doi.org/10.1002/cem.695

Uddin, M., & Okazaki, E. (2010). Applications of Vibrational Spectroscopy to the

Analysis of Fish and Other Aquatic Food Products. *Handbook of Vibrational Spectroscopy*, 1–16. https://doi.org/10.1002/0470027320.s8955

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Velzen, E. J. J., Dorsten, F. A. (2008). Assessment of PLSDA cross validation. *Metabolomics*, *4*(1), 81–89. https://doi.org/10.1007/s11306-007-0099-6

Xiccato, G., Trocino, A., Tulli, F., & Tibaldi, E. (2004). Prediction of chemical composition and origin identification of european sea bass (Dicentrarchus labrax L.) by near infrared reflectance spectroscopy (NIRS). *Food Chemistry*, *86*(2), 275–281. https://doi.org/10.1016/j.foodchem.2003.09.026

CHAPTER 3

Research Article

Near-Infrared Spectral Fingerprinting: a Tool against Origin-Related Fraud in the Sector of Processed Anchovies

Maria Olga Varrà, Sergio Ghidini, Adriana Ianieri, and Emanuela Zanardi*

Department of Food and Drug, University of Parma, Strada del Taglio 10, 43126 Parma, Italy.

*Correspondence: emanuela.zanardi.unipr.it; Tel.: +39-0521-902-760

Adapted from

Varrà, M. O., Ghidini, S., Ianieri, A., & Zanardi, E. (2020). Near infrared spectral fingerprinting: A tool against origin-related fraud in the sector of processed anchovies. Food Control, 107778.

Abstract: In the present study near-infrared (NIR) spectroscopy was used to assess the geographical traceability of salted ripened anchovies, whose raw product originated from fishing areas of Morocco, Spain, Tunisia, and Croatia. Two different products were tested: semi-finished and finished salted anchovies. The development and optimization of combined discrimination models based on orthogonal partial least square-discriminant analysis successfully led to the identification of the geographical origin of both anchovy datasets with >98% sensitivity, >99% specificity, and >99% accuracy on average. While NIR absorption bands related to proteins and degradation compounds highly characterized samples from Morocco and those of unsaturated lipids and derivatives globally contradistinguished anchovies from Tunisia, absorptions of both protein and lipid compounds were responsible for the discrimination of samples from Croatia and Spain. The proposed method is particularly helpful to guarantee the authenticity of salted ripened anchovies and, thus, to deter commercial frauds throughout the fish value chain and ensure traceability along the whole food chain

Abbreviations: Area under the Receiver Operating Characteristic Curves, AUROC; cross-validation, CV; analysis of variance testing of cross-validation predictive residuals, CV-ANOVA; Croatia, CR; Morocco, MO; multiple scatter correction, MSC; near infrared spectroscopy, NIRS; orthogonal partial last square-discriminant analysis, OPLS-DA; principal component analysis, PCA; principal component, PC; root mean square error from cross-validation, RMSECV; root mean square error of estimation, RMSEE; root mean square error of prediction, RMSEP; second derivative, SD; Savitzky-Golay smoothing, SG; Spain, SP; Tunisia, TU; variable influence on projection, VIP.

1. Introduction

Salted anchovy is obtained through the processing of the fresh anchovy, consisting of salting of the beheaded and partially gutted raw fish, followed by fermentation until the degree of ripeness required is reached. The salt-cured Mediterranean-style anchovy (Martin, Carter, Flick, & Davis, 2000) is obtained from the species Engraulis encrasicholus L., which naturally populates the Eastern North and Central Atlantic, Mediterranean, Western Black, and Azov seas coasts (FAO, 2010). Spain, Italy, Greece, Tunisia, France, and Morocco have a strong tradition in producing and consuming salted anchovies to the point of representing the main large-scale producers as well the leading exporters of the product to the whole Mediterranean area (EUMOFA, 2018).

From a commercial point of view, the supply chain of anchovy is overly complex and diversified. The salt-cured product obtained by fermentation can be considered as an intermediate product intended for further minimal or heavy processing. The anchovy semi-preserves meant for the direct sales are simply obtained by salt-packaging of the whole fermented dressed anchovy. Nevertheless, many other transformed products can be found in the marketplace, as filleted salted anchovy preserved in oil, anchovy paste, and several gastronomic preparations.

Based on both the production process adopted from the fish industry and the desired organoleptic characteristics of the final product, the duration of the ripening process may extend from a minimum of 2-3 months to an average time of 12-15 months. During the ripening process complex chemical, biochemical and physicochemical changes occur in the fish flesh, determining the conversion of the major compounds and the development of final aroma and sensory traits. The magnitude and the speed of these reactions are the results of the specific processing conditions (e.g. temperature, pH, salt, water activity) plus the original biological characteristics of the fresh fish. Fat, protein, water content, endogenous enzymatic pattern, and microbiota composition of the raw fish are in fact fully reflected in the overall qualitative attributes of the salted ripened anchovy (Hernandez-Herrero, Roig-Sagues, Lopez-Sabater, Rodriguez-Jerez, & Mora-Ventura, 2002). In turn, the whole composition of the raw fish

tissues is influenced by several factors, among which environment characteristics (e.g. temperature and salinity of the water) are determinant (Romotowska, Karlsdóttir, Gudjónsdóttir, Kristinsson, & Arason, 2016).

Although the indication of the geographical origin of semi-preserved pre-packed fish products is not mandatory under the current European legislation (European Parliament and Council of the European Union, 2013), the provision of increasingly detailed information to the consumer through the label is growing. This is due to a number of reasons, including the interest towards sustainable fisheries and the correlation between origin and perceived quality and tradition of the product. On the other hand, the fish supply chain is subjected to such a complex sector regulation that the monitoring of traceability and labelling of processed fish through all the stages of production, processing, and distribution is equally complex (Regulation (EU) No 1379/2013). These factors together have all meant that adequate and quick tools to verify the origin authenticity of raw material, semi-finished, and final products are becoming of utmost importance.

The use of NIR spectroscopy so as to find information related to the geographical origin is an elaborated and often unsuccessful process, since geographical provenance is the sum of a huge amount of different intrinsic or extrinsic factors which are difficult to be properly identified (Ghidini, Varrà, & Zanardi, 2019). Despite this, few but significant applications of NIR spectroscopy aimed at discriminating fish according to different countries or fishing areas of origin are currently available in literature (Liu et al., 2015; Guo et al., 2018; Ghidini et al., 2019). In these cases, the success in addressing the origin authenticity issues was likely due to the use of chemometrics to handle complex spectral data. In addition, these works converged upon the conclusion that the characteristics of the aquatic environment (latitude, water temperature, water exchange, salinity) reflects on the whole final product in term of flesh composition. Protein patterns, but, above all fatty acid profiles, represented the most affected fish constituents to these variations and, as such, their high detectability by NIR spectroscopy made them useful indicators of geographical origin. Other targeted investigations underlined still more significantly the close relationship between fatty acid profiles and the geographical provenance of fish and seafood (Ricardo et. al,

2015; Zhang, Liu, Li, & Zhao, 2017). Besides, some proteins were shown to be expressed as different allozymes in relation to different provenances (Mork & Giæver, 1999; Drengstig Fevolden, Galand, & Aschan, 2000) or overexpressed is response to environmental stressing conditions (El Sheikha & Montet, 2016).

Origin traceability explored in NIR-based studies was limited to raw unprocessed fish products, the characteristics of which were not altered or modified by processing and whose composition was more closely related to the provenance. Nonetheless, in several other fermented foodstuffs such as wine, cheese, tea, and soy sauce, the geographical origin of the raw material employed was correctly discriminated by NIR spectroscopy (Liu et al., 2008; Pillonel, Schaller, Picque, Cattenoz, & Bosset, 2005; Ren et al., 2013; Iizuka & Aishima, 1997). Similar applications concerning fermented fish products are not available since the technique has been exploited only to study the proximate composition and to establish optimal parameters for fermentation (Huang et al., 2001; Huang et al., 2003; Svensson, Nielsen, & Bro, 2004).

On the light of these considerations, the present work was aimed to guarantee the geographical traceability of salted ripened anchovies throughout some stages of the production chain by using a fingerprinting approach based on NIR spectroscopy and chemometrics. For this reason, both semi-finished products and finished semi-preserved products originating from different geographical origins, i.e. Morocco, Spain, Tunisia, and Croatia were considered.

2. Materials and methods

2.1. Sampling of anchovy specimens

Two sets of products of different provenances and obtained from the industrial scale processing of European anchovy (Engraulis encrasicolus L.) were considered in the present study: i) semi-finished salted anchovies (intended for further packaging or processing, i.e. filleting and oil preservation); ii) finished salted anchovies packaged in glass jars (intended for direct marketing).

The salted, ripened products were industrially manufactured according to the traditional process following a standard procedure. Briefly, the raw anchovies were firstly immersed into a saturated brine solution, manually beheaded, partially eviscerated, and then rinsed again with brine solution. Fish was then placed into barrels and layered alternatively with sodium chloride. The salt-curing process in barrels was carried out under pressure and at controlled room temperature (18 °C), for different months. After ripening, fish from each provenance and from different batches was randomly taken from barrels and directly vacuum-packaged into plastic bags to obtain the semi-finished product, or it was further processed by addition of salt and packaged into glass jars (net weight=280 g) to obtain the finished product.

The total number of anchovy specimens to be considered for the construction of statistical models aimed at discriminating fish according to different origin was chosen on the basis of a recent systematic revision of the literature dealing with the use of vibration spectroscopy to assess fish authenticity (Ghidini, Varrà, & Zanardi, 2019), according to which an average number of 20 specimens per provenance was reported as suitable for achieving more than satisfactory results in terms of robustness of the final discriminant model.

The set of semi-finished anchovies included 350 samples coming from 4 different geographical areas, namely Spain (SP, Cantabrian Sea, FAO fishing area 27.8), Morocco (MO, Eastern Central Atlantic, FAO fishing area 34.1), Croatia (CR, upper Central Mediterranean Sea, FAO fishing area 37.2.1), and Tunisia (TU, lower Central Mediterranean Sea, FAO fishing area 37.2.2). The same geographical provenances, with the exception of the Moroccan fishery zone, were also investigated for the set of finished anchovies (250 samples from different glass jars), i.e. SP (FAO fishing area 27.8), CR (FAO fishing area 37.2.1), and TU (FAO fishing area 37.2.2). Detailed information on the sampling and characteristics of the transformed fish used in the present study is reported in Table 1.

Although the experimental design was initially conceived in such a way that the same number and the same provenances of both semi-finished and finished anchovy fish were included in the final datasets, during the

experimental work some finished samples from Morocco were no longer available and, therefore, they were not provided by the fish canning company. Moreover, the choice to develop discriminant models for anchovies from Morocco, Spain, Tunisia, and Croatia but not from Italy was justified by the fact that the volumes of catch of anchovy fish are larger along the in Moroccan, Spanish, Tunisian, and Croatian areas than in Italy. By consequence, the Italian industrial transformation of local anchovy fish takes place on smaller scale compared to other countries, leading to the presence on the market of very diversified products which are less suitable for standardisation.

Table 1. Sampling plan for the anchovy products

Product	Country of origin	FAO fishing area	Fishing month	Size	Ripening age (months)	No. of samples	No. of spectra
Semi-finished anchovies	SP	27.8	05	S	10-11	200	100
	MO	34.1	05	S	10-11	200	100
	CR	37.2.1	05	S	10-11	100	50
	CR	37.2.1	08	S	9-10	50	25
	CR	37.2.1	08	M	9-10	50	25
	TU	37.2.2	09	S	9-10	50	25
	TU	37.2.2	09	XS	9-10	50	25
							TOT. 350
Finished anchovies (glass jar)	SP	27.8	05	S	10-11	200	100
	CR	37.2.1	05	S	10-11	100	50
	CR	37.2.1	08	S	9-10	50	25
	CR	37.2.1	08	M	9-10	50	25
	TU	37.2.2	09	S	9-10	50	25
	TU	37.2.2	09	XS	9-10	50	25
							TOT. 250

For country of origin: Spain (SP); Morocco (MO); Croatia (CR); Tunisia (TU). For fishing month: May (05); August (08), September (09). For size (fresh fish): medium (M); small (S); extra small (XS).

2.2. Sample preparation

Plastic bags and glass jars containing the anchovies were stored at refrigerated temperature (4 ± 2 °C). At the time of analysis each individual fish was removed from the package and patted dry with filter paper to

remove the excess salt. The skin, viscera, fins, and main bone were carefully removed, and the resulting two fillets were manually and finely chopped by knife until a doughty consistence was reached and non-homogeneous fragment of fish muscles were no longer visible. To reduce inter-individual variability, two minced fillets of the same specimens were then merged and mixed with the two minced fillets obtained from another specimen to create a unique, final sub-sample intended for NIR spectra acquisition.

The workflow summarising the experimental procedure adopted in the present work, from samples provision to spectra statistical elaboration, is illustrated in Figure 1.

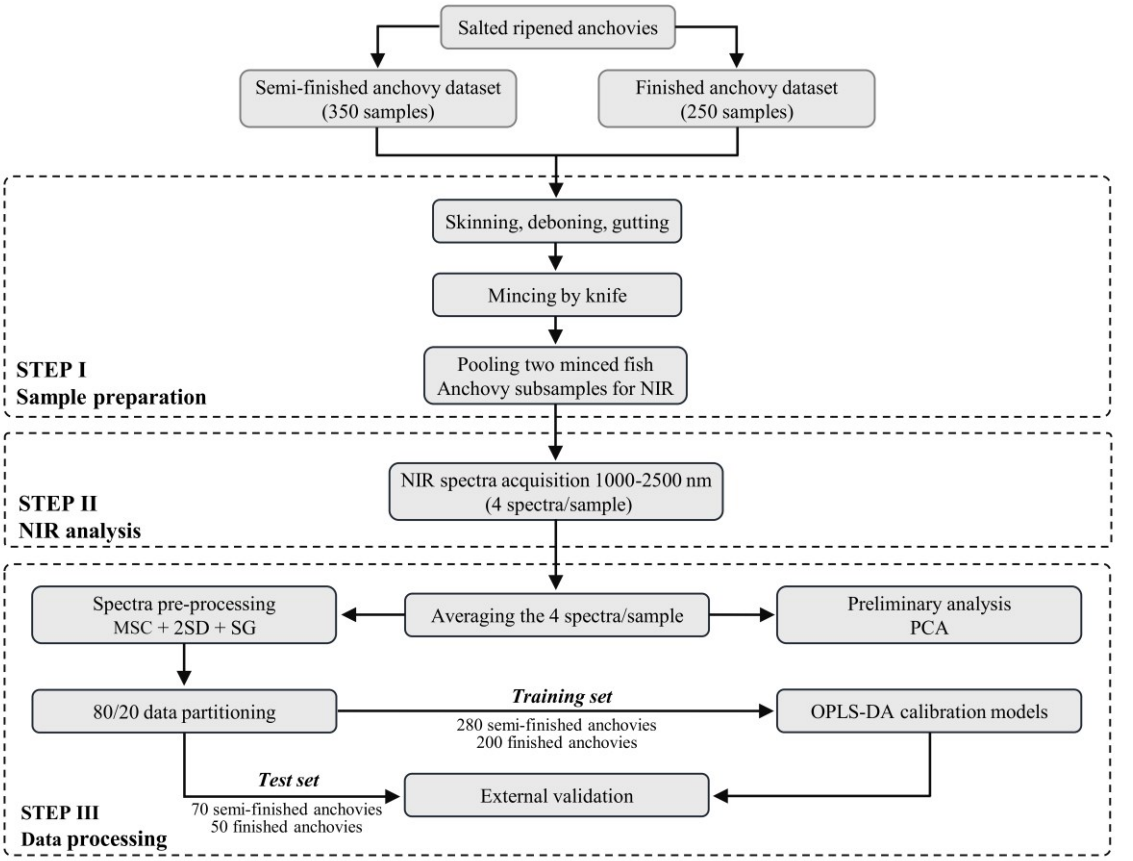


Figure 1. Workflow showing the followed **experimental** procedure.

2.3. NIR spectroscopy

Round quartz sample holders (35 mm diameter) were filled with approx. 8 g of minced sample and scanned in diffuse reflectance mode by using a NIRFlex® N-500 FT-NIR benchtop spectrometer (Büchi Labortechnik AG, Flawil, Switzerland) over the wavelength range of 1000–2500 nm and at a spectral interval of 1 nm (32 scans/spectrum). At regular intervals of acquisition (every 20 samples) the instrument was calibrated using a proper external calibration standard intended for the establishment of the accuracy of the wavelength scale and consisting of the rare earth oxide Standard Reference Material® 1920a of the National Institute of Standards and Technology. To obtain more accurate results, four replicates of the NIR spectrum were collected for each sample by rotating 90° the sample holder three times. The four spectra were further averaged.

2.4. Multivariate data analysis

Mean reflectance (R) NIR spectra were converted into absorbance (A) spectra (A = log 1/R). The range 1000–1199 nm was excluded from the subsequent chemometric elaboration, since it was characterised by a low signal-noise ratio. Therefore, the final data matrix of semi-finished salted anchovies included a total of 1301 variables (NIR absorbance values in the 1200-2500 nm range) and 350 spectra, while data matrix of finished salted anchovies included 1301 variables and 250 spectra.

Chemometrics analysis was performed by the software SIMCA-P 14.1 (Umetrics, Umeå, Sweden). Raw spectral data were firstly elaborated by means of principal component analysis (PCA) to look at the data structure, verify whether the only application of the unsupervised data modelling was suitable to achieve a separation of sample groups, and identify strong outlier samples through the Hotelling's T^2 test (at 95% confidence interval).

Multi-class orthogonal partial least square-discriminant analysis (OPLS-DA) was instead applied to create combined models for discriminating and predicting the geographical origin of the two data matrices. OPLS-DA is a qualitative regression-based supervised discriminant analysis allowing to maximize the separation among classes through the partitioning of the total variation within data into a related predictive variation (enclosing

information related to the discrimination purpose) and an orthogonal (non-related) variation (Trygg & Wold, 2002). OPLS-DA is a modification of the classical partial least square discriminant analysis (PLS-DA) which, despite also being aimed at maximising class separation, does not allow to slit the total variation into predictive and non-predictive variations. Therefore, PLS-DA may result in misinterpretation of the results (Bylesjö, Rantalainen, Cloarec, Nicholson, Holmes, & Trygg, 2007).

Prior to perform OPLS-DA, the reduction of the undesirable large baseline shifts, as well as the separation of the typical combination spectral bands, were attempted by applying the following mathematical pre-treatments to the raw spectra: multiplicative scatter correction (MSC), second derivative (2SD, second polynomial order, 15 point) and Savitzky-Golay smoothing (SG, 15 points).

While PCA was performed on the whole data, the OPLS-DA calibration models were computed on the training set including 80% of randomly selected and representative original spectra (280 spectra for semi-finished anchovies and 200 spectra for finished anchovies). The 20% of the spectra left (70 spectra for semi-finished anchovies; 50 spectra for finished anchovies) was completely independent and was used as test set to externally validate the OPLS calibration models. Considering that no general rules concerning data splitting ratio are available (Westad & Marini, 2015) and given that the number of total samples to be investigated in the present work was quite high, the 80/20 partitioning of data according to the so-called Pareto principle (Massart, Vandeginste, Buydens, Jong, Lewi, & SmeyersVerbeke, 1997) was initially investigated. Since the results achieved were more than satisfactory in terms of prediction, therefore other possible data splitting ratios were not tested. Moreover, the 80/20 splitting ratio was already reported to provide robust and stable model for the discrimination of the geographical origin of fish dataset composed by approximately the same number of samples (Guo et al., 2017).

2.4.1. Validation of the chemometric models

An internal 10-fold cross validation (CV) was employed to establish the correct number of principal component (PCs) for PCA and predictive the

orthogonal components for OPLS-DA, based on the lowest significant values of root mean square error from CV (RMSECV) achieved. The overall quality of the resulting OPLS-DA models was estimated by the following statistical indicators: $R^2X_{(cum)}$ (total amount of variation of spectra, i.e. X-matrix, collected); $Q^2_{(cum)}$ (predictive variation of X-matrix collected); $R^2Y_{(cum)}$ (total amount of variation of OPLS classifier models strictly related to class membership of samples, i.e. Y-matrix); p-values from the analysis of variance of the cross-validation residuals (CV-ANOVA). Finally, the high risk of overfitting and overestimation of the OPLS-DA models was averted by permuting the class labels of samples 200 times in CV. Models were considered valid and robust when the resulting values of R^2 and Q^2 Y-intercepts were lower than 0.4 and 0.05, respectively (Van der Voet, 1994).

As for the external validation procedure of the OPLS classifiers, sensitivity, specificity, and overall accuracy parameters were calculated for each of the groups in which the new samples of the test were classified. Sensitivity corresponded to the proportion of true positive samples, specificity to the proportion of true negative samples, and accuracy to the proportion of true positive and negative samples. To sum up how the combination of different values of sensitivity and specificity impacted the model discrimination performances, Areas under the Receiver Operating Characteristic curves (AUROC) were also calculated for each predicted group. Sensitivity, specificity, accuracy, and AUROC values equal to 1, suggested perfect outcomes in prediction of the discriminant model (Forina, Armanino, Leardi, & Drava, 1991). For further details about the equations to calculate of these parameters, readers are referred to Fawcett (2006).

Finally, the variable influence on projection (VIP) analysis for the OPLS predictive components was applied to better interpret the discrimination models and pinpoint the most relevant NIR wavelengths for the discrimination of the geographical origins of the anchovies (VIP indexes ≥ 1).

2.4.2. Identification of the NIR signature features of each geographical provenance

Since in multi-class OPLS-DA the VIP analysis does not provide information about the spectral regions bearing the separating information for single

classes, subsidiary pairwise OPLS-DA models were built (i.e. by discriminating the single classes by two, consecutively). The NIR bands with the highest power of discrimination between two classes were therefore identified via the resulting S-line plots, by looking at the extent of the predictive loadings and at the associated absolute values of the correlation coefficients (r).

2. Results

3.1. Overview of spectral characteristics and sample natural distribution by PCA

Raw NIR absorbance spectra, as such, were regarded as unfit and misleading for statistical elaboration (data not shown). The corrected averaged spectra (MSC, 2SD, SG) of the semi-finished and finished anchovies coming from different fishing areas are reported in Figure 2. No apparent differences in terms of spectral patterns and shapes were present in the spectra, although some slight changes in peak intensity were observed in the 2020–2090 nm region (whose absorption peak are linked to protein, urea, oil, and -OH group), 2180–2200 nm region (mainly associated to protein absorption), and 2400–2440 nm region (absorption of -C-H of aryl functional group) of the semi-finished anchovies (Figure 2A) (Shenk, et al., 2001; Workman & Weyer, 2012). Similarly, the 2000–2050 (protein, urea) and 2400–2450 (CH of aryl functional group) nm regions of the NIR spectra of finished anchovies showed some variations (Figure 2B). The most prominent absorption bands (negative peaks) of both anchovy products were observed as follows: around 1360 nm (hydrocarbons), 1730 nm (hydrocarbons/alcohol), 2010–2100 nm (protein, urea, oil, and -OH), 2180 nm (proteins), 2350 nm (aryl functional group), and 2390–2460 nm (aryl functional group) (Shenk, et al., 2001; Workman & Weyer, 2012).

From the application of PCA, a total of 13 PCs explaining 88.5% of total variation of the original raw spectra were calculated in CV for the semi-finished anchovy dataset ($R^2X_{(cum)}$ =0.885, $Q^2_{(cum)}$=0.733). Similarly, 15 PCs explaining 83.5% of variation were extracted in CV for the finished anchovy dataset ($R^2X_{(cum)}$ =0.835, $Q^2_{(cum)}$=0.657). Despite numerous, all the extracted components were found to carry enough spectral information to be retained in CV during the computation of the analysis, but the first two PCs

calculated from the semi-finished and finished anchovy datasets, enclosed 72% and 64% of the total variations, respectively, thus underling the redundance of the other PCs. This phenomenon is likely the consequence of large original number of spectral variables (1501 absorption values) to be compressed by PCA on the one hand, and of the natural multicollinearity proper of NIR absorption bands on the other.

As expected, no evident clusters of samples were visualized in the score plots of the first two PCs of both anchovy datasets, although for Moroccan anchovy of the semi-finished anchovy dataset and for Tunisian anchovies of the finished anchovy dataset, a distribution along the negative axis of the PC2 and a trend in separating from the other groups was observed (Figg. S1A, S2A). Spanish and Croatian semi-finished and finished samples appeared to be closer each other, but Croatian semi-finished and finished samples were found to be more widely scattered across the plot area compared to the other group, thus suggesting a greater diversification in terms of fish flesh composition.

During this preliminary data elaboration stage, some samples fell outside the 95% confidence interval (ellipse) of the PCA score plot, but they were not found to be strong outliers according to the Hotelling's T^2 test (Suppl. Figg. S1C, S2C).

By looking at the negative loadings of the PC2 of the semi-finished anchovy dataset (Supplementary Figure S1B), the largest peaks pointing downwards were found to be around 1460, 1510, 1820, 1980, and 1990, all previously attributed to amides, proteins and urea groups absorption (Workman & Weyer, 2012). Since the influence exerted by the large negative peaks in the 2140–2440 nm lipid-associated region was much lower, the protein fraction was considered the most significant in contradistinguish Moroccan specimens. By contrast, a stronger contribution of lipids to the slight separation of the specimens coming from Tunisia in the finished anchovy dataset was evident. Peaks associated with the absorption of lipids, hydrocarbon chains and urea located at 1390 and 1680 nm (hydrocarbons), 2030 and 2070 nm (urea), 2140 nm (lipids), 2350 nm (aromatic compounds), 2420 nm, and 2440 nm (CH aryl group) were found to be the most important loadings of the negative PC2 (Supplementary Figure S2B). The two peaks

located at 1530 and 2060 nm were associated to the absorption of CH group of alkynes/NH group of secondary amines and $CONH_2$ combination of amide and proteins (Workman & Weyer, 2012).

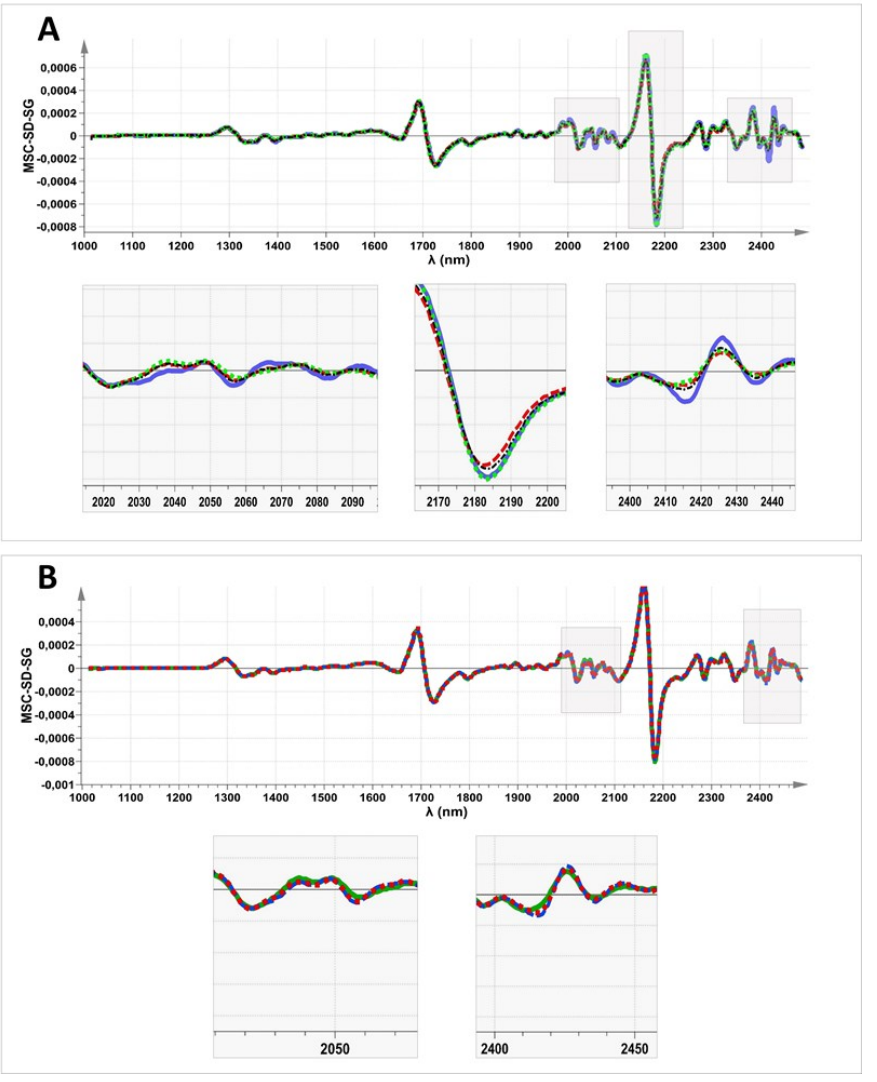


Figure 2. Pre-treated NIR spectra (MSC-SD-SG) of semi-finished (A) and finished (B) anchovies and the main differences in spectral patterns. MO= blue solid line; SP=red dashed line; TU=green dotted line; CR=black dash-dotted line.

Nevertheless, preliminary results from PCA suggested that most of the total variability in the spectral pattern was not specifically related to the fishing geographical area and stated the view that a large amount of orthogonal variability hid the discrimination. This orthogonal variability is likely the sum of many other factors influencing spectral shapes and intensities which were previously verified to be dependent on feeding habits, age, muscular activity, competition, etc. (Guo et al., 2018). For these reasons, the predictive variability within the data was selectively extracted by OPLS methods and processed to obtain accurate classification outcomes.

3.2. Combined discrimination of geographical origins using OPLS-DA algorithm

Two prediction models based on multi-class OPLS-DA were built for the two sample datasets (training sets) to fit all the geographical provenances at once.

The application of OPLS-DA to the training set of the semi-finished anchovies led to the distinction of samples into four main clusters in the resulting score plot (Supplementary Figure S3A), each corresponding to one of the four different fishing areas under investigation. A clear-cut distribution of CR and MO specimens in one specific quadrant of the score plot was not observed, but they were perfectly discriminated along the t[1], while anchovies from TU and SP groups reflected along the t[2].

Despite the fitting ability of 74% ($R^2X_{(cum)}$=0.742), it was found that only 24% of spectral variability was predictive and enclosed by 3 components; the remaining 50% was not correlated with the groups and was collected by 6 orthogonal components. Nevertheless, the high values of explained predictive variability ($R^2Y_{(cum)}$=0.878) and predictive power ($Q^2_{(cum)}$=0.812), as well as results from the permutation test and CV-ANOVA ($p<0.05$) proved the validity of the calibration model (Table 2).

Table 2. Cross-validation results of OPLS-DA applied to the two spectral datasets (1200–2500 nm) of the salted ripened anchovies

Validation	Dataset	Group	Comp	$R^2X_{(cum)}$	$R^2Y_{(cum)}$	$Q^2_{(cum)}$	RMSECV	RMSEE	R^2 Y-intercept*	Q^2 Y-intercept*	*p* CV-ANOVA**
Internal (training set)	Semi-finished anchovies	MO	3 + 6	0.742	0.878	0.812	0.126	0.106	0.239	-0.315	1.13e-16
		SP					0.168	0.141	0.244	-0.329	
		CR					0.265	0.213	0.252	-0.320	
		TU					0.152	0.127	0.250	-0.337	
	Finished anchovies (glass jar)	SP	2 + 6	0.695	0.902	0.805	0.128	0.131	0.312	-0.447	3.76e-09
		CR					0.206	0.183	0.305	-0.418	
		TU					0.163	0.124	0.310	-0.446	

For groups (country of origin): MO=Morocco; SP=Spain; CR=Croatia; TU=Tunisia. Comp= total number predictive (first) + orthogonal (second) OPLS components. $R^2X_{(cum)}$= cumulative modelled spectral variation. $R^2Y_{(cum)}$=cumulative modelled variation associated to groups. $Q^2_{(cum)}$=cumulative predictive variation. RMSECV=root mean square error from cross-validation. RMSEE=root mean square error of estimation). R^2 Y-intercept* and Q^2 Y-intercept*=calculated by permutation test (100 permutations). *p* CV-ANOVA**=*p*-value from analysis of variance of the cross-validated residuals (95% level of significance).

Results from the recognition of the unknown samples of the test set by the fitted calibration model are reported in Figure 3 and Table 3. MO, SP, and TU anchovies were 100% assigned to the correct classes. Although the natural biological diversity within specimens of the same group, only one CR sample was wrongly classified as coming from SP, as it fell in the other class-regions (Figure 3D). This misclassified sample was responsible for the lower levels of accuracy of both SP and CR classifiers (accuracy=0.98) and the lower level of specificity of 0.98 of the SP classifier compared to the other groups. By consequence, SP group also showed the lowest AUROC value of 0.69. Furthermore, RMSEP values in prediction (Table 3) were found to be remarkably similar to RMSECV values in calibration (Table 2), thus reconfirming the validity of the fitted model as an estimator of the class membership of new anchovy specimens.

As for the main spectral differences responsible for the discrimination of samples in the score plot, the absorption bands characterized by the higher VIP indexes for predictive components were located in the 2020–2120 nm region (with the maximum peak at 2030 nm, VIP=1.75), followed by the 2280–2330 and the 2360–2440 regions, all indicative of lipid absorption (Supplementary Figure S3B). Other influential variables were found at 1320–1400 and 1650–1685 nm, whose presence also disclosed the contribution of lipids to the discrimination of samples by geographical origin. The same outcomes were previously discussed for PCA applied to semi-finished products (see *Section 3.1.*), pointing out a matching between the maximum variance extracted by PCA and the maximum separating variance extracted by OPLS-DA, which corresponded mainly to amide different profiles among groups.

As for the finished anchovy dataset, the discriminant OPLS calibration model was fitted in CV with 2 predictive and 6 orthogonal components, enclosing 19% and 50% of the total spectral variation, respectively. The overall predictability reached 80% ($Q^2_{(cum)}$=0.805) (see Table 2), but the amount of variation exclusively correlated with SP, CR and TU provenances was slightly higher ($R^2Y_{(cum)}$= 0.902) compared to the former model based on semi-finished products. Similarly, RMSECV values for the three classes were

found to be lower compared to RMSECV values for SP, CR and TU semi-finished products.

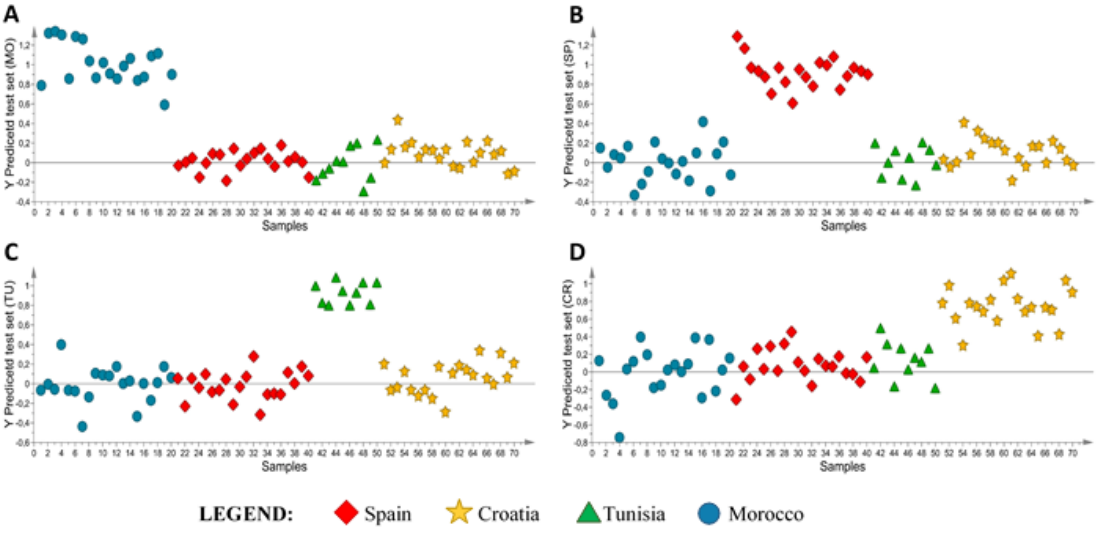


Figure 3. Values of Y variables (geographical provenances) for MO (**A**), SP (**B**), TU (**C**), and CR (**D**) test set samples of the semi-finished anchovy dataset predicted by OPLS-DA. Levels for classification in the different groups were set based on the nearest class.

Table 3. Results from classification of the test set samples.

Validation	Dataset	Group	Classif rate (%)	RMSEP	Sens	Spec	Acc	AUROC	*p*-value*
External (test set)	Semi-finished anchovies	MO	100	0.161	1	1	1	1	2.4e-18
		SP	100	0.173	1	0.98	0.98	0.69	
		CR	95	0.267	0.95	1	0.98	0.99	
		TU	100	0.159	1	1	1	1	
	Finished anchovies (glass jar)	SP	100	0.117	1	1	1	1	2.6e-11
		CR	100	0.203	1	1	1	1	
		TU	100	0.166	1	1	1	1	

For groups (country of origin): MO=Morocco; SP=Spain; CR=Croatia; TU=Tunisia. RMSEP=root mean square error of prediction. Sens=sensitivity. Spec=specificity. Acc=accuracy. AUROC=area under receiver operator characteristic curve. *p*-value*=by Fisher's Exact Test (95% level of significance).

A symmetrical separation of SP and TU samples along the t[1] was achieved, whit the first ones locating in the lower left part of the OPLS-DA score plot and the second ones in the lower right part (Supplementary Figure S4A). At the same time, CR samples were tightly clustered along the positive axis of the t[2], straddling the positive and the negative axes of the t[1]. Despite the proximity of the fishing area of CR anchovies (FAO 37.2.1) to the fishing area of TU anchovies (FAO 37.2.2), no overlapping samples were observed in the plot.

The ability of the model to perfectly recognise 100% of the finished anchovies was also confirmed by the optimal statistics resulting from the external validation procedure (Table. 3). The highest attainable values of sensitivity, specificity, accuracy, and AUROC were, in fact, reached for SP, CR and TU classes, without any sample of the test set being misclassified in the wrong geographical group, as it can be observed from the Y-predicted score plots reported in Figure 4.

Similarly to what previously reported for the discrimination of the semi-finished anchovies, the VIP plot for the predictive components showed that the most important NIR peaks influencing sample discrimination were associated to the absorption of nitrogenous compounds, since located in the 2020–2120 nm (with the highest VIP value of 2.48 at 2055 nm), in the 2280–2330, and in the 2360–2440 nm regions (Supplementary Figure S4B). Other prominent but lower peaks corresponding to lipid absorption were observed at 1375 nm (VIP=2.05), 1665 nm (VIP=1.77), and 1750 nm (VIP= 1.61). Hence, in contrast to what was observed in the loading plot of the PCA (*Section 3.1.),* the influence exerted by lipids was negligible, and likewise the OPLS model for semi-finished product discussed above, the differentiation of the geographical origin of the finished anchovies was mainly guided by variation associated to proteins and derivative molecules.

The lack of research related to the use of NIR spectroscopy to discriminate the geographical origin of fish and, in particular, to the use of VIP parameters to identify the most influential NIR absorption bands, hindered the provision of an in-depth and reliable comparison of the VIP indexes results obtained. Despite this, only one study performed on raw fish reported the use of VIP parameters but, in this case, the discrimination of

fish origin was mainly driven by the influence exerted by lipid absorption bands located in the 1620-1720 nm region (Ghidini et al., 2019). These conflicting results are likely to be the consequence of the different chemical composition and moisture content of raw fish which compared to salted ripened anchovy.

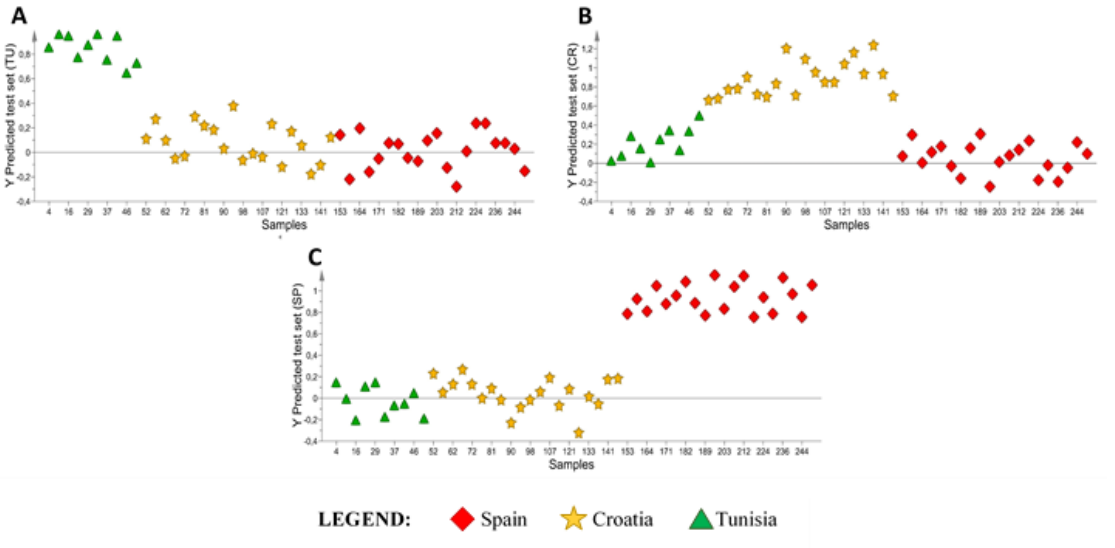


Figure 4. Values of Y variables (geographical provenances) for TU (**A**), CR (**B**), and SP (**C**) test set samples of the finished anchovy dataset predicted by OPLS-DA. Levels for classification in the different groups were set based on the nearest class.

3.3. Pairwise OPLS-DA and identification of class-related spectral hallmarks

The previously discussed multi-class models can be considered affordable and very advantageous in prevailing practice, since more than one provenance can be predicted at a time. Therefore, in the present section the whole point of applying OPLS-DA to couples of single classes was not to build classifiers to be used for the discrimination of the geographical origin, but rather to increase the understanding on how and how much single spectral variables contributed to the characterisation of each geographical provenance of both the semi-finished and the finished product.

3.3.1. Characterisation of samples from Morocco

By analysing the influence of NIR absorption bands to the separation of the semi-finished anchovies, two important loading values at 2060 and 2093 ($0.82 \leq r \geq 0.95$), linked to the absorption of proteins, peptides, and amino acids, plus one peak at 2426 nm (unassigned), always distinguished MO samples from the other three geographical groups (Figg. 5A, 5B and 5C). These loadings were specifically assigned to combination bands of CONH2 of amide A and amide I from polypeptides, combination bands of O–H groups, and stretching/bending vibration of the CH2 groups of the side chains of amino acids, respectively (Shenk, et al., 2001; Wang, Sowa, Ahmed, & Mantsch, 1994). The 1960–1990 nm NIR region (N–H combination bands of aromatic amines) (Workman & Weyer, 2012) and the peak at 2270 (N–H and C=O groups of peptide backbone referred to as the β-sheet structure) (Workman & Weyer, 2012) also showed their slight influence for the discrimination of MO from SP anchovies. In accordance with the results reported in the present study, the spectral variations in amide I absorption regions underlining modifications of the secondary structure of proteins were already reported in fish as being influenced by the ripening process of fish (Bocker, Kohler, Aursand, & Ofstad, 2008). Along with this, the possibility of effectively monitoring changes of the second structure of proteins by using NIR spectroscopy was previously demonstrated also for different matrices. As an example, increasing NIR absorption peaks intensities at 2184 nm, 2259 nm, and 2276-2280 nm and decreasing NIR absorption peaks intensities at 2167 nm, 2209 nm, and 2264 nm were attributed to protein structural alterations of a qualitative nature (Bruun, Søndergaard, & Jacobsen, 2007). Moreover, peaks at 2172 nm and 2289 nm were attributed to α-helix structure, peaks at 205 nm, 2264 nm, and 2313 nm were related to β-sheet structure, and the peak 2265 nm was found to be typical of the unordered protein structure (Robert, Devaux, Mouhous, & Dufour, 1999).

Finally, only one NIR peak assigned to the absorption of CH3 groups of hydrocarbons (1705 nm) proved to be an effective discriminant variable for

the separation of MO from CR anchovies, even though its lower level of correlation (r=0.52) (Fig 5C).

Despite the well-known dependence of the lipidic composition on the geography of the fishing area, the study of protein and peptide variations has been demonstrated to be a useful tool to track the geographical origin of food, even if poorly applied to food of animal origin (Kumari et al., 2018; Wang et al., 2009).

Modifications of the amino acid sequences of selected groups of proteins were already observed in raw hake specimens from American or African origin (Carrera & Gallardo, 2017), and the whole protein profile of different species of unprocessed shrimps showed an important correlation with their geographical provenance (Salla & Murray, 2013). Moreover, the induction of the expression of specific proteins in response to environmental pollutants has been demonstrated (Shepard & Bradley, 2000; Rodriguez-Ortega, Grosvik, Rodriguez-Ariza, Goksoyr, & Lopez-Barea, 2003). Therefore, proteomic could be further explored to identify those protein patterns more susceptible to change allow for the indirect identification of geographical origin of fish.

As for ripened products, in certain cases the relation between the peptide profiles and the country of origin was masked by fermentation process (Kumari et al., 2018), but some amino acids and degradation products of amine nature (trimethylamine, choline, and phosphocholine) were reported by other authors as useful markers of provenance of salted and dried mullet roe (Locci, Piras, Mereu, Cesare Marincola, & Scano, 2011).

Protein and polypeptide of the transformed anchovy products investigated in the present study also underwent several chemical and structural modification during the brining and the ripening processes, which made the final composition different, but reflecting the original composition of the raw products. Therefore, variations of the absorption of N–H, O–H and C–H groups assigned to proteins, indirectly supposes they mainly refer to the complex pattern of degradation compounds deriving from proteolysis, i.e. peptides, amino acids, peptide nucleotides and their decomposition fragments.

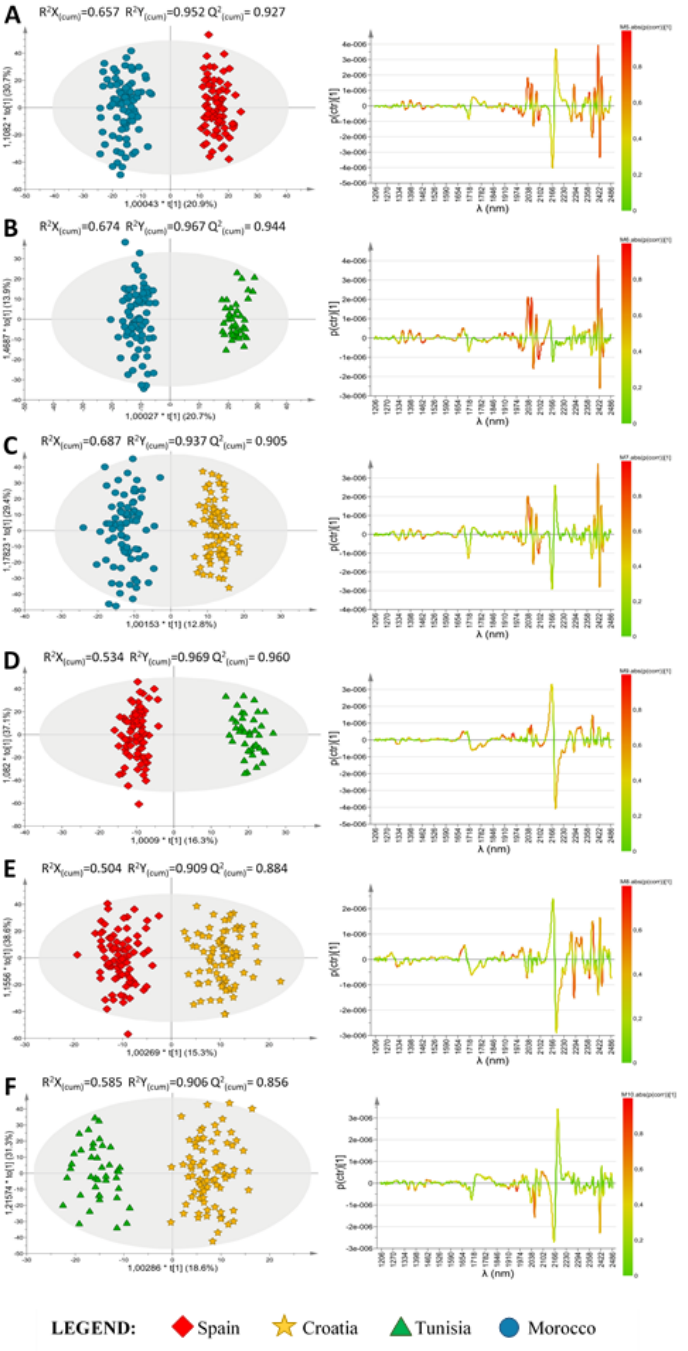


Figure. 5. Score plots of the first predictive (t[1]) and orthogonal (to[1]) components (left) and their respective S-line plots (right) for the pairwise OPLS-DA models separating the geographical origins of the semi-finished anchovy samples. (**A**) MO vs. SP; (**B**) MO vs. TU; (**C**) MO vs. CR; (**D**) SP vs. TU; (**E**) SP vs. CR; (**F**) TU vs. CR. Peaks in the positive or in the negative direction of the S-plots are coloured according to the absolute value of correlation (p(corr), from green= low values; to red = high values) and are influent in discriminating samples distributing in the positive or in the negative direction of the OPLS predictive component, respectively.

3.3.2. Characterisation of samples from Spain and Croatia

Contrary to MO samples, NIR wavelengths contributing the characterisation of SP and CR anchovies (both as semi-finished and as finished products) were highly variable based on the geographical counterpart to be compared in the two-class OPLS-DA models. For instance, SP semi-finished specimens stood out from MO specimens due to the impact of the 2030–2060 nm and the 2286–2420 nm regions, with the peaks at 2030 (C=0 stretching of urea or N-H combination bands from primary amides) (Workman & Weyer, 2012), 2055 (symmetric N–H stretching and amide I combination bands of proteins), and 2416 nm (C–H of aromatic molecules) showing the highest correlation coefficient ($r \geq 0.75$) (Figure 5A). The peak found at 2084 nm was previously attributed to the absorption of substances such as methyl oleate/linoleate hydro-peroxides (Takamura, Hyakumoto, Endo, Matoba, & Nishiike, 1995).

Correlations of specific spectral bands with the perfect separation of SP from CR semi-finished samples ($r > 0.65$) were instead observed as follows (Figure 5E): 2290 (CONH2, specifically due to the α-helix peptide structure), 2363 (CH2 methylene group of aliphatic hydrocarbons), 2440 nm (C–H of aromatic hydrocarbons) (Workman & Weyer, 2012). Similarly, both proteins and lipids were responsible for the discrimination of SP from CR finished anchovies (Figure 6C) since the SP specimens showed highly correlated peaks at 2290 nm ($r = 0.67$) and at 2363 nm ($r = 0.58$). The peak at 2090 (typical of O-H groups) was less relevant ($r = 0.55$) for the discrimination of CR from TU semi-finished samples (Figure 5F) and it was completely uncorrelated in the case of the finished anchovies (Figure 6B), for which the most significant loadings corresponded to the absorption bands at 2065, 2426, and 2440 nm.

Proteins and derivatives discriminated CR from MO semi-finished samples (peaks at 2030, 2055, and 2416 nm) (Figure 5C). Nevertheless, lipids and derivatives were more influential ($r > 0.6$) in discriminating CR from SP semi-finished products (peaks at 1900, 2310, and 2318 nm, Figure 5E) and CR from SP finished products (peak at 2380, Figure 6C). Therefore, a more balanced contribution of molecules of both protein and lipid origin justified the perfect discrimination of SP and CR anchovies from the other classes.

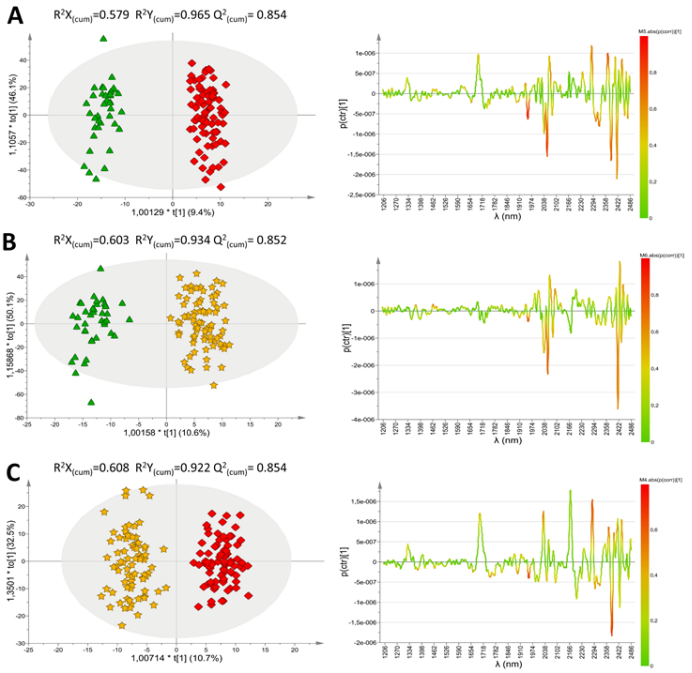


Figure. 6. Score plots of the first predictive (t[1]) and orthogonal (to[1]) components (left) and their respective S-line plots (right) for the pairwise OPLS-DA models separating the geographical origins of the finished anchovy samples. (**A**) TU vs. SP; (**B**) TU vs. CR; (**C**) CR vs. SP. Peaks in the positive or in the negative direction of the S-plots are coloured according to the absolute value of correlation (p(corr), from green= low values; to red = high values) and are influent in discriminating samples distributing in the positive or in the negative direction of the OPLS predictive component, respectively.

3.3.3. Characterisation of samples from Tunisia

The lipidic fraction has been reported in literature as the most susceptible to variation induced by environmental conditions, especially in oily fish such as anchovies. In the raw fish, unsaturated and polyunsaturated fatty acids such as oleic, eicosenoic, eicosapentaenoic, and docosahexaenoic acids and n-3/n-6 ratio were found to be the most influent lipidic molecules for the distinction of the geographical origin of anchovies (Öksuz, Özyilmaz, & Turan, 2009, Diraman & Dibeklioglu, 2009, Standal et al., 2012). Increases of the n-3 polyunsaturated fatty acids were found to be strongly linked to higher latitudes and, thus, to the water temperatures characterizing specific marine environment (Colombo, Wacker, Parrish, Kainz, & Arts, 2016).

