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

(Article begins on next page)

12 September 2024

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], Emanuela Zanardi[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
* Corresponding author.
E-mail addresses: sergio.ghidini@unipr.it (S. Ghidini), mariaolga.varra@studenti.unipr.it (M.O. Varrà), chiara.dellasta@unipr.it (C. Dall'Asta), anna.badiani@unibo.it (A. Badiani), adriana.ianieri@unipr.it (A. Ianieri), emanuela.zanardi@unipr.it (E. Zanardi).

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.

1. Introduction

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 (European Parliament and Council of the European Union, 2013), fishery and aquaculture products must be la- belled 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 in- formation 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 eco- nomically important fish species in the whole Mediterranean area: Turkey and Greece have recently become the largest producing coun- tries, while Italy, Spain and France continue to remain the world's leading importers (EUMOFA (European Market Observatory for Fisheries and Aquaculture products), 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 (FAO 2005-2018, 2005; Moretti, Fernandez-Criado, Cittolin, & Guidastri, 1999). 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 (EUMOFA, 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, Chree, Scaife, & Murray, 2005), fresh from frozen-thawed cod fillet and Atlantic salmon (Sivertsen, Kimiya, & Heia, 2011), 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 samples (Ottavian

et al., 2012) according to different intensity of farming system used (Majolini, Trocino, Xiccato, & Santulli, 2010; Xiccato, Trocino, Tulli, & Tibaldi, 2004), 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 ac- cording to production method (wild vs. farmed), farming system (in- tensive 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 were 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/m^3; semi-intensively reared ones (SI; n = 20) were reared in earthen tank at a stocking density up to 1 kg/m^3; extensively reared ones (Ext; n = 10) were reared in coastal lagoons (valliculture) at a stocking density up to 0.0025 kg/m^3.

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 1100–2500 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 opera- tions 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 sub- sequent calculations consisted of 1401 variables (i.e. wavelengths) and 144 observation (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 stan- dard 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 un- supervised 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 T^2 range values (5% level of significance).

Subsequently, OPLS-DA was employed with the aim to build dis- crimination models able to distinguish samples according to their pro- duction 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^{2}X$_{cum}$ 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^{2}Y$_{cum}$ 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 the 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 (Fig. 1). The positions of the negative peaks in the second–derivative spectra match the positions of peaks in the original spectrum (Rinnan, van den 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 (Khodabux, L'Omelette, Jhaumeer-Laulloo, Ramasami, & Rondeau, 2007; Osborne, 2000). The intensive bands at wavelengths between 1600 and 1800 nm depend on C-H and CH_2 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 pro- teins, whose content is also revealed at 1990 and 2180 nm (Aenugu et al., 2011; Cozzolino et al., 2005; Osborne, 2000). 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 de- formation combination (Liu, Zeng, & Sun, 2013; Osborne, 2000).

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 (Arechavala-Lopez et al., 2013; Lenas, Chatziantoniou, Nathanailides, & Triantafillou, 2011; Trocino et al., 2012). 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, Fernández-Segovia, Serra, & Barat, 2010; Lenas et al., 2011). A more-in depth knowledge about spectral differences among the samples was achieved by means of multivariate data analysis.

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 Fig. 2, where in Fig. 2A the observations were highlighted by production method, in Fig. 2B by farming system, and in Fig. 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 (Fig. 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 (Fig. 2B). Particular grouping behaviours did not emerge even according to the provenance (Fig. 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 Fig. 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 (ab- sorption 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.

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 (Fig. 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 (Fig. 4B). W, Ext, and SI/I subjects were well discriminated along the t[1]. Variability be- tween SI and I subjects was collected by the t[2]. Score plot for geo- graphical origins (Fig. 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.

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, Eriksson, & Trygg, 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 (see Fig. 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 avail- able literature. According to what has been suggested by most authors (Alasalvar, Taylor, Zubcov, Shahidi, & Alexis, 2002; Fuentes et al., 2010; Orban, Nevigato, Di Lena, Casini, & Marzetti, 2003; 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 1–3 predictive components and 5–7 orthogonal components (A), that captured 60–63% of the total variation in the X-matrix ($R2X_{cum}$). 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-$R2X_{cum}$), and thus indicative of a high within-class variance among samples. Values related to the predictive variation of X (p-$R2X_{cum}$), 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 ($R2Y_{cum}$), 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 pre- dictive fraction of the variation of the X block. o-R^2X_{cum} = cumulative ortho- gonal 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. No- tably, 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.

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

OPLS-DA model	Class	Single class			Overall model	
		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).

Geographical classification led to 100% of EM samples to be as- signed 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.

4. Conclusions

This work demonstrates the feasibility of a simple and rapid authentication of European sea bass by using NIRS combined with che-mometric 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 (European Parliament and Council of the European Union, 2013). 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 the sample geo- graphical 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 of the approach to enable the authenticity assessment of sea bass; it should be mentioned, how- ever, 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.

