

Article

Harnessing the Full Power of Chemometric-Based Analysis of Total Reflection X-ray Fluorescence Spectral Data to Boost the Identification of Seafood Provenance and Fishing Areas

Bernardo Duarte ^{1,2,*}, Renato Mamede ¹, João Carreiras ¹, Irina A. Duarte ¹, Isabel Caçador ^{1,2}, Patrick Reis-Santos ³, Rita P. Vasconcelos ⁴, Carla Gameiro ⁴, Pedro Ré ^{5,6}, Susanne E. Tanner ^{1,6} and Vanessa F. Fonseca ^{1,6}

- ¹ MARE—Marine and Environmental Sciences Centre & ARNET—Aquatic Research Infrastructure Network Associated Laboratory, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
 - ² Departamento de Biologia Vegetal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
 - ³ Southern Seas Ecology Laboratories, School of Biological Sciences, The University of Adelaide, Adelaide, SA 5005, Australia
 - ⁴ IPMA—Instituto Português do Mar e da Atmosfera, Av. Dr. Alfredo Magalhães Ramalho 6, 1495-165 Algés, Portugal
 - ⁵ MARE—Marine and Environmental Sciences Centre & ARNET—Aquatic Research Infrastructure Network Associated Laboratory, Laboratório Marítimo da Guia, Faculdade de Ciências, Universidade de Lisboa, Avenida Nossa Senhora do Cabo, 2750-374 Cascais, Portugal
 - ⁶ Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
- * Correspondence: baduarte@fc.ul.pt



Citation: Duarte, B.; Mamede, R.; Carreiras, J.; Duarte, I.A.; Caçador, I.; Reis-Santos, P.; Vasconcelos, R.P.; Gameiro, C.; Ré, P.; Tanner, S.E.; et al. Harnessing the Full Power of Chemometric-Based Analysis of Total Reflection X-ray Fluorescence Spectral Data to Boost the Identification of Seafood Provenance and Fishing Areas. *Foods* **2022**, *11*, 2699. <https://doi.org/10.3390/foods11172699>

Academic Editor: Gianfranco Picone

Received: 9 August 2022

Accepted: 1 September 2022

Published: 4 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Provenance and traceability are crucial aspects of seafood safety, supporting managers and regulators, and allowing consumers to have clear information about the origin of the seafood products they consume. In the present study, we developed an innovative spectral approach based on total reflection X-ray fluorescence (TXRF) spectroscopy to identify the provenance of seafood and present a case study for five economically relevant marine species harvested in different areas of the Atlantic Portuguese coast: three bony fish—*Merluccius merluccius*, *Scomber colias*, and *Sparus aurata*; one elasmobranch—*Raja clavata*; one cephalopod—*Octopus vulgaris*. Applying a first-order Savitzky–Golay transformation to the TXRF spectra reduced the potential matrix physical effects on the light scattering of the X-ray beam while maintaining the spectral differences inherent to the chemical composition of the samples. Furthermore, a variable importance in projection partial least-squares discriminant analysis (VIP-PLS-DA), with $k - 1$ components (where k is the number of geographical origins of each seafood species), produced robust high-quality models of classification of samples according to their geographical origin, with several clusters well-evidenced in the dispersion plots of all species. Four of the five species displayed models with an overall classification above 80.0%, whereas the lowest classification accuracy for *S. aurata* was 74.2%. Notably, about 10% of the spectral features that significantly contribute to class differentiation are shared among all species. The results obtained suggest that TXRF spectra can be used for traceability purposes in seafood species (from bony and cartilaginous fishes to cephalopods) and that the presented chemometric approach has an added value for coupling with classic TXRF spectral peak deconvolution and elemental quantification, allowing characterization of the geographical origin of samples, providing a highly accurate and informative dataset in terms of food safety.

Keywords: traceability; X-ray spectroscopy; chemometrics

1. Introduction

The strengthening of legal requirements in food safety in recent decades has led to the development and implementation of geographical origin authentication methodologies that allow consumers to know the origin of the food products they consume [1]. To tackle the growing issue of food fraud, the European Union (EU) published a resolution compelling all member states to develop and adopt tools to increase food traceability and prevent mislabeling [2]. Similar regulations worldwide have led to a growing number of studies focused on the development of elemental and biochemical markers that provide natural tags of the geographical location of the product capture and production, without interference from the producer's report [1]; the development of genetic approaches have greatly advanced the ability to identify species in seafood products even after food processing [3]. Various reasons underlie mislabeling, including (i) involuntary mislabeling of origin and species, (ii) misidentification of closely related species, (iii) misunderstanding of the common names of species [4], or (iv) deliberate mislabeling of species for direct commercial benefit [5], replacing labels of low-value species with economically relevant species, e.g., farmed catfish (*Pangasius* sp.) identified as wild-caught Atlantic cod (*Gadus morhua*) [6]. Moreover, restrictions on the catches of specific species enforced as management measures can lead to intentional food fraud to obfuscate collection in areas or periods where and/or when catches are forbidden [7]. Seafood is among the most economically valuable key components of the human diet, comprising 16.7% of the animal protein consumption per person on a global scale [8]. Moreover, in recent decades, societal concern to pursue a healthier lifestyle and diet has boosted seafood products' consumption from 9.9 kg per capita in the 1960s to 19.7 kg per capita in 2013 [8]. With this high value and high demand for seafood products, the risk of food fraud associated with mislabeling is greatly amplified, either unintentionally or with the intent to gain profit from illegal practices [7].

Several techniques have been employed to trace the geographical origin of food products, including elemental analysis [9–14], isotope analysis [15–17], fatty acid profiles [18–20], and optical spectroscopy techniques [21–23]. These techniques, alone or combined, produce large datasets that can be analyzed either through classical statistical techniques or by advanced chemometric approaches [14,21,22,24,25].

Considering optical spectroscopy techniques that generate large amounts of data, the use of statistical chemometric approaches for data analysis is especially valuable since these techniques provide a way to visualize variation or patterns within large multivariate data sets and enable the subsequent application of calibration or classification models [21,23,26,27]. The most common optical spectroscopy techniques used for food traceability purposes are vibration spectroscopy [27] and near-infrared or Fourier transform infrared technologies [21,23,26]. These techniques produce spectral information on the food sample, and it is the analysis of this (raw or normalized) optical data through chemometrics that allows for interpretations and identifications of chemical or biochemical compounds independently of highly specialized technicians or chemistry or biochemistry researchers [26].

Total reflection X-ray fluorescence (TXRF) spectroscopy elemental data have a high degree of accuracy in depicting elemental signatures of seafood products of different geographical origins [12–14]. Nevertheless, as with other spectroscopy techniques, the generated spectra contain more information than the one used to calculate specific indexes or compound concentrations [21,23,26,28]. Considering optical spectroscopy techniques, the use of statistical chemometric approaches for data analysis is especially valuable because these techniques generate large amounts of data. Thus, the analysis of TXRF full spectral data through chemometric approaches can improve the capability of this analytical technology for geographical traceability purposes, improving the classification accuracy of seafood provenance beyond element-concentration-based classifications [28].

In the present work, we aimed to evaluate the applicability of TXRF full spectral data coupled with chemometric models to depict the geographical provenance of muscle tissue

samples of five economically relevant marine species harvested in different areas of the Portuguese Northeast Atlantic coast.

2. Materials and Methods

2.1. Sample Collection

Five species of seafood were collected from commercial fisheries in five fishing areas along the North Atlantic Portuguese coast (Figure 1), namely European hake (*Merluccius merluccius*), Atlantic chub mackerel (*Scomber colias*), gilthead seabream (*Sparus aurata*), thornback ray (*Raja clavata*), and common octopus (*Octopus vulgaris*). A total of 649 individuals were collected, with four out of five areas sampled per species (except for *O. vulgaris* and *S. colias*, which were sampled in all five areas), and with 30 individuals (i.e., replicates) per species per area (except for the center-south area where only 19 individuals of *R. clavata* were sampled). Individuals were transported fresh to the laboratory, where they were individually dissected to collect muscle tissue samples for elemental analysis and subsequently stored at $-80\text{ }^{\circ}\text{C}$ and then freeze-dried before chemical analysis.

