Article

Near-Infrared Spectroscopy and Chemometrics for Effective Online Quality Monitoring and Process Control during Pelagic Fishmeal and Oil Processing

María Gudjónsdóttir [1,2,*] , Gudrún Svana Hilmarsdóttir [1,2], Ólafur Ögmundarson [1] and Sigurjón Arason [1,2]

[1] Faculty of Food Science and Nutrition, University of Iceland, Nýi Garður, Sæmundargata 12, 102 Reykjavík, Iceland; gudrunsvana@hi.is (G.S.H.); olafuro@hi.is (Ó.Ö.); sigurjar@hi.is (S.A.)
[2] Matis Food and Biotech R&D, Vínlandsleid 12, 113 Reykjavík, Iceland
*   Correspondence: mariagu@hi.is; Tel.: +354-698-6913

Abstract: Near-infrared spectroscopy has become a common quality assessment tool for fishmeal products during the last two decades. However, to date it has not been used for active online quality monitoring during fishmeal processing. Our aim was to investigate whether NIR spectroscopy, in combination with multivariate chemometrics, could actively predict the changes in the main chemical quality parameters of pelagic fishmeal and oil during processing, with an emphasis on lipid quality changes. Results indicated that partial least square regression (PLSR) models from the NIR data effectively predicted proximate composition changes during processing (with coefficients of determination of an independent test set at R^2_{CV} = 0.9938, RMSECV = 2.41 for water; R^2_{CV} = 0.9773, RMSECV = 3.94 for lipids; and R^2_{CV} = 0.9356, RMSECV = 5.58 for FFDM) and were successful in distinguishing between fatty acids according to their level of saturation (SFA ($R^2_{CV} = 0.9928$, $RMSECV = 0.24$), MUFA ($R^2_{CV} = 0.8291$, $RMSECV = 1.49$), PUFA ($R^2_{CV} = 0.8588$, $RMSECV = 2.11$)). This technique also allowed the prediction of phospholipids (PL $R^2_{CV} = 0.8617$, $RMSECV = 0.11$, and DHA ($R^2_{CV} = 0.8785$, $RMSECV = 0.89$) and EPA content $R^2_{CV} = 0.8689$, $RMSECV = 0.62$) throughout processing. NIR spectroscopy in combination with chemometrics is, thus, a powerful quality assessment tool that can be applied for active online quality monitoring and processing control during fishmeal and oil processing.

Keywords: near-infrared (NIR) spectroscopy; chemometrics; process monitoring; quality assessment; process optimization; product optimization

1. Introduction

Recently, pelagic fishmeal for use in fish and animal feed has been lowering in value due to the high competition with alternative protein meals. However, pelagic fishes, such as the Atlantic mackerel (*Scomber scombrus*), Atlantic herring (*Clupea harengus*), and blue whiting (*Micromesistius poutassou*), are nutritious raw materials containing several valuable fatty acids, essential proteins, and minerals [1]. This calls for increased utilization of pelagic fish raw materials and optimization of production processes of higher-value products, including products for human consumption. Traditional fishmeal and oil processes from pelagic fish species have remained unchanged for decades [1–3]. Moreover, the variation in water, protein, and lipid content in the processing streams during fishmeal processing of pelagic fishes is extreme due to the high variation in the raw material that enters the process [1,4,5]. The raw material often includes several species, cut-offs, heads, bones and guts, and the availability of species is highly seasonal [1,6]. The share of side-streams from pelagic fish processing to human consumption has increased in recent years and will increase significantly in the future. This expected demand of products from pelagic side-stream processing calls for effective raw material and process monitoring of the chemical

changes occurring at each processing step. Such monitoring is critical for process optimization and fast decision making towards the production of higher-value products. However, traditional quality assessment methods performed in the laboratory are commonly sample destructive, time consuming, and expensive, which does not fit well into the need for fast decision making within the processing environment [4].

Near-infrared spectroscopy (NIR) is a fast, non-destructive analytical method that has been used for some time for the quality assessment of seafood, including the quality of the final fishmeal products [4,7]. The main chemical constituents of fish and fishmeal are rich in C-H, O-H, N-H, and S-H chemical bonds, which overtones and combination bands absorb strongly in the NIR spectral range [8]. NIR analysis has, therefore, been shown to be useful in predicting several chemical [4,7] and derived physical [9] and sensory quality parameters [10–12] in diverse agricultural and marine food applications [13]. However, due to the complex calibration work needed and the previously mentioned variations in the chemical compositions of the streams during processing, little effort has been put into the online processing of data to allow active, live quality monitoring during fishmeal and fish oil processing. This kind of quality monitoring is, however, necessary if products of higher quality are to be produced. The use of chemometric data processing simplifies this necessary calibration work and takes multiple variables into consideration during the prediction model building [14]. The combination of spectroscopy and multivariate chemometrics is, therefore, highly effective in the development of fast online prediction tools [7,13,15].

The aim of the current study is, therefore, to investigate whether NIR spectroscopy, in combination with multivariate chemometrics, can actively predict changes in the main chemical quality parameters of pelagic fishmeal and fish oil online during processing, and to assess if the technique can be actively used for process control and optimization, potentially towards the production of high-quality products for human consumption and customized high-quality feed for aquaculture and pets.

