JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Detection and Quantification of Species Authenticity and Adulteration in Crabmeat Using Visible and Near-Infrared Spectroscopy

JAVIER GAYO AND SCOTT A. HALE*

Department of Biological and Agricultural Engineering, North Carolina State University, Box 7625, Raleigh, North Carolina 27695-0001

Seafood processing often removes morphological properties of seafood species that enable the consumer to distinguish one type of organism from another. For this reason, species substitution is the most common form of economic adulteration in the seafood industry. Visible and near-infrared spectroscopy (Vis/NIR) has been used to detect and quantify species authenticity and adulteration in crabmeat samples. Atlantic blue crabmeat was adulterated with blue swimmer crabmeat in 10% increments. Water absorption bands dominated the main features in the crabmeat spectra, with a decrease in sample absorbance with increasing adulteration percentage. Several data pretreatments, i.e., moving average, combing, first and second derivatives, and multiplicative scatter correction, in addition to the raw data, were investigated for prediction and quantitative data analysis using partial least-squares. In addition, quantitative analysis was done using the full spectrum and a sequential approach in which 50 wavelengths were added sequentially to determine a new model and find an optimal solution. The results suggest that Vis/NIR spectroscopy is a suitable technology that can be applied to detect and quantify species authenticity and adulteration in crabmeat.

KEYWORDS: Species authenticity; adulteration; VIS/NIR spectroscopy; PLS

INTRODUCTION

The issue of food authenticity has been around since the early 1800s and is mainly related to improper labeling and economic adulteration (EA), i.e., the substitution, in part or whole, of cheaper and inferior food products for high-cost foods in order to defraud the consumer (*1*). An authentic food, defined as a food that "conforms to the description provided by the producer or processor," includes the process history of a product or ingredient, its geographic region of origin, or the species or variety of the ingredient (*2*). Although rarely a health hazard, EA is driven by the demand for higher value goods, global trading, and price fluctuations, factors that provide an opportunity for illegal profits. For these reasons, food processing industries and regulatory agencies have pushed for analytical methods to confirm food product authenticity (*3*).

The development of worldwide high-seas fishing vessels, the improvement in food processing and storage, and the establishment of fishing industries in developing countries have increased the variety of seafood species, both fresh and processed, currently available in markets (*4*). These factors have contributed to an increase in total catches from fisheries and, thus, seafood consumption worldwide. The demand for a year-round seafood supply, however, has negatively impacted the number of some valued and appreciated species due to exploitation. Therefore, some have turned to illegal practices in order to meet the high demand for these valued and appreciated seafood products.

Because most consumers are not very familiar with the taxonomical and morphological characteristics of seafood species, such as skin pattern, body appearance and size, eyes, shape, and number of fins, they are subject to being defrauded by buying a seafood product that is not what it claims to be (*5*). In addition, the processing of seafood products, which often requires the removal of significant morphological characteristics, hinders species recognition. Because of these reasons, species substitution has become the main form of adulteration in the seafood-processing industry.

Detection of food authenticity, by focusing on food adulteration, has traditionally relied on wet chemistry analyses by determining the amount of compounds in a food product and comparison of the values obtained with known, i.e., previously documented, values for authentic products (*2*). Early methods for species identification have relied on using proteins as species markers. The field of electrophoresis, for example, made it possible to obtain water-soluble protein patterns, which have become a reference method for species identification to differentiate genetically related fish species (*6*). However, because of protein denaturation at high temperatures, these techniques are not effective in determining species authenticity for processed seafood products (*4*). Additional disadvantages include the large range of compounds needing quantification, discrepancies in protein patterns among members of the same species,

* To whom correspondence should be addressed. Tel: 919-515-6760. Fax: 919-515-6719. E-mail: andy_hale@ncsu.edu.

10.1021/jf061801+ CCC: $37.00
Published on Web 01/09/2007

and time-consuming and expensive wet chemistry techniques. These drawbacks, plus the fact that food adulterers are applying more sophisticated techniques to adulterate their food products, have increased the application of suitable technologies into the field of food adulteration and authenticity.

The need for fast, reliable, and on-line methods to detect species authenticity and adulteration has increased interest in the application of the use of spectroscopic research. Visible and near-infrared spectroscopy (Vis/NIR) is a noninvasive and nonspecific technique that has been used to measure spectra of different types of foods (*7*). Vis/NIR spectroscopy can be used to distinguish among biological samples through spectra describing the sample's biological composition, such as fat, protein, lipid, water, etc. In addition, it has been used to identify different species although research has focused on meat products (*8*).

The aim of this paper was to investigate the application of Vis/NIR spectroscopy to address the issue of species authenticity and adulteration in crabmeat. The use of several spectral data pretreatments, i.e., moving average smoothing, first and second derivatives, combing, and multiplicative scatter correction (MSC), was explored to determine the effects on model performance. In addition, two spectral approaches were studied to build the calibration models: a full spectrum and a sequential spectrum in which 50 wavelengths were added sequentially to the previous model to determine the best optimal solution.

MATERIALS AND METHODS

Sample Preparation. Six pounds of canned crabmeat (3 lbs. per type of crab) was obtained from a local supermarket for two species of crabs, Atlantic blue (*Callinectes sapidus*) and blue swimmer (*Portunus pelagicus*), and stored at 4 °C overnight. The imported blue swimmer crabmeat was chosen as the adulterant due to its year-round availability and reduced cost. In addition, it is the most prevalent crabmeat being imported into the United States (*9*).

Prior to sample preparation and analysis, the crabmeat was tempered individually to room temperature (25 °C) in a water bath, pooled in a separate clean container, and thoroughly mixed with gloved hands. Samples represented authentic crabmeat species, i.e., Atlantic blue and blue swimmer, and a range of adulterated samples (10−90%) containing different amounts of both crab species in 10% increments according to weight. The crabmeat of *C. sapidus* and *P. pelagicus* was weighed individually and homogenized using a blender (5 s intervals) to obtain a total sample weight of 70 g. Homogenized samples were then divided into two equal parts, each consisting of 35 g, generating a total of ten samples for each of the 11 classes, i.e., class 0, 100% Atlantic blue (0% blue swimmer); class 1, 90% Atlantic blue (10% blue swimmer); and so forth until class 10, 0% Atlantic blue (100% blue swimmer). The 110 samples were placed in labeled polyethylene bags and stored at room temperature (25 °C) until spectral analysis.