In the ripened products, however, lipolysis and oxidation processes are responsible for the modification of the lipid fraction composition, as well as for the formation of several degradation compounds (Anggo, Ma, Swastawati, & Rianingsih, 2015). Therefore, the lipid composition of the raw fish changes considerably after the ripening process, but the signature of the original composition of the fresh fish remains in the pattern of new degradation metabolites.

In the present study, the impact of lipidic molecules to the discrimination of the semi-finished product coming from TU was the most pronounced and it was retained also after the processing to the finished products. The highest correlation values in TU vs. SP and TU vs. CR models for both semi-finished (Figg. 5D, 5F) and finished anchovies (Figg. 6A, 6B), corresponded to the 1694, 1950, 2144, 2310, 2380, and 2416 nm absorption peaks ($r \geq 0.65$). These bands were assigned to CH3 second overtone of aliphatic hydrocarbons, C=O stretching of acids and ester, C-H stretching/bending of lipids, and C–H of aromatic molecules deriving from lipids (Workman & Weyer, 2012). The peaks found at 2144 and 2310 nm, were previously reported to correspond to the C=C and C–H stretch combination tone of cis unsaturated fatty acids (Cozzolino, Murray, Chree, & Scaife, 2005), thus suggesting that the unsaturated lipidic fraction of fish may be strongly influenced by the environmental conditions.

4. Conclusions

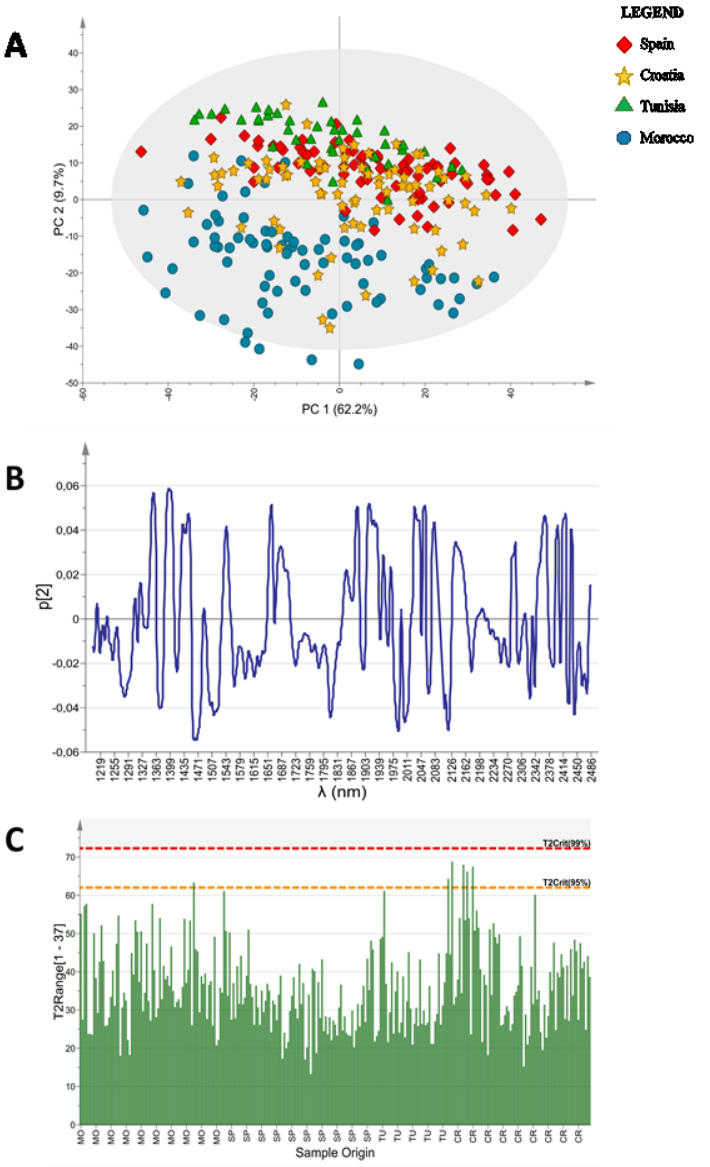
In the present work. information related to the geographical origin of the fish was found in specific regions of the 1200-2500 nm NIR spectra of both the semi-finished and the finished product. This information was successfully extracted by means of multi-class OPLS- DA and the discriminant models, whose sensibility, specificity, and accuracy values were optimal, worked well thanks the overall differences in lipid and protein pattern among classes. A more in-depth evaluation of the anchovy spectra fingerprints was further carried out, with a view to bringing to light the distinctive compositional marks related to each single provenance. Proteins, peptides, amino acids, and proteolytic compounds signatures distinguished anchovies from Morocco, while the lipid signature and,

presumably, the unsaturated lipid fraction closely characterised anchovies from Tunisia. At the same time, anchovies from Spain and Croatia were separated from the other groups, owing to the equal contribution of protein and lipid compounds. Although a more accurate characterisation of compounds involved in the discrimination of the anchovy origins by using auxiliary techniques would be helpful for the overall understanding of the NIR spectral changes observed, the present spectral fingerprinting strategy is particularly advantageous from the perspectives of rapidity, non-destructiveness, and cost-effectiveness in comparison to other well-known analytical approaches such as those based on chromatography and mass spectrometry. In addition, the proposed method, especially if transferred to handheld and portable NIR spectroscopy instrumentations, may be particularly suited for routine surveillance of transformed fish products, whose manufacturing industry is known to be characterized by extreme fast production rhythms. The authentication of anchovies is fundamental to ensure the traceability of the product along the food chain, the loss of which may have detrimental impact on the safety due to the possible exposure to not identified hazards and related risks.

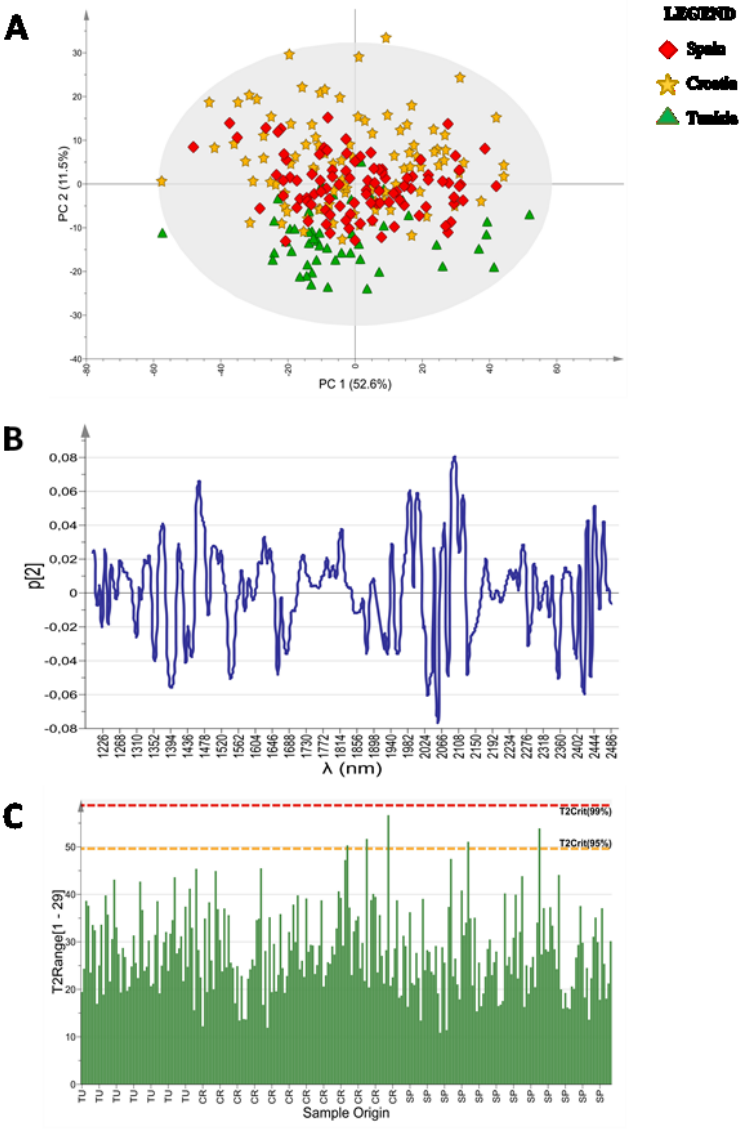
Acknowledgements

The research was carried out within the project "Development of rapid tools for anchovies authentication" funded by Rizzoli Emanuelli S.p.a. (Parma, Italy). The financial support and valuable cooperation of Rizzoli Emanuelli are gratefully acknowledged.

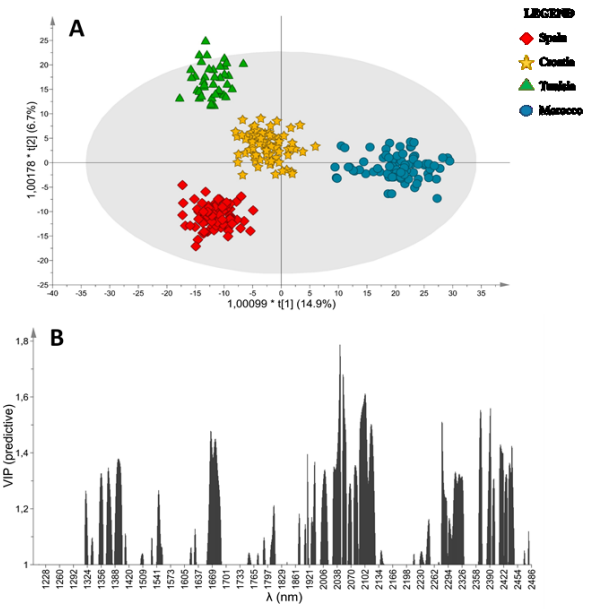
SUPPLEMENTARY MATERIALS



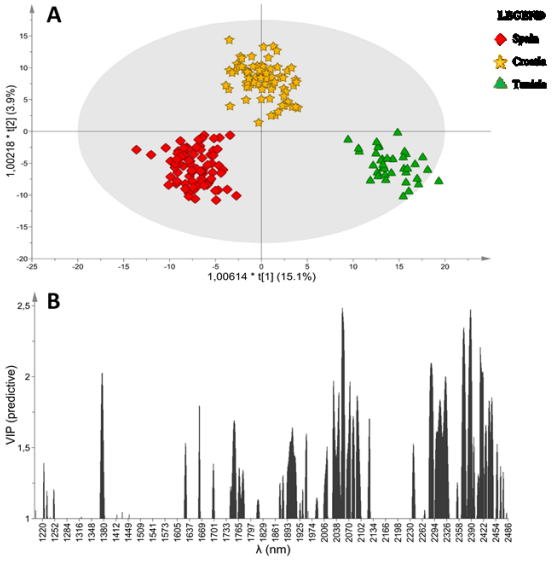
Supplementary Figure S1. Result from PCA computed on the dataset of the semi-finished anchovy samples. (A) Score scatter plot defined by PC 1 and PC 2; (B) Loading plot for the PC 2; (C) Hotelling's T^2 range plot of the distance from the origin in the score space for each selected observation.



Supplementary Figure S2. Result from PCA computed on the dataset of the finished anchovy samples. (**A**) Score scatter plot defined by PC 1 and PC 2; (**B**) Loading plot for the PC 2; (**C**) Hotelling's T^2 range plot of the distance from the origin in the score space for each selected observation.



Supplementary Figure S3. Distribution of the training sample scores (**A**) and variable importance in projection (VIP) plot for the predictive components (**B**), obtained by OPLS-DA modelling of the semi-finished anchovies according to the provenance areas.



Supplementary Figure S4. Distribution of the training sample scores (**A**) and variable importance in projection (VIP) plot for the predictive components (**B**), obtained by OPLS-DA modelling of the finished anchovies according to the provenance areas

References

Anggo, A. D., Ma, W. F., Swastawati, F., & Rianingsih, L. (2015). Changes of amino and fatty acids in anchovy (Stolephorus sp) fermented fish paste with different fermentation periods. *Procedia Environmental Sciences, 23*, 58–63. https://doi.org/10.1016/j.proenv.2015.01.009

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., & Trygg, J. (2007) OPLS Discriminant Analysis, Combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics, 20*, 341–351. https://doi.org/10.1002/cem.1006

Bocker, U., Kohler, A., Aursand, I. G., & Ofstad, R. (2008). Effects of Brine Salting with Regard to Raw Material Variation of Atlantic Salmon (Salmo salar) Muscle. *Journal of Agricultural and Food Chemistry, 56*, 5129–5137. https://doi.org/10.1021/jf703678z

Bruun, S. W., Søndergaard, I., & Jacobsen, S. (2007). Analysis of protein structures and interactions in complex food by near-infrared spectroscopy. 1. Gluten powder. *Journal of agricultural and food chemistry, 55*(18), 7234-7243. https://doi.org/10.1021/jf063680j

Carrera, M., & Gallardo, J. M. (2017). Determination of the Geographical Origin of All Commercial Hake Species by Stable Isotope Ratio (SIR) Analysis. *Journal of Agricultural and Food Chemistry, 65*(5), 1070–1077. https://doi.org/10.1021/acs.jafc.6b04972

Colombo, S. M., Wacker, A., Parrish, C. C., Kainz, M. J., & Arts, M. T. (2016). A fundamental dichotomy in long-chain polyunsaturated fatty acid abundance between and within marine and terrestrial ecosystems. *Environmental Reviews, 25*(2), 163-174. https://doi.org/10.1139/er-2016-0062

Cozzolino, D., Murray, I., Chree, A., & Scaife, J. R. (2005). Multivariate determination of free fatty acids and moisture in fish oils by partial least-squares regression and near-infrared spectroscopy. *LWT - Food Science and Technology, 38*(8), 821–828.
https://doi.org/10.1016/j.lwt.2004.10.007

Diraman, H., & Dibeklioglu, H. (2009). Chemometric characterization and classification of selected freshwater and marine fishes from turkey based on their fatty acid profiles. *JAOCS, Journal of the American Oil Chemists' Society, 86*(3), 235–246. https://doi.org/10.1007/s11746-008-1338-3

Drengstig, A., Fevolden, S. E., Galand, P. E. and Aschan, M. M. (2000). Population structure of the deep-sea shrimp (Pandalus borealis) in the north-east Atlantic based on allozyme variation. *Aquatic Living Resources, 13*(2), 121–128. https://doi.org/10.1016/S0990-7440(00)00142-X

El Sheikha, A. F., & Montet, D. (2016) How to Determine the Geographical Origin of Seafood? *Critical Reviews in Food Science and Nutrition, 56* (2), 306-317. https://doi.org/10.1080/10408398.2012.745478

EUMOFA (European Market Observatory for Fisheries and Aquaculture products) (2018). Processed anchovies in Italy. Price structure in the supply chain. https://doi.org/10.2771/181369

FAO (Food and Agriculture Organization of the United Nations) (2020). FAO Species Fact Sheets, Engraulis encrasicolus (Linnaeus, 1758). http://www.fao.org/fishery/species/2106/en. Accessed 12.09.20

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Forina, M., Armanino, C., Leardi, R., & Drava, G. (1991). A class-modelling technique based on potential functions. *Journal of Chemometrics, 5*(5), 435–453. https://doi.org/10.1002/cem.1180050504

Ghidini, S., Varrà, M. O., Dall'Asta, C., Badiani, A., Ianieri, A., & Zanardi, E. (2019). Rapid authentication of European sea bass (*Dicentrarchus labrax* L.) according to production method, farming system, and geographical origin by near infrared spectroscopy coupled with chemometrics. *Food Chemistry, 280*, 321–327. https://doi.org/10.1016/j.foodchem.2018.12.075

Ghidini, S., Varrà, M. O., & Zanardi, E. (2019). Approaching Authenticity Issues in Fish and Seafood Products by Qualitative Spectroscopy and Chemometrics. *Molecules, 24*(9), 1812. https://doi.org/10.3390/molecules24091812

Guo, X., Cai, R., Wang, S., Tang, B., Li, Y., & Zhao, W. (2018). Non-destructive geographical traceability of sea cucumber (Apostichopus japonicus) using near infrared spectroscopy combined with chemometric methods. *Royal Society Open Science, 5*(1). https://doi.org/10.1098/rsos.170714

Hernandez-Herrero, M. M., Roig-Sagues, A. X., Lopez-Sabater, E. I., Rodriguez-Jerez, J. J., & Mora-Ventura, M. T. (2002). Influence of raw fish quality on some physicochemical and microbial characteristics as related to ripening of salted anchovies (*Engraulis encrasicholus L*). *Journal of Food Science, 67*(7), 2631–2640. https://doi.org/10.1111/j.1365-2621.2002.tb08790.x

Huang, Y. H., Cavinato, A. G. C., Mayes, D. M. M., Kangas, L. J. K., Bledsoe, G. E. B., & Rasco, B. A. R. (2003). Nondestructive Determination of Moisture and Sodium Chloride in Cured Atlantic Salmon (Salmo salar) (Teijin) Using Short-wavelength. *Journal of Food Science, 68*(2), 482–486. https://doi.org/10.1111/j.1365-2621.2003.tb05698.x

Huang, Y., Rogers, T. M., Wenz, M. A., Cavinato, A. G., Mayes, D. M., Bledsoe, G. E., & Rasco, B. A. (2001). Detection of Sodium Chloride in Cured Salmon Roe by SW - NIR Spectroscopy. *Journal of Agricultural and Food Chemistry, 46*(6), 4161–4167. https://doi.org/doi.org/10.1021/jf001177f

Iizuka, K., & Aishima, T. (1997). Soy Sauce Classification by Geographic Region Based on NIR Spectra and Chemometrics Pattern Recognition. *Journal of Food Science, 62*(1), 101–105. https://doi.org/10.1111/j.1365-2621.1997.tb04377.x

Kumari, N., Grimbs, A., D'Souza, R. N., Verma, S. K., Corno, M., Kuhnert, N., & Ullrich, M. S. (2018). Origin and varietal based proteomic and peptidomic fingerprinting of Theobroma cacao in non-fermented and fermented cocoa beans. *Food Research International, 111*, 137–147. https://doi.org/10.1016/j.foodres.2018.05.010

Liu, L., Cozzolino, D., Cynkar, W. U., Dambergs, R. G., Janik, L., O'Neill, B. K.,

Colby, C. B., & Gishen, M. (2008). Preliminary study on the application of visible-near infrared spectroscopy and chemometrics to classify Riesling wines from different countries. *Food Chemistry*, *106*(2), 781–786. https://doi.org/10.1016/j.foodchem.2007.06.015

Liu, Y., Ma, D. H, Wang, X. C, Liu, L. P, Fan, Y. X, & Cao, J. X. (2015). Prediction of chemical composition and geographical origin traceability of Chinese export tilapia fillets products by near infrared reflectance spectroscopy. *LWT - Food Science and Technology*, *60*(2), 1214–1218. https://doi.org/10.1016/j.lwt.2014.09.009

Locci, E., Piras, C., Mereu, S., Cesare Marincola, F., & Scano, P. (2011). 1H NMR metabolite fingerprint and pattern recognition of mullet (Mugil cephalus) bottarga. *Journal of Agricultural and Food Chemistry*, *57*(17), 9497–9505. https://doi.org/10.1021/jf2012979

Martin, R., Carter, E., Flick, J., & Davis, L. (2000). Marine & Freshwater Products Handbook. (R. Martin, E. Carter, J. Flick, & L. Davis, Eds.). Boca Raton, FL., USA: CRC Press (Taylor & Francis group). https://doi.org/https://doi.org/10.1201/9781482293975

Massart, D.L., Vandeginste, B.G.M., Buydens, L.M.C., Jong, S.D., Lewi, P.J., & SmeyersVerbeke, P. (1997). Handbook of Chemometrics and Qualimetrics: Part A. Elsevier, Amsterdam,. ISBN 10: 0444897240

Mork, J. & Giæver, M. (1999). Genetic structure of cod along the coast of Norway: Results from isozyme studies. *Sarsia,* *84*(2),157–168. https://doi.org/10.1080/00364827.1999.10420442

Öksuz, A., Özyilmaz, A., & Turan, C. (2009). Comparative study on fatty acid profiles of anchovy from Black Sea and Mediterranean Sea (Engraulis encrasichoius L., 1758). *Asian Journal of Chemistry*, *21*(4), 3081–3086.

Pillonel, L., Schaller, E., Picque, D., Cattenoz, T., & Bosset, J. (2005). The potential of combined infrared and fluorescence spectroscopies as a method of determination of the geographic origin of Emmental cheeses. *International Dairy Journal, 15*, 287–298. https://doi.org/10.1016/j.idairyj.2004.07.005

Regulation (EU) No 1379/2013 of the European Parliament and of the Council of 11 December 2013 on the common organisation of the markets in fishery and aquaculture products, amending Council Regulations (EC) No 1184/2006 and (EC) No 1224/2009 and repealing Council Regulation (EC) No 104/2000. *Official Journal of the European Union, L354*, 1–21.

Ren, G., Wang, S., Ning, J., Xu, R., Wang, Y., Xing, Z., Wan, X., & Zhang, Z. (2013). Quantitative analysis and geographical traceability of black tea using Fourier transform near-infrared spectroscopy (FT-NIRS). *FRIN, 53*(2), 822–826. https://doi.org/10.1016/j.foodres.2012.10.032

Ricardo, R., Pimentel, T., Moreira, A. S., Rey, F., Coimbra, M. A., Domingues, M. R., Domingues, P., Costa Leal, M., & Calado, R. (2015). Potential use of fatty acid profiles of the adductor muscle of cockles (Cerastoderma edule) for traceability of collection site. *Scientific Reports, 5*, 11125. https://doi.org/10.1038/srep11125

Robert, P., Devaux, M. F., Mouhous, N., & Dufour, E. (1999). Monitoring the

secondary structure of proteins by near-infrared spectroscopy. *Applied spectroscopy, 53*(2), 226-232. https://doi.org/10.1366/0003702991946361

Rodríguez-Ortega, M. J., Grøsvik, B. E., Rodríguez-Ariza, A., Goksøyr, A., & López-Barea, J. (2003). Changes in protein expression profiles in bivalve molluscs (Chamaelea gallina) exposed to four model environmental pollutants. *Proteomics, 3*(8), 1535-1543. https://doi.org/10.1002/pmic.200300491

Romotowska, P. E., Karlsdóttir, M. G., Gudjónsdóttir, M., Kristinsson, H. G., & Arason, S. (2016). Seasonal and geographical variation in chemical composition and lipid stability of Atlantic mackerel (Scomber scombrus) caught in Icelandic waters. *Journal of Food Composition and Analysis, 49*, 9–18. https://doi.org/10.1016/j.jfca.2016.03.005

Salla, V., & Murray, K. K. (2013). Matrix-assisted laser desorption ionization mass spectrometry for identification of shrimp. *Analytica Chimica Acta, 794*, 55–59. https://doi.org/10.1016/j.aca.2013.07.014

Shenk, J. S., Workman, J. J., & Westerhaus, M. O. (2001). Application of NIR spectroscopy to agricultural products. In: D. A. Burns & E. W. Ciurczak (Eds.), Handbook of near-infrared analysis 3rd Ed., Practical Spectroscopy Series, 35, 356–357, Marcel Dekker, Inc., New York.

Shepard, J. L. & Bradley, B. P. (2000). Protein expression signatures and lysosomal stability in Mytilus edulis exposed to graded copper concentrations. *Marine Environmental Research, 50*(1), 5457–463. https://doi.org/10.1016/S0141-1136(00)00119-7

Standal, I. B., Rainuzzo, J., Axelson, D. E., Valdersnes, S., Julshamn, K., & Aursand, M. (2012). Classification of geographical origin by PNN analysis of fatty acid data and level of contaminants in oils from Peruvian anchovy. JAOCS, Journal of the American Oil Chemists' Society, 89(7), 1173–1182. https://doi.org/10.1007/s11746-012-2031-0

Svensson, V. T., Nielsen, H. H., & Bro, R. (2004). Determination of the protein content in brine from salted herring using near-infrared spectroscopy. *LWT - Food Science and Technology, 37*(7), 803–809. https://doi.org/10.1016/j.lwt.2004.03.004

Takamura, H., Hyakumoto, N., Endo, N., Matoba, T., & Nishiike, T. (1995). Determination of lipid oxidation in edible oils by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 3*(4), 219–225. https://doi.org/10.1255/jnirs.72

Trygg, J., & Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics, 16*(3), 119–128. https://doi.org/10.1002/cem.695

Van der Voet, H. (1994). Comparing the predictive accuracy of models using a simple randomioation test. *Chemometrics and Intelligent Laboratory Systems, 25*(2), 313–323. https://doi.org/10.1016/0169-7439(94)85050-X

Wang, J., Sowa, M. G., Ahmed, M. K., & Mantsch, H. H. (1994). Photoacoustic near-infrared investigation of homo-polypeptides. *Journal of Physical Chemistry, 98*(17), 4748–4755. https://doi.org/10.1021/j100068a043

Wang, J., Kliks, M. M., Qu, W., Jun, S., Shi, G., & Li, Q. X. (2009). Rapid

determination of the geographical origin of honey based on protein fingerprinting and barcoding using MALDI TOF MS. *Journal of Agricultural and Food Chemistry, 57*(21), 10081–10088. https://doi.org/10.1021/jf902286p

Westad, F., & Marini, F. (2015). Validation of chemometric models–a tutorial. *Analytica chimica acta, 893*, 14-24. https://doi.org/10.1016/j.aca.2015.06.056

Workman, Jr.J, & Weyer, L. (2012). Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy. Journal of Chemical Information and Modeling (2nd ed.,). Boca Raton, FL., USA, CRC Press (Taylor & Francis group). https://doi.org/10.1017/CBO9781107415324.004

Zhang, X., Liu, Y., Li, Y., & Zhao, X. (2017). Identification of the geographical origins of sea cucumber (Apostichopus japonicus) in northern China by using stable isotope ratios and fatty acid profiles. *Food Chemistry, 218*, 269–276. https://doi.org/10.1016/j.foodchem.2016.08.083

SECOND SECTION

CHAPTER 4

Review Article

Advances in Troubleshooting Fish and Seafood Origin Mislabelling by Inorganic Elemental Composition

Maria Olga Varrà [a], Sergio Ghidini [a], Lenka Husáková [b], Adriana Ianieri [a], and Emanuela Zanardi [a]*

[1] Department of Food and Drug, University of Parma, Strada del Taglio 10, 43126 Parma, Italy.

[2] Department of Analytical Chemistry, Faculty of Chemical Technology, University of Pardubice, Studentska 573 HB/D, Pardubice, CZ-532 10, Czech Republic

*Correspondence: emanuela.zanardi.unipr.it; Tel.: +39-0521-902-760

Under Review in Foods

Abstract: To date, the demand for fish and seafood is growing worldwide. Meanwhile, problems related to the integrity and safety of the fishery sector are increasing, leading legislators, producers, and consumers to question about the way to effectively protect themselves from fraud and health hazards related to fish consumption. What is urgently required now is the availability of reliable, truthful, and reproducible methods assuring the correspondence between the real nature of the product and label declarations accompanying the same product during its market life. The evaluation of the inorganic composition of fish and seafood appears to be one of the most promising strategies to be exploited in the near future to assist routine and official monitoring operations along the supply chain.

The present review article is focused on exploring the latest scientific achievements of using the multi-elemental composition of fish and seafood as an imprint of their authenticity and traceability, especially with regards to the geographical origin. The scientific production of the last ten years focusing on the analytical determination and statistical elaboration of elemental data (alone or in combination with methodologies targeting other compounds) to verify the identity of fishery products has been summarised and discussed.

Abbreviations: AAS=Atomic absorption spectroscopy; AES=atomic emission spectroscopy; AFS=atomic fluorescence spectroscopy; ANNs=artificial neural networks; CA=cluster analysis; CDA=canonical discriminant analysis; EDXRF=energy dispersive X-ray fluorescence spectroscopy; ICP-AES=inductively coupled plasma atomic emission spectroscopy; ICP-MS=inductively coupled plasma mass spectrometry; k-NN= k-nearest neighbours; LDA=linear discriminant analysis; NNS=neural network bagging; PCA=principal component analysis; PLS-DA=partial least square discriminant analysis; PNNs=probabilistic neural networks; QDA=quadratic discriminant analysis; RF=random forest; SIMCA=soft independent modelling of class analogy; S-LDA=stepwise-linear discriminant analysis; XRF=X-ray fluorescence spectroscopy.

1. Introduction

The verification of food authenticity and integrity is a complex topic which has come under the spotlight in recent years. This issues involves many different aspects, from the identification of mislabelling and misrepresentation to adulteration and contamination of the product. The traceability of fish and seafood and the detection of intentional or unintentional fraud is today a challenging task as the supply chain of fishery products is among the most diversified and globalised ones. As a matter of fact, fish is currently amongst the most frequently misdescribed foodstuffs worldwide, to a point that almost 20 percent of fish in the sail and restaurant sectors of fifty-five countries has been recently found to be misdescribed [Warner, Mustain, Lowell, Geren, & Talmage). Specifically, the major economic losses affecting the sector today derive from the substitution of highly valuable fish and seafood species for morphologically similar but lower-quality ones and from the increasingly common falsification in relation to the geographical origin. Albeit these fraudulent practices seem to have a negative impact only from an economical point of view, some health implications may arise, for example, from the replacement of certain fish species with cheaper but potentially poisonous ones (Giusti et al., 2019) or from the sale of illegally-caught fish originating from polluted areas (Krusche & Hanel, 2021)

In order to prevent fraud, protect producers and consumers, and promote high-quality fish products, the reinforcement of the international food monitoring programmes is not sufficient. Indeed, control measures would be required to be undertaken in synergy with the implementation of proper vulnerability assessment systems and the development rapid analytical tools, so as to confidently verify whether a product is genuine or counterfeit and guarantee the integrity of the whole production chain.

Many different biological and chemical methods have been developed over the years to ascertain the authentic nature of a wide range of foodstuffs. These methods today focus on the evaluation of the organic (DNA, proteins, lipids, sugars, and/or metabolites) and inorganic (elements, isotope ratios) fractions of food and exploit the principles of chromatography mass-spectrometry, and spectroscopy to identify in a targeted way few or

multiple compounds acting as secondary markers of authenticity (Ballin & Laursen, 2019). In this sense, the determination of the inorganic multi-elemental signatures (in terms of major, trace, and ultra-trace elements), accompanied by multivariate statistics, is increasingly applied to authenticate different foods of animal origin such as honey (Zhou, Taylor, Salouros, & Prasad, 2018), pork meat (Kim et al., 2017), and cheeses (Magdas et al., 2019), especially in relation to the geographical origin and the method of production. In this context, the evaluation of the elemental content of fish and seafood is particularly advantageous since it may allow simultaneously monitoring mislabelling as well as the maximum acceptable regulation limits for certain toxic elements established at the European level (European Commission, 2006), thus pursuing both integrity and safety objectives.

Elements found in fish tissue are scientifically recognised to be a reflection of the elemental composition of the overall surrounding environment, from the aquatic habitat to the production premises, and this is particularly advantageous when the country of origin of wild specimens is thought to be identified by using the elemental composition. On the contrary, the elemental content of farmed specimens is inevitably affected, both from a qualitative and a quantitative point of view, by feeding stuffs. Thus, using the elemental profile of the tissues of farmed specimen to trace back to their country of origin may be problematic by virtue of the fact that the same feed can be traded internationally and given to fish cultured in different parts of the world.

When working with the element composition of fish for authentication purposes, it should be taken into consideration that the presence of elements in the aquatic environment explored by the fish during life is not only dependent on the specific geochemical characteristics of the habitat, but it may be significantly influenced by other environmental factors either of natural (such as climate, water temperature, salinity, age and sexual maturity of the animal) or anthropic origin (i.e. the exogenous pollution) (Danezis, Tsagkaris, Camin, Brusic, & Georgiou, 2016). In addition, after catch, fishery products are more handled and enlivened compared to other foodstuffs; therefore, the likelihood of unwanted and misleading elements

being incorporated as contaminants throughout the whole production cycle increases many times over.

For these reasons, the overall elemental signature needs to be strictly evaluated before being used as a tool to address authenticity problems of fish and seafood. In this regard, chemometrics and machine learning have now become an essential support for increasing the strength and the reliability of high-throughput analytical techniques. As a matter of fact, the advanced statistical elaboration of elemental data has already been proved to be a straightforward and effective means to study elements' behaviour, identify common but hidden compositional characteristics among similar food samples, separate complementary, opposite, or redundant information enclosed into elemental data, define classification rules and simplify the overall methodology by extracting the effectively significant elemental markers for classification (Callao & Ruisánchez, 2018).

The present review article was aimed at discussing the last ten-years applications and advances in data mining of the multi-elemental profile of fish, molluscs, echino-derms, and crustaceans as a strategy to verify whether mandatory labelling information matches the identity of these products. The survey took into consideration the elemental measurements performed only on edible tissues of fishery products, which, albeit being more rapidly subjected to variations induced by environment compared to hard structures such as otoliths, statoliths, skeleton, and scales, are retained in the final traded products, hence potentially monitorable in every phase of the production chain. As evidenced be-low, only raw products were discussed, since, as far as we know, no considerable breakthroughs in tracing and authenticating transformed (e.g. salted, smoked, marinated) fish and seafood products have been achieved.

2. Analytical and Chemometric Methodologies for Element and Stable-Isotope Analysis of Fish and Seafood

Various analytical techniques have been used for the purpose of the determination of the elemental content of fish and seafood throughout the last years (Danezis, Tsagkaris, Camin, & Brusic, 2016; Danezis, Tsagkaris, Brusic, & Georgiou, 2016; Drivelos & Georgiu, 2012; Gopi, Mazumder,

Sammut, & Saintilan, 2019). Among these**,** atomic spectroscopic methods such as atomic absorption spectrometry (AAS) (Adebiyi, Ore, & Ogunjimi, 2020; Jinadasa, Chathurika, Jayasinghe, & Jayaweera, 2019; Stancheva, Makedonski, Peycheva, 2014; Santos, Trigueiro, Lemos, & Da Nóbrega Furtunato, 2013; De Andrade et al., 2017; Chaguri et al., 2015; Skałecki et al., 2020) with flame (Chaguri et al., 2015; Skałecki et al., 2020)
or electrothermal atomisation (Santos et al., 2013; De Andrade et al., 2017), atomic fluorescence spectrometry (AFS), inductively coupled optical emission spectrometry (ICP-OES) (Li et al., 2017), inductively coupled plasma mass spectrometry (ICP-MS) (Liu et al., 2013; Costas-Rodríguez, Lavilla, & Bendicho, 2010), and X-ray fluorescence (XRF) (Chaguri et al., 2015; Gopi et al., 2019) have been the most frequently employed ones. On the contrary, **e**lectroanalytical (Melucci, Casolari, De Laurentiis, & Locatelli, 2017) or neutron-activation-based techniques such as neutron activation analysis (NAA) have been used to a lesser extent (Moon et al., 2015; Fabiano et al., 2016; Rentería-Cano et al., 2011). These techniques offer specific advantages and, at the same time, present some limitations which make their application preferable in some cases but not in others. The main characteristics and performances of the analytical methods that can be used for comparing the multi-element or stable-isotope composition of fish and seafood samples are examined in the text below.

In the case of major and some minor elements, AAS and OES with flames (flame atomic absorption spectroscopy, FAAS, and flame optical emission spectroscopy, FOES), are still valuable and well-established techniques routinely and customarily applied in the area of fish and seafood analysis due to its robustness in relation to interferences and sample introduction problems, selectivity, straightforward, and lower cost (Skałecki et al., 2020; Brown & Milton, 2005). On the other hand, these methods still present limitations related to sensitivity, thus electrothermal AAS (ET-AAS) (Santos et al., 2013), hydride generation AAS (HG-AAS) (Rasmussen, Hedegaard, Larsen, & Sloth, 2012) and cold vapour AAS (CV-AAS) or direct thermal decomposition AAS (Panichev & Panicheva, 2015) are employed in the lower concentration range. However, the main disadvantage is that AAS is primarily limited to the determination of metallic elements and, in addition,

it is a single-element technique with the linear range typically less than two orders of magnitude. Despite used in only in a very limited number of cases (De Andrade et al., 2018), high-resolution continuum source AAS (HR-CS-AAS) is overcoming some of the limitations of AAS as it allows for the simultaneous evaluation of several absorption lines in the selected spectral range, accurate background correction, and the determination of non-metals. ICP-OES is by far the most applied technique for the analysis of food samples (Stancheva et al., 2014; De Andrade et al., 2018; Li et al., 2013; Li et al., 2017) because it offers simultaneous multi-element measurement, capabilities for sensitive determination of refractory elements, quantification of non-metals, and high analytical throughput. Microwave plasma optical emission spectrometry (MP-OES) using the magnetically excited microwave plasma source has also found recent applications to fish and seafood (De Sá et al., 2020; Ríos, Peñuela, & Botero, 2017), mainly because characterised by detection limits down to sub-ppb levels, significant cost reduction, and simpler spectra than ICP-OES. However, at present, both ICP-OES and MIP-OES cover with difficulty the needs required in routine applications, especially when the determination of elements at trace or ultra-trace concentrations is in demand.

ICP-MS is better suited to meet this task and is currently a frontline technology rapidly replacing other methods in many fields of food science. Unsurpassed advantages such as high sensitivity, selectivity, wide dynamic concentration range up to 11 orders of magnitude, high sample throughput, and multi-analyte capabilities make this method an ideal candidate for food authentication studies since they might facilitate discrimination and classification of samples (Zmozinski et al., 2014).

More detailed technical aspects of the above-mentioned methodologies can be retrieved from literature (Bulska & Ruszczyńska, 2017; Yeung, Miller, & Rutzke, 2017).

The analysis of biologic matrices such as foodstuffs by atomic and mass spectrometry methods, especially at trace and ultra-trace levels, is often a difficult and challenging task. As a matter of fact, a quite complex biological matrix poses problems related not only to sample heterogeneity, selection of

proper sample treatment, and decomposition, but also matrix interferences as well.

In the ICP-MS analysis of fish and seafood samples, both spectral and non-spectral interferences are expected to be encountered (De Andrade et al., 2018; Liu et al., 2012; Sneddon & Vincent, 2008). While non-spectral effects can be easily overcome using a proper calibration strategy including the use of an internal standard (De Andrade et al., 2018; Liu et al., 2012), standard additions, and/or the isotope dilution, spectral effects due to the overlaps by different polyatomic ions (formed from the combination of species derived from the matrix elements, plasma gas and sample solvents) are more serious and difficult to handle (Lum & Leung, 2016).

High-resolution mass spectrometers with the sector field mass analyser could be the ideal solution to bypass most of these problems; however, owing to their high price, these instruments are not easily accessible for most of the laboratories.

Time-of-flight (TOF)-ICP-MS instruments with all the advantages, such as the fast simultaneous multi-elemental analysis, improved precision of measurements of the isotope ratios, the very low volume of the sample needed for the analysis, and the tolerance to higher salinity of samples, do not however have the adequate resolution to eliminate the spectral interferences typically encountered when analysing biological samples. Then, mathematical corrections must be employed, but this approach is less effective when performing trace analysis (Husáková et al., 2011). The absence for effective solutions related to the control of problematic spectral effects, which were not accessible to users until recently, has limited the widespread diffusion of this technique in routine practice. Nevertheless, there is an increasing trend in resorting to the use of collision cell technology for interference management during sample analysis also in the TOF-ICP-MS current instrumentation (Burger et al., 2011). At the present time, the quadrupole-based ICP-MS equipped with a collision/reaction cell (CRC) for the elimination of spectral interferences is the most popular ICP-MS instrumentation on the market. In the reaction cell mode interfering ions are removed by the transformation into different species or uncharged atoms or

molecules through specific chemical reactions with a supplementary reaction gas (H_2, NH_3, O_2, N_2O or CH_4) (Lum et al., 2016).

Despite this approach is more efficient for the removal of known spectral interferences, it may lead to a formation of new unwanted interfering polyatomic ions. The collision cell mode is instead more suitable for the multi-elemental analysis of unknown samples. For this purpose, He is widely used as a collision gas to slow down polyatomic interfering ions to a larger extent than the atomic analyte ions, such that the former could be selectively discriminated against on the basis of their lower kinetic energy.

With the introduction of an ICP-tandem mass spectrometer (MS/MS, often referred to as triple quadrupole ICP-MS or ICP-QQQ) the CRC technology in quadrupole-based ICP-MS has been greatly improved (Husáková et al., 2011). This instrumentation, equipped with CRC located between two quadrupole mass filters, provides an elegant approach via a precursor ion and/or product ion scanning to solve even the most challenging cases of spectral overlap and interference. Moreover, it can determine a wider range of analytes at much lower concentrations with greater reliability and higher confidence (Balcaen et al., 2015). In addition to total element determinations, the current ICP-MS instrumentation is suited also to isotope ratio analysis, even if the isotope ratio precision is strictly dependent on the type and the design of the instrument used. Considering that the simultaneous measurement of multiple isotopes provides a better precision in isotope ratio measurement, the use of TOF-ICP-MS or multi-collector mass spectrometer with a plasma source for ionisation (MC-ICP-MS) is considerably more advantageous than the use of a single quadrupole ICP-MS for isotope analysis. However, the commonly used mass spectrometers typically do not provide the sensitivity and precision required for the determination of light isotopes ratios. In addition, they are susceptible to isotopic fractionation (mass bias). Therefore, isotope ratio mass spectrometry (IRMS) (Zhao, Liu, Li, Zhang, & Qi, 2018; Molkentin, Lehmann, Ostermeyer, & Rehbein, 2015), nuclear magnetic resonance (NMR) (Remaud et al., 2018; Aursand, Mabon, & Martin, 2020), and thermal ionisation mass spectrometry (TIMS) (Drivelos et al., 2012) are more suitable for this purpose.

Atomic fluorescence spectrometry (AFS) may represent an alternative to the other atomic and mass spectrometric techniques as it provides low detection limits, wide linear calibration range, simplicity, lower acquisition and running costs. These analytical features make AFS superior to AAS and equal to ICP-MS or ICP-OES, especially in speciation studies, as long as single element speciation studies are considered (Zmozinski et al., 2014).

Recently there has been an increase in the application of non-destructive multielement methods for analysis of seafood samples (Gopi et al., 2019; Fabiano et al., 2016; Farabegoli et al., 2018). Methods based on X-ray spectrometry such as X-ray fluorescence (XRF) (Chaguri et al., 2015; Gopi et al., 2019a; Gopi et al., 2019b), energy dispersion-XRF (ED-XRF) (Chaguri, 2015), proton induced X-ray emission (PIXE), total reflection X-ray fluorescence spectrometry (TXRF), synchrotron X-ray fluorescence (SXRF), as well as methods based on X-ray microanalysis offer several benefits (Machado et al., 2020).

Among these, the selective detection and sensitivity (about $\mu g\ g^{-1}$ and below) for most of the elements (Drivelos et al., 2012; Gopi et al, 2019b), minimal sample preparation, high sample throughputs, and accuracy in quantification are worth to be mentioned (Machado et al., 2020). In addition, field portable-XRF analysers are becoming increasingly popular for a wide variety of elemental analysis applications (Machado et al., 2020).

Laser-based techniques also play an important role for direct analysis of solid samples and in the last years they are increasingly present in the food industry. Laser-induced breakdown spectrometry (LIBS) is considered a promising micro-destructive food analysis tool for rapid qualitative and quantitative chemical analysis (Machado et al., 2020; Markiewicz-Keszycka et al., 2017). However, direct analysis of samples with complex organic matrices such as fresh food products is not easy (Machado et al., 2020). As a matter of fact, it is often not possible to analyse the sample without any preparation, since the results might be misleadingly affected by any inhomogeneity of the material. On the other hand, the sample preparation for LIBS analysis is minimal when compared to reference methods such as AAS or ICP-MS. The major limitation of LIBS for practical applications

results from its reduced sensitivity for minor mineral elements and heavy metals with very low concentrations in a complex organic matrix.

The connection of laser ablation (LA) with ICP-MS (Pozebon et al., 2014; Dunphy et al., 2011; Flem et al., 2017) represents a quite versatile analytical tool, offering the fastest analytical speed compared to all the other techniques, limit of detection approaching ppb levels, the capability for performing bulk analysis, depth profiling, and elemental/isotope mapping (Drivelos et al., 2012). Nevertheless, LA-ICP-MS still lacks sufficiently matrix-matched reference materials for each considered matrix type and the analysis accuracy is restricted by several factors, such as sensitivity drift, elemental/isotopic fractionation, matrix effects, etc. (Machado et al., 2020; Limbeck, Bonta, & Nischkauer, 2017).

Electrothermal vaporisation (ETV) is also an efficient and powerful approach for a bulk analysis where solid samples can be directly turned into aerosols (Machado et al., 2020; Limbeck, Bonta, & Nischkauer, 2017). This strategy significantly boosts ICP-MS quantitative applications in desired field (Tormen et al., 2012).

2.1 Sample Digestion Procedures for Elemental Analysis

The market of most of the above-mentioned analytical apparatus, such as AAS, AFS and those which make use of a plasma source for ionisation, offers mainly instrumentation dedicated to the analysis of liquid samples. By consequence, digestion procedures for solid samples are necessarily required. Furthermore, sample preparation is a crucial issue for food products due to their inhomogeneity and matrix complexity.

Nowadays, the most used and useful digestion technique for a wide range of analytes and sample matrices is the high-pressure digestion using a closed-vessel microwave system (Santos et al., 2013; De Andrade et al., 2017; Costas-Rodríguez et al., 2010; Zmozinski et al., 2014; Bulska et al., 2017; Bizzi et al., 2017). This technique increases the sample throughput, minimises analyte losses during the decomposition, reduces both contamination risk (especially for trace analytes) and consumption of reagents, and is more effective, resulting in low residual carbon content of digested samples as well (Bizzi et al., 2017). In addition to closed-vessel microwave digestion

also high-pressure digestion involving opened vessels or classical dry-ashing digestion is generally performed. In wet-acid digestion, HNO_3 alone (Stancheva et al., 2014; Skałecki et al., 2020; Costas-Rodríguez et al., 2010) or combined with H_2O_2 (Santos et al., 2013; De Andrade et al., 2017; Zmozinski et al., 2014) and, occasionally, HCl (Ríos et al., 2017) or $HClO_4$ (Li et al., 2017; Li et al., 2013) are the most commonly used reagents. However, several novel approaches or adaptations to established procedures for sample preparation have been recently introduced. In particular, what's now emerging is a growing interest towards the use or diluted and non-hazardous analytical reagents, in accordance green chemistry and the need to reduce the negative impact of chemical analyses on the environment (Bizzi et al., 2017).

From this standpoint, ultrasound-assisted extraction or microwave-assisted extraction seems to be very promising approaches for sample preparation in the near future, allowing to optimise working times and consumption of analytical reagents.

2.2 Multivariate Data Analysis and Machine Learning

The growing interest towards high-throughput element-based methods to characterise foodstuffs may be partly justified by the efforts in the field of multivariate data analysis and machine learning, which have significantly simplified data handling and improved the identification of food fraud. Multivariate qualitative methods are well established in the field of analytical chemistry oriented towards the authenticity and adulteration verification of foodstuffs and the development of new algorithms for classification is continuously increasing (Callao et al., 2018). Despite this, analysis of the literature revealed that the statistical analysis of the multi-elemental profile of fish is mostly limited to the classical use of principal component analysis (PCA) and cluster analysis (CA) as exploratory (unsupervised) tools. As for samples' classification purposes, hard modelling of data based on linear discriminant analysis (LDA) and canonical discriminant analysis (CDA) has been more frequently employed (see Table 1). This is probably due to the fact that the theoretical background of these data elaboration techniques is more consolidated among the

scientific community compared to other more modern hard-modelling discriminant techniques such as partial least square discriminant analysis (PLS-DA), and soft-modelling techniques as soft independent modelling of class-analogy (SIMCA) (Marini, 2010). In addition, the applied methodologies appear to lack of proper validation protocols to be followed, which are necessary for the development of reliable and transferable multivariate-based models for foods classification (López et al., 2014).

Various techniques including K-nearest neighbour (KNN), K-mean clustering and artificial neural network (ANN) are crucial for future successful development of prediction models to food authentication solutions.

For further details on chemometrics and machine learning techniques applied to food science, the reader may refer to literature (Callao et al., 2018; Jiménez-Carvelo et al., 2019).

3. Authentic elemental signature of fish and seafood

As discussed below, authentication and traceability studies were often performed by coupling elemental analysis (major, trace, and/or ultra-trace elements) with other techniques targeting other compounds, with the objective to increase the specificity of discrimination and obtain better results.

The merging of data from isotopic analysis of light (H, C, N, O and S) or heavy elements (Sr, Pb) and from elemental analysis has been the most frequently investigated analytical strategy to approach traceability problems of fish and seafood.

The rationale behind this research trend over the years laid in the strong correlation between any variation in isotope fractionation (ratio between isotopes of a specific elements) and the geological, pedological, and wheatear characteristics of a given geographical area (Kelly, Brodie, & Hilkert, 2018). Among these, the isotopic distribution of light elements such as O ($\delta^{18}O$, $^{18}O/^{16}O$) and H ($\delta^{1}H$, $^{1}H/^{2}H$) in fishery products is influenced by the original isotopic distribution of the same elements in the water basin from which the fish come from, which, in turn, is the reflection of the isotopic distribution in the rainfall of the specific area (Kelly, Heaton, &

Hoogewerff, 2005). More, the isotope ratio of C ($\delta^{13}C$, $^{13}C/^{12}C$) in fish tissues may be related to the type of vegetation eaten by the fish during its life, namely that the plants are characterized by a C3, C4, or Crassulacean Acid Metabolism (CAM) photosynthetic metabolism. Considering that each type of these plants typically grows at certain latitudes, the isotopic distribution of C may be, at first instance, indirectly exploited as a marker of origin. Since fractionation of C is expected to vary between artificial feed used to rare aquaculture fish and the natural food of wild fish, its isotopic ratio may also be exploited to distinguish the production method of fish [61]. Indeed, isotopes of N are good indicators of the feeding regime of fish and of the position occupied by the fish in the food chain, thus being ideal markers of the production methods. Wild fish at higher trophic levels is in fact characterized by a greater enrichment in $\delta^{15}N$ ($^{15}N/^{14}N$), as well as $\delta^{15}N$ enrichment in artificial feeding given to farmed fish is expected to be significantly different compared to those present into the natural food eaten by wild fish [Kelly et al., 2005)

In the present review, recent research in the field of multi-elemental analysis applied to edible tissues of fish and seafood has been taken into consideration and reviewed in what follows. The scientific literature herein includes the research articles pertinent to the topic of the present review and published in the period 2010 through 2020. Articles were retrieved from Web of Science and Scopus databases (Search terms: 'fish', 'seafood', 'authentication', 'elemental analysis', 'elemental profile', 'elemental fingerprinting', 'chemometrics').

For the sake of clarity, the next paragraphs have been structured to enclose the same type of product, therefore fish, molluscs (both bivalve and cephalopods), crustaceans, and echinoderms have been discussed separately. The most frequently measured elemental markers of both geographical origin and method of production, retrieved from the reviewed scientific literature discussed below, are graphically shown in the radial bar chart reported in Fig. 1. For a quick comparison, a summary overview of the methodological and technical aspects of the published works is given in Table 1.

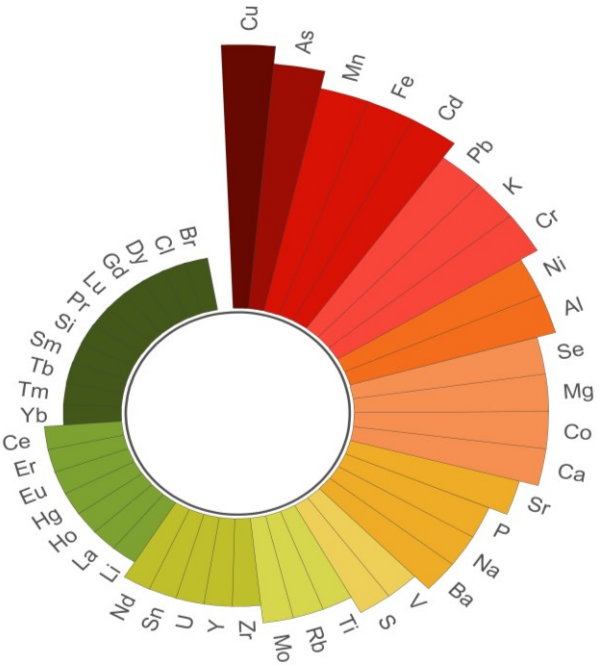


Figure 1. Radial bar chart showing the most widely used elemental markers in the last ten-years works dealing with authenticity and traceability of fish and seafood products. Data were elaborated from the scientific literature (published in the 2010–2020 years) collected from Scopus and Web of Science search engines, using 'fish', 'seafood', 'authentication', 'elemental analysis', 'elemental profile', 'elemental fingerprinting', and/or 'chemometrics' as search terms.

