Measurement of Soy Contents in Ground Beef Using Near-Infrared Spectroscopy

Hongzhe Jiang [1], Hong Zhuang [2], Miryeong Sohn [3] and Wei Wang [1,*]

[1]   College of Engineering, China Agricultural University, Beijing 100083, China; jianghongzhe@cau.edu.cn
[2]   Quality & Safety Assessment Research Unit, U.S. National Poultry Research Center, USDA-ARS, 950 College Station Rd., Athens, GA 30605, USA; hong.zhuang@ars.usda.gov
[3]   Department of Chemistry, the University of Georgia, Athens, GA 30602, USA; miryeong.sohn@ars.usda.gov
*   Correspondence: playerwxw@cau.edu.cn; Tel.: +86-10-6273-7288

Abstract: Models for determining contents of soy products in ground beef were developed using near-infrared (NIR) spectroscopy. Samples were prepared by mixing four kinds of soybean protein products (Arconet, toasted soy grits, Profam and textured vegetable protein (TVP)) with ground beef (content from 0%–100%). NIR spectra of meat mixtures were measured with dispersive (400–2500 nm) and Fourier transform NIR (FT-NIR) spectrometers (1000–2500 nm). Partial least squares (PLS) regression with full leave-one-out cross-validation was used to build prediction models. The results based on dispersive NIR spectra revealed that the coefficient of determination for cross-validation (R_{cv}^2) ranged from 0.91 for toasted soy grits to 0.99 for Arconet. The results based on FT-NIR spectra exhibited the best prediction for toasted soy grits ($R_{cv}^2 = 0.99$) and $R_{cv}^2 > 0.98$ for the other three soy types. For identification of different types of soy products, support vector machine (SVM) classification was used and the total accuracy for dispersive NIR and FT-NIR was 95% and 83.33%, respectively. These results suggest that either dispersive NIR or FT-NIR spectroscopy could be used to predict the content and the discrimination of different soy products added in ground beef products. In application, FT-NIR spectroscopy methods would be recommended if time is a consideration in practice.

Keywords: PLS; SVM; quantification; classification; dispersive NIR; FT-NIR

1. Introduction

Meat is considered the top quality protein source, not only due to its nutritional characteristics but also to its appreciated taste [1]. In recent years, since the addition of non-meat proteins to meat products may cause health problems, this practice has been forbidden in many countries [2]. Thus, the need for detecting added soy proteins in meat products is obvious. However, in the meat industry, ground meat is a further-processed product used for manufacturing a wide range of marketable products such as sausages, hams, bologna, salami, etc. It is of great importance to ensure meat products are healthy from the root material of ground meat.

As a kind of vegetable protein, soy protein products, such as flour, grits, concentrates, isolates, and textures, are widely used in processed meat, poultry and seafood products [3]. Tracing its root, on the one hand, using soy proteins as meat extenders is due to the lower price to reduce the total cost. On the other hand, soy protein products could present different functional properties in meat products such as texture-forming, fat and water emulsification and gelation [4]. However, in fact, individuals, especially infants and young children that are allergic to the added non-meat proteins, can be affected greatly by the ingestion of minute amounts of the allergen [2]. Thereby, the addition to meat products, in some cases, is not allowed. The regulatory department in the United States has been

routinely requested to determine soy contents in further-processed meat products, such as ground meat, meat patties, frankfurters and chili with meat and beans [5].

Current analytical methods, such as microscopic methods, electrophoretic methods and chromatographic methods for determining soy proteins in meat products are time-consuming or invasive. Near-infrared spectroscopy (NIRS) is a sensitive, fast and non-destructive analytical technique with simplicity in sample preparation and has received attention for use in the quantitative and qualitative analysis of quality and adulteration in meat and meat products, such as beef [6], pork [7], poultry [8] and seafood [9]. As for the near-infrared spectroscopy technique, both the dispersive system and Fourier transform (FT) system showed potential for rapid analysis of meat quality [10,11]. The objective of this study was to investigate if the NIR spectroscopic method could be used in the qualitative and quantitative analysis of soy contents in ground beef. A comparison between dispersive NIR and FT-NIR spectral models was also performed to determine which kind of NIR spectra performs better in the prediction.

2. Materials and Methods

2.1. Sample Preparation

Raw ground beef materials (4000 g) were purchased from the local supermarkets in Athens, GA, USA, with label information of 87%–93% lean meat. Four different soy products (powder form) including Arcon T U-172 Soy Protein Concentrate (abbreviated as Arconet), toasted Soy Grits, Pro-Fam H200 Hydrolyzed soy protein (abbreviated as Profam), and textured vegetable protein (TVP U-218 Minced 180, abbreviated as TVP) were procured from ADM (Archer Daniels Midland Company, Chicago, IL, USA). The meat was reground using a grinder, weighted and mixed with pre-weighted soy product powder (no water added), providing a range of soy contents from 0% to 30% (including 0%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25% and 30%) and pure soy (100% content). Among the samples, for example, the 0% was a blend of 0 g soy and 50 g meat, while the 10% was mixed with 5 g soy and 45 g meat, and so on. The beef-soy mixtures were blended with a blender (30 s at high speed), then placed in sample bags and stored in a refrigerator (0 °C) overnight for using. A total of 192 sample cells from 64 beef-soy mixture samples (16 for each soy product) were prepared for this study.

2.2. Measuring Spectra

Samples were scanned by two NIR instruments simultaneously, a Fourier transform system (4 cm^{-1} resolution, model Vector22/N, Bruker Optics, Billerica, MA, USA) and a dispersive system (10 nm bandpass, model NIRSystems 6500, FOSS NIRSystems, Inc., Laurel, MD, USA). The dispersive data were collected over the range of 400 to 2498 nm using WinISI software. However, the FT data were collected over the range of 9999.51 to 3999.80 cm^{-1} (equivalent to 1000–2500 nm) using OMNIC software (v. 3.0.19, Bruker, Madison, AL, USA, 2009). Dispersive and FT spectral data were collected in reflectance mode and triplicate spectra were collected from separately packed cells on each sample and then averaged.