Vis/NIR Spectroscopy Analysis. Absorbance spectra (400−2498 nm at 2 nm intervals) were recorded in log(1/*R*) units, for a total of 1050 wavelengths, using a NIRSystems 6500 spectrometer (FOSS NIRSystems, Silver Springs, MD) equipped with a rectangular sample cell. Crabmeat samples were scanned in random order at room temperature (25 °C), and the spectrum, an average of 32 scans, was recorded per sample to obtain a total of 110 spectra. Spectral analysis, model development, calibration, and validation were performed using The Unscrambler, software version 7.6 (CAMO Software, Inc., OR).

Spectral Data Pretreatment Methods. Because spectral data often contain noise and extra information irrelevant to the problem at hand, an appropriate model is necessary to extract the relevant information for the prediction of the response variable, i.e., adulteration percentage in this study. Several spectral pretreatment methods—derivatives, combing, smoothing, and MSC—have been used and compared in this paper to assess the best pretreatment and regression model combination for determining species authenticity. **Table 1** shows the different data pretreatments and parameters used in the convolution intervals for each algorithm.

Table 1. Data Pretreatments and Associated Parameters

data pretreatment	window segment or algorithm
first derivative	1, 5, 15, 30, 50
second derivative	1, 5, 15, 30, 50
combing	2, 4, 16, 32, 64
smoothing	none, 5, 15, 30
MSC	common offset, common amplification, full MSC

Derivatives. First and second derivatives were used to reduce peak overlap and remove constant and linear baseline drift, respectively (*7*). Differentiation was done using the Savitzky−Golay algorithm in which a moving average was applied to the spectra prior to differentiation. An appropriate window segment for the moving average was very important; the wider the window segment was, the greater the noise reduction although also the greater the distortion of the signal (*10*). Five windows composed of 1, 5, 15, 30, and 50 consecutive wavelengths were used to compute the moving average prior to obtaining the polynomial approximation for differentiation.

Data Combing. Combing was used to choose a user-defined number of data points at equal intervals from each spectrum for subsequent data analysis (*11*). Five combing intervals (2, 4, 16, 32, and 64 points wide) were chosen to examine the effect on model performance. Each data set contained 525, 263, 66, 33, and 17 data points, respectively.

Smoothing. Smoothing modified the magnitude of absorption peaks and shifted the position of asymmetric absorption bands (*11*). A moving average was used to replace each wavelength spectral data with an average of adjacent values. Four convolution intervals (none, 5, 15, and 30 points wide), applied to each spectrum using a moving average, were chosen to compare the effects on model predictability.

MSC. MSC compensated for additive (offset) and/or multiplicative (amplification) effects in Vis/NIR spectroscopy and reduced the likelihood that these effects were dominating factors in the spectral data (*7*). Three types of corrections were investigated in this paper: common offset (additive effects), common amplification (amplification effects), and full MSC (additive and multiplication effects).

Model Development: Calibration and Validation. Multivariate calibration related two data sets, X (containing the independent variables, i.e., spectral data) and Y (containing the dependent variable, i.e., adulteration percentage), via regression with the purpose of using the model for prediction. Validation, on the other hand, was used to test the model's prediction ability on a new data set, which has not been used in the model development.

For each model, the crabmeat samples were divided equally into a training set (55 samples) for calibration and a testing set (55 samples) for validation. Two different approaches were investigated in each data pretreatment to determine, if any, which method produced optimal models [i.e., lowest standard error of calibration (SEC) and standard error of prediction (SEP)]. The full spectrum approach used the complete wavelength range, i.e., 400−2498 nm (1050 data points). In the sequential approach, a 100 nm window was added to the previous consecutive spectrum window to determine a new model. In other words, the first model was developed based on the 400−500 nm wavelength range. The remaining models were developed by adding a 100 nm wavelength range at a time, i.e., 50 wavelengths at a time, so the second model corresponded to 400−600 nm, the third corresponded to 400−700 nm, etc., until model 20, which used the full spectral data range, i.e., 400−2498 nm.

There has been much debate as to the importance of finding those few wavelengths that contain significant information for optimal model development, thus reducing the number of wavelengths, variables, and model complexity. Recently, however, research has found the importance of combining some wavelengths, i.e., synergistic, although not necessarily significant by themselves, to those containing problem-dependent information, i.e., descriptive wavelengths, to improve model performance (*12*). The spectral data range was broken down in this sequential fashion in order to observe wavelength relevance on model