2. Materials and Methods

2.1. Raw Materials and Process Description

A mixture of several pelagic fish species was caught off the east and southeast coasts of Iceland with a purse seiner during 3–6 September 2017. The total landings arriving to the fishmeal factory contained a mixture of Atlantic mackerel (*Scomber scombrus*, 58%), Atlantic herring (*Clupea harengus*, 37%), blue whiting (*Micromesistius poutassou*, 4.5%), and <0.5% of by-catch from other species. The fishmeal process commenced three days post-landing. The fishmeal and oil production process, which can be seen in Figure 1, was described in detail by Hilmarsdóttir et al. (2020) [16]. Three processing runs were analysed, each with a different temperature in the cooker, i.e., 85 °C, 90 °C, and 95 °C, respectively, to give more variation to the sampling set. Samples were taken in triplicate from each processing step (as marked with red dots in Figure 1) from the raw material to the final product during each process run to follow the quality characteristics of the processing streams throughout processing. The samples were analysed for water, lipids, fat-free dry matter (FFDM), fatty acid composition (FAC), free fatty acids (FFA), and phospholipids (PL), along with their NIR spectra.
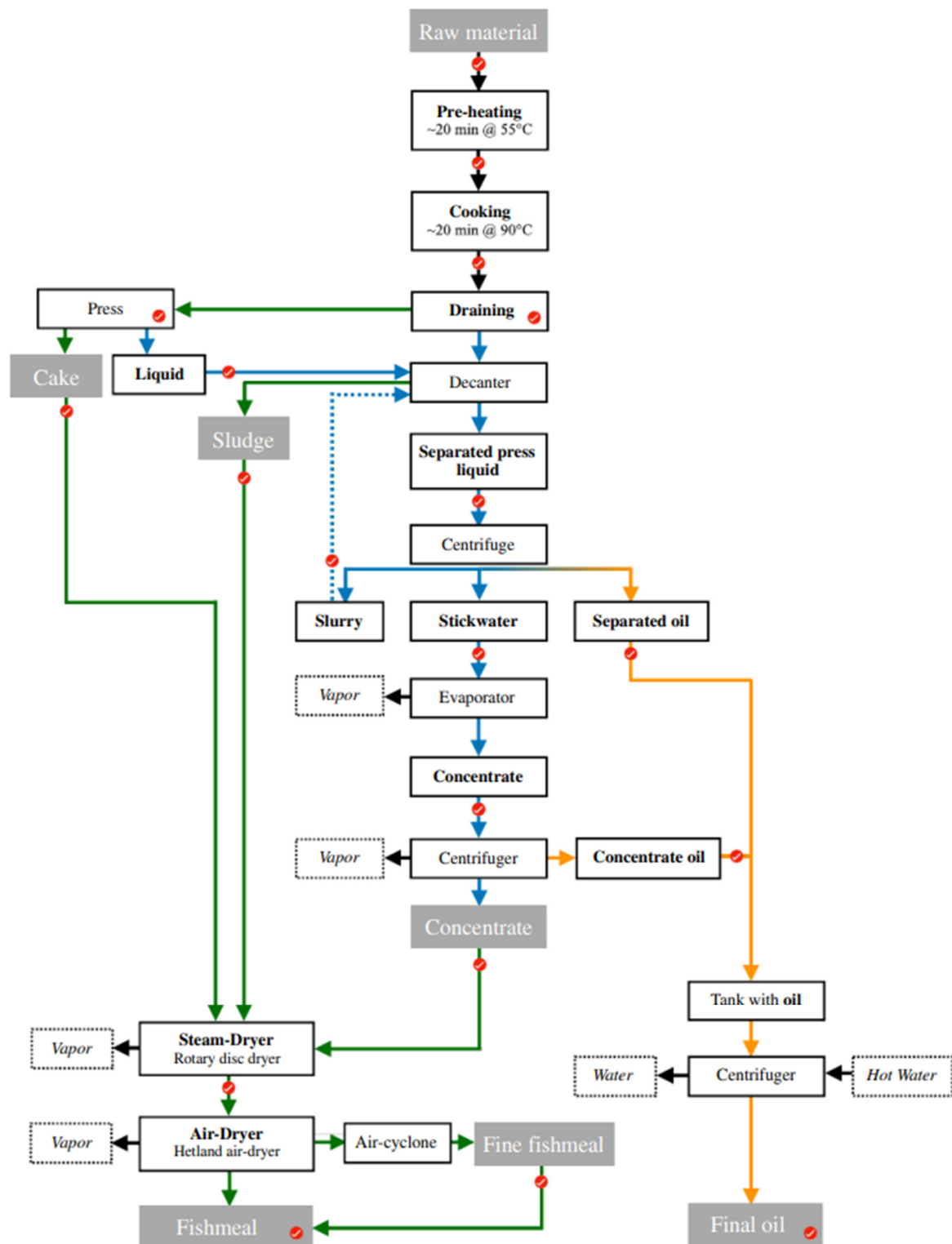
Figure 1. Process flowchart illustrating a traditional fishmeal and fish oil process. Green colour represents solid streams during processing, blue represents liquid streams, and yellow represents oil streams. Red dots indicate sampling points during processing. Green-filled boxes represent the processing streams that are joined before the drying step to form the final fishmeal, and the red dashed box highlights the cooking step, which was investigated at three different temperatures (standard 90 °C, 85 °C, and 95 °C). The flowchart is reproduced from Hilmarsdóttir et al., (2020) [16].

2.2. Near-Infrared Spectroscopy

Near-infrared (NIR) reflections of the samples were analysed with a Bruker Multi-Purpose Analyzer (MPA) spectrometer (Bruker Optics GmbH, Ettlingen, Germany), using a handheld fibre optic probe. The instrument operates in the spectral range from 4000 to 12,500 cm^{-1} and was chosen due to its robustness and easy data extraction for further data analysis. The fibre optic probe, furthermore, provides excellent opportunities for on-line, at-line and in-line quality control applications [13], and is thus a suitable instrument for the current study. NIR measurements were performed, and all NIR data were collected using the Opus software v.6.5 (Bruker Optics GmbH, Ettlingen, Germany). Five scans were performed on each sample to increase the signal-to-noise ratio (SNR), and their average spectrum was used to represent each sample. The integration time of each scan is 10 s, resulting in an overall sample integration time of less than a minute. Each sample from each sampling location (m = 30) was measured in triplicate (n = 3), giving a total of 90 NIR spectra throughout the fishmeal and oil processing operation.

2.3. Chemical Reference Measurements

The physicochemical quality parameters of water, lipids, FFDM, FAC, FFA, and PL were analysed in triplicate (n = 3) throughout the processing operation to give a detailed view of the quality changes occurring in the pelagic raw material during processing. The water, lipids, and FFDM were measured at all sampling points (m = 30), while FAC, FFA, and PL were measured at key locations during processing (m = 22 for FFA and PL, m = 17 for FAC, as indicated by Figure 1).

The water content of all processing samples was measured by the drying of 5 g samples in an oven at $104 \pm 4\,°C$ for 4 h [17]. The water content in the final oil samples was, however, measured by calorimetric titration in an 851 Titrando titrator (Metrohm, Herisau, Switzerland). The water content was measured in triplicate and expressed as g water/100 g sample.

The total lipid (TL) content of the samples was obtained by Bligh and Dyer extraction [18]. The extracts were then used for the analysis of free fatty acids (FFA) with the Lowry and Tinsley method [19] with modifications described by Bernárdez et al. (2005) [20], as well as the colorimetric analysis of phospholipids in the form of phosphatidylcholine [21]. The total lipid results were expressed as g lipids/100 g sample, while FFA and PL concentrations were expressed as g/100 g TL.

Fat-free dry matter (FFDM) was calculated as the percentage of remaining weight once the water and lipid contents had been subtracted from the total sample weight, as seen in Equation (1):

$$FFDM(\%) = 100 - water\% - TL\% \qquad (1)$$

The FFDM results were expressed as g FFDM/100 g sample.

The fatty acid compositions (FAC) of the samples were determined by gas chromatography in a Varian 3900 gas chromatography instrument (Varian Inc., Walnut Creek, CA, USA), as described in the AOAC official method [22]. The fatty acids were classified according to their level of saturation, i.e., into saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), or polyunsaturated fatty acids (PUFA), respectively. The fatty acids eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) were also analysed specifically due to their potential health benefits [23–25]. All FAC results were expressed as g FA/100 g TL.

The obtained range of each reference chemical variable can be seen in Table A2 in Appendix B. However, a detailed analysis of the chemical changes occurring during processing and an interpretation of the reference measurements were performed and discussed by Hilmarsdóttir et al. (2020) [16].

2.4. NIR PCA and PLSR Prediction Models

Data from the NIR spectral analysis and the reference measurements were imported to Unscrambler X® (Camo, Oslo, Norway) for further data analysis. Principal Compo-

nent Analysis (PCA) and Mean-Centred Partial Least Square Regression (PLSR) models were built to test the ability of the NIR spectra to predict the various analysed reference parameters. PCA and PLSR modelling are amongst the most common methods applied in spectroscopy to assess correlations between spectral data and physicochemical parameters in food and feed. PCA is an unsupervised method that assesses the total variance between variables and samples and is an effective way of identifying potential sample outliers.

Two spectra from each sampling location during processing were used for building the training set for the PLSR calibration models compared to the reference measurements. The third sample spectrum from each sampling location was used as an independent test set, used to cross-validate the prediction potential of the models. The spectral raw data showed baseline drifting, noise, peak shifting, and/or scattering in some of the variables. The effectiveness of using different spectral treatments was thus also tested, i.e., applying (i) no spectral treatments, (ii) using a baseline correction, (iii) a first derivative model using the Savitzky–Golay method with 11 smoothing points at each end of the spectra, (iv) a Savitzky–Golay second derivative model of the second polynomial order with 11 smoothing points, and finally (v) a full multiplicative scatter correction (MSC) of the spectral data [15]. The score plots of the first two principal components (PCs) and the predicted versus reference results from each obtained PLSR model were assessed both graphically and numerically to easily identify the trends, outliers, stability, and precision of each prediction model. The efficiency of the models was evaluated according to the model coefficient of calibration (R_C^2) and the coefficient of cross validation (R_{CV}^2), as well as the representative root-mean-square error of calibration (RMSEC) and the root-mean-square error of cross validation (RMSECV), as described by Gomes et al. (2022) [14].