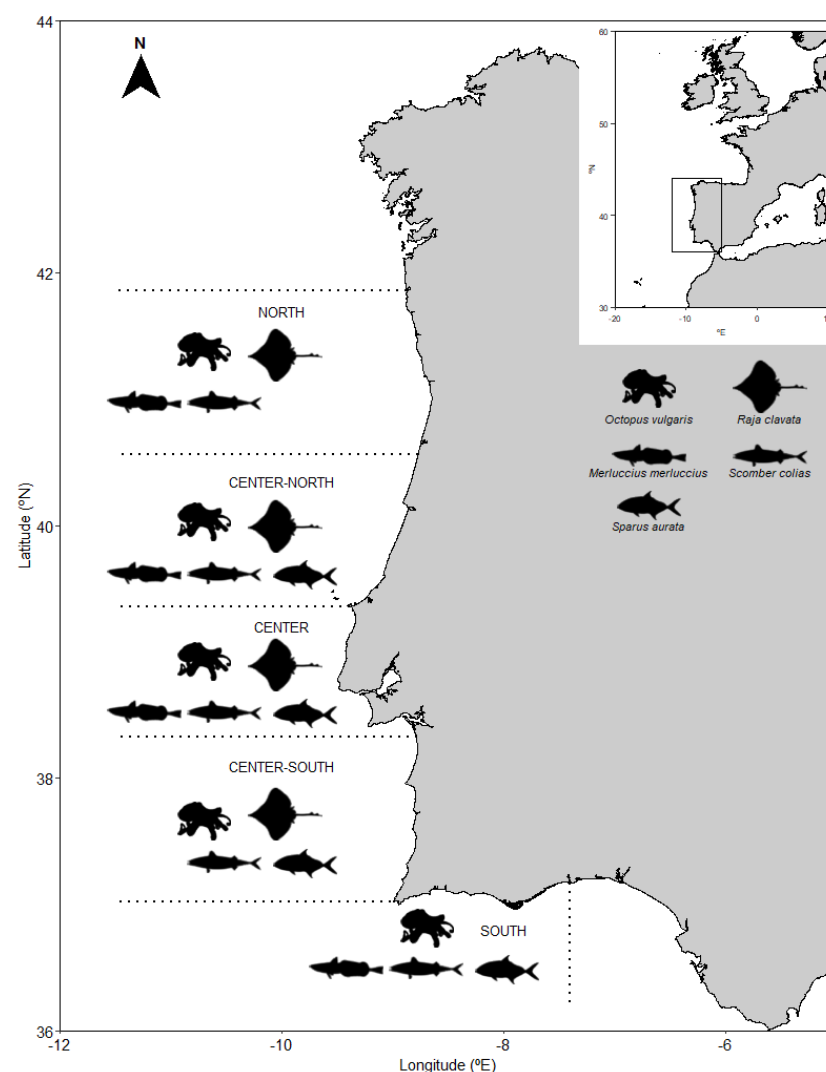


Figure 1. *Merluccius merluccius*, *Octopus vulgaris*, *Raja clavata*, *Sparus aurata*, and *Scomber colias* sampling sites along the Portuguese Atlantic coast.

2.2. Sample Processing and TXRF Analysis

All labware used for TXRF analysis was decontaminated in acid baths for 48 h before use. Freeze-dried samples (approximately 200 mg) were mineralized with HNO_3

in Teflon reactors, following a microwave digestion process (Multiwave GO, Anton Paar GmbH, Graz, Austria) according to the EPA 3052 method [29]. After cooling, an internal standard (gallium) was added to each sample, and 5 μL of each sample was then applied to a siliconized quartz disk (BrukerNano, Berlin, Germany) and dried. Total reflection X-ray fluorescence spectroscopy was performed in a TXRF S2 PICOFOX (Bruker, Germany). Instrumental recalibration (gain correction, sensitivity analysis, and multi-elemental standards) and analytical blanks were used for quality control. The data were acquired using Spectra PICOFOX Software (version 7.8.20. Bruker, Berlin, Germany).

2.3. Spectrum Data Processing and Chemometric Analysis

Specific transformations are commonly applied before the application of partial least-squares discriminant analysis, aiming to reduce the unwanted effects of light scatter caused by the intrinsic physical structure features of the medium of the sample [30]. Among the most common transformations performed on spectral data are the first and second derivatives, allowing for the removal of vertical offsets and linearly sloping baselines [30]. One of the most common algorithms used for this purpose is the Savitzky–Golay transformation [31]. This transformation is based on a localized linear regression of several neighboring points to determine the appropriate polynomial. This polynomial can be mathematically differentiated and evaluated at the x values (in this case, energy values) and, in practical terms, is a mathematical equivalent of the regression. The differentiation procedure is performed by a convolution with a set of derived coefficients [32].

A Savitzky–Golay first-order smoothing normalization coupled with the first derivative of the spectra data was performed using the *mdatools* package [33]. A size window of three points was used throughout the whole spectral range, having as a basis the raw spectral data obtained from the sample analysis. Savitzky–Golay parameters were selected according to the literature, allowing for a more direct comparison of the results obtained with previously reported results. Savitzky–Golay processing of spectral data was performed for all the biological replicates. After pre-processing, the datasets consisted of 649 individuals (generally 30 replicate individuals per sampling site * species; 4 to 5 sites per species, as described in Section 2.1) with 3025 variables in each spectral dataset.

For the chemometric approach, a partial least-squares discriminant analysis (PLS-DA) methodology was used, and a variable selection method was implemented, specifically variable importance in projection (VIP) of PLS-DA. Both analyses were performed using the *Discriminer* package [34] in R-Studio Version 1.4.1717 [35]. Cross-validation was performed using the leave-one-out function of the package, and the percentage of correct classification to the known geographical origin of the sample in cross-validation was used as a measure of model accuracy. For the leave-one-out cross-validation procedure and considering the N classes considered for each species, for each i th case in $(1, 2, \dots, N)$, the data were tested (except for the i th case) to build the classifier model. After this procedure, the model was applied to the i th case, and its classification was evaluated. This procedure was repeated N times, allowing all cases to be assigned to a classification label and the model accuracy evaluated. According to previous works [36], leave-one-out cross-validation is the most adequate for small sample size studies in comparison with resubstitution and simple split-sample estimates that lead to serious bias, with the leave-one-out cross-validation being the method with the smallest bias for discriminant analysis. For each species, the number of components for the model was set as $k - 1$, where k is the number of geographical origins of the species. The model performance was evaluated using the receiver operating characteristic (ROC) area under the curve (AUC) parameter, the goodness of fit or explained variation (R^2), and the goodness of prediction or predicted variation (Q^2). ROC is a probability curve of the false positive rate in the x -axis (i.e., $\text{FPR} = 1 - \text{specificity}$, where $\text{specificity} = \text{true negative} / (\text{true negative} + \text{false positive})$); or $\text{FPR} = \text{false positive} / (\text{true negative} + \text{false positive})$) versus true positive rate or sensitivity in the y -axis (i.e., TPR or $\text{sensitivity} = \text{true positive} / (\text{true positive} + \text{false negative})$). AUC represents the proportion of cases when the model can distinguish between classes. In the present case, the model

assigned a sample to one of several possible geographical origins, for example, for *M. merluccius* from the north: true positives (samples from the north are assigned to the north), false positive (samples not from the north are assigned to the north), true negative (samples not from the north not assigned to the north), and false negative (samples from the north not assigned to the north).

