JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

# Use of Near-Infrared Spectroscopy for Fast Fraud Detection in Seafood: Application to the Authentication of Wild European Sea Bass (*Dicentrarchus labrax*)

Matteo Ottavian,[†] Pierantonio Facco,[†] Luca Fasolato,[§] Enrico Novelli,[§] Massimo Mirisola,[⊗] Matteo Perini,[#] and Massimiliano Barolo*,[†]

[†]CAPE-Lab – Computer-Aided Process Engineering Laboratory, Dipartimento di Principi e Impianti di Ingegneria Chimica, Università di Padova, via Marzolo 9, 35131 Padova PD, Italy

[§]Dipartimento di Sanità Pubblica, Patologia Comparata e Igiene Veterinaria e [⊗]Dipartimento di Scienze Animali, Università di Padova, viale dell'Università 16, 35020 Legnaro PD, Italy

[#]IASMA, Fondazione Edmund Mach, via Mach 1, 38010 San Michele all'Adige TN, Italy

Supporting Information

**ABSTRACT:** The possibility of using near-infrared spectroscopy (NIRS) for the authentication of wild European sea bass (*Dicentrarchus labrax*) was investigated in this study. Three different chemometric techniques to process the NIR spectra were developed, and their ability to discriminate between wild and farmed sea bass samples was evaluated. One approach used spectral information to directly build the discrimination model in a latent variable space; the second approach first used wavelets to transform the spectral information and subsequently derived the discrimination model using the transformed spectra; in the third approach a cascaded arrangement was proposed whereby very limited chemical information was first estimated from spectra using a regression model, and this estimated information was then used to build the discrimination model in a latent variable space. All techniques showed that NIRS can be used to reliably discriminate between wild and farmed sea bass, achieving the same classification performance as classification methods that use chemical properties and morphometric traits. However, compared to methods based on chemical analysis, NIRS-based classification methods do not require reagents and are simpler, faster, more economical, and environmentally safer. All proposed techniques indicated that the most predictive spectral regions were those related to the absorbance of groups CH, $CH_2$, $CH_3$, and $H_2O$, which are related to fat, fatty acids, and water content.

**KEYWORDS:** *European sea bass, PLS-DA, wavelet analysis, authentication, near-infrared spectroscopy, fraud detection*

## INTRODUCTION

Assessment of seafood origin is a security measure to protect consumers and avoid fraud. Mandatory information required for a full characterization of the marketed fish (species membership, whether wild or farmed, geographic origin) is regulated by stringent laws in the European Union.[1] Regulatory interventions aim at avoiding mislabeling or substituting wild fish with farmed fish and mitigating risks for the consumer's confidence and health. Therefore, the development of novel analytical technologies, as well as the improvement of the existing ones, can be very helpful to detect fraud in the seafood industry. In particular, discrimination between wild seafood and farmed seafood is of paramount importance to achieve satisfactory quality standards.

Several techniques have been proposed in the past decade to detect the wild/farmed substitution fraud in seafood.[2] A macroscopic examination of fish is of limited value due to the lack of specific targets in terms of body integrity and loss of morphologic traits at sale time. Other types of analysis, such as genomics and proteomics patterns, present limited application because the selection of reliable markers is very difficult for different populations. Currently, the most informative methodology for discriminating between wild and farmed fish is the determination of fatty acids (FAs) fingerprinting and the ratios of isotopes of carbon and nitrogen ($^{13}C/^{12}C$ and $^{15}N/^{14}N$, expressed as $\delta^{13}C$ and $\delta^{15}N$).[3−5] Both fingerprints vary in the muscles according to the season, feeding status, and species, but some specific targets could be adopted as markers of the production system.[3] For example, Alasalvar and co-workers[6] suggested that a high arachidonic amount could be a marker for wild fish. In farmed fish, plant oil intake leads to an increase in C18 FA in muscle lipids, particularly 18:2n-6 (linoleic acid), 18:3n-3 ($\alpha$-linolenic acid), and 18:1n-9 (oleic acid), with the flesh of marine fish retaining these FAs even after a refeeding period with fish oil.[7] However, all of the methods used to characterize FAs require sample preparation for lipid extraction and gas chromatography analysis, which are expensive and time-consuming compared to the shelf life of the fish product.

Several emerging technologies have been proposed for the rapid and nondestructive analysis of fish traceability and authentication,[8−11] such as nuclear magnetic resonance (NMR), front-face fluorescence spectroscopy, and near-infrared spectroscopy (NIRS). Among the most promising techniques,

Received: August 25, 2011
Revised: December 16, 2011
Accepted: December 16, 2011
Published: December 16, 2011

high-resolution nuclear magnetic resonance (HR-NMR) was successfully applied to obtain spectral information on the classification of wild/farmed case, especially on fish lipids.[10,11] This technique provides a fingerprint of FA profiles linked to other characteristics, such as the positional distribution of polyunsaturated FA on triglycerides.[12] HR-NMR spectra combined with different chemometric strategies were used for classification purposes on different species, but this technique is not widely utilized for seafood authentication due to problems in the standardization of the procedures.[2,11,13]

NIRS is particularly favorable because it is simpler, more economical, environmentally safer, and faster than many other techniques. Several applications have been reported that use NIRS for food and beverage authentication, demonstrating versatility and high speed of analysis.[14,15] Recent studies have highlighted the potential of NIRS to differentiate sea bass (*Dicentrarchus labrax*) from different rearing systems.[8,16,17] In these applications, the classification ability among rearing systems seems to decrease according to the storage time, and typically some additional treatments such as freeze-drying are required to improve the accuracy of origin prediction.

In this paper, the possibility of using NIRS to discriminate between wild and farmed sea bass samples was investigated. Three different chemometric approaches were developed to process the available NIR spectra. In the first approach, a cascaded arrangement was proposed whereby chemical information was first estimated from spectra using latent variables regression, and this estimated information was then used to build the discrimination model in a latent variable space. The second approach used spectral information to directly build the discrimination model in a latent variable space. The third approach first used wavelets to transform the spectral information and, subsequently, developed the discrimination model using the transformed spectra. The classification results were compared to a reference classification obtained using only chemical and morphometric information.

## MATERIALS AND METHODS

**European Sea Bass Collection.** Farmed and wild fish were collected in different sales centers and different cities in 2008.[18] Samples were transported to the laboratory within 24 h from the collection time at refrigerated temperature (4 ± 1 °C, constantly monitored by a data logger Testo 174-T, Testo AG, Germany) and were immediately processed for analysis upon their arrival at the laboratory. The data set comprised 38 calibration samples with determined attribution of production method and 66 validation samples with declared methods of production (32 declared wild and 34 declared farmed). Compared to the data set presented in ref 18, two samples were removed because the corresponding spectra had not been collected. The same study (which was used as a classification reference in the present work) showed that the number of samples classified as farmed among those declared wild was 22 (corresponding to 69% of substitution fraud) and, at the same time, 6 samples among those declared farmed were formally ascribed to the wild group (i.e., misclassified).