3. Results and Discussion

3.1. Chemical Changes during Fishmeal and Oil Processing

The main chemical changes occurring during fishmeal and oil processing involve the separation of water and lipids from the fat-free dry matter, to form a low-fat fishmeal and a pure fish oil. The chemical changes observed in the present processes were described in detail by Hilmarsdóttir et al. (2020) [16], but the ranges of each of the chemical reference variables are given in Table A2 in Appendix B. The process reflects great variation in chemical composition between processing steps, varying from highly liquid streams rich in water and oil (separated press liquid, first concentrate, final oil), to more solid streams containing high dry matter content (press cake, sludge, second concentrate, and final fishmeal). The chemical analysis indicated that several processing steps in the current, traditional processes are inefficient both with regards to energy and material efficiency [16]. Lowering the temperature to 85 °C resulted in more effective lipid separation during the process, and thus, also a lower lipid content in the final fishmeal, making it both more valuable and more stable. Furthermore, a life cycle assessment (LCA) of the process indicated that lowering the cooking temperature to 85 °C had a positive impact on the evaluated environmental categories [5], indicating that small process changes can also benefit the sustainability of the process. In addition, the two studies identified that the drying and evaporation stages require optimization to allow the production of food-grade products for human consumption, instead of the current lower-value fishmeal and fish oil.

The traditional chemical reference assessment methods applied in the studies were very time consuming, required the use of environmentally harmful solvents and chemicals, and the reliability (precision, accuracy, repeatability) of the methods may in some cases also depend on the individual performing the analysis. The disadvantages of applying these methods, as well as the need for process optimization, highlights the need for improved process monitoring on-site. Spectroscopic methods, such as NIR spectroscopy, might be a solution to this problem, as investigated in the following sections.

3.2. NIR Spectra and Principal Component Analysis

The obtained NIR spectra showed clear differences between the samples according to where they were sampled from during the fishmeal process (Figure 2). Dominating absorption peaks of overtones and combination bands due the vibrations of chemical bonds such as -CH, -NH, OH, C=O, etc., which can be found in the main constituents of fishmeal, were assigned (Figure 2; Table A1 in Appendix A). The peak assignment indicated that the main differences between samples could be observed in the water and lipid contents and characteristics of the samples.
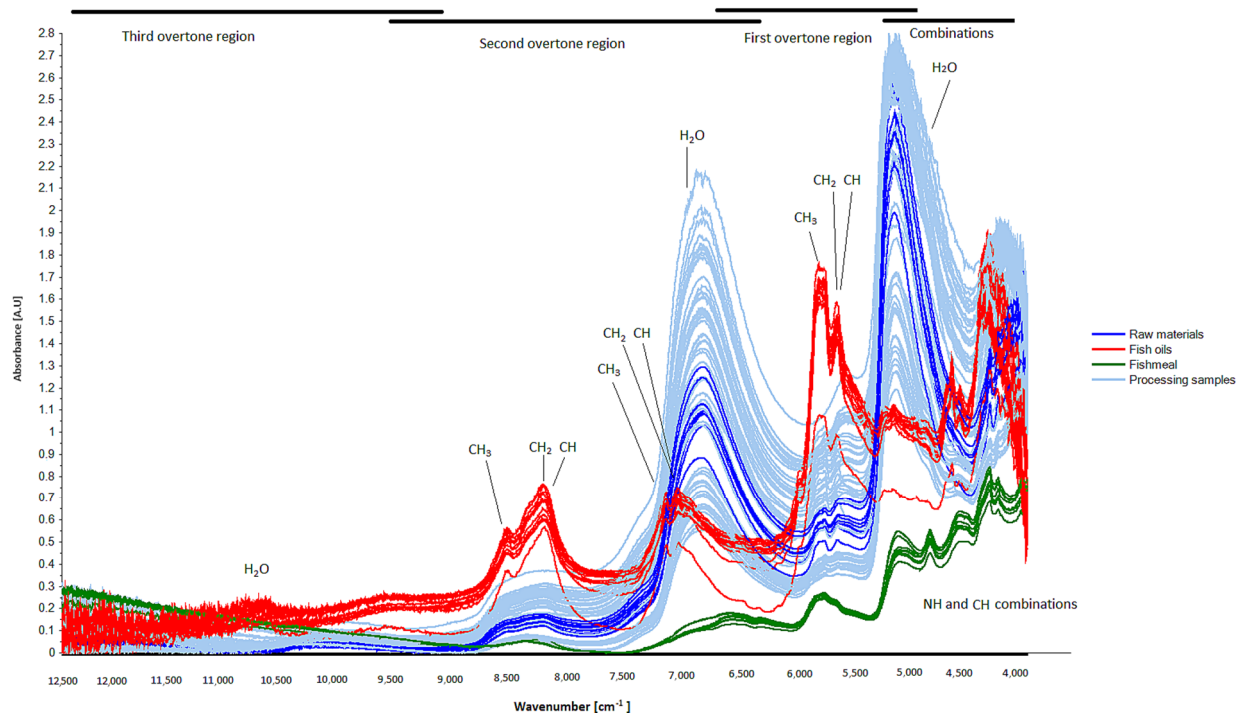

Figure 2. Baseline corrected NIR spectra obtained during fishmeal and oil processing from pelagic fish species. Chosen characteristic absorption bands from overtones and combination bands of common chemical bonds in water, lipids, and proteins are identified.