The statistical significance of the AUC parameter was evaluated using a Wilcoxon test. Component selection using ROC-AUC was performed using the *MixOmics* package [37]. After ensuring the correct number of components and high AUC values, model accuracy variable components coordinates were calculated using the *Discriminer* package [34].

The two parameters (R^2 and Q^2 : goodness of fit or explained variation (R^2) and goodness of prediction or predicted variation (Q^2), respectively) differently vary with increasing model complexity. The parameter R^2 is inflationary and approaches 1 as model complexity (number of model parameters) increases. Therefore, it is not sufficient to only consider a high R^2 . The parameter Q^2 , on the other hand, is not inflationary and, at a certain degree of complexity, will not improve any further and will then degrade. Models' performances in internal validation were evaluated in terms of accuracy (%), sensitivity (%), and specificity (%), according to [38]. The model's overall accuracy was calculated by dividing the number of correctly classified samples by the total number of samples.

3. Results and Discussion

Applying a first-order differentiation Savitzky–Golay transformation to the TXRF spectra allowed for normalization of all samples collected from different organisms, reducing the potential matrix physical effects on the light scattering of the X-ray beam while maintaining the spectral differences inherent to the chemical composition of the samples (Figure 2). The Savitzky–Golay transformation also allowed the removal of baseline effects, mainly but not entirely due to the derivative, and reduced scaling variations [39]. This normalization step was essential to remove the physical light-scattering effects of the matrix, which would have had overfitting effects on the subsequent PLS-DA of the spectra dataset.

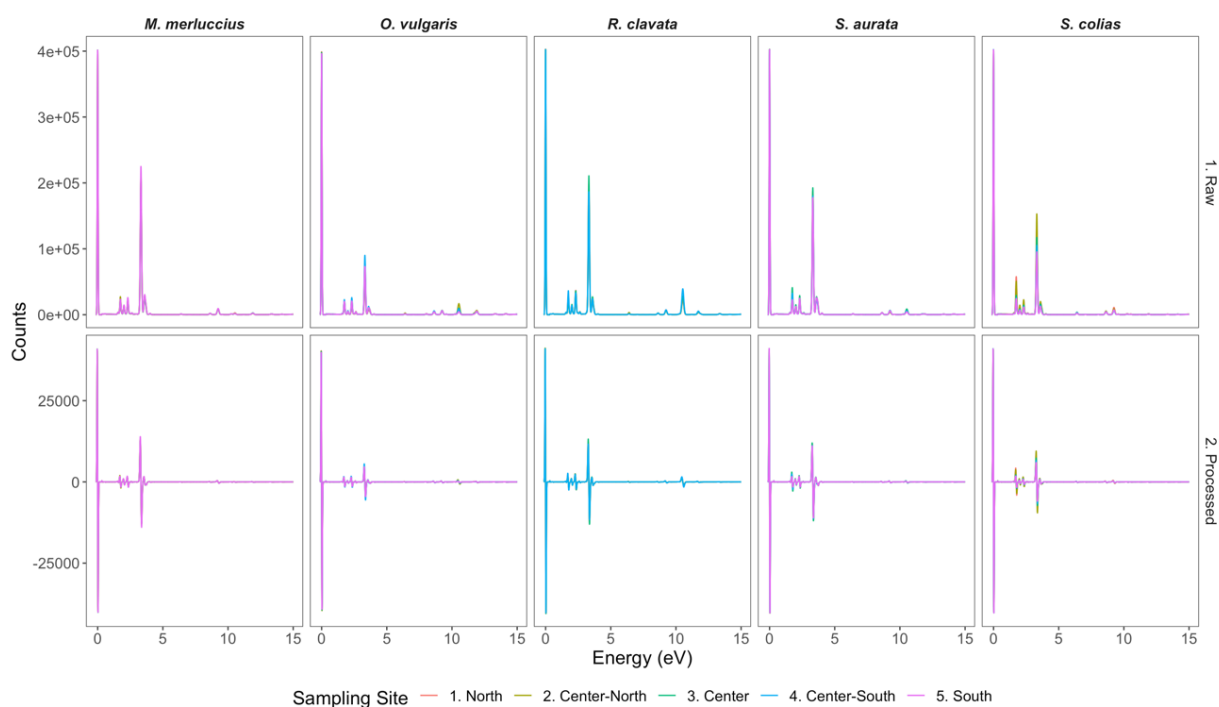


Figure 2. Average raw (upper panel) and processed (lower panel; Savitzky–Golay filter with the first-order differentiation) X-ray fluorescence reflectance spectra of the *Merluccius merluccius*, *Octopus vulgaris*, *Raja clavata*, *Sparus aurata*, and *Scomber colias* samples collected in 5 areas along the Portuguese coast (N = 30 per site per species, except for *R. clavata* in center-south area).

To select the best PLS-DA components, an evaluation of R^2 and Q^2 (Figure 3), as well as of AUC (Figure 4), was obtained from the ROC curves (Supplementary Figure S1). Moreover, the model classification accuracy was assessed. The best model was selected when both Q^2 and R^2 were maximized while maintaining an overall high classification accuracy of the model. In this sense, a number of components of $k - 1$ were selected (where k is the number of geographical origins of the species) as the best model (i.e., the model with the best combination of R^2 , Q^2 , and accuracy) (Figure 3). In contrast, Q^2 decreased when the number of components was above $k - 1$, despite continuing increases in R^2 and classification accuracy (results not shown).