A total of 35 chemical properties (fatty acids, bromatological, and isotopes) and morphometric traits were measured for each available sample. These variables are listed in Table 1 (the meaning of the VIP index will be clarified under Statistical Analysis) and will be collectively identified as "chemistry" variables in the remainder of the paper.

**Sample Preparation and NIR Analysis.** After dissection, the epiaxial white muscle portion of the fillet was ground with a Retsch Grindomix (Retsch GmbH, Hann, Germany) at 4000 rpm for 10 s. Two aliquots per sample (approximately 10 g each) were placed in a 50 mm diameter ring cup and scanned in reflectance mode at 2 nm intervals from 1100 to 2500 nm using a scanning monochromator NIRSystem 5000 (FOSS, Silver Spring, MD).

**Table 1. Measured Chemical Properties and Morphometric Traits**

| property no. | property name |
|---|---|
| 1 | fat |
| 2 | protein |
| 3 | ash |
| 4 | moisture |
| 5 | C14:0 |
| 6 | C16:0 |
| 7 | C18: 0 |
| 8 | Σsaturated |
| 9[a] | C16:1 n-7 |
| 10 | C18:1 n-9 |
| 11 | C18:1 n-7 |
| 12 | C20:1 n-9 |
| 13[a] | C22:1 n-11 |
| 14[a] | C22:1 n-9 |
| 15 | Σmonounsaturated |
| 16 | C18:2 n-6 |
| 17[a] | C18:3 n-6 |
| 18 | C18:3 n-3 |
| 19 | C20:2 n-6 |
| 20[a] | C20:3 n-6 |
| 21 | C20:3 n-3 |
| 22 | C20:4 n-6 |
| 23 | C20:5 n-3 |
| 24 | C22:5 n-3 |
| 25 | C22:6 n-3 |
| 26[a] | Σpolyunsaturated |
| 27 | Σn-3 |
| 28 | Σn-6 |
| 29 | n-3/n-6 |
| 30 | EPA + DHA |
| 31[b] | $\delta^{13}$C |
| 32 | $\delta^{15}$N |
| 33[c] | KI |
| 34[c] | HSI |
| 35[c] | CFI |

[a]Not significant according to the VIP index (VIP < 0.5). [b]Measured from fat-free extract. [c]Morphometric traits (KI , condition index; HSI, hepatosomatic index; CFI, celomatic fat index, cf. ref 18).

**Spectra Pretreatment.** For each aliquot of a sample, a mean spectrum was obtained by averaging from 32 multiple scans; then, the spectrum of the sample was obtained by averaging those of the two aliquots. Reflectance ($R$) values were converted into absorbance ($A$) values through $A = \log(1/R)$. Mathematical pretreatment reduced the light scattering caused by the sample particles and removed the additional variation in baseline shift typically present in diffused reflectance spectra. Standard normal variate and first- and second-order derivates were used to this purpose. More details on spectra pretreatment are available as Supporting Information.

**Statistical Analysis.** Principal component analysis (PCA)[19] was applied as a preliminary exploratory tool to reveal the internal structure of the available data and to check whether the validation samples could be described by the model developed for the calibration samples. After this preliminary analysis, four different strategies for sea bass classification were developed and tested. These strategies are sketched in Figure 1, in which the acronyms used to identify the proposed models are also indicated. Figure 1 clarifies that the strategies differed for the input information as well as for the type of classification model.

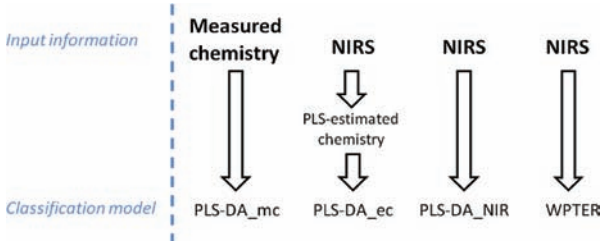


**Figure 1.** Schematic of the classification strategies considered in this study.

As far as the model inputs are concerned, either chemistry variables or NIR spectra were used. Note that, with respect to the chemistry inputs, not only the *measured* properties but also the properties *estimated* from NIRS were used as inputs to the classification model. When NIR spectra were used as inputs directly, two alternative modeling approaches were investigated: partial least-squares discriminant analysis (PLS-DA)[20] and the wavelet-based WPTER (wavelet packet transform for efficient pattern recognition) method.[21]

*Exploratory Analysis of the Available Samples.* PCA returned a compact representation of the raw data and highlighted the most predictive variables (chemistry variables or wavelengths, according to model input) through the combined use of the model scores and loadings plots. Samples that did not conform to the PCA model exhibited values of the model residuals $Q$ largely exceeding an assigned confidence limit (95%). Classification of samples based on inputs that showed very large values of the $Q$ statistic was therefore considered to be unreliable.

*Authentication Using Measured Chemistry Variables.* PLS-DA applied to the measured chemistry variables (PLS-DA_mc model) was used to discriminate between the samples belonging to the wild sea bass class and those belonging to the farmed class.

After the PLS-DA_mc model was built using the full measured chemistry calibration data set, the variable importance in projection (VIP) index[22] was evaluated to get an indication of the relative importance of each chemical variable within the discrimination model. The VIP index for the $j$th input variable of a PLS-DA model is calculated as

$$\mathrm{VIP}_j = \sqrt{\frac{\sum_{f=1}^{F} w_{jf}^2 \times \mathrm{SSY}_f \times J}{\mathrm{SSY}_{\mathrm{total}} \times F}} \qquad (1)$$

where $w_{jf}$ is the weight value for component $f$ of variable $j$, $\mathrm{SSY}_f$ is the sum of squares of explained variance for the $f$th component, $J$ is the number of variables, $\mathrm{SSY}_{\mathrm{total}}$ is the sum of squares explained of the dependent variable, and $F$ is the total number of components.

Because variables with VIP > 1 are of greater importance, whereas those with VIP < 1 are progressively less important for the model,[22] a more parsimonious PLS-DA_mc model could be developed using as inputs only those chemistry variables for which VIP > 1. Besides leading to a more robust discrimination model, this procedure can reduce the number of chemistry variables that need to be measured for any new sample requiring authentication.

*Authentication Using Estimated Chemistry Variables.* A much faster and cheaper methodology was developed by using a cascaded approach, whereby a PLS regression model was first used to *estimate* the chemistry data from a NIR spectrum, and then a PLS-DA model used the estimated chemistry variables as inputs to classify the sample (PLS-DA_ec model). In this study, a cross-validatory procedure[23] minimizing the root-mean-squared error of cross-validation was used to choose the structure of the estimation model (i.e., number of latent variables to be retained). The PLS model performance was evaluated by means of the coefficient of determination in calibration ($R^2_{\mathrm{calib}}$) and cross-validation ($R^2_{\mathrm{cv}}$), together with the ratio of prediction to deviation (RPD). The inputs to the PLS-DA_ec model were limited to chemistry variables that could be estimated with accuracy sufficient for sample classification.

