sustainability

MDPI

Article

Non-Destructive Analysis Using Near-Infrared Spectroscopy to Predict Albumin, Globulin, Glutelin, and Total Protein Content in Sunflower Seeds

Cecile Levasseur-Garcia [1,*], Pierre Castellanet [2], Camille Henry [3], Christelle Florin [3], Marion Laporte [4], Virginie Mirleau-Thebaud [5], Sandrine Plut [6] and Anne Calmon [1]

[1] Laboratoire de Chimie Agro-Industrielle (LCA), Toulouse University, INRAE, INPT, INP-Purpan, 31000 Toulouse, France
[2] Caussade Semences, 82300 Caussade, France
[3] MAS Seeds, 40280 Haut Mauco, France
[4] RAGT 2n, 12000 Rodez, France
[5] Syngenta France, 31790 Saint-Sauveur, France
[6] SOLTIS, 31700 Mondonville, France
* Correspondence: cecile.levasseur@purpan.fr

Abstract: This pilot study explores the potential of near-infrared spectroscopy (NIRS) for predicting sunflower seed protein content, focusing on both crushed and husked samples to address agricultural sustainability concerns. Sunflower seeds are renowned for their richness in both oil and protein content. The important role of sunflower seeds in the food and feed industries underscores the importance of using precise analytical tools to determine their composition. In essence, the nature of the hull of sunflower seeds, which skews the interaction between the seed and light, necessitates a sophisticated analysis. This study analyzes 326 samples using a near-infrared spectrometer to develop robust partial least squares (PLS) models. High accuracy is achieved in predicting total protein for crushed samples (r^2c = 0.97, RMSEC 0.54%, RPDc 6; r^2p = 0.78, RMSEP 1.24%, RPDp 2.1). Extending the scope to husked samples, promising results emerge for crude protein prediction (r^2c = 0.93, RMSEC 0.86%, RPDc 3.9; r^2cv = 0.83, RMSECV 1.39%, RPDcv 2.4). Additionally, this study delves into protein fractions (globulin, albumin, and glutelin) in crushed seeds, adding depth to the analysis. In conclusion, NIR spectroscopy proves valuable for rapid prescreening in breeding, especially when working with hulled grains, offering non-destructive efficiency and predictive accuracy in agricultural analysis. The novel exploration of protein fractions in sunflower seeds further enhances this study's importance, providing a valuable contribution to the field and underscoring the practical applications of NIR spectroscopy in sustainable agriculture. In conclusion, the opacity of sunflower seed hulls poses challenges in infrared spectroscopy, limiting light penetration and accuracy. Dehulled seeds are preferred for reliable results, overcoming hull-related limitations. Although grinding provides the advantages of uniformity and reproducibility for near-infrared (NIR) spectroscopy, the preference for dehulled grains persists. The practical need for accurate analysis in agriculture and breeding drives the choice of spectroscopy on dehulled seeds, allowing for replanting.

Keywords: sustainable analysis; husked seed samples; intact seed samples; protein; infrared spectroscopy; sunflower seeds; NIRS; *Helianthus annuus*

Citation: Levasseur-Garcia, C.; Castellanet, P.; Henry, C.; Florin, C.; Laporte, M.; Mirleau-Thebaud, V.; Plut, S.; Calmon, A. Non-Destructive Analysis Using Near-Infrared Spectroscopy to Predict Albumin, Globulin, Glutelin, and Total Protein Content in Sunflower Seeds. *Sustainability* **2024**, *16*, 737. https://doi.org/10.3390/su16020737

Academic Editor: Hossein Azadi

Received: 10 November 2023
Revised: 4 January 2024
Accepted: 8 January 2024
Published: 15 January 2024

1. Introduction

Sunflower (*Helianthus annuus* L.) stands as one of the most extensively cultivated crops, boasting a global production of approximately 50 million tons in 2020 (FAO Statistics Division 2023) [1]. Sunflower (*Helianthus annuus* L.) oil and meal, derived from the seeds of the sunflower plant, are commonly used in the food and feed industries globally, owing to their versatility and nutritional benefits [2]. Indeed, although sunflower seeds are mainly

cultivated for their oil, they are also used as oilcakes because of their high protein content (protein accounts for an average of 22% of the dry matter of the seeds) [3,4].

The global demand for protein is increasing worldwide. It can be explained by several factors. Firstly, constant population growth is contributing to a natural increase in the demand for food, including protein. At the same time, improved living standards in many parts of the world are leading to a shift in eating habits towards higher-protein diets. In addition, the expansion of the animal feed industry, particularly in developing countries, is contributing to an increased demand for protein in livestock production. The combination of these factors has led to a significant increase in global protein demand. Simultaneously, a rise in plant-based protein demand reflects a growing interest in sustainable and health-conscious dietary choices. To meet this growing demand, sunflower stands out as a promising candidate to meet the increasing demand for plant-based protein sources [5–7].

Significant efforts have then been devoted to developing and optimizing protein extraction and characterization techniques [8–10]. Unfortunately, few authors have studied protein fractions. For sunflower seeds, the literature describes four groups of storage proteins, which are classified according to their solubility and characterized by their sedimentation coefficient. These are globulin (helianthinin, 40% to 90% of total proteins), albumin (sunflower albumin or SFA, 10% to 35%), and, in lower concentrations, glutelin and prolamins [5,10,11]. The size of each protein fraction depends on the extraction conditions [12–15]. Indeed, chemical extraction methods use solvents or reagents to degrade cell walls and facilitate protein release, which can result in the extraction of a wide range of proteins but also in the denaturation or modification of certain protein structures depending on the chemicals used. Enzymatic extraction, on the other hand, employs specific enzymes to target and degrade cellular components, allowing for a more targeted extraction of proteins based on the enzymes' specificity. A combination of chemical and enzymatic treatments can provide a balanced approach, optimizing extraction efficiency by degrading cell barriers with chemicals and then selectively targeting specific proteins with enzymes. It depends also on the genetic variability of sunflower protein fractions.

The classical wet chemical methods used to determine the protein content of sunflower seeds are time-consuming and tedious, and the sample preparation destroys the seeds. Furthermore, these conventional methods are not sustainable, as the sample preparation involves the use of environmentally harmful chemicals. Thus, breeding projects would greatly benefit from a simple, non-destructive method to rapidly analyze sunflower seeds. One technology that may achieve this goal is near-infrared spectroscopy (NIRS), which is rapid, eco-friendly, non-destructive, and widely used to quantify nutritional components in cereals. NIRS has been used to predict water, oil, fiber, and total protein content in sunflower meal, with attractive results ($r^2 > 0.90$) [16,17]. Some groups have applied NIRS to intact seeds or husked grains [17–22] to predict oil and fatty acid content.