Table 1. Overview of the literature dealing with multielement profile for fish and seafood authenticity verification

Product	Classification objective	Input data	Technique for element determination	Elements	Data analysis	Validation	Reference
Salmon	Production method	Elemental profile Stable isotope ratio	ICP-AES	As, Ba, Be, Ca, Co, Cd, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr, Ti, Zn	PCA, CDA, LDA, QDA, ANNs, PNNs, NNB	Cross-validation External validation	(Anderson et al., 2010)
Catfish	Geographical origin	Elemental profile	ICP-AES	Al, Ca, Cr, Cu, Fe, K, Mg, Na, P, S, Zn	PCA, CDA, k-NN	Cross-validation	(Li, Boyd, Odom, & Dong, 2013)
Croacker	Geographical origin Seasonality	Elemental profile Stable isotope ratio Proximate composition	EDXRF	Hg, Cd, Pb, As, S, Cl, K Ca, Fe, Cu, Zn, Se, Br, Rb	PCA	–	(Chaguri et al., 2015)
European seabass	Geographical origin Production method Farming system	Elemental profile Stable isotope ratio Biometric measures Fatty acids	ICP-AES	As, Ca, Cd, Co, Cr, Cu, Fe, Hg, K, Mg, Mn, Na, Ni, P, Pb, S, Se, Zn	PCA	Cross-validation	(Farebegoli et al., 2018)
Asian seabass	Geographical origin Production method	Elemental profile Stable isotope ratio	XRF	Mg, Al, Si, P, S, Cl, K, Ca, Ti, Cr, Mn, Fe, Ni, Cu, Zn, As, Se, Br, Rb, Sr, Y, Zr, Cd, Sn, Sb, Nd, Hf, Pb, Bi, At, U	PCA, LDA, RF	Cross-validation External validation	(Gopi et al., 2019b)
European seabass	Geographical origin Production method	Element profile Stable isotope ratio	ICP-MS	La, Eu, Ho, Er, Lu, Tb	PCA, OLPS-DA	Cross-validation External validation	(Varrà et al., 2019)
Sea cucumber	Geographical origin	Elemental profile	ICP-MS	Al, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Mo, Cd, Hg, Pb	PCA, CA, LDA	Cross-validation	(Liu et al., 2012)
Sea cucumber	Geographical origin	Elemental profile	ICP-AES ICP-MS	Na, K, Ca, Fe, Zn, Mg, Al, Li, V, Cr, Mn, Co, Ni, Cu, As, Sn, Sr, Ag, Cd, Se, Ba, Pb, Bi, Y, La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Sc	PCA, LDA	Cross-validation	(Kang et al., 2018)
Pacific white shrimp	Geographical origin	Elemental profile	ICP-AES	Al, As, Ba, Ca, Co, Cr, Cu, Fe, K, Mg, Mn, Mo, Na, Ni, P, S, Se, Ti, Zn, Zr	PCA, CDA, k-NN	Cross-validation	(Li, Boyd, & Odom, 2014)
Shrimps	Geographical origin Production method Species	Elemental profile Stable isotope ratio	ICP-AES ICP-MS	Pb, Cd, As, P, S	PCA, CA, LDA, k-NN	Cross-validation	(Ortea & Gallardo, 2015)
Prawns	Geographical origin	Elemental profile Stable isotope ratio	ICP-MS	Li, B, Al, Ti, V, Mo, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Sr, Cd, Hg, K	PCA, CDA	Cross-validation	(Carter, Tinggi, Yang, & Fry, 2015)

(Continued)

Table 1. *(Continued)*

Product	Classification objective	Input data	Technique for element determination	Elements	Data analysis	Validation	Reference
Pacific white shrimps	Geographical origin	Elemental profile	ICP-AES	Al, As, B, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, S, Se, Si, Ti, Zn, Zr	PCA, CDA, S-LDA	Cross-validation	(Li et al., 2017)
Chinese mitten crab	Geographical origin	Elemental profile Stable isotope ratio	ICP-MS	Na, Mg, Al, K, Ca, Mn, Cu, Zn, Sr, Ba	LDA, SVM	Cross-validation External validation	(Luo et al., 2019)
Pacific white shrimps	Geographical origin	Elemental profile Stable isotope ratio	ICP-MS	Li, Al, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Ga, As, Rb, Sr, Ag, Cd, Cs, Ba, Pb, Y, Ce, Nd, Pr, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Th U	PCA, CDA, S-LDA	Cross-validation	(Li, Han, Dong, & Boyd, 2019)
Black tiger prawn	Geographical origin Production method	Elemental profile Stable isotope ratio	XRF	Mg, Al, Si, P, S, Cl, K, Ca, Ti, Cr, Mn, Fe, Ni, Cu, Zn, As, Se, Br, Rb, Sr, Y, Zr, Cd, Sn, Sb, Nd, Hf, Pb, Bi, At, U	LDA, RF	Cross-validation External validation	(Gopi et al., 2019a)
Mussels	Geographical origin	Elemental profile	ICP-MS	Ag, Ba, Cd, Co, Cr, Cu, Ga, Mn, Mo, Ni, Pb, Rb, Sb, Sn, Sr, Te, Tl, V, As, Se, Zn, Nb, Ta, Zr, Ce, Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Sm, Th, Tm, U, Y, Yb	LDA, SIMCA, ANNs	Cross-validation	(Costas-Rodríguez, Lavilla, & Bendicho, 2010)
Manila clams	Geographical origin	Elemental profile	ICP-MS	Na, Mg, Al, K, V, Mn, Fe, Co, Cu, Zn, As, Se, Rb, Sr, Mo, Pd, Cd, Sn, Sb, Cs, Ba, La, Ce, Pb, U	S-LDA	Cross-validation	(Zhao & Zhang, 2016)
Cuttlefish (ink)	Geographical origin	Elemental profile	ICP-MS	Na, Mg, P, K, Ca, V, Mn, Fe, Cu, Zn, Cr, Co, Mo, Ni, As, Cd, Pb, Hg	PCA	–	(Bua et al., 2017)

ANNs=artificial neural networks. CDA=canonical discriminant analysis. EDXRF=energy dispersive X-ray fluorescence spectroscopy. ICP-AES=inductively coupled plasma atomic emission spectroscopy. ICP-MS=inductively coupled plasma mass spectrometry. k-NN=k-nearest neighbours. LDA=linear discriminant analysis. NNS=neural network bagging. PCA=principal component analysis. PNNs=probabilistic neural networks. QDA=quadratic discriminant analysis. RF=random forest. SIMCA=soft independent modelling of class analogy. S-LDA=stepwise-linear discriminant analysis. XRF=X-ray fluorescence spectroscopy.

3.1 Fish

The maximum guarantee of transparency about the method of production, intended as catching wild fish or raising aquaculture fish, is of extreme importance, given that the two products have a differing economic value. In addition, certain farmed fish such as salmonids are reported to be more prone to accumulate environmental toxic substances, especially of organic nature (Hites, Foran, Carpenter, Hamilton, Knuth, & Schwager, 2004), thus questioning the overall wholesomeness of these products. Tracing the geographical origin of aquaculture products may be, in some ways, more complicated than tracing that of wild-caught products. In fact, despite the feeding habits and prey availability for wild fish are highly variable and cannot be controlled, it should be emphasised that feeds used in aquaculture practices (which significantly affect mineral and trace element contents of fish tissues) are not only extremely variable in terms of composition, but are frequently used worldwide to raise fish of different geographical origin (Aceto, 2016), thus masking any discriminant potential of the elemental profile.

Despite these hurdles, different species of both wild and farmed salmons corresponding to king salmon (*Oncorhynchus tshawytscha*), coho salmon (*Oncorhynchus kysutch*), and Atlantic salmon (*Salmo salar*), were analysed for their major and trace elemental content and isotope ratio profile of carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) in order to develop a model suited for their classification (Andreson, Hobbie, & Smith, 2010) As for the type of employed tracers, it was verified that using elements or isotope ratios has no bearing on the overall performances of salmon classification, but on the contrary, the outcomes are strongly influenced by the number of samples employed to train the classification model as well as by the chosen classification algorithm. On that note, using machine learning algorithms as artificial neural networks (ANNs) and neural network bagging (NNB) gave 94% and 92% correct classification rate, respectively when applied to elements only, and 94% and 87% when using stable isotope ratios only (Anderson et al., 2010).

The possibility of using rare earth elemental profile and/or light stable isotope ratios to identify fish production methods was recently investigated also for European sea bass *(Dicentrarchus labrax,* L.) samples (Varrà, Ghidini, Zanardi, Badiani, & Ianieri, 2019). Anyway, in this case, the concentrations of lanthanum, europium, holmium, erbium, lutetium, and terbium elaborated by PCA and orthogonal partial least square discriminant analysis (OPLS-DA) did not impact on the differentiation of wild from farmed specimens, in contrast to light isotope ratios of carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) which had a higher influence. However, the authors verified that holmium and lanthanum, due to their natural variability in the marine environment, had a significant influence on the discrimination of the same samples by geographic origin. As a matter of fact, almost 89% of unlabelled samples from three different fishing areas in the Mediterranean Sea (used to test the validity of the developed model) were correctly discriminated (Varrà et al., 2019).

The truthfulness of the label description of European seabass *(Dicentrarchus labrax,* L.) was also analysed in another study which took into consideration the outputs obtained through the measurement of several parameters, corresponding to biometric indices, fatty acids profile, analysis of eighteen elements, and stable isotope ratios of carbon and nitrogen (Farebegoli et al., 2018). The method of production, the intensity of farming system, and the geographical provenance of sea bass were better discriminated by using the fatty acids composition, while the use of elements alone outperformed compared to the other analytical data. Only the concentrations of Ca were in found to be significantly affected by feeding system and geographical origin, but the differences in fish tissues were not sufficient to achieve satisfying discrimination results, which settled around 79% for production method and 57% for the origin. On the contrary, stable isotope ratio data performed well in discriminating the production method of samples, due to the strong influence of the feeding inputs on these parameters, but they were not able to classify samples according to provenance (Farebegoli et al., 2018).

The merging of the results of multi-elemental and stable isotope ratios analyses has been successful also for the discrimination by origin and production method of Asian sea bass (*Lates calcarifer*) (Gopi, Mazumder,

Sammut, Saintilan, Crawford, & Gadd, 2019) and, when adding proximate composition, also for the discrimination by origin of croaker (Chaguri et al., 2015). Unlike the previously reported studies, XFR was the chosen technique to determine the elemental content of fish samples, offering mainly advantages in terms of speed of operation. In this case, although the origin discriminant models for sea bass created by applying LDA or RF to stable isotope data were more accurate than those computed using elemental data only, isotopic analysis was less performant when used alone to predict both the origin and the production method of unknown samples, thus suggesting that information provided by elements is essential to achieve satisfying discrimination accuracy for the identification of geographical provenances (Gopi et al., 2019).

One single attempt to discriminate the origin of freshwater cultured fish was found in the literature. In this case, fillets of channel catfish (*Ictalurus punctatus*) and hybrid catfish (*Ictalurus furcatus*) from three geographic areas were subjected to ICP-AES to measure a total of eleven elements (Li, Boyd, Odom, & Dong, 2013). Although the authors did not find a direct influence of water and feed used to raise the catfish on the final elemental composition of the fillets, the products were separated by origin with 100% whether canonical discriminant analysis or *K*-nearest-neighbour analysis were used. Despite this, it should be noted that provenances considered in this study are of geopolitical rather than of geochemical nature, therefore the validity of the discrimination is limited by the fact that aquaculture catfish can be raised elsewhere, but using waters with an equivalent elemental composition (Li et al., 2013).

4.2. Echinoderms and crustaceans

Mislabelling of echinoderms has been poorly treated by the scientific community, probably due to the fact that consumption of these products, however high, is mainly limited to Asian countries. Two applications regarding the authentication of sea cucumber (*Apostichopus japonicus*) through elemental profiles are currently available, aimed at classifying the samples according to three (Liu, Xue, Wang, Li, Xue, & Xu, 2012) and five (Kang, Zhao, Shang, Zhai, Ning, & Sheng, 2018) sampling areas in China,

but using a different number of elements, equal to fifteen in the first case and to thirty-nine in the second one. In both works, a stepwise-LDA was used to concomitantly sort elements by their relative importance in discrimination and build classification models. Concentrations of Al, Mn, Fe, Co, Ni, Cu, As, Se, Cd and Hg were found to be appropriate to differentiate 100% of sea cucumbers in relation to the three sampling areas (Li et al., 2014), while concentrations of Li, Na, Al, K, Co, Cu, Cd, and Sc allowed achieving 88% accuracy in differentiating samples originating from the five areas (Kang et al., 2018). So, despite a higher number of elements was measured in the second study, this is not always a straightforward matter to achieve better discrimination results. If redundant or noise elements are not strictly evaluated and removed by proper statistics, models built using many elements as variables are likely to outperform, especially with an increasing number of origins to be identified.

By reviewing the literature, crustaceans emerged as the most frequently analysed category of seafood products intended to be authenticated by their elemental composition (see Table 1). More specifically, six out of seven works analysing the multi-elemental profile of crustaceans and taken into consideration in the present review, dealt with the authentication of the origin of shrimps or prawns (Li, Boyd, & Odom, 2014; Ortea & Gallardo, 2015; Carter, Tinngi, Yang, & Fry, 2015; Li et al., 2017; Luo, Jiang, Chen, Zhenng, Liu, & Yang, 2019; Li, Han, Dong, & Boyd, 2019; Gopi, Mazumder, Sammut, Saintilan, Crawfor, & Gadd, 2019), and, among these, only two works concurrently investigated the possibility of using the same profile to address other problems such as the production method and the species identifications (Ortea et al., 2005; Gopi et al., 2019).

The use of the elemental profile alone was demonstrated to be an optimal strategy to accurately assess traceability of Pacific white shrimps (*Litopenaeus vannamei*) from different sampling sites in the USA (Li et al., 2014) to differentiate shrimps obtained from Vietnam, Thailand, and India, which represent the biggest producing countries in the world (Li et al., 2017), and, when in combination with light stable isotope ratios of carbon and nitrogen, also to discriminate shrimps according to different sampling areas in China (Li et al., 2019). In general, despite a combination of major, minor and trace

elements (especially K, Mg, Na, P, Ca, Ba, Cr, Pb, Se, Si Cd, Co, and Zr) was successful in solving origin discrimination problems in all cases, when concentrations of REEs were determined and used as discriminant variables, it was found that these elements had a greater analytical significance in determining the provenance of shrimps compared to other variables (Li et al., 2019).

The superiority, to some extent, of the element composition over stable isotope ratios of carbon and nitrogen to assess traceability of shrimps was also demonstrated when farmed and wild samples of seven different biological species, obtained from nine sampling zones, were investigated (Ortea et al., 2015). Stable isotope analysis alone yielded to 100%, 71%, and 58% of samples to be correctly classified using LDA by production method, origin, and biological species, respectively. However, with an increasing number of samples into the models, the origin discrimination accuracy decreased or did not significantly increase. On the contrary, As, Cd, Pb, P and S concentrations alone showed greater accuracy in classifying samples by origin (94%) and species (74%) and, when merged with stable isotopes ratios, the two techniques showed the maximum discrimination power (Ortea et al., 2015). Similarly, both the production method and the origin traceability of prawn (*Penaeus monodon*) were assessed with 100% accuracy when the multi-elemental profile and stable isotopes ratios were used in complementarity (Gopi et al., 2019b).

Advantages of coupling elements and light stable isotope ratio analyses outputs to verify the exact provenances of high-value crustaceans is even more evident when powerful classification machine learning techniques are applied. The contents of Na, Mg, Al, K, Ca, Mn, Cu, Zn, Sr, and Ba plus $\delta^{13}C$ and $\delta^{15}N$ measured on limited sample material and elaborated by means of SVM allowed tracing Chinese mitten crabs (*Eriocheir sinensis*) according to eight different geographical origins around China with 100% and 97% accuracy in cross-validation and external validation, respectively Luo et al., 2019).

4.3 Molluscs

At a date, molluscs appear to be the less frequently treated aquatic products in terms of the evaluation of authenticity and fraud verification. This is particularly remarkable considering that, according to the latest available data, the worldwide supply of cephalopods and other molluscs has reached values of 3535732 and of 17500801 tonnes per year, respectively (Food and Agriculture Organization, 2020).

Historically, the elemental profile of bivalve or cephalopods molluscs was employed to assess their geographical authenticity, but performing such analysis on non-edible hard parts of the animals (e.g. shells, statoliths, beaks) (Iguchi, Takashima, Namikoshi, Yamashita, & Yamashita, 2013; Northern, Smith, McKinnon, & Bolstad, 2019) and, thereby, not guaranteeing the possibility to apply the same methods to ready-to-cook products (eviscerated, beheaded, shelled), which are rising in popularity on the international markets.

An interesting study used ICP-MS in combination with LDA, SIMCA, and ANNs to quantify and elaborate a total of forty elements in order to authenticate Galician mussels (*Mytilus galloprovincialis*) under the European Protected Designation of Origin (PDO) and protect the products from similar but lower-quality mussels (Costas-Rodrìguez, Lavilla, & Bendicho, 2010). In particular, a strong relation between element composition of PDO mussels and the geomorphology and lithology of the specific production zone, as well as with external contamination sources was found. While the Se, Zn, Pb, Co, Mo, Ag, and Ba elemental signature was attributed to the metabolic activity of the animals, the Ga, Zr, Eu, Lu, Th and U signature was specifically related to mineralogical sources of the area, and the V, Cd, and Sb signature to anthropogenic pollutant activities characterising the area (Costas-Rodrìguez et al., 2010). Keeping the complementary information provided by all these elements, PDO from non-PDO products were 100% accurately classified by LDA and SIMCA. On the contrary, the use of ANNs was found to be more effective in discriminating the five different sampling zones from which the PDO mussels were obtained.

In another work, particular attention was paid toward any effect the seasonality had on the elemental composition of bivalves (Zhao & Zhang, 2016). Since season variations were misleadingly reflected on the Mg, Rb, Pd, Cd, Sn, Ba, La, and Ce distribution into the molluscs, the authors were able to authenticate samples of Manila clams (*Ruditapes philippinarum*) using a different pattern of elements composed by Mg, Rb, Pd, Cd, Sn, Ba, La, and Ce, which, in contrast, was found to be more strongly linked to the geographical origin of clams (Zhao et al., 2016).

As far as we know, no works oriented towards the evaluation of cephalopods mislabelling by measuring element composition of edible tissues such as mantles and fins are available. Nevertheless, the inorganic composition of ink derived from cuttlefish (commonly used in the Mediterranean and Japanese gastronomy) showed some potential ability to enclose geographical-related information (Bua, Albergamo, Annuario, Zammuto, Costa, & Dugo, 2017). Although no classification analysis was performed, some elements such as Cr, Ni, V, Cd, Pb, As, and Hg were significantly different among ink samples of cuttlefish (*Sepia officinalis*) of different sampling sites in the central Mediterranean Sea, suggesting that the contribution of the environmental pollution should be further investigated in this kind of studies to verify whether it can reveal actionable insights.

4. Why Aquatic Animals Are Ideal Candidates for Multi-Elemental Analysis?

The evaluation of the organic composition of foodstuffs continues to be the first choice when the identification of individual markers or patterns of markers for authenticity and traceability of fishery products is the main research goal. Nevertheless, the measurement of a high number of organic components without carefully considering their origin, significance, sources of variations, and the general framework within which they are evaluated, frequently put their outright specificity as markers of origin into question. Indeed, the concentration and distribution of certain classes of organic compounds such as fatty acids, peptides, and enzymes, are concurrently affected by so many aspects and circumstances that it is often challenging to univocally relate them to the sole species, geographic and/or farming origin.

Pre-catch conditions, such as seasonality, climatic conditions, fishing period, fish size, fish physiology and metabolism, and fishing gear, as well as human's post-catch manipulation and storage operations (storage temperature, packaging, lifetime of the product, and so forth) are just a few examples of factors affecting the organic composition of fish (Danezis et al., 2018). Similar considerations are valid also for inorganic constituents, but the correlation between the elemental compositions of fish tissues and the surrounding aquatic environment has been demonstrated to be more stable and consistent over time. Therefore, the probability that inorganic markers of fish origin are hidden by misleading factors may be considered lower compared to organic markers. Based on the concentrations found in the matrix, elements are normally categorized as major and minor (trace and ultra-trace) elements. Anyway, a detailed definition has been reported only for trace elements, intended as those elements whose concentrations in the matrix are lower than 100 mg kg^{-1} (IUPAC, 2020) and which in foods are mainly represented by B, Al, Fe, Mn, Ni, Cu, Zn, As, Se, Sr, and, sometimes, by La, and Ce. By consequence, major elements have mass fractions above 100 mg kg^{-1} (Na, Mg, P, K, Ca, Mg), while ultra-trace elements generally below 1 mg kg^{-1} [36] (Li, V, Cr, Co, Rb, Y, Zr, Mo, Ru, Pd, Cd, Sn, Ag, Cd, Sn, Sb, Cs, Ba, lanthanides, Hf, Re, Pt, Bi, Hg, Th, U, Hg). Rare earth elements (REEs), usually including Y, La, and lanthanides (Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, and Lu) (Tyler, 2004) are emerging as very promising inorganic markers of fish authenticity, despite their quantification into foodstuffs is still limited to the very low abundance and consequent obstacles in quantification by modern instrumentations. Concentrations of REEs in both surface water and groundwater were found to vary significantly in relation to the geographical areas, with the Asiatic continent (and, in particular, China) showing the highest levels, followed by Europe, Africa, USA and Australia (Adeel at al., 2019). These variations may be attributed both to the natural release of REEs from the parental soil (weathering of black shale is a common cause of increasing REEs composition in water) and to some anthropic activities (metallurgy, glass and ceramic industry, electronics) responsible for the REEs release into the aquatic environment and the consequent uptake by the aquatic fauna

Migaszewski & Galuszka, 2015). The overall major and minor elemental composition of fish is largely related to the elemental content of the eaten preys, vegetation, or fodder. In turn, the content of elements in animal and vegetable feeds is the results of the bio-available elements which have been mobilised from the soil and which reflect the overall characteristics of the geographical area (Drivelos et al., 2012; Reinholds at al., 2015). For example, being some alkaline metals (e.g. Rb and Cs) mobilised from the underlying soils with utmost ease, the probability that their incorporation into fish tissues is variable according to the geographical site is very high (Kelly et al., 2008). Other trace elements, such as B and As, naturally enter the aquatic environment from volcanic and geothermal activities (Parks & Edwards, 2005; Zkeri, Aloupi, & Gaganis, 2018) and, therefore, their concentrations into fish and seafood tissues may be exploited for discriminating animals form marine areas with specific geochemical characteristics. More, concentrations of some major and trace elements as Li, Mg, Ca, Sr, Zn, Mn, and Cu, are strictly regulated by the salinity of the marine basin and this characteristic make them suitable to be potentially used for marine fish tracing purposes (Hanson & Zdanowicz, 1999). In this setting, it should not be strange that traceability studies concerning marine fish species are, to some extent, more standardisable and, thus, reliable compared to those dealing with freshwater species. Along with some concern deriving from the closeness to anthropic environment, this aspect is attributable to the higher degree of dynamism of the marine systems compared to freshwater ones. Since this dynamism is biologically, chemically, and physically controlled, a more uniform element concentration from both a temporal and spatial point of view can therefore be found in the marine environment, especially in open ocean waters. Nevertheless, when performing authentication studies, thoughts must be given to the fact that the distribution ratio of certain elements between fish tissues and seawater is altered by the metabolic activity of animals (Aceto, 2016). Specifically, the uptake of many essential elements as Na, K, Mg, and Ca is metabolically regulated by the same fish since necessary for regulating physiological functions. Hence, the potential for variation of these elements in relation to origin is masked by

physiological 'noise' (Sturrock et al., 2015), thus being the reason why they are hardly ever used in fish authentication studies.

Finally, a greater compositional heterogeneity is encountered in waters of coastal areas compared to deep seawaters, where the proximity of anthropic releasing sources makes some trace and ultra-trace elements to be variably introduced into the marine environment. Nickel, zinc, arsenic, lead, mercury, and cadmium are well known to be more concentrated along shorelines (Rainbow, 2018) since deriving from certain agricultural practices or industrial activities. On the other side, fish and seafood are not able to physiologically regulate the concentration of these non-essential (and often toxic) elements, which, therefore, are passively accumulated into the animal's tissues. If properly evaluated, also anthropic elements can thus be used for origin authentication purposes (Cubbadda, Raggi, & Coni, 2006).

To conclude, although introduced through different sources, elements can be successfully employed as authenticity markers of fish and seafood if the same introduction sources are systematic, identifiable, manageable, and suggestive of the geographical origin or production process (Aceto, 2016). In this context, when dealing with authentication of transformed fish products, particular attention should be paid against the introduction of elements from the production chain. These obstacles may be often overcome by comparing the effective concentrations of elements in the final products with those found along as many stages as possible of the transformation process, so as to be able to verify whether distribution trends are retained along the production stages (Aceto, 2016).

5. Final Remarks and Conclusions

The application of element profiling approaches to fish and seafood products has been gaining momentum, and the scientific community has been working on the optimisation of both existing instrumentations for multi-elemental analysis and algorithms for statistical analysis. The greater thrust has come from advances in chemometrics and machine learning techniques, which now provide great support to the identification of

maximum relevant chemical information from large datasets not otherwise accessible.

From the analysis of the literature presented in this review, it is clear that the discrimination of the geographical origin has been the most frequently discussed authenticity topic, while other aspects, such as the farming systems, have been overlooked. In addition, crustaceans emerge as the most frequently investigated category of products, while less emphasis has been put on fish, echinoderms, and molluscs, especially cephalopods, probably due to difficulties in drawing up an adequate sampling plan to build representative datasets. Regarding the statistical data treatment, PCA and LDA have been more widely used, while machine learning algorithms have been neglected, despite their great potential in discovering hidden discriminant patterns among data.

As for the selected methodologies, ICP-MS, followed by ICP-OES have been the first choice, accounting for the vast majority of the published research. Especially in the last years, ICP-MS has been gaining popularity within the scientific community because less complicated, less expensive, and undoubtedly the fastest and most universal trace element technique commercially available today. This is mainly due to the advances in collision/reaction cell technology which offers an effective way of reducing spectral effects from different polyatomic ions. Quadrupole mass spectrometers, in particular, are increasingly being used and, until recently, it seemed impossible that a single technique would fit perfectly to the needs of all the laboratories. For this reason, these instruments are expected to supersede shortly most of the ICP-OES and AAS applications. In addition, it may be expected that various solid-sampling techniques such as ETV-ICP-MS, LA-ICP-MS, and XRF may succeed more in the field of food authentication taking the advantage of a reduced sample preparation.

Another peculiarity emerging from the published literature is the tendency to couple element profiles of fish and seafood with other analytical parameters, especially stable isotopes of carbon, hydrogen, and nitrogen, probably motivated by the need to increase the accuracy of discrimination. Despite benefits deriving from the fusion of complementary or synergistic information, it is worth highlighting that multi-elemental analysis may be

sufficient to achieve equivalent results with an optimal cost-performance ratio.

Looking forward, the increased use of ICP-MS-hyphenated techniques for elemental speciation and ICP-MS/MS for interference-free determination and isotope ratio measurement would represent a turning point for the high-throughput analytical characterisation of complex matrices such as food. Nevertheless, the reduction of the cost of the equipment for multi-elemental analysis would certainly be desirable to further encourage the spreading of multi-elemental analytical approaches in a different context from that of the specialized laboratories dealing with food surveillance. Before getting to this point, the validity and robustness of elemental markers to ascertain fish and seafood authenticity must be also increased. Further work on these issues is therefore encouraged, in order to integrate into adequately defined reference databases information relating to any possible variable influencing the inorganic profile of fishery products with the elemental information relating to the origin. At the same time, continuous technological improvements as well as the shift towards a progressive miniaturisation of the instruments may be a major turning point, helping to concomitantly monitor health risks associated with the occurrence of toxic metals such as cadmium, lead, mercury, and arsenic, and to meet the demand for cost-effective, energy-and reagent-saving instruments.

References

Aceto, M. (2016). The Use of ICP-MS in Food Traceability. Advances in Food Traceability Techniques and Technologies: Improving Quality Throughout the Food Chain. *Elsevier Ltd*. https://doi.org/10.1016/B978-0-08-100310-7.00008-9

Adebiyi, F. M., Ore, O. T., & Ogunjimi, I. O. (2020). Evaluation of human health risk assessment of potential toxic metals in commonly consumed crayfish (Palaemon hastatus) in Nigeria. Heliyon, 6(1), e03092. https://doi.org/10.1016/j.heliyon.2019.e03092.

Adeel, M., Lee, J.Y., Zain, M., Rizwan, M., Nawab, A., Ahmad, M.A., Shafiq, M., Yi, H., Jilani, G., Javed, R., Horton, R., Rui, Y., Tsang, D.C.W., & Xing, B. (2019). Cryptic footprints of rare earth elements on natural resources and living organisms. *Environmen International* 127, 785-800. https://doi.org/10.1016/j.envint.2019.03.022

Anderson, K. A., Hobbie, K. A., & Smith, B. W. (2010). Chemical profiling with modeling differentiates wild and farm-raised salmon. *Journal of Agricultural and Food Chemistry, 58*(22), 11768–11774. https://doi.org/10.1021/jf102046b

Aursand, M., Mabon, F., & Martin, G. J. (2000). Characterization of farmed and wild salmon (Salmo salar) by a combined use of compositional and isotopic analyses. Journal of the American Oil Chemists' Society, 77(6), 659-666. https://doi.org/0.1007/s11746-000-0106-5

Balcaen, L., Bolea-Fernandez, E., Resano, M., & Vanhaecke, F. (2015). Inductively coupled plasma–Tandem mass spectrometry (ICP-MS/MS): A powerful and universal tool for the interference-free determination of (ultra) trace elements–A tutorial review. Analytica chimica acta, 894, 7-19. https://doi.org/10.1016/j.aca.2015.08.053.

Ballin, N. Z., & Laursen, K. H. (2019, April 1). To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. *Trends in Food Science and Technology*. https://doi.org/10.1016/j.tifs.2018.09.025

Bizzi, C. A., Pedrotti, M. F., Silva, J. S., Barin, J. S., Nóbrega, J. A., & Flores, E. M. M. (2017). Microwave-assisted digestion methods: towards greener approaches for plasma-based analytical techniques. Journal of Analytical Atomic Spectrometry, 32(8), 1448-1466. https://doi.org/10.1039/C7JA00108H.

Brown, R. J., & Milton, M. J. (2005). Analytical techniques for trace element analysis: an overview. TrAC *Trends in Analytical Chemistry, 24*(3), 266-274. https://doi.org/10.1016/j.trac.2004.11.010

Bua, G. D., Albergamo, A., Annuario, G., Zammuto, V., Costa, R., & Dugo, G. (2017). High-Throughput ICP-MS and Chemometrics for Exploring the Major and Trace Element Profile of the Mediterranean Sepia Ink. *Food Analytical Methods, 10*(5), 1181–1190. https://doi.org/10.1007/s12161-016-0680-6

Burger, M., Hendriks, L., Kaeslin, J., Gundlach-Graham, A., Hattendorf, B., & Günther, D. (2019). Characterization of inductively coupled plasma time-of-flight mass spectrometry in combination with collision/reaction cell

technology–insights from highly time-resolved measurements. Journal of Analytical Atomic Spectrometry, 34(1), 135-146. https://doi.org/10.1039/C8JA00275D.

Bulska, E., & Ruszczyńska, A. (2017). Analytical techniques for trace element determination. *Physical Sciences Reviews*, *2*, 5. https://doi.org/10.1515/psr-2017-8002

Callao, M. P., & Ruisánchez, I. (2018). An overview of multivariate qualitative methods for food fraud detection. *Food Control*, *86*, 283-293 https://doi.org/10.1016/j.foodcont.2017.11.034

Carter, J. F., Tinggi, U., Yang, X., & Fry, B. (2015). Stable isotope and trace metal compositions of Australian prawns as a guide to authenticity and wholesomeness. *Food Chemistry*, *170*, 241–248. https://doi.org/10.1016/j.foodchem.2014.08.037

Chaguri, M. P., Maulvault, A. L., Nunes, M. L., Santiago, D. A., Denadai, J. C., Fogaça, F. H., Sant'Ana, L.S., Ducatti, C., Bandarra, N., & Marques, A. (2015). Different tools to trace geographic origin and seasonality of croaker (Micropogonias furnieri). *LWT - Food Science and Technology*, *61(*1), 194–200. https://doi.org/10.1016/j.lwt.2014.11.006

Commission of the European Communities (2006). Commission regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foostuffs. *Official Journal of the European Union, L364/5* (20.12.2006).

Costas-Rodríguez, M., Lavilla, I., & Bendicho, C. (2010). Classification of cultivated mussels from Galicia (Northwest Spain) with European Protected Designation of Origin using trace element fingerprint and chemometric analysis. *Analytica Chimica Acta, 664*(2), 121–128. https://doi.org/10.1016/j.aca.2010.03.003

Cubadda, F., Raggi, A., & Coni, E. (2006). Element fingerprinting of marine organisms by dynamic reaction cell inductively coupled plasma mass spectrometry. *Analytical and Bioanalytical Chemistry*, *384*(4), 887–896. https://doi.org/10.1007/s00216-005-0256-6

Danezis, G. P., Tsagkaris, A. S., Camin, F., Brusic, V., & Georgiou, C. A. (2016). Food authentication: Techniques, trends & emerging approaches. *TrAC - Trends in Analytical Chemistry, 10*, 22-31. https://doi.org/10.1016/j.trac.2016.02.026

De Andrade, R. M., de Gois, J. S., Toaldo, I. M., Batista, D. B., Luna, A. S., & Borges, D. L. (2017). Direct determination of trace elements in meat samples via high-resolution graphite furnace atomic absorption spectrometry. Food Analytical Methods, 10(5), 1209-1215. https://doi.org/10.1007/s12161-016-0659-3.

De Sá, I. P., Higuera, J. M., Costa, V. C., Costa, J. A. S., da Silva, C. M. P., & Nogueira, A. R. A. (2020). Determination of trace elements in meat and fish samples by MIP OES using solid-phase extraction. Food Analytical Methods, 13(1), 238-248. https://doi.org/10.1007/s12161-019-01615-3.

Drivelos, S. A., & Georgiou, C. A. (2012). Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. *TrAC - Trends in Analytical Chemistry, 40,* 38-51. https://doi.org/10.1016/j.trac.2012.08.003

Dunphy, B. J., Millet, M. A., & Jeffs, A. G. (2011). Elemental signatures in the shells of early juvenile green-lipped mussels (Perna canaliculus) and their potential use for larval tracking. Aquaculture, 311(1-4), 187-192. https://doi.org/10.1016/j.aquaculture.2010.12.016.

Fabiano, K. C., Lima, A. P., Vasconcellos, M. B., & Moreira, E. G. (2016). Contribution to food safety assurance of fish consumed at São Paulo city by means of trace element determination. Journal of Radioanalytical and Nuclear Chemistry, 309(1), 383-388. https://doi.org/10.1007/s10967-016-4806-5

FAO (2020). FAO yearbook, Fishery and Aquaculture Statistics 2018. Available online at http://www.fao.org/faostat/en/#data/CL

Farabegoli, F, Pirini, M., Rotolo, M., Silvi, M., Testi, S., Ghidini, S., Zanardi, E., Remondini, D., Bonaldo, A., Parma, L., & Badiani, A. (2018). Toward the Authentication of European Sea Bass Origin through a Combination of Biometric Measurements and Multiple Analytical Techniques. *Journal of Agricultural and Food Chem*istry, 66, 26, 6822–6831. https://doi.org/0.1021/acs.jafc.8b00505

Flem, B., Moen, V., Finne, T. E., Viljugrein, H., & Kristoffersen, A. B. (2017). Trace element composition of smolt scales from Atlantic salmon (Salmo salar L.), geographic variation between hatcheries. Fisheries Research, 190, 183-196. https://doi.org/10.1016/j.fishres.2017.02.010

Giusti, A., Ricci, E., Guarducci, M., Gasperetti, L., Davidovich, N., Guidi, A., & Armani, A. (2018). Emerging risks in the European seafood chain: Molecular identification of toxic Lagocephalus spp. in fresh and processed products. Food Control, 91, 311-320. https://doi.org/10.1016/j.foodcont.2018.04.013.

Gopi, K., Mazumder, D., Sammut, J., Saintilan, N., Crawford, J., & Gadd, P. (2019a). Combined use of stable isotope analysis and elemental profiling to determine provenance of black tiger prawns (Penaeus monodon). *Food Control, 95*, 242–248. https://doi.org/10.1016/j.foodcont.2018.08.012

Gopi, K., Mazumder, D., Sammut, J., Saintilan, N., Crawford, J., & Gadd, P. (2019b). Isotopic and elemental profiling to trace the geographic origins of farmed and wild-caught Asian seabass (Lates calcarifer). *Aquaculture, 502,* 56–62. https://doi.org/10.1016/j.aquaculture.2018.12.012

Hanson, P.J. & Zdanowicz, V.S. (1999). Elemental composition of otoliths from Atlantic croaker along an estuarine pollution gradient. *Journal of Fish Biology*, 54, 656–668. https://doi.org/10.1111/j.1095-8649.1999.tb00644.x.

Hites, R. A., Foran, J. A., Carpenter, D. O., Hamilton, M. C., Knuth, B. A., & Schwager, S. J. (2004). Global Assessment of Organic Contaminants in Farmed Salmon. *Science, 303*(5655), 226–229. https://doi.org/10.1126/science.1091447

Husáková, L., Urbanová, I., Šrámková, J., Černohorský, T., Krejčová, A., Bednaříková, M., ... & Pilařová, L. (2011). Analytical capabilities of inductively coupled plasma orthogonal acceleration time-of-flight mass spectrometry (ICP-oa-TOF-MS) for multi-element analysis of food and beverages. Food chemistry, 129(3), 1287-1296. https://doi.org/10.1016/j.foodchem.2011.05.047.

Iguchi, J., Takashima, Y., Namikoshi, A., Yamashita, Y., & Yamashita, M. (2013). Origin identification method by multiple trace elemental analysis of short-neck

clams produced in Japan, China, and the Republic of Korea. *Fisheries Science, 79*(6), 977–982. https://doi.org/10.1007/s12562-013-0659-9

Iguchi, J., Isshiki, M., Takashima, Y., Yamashita, Y., & Yamashita, M. (2014). Identifying the origin of Corbicula clams using trace element analysis. *Fisheries Science, 80*(5), 1089–1096. https://doi.org/10.1007/s12562-014-0775-1

International Union of Pure and Applied Chemistry (IUPAC) (1997). Compendium of Chemical Terminolgy (the "Gold Book"). Available online at http://goldbook.iupac.org/ (accessed on 20 September 2020).

Jiménez-Carvelo, A. M., González-casado, A., Bagur-gonzález, M. G., & Cuadros-rodríguez, L. (2019). Alternative data mining / machine learning methods for the analytical evaluation of food quality and authenticity – *A review. Food Research International, 122*, 25–39. https://doi.org/10.1016/j.foodres.2019.03.063

Jinadasa, B. K. K. K., Chathurika, G. S., Jayasinghe, G. D. T. M., & Jayaweera, C. D. (2019). Mercury and cadmium distribution in yellowfin tuna (Thunnus albacares) from two fishing grounds in the Indian Ocean near Sri Lanka. Heliyon, 5(6), e01875. https://doi.org/10.1016/j.heliyon.2019.e01875.

Kang, X., Zhao, Y., Shang, D., Zhai, Y., Ning, J., & Sheng, X. (2018). Elemental analysis of sea cucumber from five major production sites in China: A chemometric approach. *Food Control, 94*, 361–367. https://doi.org/10.1016/j.foodcont.2018.07.019

Kelly, S., Heaton, K., & Hoogewerff, J. (2005). Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. *Trends in Food Science and Technology*, 16, 12, 555-567. https://doi.org/10.1016/j.tifs.2005.08.008

Kelly, B. C., Ikonomou, M. G., Higgs, D. A., Oakes, J., Dubetz, C., & Canada, O. (2008). Mercury and other trace elements in farmed and wild salmon from British Columbia, Canada. *Environmental Toxicology and Chemistry, 27*(6), 1361–1370. https://doi.org/10.1897/07-527.S1

Kelly, S., Brodie, C., Hilkert, A. (2018). Isotopic-Spectroscopic Technique: Stable Isotope-Ratio Mass Spectrometry (IRMS). *Modern techniques for food authentication,* 2nd ed., Academic Press, Cambridge, U.S.A, 349-413, https://doi.org/1010.1016/B978-0-12-814264-6.00011-6.

Kim, J. S., Hwang, I. M., Lee, G. H., Park, Y. M., Choi, J. Y., Jamila, N., & Kim, K. S. (2017). Geographical origin authentication of pork using multi-element and multivariate data analyses. *Meat Science, 123*, 13–20. https://doi.org/10.1016/j.meatsci.2016.08.011

Kusche, H., & Hanel, R. (2021). Consumers of mislabeled tropical fish exhibit increased risks of ciguatera intoxication: A report on substitution patterns in fish imported at Frankfurt Airport, Germany. *Food Contro*l, 121, 107647. https://doi.org/10.1016/j.foodcont.2020.107647.

Li, L., Boyd, C. E., & Odom, J. (2014). Identification of Pacific white shrimp (Litopenaeus vannamei) to rearing location using elemental profiling. *Food Control, 45*, 70–75. https://doi.org/10.1016/j.foodcont.2014.03.013

Li, L., Boyd, C. E., Odom, J., & Dong, S. (2013). Identification of Ictalurid Catfish Fillets to Rearing Location Using Elemental Profiling. *Journal of the World*

Aquaculture Society, *44*(3), 405–414. https://doi.org/10.1111/jwas.12033

Li, L., Boyd, C. E., Racine, P., McNevin, A. A., Somridhivej, B., Minh, H. N., & Godumala, R. (2017). Assessment of elemental profiling for distinguishing geographic origin of aquacultured shrimp from India, Thailand and Vietnam. *Food Control, 80,* 162–169. https://doi.org/10.1016/j.foodcont.2017.04.046

Li, L., Han, C., Dong, S., & Boyd, C. E. (2019). Use of elemental profiling and isotopic signatures to differentiate Pacific white shrimp (Litopenaeus vannamei) from freshwater and seawater culture areas. *Food Control, 95*, 249–256. https://doi.org/10.1016/j.foodcont.2018.08.015

Liu, X., Xue, C., Wang, Y., Li, Z., Xue, Y., & Xu, J. (2012). The classification of sea cucumber (Apostichopus japonicus) according to region of origin using multi-element analysis and pattern recognition techniques. *Food Control, 23*(2), 522–527. https://doi.org/10.1016/j.foodcont.2011.08.025

Limbeck, A., Bonta, M., & Nischkauer, W. (2017). Improvements in the direct analysis of advanced materials using ICP-based measurement techniques. *Journal of Analytical Atomic Spectrometry*, 32(2), 212-232. https://doi.org/10.1039/c6ja00335d.

López, M. I., Colomer, N., Ruisánchez, I., & Callao, M. P. (2014). Validation of multivariate screening methodology. Case study: Detection of food fraud. *Analytica chimica acta, 827*, 28-33. https://doi.org/10.1016/j.aca.2014.04.019

Lum, T. S., & Leung, K. S. Y. (2016). Strategies to overcome spectral interference in ICP-MS detection. *Journal of Analytical Atomic Spectrometry*, 31(5), 1078-1088. https://doi.org/10.1039/C5JA00497G.

Luo, R., Jiang, T., Chen, X., Zheng, C., Liu, H., & Yang, J. (2019). Determination of geographic origin of Chinese mitten crab (Eriocheir sinensis) using integrated stable isotope and multi-element analyses. *Food Chemistry, 274*(March 2018), 1–7. https://doi.org/10.1016/j.foodchem.2018.08.104

Machado, R. C., Andrade, D. F., Babos, D. V., Castro, J. P., Costa, V. C., Sperança, M. A., & Pereira-Filho, E. R. (2020). Solid sampling: advantages and challenges for chemical element determination—a critical review. *Journal of Analytical Atomic Spectrometry*, 35(1), 54-77. https://doi.org/10.1039/C9JA00306A.

Magdas, D. A., Feher, I., Cristea, G., Voica, C., Tabaran, A., Mihaiu, M., & Dan, D. S. (2019). Geographical origin and species differentiation of Transylvanian cheese. Comparative study of isotopic and elemental profiling vs. DNA results. *Food Chemistry, 277*. 307–313. https://doi.org/10.1016/j.foodchem.2018.10.103

Markiewicz-Keszycka, M., Cama-Moncunill, X., Casado-Gavalda, M. P., Dixit, Y., Cama-Moncunill, R., Cullen, P. J., & Sullivan, C. (2017). Laser-induced breakdown spectroscopy (LIBS) for food analysis: A review. *Trends in Food Science & Technology,* 65, 80-93. https://doi.org/10.1016/j.tifs.2017.05.005.

Melucci, D., Casolari, S., De Laurentiis, F., & Locatelli, C. (2017). Determination of toxic metals in marine bio-monitors by multivariate analysis of voltametric and spectroscopic data. Application to mussels, algae and fishes of the Northern Adriatic Sea. *Fresenius Environmental Bullettin, 26*, 3756-3763.

Migaszewski, Z.M. & Galuszka, A. (2015). The characteristics, occurrence, and geochemical behavior of rare earth elements in the environment: a review. *Critical Reviews inEnvironmental Science and Technology*., 45, 429-471. https://doi.org/10.1080/10643389.2013.866622

Molkentin, J., Lehmann, I., Ostermeyer, U., & Rehbein, H. (2015). Traceability of organic fish–Authenticating the production origin of salmonids by chemical and isotopic analyses. Food Control, 53, 55-66. https://doi.org/10.1016/j.foodcont.2015.01.003.

Moon, J. H., Ni, B. F., Theresia, R. M., Salim, N. A., Arporn, B., & Vu, C. D. (2015). Analysis of consumed fish species by neutron activation analysis in six Asian countries. *Journal of Radioanalytical and Nuclear Chemistry*, *303*(2), 1447-1452. https://doi.org/10.1007/s10967-014-3732-7.

Marini, F. (2010). Classification methods in chemometrics. *Current Analytical Chemistry*, *6*, 72-79. https://doi.org/10.2174/157341110790069592

Northern, T. J., Smith, A. M., McKinnon, J. F., & Bolstad, K. S. R. (2019). Trace elements in beaks of greater hooked squid Onykia ingens: opportunities for environmental tracing. *Molluscan Research*, *39*(1), 29–34. https://doi.org/10.1080/13235818.2018.1495604

Ortea, I., & Gallardo, J. M. (2015). Investigation of production method, geographical origin and species authentication in commercially relevant shrimps using stable isotope ratio and/or multi-element analyses combined with chemometrics: An exploratory analysis. *Food Chemistry*, *170*, 145–153. https://doi.org/10.1016/j.foodchem.2014.08.049

Panichev, N. A., & Panicheva, S. E. (2015). Determination of total mercury in fish and sea products by direct thermal decomposition atomic absorption spectrometry. Food chemistry, 166, 432-441. https://doi.org/10.1016/j.foodchem.2014.06.032.

Parks, J. L., & Edwards, M. (2005). Critical Reviews in Environmental Science and Technology Boron in the Environment. *Critical Reviews in Environmental Science and Technology*, *35*(81), 81–114. https://doi.org/10.1080/10643380590900200

Pozebon, D., Scheffler, G. L., Dressler, V. L., & Nunes, M. A. (2014). Review of the applications of laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) to the analysis of biological samples. Journal of Analytical Atomic Spectrometry, 29(12), 2204-2228. https://doi.org/10.1039/C4JA00250D

Rainbow, P. S. (2018). Trace metals in the environment and living organisms: the British Isles as a case study. *Cambridge University Press*, UK. https://doi.org/10.1111/ejss.12809

Rasmussen, R. R., Hedegaard, R. V., Larsen, E. H., & Sloth, J. J. (2012). Development and validation of an SPE HG-AAS method for determination of inorganic arsenic in samples of marine origin. Analytical and bioanalytical chemistry, 403(10), 2825-2834. https://doi.org/10.1007/s00216-012-6006-7.

Reinholds, I., Bartkevics, V., Silvis, I.C.J., Van Ruth, S.M, & Esslinger, S. (2015). Analytical techniques combined with chemometrics for authentication and determination of contaminants in condiments: a review. *Journal of Food*

Composition and Analysis, 44, 56–72. https://doi.org/10.1016/j.jfca.2015.05.004

Remaud G.S., Giraudeau P., Lesot P., Akoka S. (2018). Isotope Ratio Monitoring by NMR: Part 1 – Recent Advances. In Modern Magnetic Resonance; Webb G. Eds; Springer: Cham, UK, pp 1353-1378.

Rentería-Cano, M. E., Sánchez-Velasco, L., Shumilin, E., Lavín, M. F., & Gómez-Gutiérrez, J. (2011). Major and trace elements in zooplankton from the Northern Gulf of California during summer. Biological trace element research, 142(3), 848-864. https://doi.org/10.1007/s12011-010-8820-0.

Ríos, S. E. G., Peñuela, G. A., & Botero, C. M. R. (2017). Method validation for the determination of mercury, cadmium, lead, arsenic, copper, iron, and zinc in fish through microwave-induced plasma optical emission spectrometry (MIP OES). Food Analytical Methods, 10(10), 3407-3414. https://doi.org/10.1007/s12161-017-0908-0.

Santos, L. F. P., Trigueiro, I. N. S., Lemos, V. A., da Nóbrega Furtunato, D. M., & Cardoso, R. D. C. V. (2013). Assessment of cadmium and lead in commercially important seafood from São Francisco do Conde, Bahia, Brazil. Food control, 33(1), 193-199. https://doi.org/10.1016/j.foodcont.2013.02.024.

Skałecki, P., Florek, M., Kędzierska-Matysek, M., Poleszak, E., Domaradzki, P., & Kaliniak-Dziura, A. (2020). Mineral and trace element composition of the roe and muscle tissue of farmed rainbow trout (Oncorhynchus mykiss) with respect to nutrient requirements: Elements in rainbow trout products. Journal of Trace Elements in Medicine and Biology, 62, 126619. https://doi.org/10.1016/j.jtemb.2020.126619.

Sneddon, J., & Vincent, M. D. (2008). ICP-OES and ICP-MS for the determination of metals: application to oysters. Analytical Letters, 41(8), 1291-1303. https://doi.org/10.1080/00032710802013991.

Stancheva, M., Makedonski, L., & Peycheva, K. (2014). Determination of heavy metal concentrations of most consumed fish species from Bulgarian Black Sea coast. Bulgarian Chemical Communications, 46(1), 195-203.

Sturrock, A.M., Hunter, E., Milton, J.A., EMIF, Johnson, R.A., Waring, C.P. & Trueman, C.N. (2015) Quantifying physiological influences on otolith microchemistry. *Methods in Ecology and Evolution. 6:* 806-816. https://doi.org/10.1111/2041-210X.12381

Tyler, G (2004). Rare earth elements in soil and plant systems - A review. *Plant Soil* 267, 191–206, https://doi.org/10.1007/s11104-005-4888-2

Tormen, L., Gil, R. A., Frescura, V. L., Martinez, L. D., & Curtius, A. J. (2012). The use of electrothermal vaporizer coupled to the inductively coupled plasma mass spectrometry for the determination of arsenic, selenium and transition metals in biological samples treated with formic acid. Analytica chimica acta, 717, 21-27. https://doi.org/10.1016/j.aca.2011.12.030.

Varrà, M. O., Ghidini, S., Zanardi, E., Badiani, A., & Ianieri, A. (2019). Authentication of European sea bass according to production method and geographical origin by light stable isotope ratio and rare earth elements analyses combined with chemometrics. *Italian Journal of Food Safety, 8*(1). https://doi.org/10.4081/ijfs.2019.7872

Warner, K., Mustain, P., Lowell,. B., Garen, S., & Talmage, S. (2016). Deceptive dishes: seafood swaps found worldwide, Oceana Reports. Available online at: http://usa.oceana.org/publications/reports/deceptive-dishes-seafood-swaps-found-worldwide.

Yeung, V., Miller, D. D., & Rutzke, M. A. (2017). Atomic Absorption Spectroscopy, Atomic Emission Spectroscopy, and Inductively Coupled Plasma-Mass Spectrometry. *Food Analysis,* 129–150. https://doi.org/10.1007/978-3-319-45776-5_9

Zhao, H., & Zhang, S. (2016). Effects of sediment, seawater, and season on multi-element fingerprints of Manila clam (Ruditapes philippinarum) for authenticity identification. *Food Control, 66*, 62–68. https://doi.org/10.1016/j.foodcont.2016.01.045

. Zhao, X., Liu, Y., Li, Y., Zhang, X., & Qi, H. (2018). Authentication of the sea cucumber (Apostichopus japonicus) using amino acids carbon stable isotope fingerprinting. Food Control, 91, 128-137. https://doi.org/10.1016/j.foodcont.2018.03.041.

Zhou, X., Taylor, M. P., Salouros, H., & Prasad, S. (2018). Authenticity and geographic origin of global honeys determined using carbon isotope ratios and trace elements. *Scientific Reports, 8*(1), 1–11. https://doi.org/10.1038/s41598-018-32764-w

Zkeri, E., Aloupi, M., & Gaganis, P. (2018). Seasonal and spatial variation of arsenic in groundwater in a rhyolithic volcanic area of Lesvos Island, Greece. *Environmental Monitoring and Assessment, 190*(1). https://doi.org/10.1007/s10661-017-6395-3

Zmozinski, A. V., Carneado, S., Ibáñez-Palomino, C., Sahuquillo, À., López-Sánchez, J. F., & Da Silva, M. M. (2014). Method development for the simultaneous determination of methylmercury and inorganic mercury in seafood. Food Control, 46, 351-359. https://doi.org/10.1016/j.foodcont.2014.05.054.

CHAPTER 5

Research Article

Stable Isotope Ratio and Rare Earth Elements Analyses for the Verification of Sea Bass Authenticity

Maria Olga Varrà[a]*, Sergio Ghidini[a], Emanuela Zanardi[a], Anna Badiani[b], and Adriana Ianieri[a]

[a]Department of Food and Drug, University of Parma, Strada del Taglio 10, 43126 Parma, Italy.

[b]Department of Veterinary Medical Science, University of Bologna, Via Tolara di Sopra 50, 40064 - Ozzano dell'Emilia, Bologna, Italy

*Correspondence: mariaolga.varra@studenti.unipr.it; Tel.: +39-0521-902-761

Adapted from

Varrà, M. O., Ghidini, S., Zanardi, E., Badiani, A., & Ianieri, A. (2019). Authentication of European sea bass according to production method and geographical origin by light stable isotope ratio and rare earth elements analyses combined with chemometrics. Italian journal of food safety, 8(1).

Abstract: In this work, stable isotope ratio (SIR) and rare earth elements (REEs) analyses, combined with multivariate data elaboration, were used to explore the possibility to authenticate European sea bass (*Dicentrarchus labrax* L.) according to: i) production method (wild or farmed specimens); ii) geographical origin (Western, Central or Eastern Mediterranean Sea).
The dataset under investigation included a total of 144 wild and farmed specimens coming from 17 different European areas located in the Mediterranean Sea basin. Samples were subjected to SIR analysis (carbon and nitrogen) and REEs analysis (lanthanum, europium, holmium, erbium, lutetium, and terbium). Then, Analytical data were handled by Principal Component Analysis (PCA) and then by Orthogonal Partial Last Square Discriminant Analysis (OPLS-DA), to obtain functional classification models to qualitatively discriminate sea bass according to the conditions under study. OPLS-DA models provided good correct classification rate both for production method and geographical origin. It was confirmed that chemometric elaboration of data obtained from SIR and REEs analyses can be a suitable tool for an accurate authentication of European sea bass.

Abbreviations: Central Mediterranean Sea, CM; Cross-validation, CV; Analysis of variance, ANOVA; Eastern Mediterranean Sea, EM; Farmed, F; Orthogonal partial last square-discriminant analysis, OPLS-DA; Principal component analysis, PCA; Principal component, PC; Rare earth elements, REEs; Root mean square error from cross-validation, RMSECV; Root mean square error of estimation, RMSEE; Root mean square error of prediction, RMSEP; Stable isotope ratio, SIR; Standard normal variate, SNV; Variable influence on projection, VIP; Wild, W; Western Mediterranean Sea, WM; Unit of Variance scaling, UV.