*Authentication with Direct Use of Spectral Data.* Following several studies in a variety of food classification examples, PLS-DA was applied directly to NIR spectra to obtain sea bass classification (PLS-DA_NIR model).

In principle, to reduce the number of model inputs (wavelengths), the VIP index could have been used as was done for the chemistry variable inputs in the PLS-DA_ec model. However, finding a procedure for variable selection on the basis of the VIP index was harder for spectra due to the very high correlation between subsequent wavelengths, and on some occasions this led to important wavelengths being missed. An alternative approach was therefore considered to build a parsimonious classification model. This approach used the wavelet-based WPTER algorithm for sample classification.[21] Because of its ability to analyze a signal at different resolution scales, wavelet analysis[24] has been receiving increasing attention for agriculture and food quality inspection.[25] Due to its complexity, a detailed description of the WPTER algorithm is not reported here; the reader should refer to ref 21 for further information. Here it suffices to say that, for each class (wild or farmed) to be discriminated, WPTER first provided a mean reconstructed signal where only the most influential wavelengths exhibit a nonzero signal. Then, classification of a sample is obtained by comparing its reconstructed NIR signal to the class mean signals.

## RESULTS

**Exploratory PCA Models.** The scores plot of the first two principal components (PCs) of a 4-PC model on the full set of measured chemistry variables of the calibration data set (Figure 2) clearly identified two clusters (open triangles vs solid triangles), which include the farmed and wild sea bass samples, respectively. Separation between the clusters occurred along PC1 (PC1 > 0 for wild samples). When projected onto the PCA model, the validation samples (circles) separated into the same clusters. However, quite a number of declared wild samples (solid circles) fell within the farmed sea bass cluster (open symbols); that is, a number of substitution frauds were highlighted in the scores plot. Although not reported here, it was verified that all validation samples fell well below the 95% confidence limit of the $Q$ statistic, which indicated that, from the point of view of the chemistry variables, the validation data set completely conformed to the calibration one.

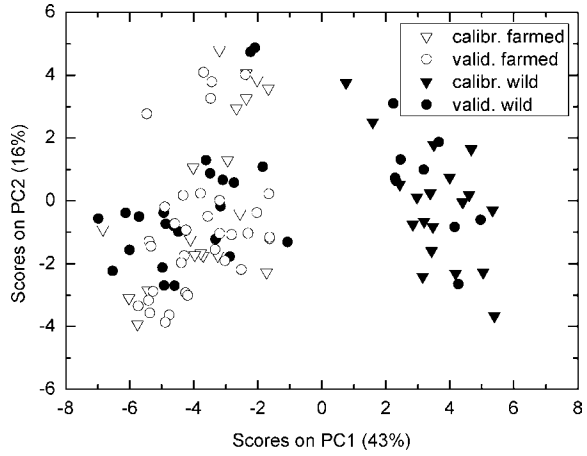


**Figure 2.** Exploratory PCA on the measured chemistry properties of Table 1. The validation samples (circles) are projected onto the model defined by the calibration samples (triangles).

Although not included here, the combined analysis of scores and loadings of the PCA model revealed that the farmed samples were characterized by higher levels of fat, protein, ash, celomatic fat index (CFI), hepatosomatic index (HIS), and,


dx.doi.org/10.1021/jf203385e | J. Agric.Food Chem. 2012, 60, 639–648

among the fatty acids listed in Table 1, C18:3 n-6, C18:2 n-6, C18:1 n-9, Σn-6, C20:3 n-3, Σmonosaturated, C20:2 n-6, and C14:0. The wild samples, instead, were characterized by higher levels of moisture, $\delta^{13}C$, and $\delta^{15}N$, and, among fatty acids, Σsaturated, C16:0, C18:1 n-7, C18:0, C20:4 n-6, EPA + DHA, Σn-3, and C22:5 n-3. These results were in agreement with other studies on sea bass.[3,6]

A similar exploratory analysis was carried out on NIR spectra. Figure 3 shows the mean spectra for the farmed and wild

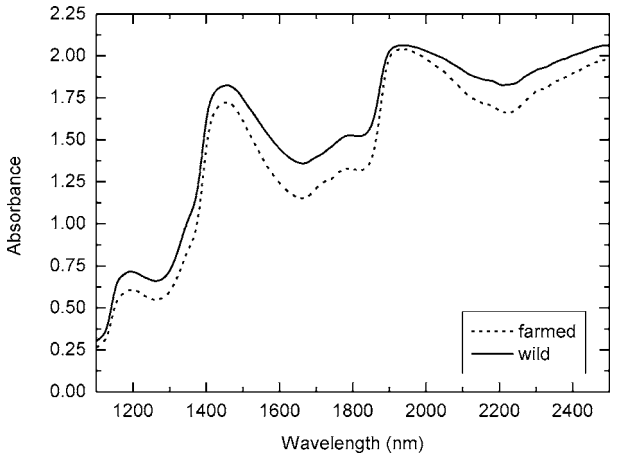


**Figure 3.** Mean raw spectra for the farmed and wild classes (calibration data set).

classes of the calibration data set. The scores plot and model residuals *Q* for the exploratory PCA model (using five PCs) built on the NIR spectra of the calibration data set are shown in Figure 4. Figure 4a indicates that the first PC is not able to separate the two classes, although it explains a much higher fraction (72 vs 43%) of the data variability than in the PCA model derived on the measured chemistry data. Furthermore, some of the declared farmed samples exhibited very high model residuals (rightmost open circles in Figure 4b), indicating that the corresponding NIR spectra did not conform to those of the calibration data set. These samples were assigned the IDs 29–34. Although a detailed discussion of these samples is beyond the purpose of this work, it is worth noting that the high model residuals indicate that these samples explored a variability on NIR spectra that is different from that described by the calibration data set.