Wang et al. analyzed automatically hulled sunflower seeds for crude protein, crude fat, and fatty acid composition. The spectral range used was 850 to 2499 nm. The authors achieved good results, with correlation coefficients ranging from 0.94 to 0.975 [17].

Sato et al. used near-infrared spectroscopy to determine the fatty acid composition of sunflower seeds. The seeds were hulled using an automatic huller or hulled manually with a knife. The spectral range used was 1100 to 2500 nm. Oil prediction performance showed a correlation coefficient r equal to −0.993. The identified wavelength corresponds to 1724 nm [18].

González Martín et al. determined the fatty acid composition of sunflower seeds by NIRS scanning between 1445 and 2348 nm using a filter instrument. They obtained models with very interesting performance (r between 0.984 and 0.999) [22].

Akkaya determined the fatty acid composition of sunflower seeds by NIRS (between 400 and 2500 nm). The coefficients of determination (r2) in the calibration ranged from 0.615 to 0.996. The author concluded that NIRS was a reliable technique that can be used as a rapid screening tool for the fatty acid composition of sunflower seeds [21].

Pérez-Vich et al. tested near-infrared reflectance spectroscopy (NIRS) between 400 and 2500 nm to estimate the oil content and fatty acid composition of sunflower seeds. Models were developed on hulled and crushed seeds, as well as on oil. The authors showed that NIRS is a reliable and accurate technique for estimating these traits in sunflower oil (validation r2 between 0.97 and 0.99), crushed seeds (r2 from 0.92 to 0.98), and hulled seeds (r2 from 0.90 to 0.97). The authors concluded that it is not necessary to grind the seeds, as the results are just as accurate when obtained by analyzing hulled seeds [23].

Fassio et al. investigated NIRS as a technique for predicting moisture, oil, and crude protein content over a spectral range from 400 to 2500 nm. The coefficient of determination in the calibration (r2) was 0.90 for crude protein content. The authors concluded that NIRS is an appropriate technique to use as a tool for the rapid preselection of quality characteristics in breeding programs [20].

Finally, very few studies have focused on near-infrared spectroscopy applied to quantify proteins in sunflower (Table 1).

Table 1. Summary of studies of near-infrared spectroscopy applied to quantification of crude protein in sunflower.

Number of Samples/Sample Preparation	Referenced Method for Crude Protein Determination	Device, Spectral Range, and Chemometrics	Performance and Characteristic Wavelengths	Reference
300 samples of intact seeds (5 to 6 akenes)	Kjeldhal method Average: 304 to 305 g kg^{-1} [89.0–424.0 g kg^{-1}]	FOSS NIRS 6500 400–2500 nm Second derivative Modified partial least squares regression	Training on 250 samples and external validation on 37 samples $r^2cal = 0.96$ SECV = 13.1 g kg^{-1}, $r^2 = 0.72$ SEP = 17.1 g kg^{-1} Strong absorption bands related to oil and water contents: 1460 nm (O-H stretch overtone bond—water), 1720 and 1764 nm (C-H stretch first overtone bonds (oil and fatty acids)), 1930 nm: O-H bonds (water), 2100 and 2278 nm: C-H combination bonds	Fassio and Cozzolino, 2004 [20]
172 mechanically dehulled seeds (10 g for each sample)	Dumas method Average: 29.09% [10.50–60%]	FOSS DS6500 850–2499 nm SNV detrend Partial least squares regression	Training on 154 samples and external validation on 18 samples, r = 0.950, SEP = 1.399%, RPD = 2.9	Wang et al. 2020 [17]

This is explained by the fact that, until now, sunflowers were selected on the basis of their oil content and fatty acid composition. To our knowledge, no study has yet focused on predicting the fractional protein content of husked seeds. Our hypothesis is that we can analyze the crude protein content and fractional protein content of husked seeds with near-infrared spectroscopy. The objective of this study is thus (1) to compare the performance of infrared models to predict the total crude protein content of crushed and husked sunflower seeds, and (2) to study their albumin, glutelin, and globulin content by applying NIRS to crushed sunflower seeds.

2. Materials and Methods

2.1. Characterization of Sunflower Seed Samples

In France, Romania, and Ukraine, 326 (n = 326) sunflower (*Helianthus annuus* L.) samples obtained from several breeding lines were harvested in 2017, reflecting a geographical diversity that may influence the chemical composition of the seeds. Three types of fertilization were performed: "fertilization 1" with 85% nitrogen (N), 85% phosphorus (P), and 85% potassium (K); "fertilization 2," presenting a different combination, with 40% nitrogen, 48% phosphorus, and 36% potassium; and, lastly, "fertilization 3", characterized by more intensive fertilization, comprising 140% nitrogen, 60% phosphorus, and 60% potassium.

The samples were split into two sets, with 285 crushed samples and the remaining 41 samples serving as manually husked samples. Of the 285 samples, 16 were selected for the analysis of their protein fractions (i.e., the albumin, globulin, and glutelin content).

2.2. Determination of Total Protein Content by Dumas Method

After freeze-drying for 48 h, subsamples of 55 (±10%) grams were ground three times for 10 s in a mill (IKA M20, IKA®-Werke GmbH & Co. KG, Staufen, Germany). Their protein content was determined by applying the protein quantification method of Dumas combustion with a FLASH 4000 Nitrogen/Protein analyzer (Thermo Fisher Scientific, Paisley, UK) to 80 mg (±5%) of homogenized ground samples at 940 °C and with an internal liquid flow rate of 300 mL/min. Next, 80 mg of lyophilized sunflower meal was inserted into 10 mm diameter, 10 mm high tin capsules with a volume of 785 µL (Thermo Fisher Scientific Inc, No. 252 08000). Weighing was performed on an Entris® precision balance (Sartorius, Göttingen, Germany) with an accuracy of 10−2 mg. All analyses were performed in triplicate, and the results were averaged.