To get a better overview of the connection between these main variables, a Principal Component Analysis (PCA) was performed on the NIR spectral data (Figure 3). The PCA, which described 99% of the variation between samples, effectively distinguished between the raw materials, products (fishmeal and oil), and other processing streams. The variation in PC1 and PC2 was primarily driven by variation in the water and lipid contents of the samples. Clear groupings could thus be seen within the processing streams, distinguishing streams with higher water and lipid content (liquid streams)—such as after cooking and draining, the separated press liquid, and the first concentrate—from streams with higher protein content (solid streams), such as from the press cake, sludge, and the second concentrate. These results indicate that the NIR spectral information obtained during processing has the potential to be used for process optimization for the main chemical components. This led to the construction and prediction potential assessment of linear partial least square models of all the chemical components, previously investigated by Hilmarsdóttir et al. (2020) [16], as discussed in the following section.
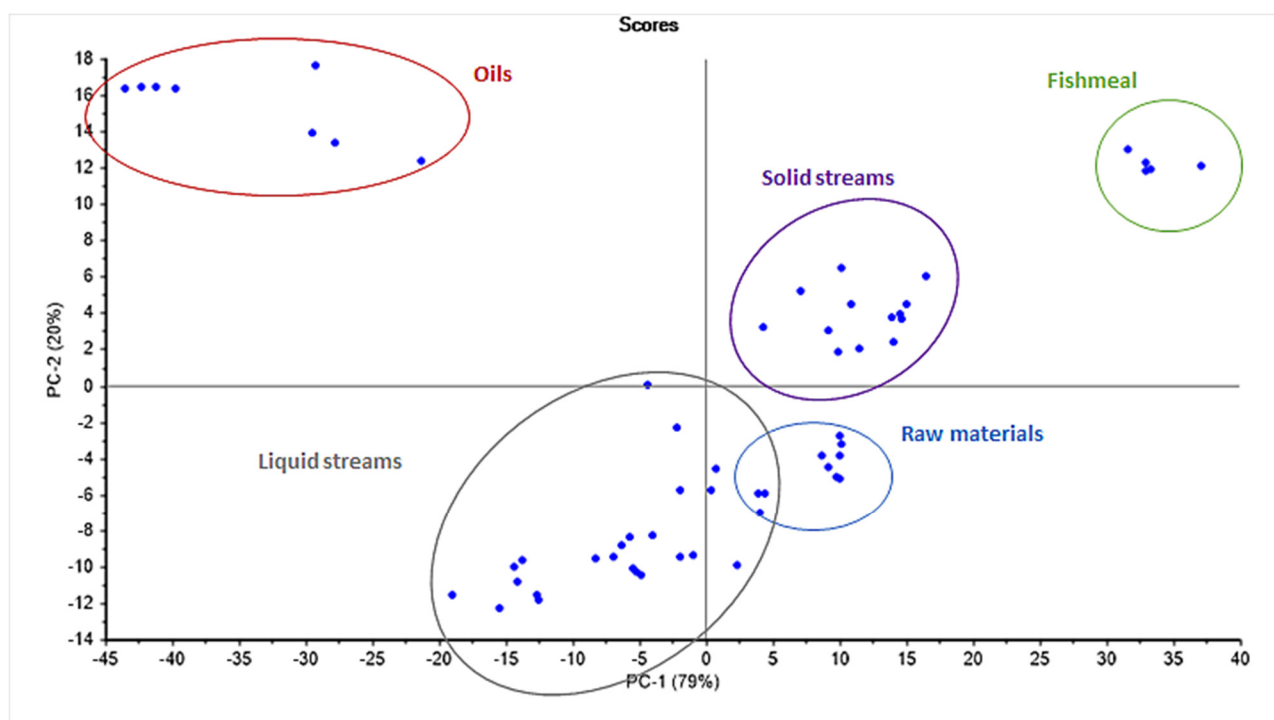
Figure 3. Principal Component Analysis (PCA) of NIR spectral data, explaining 99% of variation between samples. Sample groupings are marked with coloured ovals: blue (raw materials), purple (solid streams), grey (liquid streams), red (oils), and green (fishmeal).

3.3. PLS Quality Prediction Models

The results of valid concentration ranges; correlation factors (R^2); root-mean-square errors (RMSE); standard errors of calibration and cross validation (RMSEC and RMSECV, respectively); the optimal number of model factors (latent variables); and the number of measurements, used both for the calibration models and the independent test set cross validation modelling of the built PLS models, were listed and compared to assess which data treatments resulted in the best prediction models (Table A2 in Appendix B). An overview of the best performing PLS models for each chemical variable is shown in Figures 4–6.

Several combinations of variable selections, including either the whole spectrum or chosen spectral regions, were tested. Selecting the whole spectrum for modelling may lead to the inclusion of noise in the modelling data [14]. Appropriate data pre-processing, such as smoothing of the spectra, can be applied to decrease such detrimental effects and improve the signal-to-noise ratio [15]. Several spectral data pre-processing treatments were thus tested to find the strongest PLSR prediction models, including using the raw data directly, applying data normalization and baseline corrections, first and second derivatives, and multiplicative scatter signal correction (MSC). During the evaluation of the first and second derivative spectra, the data was smoothed by removing 11 spectral points from each end of the spectra to improve to signal-to-noise ratio and make the prediction models stronger [15].

Nevertheless, the comparison indicated that using the whole spectrum seemed to result in the strongest prediction models, potentially due to the high overlap of absorption bands within the NIR spectra, especially C-H and O-H bands relating to the moisture, protein, and lipid content of the samples [26]. Thus, selecting the whole spectrum did not seem to interfere with the prediction performance of the PLS models in the current study. Furthermore, applying no spectral pre-treatment (raw data), only baseline correction in the case of lipid content prediction, or MSC data pre-treatment provided the overall strongest prediction models. Even more precise models can potentially be obtained by applying ordered predictors selection (OPS) algorithms [26], but that is left for further

studies. The best data treatments obtained for each variable in the current study are discussed in Sections 3.3.1–3.3.5.

Figure 4. Partial least square prediction models from NIR spectra for water (**top left**), lipid (**top right**), fat-free dry matter, and phospholipid content. Calibration values are shown in blue, and model prediction test set values in red.

Figure 5. Partial least square prediction models from NIR spectra for lipid saturation class—SFA, MUFA, and PUFA—content during fishmeal and oil processing. Calibration values are shown in blue, and model prediction test set values in red.

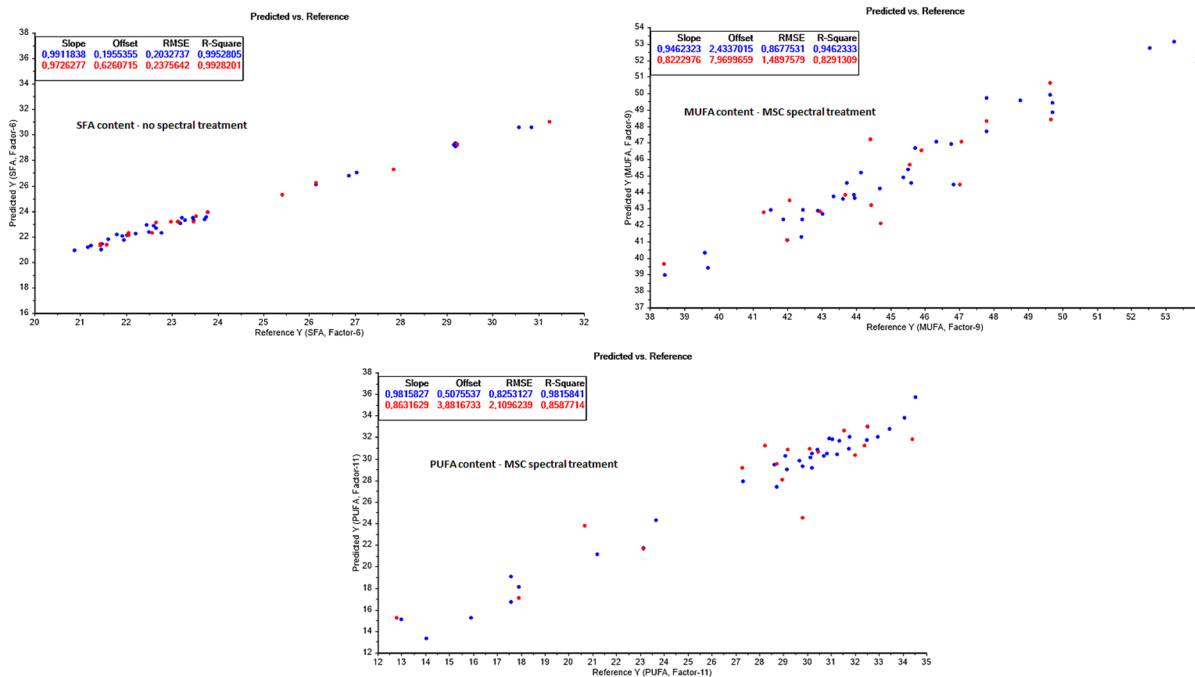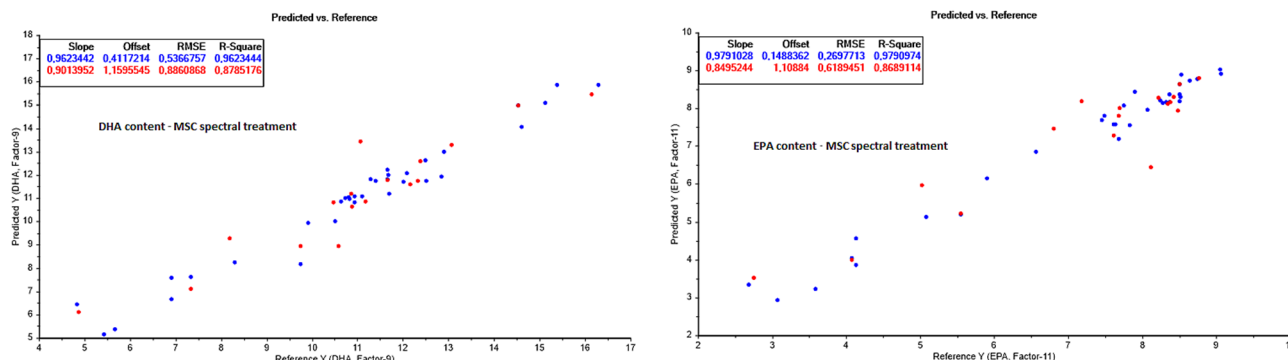**Figure 6.** Partial least square prediction models from NIR spectra for DHA and EPA content during fishmeal and oil processing. Calibration values are shown in blue, and model prediction test set values in red.