Declaration of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https:// doi.org/10.1016/j.foodchem.2018.12.075.

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*(November), 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.

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics, 17*(3), 166–173. https://doi.org/10.1002/cem.785.

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. https://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 va- lidation 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 quanti- tative 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 AQUA- culture production* (2017 ed.). The EU fish Market.

FAO 2005-2018 (2005). *Cultured AQUATIC Species INFORMATION PROGRAMME. DICENTRARCHUS LABRAX. Cultured AQUATIC Species INFORMATION PROGRAMME. Text by BAGNI, M.* Updated 18 February 2005Rome: FAO Fisheries and Aquaculture Department.

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

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.

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://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, 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.

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.

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 Science 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., & Guidastri, R. (1999). *MANUAL on HATCHERY production of SEABASS AND GILTHEAD SEABREAM, Vol. 1*. Rome: FAO.

Naes, T., Isakson, T., Fearn, T., & Davies, T. (2002). *A user-friendly guide to MULTIVARIATE CALIBRATION AND CLASSIFICATION*. Chichester, UK: NIR Publications.

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 aqua- culture 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., van den Berg, F., & 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;2-5.

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 in- frared reflectance spectroscopy (NIRS). *Food Chemistry, 86*(2), 275–281. https://doi. org/10.1016/j.foodchem.2003.09.026.
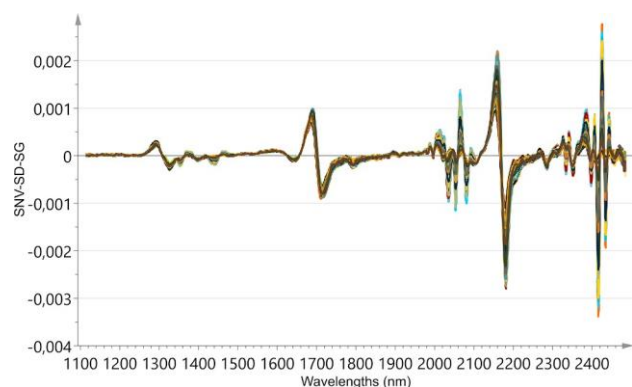
FIGURE

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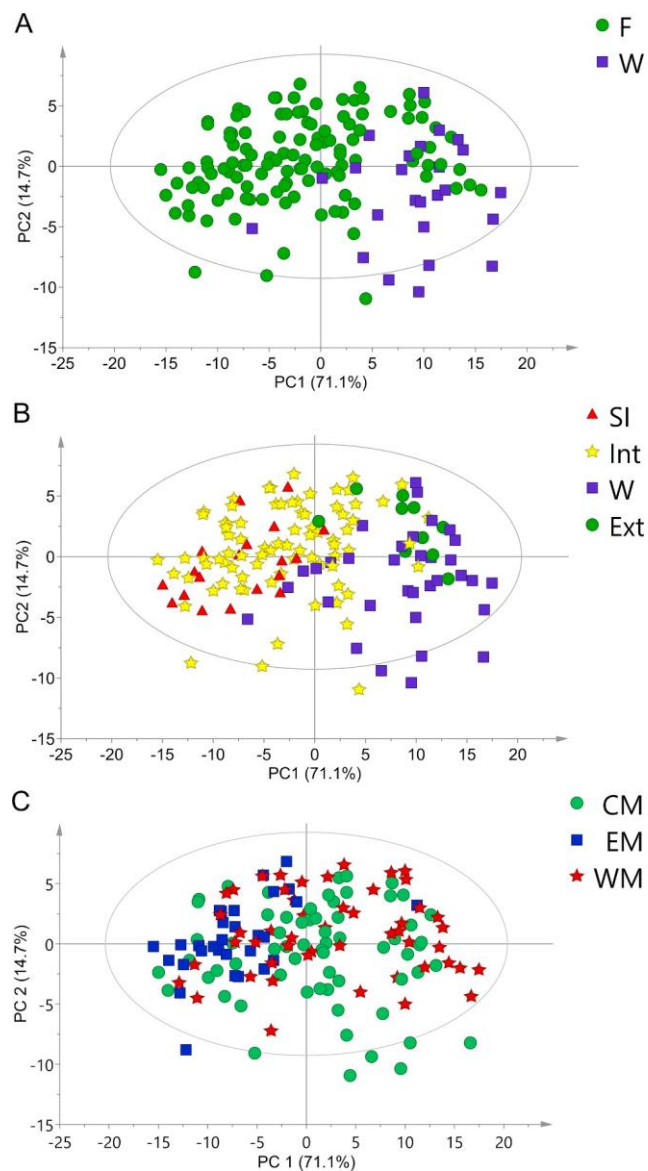