2.3. Determination of Albumin, Globulin, and Glutelin Content by Successive Extraction

Protein fractions were extracted from samples in several steps according to the Osborne protocol [11]. Specifically, 1.75 g of sunflower seed meal was placed in a 50 mL vial to which (i) 35 mL of Milli Q water to extract albumin, (ii) 35 mL of 1 M NaCl to extract globulin, and (iii) 35 mL of a pH 12 buffer to extract glutelin were then successively added. Liquid–solid separation was performed by centrifugation (10,000 rpm, 15 min, 20 °C). The supernatant was freeze-dried for 72 h before being weighed. For each protein extract, the albumin, globulin, and glutelin content was determined using a Flash 4000 instrument (Thermo Fisher Scientific Inc., Waltham, MA, USA) for homogeneous grinding. Each ground product was analyzed twice, and the average content was retained.

2.4. Collection of Near-Infrared Spectra on Crushed and Husked Sunflower Seeds

Near-infrared spectra were collected using a Bruker Fourier Transform Near-Infrared MPA™ instrument (Bruker Optics, GmbH, Ettlingen, Germany). Spectra were taken from a small cup placed on the analyzer and containing 5 g of the 285 crushed samples (FOSS, Hilleroed, Denmark). Spectra were also taken from the same small cup containing 5 g of husked sample (FOSS, Hilleroed, Denmark). The reflectance spectra spanned from 12.274 to 5.493 cm^{-1}, with an 8 cm^{-1} step and with three replicates per sample. The average of the three spectra was retained. The software Opus™ 6.5 (Bruker Optics, GmbH) served to configure the measurements and control the instrument.

2.5. Preprocessing and Statistical Analysis of Near-Infrared Spectra

The software XLSTAT (version 2020.1.1; Addinsoft, XLSTAT statistical and data analysis solution, Long Island, NY, USA; https://www.xlstat.com, 2019) was used for the descriptive analysis.

Classical mathematical preprocessing was applied to remove the effects of light scattering, highlight various spectral peaks, and eliminate multiplicative noise and baseline shifts. Six types of preprocessing were tested: no treatment (i.e., raw data were used), first-derivative (D1) and second-derivative (D2) Savitzky–Golay transformations, standard normal variate (SNV) ± D1, and SNV detrend (SNVD). Calibration equations were derived using partial least squares (PLS) regression between the infrared reflectance spectra (880 variables) and the chemical values [24].

To prevent overfitting [25], cross-validation was performed using the training data of the 41 husked samples to select both the optimal number of factors and the best preprocessing.

The 285 crushed samples were divided into two sets using the Kennard–Stone method [26]. One set was used in training to develop the model (80% of the samples—228 samples), and one set was used to test the model (20% of the samples—57 samples).

The performance of the PLS models was quantified by the coefficient of determination R2 and the root mean standard error of either training, cross-validation, or prediction (RMSEC, RMSECV, or RMSEP). The ratio of performance to deviation (RPD) served to define the predictive accuracy of the models when applied to new samples. The RPD is defined as the ratio between the standard deviation of the reference value and the RMSE of

a model [27]. RPD > 3 indicates that the model can be used to predict quantitative values, whereas 2 < RPD < 3 means that the model should be improved [28]. The best model was selected based on the highest coefficient of determination in cross-validation [29]. All data processing and modeling of infrared spectra were performed using The Unscrambler® (v. X; CAMO A/S, Oslo, Norway).

Concerning the protein fractions, the general procedure for classifying sunflower seeds according to protein fraction content involves two steps (Figure 1).

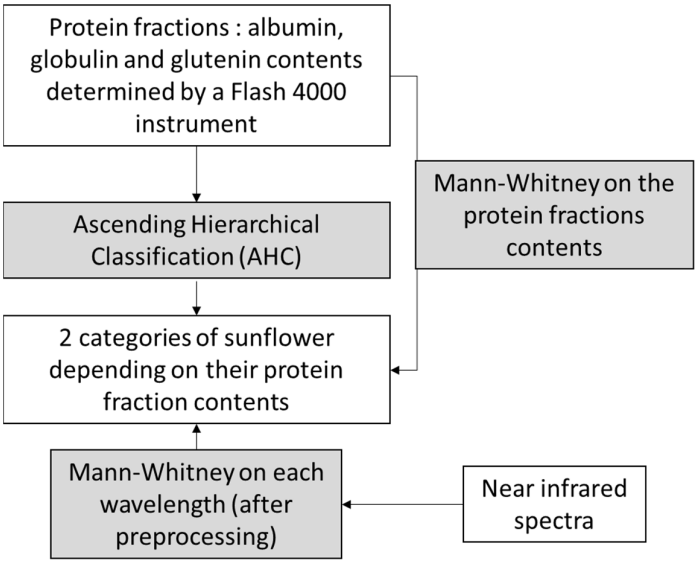

Figure 1. Procedure for classifying sunflower seeds regarding their protein fraction content.

The first step considered only the benchmark values obtained in the laboratory: the protein fraction content. The assigned values were categorized in an ascending hierarchical classification (AHC). The AHC was analyzed using XLSTAT version 2020.1.1, configured to use Euclidean distance for the dissimilarity parameter and Ward's aggregation method. Automatic truncation was performed with entropy. Differences among the groups were then investigated with the Mann–Whitney U test and p values less than or equal to 0.001 were evaluated as statistically significant. The Mann–Whitney U test was performed using Matlab 2020a Update 7 (9.8.0.1721703).

3. Results and Discussion

3.1. Compositional Characteristics of Total Content of Crude Protein, Albumin, Globulin, and Glutelin

Table 2 summarizes the descriptive statistics of the 326 sunflower samples.

Table 2. Composition of 326 sunflower samples in terms of their protein and protein fraction content.

Statistic	Crude Protein Content (%)			Albumin (%)	Globulin (%)	Glutelin (%)
	Crushed		Husked			
No. of observations	285	16 (of 285)	41	16	16	16
Minimum	10	14.6	11.5	1.7	6.1	0.8
Maximum	25.7	23.6	22.7	3.3	19.4	3.6
Mean	18.7	18.3	16.2	2.5	12.8	2
Standard deviation ($n-1$)	3.1	2.7	3.3	0.4	4.1	0.8

The total protein content of the 285 samples ranged from 10.0% to 25.7%, with a mean of 18.7% and a standard deviation of 3.1%. The minimum and maximum contents of the 41 husked samples are included in the contents of the 285 crushed samples. These results are consistent with published results [30–33].