1. Introduction

Authentication of fishery and aquaculture products is of growing interest, due to the expansion and globalization of the fish chain and, on the other hand, by the public awareness concerning food quality and safety.

According to European Legislation, label descriptions of fish must include commercial designation, proper scientific name of the species, production method, and fishing area of provenance (Regulation EU No 1379/2013). Frequently these declared information are unreliable, so they need to be properly validated to ensure compliance with food legislation and traceability requirements, and to protect consumers from commercial food frauds (FAO, 2018).

Several techniques and analytical methods are traditionally used to measure features and properties that can discriminate different origins and production process of food. Among these, stable isotope ratio (SIR) and rare earth elements (REEs) analyses represent high- sensitive and well-established methodology, even if they are not yet widely used for authentication of fish.

The natural variation in isotopic abundances of light elements, as hydrogen (H), carbon (C), nitrogen (N), oxygen (O), and sulphur (S) can be largely influenced by the overall environmental conditions of the location from where the products originate. Isotopic composition of fish tissues, in particular, is the result of several complicated factors, such as positions in food chains, kind of feed ingested throughout all life, kinetic fractionation due to metabolism, geographical origin, physiological and anatomic properties, etc. (Ghidini et al., 2006). Thus, results of SIR analysis, if properly interpreted, can provide a unique *fingerprint* of the fish investigated.

Concentration of REEs has also been proposed to identify geographical origin of different kind of food matrices, because it mainly reflects the conditions of the production environment. REEs content in fishery and aquaculture products is strongly related to the vegetation eaten by fish, which reflects bio-available and mobilized nutrients present in the underlying soils (Drivelos and Georgiou, 2012), but its analytical determination for authentication of food of animal origin is still limited. The

increasing use of REEs for high-tech applications, (production of screen displays, glass, lenses, etc.), medical applications (medical x-ray and magnetic resonance image scanning systems), and, in some country like China, for usage as fertilizer or food supplements in animal production, are responsible for losses of REEs in the aquatic system (Mittermüller et al., 2016; Rim, 2016) and for permanent alterations of REEs natural pattern of the aquatic systems (Censi et al., 2004; Li et al., 010; Noack et al., 2014). This anthropogenic inputs of REEs in seawater, if properly studied, may be used to find out from which specific location the fish originate.

Thus, the aim of the present work was to verify whether chemometric elaboration of data obtained both from SIR analysis ($^{13}C/^{12}C$ and $^{15}N/^{14}N$ isotope ratios) and REEs analysis (lanthanum, La; europium, Eu; holmium, Ho; erbium, Er; lutetium, Lu; terbium, Tb) could provide a suitable tool to build classification models to be used for authentication of European sea bass samples according to production method (wild or farmed) and geographical origin (Western, Eastern, or Central Mediterranean Sea).

2. Materials and Methods

2.1. Sea bass samples

The sampling of 144 European sea bass (*Dicentrarchus labrax* L.) was carried out during 2 periods in 2012 (spring-summer, n = 77; autumn-winter, n = 77). All the specimens came from fishing areas (wild subjects, W; n = 34) or fish farms (farmed subjects, F; n = 110) located in the Mediterranean basin. Total length (measured from the most anterior part of the fish to the most posterior part of the caudal fin) was 42.7 ± 8.0 cm and 35.3 ± 4.6 cm for wild and farmed subjects, respectively; the recorded body weight was 930.0 ± 488.6 g in wild and 517.9 ± 183.3 g in farmed specimens. Farmed samples, moreover, included intensively (submersible or floating cages located in various open sea areas), semi-intensively (earthen tanks), and extensively (coastal lagoons), reared samples, whose breeding system varied based on the stocking density declared by farmers.

From a geographical point of view, the whole dataset was composed by fifty sea bass samples originating from Western Mediterranean Sea (WM; FAO fishing subareas 37.1.2 and 37.1.3), sixty-four from Central Mediterranean

Sea (CM; FAO fishing subareas 37.2.1 and 37.2.2) and thirty were from Eastern Mediterranean Sea (EM; FAO fishing subareas 37.3.1).

2.2. Stable isotope ratio analysis

The stable isotopic composition of carbon ($^{13}C/^{12}C$, expressed as $\delta^{13}C$) and nitrogen ($^{15}N/^{14}N$, expressed as $\delta^{15}N$) in sea bass muscle was carried out through the procedure described below: 0.07 ± 0.01 and 0.7 ± 0.1 mg for $\delta^{13}C$ and $\delta^{15}N$ analyses, respectively, of freeze-dried fillets (<0.6 mm grain size) were analysed by means of an EA/NA-1100 elemental analyser with CHN configuration in helium continuous flow mode and coupled to a Finnigan Delta Plus XP mass spectrometer (Thermo Finnigan, Bremen, Germany). Sample isotope ratios were then calculated according to international reference standard V-PDB (Vienna PeeDee Belemnite) for C and atmospheric nitrogen for N. Results were expressed as follows:

$$\delta (X) = (\text{R sample} - \text{R standard} - 1) \times 1000‰$$

where δ is the notation in parts per thousands (‰) relative to the specific international standard material, X is the content of ^{13}C or ^{15}N in the sample, and R is the molar ratio of the heavy to light isotope of the element (R = $^{13}C/^{12}C$ or $^{15}N/^{14}N$).

2.3. Rare earth elements analysis

Sea bass were subjected to mineralization by adding 6 mL of nitric acid (67%, Ultrapure, Merck, Darmstadt, Germany) and 2 mL of hydrogen peroxide (31%, Ultrapure, Merck Darmstadt, Germany) to 1 g of sample, through a Milestone 1200 Mega microwave system (FKW S.r.L., Italy). Detection of the elements was obtained by means of Thermo X series inductively coupled plasma mass spectrometry (ICP−MS, Thermo Fisher Scientific, Waltham, MA, U.S.A.).

2.3. Statistical analysis

Independent t-test and analysis of variance (one-way ANOVA) with Tukey HSD post hoc tests were carried out using SPSS software (v. 23.0, SPSS Inc., Chicago, IL, USA), to determine significant differences among the means of each group. The level of significance was $p \leq 0.05$.

Multivariate data analysis of spectrometric data of SIR and REEs was performed by using SIMCA-P software (v.14.1). Data were firstly normalized by unit of variance (UV) scaling and log-transformed to reduce skewness. Subsequently, principal component analysis (PCA) was performed with the aim to take an overview of data natural structure and, eventually, eliminate strong outliers according to Hotelling's T^2 test (at 95% confidence interval). After that, classification models able to distinguish samples according to their production method and geographical origin were developed by means of a supervised orthogonal partial least squares–discriminant analysis (OPLS-DA). OPLS-DA calibration models were built using only 75% of the samples (calibration set, n = 108), while the remaining 25% of the data was reserved to externally validate OPLS-DA calibration results (prediction set, n = 36). During the external validation stage, the resulting percentage of correctly classified observations and the RMSEP (Root Mean Square Error of Prediction) were evaluated to assess overall classification performances.

PCA and OPLS-DA models computed were internally validated by a 7-fold cross-validation (CV) and their quality assessed by the statistical parameters R^2X_{cum} (representing goodness of fit), Q^2_{cum} (representing the goodness of prediction) and, only for OPLS-DA models, R^2Y_{cum} (representing total sum of variation explained in Y-matrix)

RMSECV (Root Mean Square Error from cross-validation), and RMSEE (Root Mean Square Error of Estimation) were further calculated to evaluate the reliability of the classifiers.

Finally, the Variable Influence on Projection (VIP) index for OPLS-DA components was used to identify the most discriminating variables in the OPLS classification. Variables with a VIP values ≥1 were considered significant (Galindo-Prieto et al., 2014).

3. Results

3.1. REEs concentration and SIR in sea bass samples

An overview of the results of sea bass isotopic analysis of C and N and REEs analysis from different production methods and geographic origins is reported in Table 1. As it can be observed, $\delta^{13}C$ and $\delta^{15}N$ mean values of

wild and farmed samples ranged from -20.33‰ to -17.25‰, and from 10.90‰ to 14.75‰, respectively. Both $\delta^{13}C$ and $\delta^{15}N$ values were significantly higher in wild sea bass compared with farmed ones ($p \leq 0.05$). As for geographic provenance, CM sea bass were found to be isotopically heavier in $\delta^{13}C$ values (-18.87‰) than EM group (-21,19‰; $p \leq 0.05$), while WM samples isotopically heavier in $\delta^{15}N$ value (12.80‰) compared to EM and CM group ($p \leq 0.05$).

The highest concentrations of almost all REEs were recorded in farmed sea bass on average (Table 1); only Lu values seemed to be higher in wild specimens (Lu = 0,0538 ng g^{-1}) than in farmed ones (Lu = 0,0417 ng g^{-1}), as well as Eu, Ho and Tb seemed to be higher for WM samples, and Er and Lu for CM samples. Nevertheless, significantly differences in REEs concentration were not found among all the groups ($p \geq 0.05$).

3.1. Chemometrics

PCA of the pre-treated SIR and REEs data of the 144 sea bass samples was fitted with a total of 5 principal components (PCs) ($R^2X = 0.819$, $Q^2 = 0.768$) and the first two PCs explained 78% of the total variation in the original dataset.

Despite the good statistics associated to the model, the visual separation among different groups of samples in the bidimensional score plot was not acceptable, for both wild/farmed and CM/EM/WM sea bass (data not shown). As a matter of fact, overlapping was observed between them, indicating the non-effectiveness of the unsupervised technique to successfully discriminate samples according to the authentication purposes investigated.

Anyway, PCA model don't show outliers (no samples outside the 95% confidence level according to Hotelling T^2 test) and, for this reason, none of them was removed from the dataset for the subsequent elaboration.

Two supervised OPLS-DA models were further built on calibration set (n=108) and validated by means of a 7-fold cross-validation to avoid overoptimistic and misleading results.

The first model aimed to classify samples by production method (W vs. F). This model was fitted with 1 predictive component (correlated variation)

and 2 orthogonal components (uncorrelated variation), that captured 69% of the total variation in the X-matrix (R^2X_{cum} = 0.695).

Table 1. Stable isotopic composition of carbon (‰) and nitrogen (‰) and rare earth concentrations (ng g^{-1}) of sea bass samples according to production methods (W or F) and geographical origin (CM, WM, or EM) (mean ± standard deviation).

Parameter	Production		Origin		
	W (n = 34)	F (n = 110)	CM (n = 64)	WM (n = 50)	EM (n =30)
$\delta^{13}C_{PDB-1}$	-17,25 ± 7,12^a	-20,33 ± 1,08^b	-18,87 ± 5,40^a	-19,60 ± 1,64^{ab}	-21,19 ± 0,73^b
$\delta^{15}N_{AIR}$	14,75 ± 3,42^a	10,90 ± 1,32^b	11,68 ± 2,11^a	12,80 ± 3,39^b	10,45 ± 0,86^c
La	4,6095 ± 3,4829^a	4,9649 ± 3,2167^a	5,1649 ± 3,1800^a	5,0074 ± 3,3683^a	4,0647 ± 3,3457^a
Eu	0,1555 ± 0,5055^a	0,2651 ± 0,6319^a	0,1858 ± 0,4954^a	0,3430 ± 0,7538^a	0,1802 ± 0,5382^a
Ho	0,0776 ± 0,0546^a	0,0848 ± 0,0523^a	0,0857 ± 0,0488^a	0,0876 ± 0,0600^a	0,0700 ± 0,0483^a
Er	0,3288 ± 0,2609^a	0,3655 ± 0,2180^a	0,3759 ± 0,2048^a	0,3703 ± 0,2607^a	0,2939 ± 0,2203^a
Lu	0,0417 ± 0,0306^a	0,0538 ± 0,0345^a	0,0547 ± 0,0328^a	0,0505 ± 0,0385^a	0,0438 ± 0,0279^a
Tb	1,7955 ± 1,4363^a	2,0267 ± 1,5482^a	2,0409 ± 1,5128^a	2,0863 ± 1,5607^a	1,6352 ± 1,4958^a

Mean ± SD (n = 9) followed by different letters in the same row (W and F, or CM, WM, and EM) are significantly different ($p \leq 0.05$).

The predictability power, given by the Q^2_{cum} parameter, was 0.903, indicating a good separation capability and the predictive variation contained in the data described 95% of the class membership information (R^2Y_{cum}=0.952).

The score scatter plot reporting spatial distribution of wild and farmed samples (Figure 1) revealed a high interclass variability along the first predictive component (horizontal direction): the W sea bass groudistributed on the negative side, and the F sea bass along the positive one. At the same time, a remarkable intraclass variability was observed, both for W and F samples, which largely distributed also along first orthogonal component (vertical direction). The second OPLS-DA model aimed to classify samples by geographical provenance (WM vs. CM vs. EM), and it was fitted with 2 predictive components and 3 orthogonal ones. Even if this three-class model was characterised by lower values of R^2X_{cum} (=0.599), R^2Y_{cum} (=0.907), and Q^2_{cum} (=0.893) values compared to model discriminating production methods, it was considered acceptable. The score scatter plot showing space distribution and grouping of samples according to their geographical class

membership, is reported in Figure 2. Three different clusterings were observed: WM samples were distributed along the negative axes of the first predictive component (horizontal direction) and of the second predictive component (vertical direction); EM samples were prevalently distributed along the positive axis of the first predictive component and the negative axis of the second predictive component; CM samples were distributed along the positive axis of the second predictive component.

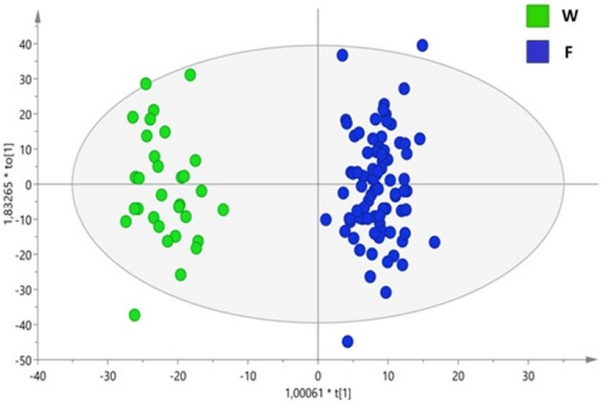


Figure 1. OPLS-DA Score Plot discriminating wild from farmed sea bass (W = wild; F = farmed).

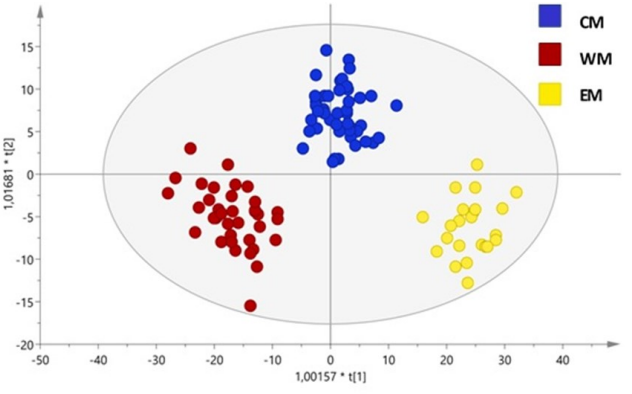


Figure 2. OPLS-DA Score Plot discriminating CM from WM from EM sea bass **(**CM = Central Mediterranean; WM = Western Mediterranean; EM= Eastern Mediterranean).

Mean RMSECV and RMSEE values calculated for model discriminating samples by production method were found to be 0.178 and 0.060, respectively, while for model discriminating samples by provenance, they were found to be 0.295 and 0.102, respectively.

Each OPLS-DA model was further tested for predictability of new sea bass samples by using the independent prediction set (n = 36). Misclassification table summarizing the proportion of correctly classified observations into the known classes of the prediction set is presented in Table 2.

As for production method, it can be observed that 100% of samples were correctly classified (mean RMSEP = 0.211), while with regards to geographical origin, some misclassifications occurred, which led to a total correct classification percentage of 90% of sea bass (mean RMSEP = 0.297).

Finally, the VIP index was employed as an indicator of the variables that mostly contributed to discriminate samples in each OPLS-DA model.

It was found that $\delta^{15}N$ (VIP = 2.30) and $\delta^{13}C$ (VIP = 1.45) strongly influenced discrimination of W from F samples. Similarly, the most relevant variables influencing discrimination by provenance were found to be $\delta^{15}N$ (VIP = 1.23) and $\delta^{13}C$ (1.38), but also La (VIP = 1.08) and Ho (VIP = 1.02).

Table 2. Classification performances tested on prediction set (n = 36)

OPLS-DA model	**Class**	**Classification rate (per class)**	**Overall classification rate**
Production method	W	100.0% (8/8)	100.0% (36/36)
	F	100.0% (28/28)	
Geographical origin	WM	92.3% (12/13)	88.9% (32/36)
	CM	81.2% (14/16)	
	EM	85.7% (6/7)	

4. Discussion

Increasing attention is given to the development of methods for ensuring fish authenticity and discouraging the spreading of fraudulent practises. Light isotopes abundances combined with rare earth elements fingerprinting may represent a valid strategy for this purposes, especially when properly

elaborated by chemometrics to develop qualitative discrimination model.

In fishery and aquaculture products, the isotopic abundances $\delta^{13}C$ and $\delta^{15}N$ are linked to the feed habitat and to the trophic position of the fish, respectively. Wild marine fish tissues are usually isotopically heavier in $\delta^{13}C$ values than farmed fish, because the dissolved carbon pool of the natural diets of wild fish ($\delta^{13}C$ = 0%) is significantly more enriched than the atmospheric carbon (CO_2) used for photosynthesis by the terrestrial plants, that can contribute to the formulation of commercial feeds ($\delta^{13}C$ = -8%) (Busetto et al., 2008). As for nitrogen isotopes, fish at higher trophic level also shows higher values of $\delta^{15}N$ (Camin et al., 2017), even if this value can be influenced by other factors, such as maturity or growth rate, seasonal isotopic variations in coastal marine environments, and possible spoilage of the fish sample between collection and analysis. Thus, particular attention should be taken to $\delta^{15}N$ data interpretation (Morrison et al., 2007).

In the present work, both $\delta^{15}N$ and $\delta^{13}C$ were found to be higher in wild specimens compared to the farmed specimens, to the point of allowing a perfect discrimination of the two populations based just on such knowledge. These results are in accordance to what has been previously reported in literature: studies based on determination of isotopes abundances of C and N successfully led to the discrimination between wild and farmed gilthead sea bream (Morrison et al., 2007; Serrano et al., 2007), shrimps (Gamboa-Delgado et al., 2014; Ortea and Gallardo, 2015), Atlantic salmon (Dempson and Power, 2004), and sea bass (Bell et al., 2007). The same method allowed to differentiate three commercial fish (mackerel, yellow croaker and pollock), originating from various countries (Li et al, 2018) and anchovy sampled from inshore habitats, from those from offshore habitats (Tanaka et al., 2010). In our study, values of $\delta^{15}N$ and $\delta^{13}C$ also made the highest contributions in discriminating sea bass by geographical origins, probably because of the differences in prey availability of the various fishing areas or the differences in feed composition used by fish farmers (Moreno Rojas et al., 2007.

The use of REEs as chemical markers for fish provide several advantages (i.e. safety of handling, bioaccumulation, long retention time in tissues, etc.), but few studies have been performed, mainly because of the high accuracy

and sensitivity of instruments required for their detection and quantification (Danezis et al., 2016). Early few studies identified samarium as a marker goldfish (Michibata, 1981) as well as dysprosium, europium, and samarium for rainbow trout (Giles and Attas, 1993). Only recently, Pérez de Nanclares et al. (2016) supplemented the fish diet with praseodymium, neodymium, dysprosium, cerium, and lanthanum and successfully discriminated production method and site of origin of salmon samples.

In this work, REEs concentrations did not influence separation of sea bass samples by production method. Despite farmed specimens were characterised by higher contents of mostly the elements studied, these contents were not strictly related to the different feeding input of the two populations. However, Ho e La were found to be influential in the discrimination by geographical provenance, as a result of their natural variability in the marine environment or as a consequence of the anthropogenic pollution.

4. Conclusions

In summary, this experiment demonstrates that SIR and REEs analyses combined with multivariate statistics can successfully be used for authentication of European sea bass according to production method and geographical origin. The classification models built were able to correctly discriminate wild from farmed specimens with 100% of accuracy and samples coming from Western, Central, and Eastern Mediterranean Sea with 89% of accuracy. Discriminant information, either way, was found to lie prevalently in isotopic abundances of carbon and nitrogen. Additionally, among six rare earth elements measured, lanthanum and holmium showed an important contribution in sample discrimination based on provenance.

Thus, the outcomes of this study demonstrate that differences in composition of the feeding sources and trophic position of fish, that influence the ratio of stable isotope of carbon and nitrogen, as well as the natural or artificial occurrence of REEs in the environment, to which the fish is exposed throughout his life, together could be adopted as an analytical strategy to effectively detect mislabelling involving the origin and thus, guarantee fish and fishery products authenticity.

References

Bell, J.G., Preston, T., Henderson, R.J., Strachan, F., Bron, J.E., Cooper, K., & Morrison, D.J. (2007). Discrimination of wild and cultured European sea bass (Dicentrarchus labrax) using chemical and isotopic analyses. *Journal of Agricultural and Food Chemistry, 55*, 5934–41.
http://dx.doi.org/10.1021/jf0704561

Busetto, M.L., Moretti, V.M., Moreno-Rojas, J.M., Caprino, F., Giani, I., Malandra, R., Bellagamba, F., & Guillou, C.J. (2008). Authentication of farmed and wild turbot (Psetta maxima) by fatty acid and isotopic analyses combined with chemometrics. *Journal of Agricultural and Food Chemistry*, *56*, 2742-50.
http://dx.doi.org/10.1021/jf0734267

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 and Technology, 61*, 176–87.
http://dx.doi.org/10.1016/j.tifs.2016.12.007

Censi, P., Mazzola, S., Sprovieri, M., Bonanno, A., Patti, B., Punturo, R., & Alonzo, G. (2004). Rare earth elements distribution in seawater and suspended particulate of the Central Mediterranean Sea. *Chemical Ecology, 20*, 323–43.
http://dx.doi.org/10.1080/02757540410001727954

Danezis, G.P., Tsagkaris, A.S., Brusic, V., & Georgiou, C.A. (2016). Food authentication: state of the art and prospects. *Current Opinion in Food Science, 10*, 22–31.
http://dx.doi.org/10.1016/j.cofs.2016.07.003

Dempson, J.B., & Power, M., (2004). Use of stable isotopes to distinguish farmed from wild Atlantic salmon, Salmo salar. *Ecology of Freshwater Fish, 13*, 176–84.
http:///doi.wiley.com/10.1111/j.1600-0633.2004.00057.x

Drivelos, S.A. & Georgiou, C.A. (2012). Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. *TrAC, Trends in Analitycal Chemistry, 40*, 38–51.
http://dx.doi.org/10.1016/j.trac.2012.08.003

FAO (2018). Overview of food fraud in the fisheries sector. Fisheries and Aquaculture Circular No. 1165. Rome, Italy.

European Parliament and Council of the European Union, 2013. Regulation of the European Parliament and of the Council of 11 December 2013 on the common organization of the markets in fishery and aquaculture products, 1379/2013/EU. In: Official Journal of the European Union, L354/1–21, 11-12-2013.

Galindo-Prieto, B., Eriksson, L., & Trygg, L. (2014). Variable influence: on projection (VIP) for orthogonal projections to latent structures (OPLS). *Journal of Chemometrics, 28*, 623-32.
https://dx.doi.org/10.1002/cem.2627

Gamboa-Delgado, J., Molina-Poveda, C., Godínez-Siordia, D.E., Villarreal-Cavazos, D., Ricque-Marie, D., Cruz-Suárez, L.E., & Gillanders, B. (2014). Application of

stable isotope analysis to differentiate shrimp extracted by industrial fishing or produced through aquaculture practices. *Canadian Journal of Fisheries and Aquatic Sciences, 71,* 1520–28.
http://www.nrcresearchpress.com/doi/abs/10.1139/cjfas-2014-0005

Ghidini, S., Ianieri, A., Zanardi, E., Conter, M., Boschetti, T., Iacumin, P., & Bracchi, P.G. (2006). Stable isotopes determination in food authentication: a review. *Annuali della Facoltà di Medicina Veterinaria di Parma, 26*, 193–204.

Giles, M.A. & Attas, E.M. (1993). Rare earth elements as internal batch marks for rainbow trout: retention, distribution, and effects on growth of injected dysprosium, europium, and samarium. *Transactions of the American Fishery Society, 122,* 289–97.

Li, J., Hong, M., Yin, X., & Liu, J. (2010). Effects of the accumulation of the rare earth elements on soil macrofauna community. *Journal of Rare Earths, 28*, 957–64.
http://dx.doi.org/10.1016/S1002-0721(09)60233-7

Li, L., Han, C., Dong, S., & Boyd, C.E. (2018). Use of elemental profiling and isotopic signatures to differentiate Pacific white shrimp (Litopenaeus vannamei) from freshwater and seawater culture areas. *Food Control, 95*, 249–256.
https://doi.org/10.1016/j.foodcont.2018.08.015

Michibata, H. (1981). Labeling Fish with an Activable Elemen Through Their Diet. *Canadian Journal of Fishery and Aquatic Sciences, 38*, 1281–82.
https://doi.org/10.1139/f81-017

Mittermüller, M., Saatz, J., &Daus, B. (2016). A sequential extraction procedure to evaluate the mobilization behavior of rare earth elements in soils and tailings materials. *Chemosphere, 147,* 155–62.
http://dx.doi.org/10.1016/j.chemosphere.2015.12.101

Moreno Rojas, J.M., Serra, F., Giani, I., Moretti, V.M., Raniero, F., & Guillou, C. (2007). The use of stable isotope ratio analyses to discriminate wild and farmed gilthead sea bream (Spaurus aurata). *Rapid Communications in Mass Spectrometry, 21,* 207-11.
https://doi.org/10.1002/rcm.2836

Morrison, D.J., Preston, T., Bron, J.E., Hemderson, R.J., Cooper, K., Strachan, F., & Bell, J.G. (2007). Authenticating production origin of gilthead sea bream (Sparus aurata) by chemical and isotopic fingerprinting. *Lipids, 42*, 537–45.
https://doi.org/10.1007/s11745-007-3055-3

Noack, C.W., Dzombak, D.A., & Karamalidis, A.K. (2014). Rare earth element distributions and trends in natural waters with a focus on groundwater. *Environmental Science and Technology, 48,* 4317–26.
https://doi.org/10.1021/es4053895

Ortea, I. & Gallardo, J.M. (2015). Investigation of production method, geographical origin and species authentication in commercially relevant shrimps using stable isotope ratio and/or multi-element analyses combined with chemometrics: An exploratory analysis. *Food Chemistry, 170*, 145–53.
http://dx.doi.org/10.1016/j.foodchem.2014.08.049

Pérez de Nanclares, M., Dessen, J.E., Rørvik, K.A., Thomassen, Y., & Thomassen, M.S. (2016). Feasibility of using rare earth elements (REEs) to mark and

identify escaped farmed Atlantic salmon Salmo salar L. *Aquaculture Research, 47*, 1885–98. http://doi.wiley.com/10.1111/are.12647

Rim, K.T. (2016). Effects of rare earth elements on the environment and human health: A literature review. *Journal of Toxicological and Environmental Health Sciences, 8,* 189–200. https://doi.org/10.1007/s13530-016-0276-y

Serrano, R., Blanes, M.A., & Orero, L. (2007). Stable isotope determination in wild and farmed gilthead sea bream (Sparus aurata) tissues from the western Mediterranean. *Chemosphere, 69*, 1075–80.
https://doi.org/10.1016/j.chemosphere.2007.04.034

Tanaka, H., Ohshimo, S., Takagi, N., & Ichimaru, T. (2010). Investigation of the geographical origin and migration of anchovy Engraulis japonicus in Tachibana Bay, Japan: A stable isotope ap
proach. *Fishery Research, 102*, 217-20.
https://doi.org/10.1016/j.fishres.2009.11.002

CHAPTER 6

Research Article

Multi-Element Signature of Cuttlefish and its Potential for the Discrimination of Different Geographical Provenances

Maria Olga Varrà[a], Lenka Husáková[b]*, Jan Patočka[b], Sergio Ghidini[a], and Emanuela Zanardi[a]*

[a]Department of Food and Drug, University of Parma, Strada del Taglio 10, 43126 Parma, Italy.

[b]Department of Analytical Chemistry, Faculty of Chemical Technology, University of Pardubice, Studentska 573 HB/D, Pardubice, CZ-532 10, Czech Republic

*Correspondence: lenka.husakova@upce.cz; Tel.: +420 466 037 029
emanuela.zanardi.unipr.it; Tel.: +39-0521-902-760

Under Review in Food Chemistry

Abstract: The measurement and analysis of more than fifty elements by quadrupole inductively coupled plasma mass spectrometry (Q-ICP-MS) and direct mercury analysis were applied to origin discrimination of Italian traditional cuttlefish (Chioggia, Venice lagoon) from Mediterranean and Atlantic samples. A total 68 specimens were analysed in triplicates to generate 204 mass spectra profiles which were statistically processed by different chemometric techniques. Loading weights from principal component analysis as input for linear discriminant analysis (LW-LDA), stepwise-LDA (S-LDA) and variable influence of projection-partial least square discriminant analysis (VIP-PLS-DA) were used to classify samples while retaining the lowest possible number of key variables. VIP-PLS-DA was found to be the best variable selection-discriminant tool combo since the selected Na–Co–B–K–Cd–V–U–Rb–Ni–Ba–Cu–As–Sr–Mn–Mo–Li–Ca–Mg–Se–Bi–Cs–P–Y elemental pattern allowed to classify samples with 100% sensitivity, specificity and accuracy.

Abbreviations: Central Mediterranean Sea, CM; Cross-validation, CV; Certified reference materials, CRMs, correlation analysis, CA; discriminant function, DF; high energy Helium mode, HE He; inductively coupled plasma mass spectrometry, ICP-MS; inductively coupled plasma optical emission spectrometry, ICP-OES; internal standard, ISTD; kinetic energy discrimination, KED; linear discriminant analysis, LDA; leave-one-out-cross-validation, LOOCV; loading weights-based linear discriminant analysis, LW-LDA; method detection limit of the method, MDL; method limit of quantification, MLOQ; microwave digestion, MWD; partial last square-discriminant analysis, PLS-DA; principal component, PC; principal component analysis, PCA; quadrupole inductively coupled plasma mass spectrometry, Q-ICP-MS; rare earth elements, REEs; relative standard deviation, RSD; root mean square error from cross-validation, RMSECV; root mean square error of estimation, RMSEE; root mean square error of prediction, RMSEP; stepwise- linear discriminant analysis, S-LDA; variable influence on projection, VIP; variable influence of projection-partial least square discriminant analysis, VIP-PLS-DA.

1. Introduction

The common cuttlefish (*Sepia officinalis,* L., 1758) is a highly valuable fishery product whose distribution extends from the Eastern Atlantic (from the Baltic and North Seas to South Africa coasts) up to the whole Mediterranean Sea. Large scale fishing operations of common cuttlefish led to an average production of 22604 tons in 2016, which ensured the product's supply all over the world (especially in Italy, Spain, Japan, and South Korea), while smaller-scale fishing is particularly profitable in some Mediterranean areas where it has a high impact on local economies (FAO, 2010). This activity is particularly developed in the fishing village of Chioggia (north-western Adriatic Sea in the Eastern Mediterranean Basin), whose *S. officinalis* production reached more than 594 tons in 2016 (Clodia database, 2017).

Chioggia cuttlefish is well known to be of high quality thanks to the preservation of a close link with the territory and the traditional processing methods by local fish plants, which make the products to be officially certified as Italian Traditional Agri-food Product (P.A.T.) (Ministerial Decree, G.U. n. 48, 2014). Foodstuffs with specific geographical ties (especially when a quality mark has been recognized) are in turn rated more highly in terms of quality by the consumer, mostly because of lower environmental impact and higher perceived safety of regional products.

Alongside with the compulsory labelling requirement established by European legislation concerning the provision of detailed information about the geographical origin of fish and seafood (European Parliament and Council of the European Union, 2013), the protection and the promotion of traditional foods must involve origin authenticity assessment both to ensure traceability and prevent commercial frauds (Ortea & Gallardo, 2015).

Complex multi-disciplinary and cross-disciplinary approaches are usually required to verify the geographical origin of fish and seafood since both environment and genetics can affect the final characteristics of the products (Abbas et al., 2018). Nevertheless, considering the association between the concentrations of elements in fish tissues and those in the surrounding aquatic environment, the determination of the multi-element profile of fish and seafoods can be regarded as a valid analytical strategy to guarantee fish

origin authenticity. Several factors such as species, size, age, sex, and sexual maturity can directly affect the elemental composition of aquatic animals, but food resources, climate, presence of contaminants, and many water quality parameters (*p*H, dissolved oxygen, alkalinity) are those which differ to a greater extent between countries (Li, Boyd, & Sun, 2016; Smith & Watts, 2009).

Several studies reported the possibility to discriminate fishery and seafood products as croaker (Chaguri et al., 2015), sea cucumber (Kang et al., 2018), shrimps (Ortea & Gallardo, 2015), crabs (Luo et al., 2019), and bivalve molluscs such as mussels (Costas-Rodríguez, Lavilla, & Bendicho, 2010) and clams (Iguchi, Isshiki, Takashima, Yamashita, & Yamashita, 2014) using element composition. In these works, the chemical characterization of samples was based on minor and/or trace elements (Costas-Rodríguez et al., 2010; Iguchi et al., 2014), a combination of major, minor, and trace elements (Kang et al., 2018), or a combination of major and trace elements plus stable isotopes analysis (Chaguri et al., 2015; Ortea et al., 2015; Luo et al., 2019).

In the range of analytical methods applied to the determination of elements, inductively coupled plasma-mass spectrometry (ICP-MS) combined with different chemometrics techniques holds a unique position by virtue of speed, sensitivity, dynamic range, and elemental coverage (Drivelos & Georgiou, 2012; Danezis, Tsagkaris, Camin, Brusic, & Georgiou, 2016). In this context, several supervised and unsupervised chemometric tools such as, principal component analysis (PCA) (Chaguri et al., 2015; Kang et al., 2018; Ortea et al., 2015), cluster analysis and *k*-means hierarchical clustering (Ortea et al., 2015), linear discriminant analysis (LDA) (Costas-Rodríguez et al., 2010; Iguchi et al., 2014; Ortea et al., 2015; Kang et al., 2018; Luo et al., 2019), support-vector machine (Luo et al., 2019), soft-independent modelling of class analogy and artificial neural networks (Costas-Rodríguez et al., 2010) were applied for authenticity and traceability purposes.

The determination of elemental composition of fish and seafood to trace their geographical provenance is particularly suitable for cephalopods molluscs. Cephalopods such as cuttlefish are in fact at high trophic levels in the aquatic food chain and the overall elements accumulated in tissues are good indicators of the surrounding habitat (Bosch, O'Neill, Sigge, Kerwath,

& Hoffman, 2015). In addition, cephalopods are non-migratory animals and, therefore, their elemental composition can reasonably be expected to be during life (Gopi et al., 2019). Nevertheless, very few studies addressed the topic of origin discrimination of cephalopods molluscs using elemental composition, in spite of being focused on the analysis of a limited number of major and trace elements included in hard structures such as statoliths (Arbuckle & Wormuth, 2014) or ink (Bua et al., 2017). These components are in fact not always retained in the final commercial product and, therefore, not always exploitable in practical food surveillance operations to trace back the origin of cephalopods.

Based on this background, the present work aimed at outlining for the first time the *S. officinalis* multi-elemental profile of edible tissues to verify whether useful information could be extracted and specifically linked to the geographical origin of the Italian traditional cuttlefish from Chioggia (FAO subarea 37.2.1) in order to be differentiated from non-Chioggia cuttlefish caught in other fishing areas of the Mediterranean Sea (FAO 37.1/37.2) and in the French Atlantic Ocean (FAO 27.7.e). For this purpose, fifty-one elements were determined by ICP-MS and Hg was quantified via single-purpose atomic absorption spectrometer AMA254. Unsupervised learning (PCA) and supervised pattern recognition methods based on LDA and partial least square-discriminant analysis (PLS-DA) merged with different independent variable selection tools were used to investigate sample characteristics, identify the key discriminant elements, and, at the same time, develop classification rules.

2. Materials and Methods

2.1. Reagents and standards

The ultrapure water (0.055 μS cm^{-1} conductivity) obtained using the Milli-Q® water purification system (Millipore, Bedford, USA) was used for the preparation of all solutions. Sub-boiled nitric acid was prepared from nitric acid (65%, w/w) of Selectipur quality (Lach-Ner, Neratovice, Czech Republic) using the distillation equipment BSB-939-IR (Berghof, Eningen, Germany). Hydrogen peroxide (Trace Select, ≥ 30%, w/w) was purchased from Fluka Chemie AG (Buchs, Switzerland). Multi-element

stock solution "A" containing 10 mg L^{-1} of Li, B, Al, V, Cr, Fe, Mn, Ni, Cu, Zn, Co, As, Se, Rb, Sr, Zr, Mo, Ru, Pd, Cd, Sn, Sb, Cs, Ba, Hf,, Re, Pt, Tl, Pb, Bi, and Th was prepared from the Supelco ICP multi-element standard solution IV (Merck, Darmstadt, Germany) and single element standards of concentration 1 ± 0.002 g L^{-1} (Analytika Ltd., Prague, Czech Republic or SCP Science, Montreal, Canada). Multi-element solution "B" containing 1 mg L^{-1} of La, Ce, Pr, Nd, U ("B1") and 0.20 mg L^{-1} of Y, Tb, Ho, Yb, Sm, Eu, Gd, Er, Tm, Lu, and Dy ("B2") was prepared from the stock solution of rare earth elements Astasol mix "M008" (Analytika Ltd., Prague, Czech Republic). Multi-element solution "C" containing 50 mg L^{-1} of Na, Mg, P, K, Ca, Mn, Cu, and Zn was prepared from single element standards of 1 g L^{-1} obtained from Analytika Ltd. The internal standard solution (ISTD) was prepared from 1 g L^{-1} stock solution of Rh obtained from SCP Science (Montreal, Canada). Carbon reference solutions were prepared from 10 g L^{-1} of stock solution prepared from urea of TraceSelect quality (Fluka Chemie AG, Buchs, Switzerland).

2.2. Sampling and preparation of cuttlefish

A total of 68 samples of common cuttlefish (*S. officinalis,* L.) including n = 17 samples from Adriatic Sea (Chioggia, Italy, FAO 37.2.1), n = 25 non-Chioggia samples from the Mediterranean Basin (FAO 37.1/37.2), and n = 26 samples from North-eastern Atlantic Ocean (France, FAO 27.7.e) were analysed. Specimens were caught by fishing trawlers during the months of September and randomly collected from different batches in local fish plants located in Chioggia (Venice, Italy), by choosing homogeneous sizes and weights (mantle lengths from 10 ± 2 cm; body weight 125 ± 25 g). After collection, samples were immediately frozen and stored at −20 °C for around 3 months prior to be further processed for multi-element analysis.

2.3. Quality assurance and quality control

The following commercially supplied reference materials (RMs) were analysed: NIST SRM 1577 Bovine Liver (National Institute of Science and Technology, NIST, Gaithersburg, MD, USA); NIST SRM 1566 Oyster Tissue (NIST, Gaithersburg, MD, USA); BCR® certified reference material (CRM)184 Bovine muscle (Institute for Reference Materials and

Measurements, IRMM, Geel, Belgium); BCR® 185 Bovine Liver (IRMM, Geel, Belgium); CRM NCS ZC73015 Milk Powder (National Research Centre for Certified Reference Materials, NRCRM, Beijing, China); P-WBF CRM 12-2-04 Essential and Toxic Elements in Wheat Bread Flour (pb-anal, Kosice, Slovakia); CRM12-2-03 P-Alfalfa Essential and toxic elements in Lucerne (pb-anal, Kosice, Slovakia); SMU CRM 12-02-01 Bovine liver (pb-anal, Kosice, Slovakia).

2.4. Sample preparation

2.4.1 Pre-processing of cuttlefish

Frozen samples were thawed overnight (16–18 h at +4 °C) before processing for multi-element analysis. Each specimen was then washed with deionized water and skin, cuttlebone, gills, reproductive and digestive tracts were carefully removed without causing their rupture. The head, arms and tentacles were excluded, while the mantle and the lateral fins were rinsed again with ultrapure water and minced with a ceramic knife.

2.4.2. Freeze-drying process

Pre-treated sample's portions of approx. 10 g were individually transferred into 100 mL cleaned round bottom flasks wherein the material was dried and placed into a deep freezer at −80 °C for 24 hours to provide a necessary conditioning for drying. Lyophilization was conducted at −50 °C and 0.01 bar for 24 h with a LIO-5P apparatus (CinquePascal srl, Trezzano sul Naviglio, Milan, Italy). Dried samples were subsequently removed from the flasks, homogeneously grounded into a powder by using ceramic mortars and pestles and then individually sealed into LDPE bags.

2.4.3. Microwave assisted digestion

Microwave digestion of samples was carried out by using the Speedwave™ MWS-3+ (Berghof, Eningen, Germany) microwave system with the maximum total output of the microwave generator (1450 W) and equipped with the optical sensor technology for contactless real-time recording of the sample temperature. The high-pressure resistant (up to 100 bar) PTFE vessels DAC-100S, together with the Multitube System (MT) (all from Berghof, Eningen, Germany), were used for sample digestion. This

arrangement allowed the simultaneous digestion of three samples in one DAC-100S PTFE vessel by placing three PFA tubes into each vessel (Husáková et al., 2015). The maximal number of used DAC-100S vessels for one digestion cycle was eight.

A powdered cuttlefish or CRM sample aliquot of 0.1 g was accurately weighed into a 10 mL PFA tube. Then, 4 mL of 16% HNO_3 (65%, w/w HNO_3, 1:3 diluted) and 1 mL of 30% H_2O_2 were added, leaving the vessels open until the initial reaction subsided. Three PFA tubes containing the same sample and reagents were placed into the outer 100mL PTFE digestion vessel previously filled with 25 mL of HNO_3 (16%, v/v) , by ensuring that the level of liquid in the outer PTFE vessel was higher than those in the PFA tubes. This way, the vapor pressures were compensated and the evaporation of the solution from the PFA tubes was avoided (Husáková et al., 2015). The samples were digested following a five-step program: (i) 20 min at 180 °C and 80 % power (ramp 5 min), (ii) 20 min at 220 °C and 95 % power (ramp 5 min), (iii–v) 5 min at 100 °C and 10 % power (ramp 1 min). The resulting colourless solutions were diluted to 25 mL with deionized water.

Each sample was mineralized in three replicates. Blanks, consisting of deionized water and reagents were subjected to a similar preparation procedure.

To assess the properness of the digestion process, the residual carbon content in the final cuttlefish digests was determined by a previously described ICP-OES method (Husáková et al., 2011) was 5.5 ± 0.4 % (n = 3).

2.5. ICP multi elemental analysis

ICP-MS measurements were performed by using the Agilent 7900 quadrupole mass spectrometer (Q-ICP-MS, Agilent Technologies, Inc., Santa Clara, CA, USA) equipped with an octopole-based collision cell for interference removal using kinetic energy discrimination (KED).

The standard sample introduction system, consisting of a glass concentric nebulizer MicroMist (400 μL min^{-1}), the Peltier-cooled (2 °C) Scott quartz spray chamber and quartz torch with 2.5 mm internal diameter injector, was used. For precise delivery of samples and ISTD, a low-pulsation, 10-roller peristaltic pump with three separate channels was employed. The internal

standard kit, including connecting tubing, connectors and the "Y" piece was used for simultaneous internal standard aspiration and its mixing with the sample. The standard sampling and skimmer nickel cones with orifices of 1 and 0.45 mm, respectively, were used. ICP-MS operating conditions were optimized during each start-up sequence by using the multi-elemental tuning solution (Agilent Technologies, Inc., Santa Clara, CA, USA) containing 1 μg L^{-1} of Ce, Co, Li, Mg, Tl and Y, in order to obtain the highest possible sensitivity for elements of low, middle and high m/z. Using the typical operating conditions summarized in Table 1, a sensitivity of 6000 counts s^{-1} per μg L^{-1} and a resolution of 0.64 amu peak width (full width at half maximum intensity) were achieved for $^{7}Li^{+}$. The same parameters were 50000 counts s^{-1} per μg L^{-1} and 0.62 for $^{89}Y^{+}$, and 30000 counts s^{-1} per μg L^{-1} and 0.60 for $^{205}Tl^{+}$. For sample analysis the "General Purpose" plasma mode included in the ICP-MS MassHunter software was used (Agilent Technologies, Inc., Santa Clara, CA, USA). The working parameters of the cell mode "no-gas" were autotuned during the instrument start-up sequence. The working parameters of the collision cell for helium ("He") and high energy He ("HE He") modes were adjusted manually. The time required for a transition between cell modes was 5 s. Parameters related to sample introduction and plasma conditions were consistent for all modes (see Table 1).

Concentrations of a total of 51 elements were evaluated from calibration curves with coefficients of determination better than 0.999 and built up within the ranges given below.

Calibration solutions: blank, 1, 5, 10, 50, 100 μg L^{-1} of Li, Be, B, Al, V, Cr, Fe, Mn, Ni, Cu, Zn, Co, Ga, Ge, As, Se, Rb, Sr, Zr, Mo, Ru, Cd, In, Sn, Sb, Te, Cs, Ba, Hf, Ta, Re, Pt, Tl, Pb, Bi, Th; 0.1, 0.5, 1, 5, 10 μg L^{-1} of La, Ce, Pr, Nd, U; 0.02, 0.1, 0.2, 1, 2 μg L^{-1} of Y, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu; 0.5, 1, 5, 10 mg L^{-1} of Na, Mg, P, K, Ca, Mn, Cu, Zn.

The solutions were prepared daily by appropriate dilution of multi-element solutions "A" (500 μg L^{-1}), "B" (50 + 10 μg L^{-1}) and "C" (50 mg L^{-1}) in 25 mL volumetric flasks (see *Section 2.1.*). To compensate possible instrumental drift and matrix effects, a 200 μg L^{-1} Rh ISTD was simultaneously aspired and mixed with samples.

Table 1. Agilent 7900 ICP-MS operating conditions

Parameter	**Setting**		
ICP			
Plasma mode	General purpose		
Rf power (27 MHz) (W)	1550		
Sampling depth (mm)	8		
Plasma gas flow (L min^{-1})	15		
Auxiliary gas flow (L min^{-1})	0.9		
Nebuliser gas flow (L min^{-1})	1		
Nebulizer pump (rps)	0.1		
Spray chamber temperature (°C)	2		
Mass spectrometer	No gas mode	He mode	HEHe mode [a]
Extract 1 (V)		0	
Extract 2 (V)		-250	
Omega bias (V)	-100	-120	-120
Omega lens (V)	9.7	7.8	9.6
Cell entrance	-30	-40	-140
Cell exit	-50	-60	-150
Deflect (V)	11.6	1	-77
Plate bias	-35	-60	-150
Helium flow (mL min^{-1})	0	5	10
OctP bias	-8	-18	-100
OctP RF		200	
Energy discrimination (V)		5	
Number of elements	39 [b]	12 [c]	5 [d]
Acquisition			
Points per peak	1		
Replicates	3		
Sweeps/replicate	100		
Total acquisition time (s)	75		

[a] HEHe mode - high energy helium mode; Monitored isotopes (integration time): [b] ^{7}Li, ^{11}B, ^{24}Mg, ^{66}Zn, ^{85}Rb, ^{88}Sr, ^{89}Y, ^{90}Zr, ^{95}Mo, ^{101}Ru, ^{103}Rh, ^{105}Pd, ^{111}Cd, ^{118}Sn, ^{121}Sb, ^{133}Cs, ^{138}Ba, ^{139}La, ^{140}Ce, ^{141}Pr, ^{146}Nd, ^{147}Sm, ^{153}Eu, ^{157}Gd, ^{159}Tb, ^{163}Dy,^{165}Ho, ^{166}Er, ^{172}Yb, ^{175}Lu, ^{178}Hf, ^{185}Re, ^{195}Pt, ^{205}Tl, $^{206+207+208}$Pb, ^{209}Bi, ^{232}Th, ^{238}U (all 0.1 s); [c] ^{23}Na (0.3 s), ^{27}Al (0.1 s), ^{39}K, ^{44}Ca (both 0.3 s), ^{51}V (1 s), ^{52}Cr, ^{55}Mn, ^{56}Fe, ^{59}Co, ^{60}Ni, ^{63}Cu, ^{103}Rh (all 0.3 s); [d] ^{31}P (0.1 s), ^{75}As, ^{78}Se (both 1 s), ^{103}Rh (0.3 s).

2.6. Mercury analysis

For the quantification of Hg content, the direct solid sampling analysis using AMA 254 (Altec Ltd., Prague, Czech Republic), a single-purpose atomic absorption spectrometer based on in situ dry-ashing followed by gold amalgamation atomic absorption spectroscopy, was used.

Samples were analysed under the following experimental conditions: typical sample mass, 50 mg; drying time, 60 s; decomposition, ~750 °C for 150 s; waiting step, 900 °C, 45 s, necessary for quantitative release of trapped mercury from the gold amalgamator to the measuring cuvette.

The peak area absorbance at 253.6 nm was monitored. The flow rate of oxygen (99.5%) carrier gas was 170 mL min^{-1}.

2.7. Data treatment

Statistics was applied to elemental concentrations referring to cuttlefish dry matter (d.m.) and carried out using IBM-SPSS (v. 23.0, SPSS Inc., Chicago, IL, USA) SIMCA-P v.14.1 (Umetrics, Umeå, Sweden), and Statgraphics Centurion v.18.1 (StatPoint Technologies, Inc, Warrenton, VA, USA) software.

The normality and homogeneity of variance of data was beforehand checked by applying the Shapiro-Wilk's and Levene's tests, respectively ($p \leq 0.05$). Since the normal distribution assumption was violated for most of the elemental data, quantitative differences between groups of cuttlefish from different origins were investigated by the nonparametric Kruskal–Wallis test with Dunn's multiple post hoc test for multiple comparison ($p \leq 0.05$). Data were presented as mean values ± margin of error (calculated at 95% confidence level), median and minimum and maximum values found.

While nonparametric statistics was used to compensate for the lack of data normal distribution, thus preserving the original information and avoiding the loose of any intuitive meanings when describing element concentrations resulting from univariate data analysis, a Box-Cox transformation was instead applied to normalize data for subsequent bivariate and multivariate data analyses.

First, the bivariate Pearson's correlation analysis (CA) was employed to Box-Cox transformed data to explore variable distribution and look for strong

positive ($r > 0.6, p \leq 0.01$) or strong negative ($r < -0.6, p \leq 0.01$) linear correlations between binary variables.

Because of highly different magnitude of values among element concentrations, the Box-Cox data were then scaled applying a Z-score standardization and further processed by multivariate data analyses. PCA was used for further investigation of covariance patterns in elemental concentrations, data first interpretation and multivariate outliers' removal based on Hotelling's T^2 test results (95% confidence interval). During this step, the number of variables was also reduced by selecting elements characterized by loading values (rescaled to 0–1) > 0.70.

Afterwards, supervised classifiers combined with different variable selection procedures were developed to achieve the discrimination of samples according to the geographical origin by using the lowest number and the most influential combination of variables. To that end, linear discrimination based on variable loading weights previously selected by PCA (LW-LDA), forward stepwise-LDA (S-LDA) and variable influence on projection-(VIP)-PLS-DA were independently tested and compared in terms of prediction ability. All the models were cross-validated using leave-one-out-cross-validation (LOOCV) to exclude overfitting. In addition to LOOCV, the validity of the PLS discriminant models was also checked by permutation testing (400 random permutations).

When supervised modelling was performed, about 33% of data (n = 68) was randomly side-lined from the calibration set (n = 136) and used for independent external validation. The overall performance of the models in internal and external validation was evaluated in terms of accuracy (%), sensitivity (%), and specificity (%).

Further details about the statistical treatments applied can be found in Supplementary material, *Supplementary Experimental Section*.

3. Results

3.1. Analytical performance and validation

The analytical determination of more than fifty major, trace and ultra-trace elements in mineralized cephalopod samples was carried out by Q-ICP-MS, taking advantage of its well-known sensitivity toward the targeted elements.

On the other hand, analysis of Hg by thermal decomposition was performed by AMA254 without pre-treatment and/or preconcentration steps, thus overcoming the most serious problems related to ICP-MS analysis of Hg, e.g. long washout time, non-linear calibration curves, and decreasing sensitivity with time (Li et al., 2006).

The microwave-assisted pressurized digestion which involved small sample amounts (100 mg) was employed to ensure rapid sample preparation for subsequent ICP-MS analysis while attaining the high decomposition efficiency. One of the most important aspects regarding the employed MWD method is the possibility to use diluted solutions for sample preparation, thus reducing the quantity of reagents, risks of contamination and generation of residues. These are important parameters for the development of greener analytical methods (Gonzalez et al., 2009). Moreover, MWD plays an important role in decreasing some kinds of spectral and non-spectral interferences, such as those deriving from carbon containing species or different chlorides, by conversion to significantly less or non-interfering species (Husáková et al., 2011; Bizzi et al., 2017). He mode with KED or HE He mode with higher cell gas pressure and higher energy discrimination voltages, were used for the analytes that suffered from spectral effects (i.e. V, Cr, Mn, Co, Ni, Fe, Co, Cu, As, and Se) caused by polyatomic ions deriving from the plasma gas, sample solvent and other sample matrix elements such as Na, K, Ca, Mg, P, S, C, and Cl (see Table 1 and Supplementary Materials, Table S2). He and/or HE He cell modes had the additional benefit of reducing the response for low mass matrix elements like Na, K, Ca or P by an order of magnitude ($^{23}Na^+$, $^{44}Ca^+$) or more ($^{39}K^+$, $^{31}P^+$), thus effectively raising the upper linear range for these elements. Thereby elements that would have needed to be analysed by ICP-OES, were included in the ICP-MS.

As for the ISTD, Rh was chosen for its mid-range mass and ionization potential and because of its absence in the analysed samples. Recoveries of Rh, relative to the initial calibration blank for the entire 8-hour sequence of sample digests, are shown in Supplementary Materials, Figure S1. ISTD behaviour across the mass range was found to be very consistent overt time and downward drift was considered acceptable.