**Authentication Using Measured Chemistry Variables (PLS-DA_mc).** A PLS-DA model using one latent variable was built using the full measured chemistry calibration data set to discriminate between farmed and wild sea bass samples. The VIP index is plotted for the most influential (VIP > 1) inputs of this model in Figure 5. Here, the input variables were classified

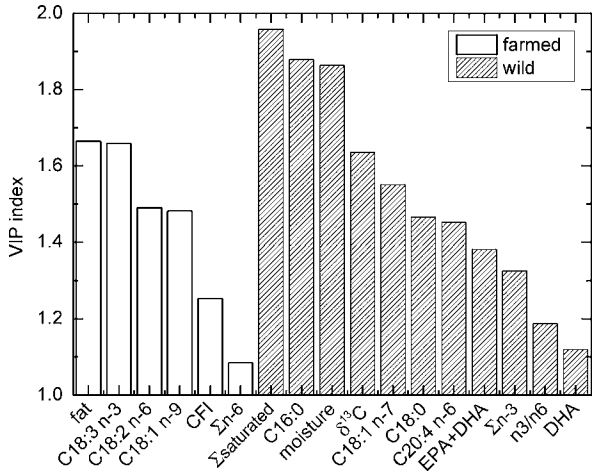


**Figure 5.** Most influential (VIP > 1) chemistry variables for the PLS-DA_mc classification model according to the VIP index.

as belonging to the farmed class or to the wild one according to the indications provided by the exploratory PCA model loadings. In terms of meaningfulness of the input variables for each class, the results of Figure 5 agreed with those obtained using a nonparametric permutation test to analyze the same data set.[18] Furthermore, the variables deemed as totally not significant by the VIP index (VIP < 0.5, see Table 1) were the same as those discarded by the nonparametric test.

As mentioned earlier, a more parsimonious model can be developed by using a subset of the chemistry variables having

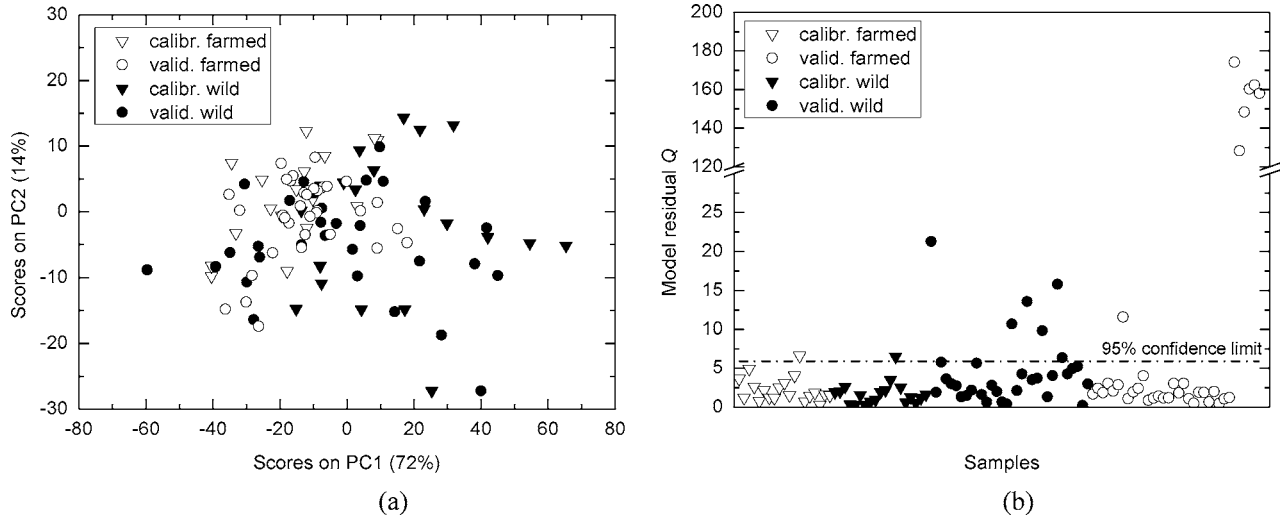


**Figure 4.** Exploratory PCA on NIR spectra: (a) scores plot; (b) model residuals. The validation samples (circles) are projected onto the model defined by the calibration samples (triangles).


dx.doi.org/10.1021/jf203385e | J. Agric.Food Chem. 2012, 60, 639–648

VIP > 1 as inputs to the model. To this purpose, a subset including only 3 (namely, fat, moisture, and $\delta^{13}C$) out of the 17 the variables shown in Figure 5 was selected. The rationale for this selection was as follows. First, according to VIP, the selected variables were among the most discriminating ones of the two classes; second, these variables could be estimated with sufficient accuracy from NIR spectra (to be discussed later). A PLS-DA model (one latent variable) using only these three measured chemistry variables as inputs was therefore developed for sea bass authentication; this model was denoted PLS-DA_mc. Although this model had the same classification ability of a PLS-DA model using the full data set of measured chemistry, reducing the number of inputs from 35 to 3 significantly reduced the distance between the clusters in the scores plot of the PCA exploratory model (Figure S1 in the Supporting Information). This means that the observable variability among samples was reduced when the number of model inputs was reduced. However, the variability described by the PCA model still allowed the two classes to be separated.

A comparison between the reference classification results provided by Fasolato and co-workers[18] and those obtained using the PLS-DA_mc model is shown in Table 2 for declared wild samples and in Table 3 for declared farmed samples. The two approaches detected the same total number of substitution frauds (Table 2), although with one sample classification shift (samples 6 and 12). On the other hand, the PLS-DA_mc model did not misclassify any of the declared farmed samples (Table 3).

**Authentication Using Estimated Chemistry Variables (PLS-DA_ec).** Fat, moisture, and $\delta^{13}C$ (i.e., the inputs to the PLS-DA_mc model) were estimated from NIR spectra using three distinct PLS models. The estimation results are shown graphically in Figure 6, whereas the estimation model characteristics and performance metrics are reported in Table 4. The models developed for fat and moisture returned accurate estimations even when extrapolated outside the calibration range, whereas the model for stable carbon isotope estimation was less accurate.

A PLS-DA model (PLS-DA_ec) with one latent variable was built using the estimated chemistry variables as inputs. Although the estimated value of $\delta^{13}C$ was not very accurate, it was used as an input to the PLS-DA_ec model because the other chemistry variables with VIP greater than that of the $\delta^{13}C$ (namely, C18:3 n-3, Σsaturated, and C16:0, cf. Figure 5) were estimated even worse (with RPDs of 1.07, 1.20, and 1.05, respectively). The classification results are reported in Tables 2 and 3. Despite the fact that one of the model inputs was not estimated as accurately as the other two from the NIR spectra, for the declared wild samples the classification results obtained using the PLS-DA_ec model totally agreed with those obtained using the PLS-DA_mc model, whereas for declared farmed samples only two misclassifications were obtained.

**Authentication with Direct Use of Spectral Data (PLS-DA_NIR and WPTER).** NIR spectra were used directly as inputs to a PLS-DA discrimination model (PLD-DA_NIR). Table 2 shows that the ability of this model to identify the substitution fraud was very good. Three misclassified farmed samples were present (Table 3); however, note that two of them (samples 29 and 30) did not conform to the calibration data set due to high values of the $Q$ residuals, and therefore the use of spectral information for these samples was questionable.