In the 16 samples analyzed for their protein fractions, albumin content varied from 1.7% to 3.3%, globulin content from 6.1% to 19.4%, and glutelin content from 0.8% to 3.6%. Expressing these contents as a percent of total protein content, albumin content averages at 14% (between 11% and 19%), glutelin content averages at 11% (between 5% and 18%), and globulin content averages at 69% (between 41% and 91%). Although these values differ slightly from those reported by Bauchot and Merrien [34] (55–60% globulin, 17–23% albumin, and 11–17% glutelin), the pattern remains the same, with globulin constituting the majority. Figure 2 shows a cumulative histogram of the albumin, glutenin, and globulin content of the 16 sunflower samples whose protein fractions were measured.

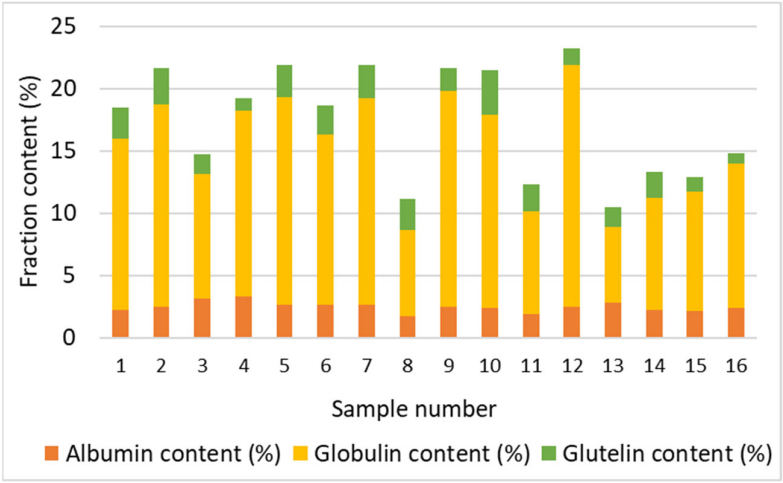

Figure 2. Albumin, globulin, and glutelin content of the 16 crushed samples of sunflower seeds.

The total of the three protein fractions varies from 11.1% to 21.9%, which is consistent with previous results [31].

3.2. Near-Infrared Reflectance Spectra of Crushed and Husked Sunflower Seeds

Figure 3 shows the average reflectance spectra of the 285 crushed samples and the 41 husked samples. Husked samples absorb more than crushed seeds, as reported by Sato et al. [18].

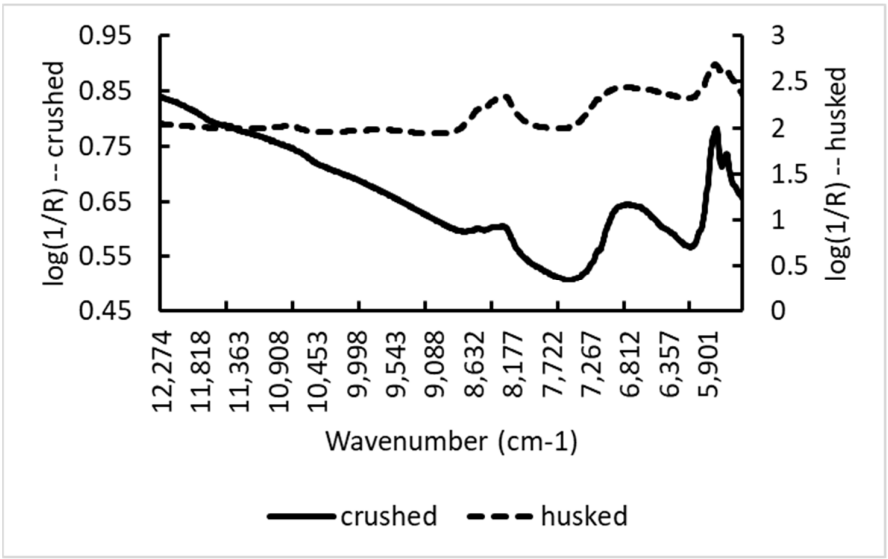

Figure 3. Average near-infrared reflectance spectra of crushed and husked sunflower seeds (285 crushed samples and 41 husked samples).

In our study, we identified characteristic wavelengths at 8237 cm^{-1}, 6930–6592 cm^{-1}, and 5780 cm^{-1}, corresponding to 1214 nm, 1443–1517 nm, and 1730 nm, respectively. These values highlight specific features in the electromagnetic spectrum. Comparing our findings with the work of Fassio and Cozzolino (2004) reveals similarities with absorption bands at 1460 nm (O-H stretch overtone bond—water) and 1720 and 1764 nm (C-H stretch first overtone bonds) [20].

3.3. Partial Least Squares Regression of Crude Protein Content of 285 Crushed Samples

The PLS prediction model for crude protein content was developed based on 80% of the samples (228 samples) and tested on the remaining 57 samples. The 228-sample training set had a mean crude protein content of 18.6% (10% minimum and 25.7% maximum, for a standard deviation of 3.2%). The mean crude protein content of the 57 samples of the test set was 19.1% (11.8% minimum and 23.3% maximum, for a standard deviation of 2.6%). Several PLS models were developed from the raw and preprocessed near-infrared reflectance spectra.

Table 3 and Figure 4 summarize the performance of the models for the 285 crushed seeds.

Table 3. Performance of best NIR models based on prediction set using partial least squares (PLS) regression for determining crude protein content % in crushed seeds for the various spectral preprocessing methods (standard normal variate (SNV), standard normal variate and detrending (SNVD), and first- and second-order Savitzky–Golay derivatives (D1 and D2)).