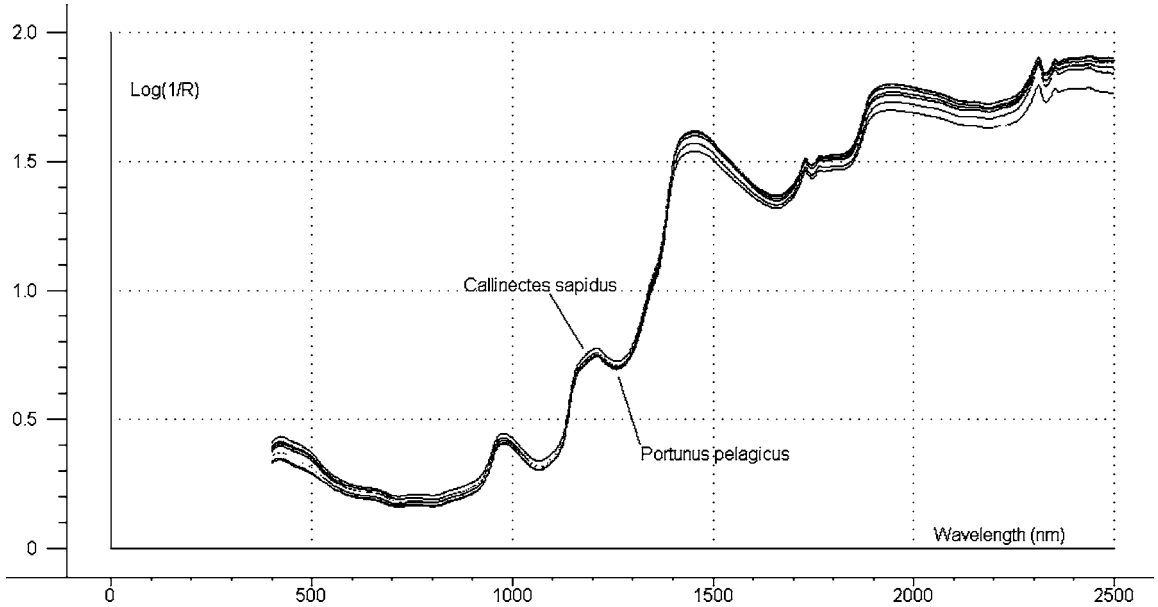

Figure 1. Average absorbance spectra of crabmeat samples. The top spectrum represents class 0 (100% Atlantic blue crabmeat), the bottom spectrum represents class 10 (100% blue swimmer crabmeat), and classes 1–9 (10–90% adulteration) are sequentially from top to bottom.

prediction and investigate synergistic and descriptive relationships among consecutive wavelength regions.

Partial Least-Squares (PLS). PLS was used instead of principal component regression because it focused on the dependent variables and used fewer latent variables to reach an optimal solution (*13*). The term latent variable, principal component (PC), or PLS loadings is often used synonymously in the literature to define the features used to extract the relevant information and reduce the dimensionality of the data. For the purpose of this paper, the extracted features will be referred to as PCs. The effects of the data pretreatment methods on the performance of PLS calibration models were evaluated in terms of the SEC, the SEP, and the corresponding coefficient of determination (r^2).

RESULTS AND DISCUSSION

Spectra of Crabmeat Species. The absorbance spectra for *C. sapidus* and *P. pelagicus* crabmeat samples (**Figure 1**) are dominated by water absorption bands at 970 (O−H bond stretching and second water overtone), 1450 (O−H bond stretching and first water overtone), and 1940 nm (O−H bond stretching and bending). The water content of biological samples, such as crabmeat, poses a limitation in the use of Vis/NIR spectroscopy because water absorbs strongly and contributes to a significant amount of light scattering (*14*). In addition, water absorption bands can interfere with the spectral features of the chemical parameter of interest, such as an adulterant. Overall, however, a higher absorbance is visible for samples containing 100% Atlantic blue crabmeat, whereas the lowest absorbance values represent samples containing 100% blue swimmer crabmeat. The sample absorbance decreases at every wavelength as the percentage of adulteration increases.

Determination of PCs in Model Development. PLS focuses on the Y matrix (the expected values) to decompose the X matrix (the spectral data). Usually, the first few PCs will describe the majority of the variation found in Y; however, this does not guarantee that these first PCs contain the problem-dependent information needed to accurately predict the desired constituent, i.e., adulteration level, in this study. PC analysis decomposes the data along directions of maximum variances. Directions of large variance in the first several components usually correspond to structure, i.e., problem-dependent information, and the directions of small variance in later components correspond mainly to noise. If the full set of PCs is used, there is no clear distinction between the structure part and the noise.

The number of PCs to use in the PLS model is very important because too few components will generate an underfitted model, i.e., fits loosely the data structure (*15*). Using too many, on the other hand, generates an overfitted model, one which fits parts of the noise of the calibration set, thus generating a low SEC but performing poorly in the validation set. The optimum number of PCs will then decompose the X matrix between the structure and the noise. For this reason, evaluation of the variance plots is needed to determine which PCs describe most of the residual variance of the Y matrix in order to determine the optimal number of PCs to use in the regression model.

Even though the calibration variance is a measure of the model fit, i.e., how well the model fits the training data, it may not be useful to rely on it solely to determine the optimal number of PCs. To ensure that the model is able to describe and predict new data, the validation variance must also be taken into account. **Figure 2** shows the residual calibration and validation variances of the untreated data. Observation of both variances indicates that the first minimum is visible using five PCs; hence, for this particular model, the optimum number of PCs used was five. The PCs for the remainder models were chosen similarly by focusing on the Y-variance plot of the calibration and validation data sets.

Quantification of Species Authenticity. Ideally, a correlation must exist between the set of variables measured instrumentally and the property to be estimated. This can be studied using a correlogram, a tool used to determine wavelength importance in terms of a given attribute (*16*). The correlogram describing the correlation between the spectral range and adulteration percentage is seen in **Figure 3**.

There is a high negative correlation ($-0.95 < R^2 < -0.80$) in the visible range (400−800 nm), and it slowly decreases to a plateau around zero correlation in the higher NIR wavelengths (1900−2500 nm). It can be assumed that using the visible wavelength range alone might be effective at detecting or quantifying adulteration. However, the sequential approach