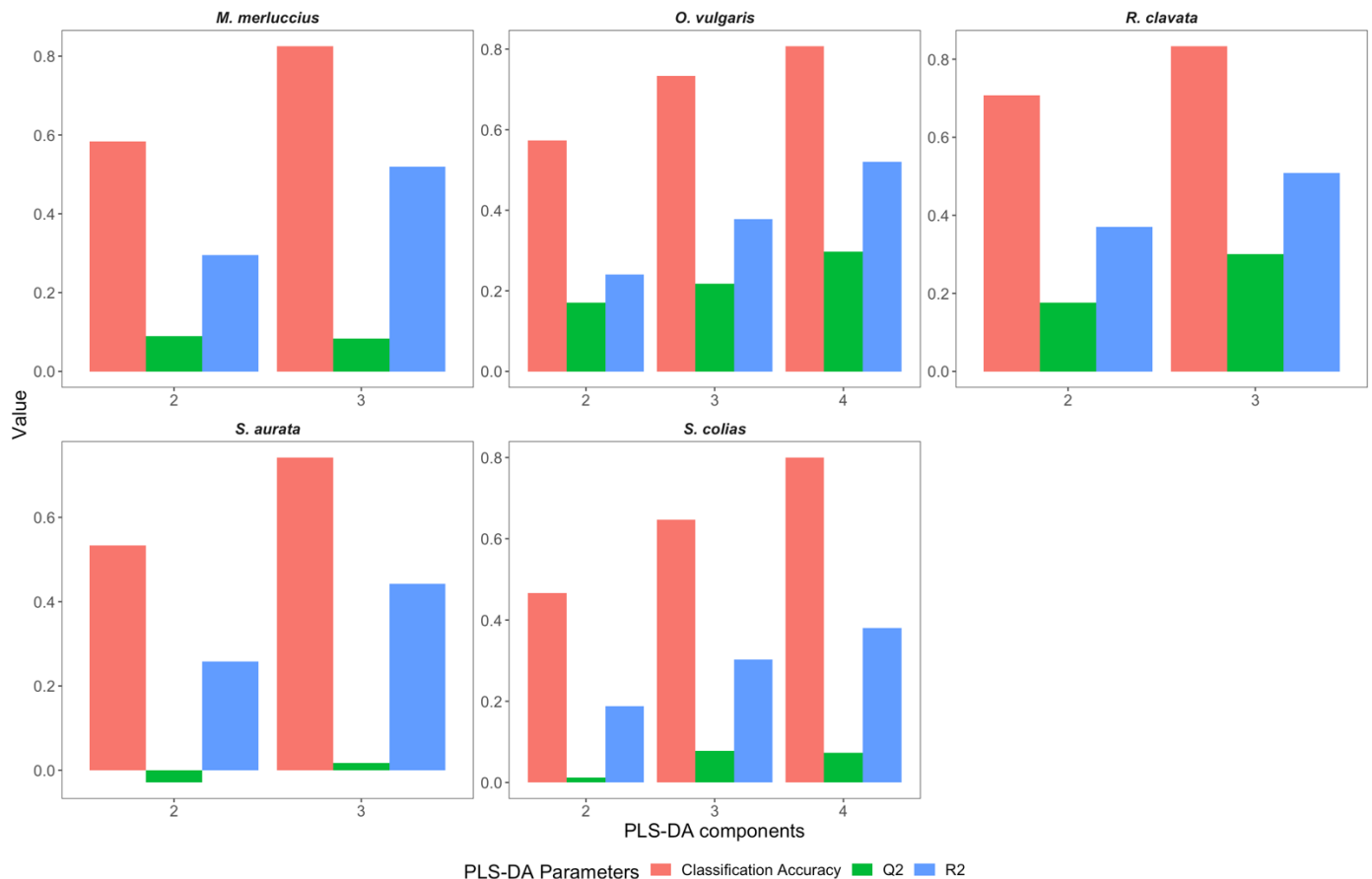


Figure 3. Overall cross-validation classification accuracy (accuracy), goodness of fit or explained variation (R^2), and goodness of prediction or predicted variation (Q^2) of the PLS-DA models, having as input the processed (Savitzky–Golay filter with the first derivative) X-ray fluorescence reflectance spectra for each number of components tested for the five studied species *Merluccius merluccius* (4 sites), *Octopus vulgaris* (5 sites), *Raja clavata* (4 sites), *Sparus aurata* (4 sites), and *Scomber colias* (5 sites,) samples collected along the Portuguese coast. Results for the number of components above $k - 1$ are not shown.

Additionally, the R^2 , Q^2 , and classification accuracy for each species and the number of PLS-DA components were also compared with the ROC-AUC (Figure 4 and Supplementary Figure S1). This comparison further supported that the best choice of components was $k - 1$ (where k is the number of geographical origins of each species), for which AUC (Figure 4) and sensitivity (Supplementary Figure S1) were the highest.

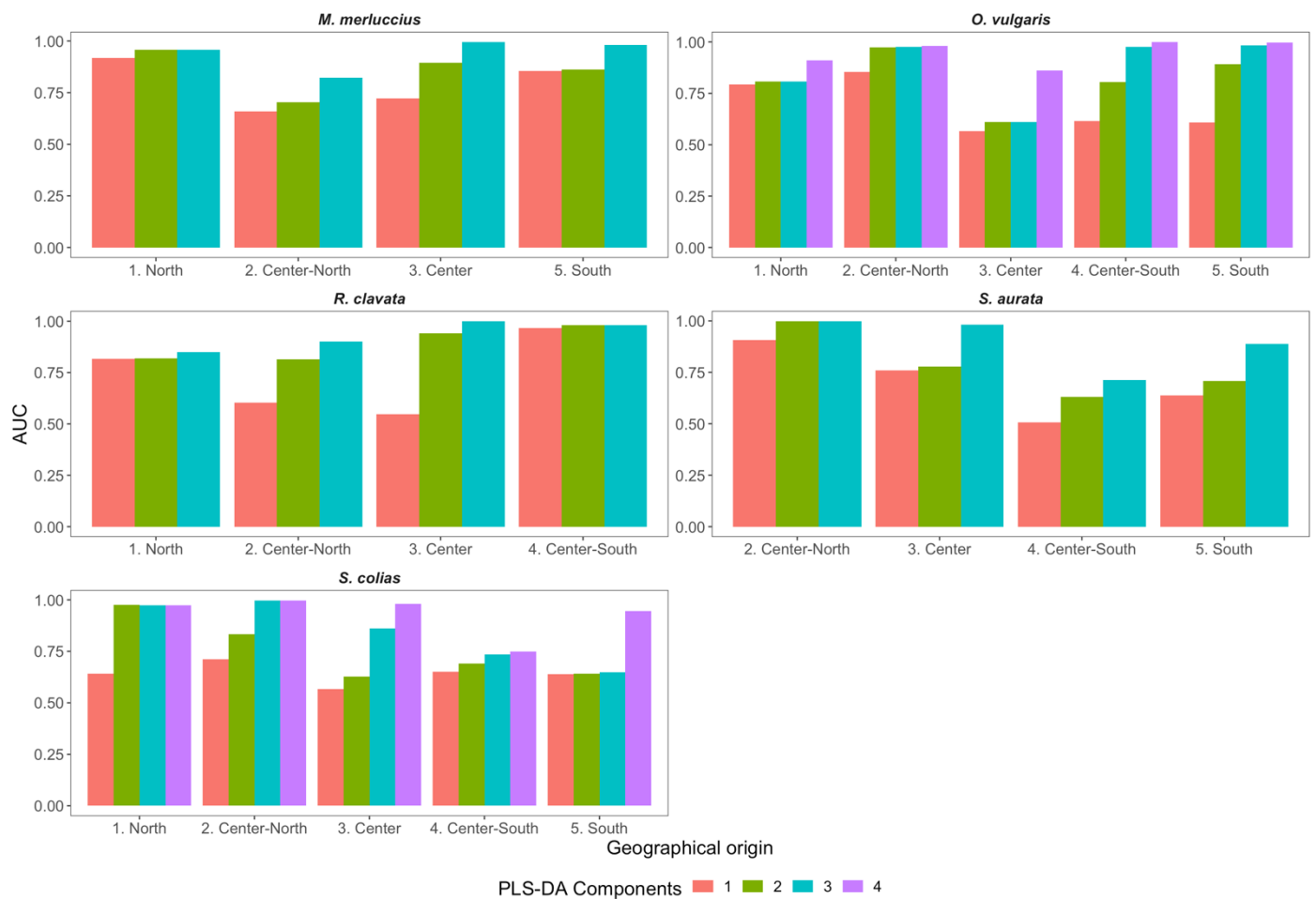


Figure 4. Area under the curve (AUC) of the receiver operating characteristic (ROC) curve of the partial least-squares discriminant analysis (PLS-DA), having as input the processed (Savitzky–Golay first-order differentiation filter) X-ray fluorescence reflectance spectra, conducted for each species (*Merluccius merluccius*, *Octopus vulgaris*, *Raja clavata*, *Sparus aurata*, and *Scomber colias*) and sample geographical origin and number of components. Results for the number of components above $k - 1$ are not shown.

Following the definition and validation of the number of components per species, PLS-DA 2D plots were generated for visualization of model dispersion, displaying the samples along with the first three components of the PLS-DA (Figure 5). Notably, the generated biplots only represent the data dispersion and grouping in the generated PLS-DA models considering the first two dimensions, whereas the generated models were all obtained using more than two components. This resulted in an apparent complete overlap between some groups, an artifact from the first two components that did not occur when considering all the dimensions used in each model. Several clusters were evidenced in each species dispersion plot, generally grouping samples from the same collection site. In agreement with this definition of clusters, the classification accuracy of models (i.e., percentage of correct classification to the known geographical origin of the samples) was generally high (Figure 6). An overall model classification accuracy above 80.0% was observed for most species, with the exception of the model generated for *S. aurata*, which had a lower overall classification accuracy (74.2%). Previous works [40] also indicated that classifiers attained in PLS-DA approaches can guarantee highly efficient classification results in cross-validation. This might be due to the existence of a direct linear relationship between TXRF spectral patterns and the geographical origin of the considered species, promoted by Savitzky–Golay spectral pre-processing operations, because even in highly complex samples, this

correlation can be easily extrapolated by traditional linear techniques [40,41]. For all species analyzed, lower accuracy of classification to geographical origin area (below 75%) was observed in the center areas (i.e., center-north, center, and center-south; Figure 6). This occurred in one area per species and was not limited to a specific type of organism, as the studied marine species have distinct habitat use and biological characteristics: demersal bony fish (*M. merluccius*), pelagic bony fish (*S. colias*), coastal demersal bony fish (*S. aurata*), demersal elasmobranch (*R. clavata*), and benthic cephalopod (*O. vulgaris*). This points to a geographical influence rather than a biological feature, possibly driven by physical-chemical similarities in these coastal areas or by possible capture in areas near the border of adjacent central areas.