	No. of Samples for Training	No. of Factors	r^2 Calc	RMSEC (%)	RPD Calc	No. of Samples for Test	r^2 Pred.	RMSEP (%)	RPD Pred.
Raw	228	10	0.97	0.54	6.0	57	0.78	1.24	2.1
D1	228	7	0.92	0.93	3.5	57	0.71	1.43	1.8
D2	228	5	0.86	1.22	2.7	57	0.56	1.76	1.5
SNV	228	10	0.97	0.55	5.9	57	0.77	1.28	2.1
SNV D1	228	7	0.88	1.13	2.9	57	0.62	1.63	1.6
SNVD	228	9	0.97	0.54	6.0	57	0.74	1.35	2.0

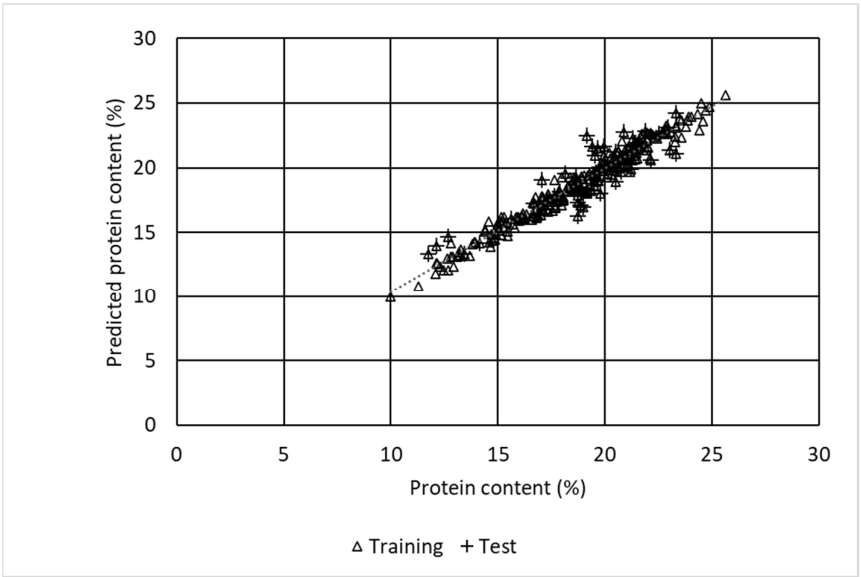

Figure 4. Scatter plot of best prediction model for crude protein content (%) of 285 samples of crushed sunflower seeds (228 samples for training and 57 samples for testing). Open triangles show the results of the training set, and the crosses show the results of the testing set.

The best preprocessing was taken to correspond to the highest calculated r^2 (which is associated with the lowest RMSEC). The best model was obtained with no preprocessing and gave $r^2 = 0.97$ and RMSEC = 0.54%, for a very high RPD of six. Blindly applying this model to the 57 testing samples produced $r^2 = 0.78$, RMSEP = 1.24%, and RPD = 2.1. In comparison, Fassio and Cozzolino's study on whole sunflower seeds yielded a calibration R^2 of 0.96 and SECV of 13.1 g kg^{-1}, while the prediction R^2 was 0.72 with an SEP of 17.1 g kg^{-1}. Our model was good in explaining the protein content within the training set; its predictive capability experienced a decline in the testing phase. Fassio and Cozzolino's approach maintained a reasonable predictive performance despite the inherent complexity of analyzing whole seeds. This comparison suggests that, although our model is promising, it still requires further improvement [28].

The advantages of quantification models using near-infrared (NIR) spectra on ground grains compared to whole grains can be attributed to several factors. Grinding the grains creates a more uniform surface, ensuring a homogeneous distribution of internal components, allowing the NIR spectrometer to obtain more representative measurements. Additionally, grinding facilitates better light penetration, eliminates surface-related interferences in whole grains such as dust or dirt, and ensures an increased reproducibility of samples. In summary, grinding enhances the quality of NIR measurements by eliminating surface variations, enabling better light penetration, reducing interferences, and ensuring optimal sample reproducibility. Let us now examine this performance by comparing it to analyses conducted on dehulled seeds.

3.4. Partial Least Squares Regression of Crude-Protein Content for 41 Husked Samples

The 41 husked samples were insufficient to be validated with an independent dataset, as done with the crushed samples. Thus, to avoid overfitting, a cross-validation was performed in this feasibility study (Table 4).

Table 4. Performance of best NIR models based on prediction set using partial least square (PLS) regression for determining concentration % of given element in husked seeds for the various spectral preprocessing methods (standard normal variate (SNV), standard normal variate and detrending (SNVD), and first- and second-order Savitzky–Golay derivatives (D1 and D2)).

	No. of Samples	**No. of Factors**	**r^2 Calc**	**RMSEC (%)**	**RPD Calc**	**r^2 CV**	**RMSECV (%)**	**RPD CV**
Raw	41	9	0.9	1.04	3.2	0.61	2.09	1.6
D1	41	7	0.8	1.46	2.3	0.43	2.57	1.3
D2	41	9	0.86	1.25	2.7	0.61	2.09	1.6
SNV	41	6	0.93	0.87	3.8	0.82	1.46	2.3
SNV D1	41	7	0.95	0.77	4.3	0.81	1.49	2.2
SNVD	41	5	0.93	0.86	3.9	0.83	1.39	2.4

The best model was obtained with SNVD (Table 4 and Figure 5). The model gave $r^2 = 0.93$ and RMSEC = 0.86%, with a good RPD value of 3.9. Cross-validating this model produced $r^2 = 0.83$, RMSECV = 1.39%, and RPD = 2.4. Cross-validation revealed a slight decrease in performance. In contrast, Wang et al. [17] presented a model trained on 154 samples and externally validated on 18 samples, achieving a high correlation coefficient ($r = 0.950$), a SEP of 1.399%, and an RPD of 2.9. Wang et al.'s model exhibited a comparable predictive performance despite a higher training dataset and an external validation. As is the case for the model developed on crushed grains, the SNVD model is promising but still needs improvement [28].