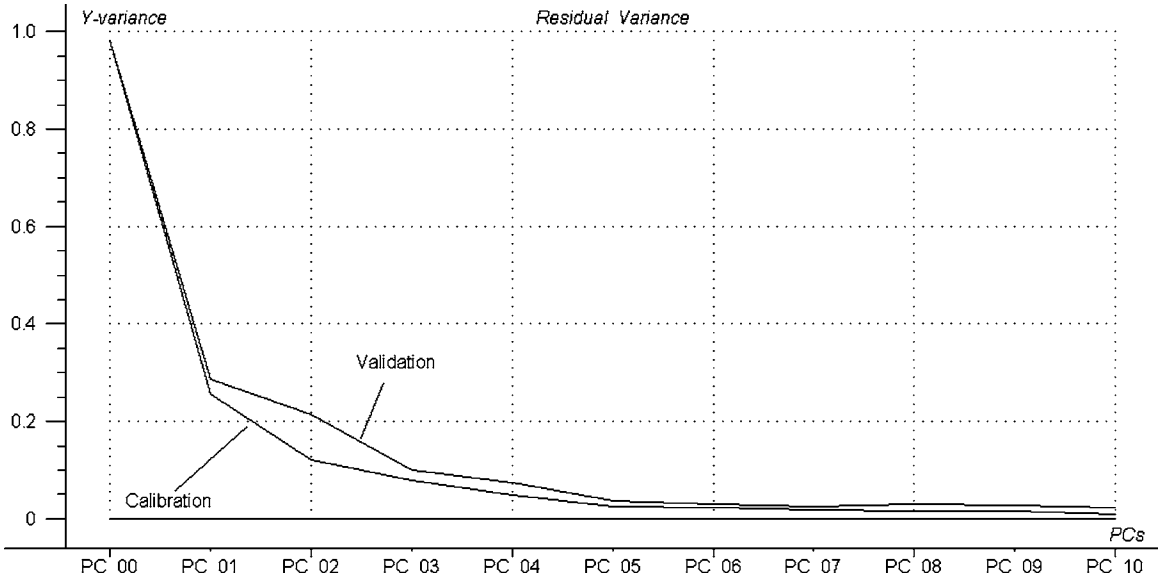

Figure 2. Residual variance of the calibration and validation data sets as a function of PLS components.

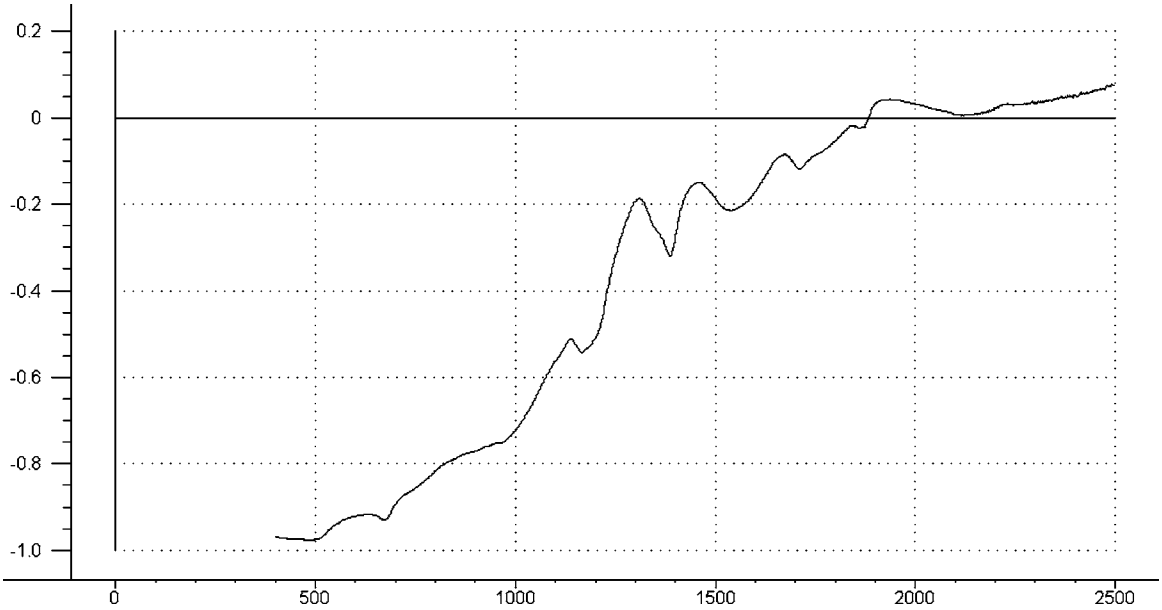

Figure 3. Correlogram showing the correlation (*y*-axis) between adulteration percentage and Vis/NIR spectral data (*x*-axis).

results, based on the visible range, do not yield the lowest SEC and SEP values. Even though this demonstrates the futility of using univariate methods for complex databases, it is a quick way to potentially determine important wavelengths for the prediction of EA of crabmeat.

Wavelength data can be classified into two categories: (i) predictive, in which the information is useful for modeling the relationship between the spectral data (X) and the response variable (Y), and (ii) synergistic, in which the information contained within the wavelengths does not improve the model but adding them to descriptive wavelengths enhances the model's predictive ability (*12*). Interestingly, there is a slight increase in the correlation coefficient at the 1400 nm wavelength ($R^2 = -0.3$). Although small, this increase in correlation can be attributed to the presence of either synergistic or descriptive wavelengths in the 1300−1400 nm region.

The results, based on SEC and SEP, of all of the models of the different data pretreatments were compared. The best model, in terms of SEC and SEP, for each data pretreatment method was then compared with the results of the untreated data (**Figures 4** and **5**). The smoothing pretreatment did not improve model performance when comparing the SEC and SEP values to those gathered from the untreated data. Using the whole spectrum, the best results of the smoothing, those gathered using a 5-point moving average, generated an SEC and SEP of 5.45 and 5.85, respectively, whereas the untreated data generated 5.43 and 5.84. A 2*SEP is regarded as a 95% confidence interval in spectral quantitative analysis (*15*, *17*). Therefore, these models were able to detect adulteration using blue swimmer crabmeat within ±5.85% for the moving average and ±5.84% for the raw data. Similar results were generated using the sequential spectrum approach.