As was noted earlier, the dimension of the calibration input data set may be extremely large when spectra are used as model

**Table 2. Declared Wild Samples (Validation Data Set): Comparison between the Reference Classification Results (Reference 18) and the Performance of the Classification Methods Considered in This Study[a]**

| | sample ID | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
| ref 18 | | | | | | + | + | + | | | + | | | | | | + | | | | | | | + | + | | + | | + | | | + |
| PLS-DA_mc | | | | | | | + | + | | | + | + | | | | | + | | | | | | | + | + | | + | | + | | | + |
| PLS-DA_ec | | | | | | | + | + | | | + | + | | | | | + | | | | | | | + | + | | + | | + | | | + |
| PLS-DA_NIR | | | | | | | + | + | | | + | + | | | | | + | | | | | | | + | + | | + | | + | | | + |
| WPTER | | | | | | | + | + | | | + | + | | | | | + | | | | + | | | | + | | + | | + | | | + |

[a]For each classification method, "+" indicates a sample classified as wild. No symbol is used if a sample is declared wild but is classified as farmed (substitution fraud).


dx.doi.org/10.1021/jf203385e | J. Agric.Food Chem. 2012, 60, 639–648

Table 3. Declared Farmed Samples (Validation Data Set): Comparison between the Reference Classification Results (Reference 18) and the Performance of the Methods Considered in This Study[a]

| | sample ID | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 |
| ref 18 | | | | | | | | X | | | | | | | | | X | | | | | | | | | | | | X | | X | | X | X |
| PLS-DA_mc | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| PLS-DA_ec | | | | | | | | | | | | | | | | | | | | | X | | | | | | X | | | | | | | |
| PLS-DA_NIR | | | | | | | | | | | | | | | | | | | | | | | | | | | X | | X[b] | X[b] | | | | |
| WPTER | | | | | | | | | | | | | | | | | | | | | | | | | | | | | X[b] | X[b] | | X[b] | | |

[a]For each classification method, "X" indicates a misclassified sample. No symbol is used if a sample is declared farmed and is classified as farmed. [b]Sample with high PCA residual.

inputs. For this reason, an attempt to reduce this dimension by extending the use of the VIP index to spectra was first attempted (Figure 7), but this approach did not lead to satisfactory results in terms of the discrimination ability of the resulting PLS-DA model. We conjecture that this can be ascribed to the much higher correlation existing between the model inputs (i.e., wavelengths) than in the measured chemistry variables case.

As an alternative approach for selective pruning of the input data set and sample classification, the WPTER algorithm[21] was used (details on the implementation of the algorithm are available as Supporting Information). Figure 8, which was obtained with a WPTER model with Daubechies-2 wavelets, shows the mean reconstructed signals for both the farmed class and the wild class. WPTER unambiguously showed that the only influential spectral regions for sample classification were those within the range ∼1600−1750 nm and at ∼2200 nm. These regions somewhat overlapped those with a VIP index much larger than 1 in Figure 7. However, note that, due to the reconstruction operation, the actual input signal used by WPTER for sample classification was different from that used by a PLS-DA model using the same intervals of wavelengths as inputs.

The classification results of WPTER were in very good agreement with those provided by the other methods considered in this work. Three misclassified farmed samples were present (Table 3), but they referred to validation samples not conforming to the calibration data set of spectral signals (IDs 29, 30, and 32).

## DISCUSSION

**NIRS-Based Approaches: Selected Wavelengths.** One advantage of using the PLS-DA_NIR model (with VIP index) or the WPTER one was the indication of the most influential wavelengths for sample classification. Both classification models pointed approximately to the same regions, namely, between 1600 and 1800 nm and around 2200 nm (2172−2226 nm). Moreover, the VIP index of Figure 7 revealed also an important region at 1200 nm and a peak at 1900 nm (1930−1938 nm).

The carbon−hydrogen (CH) stretch second overtone was represented at 1202 nm, and thus the region around 1200 nm was related to the absorbance of CH, $CH_2$, and $CH_3$ groups.[26] In the region around 1700 nm, first-overtone stretch bonds of groups CH, $CH_2$ (1722 and 1760 nm), and $CH_3$ were represented; these peaks were especially related to the lipid content of the samples.[26−28] The region around 2200 nm was characterized by CH and $CH_2$ combination bands, which could be related to fatty acids, protein, and peptide groups.[29] Eventually, the peak at 1900 nm was specific to the O−H bond, and it was related to the water content of the sample.[27]

The selection of the wavelengths related to the absorbance of fat, fatty acids, and water was in good agreement with the selection of the chemistry variables obtained through the VIP index in the PLS-DA_mc model (cf. Figure 5).

**Chemistry-Based Approaches and Estimation of Chemistry Variables.** The analysis carried out using the full set of measured chemistry variables proved to be consistent with the results reported in the literature, particularly in terms of variable importance within the classification model (Figure 5). However, this study demonstrated that the VIP index could be used to reduce the number of input variables to a PLS-DA classification model to only three, namely, fat, moisture, and $\delta^{13}C$. Although the lipid and moisture contents are known to