In terms of model performance (Table 1), the samples collected in the central part of the study area (center-north, center, and center-south) showed the lowest sensitivity (average sensitivity considering all species of 67.3%). Some of these areas also showed lower precision (average precision of 78.7%), although other areas with high sensitivity presented low precision. This finding is mostly due to the high number of false negatives (samples that were not assigned to their origin), especially evident in *O. vulgaris* and *S. aurata* samples collected in the center area. As for specificity, it was consistently high for all models, except for the *S. aurata* in center-south (74.7%) due to the high number of false positives (samples incorrectly assigned to this origin), thus reducing the specificity of the model for this location.

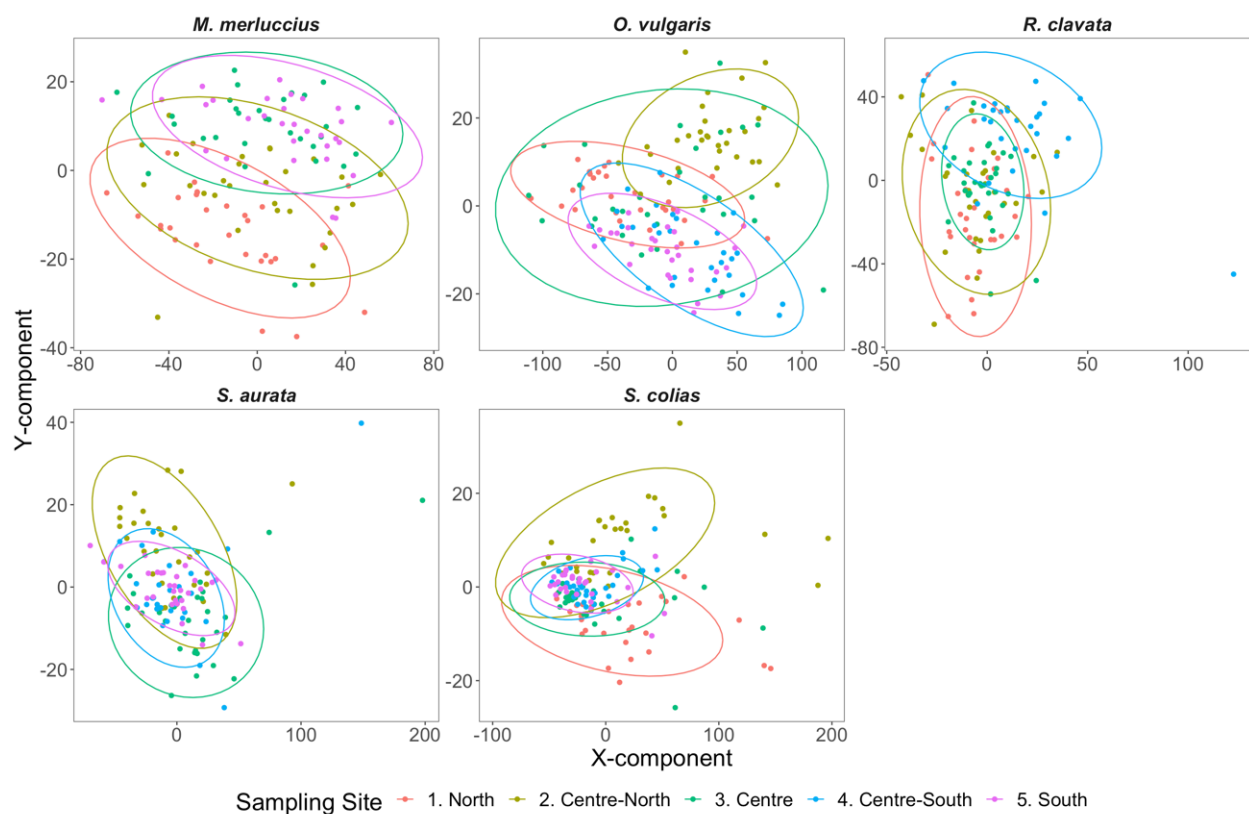


Figure 5. Partial least-squares discriminant analysis (PLS-DA) 3D plots of the first three PLS-DA components having as input the processed (Savitzky–Golay first-order differentiation filter) X-ray fluorescence reflectance spectra of *Merluccius merluccius*, *Octopus vulgaris*, *Raja clavata*, *Sparus aurata*, and *Scomber colias* samples collected along the Portuguese coast (average N = 30 per site per species).

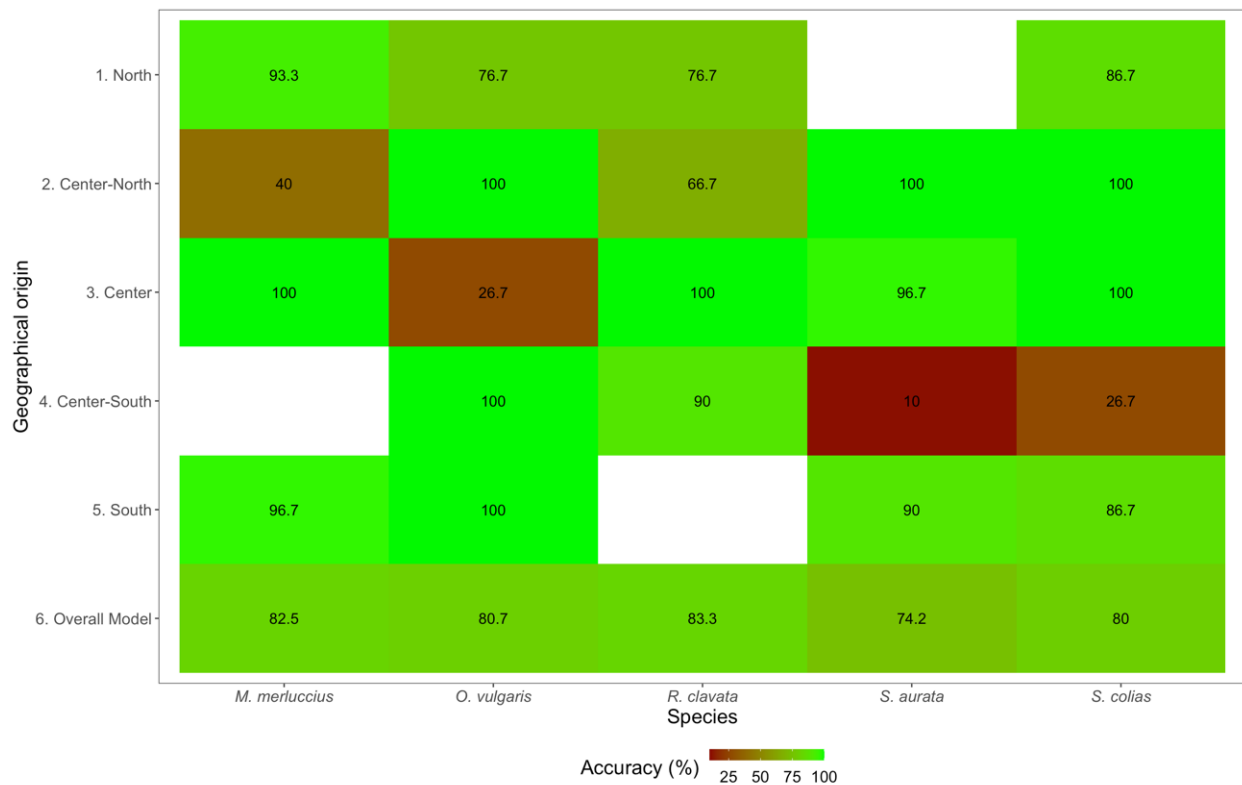


Figure 6. Partial least-squares discriminant analysis (PLS-DA) cross-validation classification accuracy heatmaps per sampling area and for all the different species considered (*Merluccius merluccius*, *Octopus vulgaris*, *Raja clavata*, *Sparus aurata*, and *Scomber colias*).

