Determination of Protein and Moisture in Fishmeal by Near-Infrared Reflectance Spectroscopy and Multivariate Regression Based on Partial Least Squares

Masoum, Saeed*+
Department of Analytical Chemistry, Faculty of Chemistry, University of Kashan, Kashan, I.R. IRAN

Alishahi, Ali Reza; Farahmand, Hamid
Department of Fisheries and Environmental Science, University of Tehran, Karaj, I.R. IRAN

Shekarchi, Maryam
Laboratory of Food and Drug, Sanitary Ministry of Iran, Tehran, I.R. IRAN

Prieto, Nuria
Estación Agrícola Experimental. CSIC. Finca Marzanas, 24346 Grulleros, León, SPAIN

ABSTRACT: The potential of Near Infrared Reflectance Spectroscopy (NIRS) as a fast method to predict the Crude Protein (CP) and Moisture (M) content in fishmeal by scanning spectra between 1000 and 2500 nm using multivariate regression technique based on Partial Least Squares (PLS) was evaluated. The coefficient of determination in calibration ($R^2_{c}$) and Standard Error of Calibration (SEC) were 0.95 and 14.03 g / kg Dry Matter (DM) and 0.80 and 3.52 g / kg, for CP and M content, respectively. This study proved that the application of NIRS using PLS is well fitted to evaluate the protein and moisture content of fishmeal.

KEY  WORDS: Near-Infrared Reflectance Spectroscopy (NIRS), Protein, moisture, Partial least squares, Fishmeal.

INTRODUCTION
Marine based ingredients, especially fishmeals, are highly sought after as the protein source of choice for many formulated diets [1]. That is because fishmeals provide feeds with high contents of essential amino and fatty acids, and low content in carbohydrates; thus being usually well digested and mainly used by feeds industry as a rich source of protein[1,2]. Hence, fishmeal industry is widely spread across the world and almost one third of the total global fish and shellfish catch is used by this industry [3]. Approximately 65% of the average annual
global production of fishmeal is used by the aquafeed industry [2], notwithstanding; it is also used in poultry, swine, ruminant, companion animal and even in human foods, as a protein feedstuff. Thereby, the percentage of fishmeal in poultry and mammalian feeds is small but the total quantity of such feeds is very large.

Nevertheless, the nutritive value of fishmeal is affected by the high variability in the protein content, which varies from 57 to 77% as a consequence of the different fish species used in the fishmeal manufacturing process. The practical and economic repercussions of this variability are very important in the feed compound manufacturing industry where a uniform product of consistent composition and quality is to be produced from inherently variable raw materials and products [4]. Hence, analytical control is essential in order to assess the protein content in fishmeal.

On the other hand, other vital parameter in fishmeal is the moisture content since it plays a major role in the shelf life and storage time of fishmeal. Indeed, it has been shown that the decrease in the moisture content induces an increase in the shelf-life and a decrease in the spoilage phenomenon [5].

The conventional methods used in order to determine both protein and moisture contents of fishmeal are the Kjeldahl and the drying in oven, respectively. However, these techniques are tedious, destructive, relatively expensive and time-consuming, as well as they require highly skilled operators [6].

Recently, the attention has focused on the development of Near Infrared Reflectance Spectroscopy (NIRS), a sensitive, fast and non-destructive analytical technique with simplicity in sample preparation [7]. In this sense, NIRS is one of the most promising techniques for large scale food evaluation, since it has been demonstrated its ability to predict chemical composition of many products such as animal [8], fish [6,9,10], milk [11], and meat samples [12,13]. Furthermore, this technology has classified successfully fresh and frozen thawed fish [14] and discriminated fishmeal batches made with different fish species [15]. However, to the best of our knowledge, just few studies [4] have been performed to evaluate the proximate chemical composition of fishmeal and no one tested the NIRS calibration models using an external validation.

Thus, the aim of the present study was to evaluate the NIRS technology as an alternative analytical method to predict the crude protein and moisture content in fishmeal samples by means of an external validation.

**EXPERIMENTAL SECTION**

*Spectra collection*

One hundred and twenty \( (n = 120) \) fishmeal samples from different industries and batches located in the north and south of Iran were collected from January to February 2007. They contained different fish species including herring \((Clupea harengus, n = 60)\), sardine \((Sardinops sagax, n = 30)\) and trash fish species \((n = 30)\). Due to produce fish meal with homogenous structure, the samples were grinded before further analysis.

All fishmeal samples in the powder form were scanned in reflectance mode over the NIR spectral range (1000 to 2500 nm) using a monochromator instrument FT-NIRSystems (Bomem, 450 St. Jean Baptiste, Quebec PQ G2E 5S5 Canada, Version 1, 1994). Twenty grams of every sample were filled into sample holder which was placed in the measurement cell to be scanned. The spectrum of each sample was the average of 64 successive scans and every sample was scanned in triplicate (repack). Thus, the area of the sample scanned could be increased and thereby the sampling error reduced [16]. Spectral data were stored as the logarithm of the reciprocal of reflectance \([\log (1/R)]\) and the spectra were averaged for chemical analysis.