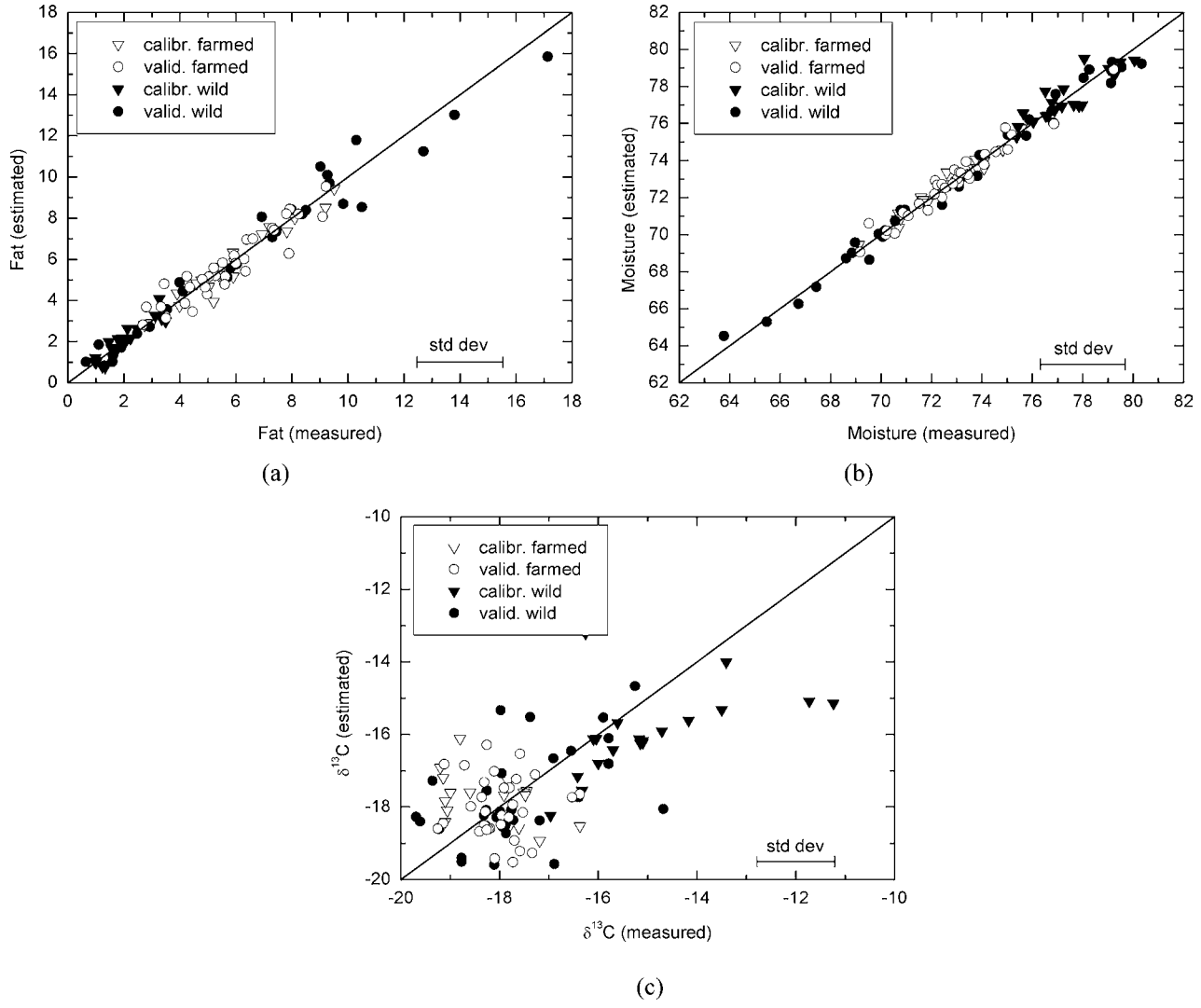


**Figure 6.** Estimated versus measured values of the three chemistry variables used as inputs to the PLS-DA_ec classification model: (a) fat; (b) moisture; (c) $\delta^{13}$C. The standard deviation (std dev) of the measured values is also indicated.

**Table 4. Number of Latent Variables (LVs) Retained in Each PLS Model for the Estimation of Fat, Moisture, and $\delta^{13}$C and Mean Performance of the Estimation Models**[a]

| | no. of LV | $R^2_{calib}$ | $R^2_{cv}$ | $R^2_p$ | RPD |
|---|---|---|---|---|---|
| fat | 7 | 0.98 | 0.97 | 0.97 | 5.69 |
| moisture | 9 | 0.99 | 0.98 | 0.98 | 6.66 |
| $\delta^{13}$C | 9 | 0.67 | 0.42 | 0.45 | 1.25 |

[a] $R^2_p$ indicates the coefficient of determination in the estimation (i.e., for samples belonging to the validation group).

vary inversely depending on the size of the animal[30] and the feeding regimen,[31,32] in marketable sea bass their simultaneous use was not redundant. The correlation between these variables was indeed high, but the variables were not perfectly collinear; therefore, the use of a correlative technique such as PLS-DA was appropriate in this respect. It must be stressed that the final selection of the variables to be used as model inputs was highly guided by the need to reduce the number of inputs of the PLS-DA_ec model. Other authors, in fact, proposed individual fatty acids such as linoleic acid (18:2 n-6),[3] arachidonic acid (20:4 n-6), and the lipid content[33] as tracers of the farmed/wild status.

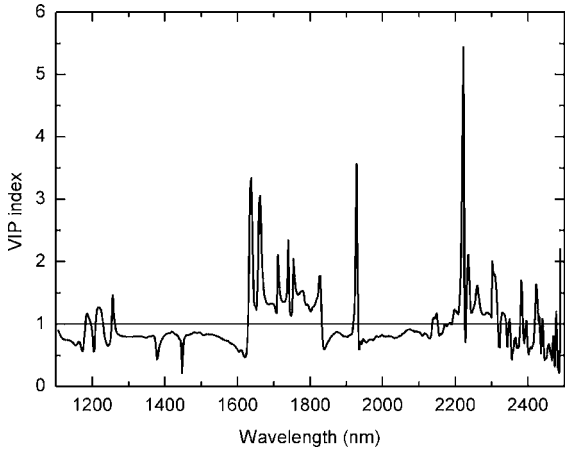


**Figure 7.** VIP index for the PLS-DA_NIR model.

In terms of predictability of each chemistry variable from NIR spectra, results were consistent with those presented in the literature. An RPD value >3, as observed for fat (Table 4), indicated the possibility of obtaining an accurate quantitative estimation.[34] In fact, the value of $R^2_p$ obtained in the estimation


dx.doi.org/10.1021/jf203385e | J. Agric.Food Chem. 2012, 60, 639–648

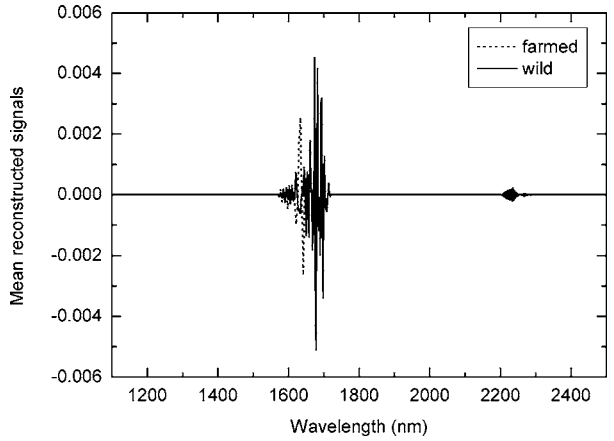


**Figure 8.** Mean reconstructed signals (in the wavelet domain) according to the WPTER algorithm. Nonzero signals indicate the wavelengths selected by the WPTER algorithm.

of fat (Table 4) was comparable to values reported in ref 35. Moreover, the estimation performances were similar to those reported for different fish species[8,36] or in intact samples using NIRS.[37] The quantitative estimation of water can be considered excellent as well.[34] Some studies reported NIR analysis as a surrogate method to evaluate the carbon isotope signature in vegetables.[38,39] The performance obtained in this study on sea bass muscles was slightly lower (with a poor RPD). The restricted variability in the calibration data set (Figure 6c) and the different application analyzed (fish, and not plant) could be reasonable explanations for the results of Table 4. In any case, the $\delta^{13}C$ estimation accuracy was shown to be adequate for the purpose of discriminating the wild and farmed populations.

Furthermore, although the regions of absorbance of FAs were responsible for classification, their predictability using a PLS model was poor: for FAs of the representative classes of FAs listed in Table 1, typical RPD values were found to range from 0.9 to 1.4. The low content of a single FA in the meat and especially in sea bass fillet[18,35] should be considered the main explanation for the poor estimations of the FA profiles. Furthermore, as suggested by Realini and co-workers[27] for ground beef, poor estimations of some individual FAs could be caused by the lack of specific absorption peaks, with a similar spectral profile related to the same absorbing molecular group ($-CH_2-$).