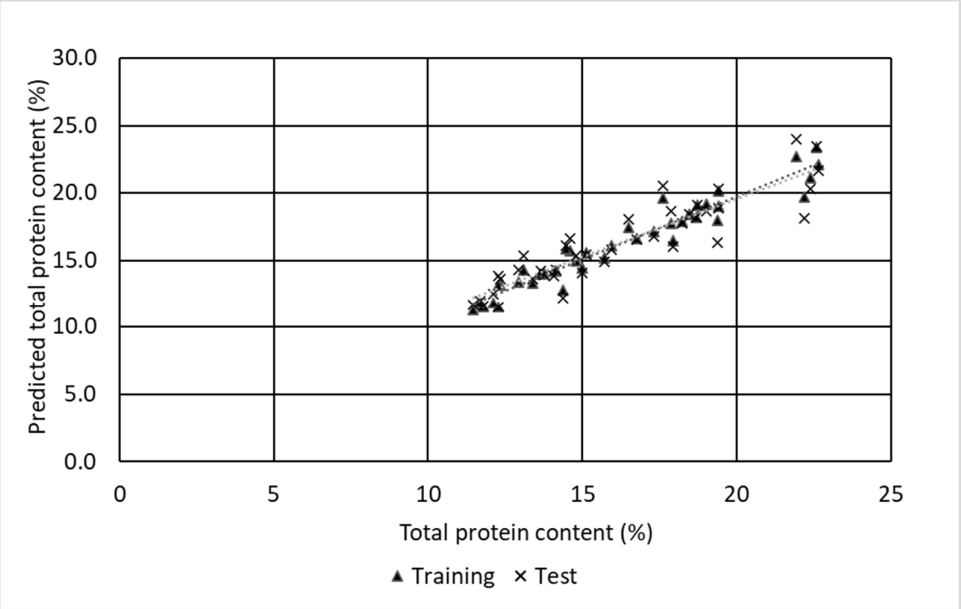

Figure 5. Scatter plot produced by the best prediction model for crude protein content (%) applied to 41 samples of husked sunflower seeds. Dark triangles show the training set, and crosses show the testing set.

The hull of the sunflower seed can pose an issue in infrared spectroscopy due to several factors. Firstly, the hull may be relatively opaque to infrared light, limiting its penetration into the sample. This can result in less accurate measurements as the light cannot effectively reach the internal components of the seed. Additionally, the hull can introduce interferences in the infrared spectra by reflecting or absorbing light non-uniformly. These interferences can distort the spectral data, complicating the interpretation and quantification of the internal components of the seed. Consequently, for more reliable results in infrared spectroscopy, it is often preferable to work with dehulled seeds to avoid the limitations imposed by the hull [17,20].

However, given that this model performs well upon training with five times fewer samples, more samples could be added to extend the feasibility study and increase the variability of the samples, thereby making the model more robust. In this way, NIRS could be used in breeding because husked seeds can be sown.

In addition to analyzing the total crude protein content, we also analyzed the albumin, globulin, and glutelin fractions.

3.5. Analysis of Albumin, Globulin, and Glutelin Content of 16 Crushed Samples

We applied an ascending hierarchical classification (AHC) to the protein fraction content of the 16 sunflower samples. The protein fraction contents were related to the sum of the three fractions. The AHC grouped the 16 sunflower samples into two groups comprising five and eleven sunflower samples (31% and 69%, respectively). These groups are labeled A and B, respectively. Samples from group A have a similar albumin content (n = 16, p-value = 0.083), significantly lower globulin and total protein content (n = 16, p-value = 0.023 and p-value = 0.0005, respectively), and higher glutenin content (n = 16, p-value = 0.027) than samples from group B (Mann–Whitney U test). Figure 6 represents a 3D plot of protein content in the 16 sunflower samples using albumin, globulin, and glutenin levels (% of total protein content—TPC).

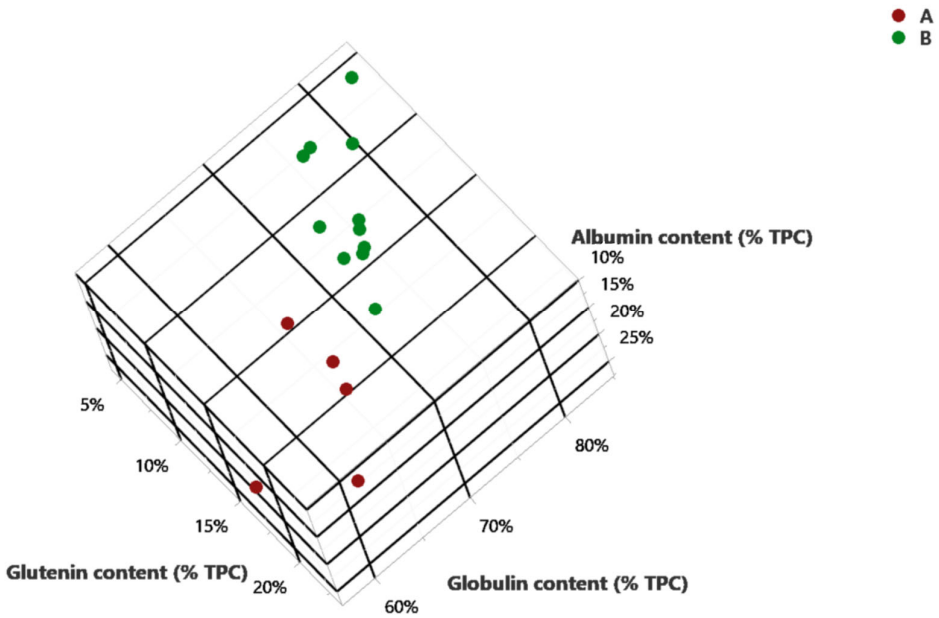

Figure 6. Analysis of protein content in 16 sunflower samples using albumin, globulin, and glutenin levels (% of total protein content—TPC): group A (in red) and group B (in green) differentiation.

Another Mann–Whitney U test was performed on each of the 2307 wavelengths of the near-infrared spectra to see if some wavelengths would be related to these protein fractions. The near-infrared spectra were considered raw and preprocessed D1, D2, SNV, SNV Detrend, and SNV D1. The corresponding wavenumbers and corresponding wavelengths are listed below: 11,788 cm^{-1} (848 nm), 9195 cm^{-1} (1087 nm), 6885 cm^{-1} (1452 nm), 6881 cm^{-1} (1453 nm), and 4216 cm^{-1} (2372 nm) for D1; 11,587 cm^{-1} (863 nm), 11,032 cm^{-1} (906 nm), 6746 cm^{-1} (1482 nm), and 5624 cm^{-1} (1778 nm) for D2; and 6881 cm^{-1} (1453 nm) for SVND1, with a *p*-value lower than 0.001. The following attribution of specific wavelengths can be made: 848 nm (C-H third overtone and C-C second overtone), 906 nm (C-H third overtone), 1087 nm (C-H second overtone and C-C second overtone), 1452–1453 nm (O-H first overtone), 1482 nm (N-H first overtone), 1778 nm (C-H first overtone), and 2372 nm (O-H second overtone) [35]. The attribution of specific wavelengths, ranging from C-H and C-C overtones to O-H and N-H overtones, highlights the complex nature of the molecules involved. The specific wavelengths identified in our study, particularly those strongly associated with protein fractions, align with the findings of Fassio and Cozzolino (2004), who observed strong absorption bands corresponding to O-H and C-H overtones [20]. The wavelengths identified as significant in this study are then consistent with those selected by various authors [36–38], which are considered relevant for protein analysis.