Fig. 2. PCA score plot of the first two principal components (PC1 and PC2), showing sea bass clustering by production method (A), farming system (B), and geographical provenance (C). The ellipse identifies the 95% confidence interval for Hotelling's T^2. For production method: wild (W, square); farmed (F, circles). For farming system: wild (W, squares); extensive (Ext, circles); semi-intensive (SI, triangles); intensive (Int, 5-point stars). For geographical provenance: Central Mediterranean (CM, circles); Eastern Mediterranean (EM, squares); Western Mediterranean (WM, 5-point stars).
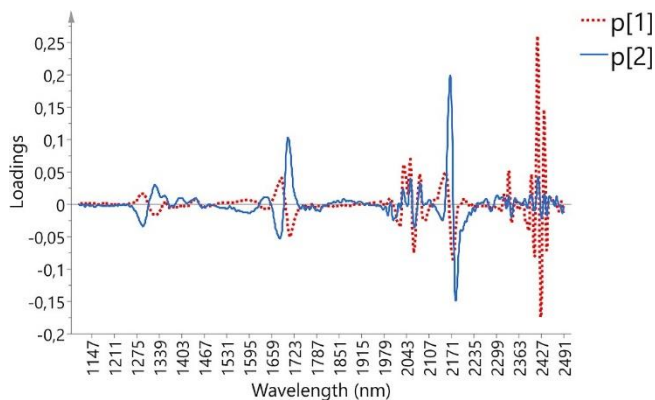
Fig. 3. Loading plot of the first two loading vectors (p1 and p2), showing the influence of NIR wavelengths on the PCA model (p1, dotted line; p2, solid line).
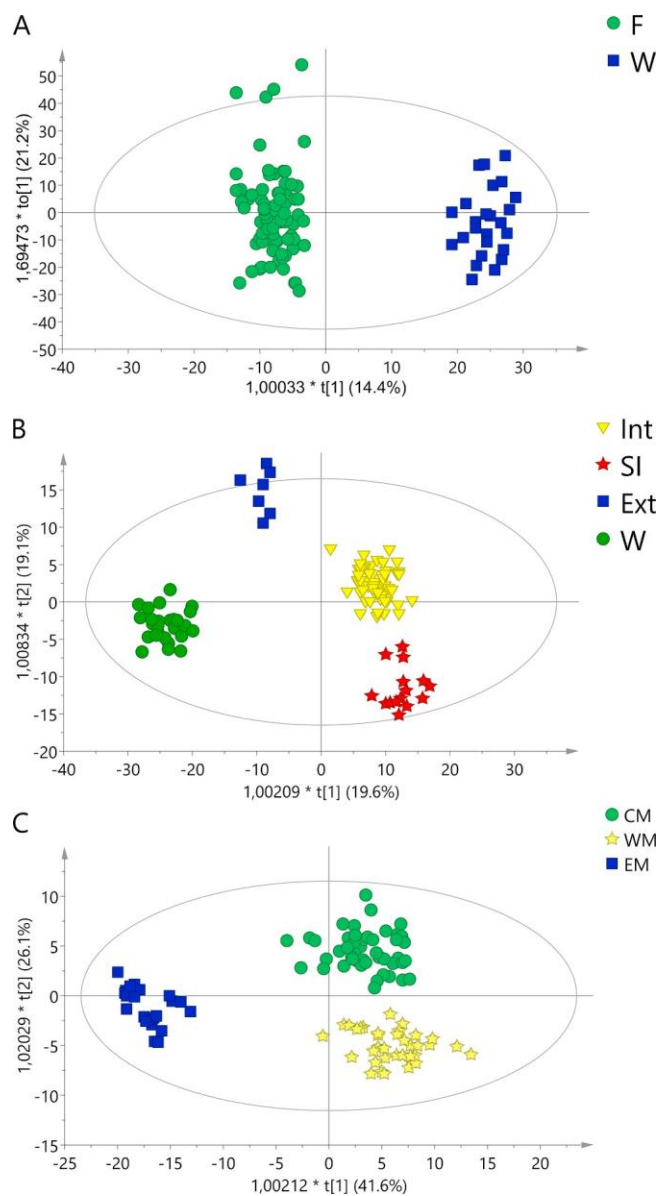

Fig. 4. OPLS-DA score plot for sea bass discrimination based on production method (A), farming system (B), and geographical provenance (C). The ellipse identifies the 95% confidence interval for Hotelling's T^2. For production method: wild (W, squares); farmed (F, circles). For farming system: wild (W, circles); extensive (Ext, squares); semi-intensive (SI, 5-point stars); intensive (Int, triangles). For geographical provenance: Central Mediterranean (CM, circles); Eastern Mediterranean (EM, squares); Western Mediterranean (WM, 5- point stars).