*Chemical analyses*

All scanned fishmeal samples were subsequently analysed for chemical composition. Crude protein content analysis was performed in accordance with the standard Kjeldahl procedure \((CP, \text{method ID 976.06})\) described by the Association of Official Analytical Chemists (AOAC 2006) [17]. Moisture content was measured by oven drying the samples at 105 °C for 18 h \((\text{method ID 934.01 AOAC 2006})\) [17]. The percent weight of water was calculated from the difference between the beginning and ending mass of fishmeal. Both protein and moisture content determinations were performed in duplicate, the average value for each sample being then used as reference value.

*NIRS data analysis*

The 120 fishmeal samples were separated into a calibration and a validation set, consisting of 100 and 20 samples, respectively. The calibration set was used to develop
the calibration model. The developed model was evaluated by using the fishmeal samples included in the validation set which had not been previously used in the calibration procedure, in order to avoid overfitting of the calibration equations. The validation set was chosen randomly so that it was an adequate representative of the calibration population. In order to ensure a fair comparison, the calibration and validation sets were the same for both parameters studied.

Before calibration and validation, CENTER algorithm was applied on spectra. The CENTER program ranks spectra according to their Mahalanobis distance (H-statistic) from the average spectra using Principal Components (PC) scores. The Mahalanobis distance indicates how different a sample spectrum is from the average spectrum of the set [18]. Thereby, a sample with an H statistic of ≥ 3.0 standardized units from the mean spectrum was defined as a global H outlier and was eliminated from the population. Because NIRS spectra are affected by particle size, light scatter and path-length variation, pre-treatments of the spectral data improve calibration accuracy. In this way, spectral pre-treatments such as Standard Normal Variation and Detrending (SNV-D) [19], Multiplicative Scatter Correction (MSC) [20] and first or second order derivatives [21] were applied to the spectra to reduce the noise and light scattering effects. Calibration equations were calculated on raw and transformed spectra by Partial Least Square Regression Type I (PLSR1) to predict crude protein (g / kg DM) and moisture (g / kg) concentration. The number of factors used as independent variables in the prediction equations was fixed at a maximum of 9 (i.e. less than 10% of the number of samples used in the calibration) in order to avoid overfitting [22]. The optimal number of factors was chosen as a function of the first local minimum in the validation residual variance plot. External validation was used to test the calibration model and the accuracy of prediction was evaluated in terms of coefficient of determination in external validation (R²E) and Standard Error of Prediction (SEP) [23].

The UNSCRAMBLER program (version 8.5.0, Camo, Trondheim, Norway) was used to perform data treatment such as MSC and derivatives and also PLS modeling of chemical data of fishmeal samples, whereas the software supplied with the NIR instrument (Bomem GRAMS/32, ABB Bomen Inc., Canada) was used for scanning, for application of SNV-D, and for calibration and external validation procedures.

RESULTS AND DISCUSSION

The CP and M contents of the fishmeal samples determined by conventional methods are shown in Table 1. The data set presented a mean M content of 95.8 g / kg (range: 80.5-112.0 g / kg). With regard to CP content, fishmeal had a requirement of 595.4 g / kg DM for industrial purposes and a range between 500.2 and 684.5 g / kg on a DM, thus representing a wide range of composition as consequence of using several fish species during the fishmeal manufacturing process. These values agree those showed in literature [4], although the range showed in our study for the CP content was consistently higher than that indicated by Cozzolino et al. [4].

The NIRS calibration and validation statistics for the CP and M contents are shown in Tables 2 and 3, respectively. As it can be observed in Table 2, the coefficient of determination of calibration and the standard error of calibration corresponding to the calibration equation selected for estimation of the CP content in fishmeal samples were satisfactory (R²C = 0.95, SEC = 14.03 g / kg DM, Table 2). Nevertheless, unknown components in fishmeal could bring about many problems during calibrating and optimizing of the model because of equations overfitting, being one of the most efficient options to get over this inconvenient applying a parallel factor analysis [24]. In the present study, in order to overcome that problem, an external validation was carried out with the fishmeal samples included in the validation set which had not been previously used in the calibration procedure. In that way, the ratio between the Standard Deviation (SD) of the population of the validation set and the Standard Error of Prediction (SEP) could be calculated [25,26]. This relationship is known as Ratio Performance Deviation (RPD) and shows how good the calibration and prediction will work for analytical purposes. A SD/SEP value greater than 2.5 is considered useful for screening purposes and a value greater than 5 is considered excellent for quality control [25,26]. In our study, the coefficient of determination of external validation was high (R²E = 0.97) and the ratio between the SD (65.61 g / kg DM) and the SEP (12.33 g / kg DM) was higher than 5 (RPD = 5.32). Hence, the equation selected to predict the CP content showed that NIRS could be used for routine analysis in order to estimate accurately this parameter.