Data accuracy was evaluated by the analysis of different CRMs (NIST SRM 1577 Bovine Liver, NIST 1566 Oyster Tissue, BCR CRM 184 Bovine muscle, BCR CRM 185 Bovine Liver, CRM 12-2-01 Bovine Liver), primarily intended for the evaluation of analytical methods and instruments used for the determination of the mass fraction values of selected elements in marine tissue, foods, or similar materials. Since the levels of most lanthanides and actinides are not certified in these CRMs, three additional certified standards (NCS ZC 73015 Milk Powder, CRM 12-2-04 Wheat bread flour, and CRM 12-2-03 Essential and toxic elements in Lucerne) were analysed to validate data for these elements. The high level of agreement between target and found values demonstrated trueness of data obtained (Supplementary Materials, Table S2).

The precision of the method was evaluated in terms of intra-day and inter-day comparison. Intra-day precision was determined by analysis of individual control materials three times during the same day. Inter-day precision was determined by analysis of the same standards on three different days over a period of one month. Within each series, every solution was analysed in triplicate. Relative standard deviation (RSD) was calculated for both series of analyses. The RSD values of intra-day and inter-day studies were mostly found to be below 14% thus showing a good precision of the method (Supplementary Materials, Table S2).

Method detection limits (MDLs) and method quantification limits (MLOQs) reported in Supplementary Material, Table S3 were evaluated as a triple and tenfold standard deviation, respectively, of ten consecutive measurements of blank signal divided by the slope of the calibration curve. Detection limits were found to be low enough that selected elements could be determined at the background level. Table S3 (Supplementary Materials) also summarizes relative sensitivities of Q-ICP-MS for analysis of individual elements with the use of Rh ISTD.

3.2. Concentrations of elements in cuttlefish samples

The concentrations of elements measured in the present study were presented and discussed as median values considering the non-normal distribution of the raw data. Results from descriptive statistics are listed in

Table 2. The most abundant elements were found to be Na, Mg, P and K, whose concentration levels were higher than 1% of weight and quite variable among cuttlefish from different countries. According to results from Kruskal–Wallis test, Na and Mg amounts were significantly higher ($p \leq 0.05$) in samples from FAO 27.7.e (French Atlantic) compared to samples from FAO 37.1/37.2 (Mediterranean area) and FAO 37.2.1 (Chioggia). At the same time, an opposite trend was observed for P and K, for which the highest concentrations ($p \leq 0.05$) were found in Chioggia cuttlefish (Table 2).

Concentrations between 1 and 0.01% of weight (10000–100 mg kg^{-1}, d.m.) were found for two elements corresponding to Li and Ca, which showed significantly higher median concentrations in French Atlantic samples (0.51 mg $kg^{-1}$and 1760 mg kg^{-1}, respectively) ($p \leq 0.05$) compared to the other groups. The wide variability in both major and minor element contents found in cuttlefish specimens of different geographical origins and within samples of the same origin can be attributed to their natural variation in seawater, but the assimilation by marine animals of Na, Mg, and Ca from the organic matter in the aquatic environment also varies with the feeding status and the age of the animal (Lall, 2002). Moreover, the presence of P in fish and seafood tissues can be directly attributable to the food sources since its concentration in seawater is lower compared to the other elements (Lall, 2002).

The main trace elements, ranging between 100 and 1 mg kg^{-1}, followed the decreasing order Zn > Al > Sr > Cu > Fe > B > Rb > Se > Mn in samples from Chioggia, Zn > Cu > Sr > Fe > Al > B > Rb > Mn > Se in samples from the Mediterranean area, and Zn > Sr > Cu > B > Fe > Al > Rb > Se > Mn in samples from the French Atlantic. Among these elements, only Se, Rb, and Sr were different in median concentrations according to all three geographical regions ($p \leq 0.05$) and the concentration ranges were in line with those previously published in literature for *S. Officinalis* (Raimundo, Pereira, Vale, & Caetano, 2005; Ayas & Ozogul, 2011).

Different median contents varying between 62.3 and 75.8 mg kg^{-1} were also found for As ($p \leq 0.05$). The wide presence of arsenic in the aquatic environment is linked both to the geochemical characteristics of the region and anthropogenic activities, especially those related to agricultural

practices (Neff, 2002). Due to toxic properties, some concerns regarding the dietary exposure to arsenic have arisen in recent times, but maximum levels for this metal in fishery products have not been established yet by European legislations. Fish and seafood, in particular, have been reported as the largest food contributors to overall total arsenic exposure, despite the largest proportion is represented by organic arsenic species which are known to be less or no toxic compared to relative inorganic species (European Food Safety Authority, EFSA, 2009). In addition, inorganic arsenic concentrations decrease with increasing content of total arsenic (European Food Safety Authority, EFSA, 2009).

Most of the elements analysed in the present paper were found to be present in cuttlefish tissues at mean concentration lower than 1 mg kg^{-1}. Levels of ultra-trace elements as rare earth elements (REEs) were lower than 100 μg kg^{-1} on average, with the highest median concentration for Ce (24.0 μg kg^{-1}) and the lowest one for Lu (0.047 μg kg^{-1}) both in in Mediterranean samples (Table 2). The natural pattern of REEs in the aquatic environment, which is associated to the mobilization of nutrients from the underlying soils, can be permanently altered by some anthropic activities and effectively used to investigate the geographical provenance of marine animals (Noack, Dzombak, & Karamalidis, 2014). In the present work, the significantly different concentrations of Tb, Dy, Ho, Er, and Yb geographically discriminated Chioggia from French specimens, while those of La, Pr, Nd, Eu, and Gd discriminated Mediterranean form French specimens ($p \leq 0.05$). In accordance with these results, the amount of both La and Ho was previously reported to be potentially useful to authenticate fish samples from different areas in the Mediterranean Sea (Varrà, Ghidini, Zanardi, Badiani, & Ianieri, 2019).

Regardless the origin, all the cuttlefish samples analysed in the present work were in line with the maximum limits for toxic heavy metals established by European regulation No. 1881/2006/EC and subsequent amendments and additions (Commission of the European Communities, 2006). Both cadmium and lead did not exceed the threshold value of 1 mg kg^{-1} set for the edible part of cephalopods, as well as Hg was always found to be below the threshold limit of 0.5 mg kg^{-1}. The sample groups under investigation did

not show significant differences in Pb and Hg concentrations ($p > 0.05$), but Cd median amount in French Atlantic specimens was found to be approximately 7 and 20 times higher than in Mediterranean and Chioggia samples, respectively (see Table 2). Furthermore, the comparison with results reported by other authors for cuttlefish from North eastern Atlantic and Adriatic Sea, showed that Hg, Pb, and Cd concentrations found in the present work fell within the same ranges (Bustamante, Lahaye, Durnez, Churlaud, & Caurant, 2006; Storelli, Giacominelli-Stuffler, Storelli, & Marcotrigiano, 2007). Even in the case of heavy toxic metals, a close link between their concentrations in seawater and environmental pollution of specific areas exists (Carvalho, Santiago, & Nunes, 2005). This, together with Ni, Co, and Bi amounts, which also varied significantly among sample groups, makes suggestions for future insights about the effective role and contribution of these elements for the geographical origin assessment of fish and seafood.

3.3. Chemometric classification of the geographical origin of cuttlefish

Since no specific information concerning the best arrangement of elements to discriminate the geographical origin of cephalopods is available in literature, the first stages of multivariate analysis took into consideration all the fifty-two elements measured, starting with the data elaboration by means of CA and PCA. Nevertheless, multicollinearity between variables as well as geographical-unrelated variability are a source of increasing noise, thus the reduction of the variables to the most predictive ones was subsequently preferred, especially in view of a future practical implementation of the methodology.

The selection of variables is strictly dependent on the sample matrix properties and dataset overall characteristics; hence its application needs to be evaluated on an individual basis. Although being the most frequently used classification technique in studies dealing with element profiling, S-LDA is coming under growing criticism because of the deceiving results deriving from the random rather than effective significance of the variables extracted, especially when the number of potential explanatory variables is high (G. Smith, 2018).

Table 2. Multi-elemental composition of cuttlefish from different geographical origins.

	Chioggia (n = 17*3)				Mediterranean Sea (n = 25*3)				French Atlantic (n = 26*3)			
Element	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**
Li	0.25 ± 0.024	0.24[a]	0.18	0.34	0.34 ± 0.011	0.33[b]	0.28	0.38	0.50 ± 0.025	0.51[c]	0.36	0.65
B	6.3 ± 0.56	6.06[ab]	4.46	8.20	5.3 ± 0.29	5.35[a]	4.11	7.15	7.6 ± 0.85	7.88[b]	4.15	11.8
Na	11164 ± 867	11235[a]	8659	14187	14334 ± 486	14290[b]	12015	16338	24632 ± 1332	24730[c]	17932	30749
Mg	2297 ± 207	2274[a]	1626	3180	2613 ± 75	2643[a]	2164	2871	3767 ± 183	3776[b]	2898	4795
Al	16 ± 5.5	10.7[a]	4.7	43.8	7 ± 1.9	5.69[b]	2.97	22.7	7 ± 2.5	5.44[b]	2.87	34.4
P	10697 ± 831	10635[a]	7459	1617	9084 ± 221	9079[b]	7440	9883	8203 ± 308	8176[c]	6791	10401
K	13471 ± 1067	13768[a]	9113	16488	10648 ± 402	10547[b]	8898	12400	8353 ± 531	8252[c]	6196	12039
Ca	820 ± 102	751[a]	619	1338	1299 ± 53	1304[b]	1019	1513	1781 ± 118	1760[c]	1239	2493
V	0.07 ± 0.036	0.040[a]	0.018	0.26	0.25 ± 0.040	0.22[b]	0.12	0.47	0.11 ± 0.023	0.11[a]	0.039	0.27
Cr	0.18 ± 0.060	0.18[a]	0.037	0.55	0.15 ± 0.041	0.13[a]	0.068	0.54	0.22 ± 0.082	0.13[a]	0.081	1.07
Fe	11 ± 4	7.85[a]	3.62	33.0	9 ± 1.9	8.19[a]	4.32	25.0	8 ± 2	5.59[a]	3.41	21.7
Mn	2.2 ± 0.87	1.28[a]	0.93	6.86	2.5 ± 0.52	2.04[a]	1.41	6.59	1.0 ± 0.23	0.83[b]	0.54	3.57
Ni	0.4 ± 0.52	0.36[a]	0.22	0.76	0.52 ± 0.056	0.50[b]	0.27	1.98	0.28 ± 0.034	0.26[a]	0.15	0.48
Co	0.04 ± 0.019	0.025[a]	0.016	0.15	0.21 ± 0.023	0.21[b]	0.11	0.31	0.14 ± 0.026	0.14[c]	0.058	0.31
Cu	10 ± 1.8	8.55[a]	5.93	20.2	24 ± 3.4	21.3[b]	12.5	42.9	20 ± 3.4	16.3[b]	8.29	36.2
Zn	62 ± 5.2	60.7[a]	41.1	86.3	73 ± 1.6	72.8[b]	63.8	80.7	74 ± 3.5	73.8[b]	58.6	94.1
As	79 ± 9	75.8[a]	56.1	121.3	69 ± 4.3	67.4[ab]	50.5	98.8	60 ± 4.7	62.3[b]	35.5	77.1
Se	1.3 ± 0.17	1.33[a]	0.881	2.11	2.0 ± 0.14	1.91[b]	1.47	2.69	1.65 ± 0.083	1.62[c]	1.26	2.16
Rb	5.3 ± 0.42	5.17[a]	3.94	6.99	4.5 ± 0.14	4.61[a]	3.93	5.13	3.6 ± 0.22	3.57[b]	2.56	5.27
Sr	9 ± 1.1	8.83[a]	6.69	14.9	14.8 ± 0.44	14.9[b]	12.0	16.8	25 ± 1.6	26.0[c]	16.3	33.2
Y	0.02 ± 0.012	0.015[a]	0.0057	0.11	0.008 ± 0.0012	0.0074[b]	0.0039	0.017	0.01 ± 0.013	0.0077[b]	0.0050	0.17
Zr	0.04 ± 0.018	0.027[a]	0.0059	0.13	0.05 ± 0.015	0.036[ab]	0.0071	0.14	0.2 ± 0.20	0.051[b]	0.0066	2.55
Mo	0.029 ± 0.0066	0.024[a]	0.018	0.060	0.10 ± 0.013	0.089[b]	0.061	0.18	0.10 ± 0.019	0.084[b]	0.044	0.23
Ru*	0.10 ± 0.029	0.087[a]	0.031	0.24	0.14 ± 0.027	0.13[a]	0.054	0.34	0.14 ± 0.023	0.13[a]	0.044	0.29
Pd	0.016 ± 0.0013	0.016[a]	0.012	0.020	0.024 ±0.0054	0.021[a]	0.012	0.067	0.029 ± 0.0033	0.027[b]	0.017	0.059
Cd	0.02 ± 0.017	0.0099[a]	0.0035	0.13	0.08 ± 0.013	0.071[b]	0.035	0.15	0.25 ± 0.059	0.22[c]	0.067	0.58
Sn	0.015 ± 0.0071	0.011[a]	0.0044	0.061	0.022 ± 0.0049	0.019[b]	0.0075	0.051	0.015 ± 0.0040	0.011[a]	0.0052	0.053
Sb	0.007 ± 0.0028	0.0053[a]	0.0023	0.021	0.009 ± 0.0012	0.0084[b]	0.0046	0.016	0.0075 ± 0.00080	0.0068[b]	0.0052	0.014

(continued)

Table 2. *(continued)*

	Chioggia (n = 17*3)				Mediterranean Sea (n = 25*3)				French Atlantic (n = 26*3)			
Element	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**
Cs	0.017 ± 0.0015	0.016^{a}	0.012	0.023	0.0146 ± 0.00085	0.014^{a}	0.012	0.023	0.012 ± 0.0014	0.012^{b}	0.0081	0.027
Ba	0.3 ± 0.13	0.18^{a}	0.087	1.17	0.22 ± 0.031	0.16^{a}	0.12	0.50	0.21 ± 0.039	0.18^{a}	0.11	0.43
La*	20 ± 11	11.9^{ab}	2.70	90.87	29 ± 12	20.3^{a}	8.03	150	16 ± 5.6	10.5^{b}	5.09	56.0
Ce*	29 ± 10	20.6^{a}	3.56	82.0	44 ± 20	24.0^{a}	11.4	250	26 ± 8.9	18.8^{a}	8.31	91.7
Pr*	2 ± 1.0	0.96^{a}	0.38	7.43	1.2 ± 0.34	0.97^{a}	0.45	4.62	0.8 ± 0.23	0.63^{b}	0.33	2.56
Nd*	8 ± 4.7	3.98^{a}	1.58	33.2	5 ± 1.5	4.45^{a}	1.74	20.1	4 ± 1.0	2.82^{b}	1.33	12.0
Sm*	1.5 ± 0.81	0.74^{a}	0.25	5.04	0.9 ± 0.28	0.82^{a}	0.34	3.73	0.7 ± 0.18	0.55^{a}	0.33	2.32
Eu*	0.4 ± 0.18	0.24^{ab}	0.12	1.19	0.28 ± 0.061	0.25^{a}	0.15	0.87	0.22 ± 0.044	0.19^{b}	0.12	0.62
Gd*	1 ± 0.5	0.69^{ab}	0.24	3.20	0.8 ± 0.19	0.72^{a}	0.32	2.52	0.6 ± 0.15	0.50^{b}	0.31	1.97
Tb*	0.20 ± 0.082	0.14^{a}	0.056	0.58	0.15 ± 0.037	0.12^{ab}	0.055	0.43	0.12 ± 0.030	0.095^{b}	0.054	0.37
Dy*	1.1 ± 0.44	0.74^{a}	0.34	3.22	0.8 ± 0.18	0.66^{ab}	0.27	2.21	0.7 ± 0.17	0.51^{b}	0.33	2.02
Ho*	0.22 ± 0.078	0.17^{a}	0.062	0.60	0.15 ± 0.030	0.14^{ab}	0.047	0.39	0.14 ± 0.037	0.11^{b}	0.063	0.45
Er*	0.7 ± 0.23	0.54^{a}	0.20	1.73	0.5 ± 0.10	0.42^{ab}	0.23	1.17	0.4 ± 0.11	0.39^{b}	0.19	1.31
Yb*	0.7 ± 0.21	0.51^{a}	0.22	1.48	0.39 ± 0.078	0.37^{ab}	0.14	0.93	0.4 ± 0.10	0.34^{b}	0.16	1.31
Lu*	0.09 ± 0.032	0.074^{a}	0.023	0.23	0.06 ± 0.015	0.047^{a}	0.021	0.18	0.06 ± 0.017	0.053^{a}	0.021	0.22
Hf*	2.5 ± 0.52	2.23^{a}	1.15	4.94	8 ± 8.4	2.89^{ab}	1.42	103	8 ± 4.9	4.19^{b}	0.92	57.3
Re*	0.10 ± 0.032	0.081^{ab}	0.016	0.24	0.12 ± 0.019	0.11^{a}	0.070	0.24	0.09 ± 0.023	0.072^{b}	0.025	0.28
Pt*	0.13 ± 0.034	0.10^{a}	0.060	0.25	0.2 ± 0.12	0.12^{a}	0.062	1.53	0.15 ± 0.061	0.10^{a}	0.019	0.60
Tl*	1.5 ± 0.11	1.42^{a}	1.07	1.94	1.4 ± 0.51	1.21^{b}	0.83	7.30	1.3 ± 0.61	1.05^{b}	0.62	8.66
Pb	0.10 ± 0.024	0.093^{a}	0.050	0.23	0.10 ± 0.019	0.099^{a}	0.063	0.31	0.09 ± 0.017	0.046^{a}	0.077	0.20
Bi	0.178 ± 0.0022	0.017^{a}	0.011	0.027	0.023 ± 0.0040	0.021^{a}	0.015	0.065	0.010 ± 0.0041	0.0066^{b}	0.0029	0.044
Th	0.002 ± 0.0011	0.0016^{a}	0.00081	0.0082	0.005 ± 0.0036	0.0022^{a}	0.0011	0.046	0.003 ± 0.0019	0.0015^{a}	0.0011	0.024
U	0.0032 ± 0.00076	0.0026^{a}	0.0019	0.0074	0.0043 ± 0.00042	0.0040^{a}	0.0027	0.0069	0.010 ± 0.0015	0.0091^{b}	0.0049	0.019
Hg$^{\#}$	0.20 ± 0.030	0.18^{a}	0.15	0.34	0.20 ± 0.023	0.17^{a}	0.14	0.38	0.21 ± 0.018	0.20^{a}	0.13	0.29

Concentrations are expressed as mg kg^{-1} (d.m.) and are reported as mean values ± margin of error (ME) at 95% confidence level. Min–Max: minimum and maximum values found. *: concentrations are expressed as μg kg $^{-1}$. Hg$^{\#}$: determined by direct mercury analyzer AMA254. Values followed by different superscript letters (a–c) in the same row are significantly different ($p \leq 0.05$).

On the other hand, one of the main problems related to the use of multivariate data analysis in food authentication studies is just the difficulty in comparing the obtained results with those already published but using a wide variety of different chemometrics techniques. Therefore, in the context of the present work, S-LDA and LW-LDA, were used for classification by variable selection, but outputs were compared with those obtained through the application of VIP-based PLS-DA. The latter, in fact, consists of a more robust and flexible algorithm, particularly suited for the classification of a large number of samples and is suggested to be a more powerful tool for reliable variable selection compared to the traditional LDA (Rashid, Hussain, Ahmad, & Abdullah, 2019)

3.3.1 Correlation and principal component analysis

Pearson's associations between all the elements were reported by plotting the correlation matrix shown in Figure 1. The main significant patterns of covariations were found among most of the REEs and, in particular, between Ce and La ($r > 0.9$), and among Nd, Pr, Tb, Ho, Gd, Er, Eu, Dy, and Sm ($r > 0.8$). Similarly, bivariate positive correlations regarding the patterns Na–Sr–Cd–Li–Mg–Ca ($r > 0.9$), P–K ($r > 0.9$), and Cs–Rb ($r > 0.8$) were also detected. Concentrations of U were positively correlated with Ca, Mg, Pd, Li, Cd, Sr, and Na ($r > 0.8$) as well as amounts of Co with Cu, Mo, and V ($r > 0.7$). The most important and significant negative correlations were instead found among K–Rb–Cs–Mn and among Na–Sr–Cd–Li–Ca–Mg–U ($r < -0.7$) patterns.

The association between the characteristic element distributions and cuttlefish origins was further investigated by PCA. The first PCA computation took into consideration all the fifty-two elements and, as a result, the number of original variables was set by LOOCV to six final principal components (PCs) enclosing 72.6% and 57.3% of the explained and predictive data variance, respectively. The first two PCs of the model contributed for 27.5% and 20.8% of the total explained variance, with the PC2 enclosing the fraction of information related to sample provenance, as reported in Figure 2A.

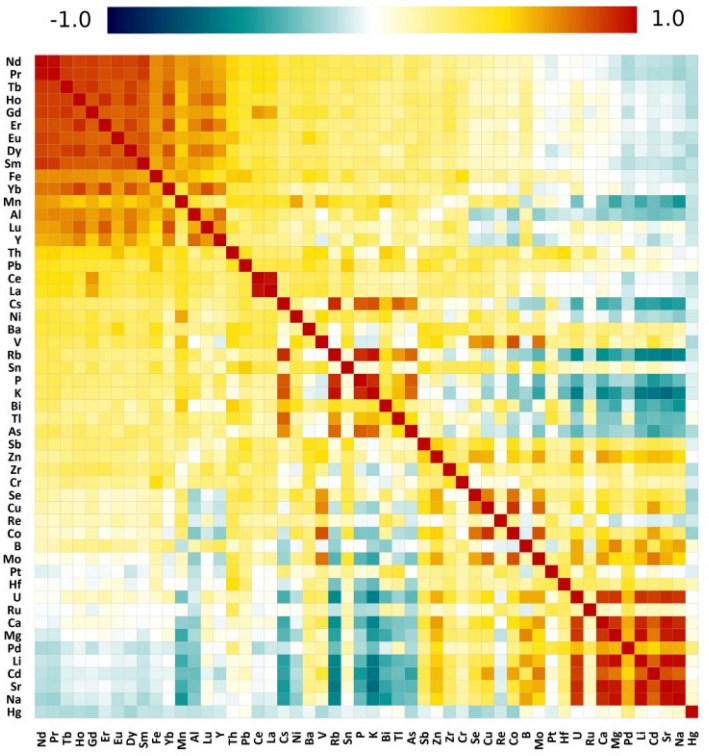


Figure 1. Pearson's correlation heat-map of the investigated elements.

The distribution of Chioggia cuttlefish along the negative axis of the PC2 was mainly dominated by the highest contribution of Rb and K, followed by Mn, Al, Cs, P, As, and, at lower levels, Bi (see loading vectors reported in Figure 2A). On the other hand, French Atlantic cuttlefish distribution was mainly guided by contribution from Na, Sr, Cd, Li, Mg, Ca, U, and Mo, whose vectors angles nearness underlined the strong covariation previously stated by CA. Likewise, a strong contribution and a high degree of covariation was confirmed in PCA for REEs, which impacted on the negative axis of PC2.

During this preliminary stage, one sample from Chioggia was also found to be a strong outlier since it fell outside the 95% confidence interval defined

by the Hotelling's T^2 range (see Figure 2A). For this reason, it was excluded from subsequent statistics.

Since the most influential loading values were found for a total of 16 out 52 elements (Li, B, Na, Mg, Ca, Co, Cu, Zn, Sr, Mo, Ru, Pd, Cd, Hf, Pt, and U), variable filtering was performed (see *Section 2.7.*) and PCA was thus recomputed. Loading weights of selected elements are reported in Supplementary Materials, Table S4. The reduced PCA model extracted three PCs that explained 84.5% of the data variability. The PC1 (explained variability, 54.1%; predictive variability, 48.3%) was mainly characterized by Zn, Cd, Pd, Ca, U, Sr, while in the PC2 (explained variability, 16.3%; predictive variability, 17.1%) Pt was the most influential element (Figure 2B). Hence, maximum variation in original dataset was retained while reducing the number of significant features. At the same time, sample separation in the score space was significantly expressed along the PC1 using the selected variables rather than the whole set of elements.

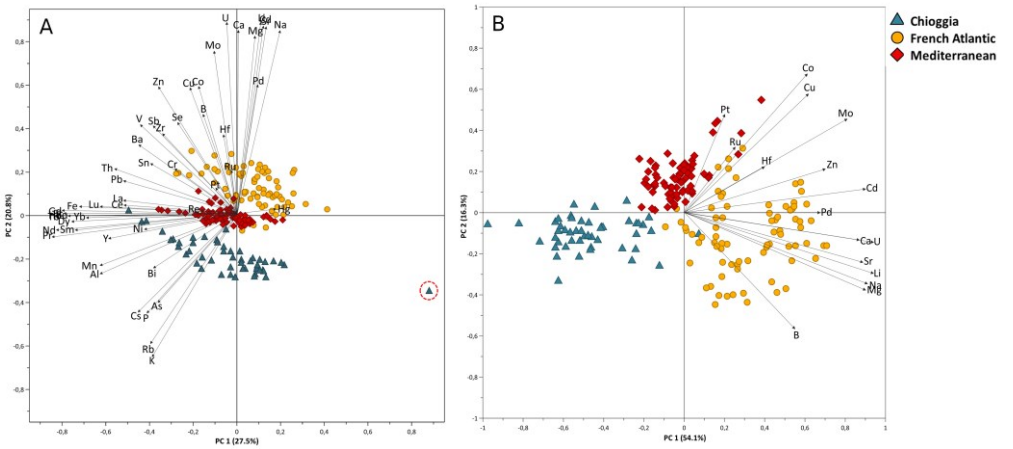


Figure 2. Biplot resulting from PCA applied to all the elements (A) and selected elements (B). The sample surrounded by the red dotted line in the biplot indicate an outlier according to Hotelling's T^2 test (95% confidence interval). Blue triangles: samples from Chioggia (FAO fishing area 37.2.1); red diamonds: samples from Mediterranean Sea (FAO fishing area 37.1/37.2); yellow circles: samples from French Atlantic (FAO fishing area 27.7.e).

3.3.2 Classification of cuttlefish by origin: LDA and PLS-DA approaches using selected variables

The LW-LDA classification model for cuttlefish origins, created by using variables previously selected by PCA (Li, B, Na, Mg, Ca, Co, Cu, Zn, Sr, Mo, Ru, Pd, Cd, Hf, Pt, and U) was described by two discriminant functions (DFs) explaining 99.7% of the data variability. The DF1 (explained variability, 83.1%; canonical correlation, 97.2%) and the DF 2 (explained variability, 16.6%; canonical correlation, 86.7%), whose statistical significance was confirmed by low Wilk's Lambda values (0.004 and 0.13, respectively, $p = 0.000$) were responsible for the clusterization of cuttlefish in the bidimensional score plot reported in Figure 3A.

Samples from French Atlantic were well discriminated from samples from Chioggia and Mediterranean Sea by the DF1 and Chioggia samples well discriminated from Mediterranean samples by the DF2. The most important loadings for the DF1 were Sr and Na, while Co was found to be the most influent element for the DF2 (Figure 3A). Based on these results, 100% of cuttlefish were correctly recognized in CV, but, when performing the external test set validation, two samples from Mediterranean area were misclassified as samples from French Atlantic and one French Atlantic sample as of Mediterranean origin, thus resulting in an overall accuracy of the model of 97% (Table 3). Fisher's discriminant function coefficients for each geographical provenance are reported in Supplementary Material, Table S5.

In S-LDA the forward selection method based on Wilk's Lambda values was chosen to select the discriminating variables, taking 3.84 as the minimum partial F value to include a variable and 2.71 as the maximum F value to exclude a variable from the model (see Supplementary materials, *Supplementary Experimental Section*). Based on this, two DFs and 12 out 52 discriminating elements were finally included into the model, corresponding to Na, Co, B, K, Cd, V, U, Rb, Ni, Ba, Cu, and As (Supplementary Materials, Table S6), of which Na, Co, B, Cd, U, and Cu were the same as those selected in LW-LDA.

The DF1 and DF2 enclosed 82.5% and 17.5% of variance of original data, with a canonical correlation of 97.4% and 88.5% respectively. The significant

discriminatory ability of each DF was confirmed not only by Wilk's Lambda values of 0.03 for the DF1 and 0.114 for the DF2 (p = 0.000), but also by 99% accuracy in LOOCV (one Mediterranean samples erroneously classified as Atlantic sample) and 100% accuracy in external validation (Table 3). Fisher's coefficients of for each provenance are reported in Supplementary Materials, Table S6.

Optimal classification results (100% sensitivity, specificity, and accuracy, Table 3) were obtained by the application of PLS-DA when 21 elements were selected on the basis of their VIP scores (VIP >1) and used as classificatory variables (see Supplementary Materials, Table S7). Among these elements, ten were found to be the same of those previously included in stepwise-LDA. Although Sr, Mn, Mo, Li, Ca, Mg, Se, Bi, Cs, P, and Y were selected in addition, some of these (Sr, Mo, Li, Ca, and Mg) were shared with LW-LDA. The highest VIP scores were 1.672 for V and 1.628 for Co, thus indicating that the maximum contribution of these elements to the geographical separation between cuttlefish.

The VIP-PLS-DA model was created by using six LVs explaining 90.1% of variance (R^2X = 0.901), of which 92.2% was directly correlated with labels of groups (R^2Y = 0.901). In addition, the overall training model's predictive power of 88.8% (Q^2 = 0.888), resulted in 100% of samples to be correctly classified both in LOOCV and external validation (Table. 3), without any possible misleading interpretation deriving from overfit or overprediction of the model as assessed by permutation test (average R^2Y-y-intercept = 0.048; average Q^2-y-intercept = -0.371). Further information about this validation are reported in Supplementary Materials, Supplementary Figure S2.

As for the RMSECV and RMSEP values, these were found to be low (0.156 and 0.159, respectively) and very close to each other, thus remarking the outstanding ability of the simplified model in categorizing cuttlefish.

Figure 3B shows the bidimensional score and loading graphs of the first two DFs obtained by S-LDA, while the first two LVs obtained by VIP-PLS-DA are reported in Figure 3C. In both cases, cuttlefish were well separated each other, with samples form Chioggia distributing in the lower right quadrant of the score plots. In addition, both the DF1 and the LV1 were mainly responsible for sample discrimination in the score plot.

Table 3. Supervised classification model performances in cross-validation (CV) and in external test set validation.

Model	Validation	Overall classification rate (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)
LW-LDA	CV	100	100	100	100
	External	96	96	98	97
S-LDA	CV	99	99	99	99
	External	100	100	100	100
VIP-PLS-DA	CV	100	100	100	100
	External	100	100	100	100

By looking at the loading plots, Chioggia cuttlefish were found to be positively correlated with U, V, Ni, Ba, As, Rb, and K in S-LDA (Figure 3B) and with Cs, Rb, K and P in PLS-DA (Figure 3C). The highest loading scores for Mediterranean cuttlefish were instead observed for Cd in S-LDA, which in turn was diagonally opposite to the elements that characterized Chioggia samples. In PLS-DA model, the highest degree of correlation with Mediterranean cuttlefish was instead found for V, Se, Co, Cu, Mo, Ni, Bi, and Mn while for French Atlantic cuttlefish the highest contribution was exerted by Na together with Mg, U, Li, Sr, Ca and Cd. Likewise, Na showed a high loading value on the DF2 negative axis and, therefore, it allowed to distinguish French samples from the other groups in the S-LDA model.

According to the results of multi-element analysis of fish and seafood products already published in literature, some of the most discriminant elements found in the present study might be linked to anthropogenic pollution (As, V, Cd and U) or to geogenic sources (Cs and Mn) (Costas-Rodríguez et al., 2010). Consistently to the results achieved, the major elements Na, Mg, and K were also previously identified among the most useful indicators of geographical origin of crabs (Luo et al., 2019) and prawns (Gopi et al., 2019). The variation in the amounts of Na, Mg and P in the tissues of Pacific shrimps was also correlated with specific sampling

sites, whose waters were characterized by higher salinity (Li, Han, Dong, & Boyd, 2019).

The concentrations of As which, in the present work, was one of the selected discriminating elements for Chioggia cuttlefish, were also reported in mussels from Venice Lagoon (in which Chioggia is located) as being positively associated to the higher degree of salinity of this area and thus valuable for the identification of local products (Cubadda, Raggi, & Coni, 2006). Moreover, also Ni, Co, Se, Mo, and V amounts showed a strong link with anthropic sources of element contamination of the Venice Lagoon (Cubadda et al., 2006), therefore the specific pattern distribution of these trace elements in Chioggia cuttlefish tissues might be supposed to specifically reflect their origin.

Based on these results, the use of multivariate classifiers in combination with the pre-selection of the most origin-discriminant elements, was proved to be an efficient, rapid and smart analytical strategy to assure the authenticity of the provenance of cuttlefish samples. Although each of the classification techniques applied showed satisfying results, suggesting that possible users may choose the most convenient methodology to suit the specific needs, superior and unequivocal outcomes were obtained by applying VIP-PLS-DA.

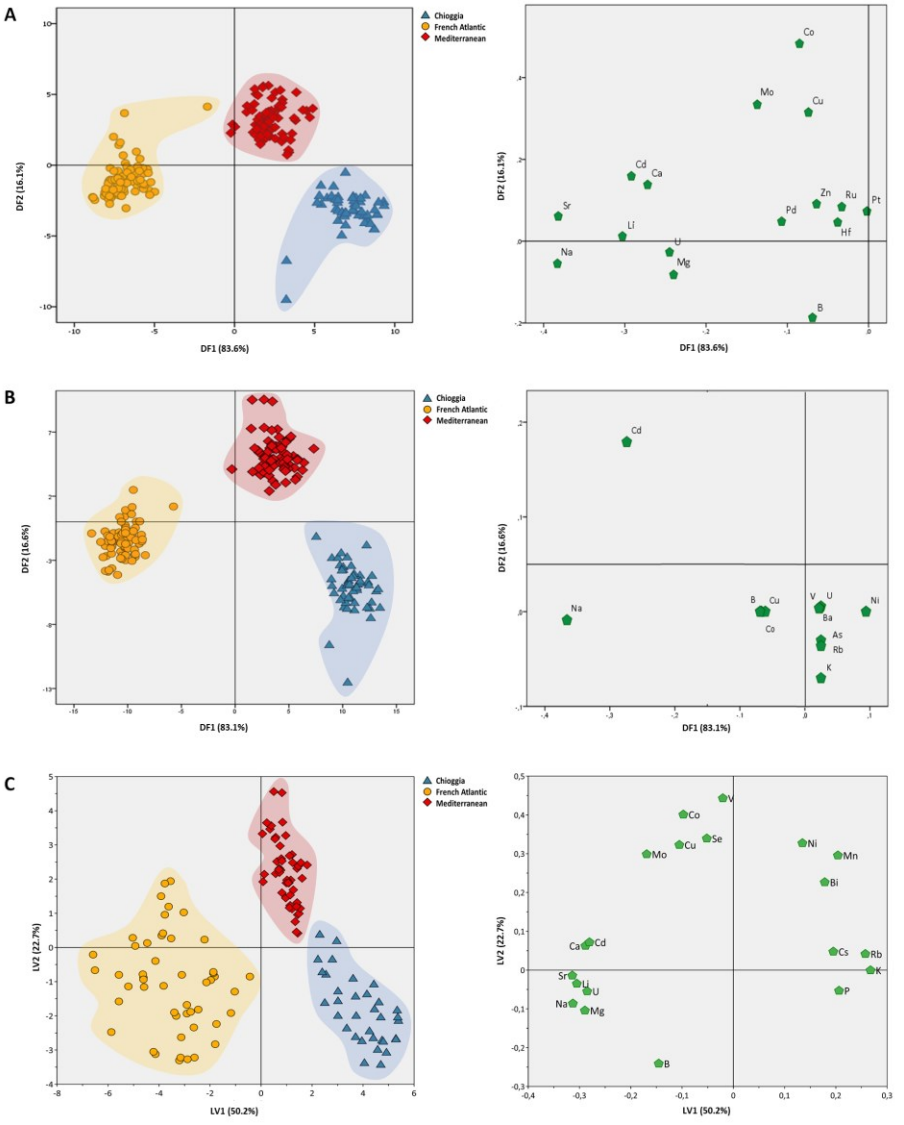


Figure 3. Comparison of score (left) and loading graphs (right) from LW-LDA (A), S-LDA (B) and VIP-PLS-DA (C) for cuttlefish samples from different origins. Blue triangles: samples from Chioggia (FAO fishing area 37.2.1); red diamonds: samples from Mediterranean Sea (FAO fishing area 37.1/37.2); yellow circles: samples from French Atlantic (FAO fishing area 27.7.e).

4. Conclusions

Cephalopods are among the fishery products more consumed and appreciated by the consumers all over the world but, to the best of our knowledge, scientific insights concerning their elemental profile and the authentication of their geographical origin are still lacking.

The results presented in this study revealed for the first time that the geographical imprint of Italian traditional cuttlefish from Chioggia can be extracted through the simultaneous quantification of more than fifty elements by ICP-MS and successfully used to discriminate the product from cuttlefish originating from different areas. The whole elemental profile was elaborated by means of different chemometric techniques. In particular, three independent variable selection strategies merged with linear or regression pattern recognition multivariate techniques (LW-LDA, S-LDA, and VIP-PLS-DA) were tested to develop classification rules while reducing the number of variables to the lowest possible, in order to find the most parsimonious way that best described sample origins. Although the application of VIP-PLS-DA led to the extraction of the highest number of variables, analysis of performance metrics suggested the best results for this methodology, since values of sensitivity, specificity, and classification accuracy of 100% were achieved in internal and in external validation thanks to the contribution of Na, K, Ca, P, Mg, Cu, Co, Mn, Se, Ni, Mo, Li, B, Ba, Bi, Sr, As, V, Rb, Cs, V, Y, U, and Cd.

In summary, the elemental pattern linked to the geographical origin of cuttlefish appears to be determined by a combination of macro, trace and ultra-trace elements of both natural and anthropogenic origin, which are known to be absorbed by the animals from the surrounding environment. In particular, the contribution of anthropogenic elements here strongly emerges as a key analytical determinant for cephalopods authenticity assessment. Of note, also concentrations of some heavy toxic metals such as Cd and As, although being so low as not to represent a potential health risk, were useful for the discrimination purpose.

Considering the permanent and continuous release of all these elements in the aquatic environment deriving from modern industrial and agricultural

practices, the quantification of anthropogenic contaminants in fishery products should therefore be encouraged, as well as further specific critical evaluations of their variation within seawater is recommended. In such a scenario, the promising results obtained may pave the way for practical application of the proposed methodology in the fishery sector to trace back the products to different geographical origins, but also for the protection and ongoing promotion of traditional local fishery products for which a quality mark has been recognized.

Acknowledgements

The authors gratefully acknowledge the financial support from the University of Pardubice (project no. SGS_2020_002) and Dr. Giovanni Muresu Ibba (DAVIMAR S.r.l, Chioggia, Italy) for providing cuttlefish samples.

SUPPLEMENTARY MATERIALS

Supplementary Experimental Section

Essential elements such as calcium, iron, copper, and zinc are frequently normally distributed in fish tissues since concentrations are metabolically regulated by the organisms. Contrarily, non-essential trace elements' concentrations are completely dependent on the exposure to the specific environment and, therefore, frequently characterized by a positive skewed lognormal distribution (Phillips & Russo, 1978). As expected, both data homoscedasticity and normality assumptions tested by the Shapiro-Wilk's and Levene's tests were violated for the largest portion of the elements in the analysed cuttlefish samples ($p \leq 0.05$), thus nonparametric statistics was used to investigate and describe elemental concentrations, taking into consideration that, alongside with this, the main advantage of nonparametric statistical tests is the low sensitivity to small samples sizes and to the potential bias of outlying data (Helsel & Hisrsch, 1992). The comparison among groups of cuttlefish was therefore reported after having performed the nonparametric Kruskal–Wallis analysis of variance followed Dunn's post hoc (Sawilowsky & Fahoome 2014) test for multiple comparison ($p \leq 0.05$).

The first chemometric technique applied to analyze the huge amount of ICP-MS data obtained from cuttlefish sample was principal component analysis (PCA). The PCA is an unsupervised projection method aimed at dimensionality reduction of the original set of correlated variables into new fewer uncorrelated latent variables extracting as much systematic variation as possible of the original data. PCA is mostly used as a preliminary explorative tool to study any intrinsic correlation among observations, variables as well as the bidirectional relation between variables and observation and it is a powerful technique for the rapid detection of strong multivariate outliers (Jolliffe & Cadima, 2016). In the present work, the application of PCA to elemental data was mainly aimed at obtaining the first global description of dataset, after having corrected the non-normal data distribution by Box-Cox transformation (Sakia, 1992) and the wide range of

magnitude of values by Z-score standardization (Shiffler, 1988). In combination with Pearson's correlation analysis applied to Box-Cox transformed data, PCA was also useful for data collinearity reduction necessary for the subsequent step in supervised classification.

To define the optimal number of principal components to retain and avoid data overfitting and overoptimistic results, leave-one-out cross validation (LOOCV) following the approach of Krzanowski was employed (Eastment & Krzanowski, 1982). According to this, one observation per time was retained and predicted by the model which, in turn, was trained on the remaining observations. This sequence was repeated until all observations were kept out one by one. Thus, considering the sample size, LOOCV statistics resulted from a number of 68*3 trained models. The selected internal validation method was chosen taking into consideration the effective number of independent samples in the dataset and because since it is faster to be computed compared to other CV methods.

In study dealing with elemental fingerprints, linear discriminant analysis (LDA) is the most frequently used supervised classification method. By using Euclidean distance, LDA is based on defining new linear combinations of the original variables able to maximize the separation between class of samples while minimizing intra-class variability. The maximum likelihood ratio criterion, referred to as Wilks' lambda ($p \leq 0.05$) was applied to verify the statistical significance of each discriminant function, where smaller values suggested greater discriminatory power associated to the function (Todorov, 2007). Since LDA may be prone to failing in classification, especially when the samples size for each class is greatly unbalanced, equal priors in estimations were not set, but probabilities were calculated based on unequal sample sizes (Sanchez, 1974; Xue & Titterington, 2008).

With the aim to define a classification model by using the minimum number of variables that increased between-group variability and, at the same time, decreased within-class variability, forward stepwise-LDA (S-LDA) was applied to Box-Cox transformed and Z-score standardized data. An *F*-statistic was used to define the statistical significance of each variable to discrimination, where *F*-to-remove value (indicating the cut off value for a

variable to be excluded) was set to 2.71 and *F*-to-enter value (indicating the cut-off value for a variable to be included) was set to 3.84.

Standard LDA and S-LDA were evaluate in terms of recognition ability in LOOCV applied to 66% of original observations (training set), i.e. the percentage of samples correctly assigned to the proper class.

Considering that the internal cross-validation is often insufficient to accurately assess the predictability of the model, an external validation of the training model was also performed using the excluded 33% of data which were randomly but consistently selected from the whole dataset. External test observations which do not uniformly cover the range of training set distribution, may led to misleading results (Consonni, Ballabio & Todeschini, 2010). Partial least square-discriminant analysis (PLS-DA) is a regression-based supervised technique aimed at finding interrelation between original variables and categorical variables by building new (latent) variables for the maximum separation between the different groups of samples. In particular, the categorical variable matrix is transformed into a dummy variable matrix, which boundaries of classification of samples into one class are defined using Bayes theorem. In the present work, the quality and of discriminant models based on PLS-DA of Box-Cox transformed and Z-score standardized data was assessed by applying LOOCV on the 66% of observation of the training set (Eastment & Krzanowski, 1982) and by evaluating the resulting estimating of fitting (R^2X and R^2Y) and predictive ability (Q^2) and the root-mean square error of cross-validation (RMSECV) (Bellabio & Consonni, 2013). Here too, the external validation of the regression model was performed using the independent 33% of observation left out during the calibration step (test set) and the number of samples correctly classified was evaluated. The reliability, in terms overoptimistic fitting and predictability results, was further checked by permutation tests (400 random permutation). The significance of the model was thus confirmed if the y-intercept values of the R^2Y were ≤ 0.4 and y-intercept values of Q^2 were ≤ 0.05 (Van der Voet, 1994). When building classification regression models, the discrimination among informative, redundant, and noisy variables is often an important step. The variable-influence on projection (VIP) index was used to identify and select the most discriminant

variables in PLS-DA (Andersen & Bro, 2010), which were afterward employed to build a new simplified model. The VIP index summarizes the cumulative importance of each variable and represents the weighted sum of squares of the PLS weights, considering the whole explained variability related to responses (sample groups) in all extracted components (Andersen & Bro, 2010). VIP indexes higher than one are considered to be the most significant for explaining the correlation of the observation to all the responses (Andersen & Bro, 2010). The overall quality and robustness of all the supervised models was assessed by taking into consideration the experimental percentages of true positive samples (sensitivity %), true negative samples (specificity%) and prediction accuracy (%). These metrics were calculated as follows, following what has been previously reported by Fawcett (2006):

$$\text{Sensitivity (\%)} = \frac{TP}{TP + FN} * 100$$

$$\text{Specificity (\%)} = \frac{TN}{TN + FP} * 100$$

$$\text{Accuracy (\%)} = \frac{TN + TP}{TN + TP + FN + FP} * 100$$

Whereas: TP = true positive samples

TN = true negative samples

FP = false positive samples

FN = false negative samples

List of supplementary references

Andersen, C. M., & Bro, R. (2010). Variable selection in regression—a tutorial. Journal of Chemometrics, 24(11-12), 728-737. https://doi.org/10.1002/cem.1360

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. Analytical Methods, 5(16), 3790-3798. https://doi.org/10.1039/C3AY40582F

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. Journal of chromatography A, 1158(1-2), 196-214. https://doi.org/10.1016/j.chroma.2007.05.024

Consonni, V., Ballabio, D., & Todeschini, R. (2010). Evaluation of model predictive ability by external validation techniques. Journal of chemometrics, 24(3-4), 194-201. https://doi.org/10.1002/cem.1290

Downey, G. (Ed.). (2016). Advances in food authenticity testing. Woodhead Publishing

Eastment, H. T., & Krzanowski, W. J. (1982). Cross-Validatory Choice of the Number of Components from a Principal Component Analysis. Technometrics, 24(1), 73-77. https://doi.org/10.1080/00401706.1982.10487712

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Helsel and Hisrsch 1992 Statisticsl methods in water resources: Amsterdam the Netherlands, Elsevier Studies in Environmental Science, 49, 522.

Jolliffe, I., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions A, 374(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Phillips, G. R., & Russo, R. C. (1978). Metal bioaccumulation in fishes and aquatic invertebrates: a literature review. U.S. Environmental Protection Agency, Washington, D.C., EPA/600/3-78/103.

Sakia, R. M. (1992). The Box-Cox transformation technique: a review. Journal of the Royal Statistical Society: Series D (The Statistician), 41(2), 169-178.
https://doi.org/10.2307/2348250

Shiffler, R. E. (1988). Maximum Z scores and outliers. The American Statistician, 42(1), 79-80. https://doi.org/10.1080/00031305.1988.10475530

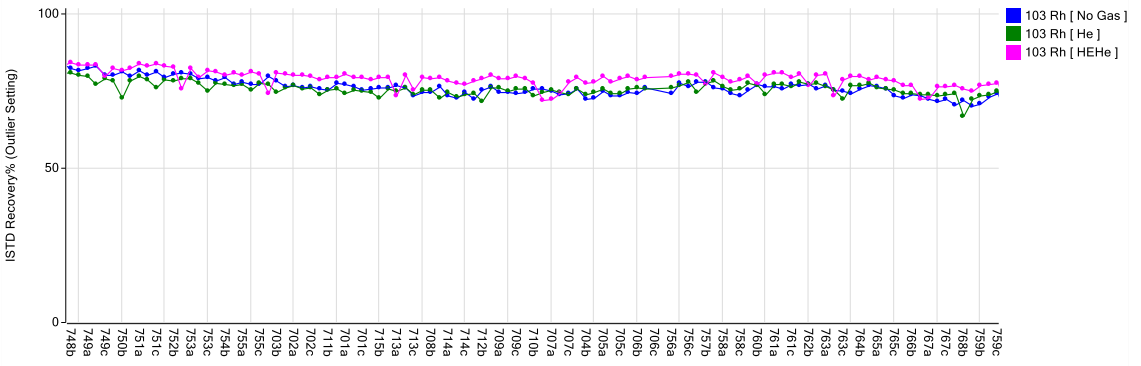
Sanchez, P. M. (1974). The unequal group size problem in discriminant analysis. Journal of the Academy of Marketing Science, 2(4), 629–633.

Todorov, V. (2007). Robust selection of variables in linear discriminant analysis. Statistical Methods and Applications, 15, 395–407. https://doi.org/10.1007/s10260-006-0032-6

Van der Voet, H. (1994). Comparing the predictive accuracy of models using a simple randomization test. Chemometrics and Intelligent Laboratory Systems, 25, 313-323. https://doi.org/10.1016/0169-7439(94)85050-X.

Xue, J. H., & Titterington, D. M. (2008). Do unbalanced data have a negative effect on LDA?. Pattern Recognition, 41(5), 1558-1571. https://doi.org/10.1016/j.patcog.2007.11.008

Sawilowsky, S., & Fahoome, G. (2014). Kruskal-Wallis Test: Basic. Wiley StatsRef: Statistics Reference Online. https://doi.org/10.1002/9781118445112.stat06567



Supplementary Figure S1. Internal standard (ISTD) stability for sequence including 111 samples of cuttlefish (sample name is displayed on the x-axis) measured during the 8-hour run. Due to limited space, not all sample names are shown in the X-axis labels. The ISTD recoveries are displayed relative to the calibration blank.

Supplementary Table S1. Principal spectral interferences encountered on the Agilent 7900 from the major components of food matrices, plasma gas and sample solvent onto the analysis of selected elements.

Interfered isotope	Interfering ion
$^{27}Al^+$	$^{12}C^{15}N^+$, $^{12}C^{14}NH^+$
$^{31}P^+$	$^{14}N^{16}OH^+$, $^{15}N^{16}O^+$,
$^{39}K^+$	$^{23}Na^{16}O^+$, $^{38}ArH^+$
$^{44}Ca^+$	$^{12}C^{16}O^{16}O^+$
$^{48}Ca^+$	$^{31}P^{17}O^+$, $^{31}P^{16}O^1H^+$
$^{51}V^+$	$^{35}Cl^{16}O^+$, $^{37}Cl^{14}N^+$
$^{52}Cr^+$	$^{40}Ar^{12}C^+$, $^{35}Cl^{16}OH^+$, $^{37}Cl^{14}NH^+$
$^{53}Cr^+$	$^{40}Ar^{13}C^+$, $^{35}Cl^{18}O^+$, $^{37}Cl^{16}O^+$, $^{36}Ar^{16}OH^+$
$^{55}Mn^+$	$^{39}K^{16}O^+$, $^{37}Cl^{18}O^+$, $^{23}Na^{32}S^+$
$^{54}Fe^+$	$^{37}Cl^{17}O^+$, $^{37}Cl^{16}O^1H^+$
$^{56}Fe^+$	$^{40}Ar^{16}O^+$, $^{40}Ca^{16}O^+$
$^{57}Fe^+$	$^{40}Ar^{16}OH^+$, $^{40}Ca^{16}OH^+$
$^{58}Ni^+$	$^{40}Ca^{18}O^+$, $^{40}Ar^{18}O^+$, $^{23}Na^{35}Cl^+$
$^{59}Co^+$	$^{43}Ca^{16}O^+$, $^{42}Ca^{16}O^1H^+$, $^{40}Ar^{18}OH^+$
$^{60}Ni^+$	$^{44}Ca^{16}O^{1+}$, $^{43}Ca^{16}O^1H^+$, $^{23}Na^{37}Cl^+$
$^{61}Ni^+$	$^{44}Ca^{16}O^1H^+$, $^{36}Ar^{25}Mg^+$, $^{38}Ar^{23}Na^+$, $^{23}Na^{37}ClH^+$
$^{62}Ni^+$	$^{23}Na^{23}Na^{16}O^+$, $^{44}Ca^{18}O^+$, $^{36}Ar^{26}Mg^+$
$^{63}Cu^+$	$^{40}Ar^{23}Na^+$, $^{31}P^{16}O^{16}O^+$, $^{12}C^{16}O^{35}Cl^+$, $^{12}C^{14}N^{37}Cl^+$
$^{64}Zn^+$	$^{31}P^{16}O^{17}O^+$, $^{32}S^{16}O_2^+$, $^{32}S_2^+$, $^{36}Ar^{12}C^{16}O^+$, $^{38}Ar^{12}C^{14}N^+$, $^{48}Ca^{16}O^+$
$^{65}Cu^+$	$^{31}P^{16}O^{18}O^+$, $^{32}S^{33}S^+$, $^{33}S^{16}O_2^+$, $^{32}S^{16}O_2{}^1H^+$, $^{32}S_2H^+$, $^{14}N^{16}O^{35}Cl^+$, $^{48}Ca^{16}O^+$
$^{66}Zn^+$	$^{34}S^{16}O_2^+$, $^{32}S^{16}O^{18}O^+$, $^{32}S^{17}O_2^+$, $^{33}S^{16}O^{17}O^+$, $^{32}S^{34}S^+$, $^{33}S_2^+$, $^{48}Ca^{18}O^+$
$^{67}Zn^+$	$^{33}S^{34}S^+$, $^{34}S^{16}O^{17}O^+$, $^{33}S^{16}O^{18}O^+$, $^{32}S^{17}O^{18}O^+$, $^{33}S^{17}O_2^+$, $^{33}S_2H^+$, $^{48}Ca^{18}OH^+$, $^{14}N^{16}O^{37}Cl^+$, $^{16}O_2{}^{35}Cl^+$
$^{68}Zn^+$	$^{34}S_2^+$, $^{32}S^{18}O_2^+$
$^{69}Ga^+$	$^{34}S_2H^+$, $^{32}S^{18}O_2H^+$, $^{16}O_2{}^{37}Cl^+$
$^{71}Ga^+$	$^{34}S^{18}O_2H^+$
$^{72}Ge^+$	$^{40}Ar^{32}S^+$, $^{35}Cl^{37}Cl^+$, $^{40}Ar^{16}O_2^+$
$^{73}Ge^+$	$^{40}Ar^{33}S^+$, $^{35}Cl^{37}ClH^+$, $^{40}Ar^{16}O_2H^+$
$^{74}Ge^+$	$^{40}Ar^{33}S^+$, $^{37}Cl_2^+$, $^{39}K^{35}Cl^+$
$^{75}As^+$	$^{43}Ca^{16}O_2^+$, $^{40}Ca^{35}Cl^+$, $^{40}Ar^{34}SH^+$
$^{77}Se^+$	$^{40}Ca^{37}Cl^+$, $^{40}Ar^{37}Cl^+$
$^{78}Se^+$	$^{41}K^{37}Cl^+$, $^{38}Ar^{40}Ca^+$
$^{80}Se^+$	$^{40}Ar_2^+$, $^{40}Ca_2^+$, $^{40}Ar^{40}Ca^+$
$^{82}Se^+$	$^{32}S^{17}O_2{}^{16}O^+$, $^{33}S^{16}O_2{}^{17}O^+$

Supplementary Table S2. Comparison of measured and certified concentrations in selected control standards and intra-day and inter-day variation coefficients (VCs) and recoveries.