3.3.1. Water Content Prediction

The water content of the samples obtained throughout processing ranged from 0.3 to 93 g water/100 g sample (Figure 4). The strongest PLRS models predicting the water contents of the samples were obtained by using the no spectral data treatment, resulting in calibration and cross validation coefficients of R_C^2 of 0.9995 and R_{CV}^2 of 0.9938, respectively. The obtained model had an RMSEC of 0.67 g/100 g sample and an RMSECV of 2.41 g/100 g sample, indicating that the model was effective for process assessment and precise enough for process decision making, even though the RMSECV could be improved. The RMSECV is still higher than the values obtained in other studies, such as Cozzolino et al. (2002) [4], who obtained standard errors of calibration and cross validation (SEC and SECV, respectively) of 3.9 g/kg dry weight in the final fishmeal alone (equivalent to an RMSEC of approximately 0.4 g/100 g sample). However, since the moisture content range in the studied samples throughout processing is much wider than in the fishmeal alone, the achieved model parameters are well within acceptable levels for water content prediction during processing.

A robust way of predicting water content while optimizing or redesigning any food production process with water content ranging from 0 to 93% is a valuable tool. With active monitoring any changes in the process can be assessed, and production can immediately be directed towards the processing of products that better fit the processing stream characteristics. As water and lipids account for 85–90% of the raw materials entering the fishmeal and fish oil production process [27], controlling and ensuring effective water and lipid removal from the solid streams remains a priority during the production of high-quality protein products [1,16]. Furthermore, the abundance of water can stimulate hydrolyzation of the raw material along with heat and oxygen [28], which can have serious degrading effects on the quality and storage stability of the products. This further emphasizes the need for immediate action if negative characteristics of the processing material are monitored during production. The developed NIR water prediction model is fully capable of predicting such quality changes during the process.

3.3.2. Total Lipid Content Prediction

The total lipid content ranged from 1.2 to 99.7 g/100 g sample in the samples during the fishmeal processing, although the raw material entering the process had an average lipid content of 19.5 ± 2.0 g/100 g sample (Figure 4). This lipid content agrees with the lipid content obtained from Atlantic herring and mackerel samples in earlier studies [6,29]. The prediction model results obtained for the lipid prediction models were similar to those of the water models, with regard to the fact that the MSC gave the overall strongest prediction model when taking both the calibration and external test set cross validation parameters into account. Interestingly, adequate models were achieved with all spectral data treatments, with each having calibration and cross validation R^2 coefficients in the range of 0.9265 to 0.9773, except the second derivative, in which the model validation was far below

expectation with a R^2_{CV} of 0.5314. The lower performance of the second derivative models may be caused by the variable selection for the modelling (full spectra with smoothing), in which the smoothing may have been insufficient, or the full second derivative spectra may contain too much noise for the modelling. However, since acceptable prediction models were achieved with other spectral treatments, the optimization of variable selection for the second derivative spectra was not investigated further. It is also possible that the choice of samples in the external test set might have influenced the prediction strengths of the models. A few different validation setups were therefore tested as well, including randomized validation and full validations, but these did not perform any better than modelling with the external, independent test set validation. Goldstein et al. (2021) [30] studied the use of FT-NIR for the quality and physiological condition assessment of farmed Pacific cod. The study obtained acceptable prediction models for screening purposes for the lipid content and lipid-based indices (r^2 = 0.68, RMSE = 0.94%). Furthermore, their models only covered a lipid content range between 1 and 8% wet weight, which showcases the limitations of that study. The current study, however, produced prediction models that are valid and acceptable for quality control from 1.2 to 99.7 g/100 g sample, as stated earlier.

Measuring lipid quality, including lipid content, with traditional laboratory methods can take several months of labour when sampling numbers are high and frequent [1], and it includes high material costs involved with glassware, chemicals, etc. Furthermore, the traditional analysis commonly uses toxic solvents, such as chloroform, which may have a negative impact on the health of the staff performing the analysis as well as having negative environmental impacts during disposal [31]. Hence, predicting the lipid content by NIR spectroscopy may save a lot of time and money, is safer to the analytical staff, and has a lower environmental impact. Furthermore, human error can be minimized by applying the NIR analysis compared to traditional lipid extraction protocols, such as the Bligh and Dyer method [18], and the staff training periods could be reduced when using NIR spectroscopy for the task. The fishmeal produced during this study using the current traditional processing methods resulted in a high lipid content (11–14 g/100 g sample), which classifies it as a Type C fish protein concentrate (FPC) according to FAO [27]. Online monitoring of the lipid content during processing could allow improved process monitoring, thus lowering the lipid content in the final fishmeal product. If the lipid content is lowered below 0.75 g/100 g, the product would be classified as a Type A FPC, which are considered fit for human consumption [32]. Direct monitoring and appropriate process adjustments could therefore substantially increase the value of the produced products, as well as increase the possibility of producing more standardized products.

3.3.3. Fat-Free Dry Matter (FFDM) Prediction

The FFDM mainly consisted of proteins, but also smaller amounts (<2 g/100 g sample) of salts, minerals, and other trace elements. The obtained FFDM content during processing ranged from 0 to 84.4% (Figure 4). Again, the MSC model performed best, resulting in an R^2_C and R^2_{CV} of 0.9183 and 0.9356, respectively. The RMSE were relatively high, at 6.23 and 5.58 g/100 g sample for the calibration and cross validation, respectively, using the MSC prediction model. Visual evaluation of the sample distribution during modelling showed that the oil samples appeared as extremes during the modelling. However, the model strength became weaker if the oil samples were removed, indicating the importance of including a wide range of FFDM content samples in order to obtain strong prediction models. Models with similar prediction capacities (with similar R^2_{CV} values >0.85, and SECV of >3.7) were obtained by Cazzolino et al. (2002) [4] for the water, lipid, and crude protein contents in fishmeal, and by Wold and Isaksson (2006) [33] in Atlantic salmon. The fishmeal study included a higher number of samples in the calibration models (n), but on the downside, the models were only part of a very limited composition range (34–140 g/kg dry matter (DM) for water content, 48–173 g/kg DM for lipids, and 605–728 g/kg DM for crude protein). A similar issue was observed in the Atlantic salmon study [33], which showed acceptable prediction models for fat in the narrower range from 8.8–19.2% (R = 0.87,

RMSECV = 1.12%) and moisture content from 61.0–70.8% (R = 0.86, RMSECV = 0.98%). The wider water, lipid, and FFDM content ranges that the prediction models obtained in the current study highlight the value of the current prediction models, which can be used to characterize much more diverse raw and processing materials than the comparative fish studies mentioned above.