Contrary to studies involving MSC, none of the methods (common offset, common amplification, and full MSC) improved model performance over the untreated data. This can be attributed to the fact that MSC is usually performed on samples that are not homogeneous and contain particles of different sizes. Because the crabmeat samples were blended, thus reducing the chances of light-scattering effects due to particle size, it is possible that this data pretreatment actually

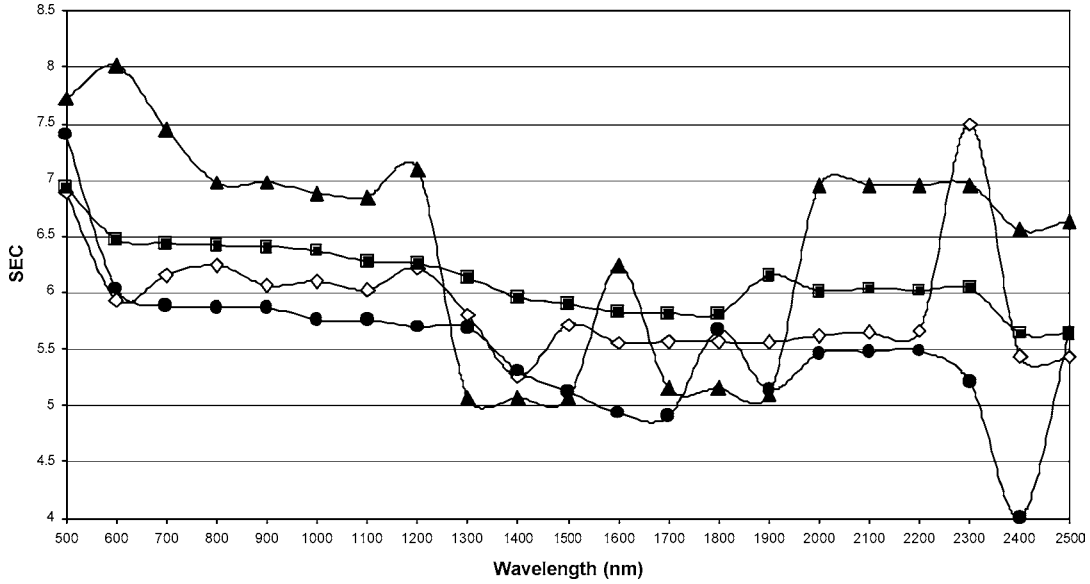

Figure 4. SEC of the best model for each data pretreatment, i.e., 64-point combing (▲), 30-point first derivative (■), 5-point second derivative (●), and the untreated (◇) data.

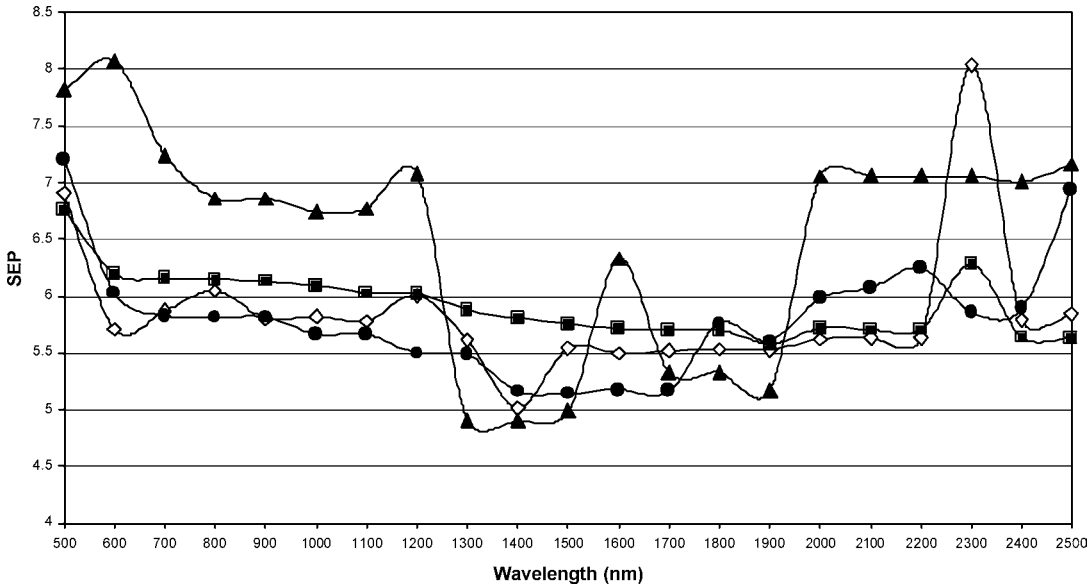

Figure 5. SEP of the best model for each data pretreatment, i.e., 64-point combing (▲), 30-point first derivative (■), 5-point second derivative (●), and the untreated (◇) data.

Table 2. Optimal Convolution Intervals for Remaining Data Pretreatments

data pretreatment	optimal convolution interval
first derivative	30 points
second derivative	5 points
combing	64 points

increased the signal-to-noise ratio. The best SEC and SEP, 5.83 and 6.23, respectively, was gathered by using a common offset MSC. Because these data pretreatments performed equally (for smoothing) and worse (for MSC), further discussion of the results will focus on the untreated data and the remaining data pretreatment methods, i.e., first and second derivative and combing (**Table 2**).

Figure 4 shows the SEC, in the sequential spectral approach, of the remaining data pretreatment methods and the untreated data across the sequential adding of spectra in 100 nm increments. In general, the SEC of the models ranges from 4.00 (5-point second derivative of a model using the 500−2400 nm range), to 8.0 (64-point combing at 500−600 and untreated data at 500−2300). Choosing an appropriate data pretreatment is problem-dependent and, as verified in **Figure 4**, depends on the wavelength range used in the model calibration. If the complete spectrum is used, both the 30-point first derivative and the 5-point second derivative perform equally in terms of SEC (5.64) and slightly higher than the untreated data SEC (5.43). The lowest SEC (4.00) across the spectrum was generated using a 5-point second derivative and a wavelength range of 400−2400 nm. With the exception visible in the 2200−2400 range, the untreated data generated a lower SEC than the 30-point first derivative data and equal to the 5-point second derivative data in the 700−1300 and 2000−2200 ranges.

Upon closer inspection of **Figure 4**, several regions are distinguishable according to the SEC values. The SEC of the untreated, first derivative, and second derivative data is initially between 7.0 and 7.5 and slowly decreases as more wavelengths are added sequentially to determine the next regression model.