Analyte	Reference sample	Declared (mg kg^{-1})	Found[a] (mg kg^{-1})	R[b] (%)	VC (%)	
					Intra	Inter
^{7}Li	NCS ZC73015 Milk Powder	0.040	0.0414 ± 0.0032	104	3.86	6.70
	CRM 12-2-04 Wheat bread flour	0.020	0.0195 ± 0.0021	98	5.45	6.07
^{11}B	NCS ZC73015 Milk Powder	1.56 ± 0.22	1.53 ± 0.08	98	2.75	4.85
	CRM 12-2-04 Wheat bread flour	< 0.5	0.49 ± 0.04	[d]	3.96	7.08
	CRM12-2-03 Lucerne	30	30.2 ± 2.0	101	3.35	9.20
^{23}Na	BCR 184 Bovine muscle	2000	2080 ± 56	104	1.35	2.07
	BCR-CRM 185 Bovine Liver	2100	2101 ± 72	100	1.71	3.32
	NIST 1577 Bovine Liver	2033 ± 64	1918 ± 88	94	2.31	3.78
	NIST 1566 Oyster Tissue	5100 ± 3000	5013 ± 250	98	2.48	3.02
	NCS ZC 73015 Milk Powder	4700 ± 300	4400 ± 752	94	8.57	6.79
	CRM12-2-03 Lucerne	474 ± 23	474 ± 24	100	2.51	3.53
^{24}Mg	CRM 12-2-01 Bovine Liver	650 ± 112	645 ± 60	99	4.68	5.54
	BCR 184 Bovine muscle	1020	1026 ± 30	101	1.39	6.38
	NIST 1577 Bovine Liver	620 ± 42	594 ± 16	96	1.30	1.40
	NIST 1566 Oyster Tissue	1280 ± 90	1316 ± 78	103	2.95	8.28
	NCS ZC 73015 Milk Powder	960 ± 70	880 ± 38	92	2.24	5.77
	CRM 12-2-04 Wheat bread flour	556 ± 29	554 ± 7.2	99	0.65	7.85
	CRM12-2-03 Lucerne	3520 ± 125	3475 ± 336	99	4.82	6.44
^{27}Al	CRM 12-2-04 Wheat bread flour	3	3.4 ± 0.2	113	1.39	3.16
	CRM12-2-03 Lucerne	330	330 ± 5	100	0.78	9.20
^{31}P	BCR 184 Bovine muscle	8300	8351 ± 328	101	1.97	2.62
	BCR-CRM 185 Bovine Liver	11700	11618 ± 452	99	1.94	11.2
	NIST 1566 Oyster Tissue	8100	8131 ± 324	101	1.99	4.91
	NCS ZC 73015 Milk Powder	7600 ± 300	7300 ± 448	96	3.17	5.30
	CRM 12-2-04 Wheat bread flour	2280 ± 85	2289 ± 95	100	2.07	2.24
	CRM12-2-03 Lucerne	3030 ± 90	2982 ± 102	98	1.71	4.60
^{39}K	CRM 12-2-04 Wheat bread flour	2550 ± 100	2596 ± 83	102	1.60	3.73
	BCR 184 Bovine muscle	16600	16354 ± 200	99	0.61	2.57
	BCR-CRM 185 Bovine Liver	11200	10807 ± 350	97	1.62	3.60
	NIST 1577 Bovine Liver	10230 ± 640	10270 ± 512	101	2.49	4.01
	NIST 1566 Oyster Tissue	9690 ± 50	9610 ± 514	99	2.67	3.83
	NCS ZC 73015 Milk Powder	12500 ± 500	11910 ± 312	95	1.33	2.74
	CRM12-2-03 Lucerne	18700 ± 650	18897 ± 638	101	1.69	1.66
^{44}Ca	BCR 184 Bovine muscle	150	151 ± 4	101	1.17	3.54
	BCR-CRM 185R Bovine Liver	131	133 ± 5	102	1.87	3.73

(continued)

Table S2. *(continued)*

Analyte	Reference sample	Declared (mg kg^{-1})	Found[a] (mg kg^{-1})	R[b] (%)	VC (%)	
					Intra	Inter
^{44}Ca	NIST 1566 Oyster Tissue	1500 ± 200	1400 ± 14	93	0.49	2.28
	NCS ZC 73015 Milk Powder	9400 ± 300	8630 ± 740	92	4.29	4.75
	CRM12-2-03 Lucerne	17500 ± 750	17348 ± 229	99	0.65	1.30
^{51}V	NIST 1566 Oyster Tissue	2.3 ± 0.1	2.1 ± 0.1	91	1.43	8.27
	CRM 12-2-01 Bovine Liver	0.26 ± 0.05	0.27 ± 0.01	104	1.73	3.14
	NCS ZC 73015 Milk Powder	0.06	0.050 ± 0.002	83	1.93	3.03
^{52}Cr	BCR 184 Bovine muscle	0.076	0.075 ± 0.003	99	2.13	7.28
	CRM 12-2-01 Bovine Liver	0.044	0.048 ± 0.002	109	2.25	9.74
	NCS ZC 73015 Milk Powder	0.39 ± 0.04	0.40 ± 0.02	103	2.22	1.98
	CRM12-2-03 Lucerne	0.900	0.74 ± 0.03	82	2.02	7.00
^{55}Mn	CRM 12-2-04 Wheat bread flour	22.6 ± 1.1	22.4 ± 0.9	99	1.93	8.13
	BCR 184 Bovine muscle	0.334 ± 0.028	0.320 ± 0.017	96	2.68	4.36
	BCR-CRM 185R Bovine Liver	9.3 ± 0.3	9.4 ± 0.2	101	0.90	10.9
	NIST 1577 Bovine Liver	10.46 ± 0.47	10.2 ± 0.4	98	2.15	2.53
	NIST 1566 Oyster Tissue	17.5 ± 1.2	16.4 ± 2.3	94	7.02	6.73
	CRM 12-2-01 Bovine Liver	7.6 ± 0.5	7.46 ± 0.09	98	0.58	4.20
	NCS ZC 73015 Milk Powder	0.51 ± 0.17	0.50 ± 0.08	98	2.85	9.56
	CRM12-2-03 Lucerne	34.2 ± 1.15	34.7 ± 0.3	101	0.44	1.86
^{56}Fe	BCR 184 Bovine muscle	79 ± 2	78 ± 3	99	1.80	2.72
	BCR-CRM 185R Bovine Liver	214 ± 5	217 ± 6	101	1.40	13.8
	NIST 1577 Bovine Liver	495 ± 32.5	501 ± 24	101	2.39	3.15
	NIST 1566 Oyster Tissue	195 ± 34	183 ± 18	94	4.81	3.62
	CRM 12-2-01 Bovine Liver	495 ± 28	501 ± 24	101	2.39	3.15
	NCS ZC 73015 Milk Powder	7.8 ± 1.3	6.8 ± 0.4	87	2.83	3.06
	CRM 12-2-04 Wheat bread flour	23.8 ± 1.5	22.7 ± 1.3	95	2.97	3.00
	CRM12-2-03 Lucerne	355 ± 18	350 ± 36	99	5.08	1.68
^{59}Co	NIST 1577 Bovine Liver	0.300 ± 0.018	0.303 ± 0.003	101	0.48	1.44
	NIST 1566 Oyster Tissue	0.400	0.320 ± 0.028	80	4.65	7.39
	CRM 12-2-01 Bovine Liver	0.37 ± 0.03	0.36 ± 0.02	97	2.71	1.37
	CRM12-2-03 Lucerne	0.193 ± 0.0185	0.175 ± 0.003	91	0.97	4.79
^{60}Ni	BCR 184 Bovine muscle	0.270	0.274 ± 0.018	102	3.28	9.96
	NIST 1566 Oyster Tissue	1.03 ± 0.19	0.94 ± 0.09	91	4.91	6.41
	CRM 12-2-04 Wheat bread flour	0.3	0.29 ± 0.05	97	3.22	5.70
	CRM12-2-03 Lucerne	2.54 ± 0.18	2.9 ± 0.2	114	2.60	5.89
^{63}Cu	BCR 184 Bovine muscle	2.36 ± 0.06[c]	2.30 ± 0.10	98	2.31	4.33
	BCR-CRM 185R Bovine Liver	189 ± 4	189 ± 3	100	0.85	9.91
	NIST 1577 Bovine Liver	275.2 ± 4.6	271 ± 8	98	1.42	1.74

(continued)

Table S2. *(continued)*

Analyte	Reference sample	Declared (mg kg^{-1})	Found[a] (mg kg^{-1})	R[b] (%)	VC (%)	
					Intra	Inter
^{63}Cu	NIST 1566 Oyster Tissue	63.0 ± 3.5	62 ± 2	98	1.68	0.17
	CRM 12-2-01 Bovine Liver	26.3 ± 1.6	25.7 ± 1.4	98	2.69	4.80
	NCS ZC 73015 Milk Powder	0.51 ± 0.13	0.49 ± 0.01	96	1.33	7.66
	CRM 12-2-04 Wheat bread flour	2.77 ± 0.03	2.74 ± 0.02	99	0.36	3.49
	CRM12-2-03 Lucerne	11.7 ± 0.75	11.3 ± 0.3	97	1.11	4.87
^{66}Zn	BCR 184 Bovine muscle	166 ± 3	168 ± 4	101	1.34	1.45
	BCR-CRM 185 Bovine Liver	142 ± 3	140 ± 35	99	0.58	10.6
	NIST 1577 Bovine Liver	181.1 ± 1.0	180 ± 5	99	1.27	2.95
	NIST 1566 Oyster Tissue	852 ± 14	851 ± 2	99	1.18	0.93
	CRM 12-2-01 Bovine Liver	162 ± 6	162 ± 10	100	3.10	2.44
	NCS ZC 73015 Milk Powder	34 ± 2	34 ± 4	100	5.95	2.49
^{75}As	BCR-CRM 185 Bovine Liver	0.024 ± 0.003	0.025 ± 0.002	104	3.45	6.77
	NIST 1566 Oyster Tissue	13.4 ± 1.9	12.9 ± 0.4	96	1.46	1.58
	CRM 12-2-01 Bovine Liver	0.110 ± 0.016	0.107 ± 0.005	97	2.42	6.20
	CRM 12-2-04 Wheat bread flour	0.017 ± 0.0046	0.0178 ± 0.001	105	2.78	0.45
	CRM12-2-03 Lucerne	0.262 ± 0.020	0.264 ± 0.017	101	3.17	2.38
^{78}Se	BCR 184 Bovine muscle	0.183 ± 0.012	0.187 ± 0.014	102	3.79	3.52
	BCR-CRM 185R Bovine Liver	0.446 ± 0.013	0.444 ± 0.040	99	4.50	5.42
	NIST 1577 Bovine Liver	2.031 ± 0.045	2.03 ± 0.06	100	1.39	5.20
	NIST 1566 Oyster Tissue	2.1 ± 0.5	2.2 ± 0.5	105	2.55	9.27
	CRM 12-2-01 Bovine Liver	0.325 ± 0.014	0.33 ± 0.03	102	3.88	10.2
	CRM 12-2-04 Wheat bread flour	0.040	0.039 ± 0.005	98	6.83	5.69
	NCS ZC 73015 Milk Powder	0.11 ± 0.03	0.104 ± 0.006	95	2.75	7.36
	CRM12-2-03 Lucerne	0.050	0.051 ± 0.002	102	0.23	2.83
^{85}Rb	NIST 1566 Oyster Tissue	4.45 ± 0.09	4.3 ± 0.3	97	2.85	3.81
	CRM 12-2-01 Bovine Liver	16.0 ± 2.7	16.1 ± 1.0	101	3.19	3.00
	NCS ZC 73015 Milk Powder	11.6 ± 0.7	11 ± 1	95	5.33	6.71
	CRM 12-2-04 Wheat bread flour	1.5	1.48 ± 0.08	99	2.61	2.04
	CRM12-2-03 Lucerne	16.1 ± 2.2	15.5 ± 0.8	96	2.49	5.11
^{88}Sr	NIST 1566 Oyster Tissue	10.36 ± 0.56	9.91 ± 0.05	96	0.27	2.90
	NCS ZC 73015 Milk Powder	5.3 ± 0.6	4.7 ± 0.1	89	1.15	1.65
	CRM 12-2-04 Wheat bread flour	1.53 ± 0.16	1.56 ± 0.11	102	2.04	1.14
^{89}Y	NCS ZC 73015 Milk Powder	8 ± 3[c]	8.9 ± 0.4c	111	1.93	3.03
^{95}Mo	NIST 1577 Bovine Liver	3.30 ± 0.13	3.40 ± 0.15	103	2.18	2.46
	NIST 1566 Oyster Tissue	0.2	0.20 ± 0.02	100	4.00	5.04
	CRM 12-2-01 Bovine Liver	3.5 ± 0.6	3.8 ± 0.3	109	3.35	5.96
	NCS ZC 73015 Milk Powder	0.28 ± 0.03	0.27 ± 0.03	96	5.43	4.16
	CRM 12-2-04 Wheat bread flour	0.2	0.203 ± 0.011	102	2.80	5.16
	CRM12-2-03 Lucerne	0.200	0.191 ± 0.011	96	2.94	3.50
^{111}Cd	BCR 184 Bovine muscle	0.013 ± 0.002	0.0132 ± 0.002	102	7.81	9.32
	BCR-CRM 185 Bovine Liver	0.298 ± 0.025	0.278 ± 0.008	93	1.38	1.34

(continued)

Table S2. *(continued)*

Analyte	Reference sample	Declared (mg kg^{-1})	Found[a] (mg kg^{-1})	R[b] (%)	VC (%)	
					Intra	Inter
^{111}Cd	NIST 1577 Bovine Liver	0.097 ± 0.0014	0.096 ± 0.005	99	2.46	2.52
	NIST 1566 Oyster Tissue	3.5 ± 0.4	3.24 ± 0.09	93	1.42	2.00
	CRM 12-2-01 Bovine Liver	0.48 ± 0.03	0.47 ± 0.03	98	3.13	3.68
	CRM 12-2-04 Wheat bread flour	0.0415 ± 0.0032	0.039 ± 0.0003	94	0.35	1.41
	CRM12-2-03 Lucerne	0.136 ± 0.0065	0.136 ± 0.005	100	2.00	2.27
^{118}Sn	CRM 12-2-04 Wheat bread flour	< 3	0.31 ± 0.04	d	5.78	4.47
^{121}Sb	CRM 12-2-01 Bovine liver	0.033	0.033 ± 0.006	100	9.76	6.99
	NCS ZC 73015 Milk Powder	6[c]	5.98 ± 0.26c	99	2.23	11.6
	CRM 12-2-03 Lucerne	0.080	0.075 ± 0.008	94	5.12	4.00
^{133}Cs	NIST 1577 Bovine Liver	21.7 ± 1.4c	22 ± 3c	101	6.91	12.1
	CRM 12-02-01 Bovine liver	0.047	0.044 ± 0.001	94	1.12	3.88
	NCS ZC 73015 Milk Powder	0.034 ± 0.005	0.031 ± 0.005	91	7.54	10.1
	CRM 12-2-04 Wheat bread flour	5[c]	4.9 ± 0.8c	98	8.59	14.6
	CRM 12-2-03 Lucerne	0.090	0.086 ± 0.002	96	1.37	1.75
^{138}Ba	NCS ZC 73015 Milk Powder	1.0 ± 0.3	0.86 ± 0.03	86	1.77	8.89
	CRM 12-2-04 Wheat bread flour	1.5	1.58 ± 0.22	105	7.26	0.15
	CRM 12-2-03 Lucerne	23.4 ± 2.1	23.6 ± 0.7	101	1.57	4.06
^{139}La	CRM 12-2-01 Bovine Liver	0.070	0.071 ± 0.004	101	2.66	7.36
	CRM12-2-03 Lucerne	0.940	0.95 ± 0.03	101	1.73	4.00
^{140}Ce	CRM12-2-03 Lucerne	1.0	1.04 ± 0.06	104	3.01	4.06
^{141}Pr	NCS ZC 73015 Milk Powder	0.7[c]	0.59 ± 0.02c	85	1.35	12.0
^{146}Nd	NCS ZC 73015 Milk Powder	2[c]	2.2 ± 0.4c	110	8.82	12.7
^{147}Sm	NCS ZC 73015 Milk Powder	0.5[c]	0.46 ± 0.09c	92	11.4	13.6
	CRM12-2-03 Lucerne	0.140	0.141 ± 0.006	101	2.16	3.27
^{153}Eu	NCS ZC 73015 Milk Powder	0.4c	0.45 ± 0.09c	112	2.28	8.98
	CRM12-2-03 Lucerne	0.030	0.031 ± 0.001	103	1.53	2.39
^{159}Tb	NCS ZC 73015 Milk Powder	0.7c	0.61 ± 0.07c	87	5.70	12.1
	CRM12-2-03 Lucerne	0.020	0.018 ± 0.001	90	2.93	3.63
^{163}Dy	NCS ZC 73015 Milk Powder	0.45c	0.40 ± 0.03c	89	4.49	10.9
	CRM12-2-03 Lucerne	0.090	0.087 ± 0.007	97	3.89	4.96
^{165}Ho	NCS ZC 73015 Milk Powder	0.07c	0.073 ± 0.004c	104	2.46	11.8
^{166}Er	NCS ZC 73015 Milk Powder	0.16c	0.19 ± 0.004c	118	0.84	1.10
^{175}Lu	CRM12-2-03 Lucerne	5[c]	4 ± 0.2c	80	3.22	6.83
^{178}Hf	CRM12-2-03 Lucerne	0.100	0.082 ± 0.002	82	1.77	5.83
^{202}Hg	BCR 185 Bovine Liver	0.044 ± 0.003	0.0460 ± 0.0005	105	0.54	d
	NIST 1577c Bovine Liver	5.36 ± 0.17c	5.0 ± 0.4c	93	3.80	7.06
	NIST 1566 Oyster Tissue	0.057 ± 0.015	0.0528 ± 0.0004	93	0.38	d
	CRM 12-2-01 Bovine Liver	0.37 ± 0.02	0.35 ± 0.02	95	3.14	5.92
	NCS ZC 73015 Milk Powder	2.2[c]	2.0 ± 0.2	91	4.50	7.73
^{205}Tl	NIST 1566 Oyster Tissue	5[c]	4.92 ± 0.05c	98	1.04	10.9
	NCS ZC 73015 Milk Powder	0.9[c]	0.89 ± 0.03c	99	1.66	6.40

(continued)

Table S2. *(continued)*

Analyte	Reference sample	Declared (mg kg^{-1})	Found[a] (mg kg^{-1})	R[b] (%)	VC (%)	
					Intra	Inter
Pb[e]	CRM 12-2-04 Wheat bread flour	0.041 ± 0.0078	0.043 ± 0.0076	104	9.97	8.64
	BCR 184 Bovine muscle	0.239 ± 0.011	0.2343 ± 0.0096	98	2.04	2.35
	BCR-CRM 185 Bovine Liver	0.501 ± 0.027	512 ± 72c	102	7.03	5.92
	NIST 1566 Oyster Tissue	0.480 ± 0.040	0.475 ± 0.040	99	4.20	4.95
	CRM 12-2-01 Bovine Liver	0.71 ± 0.08	0.71 ± 0.05	100	3.79	4.54
	NCS ZC 73015 Milk Powder	0.07 ± 0.02	0.071 ± 0.003	101	1.77	6.55
	CRM12-2-03 Lucerne	1.84 ± 0.17	1.89 ± 0.07	103	1.74	4.86
^{209}Bi	NCS ZC 73015 Milk Powder	1.2[c]	1.27 ± 0.22c	106	8.89	13.8
^{232}Th	NCS ZC 73015 Milk Powder	2.8[c]	2.64 ± 0.05c	99	1.00	10.5
	CRM12-2-03 Lucerne	0.110	0.109 ± 0.006	99	2.78	3.68
^{238}U	NIST 1566 Oyster Tissue	0.116 ± 0.006	0.112 ± 0.006	97	2.79	7.58
	NCS ZC 73015 Milk Powder	3[c]	3.12 ± 0.14c	104	2.18	5.18

[a] Mean ± 2 S.D. (n = 3).

[b] Recovery (%) = (Found value/Declared value)×100.

[c] ◎g kg^{-1}

[d] Not determined.

[e] Pb was measured as the sum of the three most abundant isotopes, ^{206}Pb^{+}, ^{207}Pb^{+} and ^{208}Pb^{+}.

Table S3. Method Detection Limits (MDL)[a] and Method Limits of Quantification (MLOQ)[a] (μg kg^{-1}) and normalized calibration slopes (NCS) (1/μg L^{-1}) of Agilent 7900 Q-ICP-MS for analysis of different elements in cuttlefish samples with the use of Rh as internal standard.

Analyte	Cell mode	NCS	MDL	MLOQ	Analyte	NCS	Cell mode	MDL	MLOQ
$^{7}Li^{+}$	No gas	5.2×10^{-3}	0.59	1.98	$^{118}Sn^{+}$	1.73×10^{-2}	No gas	0.49	1.6
$^{11}B^{+}$	No gas	1.4×10^{-3}	5.7	19	$^{121}Sb^{+}$	2.3×10^{-2}	No gas	0.15	0.49
$^{23}Na^{+}$	He	1.2×10^{-3}	3500	11667	$^{133}Cs^{+}$	6.4×10^{-2}	No gas	0.015	0.051
$^{24}Mg^{+}$	No gas	1.3×10^{-2}	15	51	$^{138}Ba^{+}$	5.6×10^{-2}	No gas	0.55	1.83
$^{27}Al^{+}$	He	1.1×10^{-4}	8.6	29	$^{139}La^{+}$	6.4×10^{-2}	No gas	0.013	0.043
$^{31}P^{+}$	HE He	3.4×10^{-5}	161	536	$^{140}Ce^{+}$	6.4×10^{-2}	No gas	1.50	5.0
$^{39}K^{+}$	He	4.0×10^{-4}	735	2448	$^{141}Pr^{+}$	7.8×10^{-2}	No gas	0.010	0.033
$^{44}Ca^{+}$	He	1.9×10^{-5}	1291	4304	$^{146}Nd^{+}$	1.2×10^{-2}	No gas	0.07	0.22
$^{51}V^{+}$	He	8.1×10^{-3}	0.06	0.21	$^{147}Sm^{+}$	1.1×10^{-2}	No gas	0.27	0.88
$^{52}Cr^{+}$	He	1.1×10^{-2}	0.97	3.2	$^{153}Eu^{+}$	4.0×10^{-2}	No gas	0.019	0.063
$^{55}Mn^{+}$	He	4.3×10^{-3}	2.4	8.1	$^{157}Gd^{+}$	1.8×10^{-2}	No gas	0.046	0.153
$^{56}Fe^{+}$	He	7.7×10^{-3}	6.5	22	$^{159}Tb^{+}$	8.1×10^{-2}	No gas	0.0095	0.032
$^{59}Co^{+}$	He	2.1×10^{-2}	0.12	0.41	$^{163}Dy^{+}$	1.9×10^{-2}	No gas	0.070	0.23
$^{60}Ni^{+}$	He	5.7×10^{-3}	6.3	21	$^{165}Ho^{+}$	7.7×10^{-2}	No gas	0.054	0.18
$^{63}Cu^{+}$	He	1.7×10^{-2}	2.0	6.7	$^{166}Er^{+}$	2.6×10^{-2}	No gas	0.031	0.102
$^{66}Zn^{+}$	No gas	5.8×10^{-3}	159	528	$^{172}Yb^{+}$	1.7×10^{-2}	No gas	0.047	0.157
$^{75}As^{+}$	HE He	2.5×10^{-3}	0.49	1.6	$^{175}Lu^{+}$	7.2×10^{-3}	No gas	0.011	0.035
$^{78}Se^{+}$	HE He	3.7×10^{-4}	1.6	5.5	$^{178}Hf^{+}$	2.1×10^{-3}	No gas	0.0023	0.0075
$^{85}Rb^{+}$	No gas	4.3×10^{-2}	0.09	0.31	$^{185}Re^{+}$	2.3×10^{-3}	No gas	0.12	0.39
$^{88}Sr^{+}$	No gas	5.6×10^{-2}	0.19	0.62	$^{195}Pt^{+}$	1.6×10^{-3}	No gas	0.18	0.59
$^{89}Y^{+}$	No gas	6.8×10^{-2}	0.03	0.10	$^{205}Tl^{+}$	4.1×10^{-2}	No gas	0.027	0.088
$^{90}Zr^{+}$	No gas	3.5×10^{-2}	0.11	0.35	Pb [b]	5.3×10^{-2}	No gas	0.12	0.40
$^{95}Mo^{+}$	No gas	1.0×10^{-2}	0.58	1.93	$^{209}Bi^{+}$	4.4×10^{-2}	No gas	0.033	0.11
$^{101}Ru^{+}$	No gas	1.2×10^{-2}	0.07	0.22	$^{232}Th^{+}$	4.7×10^{-2}	No gas	0.033	0.11
$^{105}Pd^{+}$	No gas	1.34×10^{-2}	0.14	0.45	$^{238}U^{+}$	4.8×10^{-2}	No gas	0.0005	0.0017
$^{111}Cd^{+}$	No gas	6.0×10^{-3}	0.0038	0.013	Hg [c]	2.8×10^{-2}		0.2	0.7

[a] Values were calculated assuming a sample mass of 0.100 g.

[b] Pb is measured as the sum of the three most abundant isotopes, $^{206}Pb^{+}$, $^{207}Pb^{+}$ and $^{208}Pb^{+}$.

[c] Values were evaluated for direct analysis of Hg by single purpose atomic absorption spectrometer AMA 254.

Table S4. Loading weights on the first two principal components from PCA performed using all the elements. Variables with values higher than 0.7 were selected and used for the creation of LW-LDA model.

Variable	Loading Weight		Variable	Loading Weight	
	PC 1	PC 2		PC 1	PC 2
Li	0.9161	0.9828	Sn	0.4052	0.5677
B	0.6713	0.7207	Sb	0.4487	0.6867
Na	0.9835	0.9617	Cs	0.3746	0.1279
Mg	0.8714	0.9447	Ba	0.3775	0.6344
Al	0.2004	0.2489	La	0.3005	0.4644
P	0.4261	0.1216	Ce	0.2928	0.4470
K	0.4533	0.0000	Pr	0.0005	0.3591
Ca	0.8132	0.9881	Nd	0.0000	0.3737
V	0.3980	0.6934	Sm	0.0758	0.3760
Cr	0.5316	0.5639	Eu	0.0588	0.4170
Fe	0.1041	0.4436	Gd	0.0242	0.4294
Mn	0.1649	0.2719	Tb	0.0233	0.4141
Ni	0.3848	0.3796	Dy	0.0654	0.3997
Co	0.6354	0.8007	Ho	0.0248	0.4180
Cu	0.5961	0.7946	Er	0.0443	0.4210
Zn	0.4682	0.8007	Yb	0.1320	0.4139
As	0.4623	0.1563	Lu	0.1959	0.4488
Se	0.5489	0.6970	Hf	0.7509	0.6634
Rb	0.4229	0.0328	Re	0.6194	0.4359
Sr	0.9352	1.0000	Pt	0.7033	0.5132
Y	0.2576	0.3511	Tl	0.4505	0.2269
Zr	0.4817	0.6681	Pb	0.2908	0.5203
Mo	0.6908	0.9030	Bi	0.4470	0.2509
Ru	0.7650	0.5617	Th	0.2504	0.5523
Pd	0.8914	0.8030	U	0.7600	0.9832
Cd	0.9222	0.9777	Hg	0.6800	0.4359

Table S5. Fisher's classification coefficients for the LW-LDA model.

Variable selected	Fisher's classification coefficients		
	CH	**MED**	**AT**
Li	–13.496	3.080	5.994
B	14.378	0.615	–10.118
Na	–24.484	–6.308	23.663
Mg	30.145	1.805	–21.952
Ca	19.229	11.045	–24.197
Co	1.275	9.925	–9.745
Cu	3.603	2.396	–4.396
Zn	–0.484	–0.790	1.294
Sr	–49.228	–14.310	45.787
Mo	–3.046	1.769	0.438
Ru	0.305	–0.666	0.142
Pd	–0.615	1.774	–0.946
Cd	–6.960	–11.474	13.922
Hf	1.132	–0.357	–0.287
Pt	0.882	–0.344	–0.211
U	–2.835	–3.589	5.752
(Constant)	–31.893	–8.675	–24.855

CH = Cuttlefish from Chioggia (FAO 37.2.1). MED = cuttlefish from Mediterranean Sea (FAO 37.1/37.2) AT = cuttlefish from French Atlantic Ocean (FAO 27.7.e).

Table S6. Variables selected according to Wilk's Lambda values used for the creation of stepwise-LDA models and relative Fisher's classification coefficients for each group.

Variable selected	Lambda value	*T*	*F* to remove	*p*	Fisher's classification coefficients		
					CH	MED	AT
Na	0.011	0.199	153.030	0.000	–38.122	–1.622	26.301
Co	0.004	0.219	12.568	0.000	0.879	7.071	–7.012
B	0.006	0.207	65.093	0.000	15.871	0.960	–11.211
K	0.005	0.110	34.483	0.000	24.739	2.006	–17.469
Cd	0.004	0.171	24.886	0.000	–6.304	–12.504	14.739
V	0.005	0.319	29.741	0.000	–3.456	6.218	–3.797
U	0.004	0.194	14.543	0.000	1.876	–7.844	6.759
Rb	0.003	0.138	5.703	0.000	–9.350	0.011	5.943
Ni	0.003	0.740	7.558	0.000	1.404	2.285	–2.878
Ba	0.003	0.700	4.316	0.000	2.614	–0.112	–1.444
Cu	0.003	0.274	5.639	0.000	1.087	3.725	–3.948
As	0.003	0.428	4.393	0.000	–2.660	–1.828	3.321
(Constant)					–33.155	–10.750	–27.163

T = tolerance. F to remove = maximum F set to 2.71. p = significant level of 0.05. CH = Cuttlefish from Chioggia (FAO 37.2.1). MED = cuttlefish from Mediterranean Sea (FAO 37.1/37.2) AT = cuttlefish from French Atlantic Ocean (FAO 27.7.e)

Table S7. VIP values used for the creation of VIP-PLS-DA model and relative regression coefficients for each group.

Variable selected	VIP value	Regression coefficients		
		CH	MED	AT
V	1.672	–0.2584	0.4376	–0.2115
Co	1.628	–0.1722	0.2502	–0.0995
Na	1.477	–0.1700	0.0678	0.0809
Sr	1.447	–0.2113	0.1065	0.0783
Mn	1.420	0.1365	0.0292	–0.1487
Mo	1.416	–0.0867	0.0393	0.0365
Cd	1.411	0.1963	–0.4848	0.3131
Li	1.391	–0.2127	0.7827	0.0257
U	1.386	0.2916	–0.5152	0.2601
Cu	1.375	–0.0285	0.0930	–0.0681
Ni	1.370	0.0056	0.0572	–0.0620
Ca	1.368	–0.2666	0.2128	0.0204
Mg	1.361	–0.0263	–0.1046	0.1276
B	1.342	0.3117	–0.1665	–0.1062
Se	1.273	–0.1543	0.1802	–0.0451
K	1.242	0.1371	–0.0803	–0.0397
Bi	1.227	0.3117	–0.1665	–0.1062
Rb	1.174	–0.0167	0.1343	–0.1197
Cs	1.060	0.0269	–0.0675	0.0440
P	1.046	0.1007	–0.1280	0.0398
Y	1.003	0.0509	0.1523	0.0212
(Constant)		0.5546	0.7827	0.7827

CH = Cuttlefish from Chioggia (FAO 37.2.1). MED = cuttlefish from Mediterranean Sea (FAO 37.1/37.2) AT = cuttlefish from French Atlantic Ocean (FAO 27.7.e).

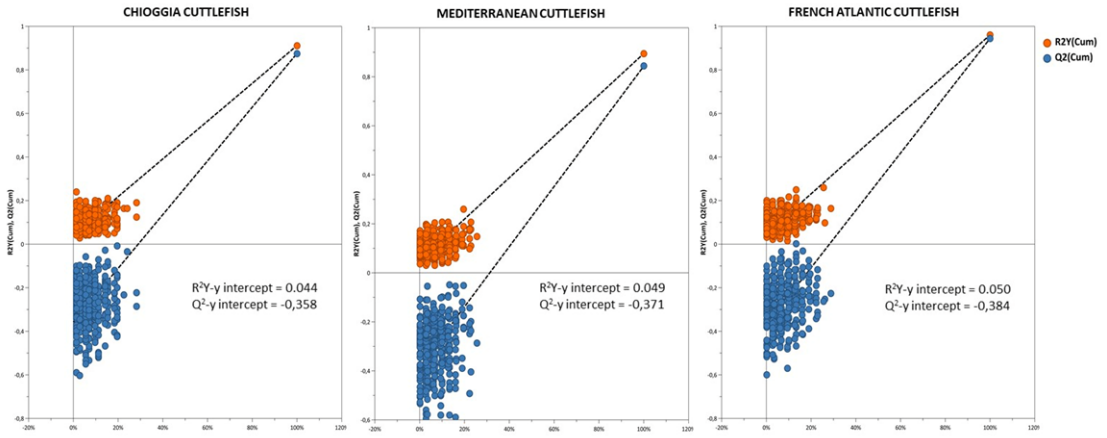


Figure S2. Permutation plot (400 random permutations) assessing the validity of VIP-PLS-DA training model. The intercept to the y-axis of the regression line, correlating the R^2Y and Q^2 values between the original and the permuted y-variables (displayed on the X-axis) and the cumulative value of R^2Y and Q^2 values (displayed on the X-axis), outline the degree of model's overfitting.

References

Abbas, O., Zadravec, M., Baeten, V., Mikuš, T., Lešić, T., Vulić, A., Prpić, J., Jemeršić, L., & Pleadin, J. (2018). Analytical methods used for the authentication of food of animal origin. *Food Chemistry, 246*, 6–17.
https://doi.org/10.1016/j.foodchem.2017.11.007

Arbuckle, N. S. M., & Wormuth, J. H. (2014). Trace elemental patterns in Humboldt squid statoliths from three geographic regions. *Hydrobiologia, 725*(1), 115–123. https://doi.org/10.1007/s10750-013-1608-4

Ayas, D., & Ozogul, Y. (2011). The Effects of Season and Sex in the Metal Levels of Mature Common Cuttlefish (Sepia officinalis) in Mersin Bay, Northeastern Mediterranean. *Journal of Food Science, 76*(4), 121–124.
https://doi.org/10.1111/j.1750-3841.2011.02152.x

Bizzi, C. A., Pedrotti, M. F., Silva, J. S., Barin, J. S., Nóbrega, J. A., & Flores, E. M. M. (2017). Microwave-assisted digestion methods: Towards greener approaches for plasma-based analytical techniques. *Journal of Analytical Atomic Spectrometry. 32*(8), 1448–1466. https://doi.org/10.1039/c7ja00108h

Bosch, A. C., O'Neill, B., Sigge, G. O., Kerwath, S. E., & Hoffman, L. C. (2015). Heavy metals in marine fish meat and consumer health: a review. *Journal of the Science of Food and Agriculture*, *96*, 32–48.
https://doi.org/10.1002/jsfa.7360

Bua, G. D., Albergamo, A., Annuario, G., Zammuto, V., Costa, R., & Dugo, G. (2017). High-Throughput ICP-MS and Chemometrics for Exploring the Major and Trace Element Profile of the Mediterranean Sepia Ink. *Food Analytical Methods, 10*(5), 1181–1190. https://doi.org/10.1007/s12161-016-0680-6

Bustamante, P., Lahaye, V., Durnez, C., Churlaud, C., & Caurant, F. (2006). Total and organic Hg concentrations in cephalopods from the North Eastern Atlantic waters: Influence of geographical origin and feeding ecology. *Science of the Total Environment, 368*(2–3), 585–596.
https://doi.org/10.1016/j.scitotenv.2006.01.038

Carvalho, M. L., Santiago, S., & Nunes, M. L. (2005). Assessment of the essential element and heavy metal content of edible fish muscle. *Analytical and Bioanalytical Chemistry, 382*, 426–432. https://doi.org/10.1007/s00216-004-3005-3

Chaguri, M. P., Maulvault, A. L., Nunes, M. L., Santiago, D. A., Denadai, J. C., Fogaça, F. H., Sant'Ana, L. S., Ducatti, C., Bandarra, N., Carvalho, M. L., Marques, A. (2015). Different tools to trace geographic origin and seasonality of croaker (Micropogonias furnieri). *LWT - Food Science and Technology, 61*(1), 194–200. https://doi.org/10.1016/j.lwt.2014.11.006

Clodia database (2017). Database of Fishery Data from Chioggia, Northern Adriatic Sea. Available online at: http://chioggia.biologia.unipd.it/banche-dati/.
Accessed 26.03.20

Commission of the European Communities (2006). Commission regulation (EC) No 1881/2006 of 19 December 2006 setting maximum levels for certain contaminants in foostuffs. *Official Journal of the European Union, L364/5* (20.12.2006).

Costas-Rodríguez, M., Lavilla, I., & Bendicho, C. (2010). Classification of cultivated mussels from Galicia (Northwest Spain) with European Protected Designation of Origin using trace element fingerprint and chemometric analysis. *Analytica Chimica Acta, 664*(2), 121–128.
https://doi.org/10.1016/j.aca.2010.03.003

Cubadda, F., Raggi, A., & Coni, E. (2006). Element fingerprinting of marine organisms by dynamic reaction cell inductively coupled plasma mass spectrometry. *Analytical and Bioanalytical Chemistry, 384*(4), 887–896.
https://doi.org/10.1007/s00216-005-0256-6

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

Drivelos, S. A., & Georgiou, C. A. (2012). Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. *TrAC - Trends in Analytical Chemistry, 40,* 38–51.
https://doi.org/10.1016/j.trac.2012.08.003

European Food Safety Authority (EFSA), Panel on Contaminants in the Food Chain (CONTAM), (2009). Scientific Opinion on Arsenic in Food. *EFSA Journal, 7(*10), 1351. https://doi.org/10.2903/j.efsa.2009.1351.

European Parliament and Council of the European Union. (2013). Regulation (EU) No 1379/2013 of the the European Parliament and of the Council of 11 December 2013 on the common organization of the markets in fishery and aquaculture products, amending Council Regulations (EC) No 1184/2006 and (EC) No 1224/2009 and repealin Council Regulation (EC) No 104/2000. *Official Journal of the European Union, L354*(28.12.2013), 1–21.

FAO (Food and Agriculture Organization of the United Nations) (2010). FAO Species Fact Sheets, Sepia officinalis (Linnaeus, 1758).
Available online at: http://www.fao.org/fishery/species/2711/en.

Gonzalez, M. H., Souza, G. B., Oliveira, R. V., Forato, L. A., Nóbrega, J. A., & Nogueira, A. R. A. (2009). Microwave-assisted digestion procedures for biological samples with diluted nitric acid: Identification of reaction products. Talanta, 79(2), 396–401. https://doi.org/10.1016/j.talanta.2009.04.001

Gopi, K., Mazumder, D., Sammut, J., Saintilan, N., Crawford, J., & Gadd, P. (2019b). Combined use of stable isotope analysis and elemental profiling to determine provenance of black tiger prawns (Penaeus monodon). Food Control, 95, 242–248. https://doi.org/10.1016/j.foodcont.2018.08.012

Husáková, L., Urbanová, I., Šrámková, J., Černohorský, T., Krejčová, A., Bednaříková, M., Frýdová, E., Nedělková, I., & Pilařová, L. (2011). Analytical capabilities of inductively coupled plasma orthogonal acceleration time-of-flight mass spectrometry (ICP-oa-TOF-MS) for multi-element analysis of food and beverages. *Food Chemistry, 129*(3), 1287–1296.
https://doi.org/10.1016/j.foodchem.2011.05.047

Husáková, L., Urbanová, I., Šídová, T., Cahová, T., Faltys, T., & Šrámková, J. (2015). Evaluation of ammonium fluoride for quantitative microwave-assisted extraction of silicon and boron from different solid samples. International

Journal of Environmental Analytical Chemistry, *95*(10), 922–935. https://doi.org/10.1080/03067319.2015.1070409

Iguchi, J., Isshiki, M., Takashima, Y., Yamashita, Y., & Yamashita, M. (2014). Identifying the origin of Corbicula clams using trace element analysis. *Fisheries Science*, *80*(5), 1089–1096. https://doi.org/10.1007/s12562-014-0775-1

Kang, X., Zhao, Y., Shang, D., Zhai, Y., Ning, J., & Sheng, X. (2018). Elemental analysis of sea cucumber from five major production sites in China: A chemometric approach. *Food Control*, *94*, 361–367. https://doi.org/10.1016/j.foodcont.2018.07.019

Lall, S.P. (2002). The minerals. J.E. Halver, R.P. Hardy (Eds.), Fish Nutrition (3rd ed.), Academic Press, London, pp. 260–308

Li, L., Boyd, C. E., & Sun, Z. (2016). Authentication of fishery and aquaculture products by multi-element and stable isotope analysis. *Food Chemistry*, *194*, 1238–1244. https://doi.org/10.1016/j.foodchem.2015.08.123

Li, L., Han, C., Dong, S., & Boyd, C. E. (2019). Use of elemental profiling and isotopic signatures to differentiate Pacific white shrimp (Litopenaeus vannamei) from freshwater and seawater culture areas. *Food Control*, *95*, 249–256. https://doi.org/10.1016/j.foodcont.2018.08.015

Li, Y., Chen, C., Li, B., Sun, J., Wang, J., Gao, Y., Zhao, Y., Chai, Z. (2006). Elimination efficiency of different reagents for the memory effect of mercury using ICP-MS. *Journal of Analytical Atomic Spectrometry*, *21*(1), 94–96. https://doi.org/10.1039/b511367a

Luo, R., Jiang, T., Chen, X., Zheng, C., Liu, H., & Yang, J. (2019). Determination of geographic origin of Chinese mitten crab (Eriocheir sinensis) using integrated stable isotope and multi-element analyses. *Food Chemistry*, *274*, 1–7. https://doi.org/10.1016/j.foodchem.2018.08.104

Ministerial Decree 5 June 2014. XIV Revisione dell'Elenco Nazionale dei Prodotti Agroalimentari Tradizionali. G.U. Serie Generale n.141 del 20-06-2014 - Suppl. Ordinario n. 48, Ministero delle Politiche Agricole Alimentari e Forestali, MIPAAF: Rome, Italy, 2014. Available online at: https://www.gazzettaufficiale.it/eli/id/2014/06/20/14A04557/sg

Neff, J. M. (2002). Chapter 3 - Arsenic in the Ocean. Bioaccumulation in Marine Organisms, Effect of Contaminants from Oil Well Produced Water (1st ed.). Elsevier Science, pp. 57–78. https://doi.org/10.1016/B978-008043716-3/50004-X

Noack, C. W., Dzombak, D. A., & Karamalidis, A. K. (2014). Rare earth element distributions and trends in natural waters with a focus on groundwater. *Environmental Science and Technology*, *48*(8), 4317–4326. https://doi.org/10.1021/es4053895

Ortea, I., & Gallardo, J. M. (2015). Investigation of production method, geographical origin and species authentication in commercially relevant shrimps using stable isotope ratio and/or multi-element analyses combined with chemometrics: An exploratory analysis. *Food Chemistry*, *170*, 145–153. https://doi.org/10.1016/j.foodchem.2014.08.049

Raimundo, J., Pereira, P., Vale, C., & Caetano, M. (2005). Fe , Zn , Cu and Cd concentrations in the digestive gland and muscle tissues of Octopus vulgaris

and Sepia officinalis from two coastal areas in Portugal. *Ciencias Marinas, 31*(1B), 243–251. https://doi.org/10.7773/cm.v31i12.91

Rashid, N. A., Hussain, W. S. E. C., Ahmad, A. R., & Abdullah, F. N. (2019). Performance of classification analysis: A comparative study between PLS-DA and integrating PCA+LDA. *Mathematics and Statistics, 7*(4), 24–28.
https://doi.org/10.13189/ms.2019.070704

Smith, G. (2018). Step away from stepwise. *Journal of Big Data, 5(*1). https://doi.org/10.1186/s40537-018-0143-6

Smith, R. G., & Watts, C. A. (2009). Determination of the country of origin of farm-raised shrimp (family penaeide) using trace metal profiling and multivariate statistics. *Journal of Agricultural and Food Chemistry, 57*(18), 8244–8249.
https://doi.org/10.1021/jf901658f

Storelli, M. M., Giacominelli-Stuffler, R., Storelli, A., & Marcotrigiano, G. O. (2007). Cadmium and mercury in cephalopod molluscs : Estimated weekly intake. *Food Additives and Contaminants, 23*(1), 25–30.
https://doi.org/10.1080/02652030500242023

Varrà, M. O., Ghidini, S., Zanardi, E., Badiani, A., & Ianieri, A. (2019). Authentication of European sea bass according to production method and geographical origin by light stable isotope ratio and rare earth elements analyses combined with chemometrics. *Italian Journal of Food Safety, 8*(1).
https://doi.org/10.4081/ijfs.2019.7872

CHAPTER 7

Research Article

Classification of Transformed Anchovy Products Based on the Use of Element Patterns and Decision Trees to Assess Traceability and Country of Origin Labelling

Maria Olga Varrà[a], Lenka Husáková[b]*, Jan Patočka[b], Sergio Ghidini[a], and Emanuela Zanardi[a]

[a] Department of Food and Drug, University of Parma, Strada del Taglio 10, 43126 Parma, Italy.

[b] Department of Analytical Chemistry, Faculty of Chemical Technology, University of Pardubice, Studentska 573 HB/D, Pardubice, CZ-532 10, Czech Republic

*Correspondence: lenka.husakova@upce.cz; Tel.: +420 466 037 029
emanuela.zanardi.unipr.it; Tel.: +39-0521-902-760

Abstract: In the present work, quadrupole inductively coupled plasma mass spectrometry (Q-ICP-MS) and direct mercury analysis were used to determine the elemental composition of 180 transformed (salt-ripened) anchovies before and after packaging, whose raw fish originated from three different fishing areas. After pre-screening and selection of the most geographical-related elements, four decision trees-based algorithms corresponding to C5.0, classification and regression trees (CART), chi-square automatic interaction detection (CHAID), and quick unbiased efficient statistical tree (QUEST) were applied to the elemental dataset and validated, to find the most accurate data mining procedure to trace back the provenance of fish. Classification rules generated by the trained CHAID model optimally identified unlabelled testing bulk anchovies (93.9% F-score) by using just 6 out 52 elements, corresponding to As, K, P, Cd, Li, and Sr. The finished packaged product was better modelled by the QUEST algorithm which recognized the origin of anchovies with F-score of 97.7%, taking into account the information carried out by just 5 elements (B, As, K. Cd, and Pd).

The results suggested that the traceability system in the fishery sector may be efficiently supported by simplified and accessible machine learning techniques applied to a limited but effective number of inorganic predictors of origin.

Abbreviations: Certified reference materials, CRMs; classification and regression trees, CART; chi-square automatic interaction detection, CHAID; hierarchical cluster analysis, HCA; high energy Helium mode, HEHe; inductively coupled plasma mass spectrometry, ICP-MS; inductively coupled plasma optical emission spectrometry ICP-OES; internal standard, ISTD; kinetic energy discrimination, KED; method detection limit of the method, MDL; method limit of quantification, MLOQ; quick unbiased efficient statistical tree, QUEST.

1. Introduction

Foodstuff free-trade between nations all over the world, together with increasing diversification into food-related products, recently made the development of easy, rapid, cheap, and robust tools to assess traceability of foodstuffs to become a hot topic in the scientific community as well as in industrial context.

The fishery sector is particularly prone to fraudulent practices but, on the other hand, it is insufficiently protected. The high complexity of the fish supply chain, the high number stakeholders involved, and the fast perishability of fish are a few of the many factors hampering the fight against fraud, which in turn reflect negatively on producers, transformers and final consumers from both economical and sanitary point of view.

The perception of quality of fish and seafood fresh or processed products by consumers is the sum of several different objective and subjective factors and it directly influences the global economic and market values of the product. At present, mislabelling or misrepresenting the origin of fish products keep getting encouraged by the so-called country-of-origin effect, according to which the consumers increasingly tend to associate high quality fish products with specific production areas because of specific sensorial characteristics, ethical or ecological motivations.

In this context, processed seafood products deriving from the industrial transformation of the highly valuable European anchovy (*Engraulis encrasicolus*, L., 1758) are frequently subjected to fraud (Velasco, Aldrey, Pérez-Martín, & Sotelo, 2016).

European anchovy is a small pelagic fish that is mainly fished in the Mediterranean Sea and Black Sea as well as in Eastern Central Africa alongside the Moroccan coasts and in Northeast Atlantic, especially in the Cantabrian Sea (Food and Agriculture Organization of the United Nations, 2020). In addition to the direct consumption as fresh fish, the product is frequently found in the European marketplace in the form of transformed, brine-fermented anchovy or filleted and canned (preserved in oil) anchovy (Food and Agriculture Organization of the United Nations, 2020).

The traditional anchovy transformation process by brine-ripening finds a long tradition in southern Europe. The fish, typically caught by purse seines, is quicky transported to the fish canning industry where it is beheaded, partially eviscerated and punt into ripening containers (barrels), alternating layers of fish and salt. A pressure is then applied on the top layer to facilitate the progressive elimination of water. The fish is ripened until the desired degree of maturation is reached (from 3 up to 11-12 month on average), to then, be moved from the barrels. From that point on, the bulk ripened anchovies can be preserved and packaged in salt to be commercialized or further processed to obtain different products and preparations, for example by filleting and packaging-in-oil.

During the ripening, several chemical and physical modifications occur, including lipolysis, lipid oxidation, and proteolysis (Hernandez-Herrero, Roig-Sagués, López-Sabater, Rodriguez-Jerez, & Mora-Ventura, 1999; Czerner, Agustinelli, Guccione, & Yeannes, 2015). These modifications are of fundamental importance to prolong the shelf life and reduce the microbiological-associated risks and, at the same time, influence the final organoleptic characteristics of the products (Besteiro, Rodríguez, Tilve-Jar, & Pascual López, 2000).

Salted anchovy from the Cantabrian Sea (Northern Spain) is worldwide appreciated as a high-quality product by consumers, thanks to the sensorial characteristics of the raw fish, the strong link with the territory, and the long artisanal tradition behind in manufacturing (Laso et al., 2017). Taking into consideration the Cantabrian anchovy overall reputation and its high commercial value, it is therefore assumed to be object of fraud by substitution with fish from other sources. Therefore, developing methods that aim at providing concrete protection to the product is a matter of the utmost importance.

Up to now, scientific research dealing with the identification of fish and seafood origin was mainly focused on raw untransformed fish and seafood and make use of different approaches. Among these, approaches based on the use of the inorganic components, such as stable isotopes (Carrera & Gallardo, 2017), mineral, trace- and/or ultra-trace elements (Smith & Watts, 2009) and a combination of stable isotopes and trace elements (Li, Han,

Dong, & Boyd, 2019; Varrà, Ghidini, Zanardi, Badiani, & Ianieri, 2019) have been demonstrated to be successful strategies and offer several advantages depending on the reflection of seawater overall compositions on fish flesh.

Tracing back to the origin of processed or highly processed products is considerably more difficult because of the manipulation and the addition of several compounds during preparation. The use of salt during anchovy preparation may represent the most critical point, since it can potentially mask the natural elemental content of fish. Nevertheless, the multiple identification of elements using mass spectrometry (MS)-based techniques such as inductively coupled plasma-optical emission spectroscopy (ICP-OES) and inductively coupled plasma-mass spectrometry (ICP-MS) have been successfully applied to identify the origin of transformed food products such processed tomato products (Lo Feudo, Naccarato, Sindona, & Tagarelli, 2010; Fragni, Trifirò, & Nucci, 2015), fruit juices (Turra et al., 2017), wines (da Costa, Ximenez, Rodrigues, Barbosa, & Barbosa, 2020), dried beef (Franke et al., 2008), hams (Epova et al., 2018), and different types of cheese (Suhaj & Kore, 2008; Moreno-Rojas, Cámara-Martos, Sánchez-Segarra, & Amaro-López, 2012; Magdas et al., 2019). One application dealing with the use of multi-elemental analysis to authenticate seafood products is also available and it concerns the identification of caviar from different origin, which, as anchovy, is a salted product (Rodushkin et al., 2007).

The success of most of these applications was anyway strictly dependent on the support provided by chemometrics and machine learning methods for the identification of those elemental patterns echoing the original environment.

In this study, four decision tree algorithms corresponding to C5.0, classification and regression trees (CART), chi-square automatic interaction detection (CHAID), and quick unbiased efficient statistical tree (QUEST) were applied to elemental content of transformed anchovy products to predict the country of origin of the raw materials and support traceability of the products before and packaging. The reasons behind the selection of decision trees as the basis method of this study are due to the fact that decision tree models are easily understandable and interpretable, quick to build and, in general, low training times are associated with their use.

Moreover, these techniques have high prediction accuracy in many fields so that makes them preferable and trustable choices for this kind of task.

2. Materials and Methods

2.1. Chemicals, standards, and reference materials

All the aqueous solutions employed for analyses were prepared using ultrapure water (0.05 μS cm^{-1}) obtained by the Milli-Q® water purification system (Millipore, Bedford, USA).

For microwave digestion, hydrogen peroxide (H_2O_2, ≥ 30% w/w) for ultra-trace analysis (Fluka Chemie AG, Buchs, Switzerland) and sub-boiled nitric acid prepared from nitric acid (65%, w/w, Selectipur quality, Lach-Ner, Neratovice, Czech Republic) by means of the sub-boiling distillation apparatus Distillacid™ BSB-939-IR (Berghof, Eningen, Germany) were used.