4. Conclusions

Firstly, this study achieves high accuracy in predicting the total protein content of crushed 326 samples (r^2c = 0.97, RMSEC 0.54%, RPDc 6; r^2p = 0.78, RMSEP 1.24%, RPDp 2.1). Grinding ground grains for near-infrared (NIR) spectroscopy improves measurement quality by creating a uniform surface, enhancing light penetration, reducing interferences, and ensuring optimal sample reproducibility.

Secondly, this study shows promising outcomes (n = 41, r^2c = 0.93, RMSEC 0.86%, RPDc 3.9; r^2cv = 0.83, RMSECV 1.39%, RPDcv 2.4) extended to husked samples. Sunflower seed hulls can hinder infrared spectroscopy due to their opacity, limiting light penetration and causing interferences in spectral data. Dehulled seeds are preferred for more reliable results. The most notable aspect of this study lies in the possibility of replanting sunflower seeds after the shelling process. This observation underscores the sustainable and practical

nature of the adopted approach and is of considerable importance as it broadens the application of near-infrared spectroscopy (NIRS) in crop selection.

This study reveals valuable insights, but it comes with certain limitations. Firstly, the sample size for husked samples is limited to 41, hindering robust statistical validation. Additionally, while the partial least squares (PLS) regression models show promising predictive capabilities for crude protein content in crushed seeds, the decline in performance when applied to the testing set indicates a need for further refinement. Furthermore, the reliance on crushed seeds for NIR analysis may not fully capture the complexities introduced by the presence of the hull in whole seeds.

Thirdly, the analysis categorized 16 crushed samples into two groups based on their compositions of protein fractions (albumin, globulin, and glutelin), using significant wavelengths attributed to proteins. This feasibility study presents exciting possibilities for large-scale screening categorizing content levels (low, medium, and high). These findings are essential for the food and animal feed sectors, where a detailed knowledge of protein fractions is crucial for adjusting formulations while limiting environmental impact. Despite promising results, improvements are needed and can be made by adding samples of various origins.

Finaly, the potential development of a field device could simplify the work of plant breeders, making this technology even more accessible and impactful in real-world green agricultural settings.

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

Funding: This work was financed by a grant from Sofiprotéol, France.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is unavailable due to privacy considerations.

Conflicts of Interest: Author P.C. is employed by Caussade Semences; Authors C.H. and C.F. are employed by MAS Seeds; Author M.L. is employed by RAGT 2n; Author V.M.-T. is employed by Syngenta France; Author S.P. is employed by SOLTIS. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. FAO Statistics Division. FAOSTAT-Production, Last Update 24 March 2023. Available online: http://www.fao.org/faostat/fr/#data/QC (accessed on 15 December 2023).
2. Giannini, V.; Maucieri, C.; Vamerali, T.; Zanin, G.; Schiavon, S.; Pettenella, D.M.; Bona, S.; Borin, M. Sunflower: From Cortuso's Description (1585) to Current Agronomy, Uses and Perspectives. *Agriculture* **2022**, *12*, 1978. [CrossRef]
3. Tkeshelashvili, M.; Bobozhonova, G.; Ananiev, M. The Use of High-Protein Flour from Sunflower "Bioprotein" for the Development of Vegetarian Food Products. *J. Food Chem. Nanotechnol.* **2022**, *8*, 21–25.
4. Povod, M.; Mykhalko, O.; Povoznikov, M.; Gutyj, B.; Koberniuk, V.; Shuplyk, V.; Ievstafiieva, Y.; Buchkovska, V. Efficiency of Using High-Protein Sunflower Meal Instead of Soybean Meal in Feeding of Growing Piglets. *Sci. Pap. Manag. Econ. Eng. Agric. Rural Dev.* **2022**, *22*, 595–602.
5. Sara, A.S.; Mathé, C.; Basselin, M.; Fournier, F.; Aymes, A.; Bianeis, M.; Galet, O.; Kapel, R. Optimization of sunflower albumin extraction from oleaginous meal and characterization of their structure and properties. *Food Hydrocoll.* **2020**, *99*, 105335. [CrossRef]
6. Yuldasheva, S.; Gofurova, O.; Askarova, G. Prospects of crop growing and significance. *Sci. Innov.* **2022**, *1*, 298–302.
7. Kumar, P.; Mehta, N.; Abubakar, A.A.; Verma, A.K.; Kaka, U.; Sharma, N.; Sazili, A.Q.; Pateiro, M.; Kumar, M.; Lorenzo, J.M. Potential alternatives of animal proteins for sustainability in the food sector. *Food Rev. Int.* **2023**, *39*, 5703–5728. [CrossRef]