Table 3. Model Performance for Full and Partitioned Spectrum Wavelength Data for Each Data Pretreatment at Optimal Convolution Intervals[a]

		full spectrum			partitioned spectrum			
data pretreatment	data range	PCs	SEC(r2)	SEP(r2)	data range	PCs	SEC(r2)	SEP(r2)
none	400–2500	6	5.43 (0.985)	5.88 (0.983)	400–1400	5	5.26 (0.986)	5.02 (0.988)
first derivative	**400–2500**	**3**	**5.64 (0.984)**	**5.64 (0.984)**				
second derivative	400–2500	4	6.63 (0.978)	7.16 (0.975)	**400–1700**	**3**	**4.91 (0.988)**	**5.17 (0.987)**
64-point combing	400–2500	5	5.64 (0.984)	6.94 (0.979)	400–1400	6	5.07 (0.987)	4.90 (0.988)

[a] The best models are in bold.

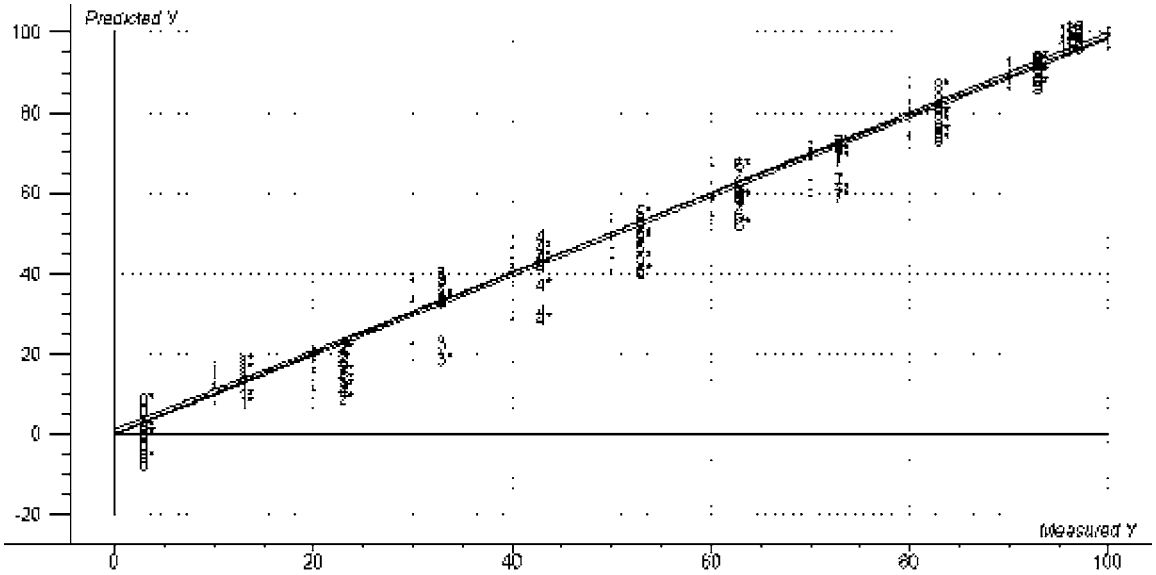

Figure 6. Actual vs predicted adulteration content of PLS using a 5-point second derivative data. Validation samples are shown with an asterisk (*).

Models developed with wavelengths of two regions, i.e., 400−1300 and 400−2200 nm, generate an SEC between 5.5 and 6.5 and 5.4 and 6.0, respectively. Models generated by including 1400−1900 nm produced an SEC between 4.8 and 5.8, and those generated after including the 2200 nm region produced varying SEC values (between 4.0 and 7.5).

It is interesting to note that the 64-point combing data pretreatment, although considerably higher in the 500−1200 and 2000−2500 nm ranges, has a minimum SEC of 5.0 in the 1300−1500 nm region, a region that coincides with an increase in negative correlation visible in **Figure 3**. Similarly, the untreated data also have a minimum SEC of 5.0 at 1400 nm. In addition, the combing data pretreatment shows a second low SEC value in the 1700−1900 nm region. This phenomenon could help explain the importance of synergic wavelengths in model development.

The SEP of the models using the testing set is shown in **Figure 5**. The SEP of the 64-point combing data pretreatment shows a similar trend to that seen in **Figure 4**. Overall, the model's performance, in terms of SEP, is higher than any other pretreatment with the exception of the models generated using wavelength ranges from 400 to 1300/1500 and even 1800/1900. The second derivative data give a lower SEP in the 400−1200 data range than the first derivative data but perform worse than the first derivative in the models that include the 2000−2500 wavelength data.

Table 3 shows the best model performance in terms of SEC, SEP, and coefficient of determination for each data pretreatment in addition to the untreated data. Utilizing the full spectral wavelength data, i.e., 400−2500 nm, the 5-point first derivative data give the best model performance in terms of SEC (5.64), SEP (5.64), and the number of factors (*3*) used in the model. Even though the untreated data generate a lower SEC, a higher number of factors are required, and there is a slight increase in the SEP. Therefore, the true adulteration of a sample, using a 5-point second derivative data pretreatment, would be predicted to be within ±5.64%.

The best model for the sequential spectral approach, however, is generated by using the second derivative on the 400−1700 data range, not only in terms of SEC and SEP but also in terms of a number of factors (**Table 3**). The first derivative data did not have a better model performance using a sequential approach so the best SEC and SEP were generated using the full spectrum. This exception aside, models using a partitioned spectrum generated lower SECs and SEPs than those using the full spectrum. Two of the models, i.e., untreated and 64-point combing, gave the optimal performance when including data contained in the 1300−1400 range, whereas the 5-point second derivative gave the optimal performance when including data in the 1600−1700 range. This phenomenon suggests that there are synergistic wavelengths in these regions that, when sequentially combined with predictive wavelengths, enhance the model's ability to predict species authenticity and adulteration. Even though the untreated data and the 64-point combing data pretreatment generated a lower SEP (5.02 and 4.90, respectively), the 5-point second derivative data are preferred because they generated a lower SEC (4.91) by using fewer PCs in order to achieve a similar SEP (5.17). **Figure 6** shows the actual vs predicted adulteration content of the model developed using the 5-point second derivative data. The slope, offset, correlation,