The working calibration solutions for ICP-MS analysis were prepared daily using the multi-element stock solutions "A", "B1", "B2", and "C". Stock solution "A" (10 mg L^{-1} of of Li, B, Al, V, Cr, Fe, Mn, Ni, Cu, Zn, Co, As, Se, Rb, Sr, Zr, Mo, Ru, Pd, Cd, Sn, Sb, Cs, Ba, Hf, Re, Pt, Tl, Pb, Bi, and Th) was prepared from the Supelco ICP multi-element standard solution IV (Merck, Darmstadt, Germany) and single element standards of concentration 1 ± 0.002 g L^{-1} (Analytika Ltd., Prague, Czech Republic or SCP Science, Montreal, Canada).

Stock solutions "B1" (1 mg L^{-1} of La, Ce, Pr, Nd, and U) and "B2" (0.20 mg L^{-1} of Y, Tb, Ho, Yb, Sm, Eu, Gd, Er, Tm, Lu, and Dy) were prepared from the stock solution of rare earth elements Astasol mix "M008" (Analytika Ltd., Prague, Czech Republic). Stock solution "C" (50 mg L^{-1} of Na, Mg, P, K, Ca, Mn, Cu, and Zn) was prepared from single element standards of 1 g L^{-1} (Analytika Ltd., Prague, Czech Republic).

A 1 g L^{-1} stock solution of Rh (SCP Science, Montreal, Canada) was used to prepare the internal standard solution (ISTD) at concentration of 200 μg L^{-1}.

A 10 g L^{-1} stock solution prepared from urea (TraceSelect quality, Fluka Chemie AG, Buchs, Switzerland) was used to prepare carbon reference solutions.

The element quantification accuracy was evaluated using the following certified reference materials (CRMs): NIST SRM 1577 Bovine Liver (National Institute of Science and Technology, NIST, Gaithersburg, MD, USA); NIST SRM 1566 Oyster Tissue (NIST, Gaithersburg, MD, USA); BCR® certified reference material (CRM)184 Bovine muscle (Institute for Reference Materials and Measurements, IRMM, Geel, Belgium); BCR® 185 Bovine Liver (IRMM, Geel, Belgium); CRM NCS ZC73015 Milk Powder (National Research Centre for Certified Reference Materials, NRCRM, Beijing, China); P-WBF CRM 12-2-04 Essential and Toxic Elements in Wheat Bread Flour (pb-anal, Kosice, Slovakia); CRM12-2-03 P-Alfalfa Essential and toxic elements in Lucerne (pb-anal, Kosice, Slovakia); SMU CRM 12-02-01 Bovine liver (pb-anal, Kosice, Slovakia).

2.2. Anchovy sampling and processing

Salt-ripened anchovies as bulk (semi-finished, non-packaged) product and as packaged (finished) product were obtained from the processing of European anchovy (*Engraulis encrasicolus* L.) and provided by the same fish preserves company.

A total of 90 bulk specimens were randomly collected from different ripening barrels' batches after maturation and suddenly vacuum-packaged into plastic bags. Similarly, a total of 90 finished specimens were obtained after salt-packaging of bulk anchovies and provided packaged into glass jars. Both types of products were prepared from salting process (sea salt, not iodized of the same origin of fish) of raw fish caught in the following geographical areas: Cantabrian Sea (Spain, FAO fishing area 27.8, n=30), upper Central Mediterranean Sea (Croatia, FAO fishing area 37.2.1, n=30) and lower Central Mediterranean Sea, (Tunisia, FAO fishing area 37.2.2, n=30).

Detailed information on the sampling and characteristics of the transformed fish used in the present study is reported in Table 1.

Before analysis, each individual fish was carefully cleansed with filter paper to remove external salt and manually peeled, eviscerated, and deboned, and finally minced with a ceramic knife. After that, samples were individually stored into glass vials and frozen at -20 °C.

Table 1. Sampling specification of salted ripened anchovies

Properties	Transformed anchovy product					
	Bulk			Packaged		
Origin:						
Country	Spain	Croatia	Tunisia	Spain	Croatia	Tunisia
FAO fishing zone	27.8	37.2.1	37.2.2	27.8	37.2.1	37.2.2
Number of samples	30	30	30	30	30	30
Ripening time:						
months in barrels	10-11	9-11	9-10	10-11	9-11	9-10

2.2.1. Lyophilization process

Around 3.5 g of the pre-treated fish sample was transferred into 5mL lyophilization vials (borosilicate glass Vacule® equipped with 3-leg stopper, Wheaton, USA) wherein the material was dried. Before the freeze-drying process, the samples were deep-frozen at −80 °C for 24 hours to provide a necessary conditioning for low temperature drying. CoolSafe 4-15 L bench-top freeze dryer (LaboGene, Lynge, Denmark) was employed for the lyophilisation of samples, with the CoolSafe condenser working temperature held at -110°C and a total chamber pressure of 3 hPa. The freeze-dried samples were subsequently homogenized directly inside the glass vials using a plastic rod to obtain a fine powder.

2.2.2. Microwave-assisted acid digestion

For subsequent ICP-MS analysis, 0.1 g of samples or CRMs were weighted (in triplicate) into a 10 mL perfluoroalkoxy (PFA) tube and 4 mL of 16% HNO3 (65%, w/w HNO3, 1:3 diluted) and 1 mL of 30% H2O2 were added. Three PFA tubes were placed into DAC-100S polytetrafluoroethylene vessels (Berghof, Eningen, Germany) previously filled with 25 mL of HNO3 (16%, v/v), by ensuring that the level of liquid in the outer polytetrafluoroethylene vessel was higher than those in the PFA tubes. This way, the vapor pressures were compensated and the evaporation of the solution from the PFA tubes was avoided (Husáková et al., 2015).

Samples were decomposed by the use of a Berghof Speedwave™ MWS-3+ microwave digestion system (Berghof, Germany) with the maximum total

output of the microwave generator (1450 W) via the following multistep program: step 1, 20 min at at 180 °C (ramp 5min); step 2, 20 min at 220 °C (ramp 5 min); steps 3, 5 min at 100 °C (ramp 5 min).

The clear digested samples were diluted with deionized water up to 25 mL and the residual carbon content quantified at 5.58 ± 0.12% by ICP-OES, following the method previously reported by Husáková et al. (2011).

2.3. Mercury analysis

Total Hg content was determined directly on lyophilised solid samples or CRMs using a single-purpose atomic absorption spectrometer AMA 254 (Altec Ltd., Prague, Czech Republic).

Analytical operation conditions as follows: sample mass, 50 mg; drying step, 60 s at 120 °C; decomposition step, 150 s at 750 °C; Hg release step, 45 s at 900 °C; reading step, 60 s monitoring the 253.6 nm absorbance peak. The flow rate of oxygen (99.5%) carrier gas was 170 mL min^{-1}.

2.4. ICP-MS analysis

Element quantification in samples was performed by using an Agilent 7900 quadrupole ICP-MS apparatus (Agilent Technologies, Inc., Santa Clara, CA, USA) equipped with a quartz concentric nebulizer MicroMist (400 μL min^{-1}), the Peltier-cooled (2 °C) Scott quartz spray chamber, quartz torch with 2.5 mm internal diameter injector, standard sampling and skimmer nickel cones with orifices of 1 and 0.45 mm, and an octopole collision/reaction cell for interference removal using kinetic energy discrimination.

Analytical conditions were enhanced before staring sample measurement by using the multi-elemental tuning solution (Agilent Technologies, Inc., Santa Clara, CA, USA) containing 1 μg L^{-1} of Ce, Co, Li, Mg, Tl and Y, in order to obtain the highest possible sensitivity for elements of low, middle and high m/z. A sensitivity of 6000 counts s^{-1} per μg L^{-1} and a resolution of 0.64 amu peak width (full width at half maximum intensity) were achieved for $^{7}Li^{+}$. The same parameters were 50000 counts s^{-1} per μg L^{-1} and 0.62 for $^{89}Y^{+}$, and 30000 counts s^{-1} per μg L^{-1} and 0.60 for $^{205}Tl^{+}$.

While the quantification of certain elements was performed without pressurizing the collision cell (i.e. “no-gas” mode), “Helium” mode (He) and

"High Energy Helium" mode (HE He) were instead used for the quantification of problematic elements mostly suffering from polyatomic interferences (see the following 2.4.1. Section).

The calibration curves for the quantification of 51 elements (R^2>0.999) resulted from the acquisition of working calibration solutions prepared from multi-element stock solutions "A", "B1", "B2", and "C" described in *Section 2.1.* The concentration of elements for calibration were as follows: blank, 1, 5, 10, 50, 100 μg L^{-1} of Li, Be, B, Al, V, Cr, Fe, Mn, Ni, Cu, Zn, Co, Ga, Ge, As, Se, Rb, Sr, Zr, Mo, Ru, Cd, In, Sn, Sb, Te, Cs, Ba, Hf, Ta, Re, Pt, Tl, Pb, Bi, Th; 0.1, 0.5, 1, 5, 10 μg L^{-1} of La, Ce, Pr, Nd, U; 0.02, 0.1, 0.2, 1, 2 μg L^{-1} of Y, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu; 0.5, 1, 5, 10 mg L^{-1} of Na, Mg, P, K, Ca, Mn, Cu, Zn.

To compensate possible instrumental drift and matrix effects, a 200 μg L-1 Rh ISTD was simultaneously aspired and mixed with samples. The percent recovery of the ISTD responses for the entire 12-hour sequence normalized to the calibration blanks shows that there was no gradual loss of sensitivity over time, even when running high matrix samples. Long term stability measured by comparing ISTD responses from the beginning of the sequence to the end was better than 5%.

Samples were measured in triplicate. Calibration solutions, CRMs and blanks were repeatedly acquired throughout ICP-MS analysis to exclude contamination and define accuracy and precision.

2.4.1. ICP-MS working conditions

The following instrumental parameters were employed for ICP-MS analysis: plasma mode, general purpose, RF power (27 MHz), 1550 W; sampling depth, 8 mm; plasma gas flow, 15 L min^{-1}; auxiliary gas flow, 0.9 L min^{-1}; nebulizer gas flow, 1 L min^{-1}; nebulised pump, 0.1 rps; spray chamber temperature, 2°C; isotopes acquired in "No gas mode": ^{7}Li, ^{11}B, ^{24}Mg, ^{66}Zn, ^{85}Rb, ^{88}Sr, ^{89}Y, ^{90}Zr, ^{95}Mo, ^{101}Ru, ^{103}Rh, ^{105}Pd, ^{111}Cd, ^{118}Sn, ^{121}Sb, ^{133}Cs, ^{138}Ba, ^{139}La, ^{140}Ce, ^{141}Pr, ^{146}Nd, ^{147}Sm, ^{153}Eu, ^{157}Gd, ^{159}Tb, ^{163}Dy,^{165}Ho, ^{166}Er, ^{172}Yb, ^{175}Lu, ^{178}Hf, ^{185}Re, ^{195}Pt, ^{205}Tl, $^{206+207+208}$Pb, ^{209}Bi, ^{232}Th, ^{238}U , all with 0.1 s integration time; isotopes acquired in "He mode": ^{23}Na (0.3 s integration time), ^{27}Al (0.1 s integration time), ^{39}K, ^{44}Ca (both 0.3 s integration time), ^{51}V (1 s integration

time), ^{52}Cr, ^{55}Mn, ^{56}Fe, ^{59}Co, ^{60}Ni, ^{63}Cu, ^{103}Rh (all 0.3 s integration time); isotopes acquired in "HEHe mode": ^{31}P (0.1 s), ^{75}As, ^{78}Se (both 1 s integration time), ^{103}Rh (0.3 s integration time).

2.4.2. Quality assurance and validation of the method

The accuracy of data was checked by analysing the five certified reference materials (NIST 1566 Oyster Tissue, BCR CRM 184 Bovine muscle, BCR CRM 185 Bovine Liver, CRM 12-2-01 Bovine Liver, NIST SRM 1577 Bovine Liver). In addition, CRM 12-2-03 Essential and toxic elements in Lucerne, NCS ZC 73015 Milk Powder, and CRM 12-2-04 Wheat bread flour were analysed to assess accuracy in determining lanthanides and actinides. The recovery of all the elements was found to range between 80 and 118% (n=3), thus underlying the good accuracy of data.

Intra-day and inter-day precisions were calculated to assess the overall precision of the method and were determined by analysing single CRMs three times during the same day and during three different days over one month, respectively. The method was found to be precise enough due to the relative standard deviations (RDS %) of intra-day and inter-day precision which were mostly below 14%.

The detection limit of the method (MDL) and the quantification limits of the methods (MLOQs) are reported in Table S1 (Supplementary Materials) and were experimentally determined through the measurement of ten replicates measurements of a blank sample divided by the slope of the calibration curve and by calculating triple and tenfold standard deviation, respectively. In all cases, detection limits were found significantly below the typical requirements for this analysis so that selected elements could be determined at the background level.

2.5. Data elaboration

Statistics was applied to elemental concentrations referring to anchovy dry matter (d.m.) and carried out using IBM SPSS software (v. 23.0, SPSS Inc., Chicago, IL, USA), SIMCA software (v.14.1, Umetrics, Umeå, Sweden), and IBM SPSS Modeler software (v. 18.2, SPSS Inc., Chicago, IL, USA).

Univariate data analysis, consisting of nonparametric Mann-Whitney test ($p \leq 0.05$) and Kruskal–Wallis test plus Dunn's post hoc test ($p \leq 0.05$) were applied to the whole data of the two set of samples (30 samples*3replicates*3 provenances each) to investigate any significant difference between groups of samples.

As a classical unsupervised chemometric method, hierarchical cluster analysis (HCA) using factorial coordinates of principal component analysis was applied to the two anchovy datasets in order to scout data structures and identify group of samples by the similarity of their variables (Drab & Daszykowski, 2014).

To create classification models with good validity and good consistency, the holdout technique (stratified randomly sampling) was adopted when machine learning was applied. The bulk and the packaged anchovy datasets were organized into three different data matrices each in a 70:15:15 partition ratio to create training, validation, and testing and sets, respectively. The training sets were used to estimate the models, the validation set to test and select the best models and the testing sets to confirm the reliability of the selected models.

Different learning algorithms relying on the principle of decision trees and which do not require assumptions about data distribution were chosen to learn how the data of the training set were classifiable according to origin of samples and to create proper prediction models. These were C5.0, CART, CHAID, and QUEST.

Briefly, decision trees work on the division /classification of samples driven by the values of the variables under examination. The output of the method is a set of concatenated classification rules in the form of a decision tree composed by nodes (identifying the features that need to be sorted) and branches (identifying the values assumed by nodes) (Han, Kamber, & Pei, 2011). CART, C5.0, and CHAID and QUEST represent different methods by which the architecture of the decision trees can be built up and differ each other mainly for the segmentation rules applied and the tree optimization method The C5.0 algorithm is based on binary splitting and works by choosing progressively the instances that allow to gain the maximum partitioning information and stopping via the pruning rule, i.e. by removing

from the splits which do not add significant information. CART algorithm is also based on binary splitting but differs from C5.0 essentially for a different stopping rule in the creation of the tree, consisting of the evaluation of the purity of the node, i.e the maximum degree of homogeneity between categories. The CHAID algorithm is a multiway splitting system based on Chi-square statistics to decide for tree ramifications and is based on the measures of the impurity of the nodes. Finally, the QUEST is a binary-split algorithm which, in the case of continuous variables, uses an ANOVA F-Test to create tree nodes. Further information can be retrieved from Han, Kamber, & Pei (2011) and from Rokach & Maimon (2008).

In the present work, CART was built using the Gini Impurity Index to determine the nodes impurity and select input variables. For CART, CHAID, and QUEST a maximum of five tree levels was set to avoid excessive splitting. Building settings for CHAID included the use of Pearson's Chi-square statistics and a Bonferroni adjustment to calculate the adjusted p-values. Significant level for splitting for both CHAID and QUEST was set at 95%.

Training models were compared each other by analysing classical metrics in multivariate classification methods, corresponding to accuracy (%), sensitivity (%), specificity (%), precision (%), and F-score (%) indexes, calculated from the unlabelled test set. Calculations were performed according to Cuadros-Rodríguez, Pérez-Castaño, & Ruiz-Samblás (2016).

3. Results

3.1. Initial data evaluation

3.1.1 Global elemental profiles of transformed anchovy products

The anchovy samples distributions, according to the measured elemental concentrations which varied significantly in relation to the three investigated geographical provenances, are shown Figure 1. The beeswarm box-plots reported revealed that, according to Kruskall-Wallis and Dunn's multiple comparison test, concentrations of B, V, As, and Hg bulk anchovy were ($p<0.05$) different among Cantabrian, Tunisian and Croatian bulk anchovies (Figure 1A). Packaged anchovies, indeed, differed for the same

element concentrations plus those of Li were instead (Figure 1B). Regardless the type of product, the highest amounts of B and V were found in Tunisian samples and those of As and Hg in Croatian samples (Figure 1A, Figure 1B). Complete data matrices reporting the whole concentrations of the 52 elements measured can be found in Tables S2 and S3 (Supplementary Materials). The most abundant element was found to be Na (whose median concentrations ranged from 142 mg kg-1 in bulk Croatian anchovies to 177 mg kg-1 in Tunisian packaged anchovies) followed by P, K, Ca, and Mg. Bulk and packaged products from Cantabrian Sea differed from samples of Mediterranean origin (Croatian and Tunisian) because of significantly higher concentrations of P and K. At the same time, no differences in Na and Ca contents between Cantabrian and Croatian anchovies was encountered, which, instead, were both significantly lower compared to those found in samples originating from Tunisia (Tables S2 and S3, Supplementary Materials).

Similarly, some minor, trace- and ultra-trace elements showed significant variations in relation to the provenance Besides, Cantabrian anchovies were characterized by higher Ni, Mo, Cs, and Tl concentrations and lower Al concentrations than those encountered in the two Mediterranean products (Supplementary Tables S2 and S3). Considering the peculiarity of the products investigated, no published works dealing with the same food matrices and purposes were found, except for similar canned products including, however, other ingredients (Ikem & Egiebor, 2005). For this reason, a more in-depth analysis of the elemental concentrations was not possible.

Beyond statistical differences related to the geographical origin, a high degree of variations of elemental concentrations within fish of the same provenance was however found. Considering equal characteristics of marine environment for each group, this variation is likely to be attributable to the natural biological diversity among individuals. Moreover, the technology behind the processing of anchovies does suggest that all manufacturing operations (including handling, treatment, production, and distribution) between the time of fishing and the end-product stage can impact the final

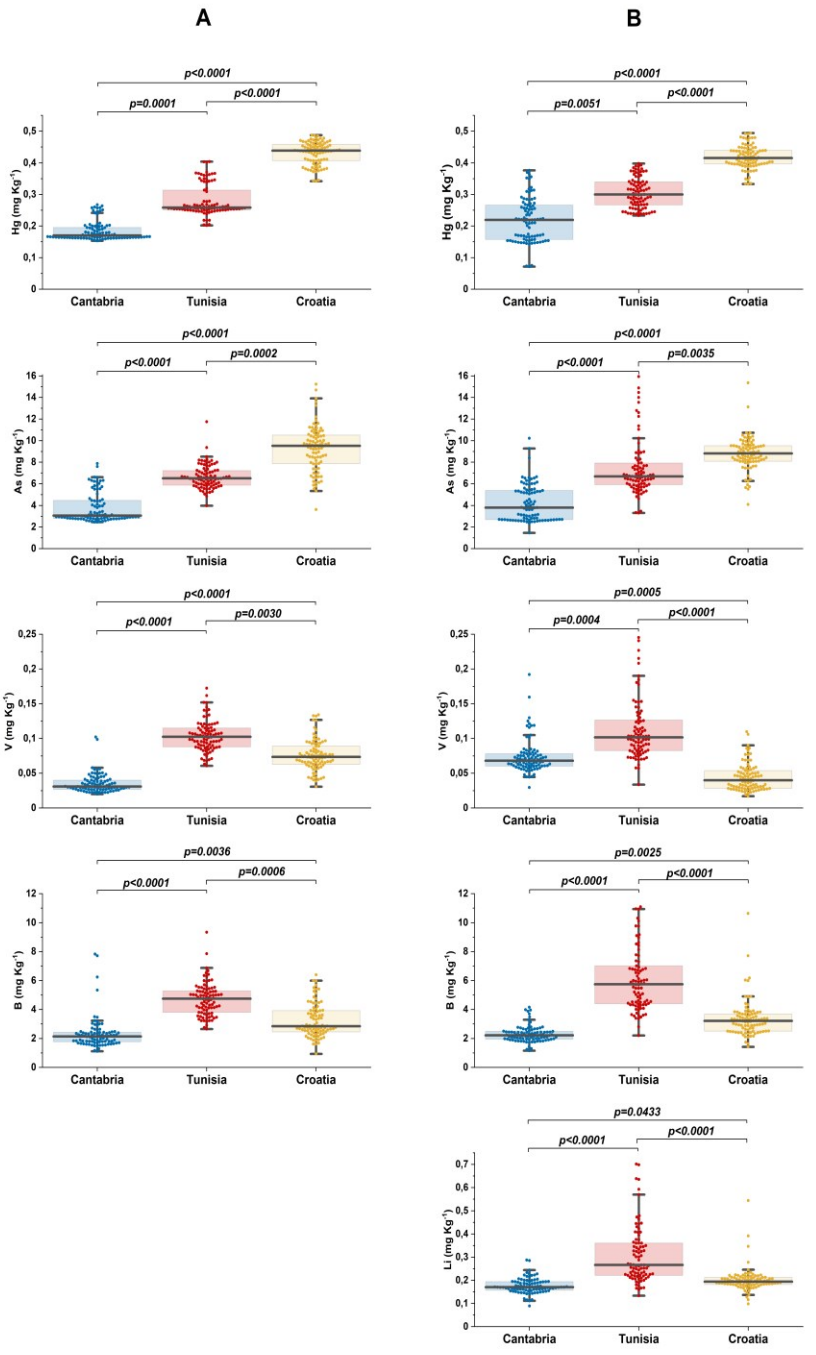


Figure 1. Beeswarm box-plots with Kruskall-Wallis and Dunn's multiple comparison test results (median and quartiles) showing elements in bulk (column A) and packaged (column B) anchovy products varying significantly in relation to the origin ($p < 0.05$).

elemental profile of anchovies. Therefore, potential markers of geographical origin need to go through a more-in-depth evaluation in order to be correctly identified, thereby preventing misinterpretations.

3.1.2 Pre-selection of elements as potential indicators of origin

The use of salt in the form of saturated brine during the processing and packaging of ripened anchovies is one of the main complicating factors which might limit the proper identification of markers of geographical origin since it inevitably modifies the natural element profile of the product. For instance, the use of a brine for fermentation purposes was reported to decrease the total amount of certain trace elements in transformed fish roe because of partitioning phenomena between the solid and the liquid phases occurring during fermentation (Bekhit, Morton, & Dawson, 2008).

Despite the addition of salt as an exogenous source of contaminants, the possibility of using the elemental profile to discriminate the origin of processed salted cheeses has been previously investigated and it was found that some elements as Sr, Li, Mg, Rb and K (Magdas et al., 2019) as well as Tl, Li (Epova et al., 2018) can still be used as powerful tracers since retain a strong link with the place of origin of milk. In addition, concentrations of As, Ba, Br, I, Mo, and Se, were proven to stable although processing and were identified as useful markers for the geographic origin of caviar (Rodushkin et al., 2007) as well as to distinguish caviar obtained from wild sturgeon (Depeters, Puschner, Taylor, & Rodzen, 2013). On the contrary, the concentration of as Fe, Al, Ti, V were found to be heavily affected by handling and packaging operations and, therefore, useless for authentication of transformed products (Rodushkin et al., 2007).

In the present work, the elemental profile of the bulk anchovies of each origin was compared to that of the finished packaged products, in order to highlight possible modifications occurring during the end stages of anchovies processing. Results from Mann-Whitney test (two-tailed, confidence level 95%) highlighted the concentrations of following elements to be significantly different between the two types of products at least in two out of three origin groups: Na, Mg, Ca, Cu, Cr, V, Ru, Rb, La, Ce, Pr, Gd, Re and Tl (see Supplementary Materials, Table S4).

Considering that the main aim of the present research was to create machine learning based models as robust as possible in classifying samples by origin, highly variable elements identified both in the present study (Na, Mg, Ca, Cu, Cr, V, Ru, Rb, La, Ce, Pr, Gd, Re and Tl), and retrieved from the literature (Fe, Ti, and Al) were excluded and a total of 35 input variables, i.e. elements, were retained for subsequent classification analyses.

3.1.3. Hierarchical cluster analysis (HCA)

The inner potential of the elemental profile in guiding the creation of groups of anchovy was at first instance explored. Any possible natural difference or similarity among anchovies of both datasets was therefore uncovered by the application of HCA. Since different methods to measure distances (i.e. Euclidean and Manhattan distances) and to perform grouping (i.e. Ward's minimum variance, single linkage, complete linkage, simple average group average, median, and centroid clustering methods) are applicable for HCA building, the cophenetic correlation coefficient (CCC) was employed as an index to compare the methods with each other and to evaluate the validity of the resulting dendrograms (Sokhal & Rohlf, 1962; Saraçli, Doğan, & Doğan, 2013). Further details about the tested methods can be retrieved from Everitt, Landau, Leese, & Stahl (2011).

By evaluating the CCC reported in Table S5 (Supplementary Materials) and taking into consideration that the closer is this index to 1, the higher is the degree of fit of clustering (Saraçli et al., 2013) it was found that the use of the Euclidean inter-point distance and the Ward's aggregation method to compute HCA was the most performant HCA-based methods both for bulk and packaged anchovies, with CCC values of 0.9888 and 0.9836, respectively (see Table S5, Supplementary Materials). Dendrograms resulting from the proposed methodology are shown in Figure 2. As it can be observed, bulk anchovies (Figure 2A) and packaged anchovies (Figure 2B) were gathered into three major clusters at dissimilarity values of 65% and 70%, respectively. These three clusters mostly enclosed samples of the same origin, thus suggesting the existence elemental patterns strong enough to reflect on the presence of geographical origin-driven groupings. Nevertheless, a few bulk samples from Croatia drifted apart from the others

(Figure 2A), as well as some samples from Croatia and Tunisia were mingled together in the second cluster of the dendrogram of packaged anchovies (Figure 2B). Thus, Cantabrian products were in general better clustered than samples from Croatia and Tunisia which, on the contrary, were more connected each other. This is easily explained by the fact that the two fishing areas of anchovy from Croatia and Tunisia are neighbouring zones of the Mediterranean Sea (FAO fishing area 37.2.1 and 37.2.2, respectively). Therefore, these samples may share lots of compositional features linked to the similar environmental characteristics.

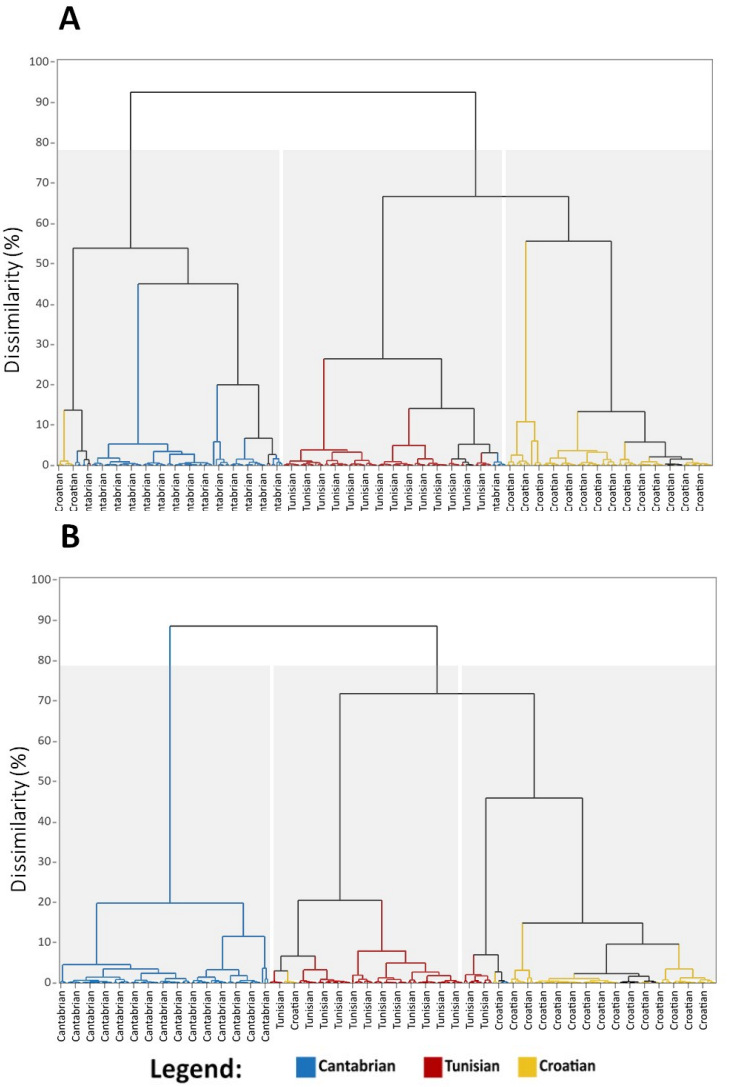


Figure 2. Hierarchical cluster analysis simplified dendrograms for semi-finished anchovy dataset (A) and finished anchovy dataset (B) based on 35 elements.

3.2. *Data mining for the geographic origin evaluation*

Considering that the main task of the present research was to verify the usefulness of elemental profile of anchovy products to the development of rapid methods for geographical origin verification before and after the product packaging, different machine learning algorithms were explored in order to identify the best-suited one to this purpose.

Four decision trees algorithm (C5.0, CART, CHAID, and QUEST) trained on 70% of original 270 data were examined, in search of the most accurate models in identifying the origin labels of bulk and packaged anchovy samples in the validation set. One of the main advantages of using decision tree is that the selection of the most important variables for classification is performed automatically during training stage. Therefore, computation time is reduced, while interpretability, accessibility and handiness of the models is improved.

Performance outcomes of the trained and validated classification models are summarised in Table 2. The rate of classification accuracy of the samples of the training sets ranged between 89.9% and 99.4%, with better results shown up by C5.0 both for the bulk and the packaged anchovy dataset. When validation of the models was performed, 98.3% of bulk samples was correctly classified by CHAID. As for the packaged products, the most accurate model for classifying anchovies of the validation set was found to be QUEST (96.0% accuracy) (Table 2). Based on the accuracy outcomes obtained during the validation phase, CHAID and QUEST were selected as the most appropriate algorithms to classify the origin of bulk and packaged anchovy samples, respectively. A short summary of the outputs obtained by the application of the other algorithms is however reported in Supplementary Material (Supplementary Tables S6 and S7).

By looking at the ranks of each predictive element selected by the four models (Figure 3) it is possible to highlight that C5.0 and CHAID models extracted a lower number of attributes compared to CART and QUEST models. Li, B, P, K, As, Sr, Zr, Pd, Cd, Cs, and Ba were shared as predictors within the two anchovy datasets. By contrast, Sb and Pb were influent only for bulk products (Figure 3A), while Ni e U were extracted only for

packaged products. Interestingly, As emerged as the variable showing the highest impact for all the classification models of bulk anchovies (Figure 3A) and for QUEST and CHAID models of packaged anchovies (Figure 3B). B and Cs were instead found to be the most important attributes in the CART and C5.0 models of packaged anchovies, respectively (Figure 3B).

Arsenic contamination of seawaters can be related to anthropogenic pollutant activities as well as to the natural geological characteristics of the area (Garelick, Jones, Dybowska, & Valsami-Jones, 2008). As an example, As (together with Cr, Cu, Hg, Mn, Ni, Pb, Se, and V) concentrations were reported to be higher in seawater where volcanic activities exist such as the Mediterranean Basin (Juncos et al., 2016; Zkeri, Aloupi, & Gaganis, 2018).

Table 2 Classification results for decision tree and neural network learning methods for training and validation data.

Product	Dataset	Model predictive accuracy (%)			
		C5.0	CHAID	CART	QUEST
Semi-finished anchovies	Training	99.4	98.3	94.4	89.9
	Validation	96.6	98.3	94.8	93.1
Finished anchovies	Training	96.0	94.6	91.3	95.2
	Validation	91.4	94.2	89.7	96.0

Moreover, the uptake of As by fish is influenced by several natural factors including water temperature and salinity, cooccurrence of phosphate, and seasonal differences of the distribution of the inorganic and organic forms of As in the aquatic environments (Ferrante et al., 2019). Regardless the natural or anthropogenic nature of As, releasing sources, the reduced exchange of water in the Mediterranean Basin and, especially in the Adriatic Sea, can facilitate the accumulation of As in the environment (Ferrante et al., 2019). This can justify the higher amounts of As in anchovy from the Mediterranean Sea compared to Cantabrian (Atlantic Ocean) anchovy which were reported in the present work (see Figure 1A, Figure 1B). Moreover, in pelagic fish species from the Adriatic Sea higher amounts of As compared to

other sampling zones was previously reported (Storelli & Marcotrigiano, 2004).

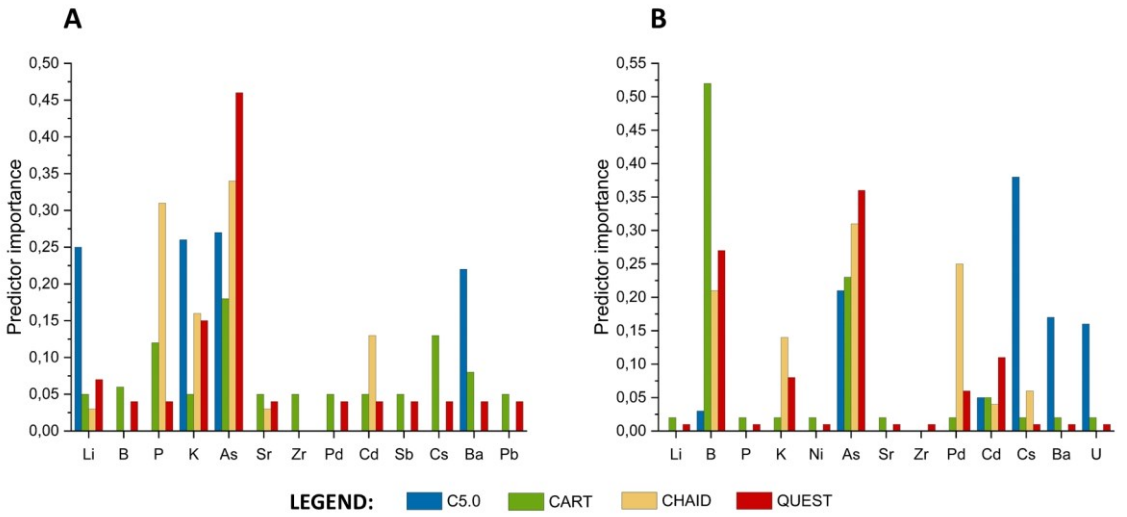


Figure 3. Comparison of the most important elemental predictors in C5.0, CART, CHAID, and QUEST models for semi-finished anchovies (A) and finished anchovies (B). Values are scaled from 0 (no influence) to 1 (maximum influence).

3.2.1. Decision tree by CHAID algorithm for origin authenticity of semi-finished anchovies

In accordance with the optimal accuracy results achieved in training and validation, CHAID model was found to be characterised by optimal accuracy (94.1%), sensitivity (95.6%), and specificity (97.4%) values also when used to classify unlabelled samples of the bulk anchovy test set (Table 3). Therefore, the method used can be effectively considered powerful enough when the analytical goal is the identification of Tunisian, Cantabrian, and Croatian bulk anchovy products origins.

The architecture of the decision tree obtained is illustrated in Figure 4. As it can be observed, the tree was a four-level structure, with a total of 19 decision nodes and 12 classification rules created by using 6 elements only. The decision rules generated from the root node were based on 5 concentration ranges of As, which was confirmed to be the most influent

element for first sample discrimination by CHAID (see Figure 3A). CHAID-decision trees are generally more complex than those generated by other technique since it relies on a multiway splitting principle, but the higher degree of segmentation can help reducing the tree depth and speed up the classification of samples.

Concentrations of As ≤ 3.38 mg kg^{-1} classified Cantabrian anchovies just at level 1 with 100% probability. In general, decreasing As concentrations (from 8.16 mg kg^{-1} downwards) together with increasing concentrations of K (from 4084 mg kg^{-1} upwards) and P (from 5078 mg kg^{-1} upwards) were associated to the highest probability of identifying Cantabrian samples. Tunisian samples were better classified by descending As concentration ranges coupled with higher Li (> 0.16 mg kg^{-1}), Sr (>29.75 mg kg^{-1}), or Cd amounts (>0.07 mg kg^{-1}). Finally, when P, Li, Cd got lower the occurrence of Croatian samples become more probable.

Table 3. Summary of performance parameters of decision tree models for classification of anchovy products of the testing datasets.

Performance index	Semi-finished anchovies				Finished anchovies			
	C5.0	CHAID	CART	QUEST	C5.0	CHAID	CART	QUEST
Accuracy (%)	91.2	94.1	91.2	94.1	90.5	91.9	96.8	97.7
Sensitivity (%)	91.4	95.6	89.4	93.6	92.6	92.3	96.7	97.6
Specificity (%)	95.3	97.4	95.1	96.7	95.4	95.5	98.5	98.9
Precision (%)	90.7	93.3	92.6	95.0	92.2	92.7	96.7	97.9
F-score (%)	90.9	93.9	90.2	94.1	91.4	91.8	96.5	97.7

3.2.2. Decision tree by QUEST algorithm for origin authenticity of finished anchovies

The QUEST-based decision trees applied to packaged anchovy was composed by 13 decisions nodes stratified into 5 levels. B was selected as the first binary splitting variable. The outcomes related to predictor importance reported in Figure 3B (according to which B had the highest influence in prediction) were confirmed by analysing the splitting variables used to generate the QUEST decision tree, where B was just selected as the first binary splitting variable (Figure 5). In particular, B value higher than 5.13

mg kg^{-1} generated a leaf (final) node with 91.3% probability of predicting samples originating from Tunisia. Globally, 100% probability of correctly identifying Croatian samples was reached at the last tree-level using the classification rule based on B, As, and Cd or the classification rule based on B, As, K, and Pd. With increased B (>5.13 mg kg^{-1}), As (>7.14 mg kg^{-1}), and K (>5669 mg kg-1) concentrations, also the likelihood of recognizing Cantabrian anchovies increased. The set of decision rules established by QUEST further proved to be reliable and effective for the classification of packaged anchovy origin, owing to the good ability in predicting the unknown origin of samples of the test set. Compared to other decision tree algorithm, QUEST ranked first in terms of accuracy (97.7%), sensitivity (97.6%), specificity (98.9%), and precision (97.97%) (Table 3).

Even though transformed anchovy implicitly represents a complex processed foodstuff, it is important to stress that using decision trees may be the quickest and the most intelligible way to solve problems related to classification of foodstuffs.

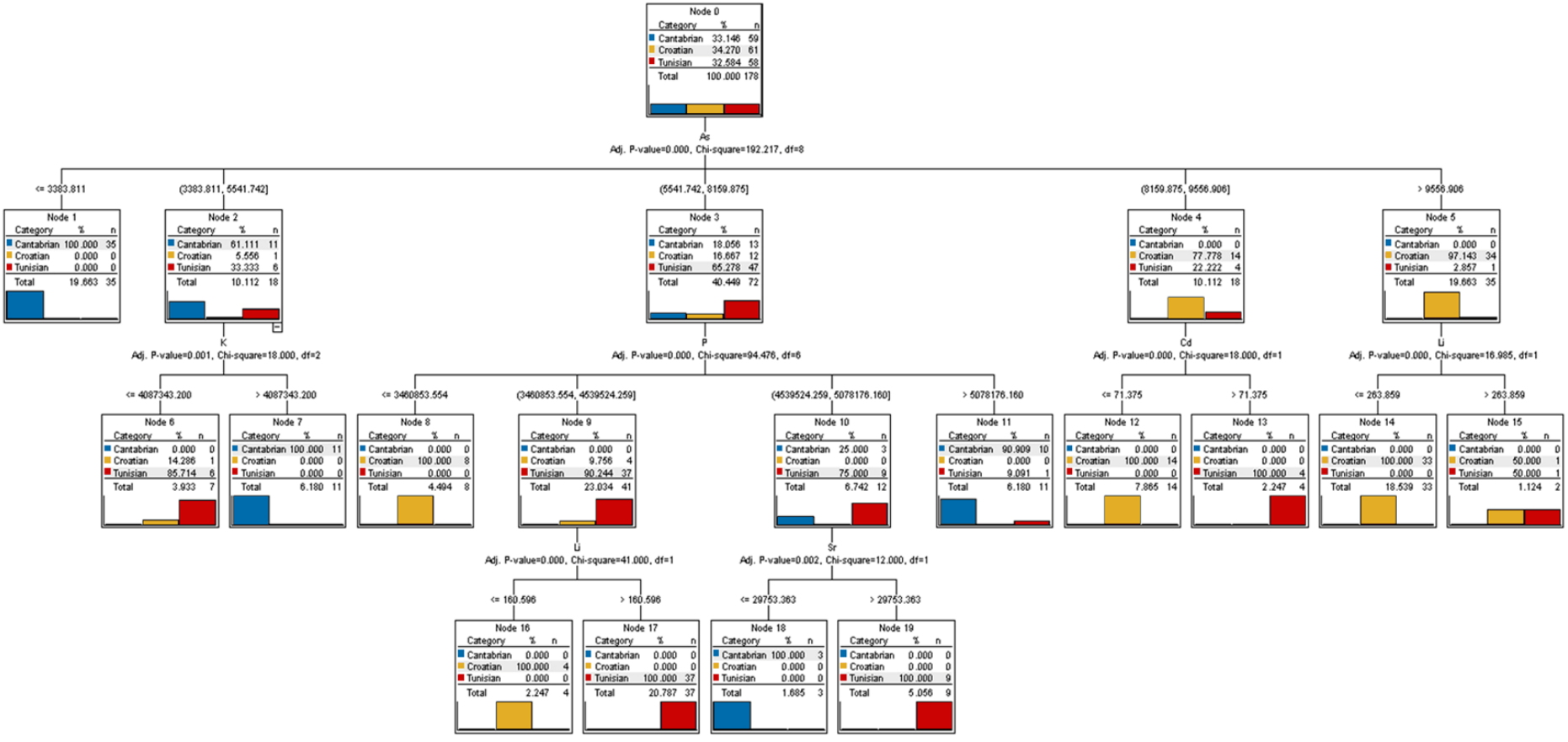


Figure 4. Decision classification tree resulting from the application of the CHAID algorithm for the classification of semi-finished anchovies using the element profile (concentrations are reported in μg kg^{-1}).

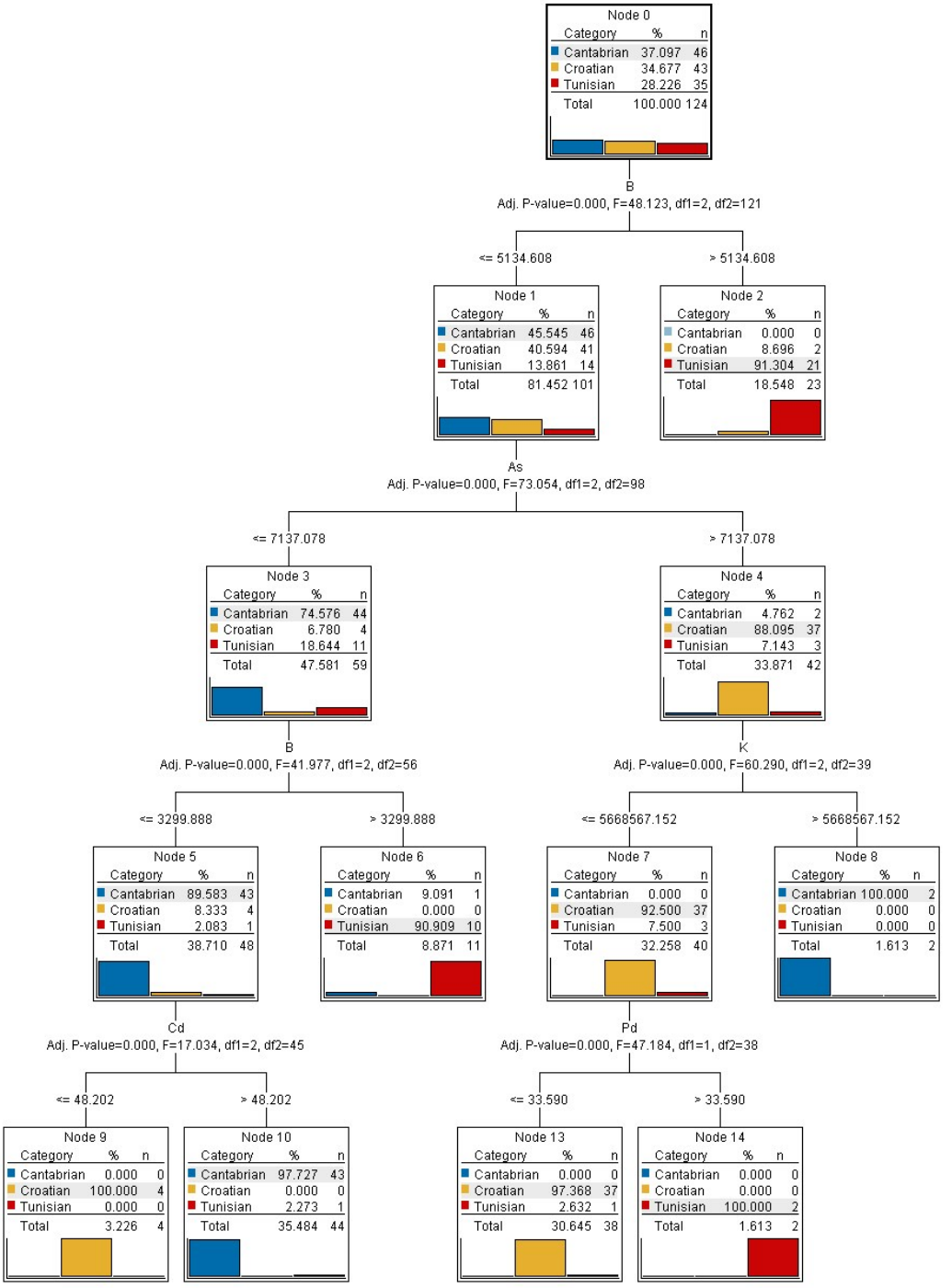


Figure 5. Decision classification tree resulting from the application of the QUEST algorithm for the classification of semi-finished anchovies using the element profile (concentrations are reported in $\mu g\ kg^{-1}$).

4. Conclusions

In this work, data mining techniques were applied to transformed anchovy products, in order to verify whether the origin of fish could be identified through the elemental patterns measured by ICP-MS and direct mercury analysis. Different machine learning algorithms relying on the principle of decision trees were applied to data and classification rules to distinguish anchovy fish of Cantabrian Sea from Tunisian and Croatian anchovies were created.

Firstly, differences of elemental composition between anchovies at two stages of the production chain were investigate to verify whether misleading elemental inclusion from the manufacturing environment was introduced. After having excluded problematic elements based on literature review and direct comparison of bulk and packaged anchovy profile and after having explored the effective presence of fish clusters related to origin, C5.0, CART, CHAID and QUEST decision trees were trained. This way, the selection of the most important variables and the identification of cut-off limits for each element concentration to describe a specific group of samples were performed in tandem.

The results obtained showed that the concentrations of 6 elements only (As, K, P, Li, Cd, and Sr) are required to identify through the use of the CHAID algorithm the origin of anchovy fish under the form of bulk product. In particular, arsenic was found to be the first sorting element, whose contribution to geographical origin differentiation was remarkably reflected in the ability of the decision tree to identify the unknown label of bulk fish with accuracy, specificity, sensitivity, and precision values above 93% on average.

The origin of the packaged anchovy product for sale, was better recognized by the set of classification rules generated by the QUEST algorithm. In this case, 5 elements were sufficient to achieve accuracy, sensitivity, specificity, and precision outcomes higher than 96%. In this case, the splitting of samples into groups was driven by the predictive influence of B, followed by As, K, Cd, and Pd.

In view of the above results, decision tree-based methods applied to elemental profiles of fishery products after industrial processing might be postulated as an immediate, practical and easy-to-handy procedure to figure out how the elemental composition can help in solving many actual challenges related to fish authenticity and commercial fraud. Moreover, the cost-effectiveness of the methodology, reached by doing away with the irrelevant elements, may finally disconnect this kind of applications from the scientifical research and lead to the application in the primary and secondary production sectors.

Future research including anchovy fish obtained from different countries and production systems is however desirable to clarify the involvement of multiple environmental factors on the stability over space and time of element profile of processed fish products.

Acknowledgements

The research was carried out within the project "Development of rapid tools for anchovies authentication" funded by Rizzoli Emanuelli S.p.a. (Parma, Italy). The authors also gratefully acknowledge the University of Pardubice for the financial support (project no. SGS_2020_002).

SUPPLEMENTARY MATERIALS

Supplementary Table S1. Method Detection Limits (MDL)a and Method Limits of Quantification (MLOQ)a (μg kg-1) and normalized calibration slopes (NCS) (1/μg L-1) of Agilent 7900 Q-ICP-MS for analysis of different elements using Rh as internal standard

Analyte	Cell mode	NCS	MDL	MLOQ	Analyte	NCS	Cell mode	MDL	MLOQ
$^{7}Li^{+}$	No gas	5.2×10^{-3}	0.59	1.98	$^{118}Sn^{+}$	1.73×10^{-2}	No gas	0.49	1.6
$^{11}B^{+}$	No gas	1.4×10^{-3}	5.7	19	$^{121}Sb^{+}$	2.3×10^{-2}	No gas	0.15	0.49
$^{23}Na^{+}$	He	1.2×10^{-3}	3500	11667	$^{133}Cs^{+}$	6.4×10^{-2}	No gas	0.015	0.051
$^{24}Mg^{+}$	No gas	1.3×10^{-2}	15	51	$^{138}Ba^{+}$	5.6×10^{-2}	No gas	0.55	1.83
$^{27}Al^{+}$	He	1.1×10^{-4}	8.6	29	$^{139}La^{+}$	6.4×10^{-2}	No gas	0.013	0.043
$^{31}P^{+}$	HE He	3.4×10^{-5}	161	536	$^{140}Ce^{+}$	6.4×10^{-2}	No gas	1.50	5.0
$^{39}K^{+}$	He	4.0×10^{-4}	735	2448	$^{141}Pr^{+}$	7.8×10^{-2}	No gas	0.010	0.033
$^{44}Ca^{+}$	He	1.9×10^{-5}	1291	4304	$^{146}Nd^{+}$	1.2×10^{-2}	No gas	0.07	0.22
$^{51}V^{+}$	He	8.1×10^{-3}	0.06	0.21	$^{147}Sm^{+}$	1.1×10^{-2}	No gas	0.27	0.88
$^{52}Cr^{+}$	He	1.1×10^{-2}	0.97	3.2	$^{153}Eu^{+}$	4.0×10^{-2}	No gas	0.019	0.063
$^{55}Mn^{+}$	He	4.3×10^{-3}	2.4	8.1	$^{157}Gd^{+}$	1.8×10^{-2}	No gas	0.046	0.153
$^{56}Fe^{+}$	He	7.7×10^{-3}	6.5	22	$^{159}Tb^{+}$	8.1×10^{-2}	No gas	0.0095	0.032
$^{59}Co^{+}$	He	2.1×10^{-2}	0.12	0.41	$^{163}Dy^{+}$	1.9×10^{-2}	No gas	0.070	0.23
$^{60}Ni^{+}$	He	5.7×10^{-3}	6.3	21	$^{165}Ho^{+}$	7.7×10^{-2}	No gas	0.054	0.18
$^{63}Cu^{+}$	He	1.7×10^{-2}	2.0	6.7	$^{166}Er^{+}$	2.6×10^{-2}	No gas	0.031	0.102
$^{66}Zn^{+}$	No gas	5.8×10^{-3}	159	528	$^{172}Yb^{+}$	1.7×10^{-2}	No gas	0.047	0.157
$^{75}As^{+}$	HE He	2.5×10^{-3}	0.49	1.6	$^{175}Lu^{+}$	7.2×10^{-3}	No gas	0.011	0.035
$^{78}Se^{+}$	HE He	3.7×10^{-4}	1.6	5.5	$^{178}Hf^{+}$	2.1×10^{-3}	No gas	0.0023	0.0075
$^{85}Rb^{+}$	No gas	4.3×10^{-2}	0.09	0.31	$^{185}Re^{+}$	2.3×10^{-3}	No gas	0.12	0.39
$^{88}Sr^{+}$	No gas	5.6×10^{-2}	0.19	0.62	$^{195}Pt^{+}$	1.6×10^{-3}	No gas	0.18	0.59
$^{89}Y^{+}$	No gas	6.8×10^{-2}	0.03	0.10	$^{205}Tl^{+}$	4.1×10^{-2}	No gas	0.027	0.088
$^{90}Zr^{+}$	No gas	3.5×10^{-2}	0.11	0.35	Pb [b]	5.3×10^{-2}	No gas	0.12	0.40
$^{95}Mo^{+}$	No gas	1.0×10^{-2}	0.58	1.93	$^{209}Bi^{+}$	4.4×10^{-2}	No gas	0.033	0.11
$^{101}Ru^{+}$	No gas	1.2×10^{-2}	0.07	0.22	$^{232}Th^{+}$	4.7×10^{-2}	No gas	0.033	0.11
$^{105}Pd^{+}$	No gas	1.34×10^{-2}	0.14	0.45	$^{238}U^{+}$	4.8×10^{-2}	No gas	0.0005	0.0017
$^{111}Cd^{+}$	No gas	6.0×10^{-3}	0.0038	0.013	Hg [c]	2.8×10^{-2}		0.2	0.7

[a] Values were calculated assuming a sample mass of 0.100 g.
[b] Pb is measured as the sum of the three most abundant isotopes, $^{206}Pb^{+}$, $^{207}Pb^{+}$ and $^{208}Pb^{+}$.
[c] Values were evaluated for direct analysis of Hg by single purpose atomic absorption spectrometer AMA 254.

Supplementary Table S2. Results of multi-elemental analysis applied to semi-finished salted anchovies (3 replicates for 30 samples in each group) according to geographical origin.