8. Jayasena, A.S.; Franke, B.; Rosengren, J.; Mylne, J.S. A tripartite approach identifies the major sunflower seed albumins. *Theor. Appl. Genet.* **2016**, *129*, 613–629. [CrossRef]
9. Albe Slabi, S.; Mathe, C.; Basselin, M.; Framboisier, X.; Ndiaye, M.; Galet, O.; Kapel, R. Multi-objective optimization of solid/liquid extraction of total sunflower proteins from cold press meal. *Food Chem.* **2020**, *317*, 126423. [CrossRef]
10. Kaur, R.; Ghoshal, G. Sunflower protein isolates-composition, extraction and functional properties. *Adv. Colloid Interface Sci.* **2022**, *306*, 102725. [CrossRef]
11. Osborne, T.B. *The Vegetable Proteins*; Longmans, Green and Co.: London, UK, 1924.
12. Sosulski, F.; Bakal, A. Isolated proteins from rapeseed, flax and sunflower meals. *Can. Inst. Food Technol. J.* **1969**, *2*, 28–32. [CrossRef]
13. Bongartz, V.; Brandt, L.; Gehrmann, M.L.; Zimmermann, B.F.; Schulze-Kaysers, N.; Schieber, A. Evidence for the formation of benzacridine derivatives in alkaline-treated sunflower meal and model solutions. *Molecules* **2016**, *21*, 91. [CrossRef] [PubMed]
14. Fetzer, A.; Herfellner, T.; Stäbler, A.; Menner, M.; Eisner, P. Influence of process conditions during aqueous protein extraction upon yield from pre-pressed and cold-pressed rapeseed press cake. *Ind. Crops Prod.* **2018**, *112*, 236–246. [CrossRef]
15. Tanger, C.; Engel, J.; Kulozik, U. Influence of extraction conditions on the conformational alteration of pea protein extracted from pea flour. *Food Hydrocoll.* **2020**, *107*, 105949. [CrossRef]
16. Fontaine, J.; Hörr, J.; Schirmer, B. Near-infrared reflectance spectroscopy enables the fast and accurate prediction of the essential amino acid contents in soy, rapeseed meal, sunflower meal, peas, fishmeal, meat meal products, and poultry meal. *J. Agric. Food Chem.* **2001**, *49*, 57–66. [CrossRef] [PubMed]
17. Wang, L.; Tan, M.-L.; Fu, C.-L.; Wang, W.; Wang, L.-J.; Yin, Z.-Y.; Yan, X.-C. Prediction of qualitative characteristics of sunflower husked seed by near infrared spectroscopy. *Chin. J. Oil Crop Sci.* **2020**, *42*, 147.
18. Sato, T.; Takahata, Y.; Noda, T.; Yanagisawa, T.; Morishita, T.; Sakai, S. Nondestructive determination of fatty acid composition of husked sunflower (*Helianthus annua* L.) seeds by near-infrared spectroscopy. *J. Am. Oil Chem. Soc.* **1995**, *72*, 1177–1183. [CrossRef]
19. Moschner, C.R.; Biskupek-Korell, B. Estimating the content of free fatty acids in high-oleic sunflower seeds by near-infrared spectroscopy. *Eur. J. Lipid Sci. Technol.* **2006**, *108*, 606–613. [CrossRef]
20. Fassio, A.; Cozzolino, D. Non-destructive prediction of chemical composition in sunflower seeds by near infrared spectroscopy. *Ind. Crops Prod.* **2004**, *20*, 321–329. [CrossRef]
21. Akkaya, M.R. Prediction of fatty acid composition of sunflower seeds by near-infrared reflectance spectroscopy. *J. Food Sci. Technol.* **2018**, *55*, 2318–2325. [CrossRef]
22. González Martín, I.; Villaescusa García, V.; López González, F.; Oiz Jiménez, C.; Lobos Ortega, I.A.; Gordillo Arrobas, B.; Hernández Hierro, J.M. Control of quality and silo storage of sunflower seeds using near infrared technology. *Grasas Aceites* **2013**, *64*, 30–35.
23. Perez-Vich, B.; Velasco, L.; Fernandez-Martinez, J.M. Determination of seed oil content and fatty acid composition in sunflower through the analysis of intact seeds, husked seeds, meal and oil by near-infrared reflectance spectroscopy. *J. Am. Oil Chem. Soc.* **1998**, *75*, 547–555. [CrossRef]
24. Martens, H.; Naes, T. *Multivariate Calibration*; John Wiley & Sons Ltd.: Chichester, UK; Wiley: New York, NY, USA, 1989; p. 419.
25. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. *A User-Friendly Guide to Multivariate Calibration and Classification*; Nir Pubns: Chichester, UK, 2002.
26. Kennard, R.W.; Stone, L.A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11*, 137–148. [CrossRef]
27. Duthen, S.; Levasseur-Garcia, C.; Kleiber, D.; Violleau, F.; Vaca-Garcia, C.; Tsuchikawa, S.; Delgado Raynaud, C.; Daydé, J. Using near-infrared spectroscopy to determine moisture content, gel strength, and viscosity of gelatin. *Food Hydrocoll.* **2021**, *115*, 106627. [CrossRef]
28. Kapper, C. Rapid Prediction of Pork Quality: Correlation of Fresh Meat Measurements to Pork Water Holding Capacity and Its Technological Quality. Ph.D. Thesis, Wageningen University & Research, Wageningen, The Netherlands, 2012.
29. Agelet, L.E.; Hurburgh, C.R. A Tutorial on Near Infrared Spectroscopy and Its Calibration. *Crit. Rev. Anal. Chem.* **2010**, *40*, 246–260. [CrossRef]
30. Dorrell, D.G.; Vick, B.A. Properties and Processing of Oilseed Sunflower. In *Sunflower Technology and Production*; Schneiter, A.A., Ed.; Wiley: Hoboken, NJ, USA, 2015; pp. 709–745.
31. González-Pérez, S.; Vereijken, J.M. Sunflower proteins: Overview of their physicochemical, structural and functional properties. *J. Sci. Food Agric.* **2007**, *87*, 2173–2191. [CrossRef]
32. Nesterenko, A.; Alric, I.; Violleau, F.; Silvestre, F.; Durrieu, V. A new way of valorizing biomaterials: The use of sunflower protein for alpha-tocopherol microencapsulation. *Food Res. Int.* **2013**, *53*, 115–124. [CrossRef]
33. Žilić, S.; Barać, M.; Pešić, M.; Crevar, M.; Stanojević, S.; Nišavić, A.; Saratlić, G.; Tolimir, M. Characterization of sunflower seed and kernel proteins. *Helia* **2010**, *33*, 103–114. [CrossRef]
34. Bauchot, A.; Merrien, A. Teneur en protéines des graines de tournesol et état protéique foliaire: Revue bibliographique. *Inf. Tech.-CETIOM* **1988**, *102*, 18–28.
35. Osborne, B.; Fearn, T.; Hindle, P.H. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, 2nd ed.; Addison-Wesley Longman Ltd.: Harlow, UK, 1993; p. 227.
36. Cozzolino, D.; Chree, A.; Murray, I.; Scaife, J.R. The assessment of the chemical composition of fishmeal by near infrared reflectance spectroscopy. *Aquac. Nutr.* **2002**, *8*, 149–155. [CrossRef]

37. Daszykowski, M.; Wrobel, M.S.; Czarnik-Matusewicz, H.; Walczak, B. Near-infrared reflectance spectroscopy and multivariate calibration techniques applied to modelling the crude protein, fibre and fat content in rapeseed meal. *Analyst* **2008**, *133*, 1523–1531. [CrossRef]
38. Chen, L.; Yang, Z.; Han, L. A Review on the Use of Near-Infrared Spectroscopy for Analyzing Feed Protein Materials. *Appl. Spectrosc. Rev.* **2013**, *48*, 509–522. [CrossRef]