Table 1. Partial least-squares discriminant analysis (PLS-DA) model cross-validation performance (precision, sensitivity, and specificity) per geographical origin area and overall, based on the processed (Savitzky–Golay first-order differentiation filter) X-ray fluorescence reflectance spectra of the *Merluccius merluccius*, *Octopus vulgaris*, *Raja clavata*, *Sparus aurata*, and *Scomber colias* samples collected along the Portuguese coast.

Species	Geographical Origin	Group Precision	Overall Precision	Group Sensitivity	Overall Sensitivity	Group Specificity	Overall Specificity
<i>M. merluccius</i>	North	70.0%	82.5%	93.3%	82.5%	85.5%	82.5%
	Center-North	92.3%		40.0%		98.9%	
	Center	93.8%		100.0%		97.2%	
	Center-South	-		-		-	
	South	82.9%		96.7%		92.1%	
<i>O. vulgaris</i>	North	59.0%	80.7%	76.7%	80.7%	86.0%	80.7%
	Center-North	83.3%		100.0%		93.8%	
	Center	53.3%		26.7%		94.2%	
	Center-South	100.0%		100.0%		100.0%	
	South	100.0%		100.0%		100.0%	
<i>R. clavata</i>	North	76.7%	83.3%	76.7%	83.3%	91.7%	83.3%
	Center-North	71.4%		66.7%		90.9%	
	Center	100.0%		100.0%		100.0%	
	Center-South	84.4%		90.0%		93.6%	
	South	-		-		-	

Table 1. Cont.

Species	Geographical Origin	Group Precision	Overall Precision	Group Sensitivity	Overall Sensitivity	Group Specificity	Overall Specificity
<i>S. aurata</i>	North	89.7%		100.0%		96.7%	
	Center-North	88.2%	73.3%	96.7%	73.3%	89.4%	73.3%
	Center	73.2%		10.0%		97.7%	
	Center-South	88.9%		86.7%		74.7%	
	South	-		-		-	
<i>S. colias</i>	North	89.7%		86.7%		96.9%	
	Center-North	88.2%	80.0%	100.0%	80.0%	95.7%	80.0%
	Center	73.2%		100.0%		89.1%	
	Center-South	88.9%		26.7%		99.1%	
	South	70.3%		86.7%		89.5%	

Most chemometric approaches based on spectral data use near-infrared (NIR) as a basis [21,23,40], where some spectral regions correspond to specific groups of compounds present in the sample matrix (e.g., lipids, carbohydrates, and proteins) [22]. In X-ray fluorescence-based spectral data, peaks result from the excitation of certain chemical elements present in the sample matrix by X-ray photons, generating fluorescence emission peaks, with each element generating two or more tails [28]. Analyzing the spectral features with VIP scores above one from all the models generated for the considered species, it is possible to observe that certain areas of the spectra appear with a higher density of points corresponding to spectral features with $VIP > 1$ (Figure 7). The Venn diagram revealed that the 311 spectral data points (10% of the spectral features in each dataset) with $VIP > 1$ were shared among the different species datasets analyzed by the PLS-DA approach, indicating that they are key features for sample class differentiation throughout PLS-DA. On the other hand, the unique features highlighted for each species dataset that were not shared by any other were much more reduced in number, ranging from 159 to 193 (5.3% to 6.4% of the spectral features in each dataset). Although each element can have several fluorescence peaks, certain spectral windows can be associated with groups of elements, whereas for elemental concentration calculation, two or more peaks are normally used for deconvolution. Observing the higher data point density regions, it is possible to distinguish four main peaks with a particularly high density of VIP scores with noticeable high values ($VIP > 1.5-2$). The first observable peak area corresponds to the beginning of the spectra, where it corresponds to low-atomic-number (Z) elements such as Na, K, and Ca, highly abundant in marine species [12–14,42,43]. The last three observable peaks correspond to an area where Cu, Zn, Br, Sr, Pb, and other high-Z elements, have one of the main fluorescence peaks, with these elements also being very abundant in marine seafood samples [12–14,42,43]. Nevertheless, the use of specific elements instead of the full TXRF spectra greatly reduces the number of features used as input for the chemometric approaches, from several thousand to a few dozen, as it is observable between elemental analysis and other spectral fingerprinting approaches [40,43]. While for food safety and nutrition analysis, elemental concentration in edible seafood tissues is essential, for provenance and traceability, we can use the full power of the spectral analysis to amplify discrimination and classification success. Nonetheless, XRF spectral data were already previously included in chemometric approaches in several areas from geochemistry to ecology, archaeology, agriculture, material, forensic sciences and medicine [28]. To the best of our knowledge, this is the first time this approach has been employed for food traceability purposes.