**General Comments on the Proposed Classification Methods.** As shown in Table 2, the classification of declared wild samples was very similar (with few exceptions) among all of the techniques analyzed in this work, considering both those based on chemistry data (PLS-DA_mc, PLS-DA_ec and ref 18) and those based on spectral data (PLS-DA_NIR and WPTER). On the contrary, the classification of declared farmed samples (Table 3) exhibited fewer differences from the one in ref 18, in which the chemistry data set was correlated with a profile type generated with data from the literature. It should be noted that the comparison with a class profile might be a source of misclassification if, as a result of the farming condition, the composition of a farmed fish was close to the one of a wild fish.

Because the classifier of ref 18 was used only as a reference for the classification results of the NIRS-based ones, this work showed that NIR spectroscopy is a reliable tool for quickly detecting substitution fraud in marketable sea bass. Particularly, WPTER seemed to be the most informative way to derive a sea bass classification model. In fact, WPTER performed not only a correct classification of the samples but also unambiguously highlighted the most informative wavelength ranges for sample classification. One drawback of this method was that it was the hardest to implement among those considered in this work. The VIP index applied to a PLS-DA_NIR model provided similar results in terms of wavelength importance detection, although this information was somewhat less clearly outlined than in the WPTER (i.e., no direct wavelength selection could be carried out). The PLS-DA_ec cascaded arrangement was an interesting hybrid (in terms of required inputs) approach that showed the same discrimination capability of the other approaches, providing the additional advantage of a quick estimation of fat, moisture, and $\delta^{13}C$.

When spectra were used as inputs, a preliminary PCA assessment was very useful to determine whether or not the sample spectrum conformed to the calibration data set. A high PCA model residual for the sample under investigation indicated that the classification obtained using NIRS was not fully reliable for that sample.

In conclusion, the results illustrated in this work clearly showed that NIRS can be very effective in assessing the authenticity of wild European sea bass. Classification results obtained from NIR information (whether using PLS-DA_ec, PLS-DA_NIR, or WPTER) were almost identical to that obtained from the combined use of chemical properties and morphometric traits.

Due to the minimal processing of the fish sample and the ease of detection of NIR spectra, NIRS-based sample classification requires negligible processing time and is therefore particularly suitable for real-time, cost-effective applications.

## ■ ASSOCIATED CONTENT

### Supporting Information

Figure S1 and Table S1. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*Phone: +39.049.827.5473. Fax: +39.049.827.5461. E-mail: max.barolo@unipd.it.

### Funding

Financial support of this work has been provided by CARIPARO under Project PARO104725 (2010).

## ■ ABBREVIATIONS USED

FA, fatty acids; HR-NMR, high-resolution nuclear magnetic resonance; NIRS, near-infrared spectroscopy; NMR, nuclear magnetic resonance; PC, principal component; PCA, principal component analysis; PLS-DA_ec, partial least-squares discriminant analysis_estimated chemistry; PLS-DA_mc, partial least-squares discriminant analysis_measured chemistry; PLS-DA_NIR, partial least-squares discriminant analysis_near-infrared; RPD, ratio of prediction to deviation; VIP, variable importance in projection; WPTER, wavelet packet transform for efficient pattern recognition.

## ■ REFERENCES

(1) Commission Regulation (EC) No. 2065/ 2001 of October 2001 laying down detailed rules for the application of Council Regulation