The main FFDM component in fishmeal and its production streams is crude protein, which is traditionally measured by determining the nitrogen content, e.g., with the Kjeldahl method [34] or the Dumas method [35]. Neither the Kjeldahl nor the Dumas method distinguish between protein and non-protein nitrogen [35]. However, since the NIR absorption is highly dependent on the molecular composition, the micro-structure of the molecule, and the environment that the molecule is in [15], NIR spectroscopy has the potential to assess the protein content with higher precision than the traditional methods. This also highlights the importance of choosing the reference method wisely when building the NIR prediction models, as this could affect the prediction ability and precision of the models.

The FFDM in this study is expressed as what is left when the water and lipid content has been removed from the total weight of the sample. Even more precise prediction of the protein content and FFDM might have been achieved if direct measurements of the protein content through the Kjeldahl or Dumas method had been applied as reference data during the FFDM NIR prediction model building, along with more precise analyses of the vitamins and minerals potentially present in the samples. The precision of the Kjeldahl and Dumas methods in assessing protein content has, however, also been criticized lately, and there is not a uniform consensus on which multiplication factors should be used when converting the analysed nitrogen content to crude protein content. Most studies refer to the 6.25 multiplication factor, as derived from the standard Kjeldahl method, while others have shown that the factor requires modifications dependent on the characteristics of the sample, as well as the choice of analytical method (Kjeldahl or Dumas) [35]. However, this comparison between the precision of different protein assessment methods is outside the scope of the current study, but is worthy of further comparisons in the future. Nevertheless, the current study clearly shows that both time and cost can be saved by applying NIR spectroscopy to analyse FFDM during fishmeal processing. As the value of fishmeal depends on a high protein content, an online system monitoring the protein content throughout each operational step could help keep the final product's quality consistent, which is of great importance to both the producers and the consumers [36].

3.3.4. Fatty Acid Composition (FAC) Prediction

The fatty acid composition was analysed and NIR prediction models were built for the fatty acid classes SFA, MUFA, and PUFA, along with DHA and EPA. Excellent prediction models were achieved for all parameters (Figure 5), especially SFA, DHA, and EPA, which had high R^2 parameters (>0.98) and low prediction errors (<1 g/100 g lipid) for all lipid class parameters. Slightly higher prediction errors were obtained for MUFA (RMSECV = 1.5 g/100 g lipid) and PUFA (RMSECV = 2.1 g/100 g lipid) compared to SFA (RMSECV = 0.23 g/100 g lipid), potentially due to the spectral similarities and high absorption peak overlap between the unsaturated bonds in the two fatty acid classes, making it difficult to distinguish between the levels of unsaturation. However, the performances of the MUFA and PUFA prediction models are still within satisfactory limits for process monitoring and optimization purposes. This agrees with earlier studies, showing the potential of using FT-NIR for rapid screening and monitoring of fatty acid quantification in fats and oils [37].

Prediction models for DHA and EPA using the MSC spectral data treatment turned out to be highly efficient during processing, showing R^2_{CV} parameters of 0.8785 and 0.8689, and RMSECV values of 0.89 and 0.62 g/100 g sample for the DHA and EPA classes, respectively (Figure 6). DHA and EPA are desirable omega-3 fatty acids which can be destroyed during drying, as high temperatures drastically affect the long-chain PUFAs [34]. Hence, heating and drying techniques need to be monitored closely for high yield. As

DHA and EPA are among the desirable PUFAs due to various health effects [23–25], a higher yield would be beneficial. The prediction of these fatty acids during processing is therefore especially important due to their potential health effects and their influence on product value. Furthermore, no other studies were found in the literature predicting these parameters in fishmeal or during fishmeal production, supporting the novelty of the current study results.

3.3.5. Free Fatty Acids (FFA) and Phospholipids (PL) Prediction

No significant correlations were obtained for the free fatty acids in the samples, possibly due to the narrow range of obtained FFA values (ranging between 0 and 5 g FFA/100 g sample). However, a good PLS model was achieved for the phospholipid content (PL) in the range between 0 and 1.4 g PL/100 g sample with the MSC spectral data treatment, resulting in a prediction model with an R_C^2 and R_{CV}^2 of 0.9617 and 0.8617, respectively (Figure 4). The low RMSEC and RMSECV of 0.06 and 0.11 g/100 g sample furthermore indicate the excellent prediction precision of the model compared to the reference measurements.

Free fatty acid and phospholipid content are strong indications of lipid and protein degradation, and analysis thereof can emphasize and identify problematic areas or processing steps which involve hydrolyzation- or lipid oxidation-induced degradation [1,16,28]. However, further investigations are proposed with samples containing a wider range of FFAs, in order to build potential FFA prediction models using the NIR technology.

4. Conclusions

NIR spectroscopy-based PLS modelling was shown to be a precise and accurate tool for the simultaneous assessment of water, lipid, and FFDM content, as well as the fatty acid composition, including EPA, DHA, and PLs during fishmeal processing. Overall, two spectral treatments resulted in stronger prediction models than the others, i.e., the spectral model using no spectral treatments (raw spectra), and the model using MSC-treated spectra, respectively. The models can potentially be improved even further by adding an even wider range of raw materials (i.e., including data on more species during processing) or applying more detailed variable selections as inputs for the multivariate PLS model building. Investigations of this is left for future studies. However, the developed composition models can substantially decrease labour and material costs. Integrating NIR for process monitoring can be promoted as a more sustainable method to monitor processing-induced changes, with lower environmental impacts compared to the traditional quality measurements and processing standards. More effective process control leads to further increases in production yield and better utilization of the available biomass, which also has substantial environmental benefits. How large these potentially positive environmental impacts are, and which categories are most affected, is left for further studies.

Overall, NIR online monitoring allows better process control and the direction of the process towards the production of higher-value products, potentially even for human consumption, and it also facilitates the production of more specialized feed for aquaculture, agriculture, and pets.

Author Contributions: Conceptualization, M.G., G.S.H., S.A. and Ó.Ö.; methodology, M.G. and G.S.H.; software, M.G.; validation, M.G.; formal analysis, M.G. and G.S.H.; investigation, M.G. and G.S.H.; resources, M.G. and S.A.; data curation, M.G. and G.S.H.; writing—original draft, M.G.; writing—review and editing, M.G., G.S.H., Ó.Ö. and S.A.; visualization, M.G.; supervision, M.G., S.A. and Ó.Ö.; project administration, M.G.; funding acquisition, M.G., S.A. and Ó.Ö. All authors have read and agreed to the published version of the manuscript.

Appendix A. Assignment of NIR Overtones and Combination Bands

Table A1. Assignment of NIR overtone and combination bands obtained during pelagic fishmeal and oil processing. Adapted and modified from Shiroma and Rodriguez-Saona (2009) [38] and Cozzolini et al. (2009) [7].

Approximate Peak Maximum Wave Numbers (cm^{-1})	Functional Group Assignment	Dominant Chemical Component
10,600	O-H stretch 2nd overtone	Water
8600–8150	C-H 2nd overtone	Lipids
7400–7000	2nd overtone of C-H combinations	Lipids
6800	O-H 1st overtone	Water
5750	C-H 1st overtone of CH_3	Lipids
5500	C-H 1st overtone of -HC=CH-	Lipids
5200–5000	O-H stretch 1st overtone of O-H combinations	Water
5100–4500	N-H combination peptide absorption bands of amide groups	Proteins
4370	C-H+C-H combination bands -CH_2, -CH_3	Lipids
4350–4150	C-H combination bands C=C stretch, -CH, -CH_2	Lipids

Appendix B. NIR Prediction Models

Table A2. The NIR prediction model summary for the calibration and cross validation parameters for the prediction of the chemical compositions of fishmeal and oil during processing.