As far as spectral data pre-treatments is concerned, the CP content could be predicted accurately when first-
Table 1: Range, mean and standard deviation of the crude protein and moisture content of fishmeal samples.

<table>
<thead>
<tr>
<th>Chemical data</th>
<th>Calibration set (n=100)</th>
<th>External validation set (n=20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range</td>
<td>Mean</td>
</tr>
<tr>
<td>Crude Protein (g / kg DM)</td>
<td>500.2 - 684.5</td>
<td>595.4</td>
</tr>
<tr>
<td>Moisture (g / kg)</td>
<td>80.5 - 112.0</td>
<td>95.8</td>
</tr>
</tbody>
</table>

n: number of samples, SD: standard deviation.

Table 2: Prediction of crude protein content corresponding to the fishmeal samples (g / kg DM).

<table>
<thead>
<tr>
<th>Spectra data pre-treatments</th>
<th>Calibration</th>
<th>External validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>p</td>
</tr>
<tr>
<td>None [Log (1/R)]</td>
<td>99</td>
<td>3</td>
</tr>
<tr>
<td>D1</td>
<td>99</td>
<td>2</td>
</tr>
<tr>
<td>D2</td>
<td>97</td>
<td>1</td>
</tr>
<tr>
<td>MSC</td>
<td>98</td>
<td>3</td>
</tr>
<tr>
<td>MSC+ D1</td>
<td>99</td>
<td>1</td>
</tr>
<tr>
<td>MSC+ D2</td>
<td>97</td>
<td>3</td>
</tr>
<tr>
<td>SNV-D</td>
<td>94</td>
<td>4</td>
</tr>
</tbody>
</table>

Log (1/R): raw absorbance data, D1: first-order derivative, D2: second-order derivative, p: number of PLS terms utilized in the calibration equation.

Table 3: Prediction of moisture content corresponding to the fishmeal samples (g / kg).

<table>
<thead>
<tr>
<th>Spectra data pre-treatments</th>
<th>Calibration</th>
<th>External validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>p</td>
</tr>
<tr>
<td>None [Log (1/R)]</td>
<td>99</td>
<td>5</td>
</tr>
<tr>
<td>D1</td>
<td>99</td>
<td>2</td>
</tr>
<tr>
<td>D2</td>
<td>98</td>
<td>3</td>
</tr>
<tr>
<td>MSC</td>
<td>97</td>
<td>4</td>
</tr>
<tr>
<td>MSC+ D1</td>
<td>97</td>
<td>2</td>
</tr>
<tr>
<td>MSC+ D2</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td>SNV-D</td>
<td>94</td>
<td>6</td>
</tr>
</tbody>
</table>

Log (1/R): raw absorbance data, D1: first-order derivative, D2: second-order derivative, p: number of PLS terms utilized in the calibration equation.
order derivative was applied to the raw spectra. This could be due to the fact that physical effects gave rise to baseline shift which might have been corrected by the use of the first-order derivative [21], thus enhancing the information related to chemical composition of the fishmeal samples. Furthermore, derivatives separate overlapping peaks of the spectrum; hence a band resolution enhancement takes place resulting in spectra with a better definition than raw spectra.

As it has been showed so far, the CP content was estimated by NIRS successfully; however, it is also important to know why. Finding wavelength regions relevant to the CP content prediction are essential to find a reliable explanation. To do this, the derivative spectrum of each sample was plotted against the crude protein content data studied, so the correlation between the absorbance at each wavelength and the CP content could be observed (Fig. 1). In this sense, the absorbance data showing the highest positive correlation with the CP content were located at 1510 nm, 2050-2060 nm and 2180 nm, which correspond to N-H stretch first overtone, N-H stretch combination bands, and N-H bend second overtone, respectively [21,27,28]. These wavelengths are typical of protein absorption whose chemical structure is based on N-H bonds, thus explaining the wavelengths previously described. In addition, there was a high correlation around 2300 nm, which corresponds to the absorption of the C-H bend second overtone of protein. It has to be noted that all these correlations were high, reaching values higher than 0.9; hence the reason why NIR was well suited when assessing the CP content in fishmeal.

Fig. 2 plots the chemical reference data against the NIRS predicted values for the CP content in both calibration and validation fishmeal samples and it shows a strong relationship between them. Nevertheless, despite the successful predictive ability of the equation for the CP content, a further broaden the database by incorporating new samples from different years and fish species would be desirable. It must be pointed out that those samples will become available in the future.