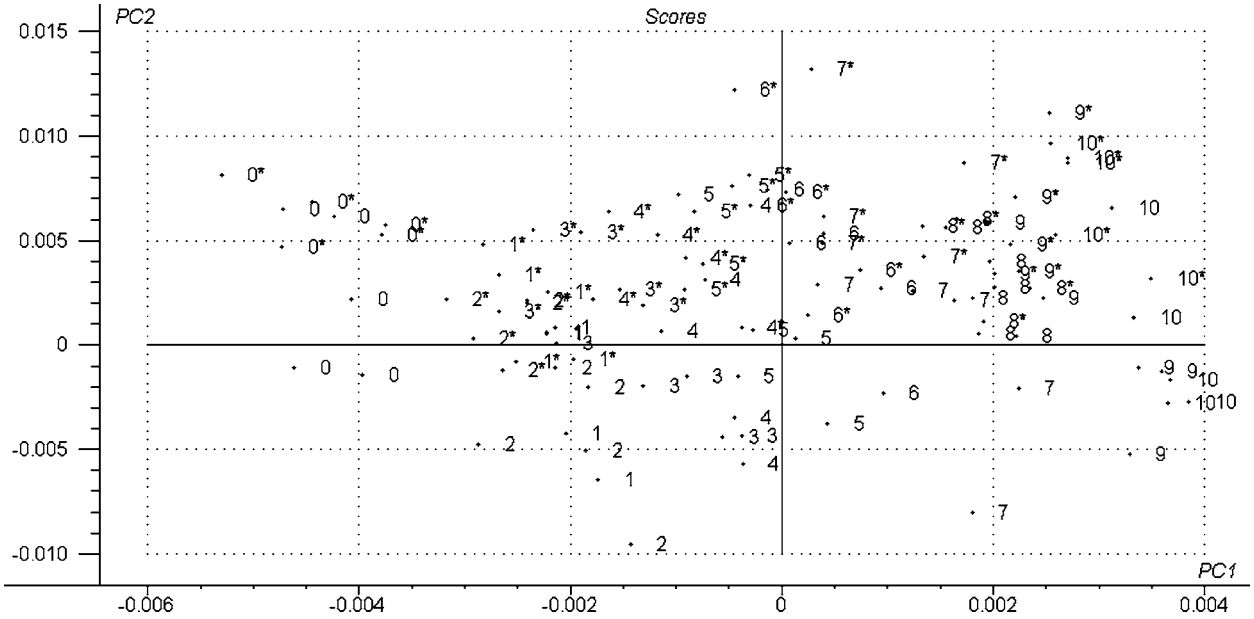

Figure 7. PCA plot of the first two PCs of the first derivative data using the full spectrum. Explained variations in PC1 and PC2 were 22 and 62%, respectively. Samples were named according to crabmeat species (0, 100% Atlantic blue; 10, 100% Blue swimmer) and adulteration percentage (1–9, 10–90% Blue swimmer). Validation samples are shown with an asterisk (*).

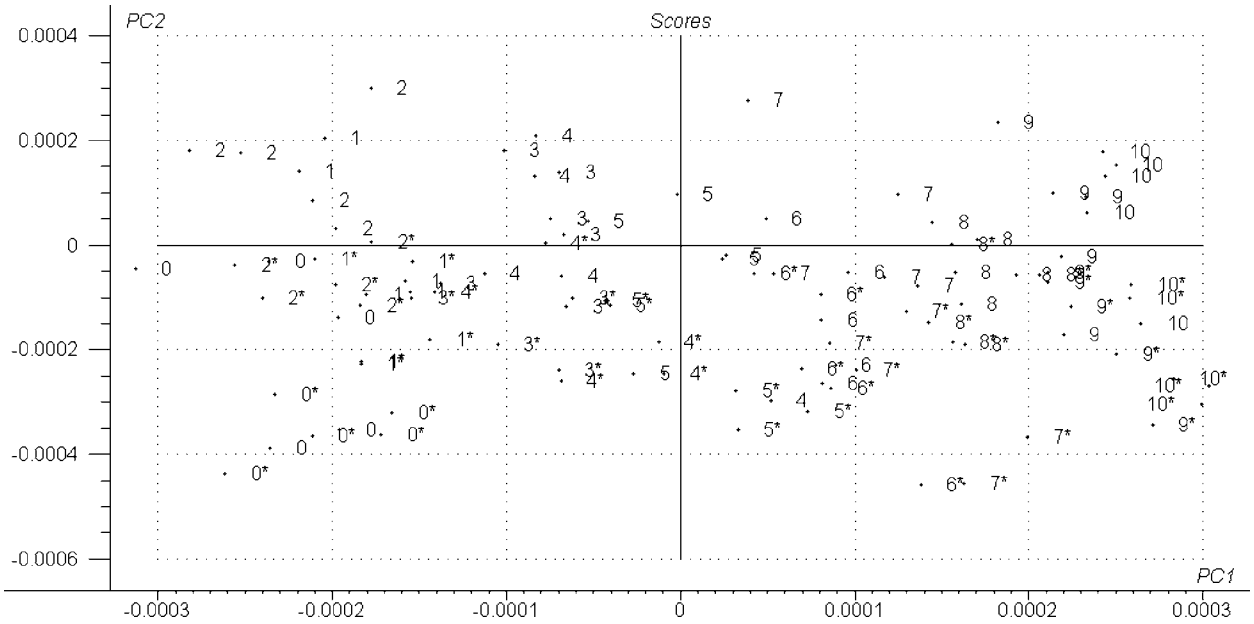

Figure 8. PCA plot of the first two PCs of the second derivative data using a partitioned spectrum (400–1700 nm). Explained variations in PC1 and PC2 were 38 and 35%, respectively. Samples were named according to crabmeat species (0, 100% Atlantic blue; 10, 100% Blue swimmer) and adulteration percentage (1–9, 10–90% Blue swimmer). Validation samples are shown with an asterisk (*).

and bias are 0.976, 1.182, 0.988, and −1.04e−06, respectively, for the calibration set and 0.985, −0.199, 0.987, and −0.970, respectively, for the validation set.