Spectral Data Pre-Treatment	Valid Range of Reference Parameter	Calibration			RMSEC	Cross Validation with Independent Test Set		RMSECV
		n	Optimal LV	R_C^2		n	R_{CV}^2	
Water (g/100 g sample)	**0.3–93.0**							
None		**60**	**2**	**0.9995**	**0.67**	**30**	**0.9938**	**2.41**
Baseline correction		60	3	0.9666	5.56	30	0.9631	5.88
1st derivative		60	3	0.9404	7.42	30	0.9348	7.82
2nd derivative		60	5	0.9793	4.38	30	0.8142	13.20
MSC		60	3	0.9498	6.81	30	0.9477	7.00
Lipids (g/100 g sample)	**0–100**							
None		60	4	0.9640	4.11	30	0.9749	4.97
Baseline correction		**60**	**3**	**0.9669**	**3.94**	**30**	**0.9773**	**3.94**
1st derivative		60	2	0.9265	5.87	30	0.9404	7.66
2nd derivative		60	3	0.9828	2.84	30	0.5314	21.49
MSC		60	3	**0.9681**	**3.86**	30	**0.9592**	**6.34**

Table A2. *Cont.*

Spectral Data Pre-Treatment	Valid Range of Reference Parameter	Calibration			RMSEC	Cross Validation with Independent Test Set		
		n	Optimal LV	R^2_C		n	R^2_{CV}	RMSECV
FFDM (g/100 g sample)	**0–84.4**							
None		60	4	0.8899	8.36	30	0.9063	7.01
Baseline correction		60	4	0.8473	8.52	30	0.8969	7.06
1st derivative		60	5	0.6743	12.44	30	0.8590	8.25
2nd derivative		60	4	0.8570	8.24	30	0.8003	9.82
MSC		60	7	**0.9183**	**6.23**	30	**0.9356**	**5.58**
FFA (g/100 g sample)	**0–5**							
None		52	1	0.0911	0.91	25	0.0242	0.99
MSC		52	2	0.0785	0.92	25	0.0583	0.98
PL (g/100 g sample)	**0–1.4**							
None		44	6	0.4476	0.24	22	0.4841	0.22
Baseline correction		44	8	0.6506	0.19	22	0.5453	0.20
1st derivative		44	9	0.8888	0.11	22	0.6033	0.19
2nd derivative		44	1	0.3427	0.26	22	0.3055	0.25
MSC		44	10	**0.9617**	**0.06**	22	**0.8617**	**0.11**
SFA (g/100 g lipids)	**20–32**							
None		34	6	**0.9953**	**0.20**	17	**0.9928**	**0.24**
Baseline correction		34	9	0.9464	0.68	17	0.6922	1.56
1st derivative		34	15	0.9989	0.10	17	0.6668	1.62
2nd derivative		34	1	0.4147	2.26	17	0.3149	2.32
MSC		34	12	**0.9889**	**0.31**	17	**0.8363**	**1.13**
MUFA (g/100 g lipids)	**38–54**							
None		34	9	**0.9062**	**1.15**	17	**0.7968**	**1.62**
Baseline correction		34	5	0.6815	2.11	17	0.6200	2.22
1st derivative		34	4	0.3882	2.93	17	0.4230	2.74
2nd derivative		34	1	0.2863	3.16	17	0.2532	3.11
MSC		34	9	**0.9462**	**0.87**	17	**0.8291**	**1.49**
PUFA (g/100 g lipids)	**12–36**							
None		34	9	**0.9371**	**1.53**	17	**0.8461**	**2.20**
Baseline correction		34	6	0.7060	3.30	17	0.6051	3.53
1st derivative		34	4	0.4791	4.39	17	0.4924	4.00
2nd derivative		34	1	0.3196	5.02	17	0.3087	4.67
MSC		34	11	**0.9816**	**0.83**	17	**0.8588**	**2.11**
DHA (g/100 g lipids)	**4–17**							
None		34	8	**0.9073**	**0.84**	17	**0.8146**	**1.09**
Baseline correction		34	6	0.7320	1.43	17	0.6302	1.55
1st derivative		34	4	0.5864	1.78	17	0.6197	1.57
2nd derivative		34	14	0.9982	0.12	17	0.3380	2.07
MSC		34	9	**0.9623**	**0.54**	17	**0.8785**	**0.89**

Table A2. *Cont.*

Spectral Data Pre-Treatment	Valid Range of Reference Parameter	Calibration			RMSEC	Cross Validation with Independent Test Set		
		n	Optimal LV	R_C^2		n	R_{CV}^2	RMSECV
EPA (g/100 g lipids)	**2–10**							
None		**34**	**10**	**0.9536**	**0.40**	**17**	**0.8278**	**0.71**
Baseline correction		34	4	0.5110	1.30	17	0.6366	1.03
1st derivative		34	4	0.5146	1.30	17	0.5077	1.20
2nd derivative		34	1	0.3437	1.51	17	0.3335	1.40
MSC		**34**	**11**	**0.9791**	**0.27**	**17**	**0.8689**	**0.62**

n = number of samples, LV = latent variables, R_C^2 = calibration correlation factor, R_{CV}^2 = cross validation correlation factor obtained in independent test set, RMSEC = root-mean-square error of calibration, RMSECV = root-mean-square error of cross validation.

References

1. Hilmarsdóttir, G.S.; Ogmundarson, Ó.; Arason, S.; Gudjónsdóttir, M. Efficiency of fishmeal and fish oil processing of different pelagic fish species: Identification of processing steps for potential optimization toward protein production for human consumption. *J. Food Proc. Pres.* **2021**, *45*, e15294. [CrossRef]

2. Hall, G.M. Fishmeal production and sustainability. In *Fish Processing: Sustainability and New Opportunities*; Hall, G.M., Ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2010; pp. 207–235.

3. Oterhals, Å.; Vogt, G. Impact of extraction, refining and concentration stages on the stability of fish oil. In *Food Enrichment with Omega-3 Fatty Acids*; Jacobsen, C., Nielsen, N.S., Horn, A.F., Sørensen, A.-D.M., Eds.; Woodhead Publishing: Sawston, UK, 2013; pp. 111–129.

4. Cozzolino, D.; Chree, A.; Murray, I.; Scaife, J.R. The assessment of the composition of fishmeal by near infrared reflectance spectroscopy. *Aquacult. Nutr.* **2002**, *8*, 147–155. [CrossRef]

5. Hilmarsdóttir, G.S.; Ogmundarson, Ó.; Arason, S.; Gudjónsdóttir, M. Identification of environmental hotspots in fishmeal and fish oil production for process optimization purposes. *J. Clean. Prod.* **2022**, *343*, 130880. [CrossRef]

6. Romotowska, P.E.; Karlsdóttir, M.G.; Gudjónsdóttir, M.; Kristinsson, H.G.; Arason, S. Seasonal and geographical variation in chemical composition and lipid stability of Atlantic mackerel (*Scomber scombrus*) caught in Icelandic waters. *J. Food Comp. Anal.* **2016**, *49*, 9–18. [CrossRef]

7. Cozzolino, D.; Chree, A.; Murray, I.; Scaife, J.R. Usefulness of near infrared spectroscopy to monitor the extent of heat treatment in fish meal. *Int. J. Food Sci. Technol.* **2009**, *44*, 1579–1584. [CrossRef]

8. McClure, F.W. 204 years of near infrared technology: 1800–2003. *J. Near Infrared Spectro* **2003**, *11*, 487–518. [CrossRef]