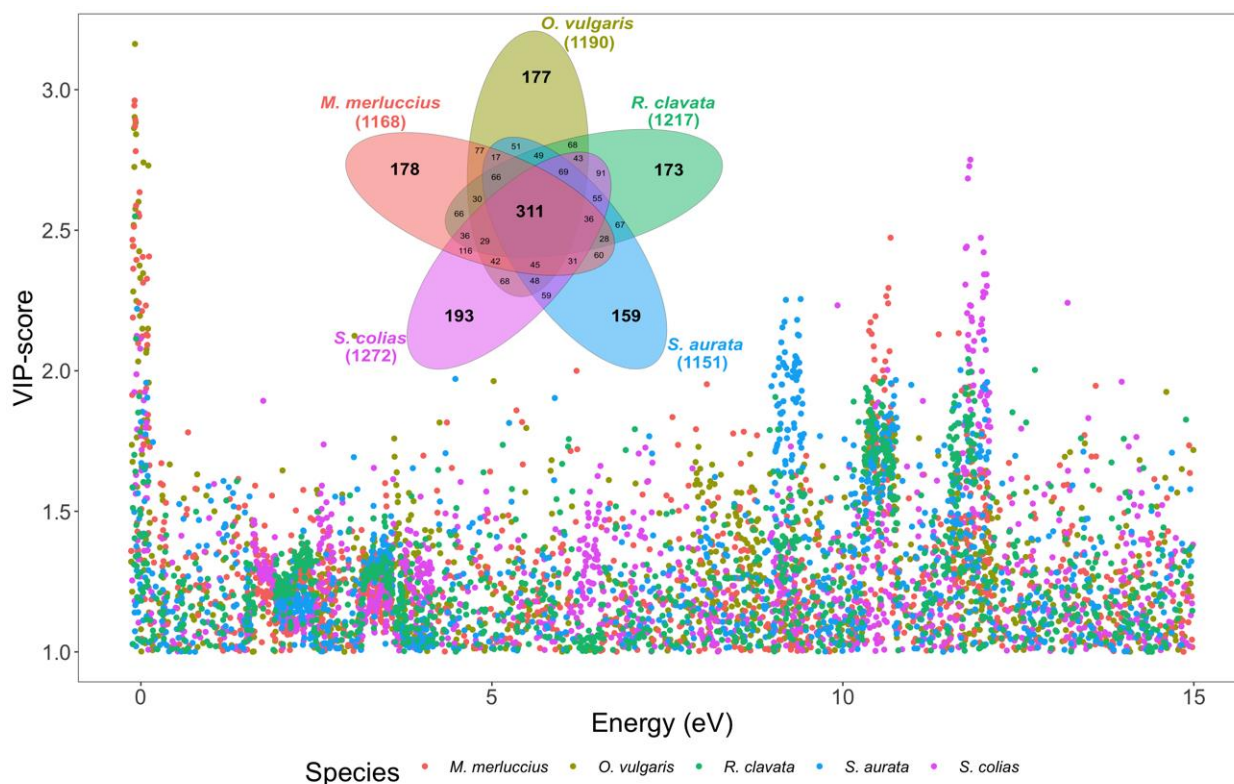


Figure 7. Scatter plot of the VIP scores (only variables with VIP score > 1) attained from the partial least-squares discriminant analysis (PLS-DA) of the processed (Savitzky–Golay first-order differentiation filter) X-ray fluorescence reflectance spectra of the *Merluccius merluccius*, *Octopus vulgaris*, *Raja clavata*, *Sparus aurata*, and *Scomber colias* samples collected along the Portuguese coast, and Venn diagram of the selected variables (VIP score > 1) between the different analysed species.

4. Conclusions

Total X-ray fluorescence (TXRF) analysis provides important elemental data that can be used for food safety and nutritional purposes but also provides a valuable source of spectral data that can be leveraged to boost traceability and provenance applications. Similar to the approach used for infrared spectroscopy, using Savitzky–Golay smoothing normalization coupled with the first derivative approach, it is possible to produce TXRF spectra with significant noise reduction while maintaining discriminant features. Applying PLS-DA to these smoothed spectral datasets was found to be a highly efficient approach to discriminate samples from each species from different sampling areas, with minimum overall model accuracies of 74.2% and individual geographical origins identified with 100% accuracy. It should be emphasized that this approach was achieved for seafood species with very different sample matrixes (muscle tissue from bony and cartilaginous fishes to cephalopods) and habitat use (demersal, pelagic, benthic, and coastal), highlighting the broad applicability of the present methodology. The present methodology is proposed for provenance and traceability purposes. If coupled with classic TXRF spectral peak deconvolution and elemental quantification, it additionally allows for the characterization of different samples in terms of their elemental profiles, providing a highly accurate and informative dataset in terms of food safety.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/foods11172699/s1>, Figure S1: Receiver operating characteristic (ROC) curves for the number of selected partial least-squares discriminant analysis (PLS-DA) components having as input the processed (Savitzky–Golay filter with the first derivative) X-ray fluorescence reflectance spectra of *Merluccius merluccius*, *Octopus vulgaris*, *Raja clavata*, *Sparus aurata*, and *Scomber colias* samples collected along the Portuguese coast.

Author Contributions: Conceptualization, B.D. and V.F.F.; methodology, B.D., R.M. and J.C.; validation, formal analysis, B.D.; investigation, B.D. and V.F.F.; resources, B.D. and V.F.F.; data curation, B.D.; writing—original draft preparation, B.D.; writing—review and editing, R.M., J.C., I.A.D., I.C., P.R.-S., R.P.V., C.G., P.R., S.E.T. and V.F.F.; project administration, B.D. and V.F.F.; funding acquisition, B.D. and V.F.F. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to thank Fundação para a Ciência e a Tecnologia (FCT) for funding MARE (Marine and Environmental Sciences Centre, UIDB/04292/2020 and UIDP/04292/2020) and ARNET (Aquatic Research Infrastructure Network Associated Laboratory, LA/P/0069/2020). The work was also funded by MAR2020 program via the Projects MarCODE (MAR-01.03.01-FEAMP-0047). B.D., S.E.T. and V.F.F. were supported by research contracts (CEECIND/00511/2017, 2021.02710.CEECIND, and 2021.00244.CEECIND). R.P.V. and C.G. were funded by the Programa Nacional de Amostragem Biológica (EU Data Collection Framework). The authors would like to thank scientific observers of the Programa Nacional de Amostragem Biológica from IPMA for their collaboration in the collection of biological samples.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Leal, M.C.; Pimentel, T.; Ricardo, F.; Rosa, R.; Calado, R. *Seafood Traceability: Current Needs, Available Tools, and Biotechnological Challenges for Origin Certification*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 33.
2. European Commission. *European Commission Regulation (EC) No. 178/2002 of the European Parliament and of the Council*; European Commission: Brussels, Belgium, 2002; pp. L31/24–L31/31.
3. Rasmussen, R.S.; Morrissey, M.T. DNA-Based Methods for the Identification of Commercial Fish and Seafood Species. *Compr. Rev. Food Sci. Food Saf.* **2008**, *7*, 280–295. [[CrossRef](#)] [[PubMed](#)]
4. Cawthorn, D.-M.; Mariani, S. Global Trade Statistics Lack Granularity to Inform Traceability and Management of Diverse and High-Value Fishes. *Sci. Rep.* **2017**, *7*, 12852. [[CrossRef](#)] [[PubMed](#)]
5. Barendse, J.; Roel, A.; Longo, C.; Andriessen, L.; Webster, L.M.I.; Ogden, R.; Neat, F. DNA Barcoding Validates Species Labelling of Certified Seafood. *Curr. Biol.* **2019**, *29*, R198–R199. [[CrossRef](#)] [[PubMed](#)]
6. Christiansen, H.; Fournier, N.; Hellemans, B.; Volckaert, F.A.M. Seafood Substitution and Mislabeling in Brussels' Restaurants and Canteens. *Food Control* **2018**, *85*, 66–75. [[CrossRef](#)]
7. He, J. From Country-of-Origin Labelling (COOL) to Seafood Import Monitoring Program (SIMP): How Far Can Seafood Traceability Rules Go? *Mar. Policy* **2018**, *96*, 163–174. [[CrossRef](#)]
8. Reilly, A. *Overview of Food Fraud in the Fisheries Sector*; Fisheries and Aquaculture Circular; Food and Agriculture Organization of the United Nations (FAO): Rome, Italy, 2018; p. 32.
9. Albuquerque, R.; Queiroga, H.; Swearer, S.E.; Calado, R.; Leandro, S.M. Harvest Locations of Goose Barnacles Can Be Successfully Discriminated Using Trace Elemental Signatures. *Sci. Rep.* **2016**, *6*, 27787. [[CrossRef](#)]
10. Arbuckle, N.S.M.; Wormuth, J.H. Trace Elemental Patterns in Humboldt Squid Statoliths from Three Geographic Regions. *Hydrobiologia* **2014**, *725*, 115–123. [[CrossRef](#)]
11. Bennion, M.; Morrison, L.; Shelley, R.; Graham, C. Trace Elemental Fingerprinting of Shells and Soft Tissues Can Identify the Time of Blue Mussel (*Mytilus Edulis*) Harvesting. *Food Control* **2021**, *121*, 107515. [[CrossRef](#)]
12. Duarte, B.; Carreiras, J.; Mamede, R.; Duarte, I.A.; Caçador, I.; Reis-Santos, P.; Vasconcelos, R.P.; Gameiro, C.; Rosa, R.; Tanner, S.E.; et al. Written in Ink: Elemental Signatures in Octopus Ink Successfully Trace Geographical Origin. *J. Food Compos. Anal.* **2022**, *109*, 104479. [[CrossRef](#)]
13. Duarte, B.; Duarte, I.A.; Caçador, I.; Reis-Santos, P.; Vasconcelos, R.P.; Gameiro, C.; Tanner, S.E.; Fonseca, V.F. Elemental Fingerprinting of Thornback Ray (*Raja Clavata*) Muscle Tissue as a Tracer for Provenance and Food Safety Assessment. *Food Control* **2022**, *133*, 108592. [[CrossRef](#)]
14. Duarte, B.; Mamede, R.; Duarte, I.A.; Caçador, I.; Tanner, S.E.; Silva, M.; Jacinto, D.; Cruz, T.; Fonseca, V.F. Elemental Chemometrics as Tools to Depict Stalked Barnacle (*Pollicipes Pollicipes*) Harvest Locations and Food Safety. *Molecules* **2022**, *27*, 1298. [[CrossRef](#)] [[PubMed](#)]
15. Drivelos, S.A.; Georgiou, C.A. Multi-Element and Multi-Isotope-Ratio Analysis to Determine the Geographical Origin of Foods in the European Union. *TrAC Trends Anal. Chem.* **2012**, *40*, 38–51. [[CrossRef](#)]
16. Kelly, S.; Heaton, K.; Hoogewerff, J. Tracing the Geographical Origin of Food: The Application of Multi-Element and Multi-Isotope Analysis. *Trends Food Sci. Technol.* **2005**, *16*, 555–567. [[CrossRef](#)]