	Cantabria (n = 30*3)				Tunisia (n = 30*3)				Croatia (n = 30*3)			
Element	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**
Li	0.17 ± 0.018	0.16[a]	0.099	0.35	0.27 ± 0.022	0.26[b]	0.19	0.41	0.16 ± 0.011	0.16[a]	0.094	0.22
B	2.3 ± 0.35	2.20[a]	1.19	6.43	4.7 ± 0.38	4.69[b]	2.88	6.63	3.2 ± 0.39	2.80[c]	1.65	5.82
Na	156119 ± 16113	149810[a]	93121	307227	167082 ± 7341	168095[b]	122563	218388	141282 ± 8783	141666[a]	93927	189236
Mg	1150 ± 131	1069[a]	639	2425	1586 ± 75	1614[b]	1173	2248	986 ± 456	961[a]	556	1315
Al	2.7 ± 0.50	2.24[a]	1.31	6.59	4.8 ± 0.52	4.69[b]	2.82	9.70	5.9 ± 0.85	5.55[b]	3.18	14.7
P	5461 ± 507	5267[a]	3327	9852	4239 ± 173	4306[b]	3359	5106	4170 ± 258	4177[b]	2770	5520
K	6136 ± 659	5051[a]	3345	11778	3977 ± 209	3686[b]	2664	4762	4394 ± 474	4467[b]	2986	6214
Ca	3251 ± 343	3108[a]	1798	6301	4120 ± 186	4092[b]	3272	5375	3026 ± 249	3161[a]	1890	4349
V	0.035 ± 0.0041	0.031[a]	0.021	0.071	0.10 ± 0.0085	0.10[b]	0.071	0.16	0.078 ± 0.0094	0.073[c]	0.041	0.18
Cr	0.17 ± 0.072	0.11[a]	0.060	1.04	10.8 ± 0.018	0.10[a]	0.036	0.24	0.48 ± 0.53	0.11[a]	0.030	7.59
Fe	44 ± 4.2	42.0[a]	28.3	87.5	46 ± 2.1	46.3[ab]	31.6	55.8	49 ± 4.9	46.9[b]	31.8	101
Mn	1.8 ± 0.20	1.70[a]	0.99	3.66	1.60 ± 0.076	1.57[a]	1.20	2.13	1.7 ± 0.15	1.68[a]	1.00	2.70
Ni	1 ± 1.2	0.44[a]	0.41	17.2	0.36 ± 0.14	0.22[b]	0.096	1.71	0.20 ± 0.031	0.17[b]	0.76	0.44
Co	0.041 ± 0.0035	0.038[a]	0.024	0.071	0.033 ± 0.0011	0.033[b]	0.027	0.044	0.037 ± 0.0026	0.037[a]	0.024	0.057
Cu	3.9 ± 0.45	3.68[a]	2.43	9.30	3.9 ± 0.18	3.90[a]	2.97	5.57	3.9 ± 0.26	4.02[a]	2.55	5.19
Zn	83 ± 7.2	81.2[ab]	50.7	156	82 ± 3.4	82.5[a]	66.0	114	75 ± 4.6	73.3[b]	55.4	105
As	3.8 ± 0.49	3.01[a]	2.46	6.68	6.6 ± 0.35	6.53[b]	5.06	8.20	9.3 ± 0.59	9.28[c]	6.27	13.1
Se	1.1 ± 0.12	1.07[a]	0.61	2.34	1.01 ± 0.037	1.05[a]	0.81	1.33	1.03 ± 0.059	1.01[a]	0.61	1.33
Rb	1.03 ± 0.11	1.02[a]	0.57	2.01	0.61 ± 0.023	0.61[b]	0.48	0.76	0.73 ± 0.049	0.72[c]	0.50	1.02
Sr	34 ± 3.6	32.0[a]	18.8	67.9	35 ± 1.8	35.0[a]	25.9	44.5	22 ± 2.4	20.4[b]	13.0	38.4

(continued)

Supplementary Table S2. *(continued)*

	Cantabria (n = 30*3)				Tunisia (n = 30*3)				Croatia (n = 30*3)			
Element	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**
Y	0.004 ± 0.0012	0.024[a]	0.001	0.013	0.002 ± 0.00067	0.0017[a]	0.00066	0.0093	0.0071 ± 0.0020	0.0057[b]	0.0011	0.024
Zr	0.077 ± 0.0049	0.031[a]	0.0065	0.67	0.017 ± 0.012	0.013[b]	0.0048	0.086	0.019 ± 0.0032	0.016[b]	0.0079	0.043
Mo	0.047 ± 0.0073	0.043[a]	0.024	0.12	0.31 ± 0.0017	0.030[b]	0.025	0.041	0.043 ± 0.024	0.027[b]	0.015	0.37
Ru*	0.4 ± 0.14	0.27[a]	0.11	1.75	0.6 ± 0.10	0.61[b]	0.27	1.38	0.30 ± 0.065	0.25[a]	0.13	0.75
Pd	0.06 ± 0.032	0.037[a]	0.021	0.44	0.030 ±0.0018	0.031[b]	0.020	0.041	0.025 ± 0.0040	0.023[b]	0.011	0.063
Cd	0.09 ± 0.012	0.089[a]	0.052	0.22	0.084 ± 0.0046	0.085[a]	0.060	0.11	0.057 ± 0.0051	0.056[b]	0.026	0.088
Sn	0.49 ± 0.039	0.12[ab]	0.031	5.10	0.45 ± 0.25	0.21[a]	0.061	3.43	0.30 ± 0.15	0.11[b]	0.031	1.50
Sb	0.008 ± 0.0046	0.0042[a]	0.0023	0.060	0.004 ± 0.0010	0.0042[ab]	0.0023	0.0060	0.003 ± 0.0010	0.0032[b]	0.0014	0.063
Cs	0.022 ± 0.0025	0.021[a]	0.013	0.048	0.012 ± 0.0016	0.011[b]	0.0083	0.027	0.014 ± 0.0010	0.014[b]	0.0083	0.027
Ba	0.51 ± 0.18	0.42[a]	0.22	2.87	0.57 ± 0.066	0.52[a]	0.39	1.21	1.14 ± 0.13	1.12[b]	0.49	176
La*	53 ± 51	13.0[a]	2.83	658	4 ± 1.2	3.40[b]	1.34	15.3	8 ± 5.0	4.07[b]	1.61	75.5
Ce*	95 ± 93	19.9[a]	4.46	1070	7 ± 2.0	4.93[b]	1.58	27.2	13 ± 8.6	6.34[b]	2.79	130
Pr*	0.23 ± 0.044	0.21[a]	0.094	0.57	0.19 ± 0.027	0.17[a]	0.075	0.36	0.36 ± 0.13	0.25[b]	0.12	2.02
Nd*	0.9 ± 0.15	0.83[a]	0.52	2.11	0.69 ± 0.11	0.66[a]	0.29	1.59	1.5 ± 0.57	1.06[b]	0.50	9.10
Sm*	0.19 ± 0.033	0.17[a]	0.074	0.38	0.18 ± 0.024	0.18[a]	0.063	0.28	0.29 ± 0.11	0.20[a]	0.081	1.76
Eu*	0.21 ± 0.045	0.18[ab]	0.11	0.76	0.20 ± 0.028	0.18[a]	0.12	0.47	0.33 ± 0.036	0.32[b]	0.15	0.50
Gd*	0.6 ± 0.34	0.29[a]	0.11	4.28	0.16 ± 0.033	0.15[b]	0.037	0.45	0.28 ± 0.070	0.23[a]	0.064	1.01
Tb*	0.05 ± 0.011	0.041[a]	0.017	0.14	0.040 ± 0.0092	0.038[a]	0.0015	0.15	0.041 ± 0.0092	0.035[a]	0.0083	0.14
Dy*	0.25 ± 0.082	0.18[a]	0.061	1.00	0.14 ± 0.022	0.13[b]	0.067	0.36	0.23 ± 0.038	0.20[a]	0.10	0.49
Ho*	0.07 ± 0.028	0.047[a]	0.0050	0.34	0.041 ± 0.0046	0.040[a]	0.023	0.78	0.049 ± 0.0089	0.045[a]	0.013	0.11
Er*	0.3 ± 0.13	0.13[a]	0.059	1.62	0.10 ± 0.015	0.090[b]	0.045	0.24	0.15 ± 0.027	0.14[a]	0.037	0.36
Yb*	0.4 ± 0.24	0.16[a]	0.071	2.23	0.10 ± 0.015	0.086[b]	0.037	0.24	0.09 ± 0.012	0.43[b]	0.071	0.24
Lu*	0.08 ± 0.060	0.029[a]	0.011	0.046	0.026 ± 0.0033	0.044[a]	0.014	0.046	0.035 ± 0.0054	0.053[a]	0.031	0.072
Hf*	4.3 ± 1.84	2.19[a]	0.39	19.1	4 ± 1.2	2.26[a]	0.39	39.9	2.4 ± 0.58	2.02[a]	0.47	6.51
Re*	0.37 ± 0.033	0.36[a]	0.23	0.71	0.67 ± 0.058	0.61[b]	0.27	1.66	0.60 ± 0.085	0.53[b]	0.28	1.30

(continued)

Supplementary Table S2. *(continued)*

	Cantabria (n = 30*3)				Tunisia (n = 30*3)				Croatia (n = 30*3)			
Element	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**
Pb	0.4 ± 0.25	0.084[a]	0.030	3.22	0.097 ± 0.044	0.062[a]	0.035	0.67	0.11 ± 0.034	0.069[a]	0.028	0.46
Pt*	0.6 ± 0.29	0.31[a]	0.14	3.65	0.45 ± 0.12	0.37[a]	0.022	1.24	0.19 ± 0.053	0.14[b]	0.053	0.65
Tl*	3.9 ± 0.56	3.76[a]	2.01	9.71	1.60 ± 0.63	1.10[b]	0.31	8.32	1.35 ± 0.096	1.40[b]	0.84	1.88
Pb	0.4 ± 0.25	0.084[a]	0.030	3.22	0.097 ± 0.044	0.062[a]	0.035	0.67	0.11 ± 0.034	0.069[a]	0.028	0.46
Bi*	2.7 ± 0.91	1.62[a]	0.62	9.45	2 ± 1.2	1.19[a]	0.27	12.6	3 ± 1.0	2.04[a]	0.41	9.56
Th*	1.81 ± 0.66	0.91[a]	0.29	23.43	1.8 ± 0.70	1.16[a]	0.30	8.78	1.11 ± 0.23	0.88[a]	0.36	2.76
U*	3.6 ± 0.81	3.06[a]	1.91	14.03	3.9 ± 0.34	4.02[a]	2.23	6.08	2.48 ± 0.28	2.31[b]	1.34	4.90
Hg[#]	0.18 ± 0.012	0.17[a]	0.16	0.26	0.28 ± 0.019	0.26[b]	0.20	0.40	0.43 ± 0.014	0.44[c]	0.35	0.49

Concentrations are reported as mg kg^{-1} (d.m.). ME: margin of error at 95% confidence level. *: concentrations reported as μg kg^{-1}. Min–Max: minimum and maximum values. Hg[#]: determined by direct mercury analyzer AMA254. Different superscript letters (a–c) in the same row indicate statistically significant differences among median values ($p \leq 0.05$).

Supplementary Table S3. Results of multi-elemental analysis applied to finished salted anchovies 3 replicates for 30 samples in each group) according to geographical origin.

	Cantabria (n = 30*3)				Tunisia (n = 30*3)				Croatia (n = 30*3)			
Element	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**
Li	0.18 ± 0.012	0.17[a]	0.12	0.27	0.31 ± 0.033	0.32[b]	0.17	0.49	0.20 ± 0.017	0.19[c]	0.12	0.41
B	2.3 ± 0.21	2.18[a]	1.23	3.88	6.3 ± 0.74	6.25[b]	3.35	11.2	3.3 ± 0.42	3.20[c]	1.74	7.31
Na	160786 ± 8084	160218[a]	97975	225942	194659 ± 17158	176992[b]	139921	284778	161742 ± 8492	162517[a]	96381	226690
Mg	1155 ± 75	1172[a]	631	1945	1829 ± 187	1724[b]	1137	2842	1448 ± 90	1486[b]	926	2226
Al	6 ± 4.8	3.1[a]	1.31	72.6	6 ± 1.0	5.21[b]	3.17	14.7	6.3 ± 0.90	5.80[b]	3.51	13.2
P	5085 ± 294	5074[a]	2763	6097	4315 ± 332	4163[b]	2980	6097	4165 ± 269	4176[b]	2538	6686
K	5663 ± 390	5651[a]	2717	7725	4146 ± 428	3792[b]	2606	6621	4216 ± 300	4203[b]	2443	7203
Ca	3637 ± 258	3582[a]	2026	6196	4548 ± 431	4184[b]	2667	6781	3559 ± 229	3459[a]	2532	5044
V	0.074 ± 0.0075	0.068[a]	0.062	0.13	0.11 ± 0.011	0.11[b]	0.54	0.16	0.04 ± 0.011	0.042[c]	0.022	0.91
Cr	0.4 ± 0.19	0.18[a]	0.088	2.22	2.52 ± 0.25	0.081[b]	0.050	3.92	0.23 ± 0.065	0.17[a]	0.061	0.69
Fe	48 ± 3.5	46.8[a]	26.1	77.7	54 ± 5.2	50.1[a]	33.1	81.1	54 ± 2.6	44.9[a]	32.1	64.3
Mn	1.9 ± 0.27	1.84[a]	0.99	2.86	1.8 ± 0.18	1.73[a]	0.98	2.77	1.9 ± 0.13	1.9[a]	1.26	2.95
Ni	1.8 ± 1.83	0.36[a]	0.15	24.7	0.28 ± 0.079	0.19[b]	0.11	0.94	0.4 ± 0.16	0.21[b]	0.12	1.91
Co	0.037 ± 0.0021	0.037[a]	0.021	0.055	0.041 ± 0.0039	0.038[a]	0.028	0.062	0.036 ± 0.0031	0.035[a]	0.021	0.066
Cu	3.8 ± 0.23	3.76[a]	2.53	5.77	4.6 ± 0.42	4.44[b]	3.08	6.70	4.6 ± 0.33	4.43[b]	2.50	6.92
Zn	83 ± 4.9	81.8[a]	47.8	129.7	85 ± 9.9	77.4[a]	49.8	142	86 ± 6.8	85.0[a]	45.7	149
As	4.2 ± 0.64	3.94[a]	2.29	9.29	7.7 ± 0.77	6.91[b]	4.81	12.4	8.9 ± 0.58	8.83[c]	5.34	14.5
Se	1.07 ± 0.066	1.07[a]	0.60	1.81	1.16 ± 0.099	1.07[a]	0.60	1.81	1.095 ± 0.075	1.08[a]	0.59	1.83
Rb	0.91 ± 0.050	0.917[a]	0.46	1.24	0.70 ± 0.075	0.61[b]	0.43	1.10	0.67 ± 0.056	0.68[b]	0.39	1.30
Sr	38 ± 2.4	36.6[a]	18.7	53.1	40 ± 4.2	36.2[a]	24.4	69.2	30 ± 1.9	28.7[b]	17.5	39.2
Y	0.004 ± 0.0033	0.0020[a]	0.00085	0.049	0.010 ± 0.0049	0.0052[b]	0.0010	0.053	0.0059 ± 0.001	0.0045[b]	0.00099	0.014
Zr	0.04 ± 0.012	0.25[a]	0.0086	0.15	0.03 ± 0.0059	0.018[ab]	0.0059	0.074	0.02 ± 0.033	0.016[b]	0.0099	0.043
Mo	0.049 ± 0.011	0.040[a]	0.026	0.18	0.038 ± 0.0081	0.031[b]	0.016	0.14	0.032 ± 0.0034	0.030[b]	0.018	0.069
Ru*	0.5 ± 0.11	0.36[a]	0.088	10.08	0.5 ± 0.12	0.36[a]	0.15	1.55	0.71± 0.084	0.77[b]	0.19	1.06
Pd	0.041 ± 0.0061	0.036[a]	0.026	0.11	0.043 ±0.0049	0.040[a]	0.021	0.075	0.024 ± 0.0021	0.023[b]	0.013	0.038
Cd	0.089 ± 0.0059	0.090[a]	0.0053	0.12	0.086 ± 0.0091	0.079[a]	0.056	0.15	0.061 ± 0.0074	0.062[b]	0.027	0.14
Sn	0.5 ± 0.53	0.13[a]	0.050	8.38	0.4 ± 0.21	0.27[b]	0.085	8.38	0.85 ± 0.86	0.19[ab]	0.0056	16.9
Sb	0.014 ± 0.0014	0.0049[a]	0.0030	0.23	0.0049 ± 0.0058	0.0048[ab]	0.0026	0.0094	0.008 ± 0.0066	0.0033[b]	0.0024	0.098

(continued)

Table S3 *(continued)*

	Cantabria (n = 30*3)				Tunisia (n = 30*3)				Croatia (n = 30*3)			
Element	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**	**Mean ± ME**	**Median**	**Min**	**Max**
Cs	0.020 ± 0.0026	0.019^{a}	0.0093	0.048	0.013 ± 0.0015	0.011^{b}	0.0071	0.048	0.014 ± 0.0018	0.014^{b}	0.0093	0.048
Ba	0.54 ± 0.095	0.48^{a}	0.26	1.62	1.18 ± 0.21	1.01^{b}	0.35	2.74	1.1 ± 0.12	0.97^{b}	0.70	2.24
La*	175 ± 165	6.83^{a}	2.61	4864	9 ± 4.7	5.71^{ab}	1.83	69.6	6 ± 1.5	4.1^{b}	1.59	16.7
Ce*	405 ± 791	10.0^{a}	3.97	1162	16 ± 9.2	9.13^{a}	2.72	136	10 ± 2.8	7.21^{a}	2.62	33.1
Pr*	0.5 ± 0.28	0.26^{a}	0.16	3.41	0.36 ± 0.086	0.28^{a}	0.13	1.27	1 ± 1.2	0.93^{a}	0.62	13.4
Nd*	4 ± 4.9	0.90^{a}	0.55	72.7	3.49 ± 0.34	1.17^{b}	0.51	4.97	3.03 ± 1.1	2.82^{ab}	1.33	12.0
Sm*	0.7 ± 0.91	0.21^{a}	0.10	13.6	0.32 ± 0.063	0.27^{a}	0.13	0.93	0.4 ± 0.19	0.22^{a}	0.13	2.56
Eu*	0.3 ± 0.19	0.19^{a}	0.12	3.08	0.37 ± 0.054	0.33^{b}	0.14	0.67	0.32 ± 0.031	0.31^{b}	0.21	0.58
Gd*	2 ± 3.0	0.20^{a}	0.092	43.8	0.30 ± 0.068	0.28^{a}	0.094	1.01	0.26 ± 0.091	0.19^{a}	0.13	1.14
Tb*	0.1 ± 0.13	0.037^{a}	0.023	1.49	0.055 ± 0.0099	0.047^{a}	0.021	0.13	0.051 ± 0.011	0.045^{a}	0.018	0.15
Dy*	0.5 ± 0.55	0.16^{a}	0.097	8.27	0.27 ± 0.048	0.22^{b}	0.11	0.61	0.21 ± 0.043	0.18^{ab}	0.042	0.54
Ho*	0.10 ± 0.10	0.046^{a}	0.020	1.60	0.065 ± 0.011	0.062^{a}	0.023	0.51	0.061 ± 0.0096	0.056^{a}	0.018	0.14
Er*	0.37 ± 0.33	0.12^{a}	0.072	4.54	0.18 ± 0.032	0.17^{a}	0.072	0.43	0.15 ± 0.25	0.14^{a}	0.076	0.39
Yb*	0.31 ± 0.28	0.15^{a}	0.056	4.27	0.22 ± 0.043	0.20^{a}	0.077	0.56	0.4 ± 0.028	0.15^{a}	0.077	0.48
Lu*	0.06 ± 0.042	0.030^{ab}	0.0084	0.65	0.04 ± 0.0071	0.035^{a}	0.0016	0.10	0.03 ± 0.043	0.025^{b}	0.0056	0.061
Hf*	4 ± 3.1	2.33^{a}	0.67	47.3	4 ± 1.2	1.95^{a}	0.45	39.9	3 ± 1.1	2.08^{a}	0.48	14.6
Re*	0.54 ± 0.086	0.44^{a}	0.27	1.06	0.71 ± 0.12	0.65^{a}	0.27	1.66	0.87 ± 0.071	0.91^{b}	0.049	1.19
Pt*	0.5 ± 0.12	0.39^{a}	0.077	1.44	0.5 ± 0.34	0.13^{b}	0.033	3.62	0.5 ± 0.19	0.48^{a}	0.059	2.13
Tl*	3.49 ± 0.85	3.04^{a}	1.29	14.8	1.4 ± 0.51	1.74^{b}	0.65	7.69	1.9 ± 0.36	1.65^{b}	0.89	6.24
Pb	0.16 ± 0.051	0.12^{a}	0.047	0.73	0.14 ± 0.029	0.11^{a}	0.045	0.37	0.17 ± 0.13	0.071^{b}	0.036	1.94
Bi*	48 ± 90.6	1.50^{a}	0.39	1333	1.8 ± 0.095	0.68^{b}	0.29	11.1	1.22 ± 0.52	0.86^{b}	0.42	8.16
Th*	2 ± 2.1	0.87^{a}	0.031	23.5	1.3 ± 0.58	0.81^{a}	0.33	8.02	1.73 ± 0.63	1.20^{a}	0.37	6.99
U*	2.97 ± 0.17	2.96^{a}	2.29	4.22	4.47 ± 0.52	4.41^{b}	1.99	7.52	2.66 ± 0.22	2.62^{a}	1.79	4.60
Hg$^{\#}$	0.22 ± 0.026	0.22^{a}	0.073	0.37	0.31 ± 0.018	0.30^{b}	0.24	0.39	0.42 ± 0.014	0.42^{c}	0.33	0.49

Concentrations are reported as mg kg^{-1} (d.m.). ME: margin of error at 95% confidence level. *: concentrations reported as μg kg $^{-1}$. Min–Max: minimum and maximum values. Hg$^{\#}$: determined by direct mercury analyzer AMA254. Different superscript letters (a–c) in the same row indicate statistically significant differences among median values ($p \leq 0.05$).

Supplementary Table S4. Significantly different elements according to Mann-Whitney test (comparison $p \leq 0.05$) between semi-finished and finished anchovy products of the same origin.

Spain (Cantabria)		**Croatia (Mediterranean)**		**Tunisia (Mediterranean)**	
Element	**Significance (*p*-value)**	**Element**	**Significance (*p*-value)**	**Element**	**Significance (*p*-value)**
Na	0.0189	Na	< 0.0001	Na	< 0.0001
Al	0.0225	Mg	0.0005	Mg	< 0.0001
Ca	0.0009	Co	0.0005	Ca	0.0034
V	< 0.0001	Cu	0.0191	V	< 0.0001
Cr	0.0006	Y	< 0.0001	Cr	0.0358
Fe	0.0029	Zr	0.0039	Mn	0.0176
Rb	0.0039	Ru	0.0019	Ni	0.0125
Ru	0.0429	Pd	0.0001	Cu	0.0002
Cs	0.0176	Sb	0.0496	Zn	0.0125
La	0.0092	Ba	< 0.0001	Rb	0.0400
Ce	0.0023	La	0.0007	Sr	< 0.0001
Pr	0.0169	Ce	0.0005	Ru	< 0.0001
Gd	0.0430	Pr	< 0.0001	Sn	0.0430
Re	0.0004	Nd	< 0.0001	Ho	0.0319
Tl	0.0479	Gd	< 0.0001	Yb	< 0.0001
		Tb	0.0080	Lu	< 0.0001
		Dy	< 0.0001	Re	< 0.0001
		Ho	< 0.0001	Pt	< 0.0001
		Er	< 0.0001	Tl	0.0007
		Yb	< 0.0001		
		Pt	0.01		

Supplementary Table S5. Cophenetic correlation coefficient values resulting from application of HCA to bulk and packaged anchovies.

Distance type	Clustering method	Cophenetic Correlation Coefficients	
		Bulk anchovies	**Packaged anchovies**
Euclidean	Ward's Minimum Variance	0.9988	0.9836
	Single Linkage (Nearest Neighbor)	0.9221	0.9565
	Complete Linkage (Furthest Neighbor)	0.8170	0.8569
	Simple Average (Weighted Pair-Group)	0.9337	0.9655
	Group Average (Unweighted Pair-Group)	0.9757	0.9830
	Median (Weighted Pair-Group Centroid)	0.9851	0.9469
	Centroid (Unweighted Pair-Group Centroid)	0.9716	0.9777
Manhattan	Ward's Minimum Variance	0.6344	0.8505
	Single Linkage (Nearest Neighbor)	0.8940	0.9717
	Complete Linkage (Furthest Neighbor)	0.7351	0.9065
	Simple Average (Weighted Pair-Group)	0.8377	0.8495
	Group Average (Unweighted Pair-Group)	0.9686	0.9785
	Median (Weighted Pair-Group Centroid)	0.8969	0.9760
	Centroid (Unweighted Pair-Group Centroid)	0.9457	0.9796

Supplementary Table S6. Classification rules extracted by C5.0, CART, and QUEST decision trees to classify semi-finished anchovies (concentrations are expressed in μg kg^{-1}).

Algorithm	Rules	Origin	Percentage (%)
C5.0	As ≤ 8320 AND K ≤ 4994485 AND Li ≤ 163 AND Ba ≤ 498	Cantabrian	100
	As ≤ 8320 AND K ≤ 4994485 AND Li ≤ 163 AND Ba > 498	Croatian	100
	As ≤ 8320 AND K ≤ 4994485 AND Li > 163	Tunisian	100
	As ≤ 8320 AND K > 4994485	Cantabrian	100
	As > 8320 AND Li ≤ 260	Croatian	100
	As > 8320 AND Li > 260	Tunisian	66.7
CART	As ≤ 4709	Cantabrian	97.1
	As > 4709 AND Li ≤ 217 AND Ba ≤ 633 AND P ≤ 4721699 AND B ≤ 2745	Croatian	100
	As > 4709 AND Li ≤ 217 AND Ba ≤ 633 AND P ≤ 4721699 AND B > 2745	Tunisian	100
	As > 4709 AND Li ≤217 AND Ba ≤ 633AND P > 4721699	Cantabrian	100
	As > 4709 AND Li ≤ 217 AND Ba > 633	Croatian	96.9
	As > 4709 AND Li > 217 AND Cs ≤ 14.6	Tunisian	100
	As > 4709 AND Li > 217 AND Cs > 14.6 AND Pd ≤ 46.0	Croatian	100
	As > 4709 AND Li > 217 AND Cs > 14.6 AND Pd > 46.0	Cantabrian	50
QUEST	As ≤ 7752 AND As ≤ 5343 AND As ≤ 4917	Cantabrian	100
	As ≤ 7752 AND As ≤ 5343 AND As > 4917	Tunisian	66.7
	As ≤ 7752 AND As > 5343 AND Li ≤ 151	Croatian	87.5
	As ≤ 7752 AND As > 5343 AND Li > 150.517 AND K ≤ 5228787	Tunisian	100
	As ≤ 7752 AND As > 5343 AND Li > 150.517 AND K > 5228787	Cantabrian	100
	As > 7752	Croatian	85.4

Supplementary Table S7. Classification rules extracted by C5.0, CAR, and CHAID decision trees to classify finished anchovies (concentrations are expressed in μg kg^{-1}).

Algorithm	Rules	Origin	Percentage (%)
C5.0	B ≤ 3536 AND As ≤ 6694 AND Cd ≤ 46.9	Croatian	100
	B ≤ 3536 AND As ≤ 6694.200 AND Cd > 46.9	Cantabrian	100
	B ≤ 3536 AND As > 6694	Croatian	100
	B > 3536 AND Cs <= 12.1	Tunisian	100
	B > 3536 AND Cs > 12.1 AND U ≤ 4.80 AND Ba ≤ 830	Cantabrian	100
	B > 3536 AND Cs > 12.1 AND U ≤ 4.80 AND Ba > 830	Croatian	95
	B > 3536 AND Cs > 12.1 AND U > 4.80	Tunisian	93.3
CART	B ≤ 3990 AND As ≤ 6694 AND Cd ≤ 49.6	Croatian	71.4
	B ≤ 3990 AND As ≤ 6694 AND Cd > 49.6	Cantabrian	97.9
	B ≤ 3990 AND As > 6694	Croatian	96.9
	B > 3990 AND Cd ≤ 55.5	Croatian	60
	B > 3990 AND Cd > 55.5	Tunisian	94.4
CHAID	B ≤ 2129 AND Cd ≤ 48.0	Croatian	100
	B ≤ 2129 AND Cd > 48.0 and Cd ≤ 57.4	Cantabrian	80
	B ≤ 2129 AND Cd > 57.4	Cantabrian	100
	B > 2129 and B ≤ 2966 AND As ≤ 7485 AND K ≤ 3085761	Croatian	100
	B > 2129 and B ≤ 2966 AND As ≤ 7485 AND K > 3085761	Cantabrian	100
	B > 2129 and B ≤ 2966 AND As > 7485	Croatian	100
	B > 2966 and B ≤ 4076 AND K ≤ 3568	Tunisian	100
	B > 2966 and B ≤ 4076 AND K > 3568 and K ≤ 5419315	Croatian	100

(continued)

Table S6 *(continued)*

Algorithm	Rules	Origin	Percentage (%)
CHAID	B > 2966 and B ≤ 4076 AND K > 5419315	Cantabrian	83.3
	B > 4076 AND Pd ≤28.8 AND Cs ≤ 12.4	Tunisian	100
	B > 4076 AND Pd ≤ 28.8 AND Cs > 12.4	Croatian	100
	B > 4076 AND Pd > 28.8 AND Cs ≤ 21.8	Tunisian	100
	B > 4076 AND Pd > 28.8 AND Cs > 21.8	Tunisian	57.1

References

Bekhit, A. E. D. A., Morton, J. D., & Dawson, C. O. (2008). Effect of processing conditions on trace elements in fish roe from six commercial New Zealand fish species. *Journal of Agricultural and Food Chemistry*, *56*(12), 4846–4853. https://doi.org/10.1021/jf8005646

Besteiro, I., Rodríguez, C. J., Tilve-Jar, C., & Pascual López, C. (2000). Selection of attributes for the sensory evaluation of anchovies during the ripening process. *Journal of Sensory Studies*, *15*(1), 65–77. https://doi.org/10.1111/j.1745-459X.2000.tb00410.x

Carrera, M., & Gallardo, J. M. (2017). Determination of the Geographical Origin of All Commercial Hake Species by Stable Isotope Ratio (SIR) Analysis. *Journal of Agricultural and Food Chemistry*, *65*(5), 1070–1077. https://doi.org/10.1021/acs.jafc.6b04972

Cuadros-Rodríguez, L., Pérez-Castaño, E., & Ruiz-Samblás, C. (2016). Quality performance metrics in multivariate classification methods for qualitative analysis. *Trends in Analytical Chemistry*, *80*, 612–624. https://doi.org/10.1016/j.trac.2016.04.021

Czerner, M., Agustinelli, S. P., Guccione, S., & Yeannes, M. I. (2015). Effect of different preservation processes on chemical composition and fatty acid profile of anchovy (Engraulis anchoita). *International Journal of Food Sciences and Nutrition*, *66*(8), 887–894. https://doi.org/10.3109/09637486.2015.1110687

da Costa, N. L., Ximenez, J. P. B., Rodrigues, J. L., Barbosa, F., & Barbosa, R. (2020). Characterization of Cabernet Sauvignon wines from California: determination of origin based on ICP-MS analysis and machine learning techniques. *European Food Research and Technology*, *246*(6), 1193–1205. https://doi.org/10.1007/s00217-020-03480-5

Depeters, E. J., Puschner, B., Taylor, S. J., & Rodzen, J. A. (2013). Can fatty acid and mineral compositions of sturgeon eggs distinguish between farm-raised versus wild white (Acipenser transmontanus) sturgeon origins in California ? Preliminary report. *Forensic Science International*, *229*(1–3), 128–132. https://doi.org/10.1016/j.forsciint.2013.04.003

Drab, K., & Daszykowski, M. (2014). Clustering in analytical chemistry. *Journal of AOAC International*, *97*(1), 29–38. https://doi.org/10.5740/jaoacint.SGEDrab

Epova, E. N., Bérail, S., Zuliani, T., Malherbe, J., Sarthou, L., Valiente, M., & Donard, O. F. X. (2018). 87Sr/86Sr isotope ratio and multielemental signatures as indicators of origin of European cured hams: The role of salt. *Food Chemistry*, *246*, 313–322. https://doi.org/10.1016/j.foodchem.2017.10.143

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster analysis 5th ed. Wiley Series in Probability and Statistics. ISBN: 978-0-470-97844-3

Ferrante, M., Napoli, S., Grasso, A., Zuccarello, P., Cristaldi, A., & Copat, C. (2019). Systematic review of arsenic in fresh seafood from the Mediterranean Sea and European Atlantic coasts: A health risk assessment. *Food and Chemical Toxicology*, *126*, 322–331. https://doi.org/10.1016/j.fct.2019.01.010

Fragni, R., Trifirò, A., & Nucci, A. (2015). Towards the development of a multi-element analysis by ICP-oa-TOF-MS for tracing the geographical origin of processed tomato products. *Food Control*, *48*, 96–101. https://doi.org/10.1016/j.foodcont.2014.04.027

Franke, B. M., Haldimann, M., Gremaud, G., Bosset, J. O., Hadorn, R., & Kreuzer, M. (2008). Element signature analysis: Its validation as a tool for geographic authentication of the origin of dried beef and poultry meat. *European Food Research and Technology*, *227*(3), 701–708. https://doi.org/10.1007/s00217-007-0776-8

Garelick, H., Jones, H., Dybowska, A., & Valsami-Jones, E. (2008). Arsenic pollution sources. *Reviews of Environmental Contamination and Toxicology*, *197*(January 2017), 17–60. https://doi.org/10.1007/978-0-387-79284-2_2

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining : Concepts and Techniques* (Third Edit). Morgan Kaufmann Publishers. https://doi.org/10.1016/B978-0-12-381479-1.00001-0

Hernandez-Herrero, M. M., Roig-Sagués, A. X., López-Sabater, E. I., Rodriguez-Jerez, J. J., & Mora-Ventura, M. T. (1999). Total Volatile Basic Nitrogen and other Physico- chemical and Microbiological Characteristics as. *Journal of Food Science*, *64*(2), 344–347.

Ikem, A., & Egiebor, N. O. (2005). Assessment of trace elements in canned fishes (mackerel, tuna, salmon, sardines and herrings) marketed in Georgia and Alabama (United States of America). *Journal of Food Composition and Analysis*, *18*(8), 771–787. https://doi.org/10.1016/j.jfca.2004.11.002

Juncos, R., Arcagni, M., Rizzo, A., Campbell, L., Arribére, M., & Guevara, S. R. (2016). Natural origin arsenic in aquatic organisms from a deep oligotrophic lake under the influence of volcanic eruptions. *Chemosphere*, *144*, 2277–2289. https://doi.org/10.1016/j.chemosphere.2015.10.092

Laso, J., Margallo, M., Fullana, P., Bala, A., Gazulla, C., Irabien, A., & Aldaco, R. (2017). Introducing life cycle thinking to define best available techniques for products: Application to the anchovy canning industry. *Journal of Cleaner Production*, *155*, 139–150. https://doi.org/10.1016/j.jclepro.2016.08.040

Li, L., Han, C., Dong, S., & Boyd, C. E. (2019). Use of elemental profiling and isotopic signatures to differentiate Pacific white shrimp (Litopenaeus vannamei) from freshwater and seawater culture areas. *Food Control*, *95*, 249–256. https://doi.org/10.1016/j.foodcont.2018.08.015

Lo Feudo, G., Naccarato, A., Sindona, G., & Tagarelli, A. (2010). 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*, *58*(6), 3801–3807. https://doi.org/10.1021/jf903868j

Magdas, D. A., Feher, I., Cristea, G., Voica, C., Tabaran, A., Mihaiu, M., … Dan, D. S. (2019). Geographical origin and species differentiation of Transylvanian cheese. Comparative study of isotopic and elemental profiling vs . DNA results. *Food Chemistry*, *277*, 307–313. https://doi.org/10.1016/j.foodchem.2018.10.103

Moreno-Rojas, R., Cámara-Martos, F., Sánchez-Segarra, P. J., & Amaro-López, M. Á. (2012). Influence of manufacturing conditions and discrimination of Northern Spanish cheeses using multi-element analysis. *International Journal of Dairy Technology*, *65*(4), 594–602. https://doi.org/10.1111/j.1471-0307.2012.00853.x

Rodushkin, I., Bergman, T., Douglas, G., Engström, E., Sörlin, D., & Baxter, D. C. (2007). Authentication of Kalix (N.E. Sweden) vendace caviar using inductively coupled plasma-based analytical techniques: Evaluation of different approaches. *Analytica Chimica Acta*, *583*(2), 310–318. https://doi.org/10.1016/j.aca.2006.10.038

Rokach, L., & Maimon, O. Z. (2008). *Data mining with decision trees. Theory and applications* (2nd Editio). Singapore: World Scientific Publishing.

Saraçli, S., Doğan, N., & Doğan, İ. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. Journal of Inequalities and Applications, 2013(1), 203. https://doi.org/10.1186/1029-242X-2013-203

Smith, R. G., & Watts, C. A. (2009). Determination of the country of origin of farm-raised shrimp (family penaeide) using trace metal profiling and multivariate statistics. *Journal of Agricultural and Food Chemistry*, *57*(18), 8244–8249. https://doi.org/10.1021/jf901658f

Sokal, R. R., & Rohlf, F. J. (1962). The comparison of dendrograms by objective methods. Taxon, 33-40.

Storelli, M. M., & Marcotrigiano, G. O. (2004). Interspecific variation in total arsenic body concentrations in elasmobranch fish from the Mediterranean Sea. *Marine Pollution Bulletin*, *48*(11–12), 1145–1149. https://doi.org/10.1016/j.marpolbul.2004.03.005

Suhaj, M., & Kore, M. (2008). Study of some European cheeses geographical traceability by pattern recognition analysis of multielemental data. *European Food Research and Technology*, 1419–1427. https://doi.org/10.1007/s00217-008-0861-7

Turra, C., Dias de Lima, M., Fernandes, E. A. D. N., Bacchi, M. A., Barbosa, F., & Barbosa, R. (2017). Multielement determination in orange juice by ICP-MS

associated with data mining for the classification of organic samples. Information Processing in Agriculture, 4(3), 199–205. https://doi.org/10.1016/j.inpa.2017.05.004

Varrà, M. O., Ghidini, S., Zanardi, E., Badiani, A., & Ianieri, A. (2019). Authentication of European sea bass according to production method and geographical origin by light stable isotope ratio and rare earth elements analyses combined with chemometrics. *Italian Journal of Food Safety*, *8*(1). https://doi.org/10.4081/ijfs.2019.7872

Velasco, A., Aldrey, A., Pérez-Martín, R. I., & Sotelo, C. G. (2016). Assessment of the labelling accuracy of spanish semipreserved anchovies products by FINS (forensically informative nucleotide sequencing). *Heliyon*, *2*(6), e00124. https://doi.org/10.1016/j.heliyon.2016.e00124

Zkeri, E., Aloupi, M., & Gaganis, P. (2018). Seasonal and spatial variation of arsenic in groundwater in a rhyolithic volcanic area of Lesvos Island, Greece. *Environmental Monitoring and Assessment*, *190*(1). https://doi.org/10.1007/s10661-017-6395-3

GENERAL CONCLUSIONS AND FUTURE OUTLOOK

Despite the major steps taken in recent years to ensure better controls from a regulatory, scientific, and analytic point of view, opportunities for fraud in the fishery and aquaculture sector are growing significantly.
Hence, there is clearly a need for tightening up fish controls on the one hand and shifting towards a risk-based control of food authenticity on the other, through the implementation of up-to-date, efficient, and standardised food inspection tools.

Although NIR spectral fingerprint has become more widely recognised as an effective analytical technique, it still deals with a lot of lingering prejudices because of its empirical-based nature that led it to be regarded as a black box tool. In the *First Section* of this PhD Thesis, however, the possibility to unravel the hypothetically complex physical-chemical features embedded within the NIR spectra with the support of chemometrics was demonstrated and the full spectra were found to be highly promising for the quick resolution of authenticity problems along the fish production chain, both in unprocessed and processed samples.
Therefore, resuming earlier questions No. 1.*a*. and No. 1.*b*. posed in the *Aims and Objectives of the Thesis* section and based on the achieved results reported in *Chapters 2* and *3*, the following major conclusions can be drawn:

1.*a.* The 1100-2500 nm full scan NIR spectra of European sea bass can be successfully used to distinguish wild from farmed fish, the intensity of the rearing system of farmed fish, and the subareas of origin in the Mediterranean Sea by performing one single spectral analysis. Specifically, the application of orthogonal partial least square discriminant analysis (OPLS-DA) appears to be very useful since it allows to effectively separate non-relevant from relevant spectral information related to the specific classification purpose, thus improving classification accuracy of sea bass. The analysis of the NIR spectral fingerprint revealed that the variability of wavelengths associated to protein absorption was unusually significant towards sea bass discrimination based on production method/farming system. This was likely attributed to the higher influence of the muscular activity over the external feeding input of wild specimens compared to farmed

ones. Instead, NIR bands associated to lipid absorption had a major contribution to identifying the geographical provenance of fish, thus suggesting further deepening of the existing relationship between fatty acid composition and environmental conditions such as water temperature and salinity.

1.*b.* The ripening period the salted anchovy undergoes inevitably alters or modifies the original composition of fish tissue, thus masking potential markers related to the provenance. Although this, handling 1200-2500 nm NIR spectra of salt-ripened anchovies by means of OPLS-DA can represent a reliable strategy to monitor traceability of these products, whose raw fish originates from Spanish, Tunisian, Moroccan, and Croatian fishing areas. Indeed, the discrimination by provenance of both industrial intermediate and finished anchovy products can be attributed to organic fragmentation patterns generated during the manufacturing process and reflecting the original composition of the untransformed fish, which can be reliably monitored by NIR spectroscopy and chemometrics. An in-depth evaluation of the spectral signatures of fish indicated that a complex pattern of degradation compounds deriving from proteolysis (i.e. peptides, amino acids and low-molecular-weight nitrogen compounds and other decomposition fragments) is useful for the identification of anchovy from Morocco. On the other hand, the unsaturated lipid fraction seems to represent a distinctive compositional mark for anchovies from Tunisia, while anchovies from Spain and Croatia can be accurately identified owing to the equal contribution of protein and lipid compounds. Therefore, the proposed methodology is suggested to be implemented in the routine surveillance operations to monitor the integrity of the fish production chain.

The evidence that inorganic composition of fish and seafood can be exploited as an alternative strategy for simultaneously establishing the origin authenticity of fish and seafood and monitor safety issues related to the presence of metal contaminants as well, was provided by the studies reported in the *Second Section* of this PhD Thesis. In this context, recent

technological developments that have affected ICP-MS instrumentations and computer science in terms of polyatomic interferences management, linear dynamic range extension, and data handling, have represented a turning point for the accurate identification of elemental markers reflecting specific fish conditions.

Reiterating questions No. 2.*a*., No. 2.*b*., and No. 2.*c*. raised in the *Aims and Objectives of the Thesis* section, the following major findings obtained from experiments performed in Chapters 5,.6, and 7:

2.*a.* The combination of multivariate discriminant analysis with a fusion of data deriving from light stable isotope ratio and rare earth elements revealed that higher isotopic abundances of carbon and nitrogen of wild-caught fish compared to aquaculture fish can be used alone as discriminant markers for sea bass populations, thus remarking the important influence of the feeding sources and the trophic position of the fish on the final flesh composition.

Since the geographical origin is known to have a more limited impact on flesh composition compared to the production method, the isotopic signature did not appear sufficient to discriminate fish by provenance. Anyway, when coupled with lanthanum and holmium concentrations, whose presence in seawater are notoriously affected by natural and anthropogenic activities, the accuracy in discrimination increased notably to a point that fish from three different fishing areas in the Mediterranean Sea are reliably discernible and mislabelling potentially identifiable.

2.*b.* The protection and promotion of Chioggia cuttlefish, an Italian traditional product for which a quality mark has been recognised, was demonstrated to be feasible through the determination of the multielement profile by inductively coupled plasma mass spectrometry (ICP-MS) and the specific identification of the geographical inorganic imprint by multivariate discriminant data analysis and variable selection.

The elemental pattern linked to the geographical origin of cuttlefish appears to be determined by a combination of macro, trace, and ultra-

trace elements which are known to be absorbed by the animals from the surrounding environment. In particular, anthropogenic elements linked to the specific production area here strongly emerged as a key analytical determinant for cephalopods authenticity assessment. Of note, also concentrations of some heavy toxic metals such as cadmium and arsenic, although being so low as not to represent a potential health risk, were useful for discrimination, therefore their quantification for a dual safety-authenticity purpose must be encouraged.

2.*c.* Machine learning applied to elemental profiles of high-quality Cantabrian anchovy was verified as an optimal strategy to differentiate the products from competing products from other countries in the Mediterranean area.

The selection of a limited number of highly informative elements measured by ICP-MS (other than those verified to be misleading included into the product and originating from the manufacturing environment) together with the concomitant creation of classification rules by applying four decision tree-based machine learning algorithms ended up being an immediate, cheap, practical, and easy-to-handy procedure to figure out how the elemental composition can help in solving many actual challenges related fish quality promotion. Specifically, the results obtained suggest that only five out fifty-two and six out fifty-two originally measured elements are sufficient to effectively authenticate Cantabrian anchovies before and after packaging, respectively. Arsenic, boron, potassium, phosphorous cadmium demonstrated to be leading markers of origin, thereby emphasizing how their measurement can represent an effective and cost-effective strategy to be applied in the primary and secondary fish production sectors.

By way of conclusion, the results presented in this PhD Thesis clearly highlight the effective support in fish authenticity and safety assessment provided by analytical approaches based on fingerprinting of organic component by NIR spectroscopy on the one hand and on profiling of the

inorganic components of fish on the other. Although the proposed methodologies address problems of the same nature, it is worth noting how the field of application is supposed to be different, depending first and foremost on the fish sample characteristic and response to the technique and, secondly, the actual need of the users.

For instance, rapidity, non-destructive nature, ease of use and high-throughput measurement ability, make NIR spectroscopic non-targeted approach an ideal tool for quality control operations, especially in the context of daily routine and screening analysis in the food industry. Multielement-based strategies have been experienced for much longer than spectroscopic ones, and the greater sensitivity, specificity, and precision both in qualitative and quantitative determination make them particularly suitable for use within the official controls. ICP-MS equipment for multielement analysis is more expensive and less environmentally friendly than NIR spectroscopy instrumentations, but on the upside the lower number of samples required for proper calibration is relevant for expensive fish samples.

Nevertheless, these analytical strategies are still far from being effectively applied in the context of official and routine controls of foodstuffs, largely due to the need for a strict validation to assure further reliability and robustness of the results before implementation.

New-borne multivariate analysis of multiple data obtained through the application of spectroscopic as well spectrometric techniques still lacks harmonisation of the working conditions, from sample preparation to statistical data elaboration and interpretation. This is probably the major challenge to be faced in the near future, where only the effective comparability and portability of analytical data among different laboratories, supported by the creation and constant updating of larger and up-to-date databases of authentic reference foodstuffs within different laboratories could provide the sound evidence base required for the practical application of these methodologies to ensure the integrity of the fisheries and aquaculture value chain.

ACKNOWLEDGMENTS

This PhD Thesis is just the last major milestone achieved thanks to the valuable experience, professionalism, and support of Prof. Sergio Guiding and Prof. Emanuela Zanardi. I am extremely grateful not only for their supervision and guidance throughout this this three-years' research programme, but also for their faith in me and profound belief in this work which have been central for my personal and professional growth.

I am extremely grateful to prof. Adriana Ianieri for guiding me through the many obstacles and for her generous advice and encouragements which helped keep me resolute during this path.

My endless thank to Prof. Lenka Husáková for the kindness and cordiality with which she welcomed me to the University of Pardubice and into her research group, for sharing her experience and for providing a central and essential scientific support to this PhD Dissertation. I would also like to acknowledge Dr Jan Patočka for having offered me his help and making his expertise available for the success of the research project.

My thankfulness is also to my colleagues, Pedro, Silvio, and Cezara for their understanding and supportiveness in the most difficult moments. The good and bad times spent together will be a nice memory for me.

Finally, my eternal appreciation towards my mom, dad, and my brother who have passed on honesty, love, and the courage to face any difficulty to me and have always been by my side, no matter how far apart they are.

A big thank also to Francesco just for sharing this journey with me. This has meant everything to me.

AUTHOR

Maria Olga Varrà

Email: mariaolga.varra@unipr.it
olga.varr@gmail.com
Skype: olga.varr
Researchgate: https://www.researchgate.net/profile/Maria_Olga_Varra
ORCID iD: https://orcid.org/0000-0001-9068-0254

Research Articles

Sergio Ghidini, **Maria Olga Varrà,** Chiara Dall'Asta, Anna Badiani, Adriana Ianieri, Emanuela Zanardi (2019).
Rapid authentication of European sea bass (Dicentrarchus labrax L.) according to production method, farming system, and geographical origin by near infrared spectroscopy coupled with chemometrics.
Food Chemistry, 280, 321-327.
DOI: 10.1016/j.foodchem.2018.12.075

Maria Olga Varrà, Sergio Ghidini, Emanuela Zanardi, Anna Badiani, Adriana Ianieri (2019).
Authentication of European sea bass according to production method and geographical origin by light stable isotope ratio and rare earth elements analyses combined with chemometrics.
Italian Journal of Food Safety, 8 (1).
DOI: 10.4081/ijfs.2019.7872

Sergio Ghidini, **Maria Olga Varrà**, Emanuela Zanardi
Approaching Authenticity Issues in Fish and Seafood Products by Qualitative Spectroscopy and Chemometrics (2019).
Molecules, 24 (9), *1812.*
DOI: 10.3390/molecules24091812

Maria Olga Varrà, Luca Fasolato, Lorenza Serva, Sergio Ghidini, Enrico Novelli, Emanuela Zanardi
Detection of irradiated dry fermented sausages by near infrared spectroscopy coupled with chemometrics (2020).
Food Control, 110, 107009
DOI: 10.1016/j.foodcont.2019.107009

Maria Olga Varrà, Sergio Ghidini, Adriana Ianieri, Emanuela Zanardi.
Near-infrared spectral fingerprinting: a tool against origin-related fraud in the sector of processed anchovies (2020).
Food Control
DOI: 10.1016/j.foodcont.2020.107778

Poster presentations

Varrà, M.O.; Zanardi, E.; Ghidini, S.; Ianieri, A.

Tecnica rapida per la caratterizzazione d'origine geografica di acciughe salate
XXIX Convegno Nazionale AIVI (Associazione Italiana Veterinari Igienisti), Bari (September 2019).

Ghidini, S.; Zanardi, E.; Varrà, M.O.; Colagiorgi, A.; Simon, A.C.; Ianieri, A.
VIVALDI—veterinary validation of point-of-care detection instrument
3rd Safepork 2019, ONE HEALTH - Tear down interdisciplinary walls, Berlin (August 2019).

Varrà, M.O.; Fasolato, L.; Serva, L.; Ghidini, S.; Novelli, E.; Zanardi, E.
Detection of irradiated dry fermented sausages by near infrared spectroscopy
65th ICOMST (International Congress of Meat Science and Technology), Berlin (August 2019).

Ghidini, S.; Maisano, A.M.; Scali, F.; Santucci, G.; Varrà, M.O.; Zanardi, E.; Ianieri, A.; Alborali, L.G.; Borrello, S.
Tail lesions at slaughterhouse in Italian heavy pigs: difference between intact tail and docked
73° Convegno SISVet (Società Italiana delle Scienze Veterinarie), Olbia (June 2019).

Varrà, M.O.; Ghidini, S; Zanardi, E; Ianieri, A.
Authentication of European sea bass (Dicentrarchus labrax L.) according to production method, farming system, and geographical origin by near infrared spectroscopy coupled with chemometrics
Rapid Methods Europe 2018, Amsterdam (November 2018).

Varrà, M.O.
Development of rapid methods for fish quality, safety, and authenticity
23rd Workshop on the Developments in the Italian PhD Research on Food Science Technology and Biotechnology, Oristano (September 2018)

Varrà, M.O.; Ghidini, S.; Zanardi, E.; A. Ianieri, A
Analisi chemiometrica degli isotopi stabili e delle terre rare per l'autenticazione d'origine geografica e del metodo di produzione di branzini (Dicentrarchus Labrax L.)
XXVIII Convegno Nazionale AIVI (Associazione Italiana Veterinari Igienisti), Milan (September 2018).

Ghidini, S.; Scalfi, F.; Borrello, S.; Colagiorgi, A., Ianieri, A.; Di Ciccio, P.A.; Varrà, M.O.; Zanardi, E.; Maisano, A. M.; Giacomini, E.; Alborali, G.L.
Applicazione del sistema CLASSYFARM su allevamenti suinicoli da ingrasso della Lombardia: relazioni tra consumo di antibiotici, benessere e lesioni al macello
XXVIII Convegno Nazionale AIVI (Associazione Italiana Veterinari Igienisti), Milan (September 2018).

Colagiorgi, A.; Baldo, V.; Scali, F.; Varrà, M.O.; Zanardi, E.; Ghidini, S.; Alborali, G.L.; Ianieri, A.
Antibiotic resistance of methicillin-resistant Staphylococcus aureus (MRSA) from Italian swine chain
72° Convegno SISVet (Società Italiana delle Scienze Veterinarie), Turin (June 2018).

Scientific prizes/awards

"**Best Poster Award winner**": *Analisi chemiometrica degli isotopi stabili e delle terre rare per l'autenticazione d'origine geografica e del metodo di produzione di branzini (Dicentrarchus Labrax L.).* Varrà, M.O.; Ghidini, S.; Zanardi, E.; A. Ianieri, A.
XXVIII Convegno Nazionale AIVI (Associazione Italiana Veterinari Igienisti), Milan (September 2018).

"**Best Poster Award winner**": *Tail lesions at slaughterhouse in Italian heavy pigs: difference between intact tail and docked.*
Ghidini, S.; Maisano, A.M.; Scali, F.; Santucci, G.; Varrà, M.O.; Zanardi, E.; Ianieri, A.; Alborali, L.G.; Borrello, S.
73° Convegno SISVet (Società Italiana delle Scienze Veterinarie), (June 2018).

Workshops and Schools

23rd Workshop on the Developments in the Italian PhD Research on Food Science Technology and Biotechnology, Oristano, Italy (September 2018).

Workshop: New tools for handling animal health and food safety: the VIVALDI project University of Parma (January 2019)

Workshop: Quality by Design, Scienza e tecnologia per la qualità, sicurezza ed innovazione in campo alimentare. University of Parma (May 2019).

Training school: Identification and Traceability of Fishery Products for the Official Inspection Activities. Mercato Ittico di Milano (May-October 2019)

Parma Summer School "One Health". School of Advanced Studies on Food and Nutrition, University of Parma, European Food Safety Authority and Università Cattolica del Sacro Cuore of Piacenza (June 2020).