9. Stevik, A.M.; Duun, A.S.; Rustad, T.; O´Farrel, M.; Schulerud, H.; Ottestad, S. Ice fraction assessment by near infrared spectroscopy enhancing automated superchilling process lines. *J. Food Eng.* **2010**, *100*, 169–177. [CrossRef]

10. Bøknæs, N.; Jensen, K.N.; Andersen, C.M.; Martens, H. Fresh assessment of thawed and chilled cod fillets packed in Modified atmosphere using near infrared spectroscopy. *LWT–Food Sci. Technol.* **2002**, *35*, 628–634. [CrossRef]

11. Nilsen, H.; Esaiassen, M. Predicting sensory score of cod (*Gadus morhua*) from visible spectroscopy. *LWT–Food Sci. Technol.* **2005**, *38*, 95–99. [CrossRef]

12. Ellekjær, M.R.; Isaksson, T.; Solheim, R. Assessment of sensory quality of meat sausages using near infrared spectroscopy. *J. Food Sci.* **2006**, *59*, 456–464. [CrossRef]

13. Blanco, M.; Villaroya, I. NIR spectroscopy: A rapid-response analytical tool. *Trends Anal. Chem.* **2002**, *21*, 240–250. [CrossRef]

14. Gomes, A.A.; Azcarate, S.M.; Dias Diniz, P.H.G.; de Sousa Fernandes, D.D.; Veras, G. Variable selection in the chemometric treatment of food data: A tutorial review. *Food Chem.* **2022**, *370*, 131072. [CrossRef] [PubMed]

15. Rinnan, Å.; van den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *Trends Anal. Chem.* **2009**, *28*, 1201–1222. [CrossRef]

16. Hilmarsdóttir, G.S.; Ogmundarson, Ó.; Arason, S.; Gudjónsdóttir, M. The effects of varying heat treatments on lipid composition during pelagic fishmeal production. *Processes* **2020**, *8*, 1142. [CrossRef]

17. *ISO 6496*; Animal Feeding Stuffs—Determination of Moisture and Other Volatile Matter Content. International Organization for Standardization: Geneva, Switzerland, 1999. Available online: https://www.iso.org/obp/ui/#iso:std:iso:6496:ed-2:v1:en (accessed on 10 March 2020).

18. Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Phys.* **1959**, *37*, 911–917. [CrossRef]

19. Lowry, R.R.; Tinsley, I.J. Rapid colorimetric determination of free fatty acids. *J. Am. Oil Chem. Soc.* **1976**, *53*, 470–472. [CrossRef] [PubMed]

20. Bernárdez, M.; Pastoriza, L.; Sampedro, G.; Herrera, J.J.R.; Cabo, M.L. Modified method for the analysis of free fatty acids in fish. *J. Agric. Food Chem.* **2005**, *53*, 1903–1906. [CrossRef]

21. Stewart, J.C.M. Colorimetric determination of phospholipids with ammonium ferrothiocyanate. *Anal. Biochem.* **1980**, *104*, 10–14. [CrossRef]

22. AOCS. *Official Method and Recommended Practices of the AOC*, 7th ed.; AOCS Official Method Ce 1b-89; AOCS: Champaign, IL, USA, 1998.

23. Fournier, V.; Destaillats, F.; Juanéda, P.; Dionisi, F.; Lambelet, P.; Sébédio, J.-L.; Berdeaux, O. Thermal degradation of long-chain polyunsaturated fatty acids during deodorization of fish oil. *Eur. J. Lipid Sci. Technol.* **2006**, *108*, 33–42. [CrossRef]

24. Larsen, R.; Eilertsen, K.-E.; Elvevoll, E.O. Health benefits of marine foods and ingredients. *Biotech. Adv.* **2011**, *29*, 508–518. [CrossRef]

25. Zheng, J.-S.; Hu, X.-J.; Zhao, Y.-M.; Yang, J.; Li, D. Intake of fish and marine n-3 polyunsaturated fatty acids and risk of breast cancer: Meta-analysis of data from 21 independent prospective cohort studies. *BMJ* **2013**, *346*, f3706. [CrossRef] [PubMed]

26. Díaz, E.O.; Kawamura, S.; Matsuo, M.; Kato, M.; Koseki, K. Combined analysis of near-infrared spectra, colour, and physicochemical information of brown rice to develop accurate calibration models for determining amylose content. *Food Chem.* **2019**, *286*, 297–306. [CrossRef] [PubMed]

27. FAO. *The Production of Fish Meal and Oil*; FAO Fisheries Technical Paper—142; Aquaculture Feed and Fertilizer Resources Information System; FAO: Rome, Italy, 1986; ISBN 92-5-102464-2.

28. Jacobsen, C. Some strategies for the stabilization of long chain n-3 PUFA-enriched foods: A review. *Eur. J. Lipid Sci. Technol.* **2015**, *117*, 1853–1866. [CrossRef]

29. Dang, T.T.H.; Gudjónsdóttir, M.; Ren, D.; Karlsdóttir, M.G.; Nguyen, M.V.; Tómasson, T.; Arason, A. Effects of pre and postrigor freezing and temperature stability during frozen storage on physicochemical stability of Atlantic herring (*Clupea harengus*) muscle). *J. Food Proc. Pres.* **2018**, *42*, e13754. [CrossRef]

30. Goldstein, E.D.; Helser, T.E.; Vollenweider, J.J.; Sreenivasan, A.; Sewall, F.F. Rapid and reliable assessment of fish physiological condition for fisheries research and management using Fourier Transform Near-infrared spectroscopy. *Front. Mar. Sci.* **2021**, *8*, 690934. [CrossRef]

31. Watts, P.; Long, G.; Meek, M.E. Chloroform. In *Concise International Chemical Assessment Document 58*; World Health Organization: Geneva, Switzerland, 2004.

32. Windsor, M.L. Fish Protein Concentrate. FAO Fisheries and Aquaculture Department. Available online: http://www.fao.org/3/x5917e00.htm#Contents (accessed on 7 February 2020).

33. Wold, J.P.; Isaksson, T. Non-destructive determination of fat and moisture in whole Atlantic salmon by near-infrared diffuse spectroscopy. *J. Food Sci.* **2006**, *62*, 734–736. [CrossRef]

34. *ISO 5983-2*; Animal Feeding Stuffs—Determination of Nitrogen Content and Calculation of Crude Protein Content—Part 2: Block Digestion and Steam Distillation Method. International Organization for Standardization: Geneva, Switzerland, 2009. Available online: https://www.iso.org/standard/52199.html (accessed on 5 January 2022).

35. *ISO 16634-1*; Food products—Determination of the Total Nitrogen Content by Combustion According to the Dumas Principle and Calculation of the Crude Protein Content—Part 1: Oilseeds and Animal Feeding Stuffs. International Organization for Standardization: Geneva, Switzerland, 2008. Available online: https://www.iso.org/obp/ui/#iso:std:iso:16634:-1:ed-1:v1:en (accessed on 5 January 2022).

36. Kristinsson, H.G. Aquatic food protein hydrolysates. In *Maximising the Value of Marine By-Products*; Shahidi, F., Ed.; Woodhead Publishing Series in Food Science; Technology and Nutrition: Sawston, UK, 2007; pp. 229–248.

37. Azizian, H.; Kramer, J.K.G. A rapid method for the quantification of fatty acids in fats and oils with emphasis on trans fatty acids using fourier transform near infrared spectroscopy (FT-NIR). *Lipids* **2005**, *40*, 855–867. [CrossRef]

38. Shiroma, C.; Rodriguez-Saona, L. Application of NIR and MIR spectroscopy in quality control of potato chips. *J. Food Comp. Anal.* **2009**, *22*, 596–605. [CrossRef]