(EC) No. 104/2000 as regards informing consumers about fishery and aquaculture products.
(2) Martinez, I.; Stendhal, I.; Aursand, M.; Yamashita, Y.; Yamashita, M. Chapter 14: Analytical Methods to Differentiate Farmed from Wild Seafood. In *Handbook of Seafood and Seafood Products Analysis*, 1st ed.; Nollet, L. M. L., Toldrá, F., Eds.; CRC Press: Boca Raton, FL, 2009; pp 215−232.
(3) Bell, J. G.; Preston, T.; Henderson, R. J.; Strachan, F.; Bron, J. E.; Cooper, K.; Morrison, D. J. Discrimination of wild and cultured European sea bass (*Dicentrarchus labrax*) using chemical and isotopic analyses. *J. Agric. Food Chem.* **2007**, *55*, 5934−5941.
(4) Martinez, I. Revision of analytical methodologies to verify the production method of fish. In *Seafood from Fish to Dish, Quality, Safety and Processing of Wild and Farmed Fish*, 1st ed.; Luten, J. B., Jacobsen, C., Bekaert, K., Sæbo, A., Oehlenschlager, J., Eds.; Academic Publishers: Wageningen, The Netherlands, 2006; pp 541−550.
(5) Morrison, J.; Preston, T.; Bron, J. E.; Henderson, R. J.; Cooper, K.; Strachan, F.; Bell, J. G. Authenticating production origin of gilthead sea bream (*Sparus aurata*) by chemical and isotopic fingerprinting. *Lipids* **2007**, *42*, 537−545.
(6) Alasalvar, C.; Taylor, K. D. A.; Zubcov, E.; Shahidi, F.; Alexis, M. Differentiation of cultured and wild sea bass (*Dicentrarchus labrax*): total lipid content, fatty acid and trace mineral composition. *Food Chem.* **2002**, *79*, 145−150.
(7) Montero, D.; Robaina, L.; Caballero, M. J.; Ginés, R.; Izquierdo, M. S. Growth, feed utilization and flesh quality of European sea (*Dicentrarchus labrax* L.) fed diets containing vegetable oils: a timecourse study on the effect of a re-feeding period with a 100% fish oil diet. *Aquaculture* **2005**, *248*, 121−134.
(8) Xiccato, G.; Trocino, A.; Tulli, F; Tibaldi, E. Prediction of chemical composition and origin identification of european sea bass (*Dicentrarchus labrax*) by near infrared reflectance spectroscopy (NIRS). *Food Chem.* **2004**, *86*, 275−281.
(9) Arvanitoyannis, I. S.; Tsitsika, E. V.; Panagiotaki, P. Implementation of quality control methods (physicochemical, microbiological and sensory) in conjunction with multivariate analysis towards fish authenticity. *Int. J. Food Sci. Technol.* **2005**, *40*, 237−263.
(10) Masoum, S.; Malabat, C.; Jalali-Heravi, M.; Guillou, C.; Rezzi, S.; Rutledge, D. N. Application of support vector machines to $^{1}$H NMR data of fish oils: methodology for the confirmation of wild and farmed salmon and their origins. *Anal. Bioanal. Chem.* **2007**, *387*, 1499−1510.
(11) Rezzi, S; Giani, I.; Héberger, K.; Axelson, D. K.; Moretti, V. M.; Reniero, F.; Guillou, C. Classification of gilthead sea bream (*Sparus aurata*) from $^{1}$H NMR lipid profiling combined with principal component and linear discriminant analysis. *J. Agric. Food Chem.* **2007**, *55*, 9963−9968.
(12) Aursand, M.; Jørgensen, L.; Grasdalen, H. Positional distribution of w3 fatty acids in marine lipid triacylglycerols by high-resolution $^{13}$C nuclear magnetic resonance spectroscopy. *J. Am. Oil Chem. Soc.* **1995**, *72*, 293−297.
(13) Forshed, J.; Schuppe-Koistinen, I.; Jacobsson, S. P. Peak alignment of NMR signals by means of a genetic algorithm. *Anal. Chim. Acta* **2003**, *487*, 189−199.
(14) Martinez, I.; Aursand, M.; Erikson, U.; Singstad, T. E.; Veliyulin, E.; van der Zwaag, C. Destructive and non-destructive analytical techniques for authentication and composition analyses of foodstuffs. *Trends Food Sci. Technol.* **2003**, *14*, 489−498.
(15) Manley, M.; Downey, G.; Baeten, V. Spectroscopic technique: near-infrared (NIR) spectroscopy In *Modern Techniques for Food Authentication*, 1st ed.; Sun, D. W., Ed.; Elsevier/Academic Press: Boston, MA, 2008; pp 65−115.
(16) Majolini, D.; Trocino, A.; Xiccato, G.; Santulli, A. Near infrared reflectance spectroscopy (NIRS) characterization of European sea bass (*Dicentrarchus labrax*) from different rearing systems. *Ital. J. Anim. Sci.* **2009**, *8*, 860−862.
(17) Costa, C.; D'Andrea, S.; Russo, R.; Antonucci, F.; Pallottino, F.; Menesatti, P. Application of non-invasive techniques to differenziate sea bass (*Dicentrarchus labrax*, L. 1758) quality cultured under different conditions. *Aquacult. Int.* **2010**, *19*, 765−778.
(18) Fasolato, L.; Novelli, E.; Salmaso, L.; Corain, L.; Camin, F.; Perini, M.; Antonetti, P.; Balzan, S. Application of nonparametric multivariate analysis to the authentication of wild and farmed European sea bass (*Dicentrarchus labrax*). Results of a survey on fish sampled in the retail trade. *J. Agric. Food Chem.* **2010**, *58*, 10979−10988.
(19) Jackson, J. E. *A User'S Guide to Principal Component*; Wiley: New York, 1991.
(20) Barker, M.; Rayens, W. Partial least squares for discrimination. *J. Chemom.* **2003**, *17*, 166−173.
(21) Cocchi, M.; Seeber, R.; Ulrici, A. WPTER: wavelet packet transform for efficient pattern recognition of signals. *Chemom. Intell. Lab. Syst.* **2001**, *57*, 97−119.
(22) Andersen, C. M.; Bro, R. Variable selection in regression − a tutorial. *J. Chemom.* **2010**, *24*, 728−737.
(23) Wold, S. Cross-validatory estimation of number of components in factor and principal components models. *Technometrics* **1978**, *20*, 397−405.
(24) Addison, P. S. *The Illustrated Wavelet Transform Handbook*; Taylor and Francis: New York, 2002.
(25) Singh, C. B.; Choudhary, R.; Jayas, D. S.; Paliwal, J. Wavelet analysis of signals in agriculture and food quality inspection. *Food Bioprocess Technol.* **2010**, *3*, 2−12.
(26) McClure, W. F.; Stanfield, D. L. Near-infrared spectroscopy of biomaterials In *Handbook of Vibrational Spectroscopy*; Chalmers, J. M., Griffiths, P. R., Eds.; Wiley: Chichester, U.K., 2002; Vol. *1*, pp 212−225.
(27) Realini, C. E.; Duckett, S. K.; Windham, W. R. Effect of vitamin C addition to ground beef from grass-fed or grain-fed sources on color and lipid stability, and prediction of fatty acid composition by near-infrared reflectance analysis. *Meat Sci.* **2004**, *68*, 35−43.
(28) Osborne, B. J. Near-infrared spectroscopy in food analysis. In *Encyclopedia of Analytical Chemistry*; Wiley: Chichester, U.K., 2000.
(29) Cozzolino, D.; Murray, I. Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. *Lebensm. Wiss. Technol.* **2004**, *37*, 447−452.
(30) Poli, B. M.; Parisi, G.; Zampacavallo, G.; Mecatti, M.; Lupi, P.; Gualtieri, M.; Franci, O. Quality outline of European sea bass (*Dicentrarchus labrax*) reared in Italy: shelf life, edible yield, nutritional and dietetic traits. *Aquaculture* **2001**, *202*, 303−315.
(31) Lanari, D.; Poli, B. M.; Ballestrazzi, R.; Lupi, P.; D'Agaro, E.; Mecatti, M. The effects of dietary fat and NFE levels on growing European sea bass (*Dicentrarchus labrax L.*): growth rate, body and fillet composition, carcass traits and nutrient retention efficiency. *Aquaculture* **1999**, *179*, 351−364.
(32) Grigorakis, K. 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* **2007**, *272*, 55−75.
(33) Sales, J. Quantification of the differences in flesh fatty acid components between farmed and wild fish. *J. Aquat. Food Prod. Technol.* **2010**, *19*, 298−309.
(34) Mouazen, A. M.; Saeys, W.; Xing, J.; De Baerdemaeker, J.; Ramon, H. Near infrared spectroscopy for agricultural materials: an instrument comparison. *J. Near Infrared Spectrosc.* **2005**, *13*, 87−97.
(35) Weeranantanaphan, J.; Downey, G.; Allen, P.; Sun, D. W. A review of near infrared spectroscopy in muscle food analysis: 2005−2010. *J. Near Infrared Spec.* **2011**, *19*, 61.
(36) Sollid, H. E.; Solberg, C. H. Salmon fat content estimation by near infrared transmission spectroscopy. *J. Food Sci.* **1992**, *57*, 792−793.
(37) Khodabux, K.; L'Omelette, M. S. S.; Jhaumeer-Laulloo, S.; Ramasami, P.; Rondeau, P. Chemical and near-infrared determination of moisture, fat and protein in tuna fishes. *Food Chem.* **2007**, *102*, 669−675.
(38) Ferrio, J. P.; Bertran, E.; Nachit, M.; Royo, C.; Araus, J. L. Near infrared reflectance spectroscopy as a potential surrogate method for

the analysis of $\Delta^{13}C$ in mature kernels of durum wheat. *Aust. J. Agric. Res.* **2001**, *52*, 809−816.

(39) Kleinebecker, T.; Schmidt, S. R.; Fritz, C.; Smolders, A. J. P.; Hölzel, N. Prediction of $d^{13}C$ and $d^{15}N$ in plant tissues with near-infrared reflectance spectroscopy. *New Phytol.* **2009**, *184*, 732−739.