The results showed in the present study are better than those reported by Cozzolino et al. [4], for fishmeal (R² = 0.85, RPD = 2.77), probably as consequence of a higher range of the reference data in the current research. According to Cozzolino et al. [4], a very narrow range in the CP content (range: 604-708 g / kg DM) could have endangered the prediction and it could be explained by the careful selection of the raw material used to make the fishmeal and the high quality standard maintained and imposed by the factory in the final product. A validation of our findings with other studies was not possible as no more literature data concerning the estimation of chemical composition of fishmeal by NIRS was found. Nevertheless, our results are in broad agreement with other studies applying the NIR spectroscopy to assess the CP content in animal feeds and surimi and are better than those showed by Xiccato et al. [10], in sea bass (R² < 0.7, RPD < 2), probably due to wide variations of chemical characteristics in the set of sea bass which included great differences in fish rearing systems and weight.

Table 3 shows that the prediction of the M content was not as accurate as that previously described for CP, but it still showed a relatively high coefficient of determination of calibration and a low standard error of calibration (R²C = 0.80, SEC = 3.52 g / kg). Indeed, as an external validation was performed the coefficient of determination of validation was relatively high (R²V = 0.81) and the standard error of prediction was still low enough to give rise a RPD statistic closed to that considered by the literature as acceptable (SEP = 3.24 g / kg, RPD = 2.28). Furthermore, as it can be observed in the Fig. 3, the wavelengths located at 1450 nm (O-H stretch first overtone), 1790 nm (O-H combination bands) and 1940 nm (O-H bend second overtone) showed high correlations between absorbance data and M content, reaching values up to 0.8. Consequently, the accuracy of
prediction of NIRS for the M content might be useful for screening purposes.

In relation to the spectral data pre-treatments, the best equation to predict the M content was that using no one, that is, the highest prediction performance of M content was obtained when the raw spectra \[\log (1/R)\] were used.

The chemical reference data against the NIRS predicted values for the M content in both calibration and validation fishmeal samples are showed in the Fig. 4. As it can be observed, the samples were nicely positioned along the regression line; however, the relationship between the reference and predicted values was not as good as that found for the CP content. One of the reasons explaining the lowest level of accuracy to estimate the M content could be the effect of temperature on the structure of water in fishmeal, since it is well known that the NIR spectrum of water is sensitive to temperature. The large spectral variations of the water NIR absorption spectrum induced by temperature are interpreted in the literature as being due to changes of the hydrogen-bonded water structure [29]. Indeed, all water molecules undergo random motion, making or breaking the hydrogen bond. Accordingly, oscillations in temperature along both the production of fishmeal and the spectra collection in our study could have brought about that the vibration of O-H bond (especially hydrogenous bonds) varied and ultimately gave rise to several forms of spectra [30]. These changes in the spectra are not consequence of changes in the composition chemical, hence jeopardizing the prediction of the M content. Other promising reason could be the different procedures and styles of production of fishmeal and the different raw materials used in its production. All of those bring about that the particle size and performance of the fishmeal become different; thus scanned spectra of the samples could have been no precise and reliable representatives of reality [31].

In comparison with our results, Cozzolino et al. [4], Xiccato et al. [10], and Uddin et al. [6], provided more reliable calibrations for the M content prediction (\(R^2 > 0.93\), RPD > 3.8) when analysing fishmeal, European sea bass and surimi samples, respectively; probably because of wider ranges of M content. Nevertheless, our results are in accordance with those reported by Cozzolino et al. [32], in fish oils (RPD = 2.2) who indicated some features
including the species of fish used to produce the fishmeal, the seasonality of the fish, and modifications in the steps of the manufacturing process during the period of sampling, which could have endangered the prediction.

It is worth mentioning the study achieved by Brüll et al. [33], who showed the spectral area analysis as a viable method in NIRS moisture assays. According to these authors, the use of that method allowed to obtain accurate predictions, being more robust and requiring fewer calibration samples than the conventional partial least squares regression method. Hence, the peak area method could be especially suitable for moisture assays in early formulation development and in-situ process monitoring.

CONCLUSIONS

The results observed in this study show that NIRS technology can provide an accurate prediction of the CP content in fishmeal samples. The equation for the M content showed lower predictive ability, but nevertheless it was useful to predict the M content under industrial conditions. Hence, NIRS technology could be a suitable replacement of traditional chemical methods for the rapid and non-destructive assessment of the CP and M content in fishmeal.

Acknowledgements

The authors are grateful to Dr. Bro and Dr. Cozzolino for their collaboration during the writing of the paper. Also, we thank Dr. Pirali, the head of the food and drug center in the sanitary ministry of Iran, the chairman of the environmental and fisheries department, M. Abasi and S. Fekri.

REFERENCES