PCA Analyses. PCA analyses were done to graphically determine grouping patterns in an effort to classify samples according to adulteration percentage. **Figure 7** shows the graph of the first two components of the first derivative data using the full spectrum. The explained variances in the first and second PC are 22 and 62%, respectively. On the basis of just two components, a visual relationship among the samples can be gathered. The PCA plot shows a trend from samples of authentic Atlantic blue crabmeat (left) to samples of authentic blue swimmer crabmeat (right). This horizontal trend indicates that the first component contains information that pertains to species authenticity and adulteration percentage. As the percentage of blue swimmer increases, the samples are plotted toward the right side of the graph.

Figure 8 is the PCA plot of the second derivative data of the sequential spectrum approach, which shows a similar trend to that seen in **Figure 7**. The explained variances for the first and second PC are 38 and 34%, respectively. On the basis of both the model performance, in terms of SEC, SEP, associated coefficients of determination, and reduced complexity, in addition to the PCA plots, the best model to detect species authenticity uses the sequential spectrum approach (PLS for

spectral data in the 400−1700 nm wavelength range) and a 5-point second derivative as a data pretreatment.

In conclusion, adulteration, especially species substitution, of high-quality and high-priced food products, such as crabmeat, is an existing problem in the seafood industry. Seafood processing, which often removes morphological properties that enable consumers to distinguish one species from another, makes it easier for species substitution to occur. Results from this study encompass the possibility of using Vis/NIR spectroscopy to detect species authenticity and EA of crabmeat. The 30-point first derivative data generated the lowest error for the full spectral approach (SEC = 5.64 and SEP = 5.64) of the different parameters investigated for spectral data pretreatments. A problem in determining species authenticity using Vis/NIR spectroscopy, however, is the large number of data points per spectrum. Even though multivariate analyses, such as PC analysis, are used to reduce the dimensionality, it is important to distinguish the wavelength regions and features that contain relevant information from the ones that do not, thereby reducing sources of noise and creating more robust regression models. Utilizing a sequential approach, a data set containing the 400−1700 nm spectral range with a 5-point second derivative data generated the best model to determine species authenticity and adulteration (SEC = 4.91 and SEP = 5.17). Regardless of the type of approach to analyze the spectral data, i.e., full or sequential, used in generating the model, results from this study indicate that it is possible to detect species authenticity and adulteration with less than ±6% error.

LITERATURE CITED

(1) Hargin, K. D. Authenticity issues in meat and meat products. *Meat Sci.* **1996**, *43* (S), S227−S289.

(2) Downey, G. Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics. *Trends Anal. Chem.* **1998**, *17* (7), 418−424.

(3) Lees, M. 2nd European symposium on food authenticity—Isotope analysis and other advanced analytical techniques. *Trends Food Sci. Technol.* **1994**, *51*, 200−203.

(4) Sotelo, C. G.; Perez-Martin, R. I. Species identification in processed seafoods. In *Food Authenticity and Traceability*; Lees, M., Ed.; Woodhead Publishing Ltd.: CRC Press: New York, 2003.

(5) Mackie, I. M. Authenticity of fish. In *Food Authentication*; Ashurst, P. R., Dennis, M. J., Eds.; Blackie Academic & Professional; Chapman and Hall: New York, 1996.

(6) AOAC. *Official Methods of Analysis*; Association of Official Analytical Chemists: Washington, DC, 1990; pp 883−889.

(7) Osborne, B. G.; Fearn, T.; Hindle, P. H. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, 2nd ed.; Longnam Scientific & Technical: Harlow, Essex, England, 1993.

(8) McElhinney, J.; Downey, G.; O'Donnell. Quantification of lamb content in mixtures with raw minced beef using visible, near and mid-infrared spectroscopy. *J. Food Sci.* **1999**, *64* (4), 587−591.

(9) Oesterling, M. J. "Blue" crab resources in other countries: Implications for the U.S. industry. *J. Shellfish Res.* **1998**, *17* (2), 375−378.

(10) Zeaiter, M.; Roger, J.-M.; Bellon-Maurel, V. B. Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods. *Trends Anal. Chem.* **2005**, *24* (5), 437−445.

(11) McClure, W. F.; Crowell, B. Quantitative information in near infrared spectra. Part 1. Effects of smoothing and combing. *J. Near Infrared Spectrosc.* **1996**, *4*, 129−137.

(12) Leardi, R. Genetic algorithm-PLS as a tool for wavelength selection in spectral data sets. In *Nature-Inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks*; Leardi, R., Ed.; Elsevier: New York, 2003.

(13) De Jong, S. PLS fits closer than PCR. *J. Chemom.* **1993**, *7*, 551−557.

(14) Bechmann, I. E.; Jorgensen, B. M. Rapid assessment of quality parameters for frozen cod using near infrared spectroscopy. *Lebensm.-Wiss. Technol.* **1998**, *31*, 648−652.

(15) Esbensen, K.; Schonkopf, S.; Midtgaard, T. *Multivariate Analysis in Practice*; CAMO AS, Wennbergs Trykkeri AS: Trondheim, Norway, 1994.

(16) McClure, W. F. Near-infrared spectroscopy. In *Spectroscopic Techniques for Food Analysis*; Wilson, R. H., Ed.; VCH Publishers, Inc.: New York, 1994; pp 13−52.

(17) Kelly, J. F. D.; Downey, G.; Fouratier, V. Initial study of honey adulteration by sugar solutions using midinfrared (MIR) spectroscopy and chemometrics. *J. Agric. Food Chem.* **2004**, *52*, 33−39.

Received for review June 27, 2006. Revised manuscript received November 13, 2006. Accepted November 16, 2006.

JF061801+