17. Varrà, M.O.; Ghidini, S.; Zanardi, E.; Badiani, A.; Ianieri, A. 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. *Ital. J. Food Saf.* **2019**, *8*, 7872. [CrossRef] [PubMed]
18. Fonseca, V.F.; Duarte, I.A.; Matos, A.R.; Reis-Santos, P.; Duarte, B. Fatty Acid Profiles as Natural Tracers of Provenance and Lipid Quality Indicators in Illegally Sourced Fish and Bivalves. *Food Control* **2021**, *134*, 108735. [CrossRef]
19. Mottese, A.F.; Albergamo, A.; Bartolomeo, G.; Bua, G.D.; Rando, R.; De Pasquale, P.; Saija, E.; Donato, D.; Dugo, G. Evaluation of Fatty Acids and Inorganic Elements by Multivariate Statistics for the Traceability of the Sicilian *Capparis spinosa* L. *J. Food Compos. Anal.* **2018**, *72*, 66–74. [CrossRef]
20. Ricardo, F.; Gonçalves, D.; Pimentel, T.; Mamede, R.; Rosário, M.; Domingues, M.; Lillebø, A.I.; Calado, R. Prevalence of Phylogenetic over Environmental Drivers on the Fatty Acid Profiles of the Adductor Muscle of Marine Bivalves and Its Relevance for Traceability. *Ecol. Indic.* **2021**, *129*, 108017. [CrossRef]
21. Ghidini, S.; Varrà, M.O.; Dall'Asta, C.; Badiani, A.; Ianieri, A.; Zanardi, E. 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 Chem.* **2019**, *280*, 321–327. [CrossRef]
22. Ghidini, S.; Varrà, M.O.; Zanardi, E. Approaching Authenticity Issues in Fish and Seafood Products by Qualitative Spectroscopy and Chemometrics. *Molecules* **2019**, *24*, 1812. [CrossRef]
23. Varrà, M.O.; Ghidini, S.; Ianieri, A.; Zanardi, E. Near Infrared Spectral Fingerprinting: A Tool against Origin-Related Fraud in the Sector of Processed Anchovies. *Food Control* **2021**, *123*, 107778. [CrossRef]
24. Bua, G.D.; Albergamo, A.; Annuario, G.; Zammuto, V.; Costa, R.; Dugo, G. High-Throughput ICP-MS and Chemometrics for Exploring the Major and Trace Element Profile of the Mediterranean Sepia Ink. *Food Anal. Methods* **2017**, *10*, 1181–1190. [CrossRef]
25. Costas-Rodríguez, M.; Lavilla, I.; Bendicho, C. Classification of Cultivated Mussels from Galicia (Northwest Spain) with European Protected Designation of Origin Using Trace Element Fingerprint and Chemometric Analysis. *Anal. Chim. Acta* **2010**, *664*, 121–128. [CrossRef] [PubMed]
26. Cozzolino, D. An Overview of the Use of Infrared Spectroscopy and Chemometrics in Authenticity and Traceability of Cereals. *Food Res. Int.* **2014**, *60*, 262–265. [CrossRef]
27. Power, A.; Cozzolino, D. How Fishy Is Your Fish? Authentication, Provenance and Traceability in Fish and Seafood by Means of Vibrational Spectroscopy. *Appl. Sci.* **2020**, *10*, 4150. [CrossRef]
28. Panchuk, V.; Yaroshenko, I.; Legin, A.; Semenov, V.; Kirsanov, D. Application of Chemometric Methods to XRF-Data—A Tutorial Review. *Anal. Chim. Acta* **2018**, *1040*, 19–32. [CrossRef]
29. Environmental Protection Agency (EPA). *Test Method 3052: Microwave Assisted Acid Digestion of Siliceous and Organically Based Matrices*; Environmental Protection Agency (EPA): Washington, DC, USA, 1996; p. 20.
30. Delwiche, S.R.; Reeves, J.B. A Graphical Method to Evaluate Spectral Preprocessing in Multivariate Regression Calibrations: Example with Savitzky–Golay Filters and Partial Least Squares Regression. *Appl. Spectrosc.* **2010**, *64*, 73–82. [CrossRef]
31. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
32. Steinier, J.; Termonia, Y.; Deltour, J. Smoothing and Differentiation of Data by Simplified Least Square Procedure. *Anal. Chem.* **1972**, *44*, 1906–1909. [CrossRef]
33. Kucheryavskiy, S. Mdatools—R Package for Chemometrics. *Chemom. Intell. Lab. Syst.* **2020**, *198*, 103937. [CrossRef]
34. Sanchez, G. Package ‘DiscrMiner’. 2013. Available online: <https://mran.microsoft.com/snapshot/2015-10-02/web/packages/DiscrMiner/DiscrMiner.pdf> (accessed on 31 August 2022).
35. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2021.
36. Molinaro, A.M.; Simon, R.; Pfeiffer, R.M. Prediction Error Estimation: A Comparison of Resampling Methods. *Bioinformatics* **2005**, *21*, 3301–3307. [CrossRef]
37. Rohart, F.; Gautier, B.; Singh, A.; Lê Cao, K.-A. MixOmics: An R Package for ‘omics Feature Selection and Multiple Data Integration. *bioRxiv* **2017**, *13*, 108597. [CrossRef] [PubMed]
38. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
39. Zimmermann, B.; Kohler, A. Optimizing Savitzky–Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy. *Appl. Spectrosc.* **2013**, *67*, 892–902. [CrossRef] [PubMed]
40. Varrà, M.O.; Ghidini, S.; Fabrile, M.P.; Ianieri, A.; Zanardi, E. Country of Origin Label Monitoring of Musky and Common Octopuses (*Eledone* spp. and *Octopus vulgaris*) by Means of a Portable near-Infrared Spectroscopic Device. *Food Control.* **2022**, *138*, 109052. [CrossRef]
41. Zareef, M.; Chen, Q.; Hassan, M.M.; Arslan, M.; Hashim, M.M.; Ahmad, W.; Kutsanedzie, F.Y.H.; Agyekum, A.A. An Overview on the Applications of Typical Non-Linear Algorithms Coupled With NIR Spectroscopy in Food Analysis. *Food Eng. Rev.* **2020**, *12*, 173–190. [CrossRef]
42. Mamede, R.; Ricardo, F.; Gonçalves, D.; Ferreira da Silva, E.; Patinha, C.; Calado, R. Assessing the Use of Surrogate Species for a More Cost-Effective Traceability of Geographic Origin Using Elemental Fingerprints of Bivalve Shells. *Ecol. Indic.* **2021**, *130*, 108065. [CrossRef]
43. Varrà, M.O.; Husáková, L.; Patočka, J.; Ghidini, S.; Zanardi, E. Multi-Element Signature of Cuttlefish and Its Potential for the Discrimination of Different Geographical Provenances and Traceability. *Food Chem.* **2021**, *356*, 129687. [CrossRef]