2.3. Data Processing and Chemometrics

All chemometrics analyses were conducted using Matlab software with partial least squares (PLS) and support vector machine (SVM) Toolbox (ver. 2012b, Mathworks, Inc., Natick, MA, USA, 2012). PLS regression with full leave-one-out cross-validation (LOOCV) was used for development of content prediction models of soy contents. With this method, samples were all used for each model training to gain the highest robustness. In order to determine and compare the proficiency of predictive models, root mean square error of calibration and cross-validation (RMSEC and RMSECV) as well as the coefficient of determination of calibration and cross-validation ($R_c{}^2$ and $R_{cv}{}^2$) were used. SVM classification method was also used for discriminating different soy products in ground beef.

3. Results and Discussion

3.1. Spectra Information

The relationship information between the raw NIR spectra and the mixed soy content classes is shown in Figure 1. The wavelength used here ranged from 400 to 2500 nm for dispersive NIR (Figure 1a) and 1000 to 2500 nm for FT-NIR (Figure 1b).
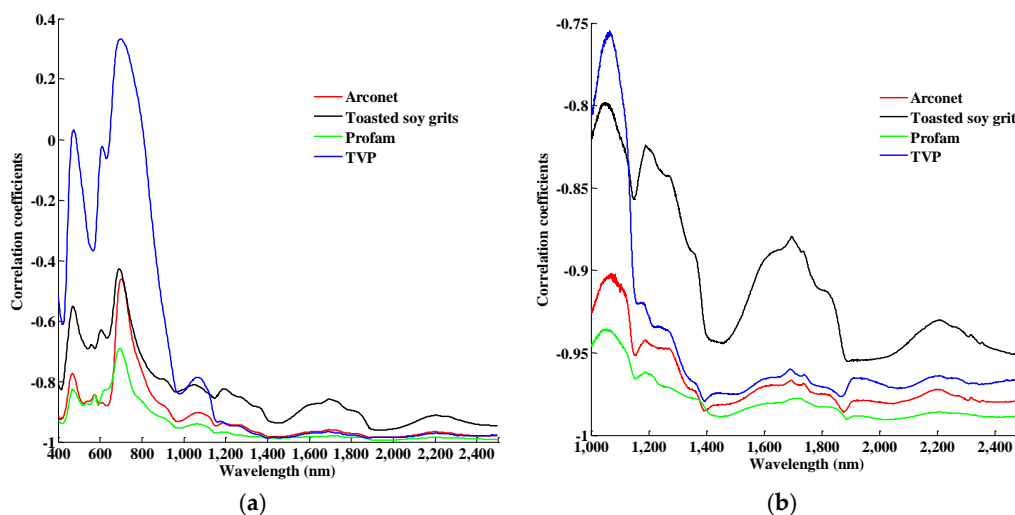

Figure 1. Correlogram between spectral data and mixed soy content classes of soy products in ground beef. (**a**) Dispersive spectral correlogram; (**b**) FT-NIR spectral correlogram, NIR: near-infrared spectroscopy; FT-NIR: Fourier transform NIR; TVP: textured vegetable protein.

As for Figure 1a, the curve patterns resembled each other for the four types of soy protein in the full waveband. However, in the section of the visible and part of the shortwave infrared wavelengths (400–1000 nm), the correlation coefficients showed a poor relationship compared to the rest of the wavebands regardless of soy type. This indicates that spectra between 400 and 1000 nm were less suitable for the prediction of soy contents by dispersive NIR methods. In the majority of the curves, the correlation was negative with a high absolute value (above 0.8) in the region of 1000–2500 nm, indicating that these spectra might provide a better prediction of soy contents. Overall, toasted soy grits showed a generally lower correlation than the other three soy protein types in this region, suggesting that the ability for determining contents quantitatively for toasted soy grits was slightly worse. In the case of Vis/NIR applications, the use of a set of optimal wavelengths that contains the most relevant information is desired [9]. Thus, in this study, a data set of 1000–2500 nm for dispersive NIR spectra was chosen for quantitative analysis. Interestingly enough, the selected spectra were conducive to the same range measured with the FT-NIR method. Figure 1b shows the correlogram between the spectra measured with the FT-NIR instrument and the contents of soy products in the mixtures. The spectral data revealed an obvious relationship with the mixed level by showing a high coefficient of correlation ($|r| > 0.75$). Among the four soy products, toasted soy grits again showed the lowest correlation and greatest fluctuation overall. This is similar to the results shown in Figure 1a. In the region of 1000 to 1200 nm, TVP also displayed a slightly poor relationship ($|r| < 0.8$) compared with the other types. A similar pattern was also noticed in the same region for dispersive NIR spectra (Figure 1a). Through comparison, it indicates that both NIR methods would provide similar predictions in general.

Figure 2 shows the average raw NIR spectra of ground beef and soy protein samples measured separately with the dispersive instrument (top) and the FT-NIR instrument (bottom). Due to the effect of the wavelength on the correlation coefficients (poor relationship at a wavelength less than 1000 nm) and for a better comparison between both spectral data sets (only 1000–2500 nm for FT-NIR spectra),

only the region of 1000 to 2500 nm of the dispersive data was used for the analysis. The difference between the two spectra results from the different operational principle of light splitting by diffraction grating and Michelson interference of the two instruments. The results showed that there was some degree of differences between the two spectra, such as the locations of the absorbance peaks and the absorbance extent at different wavelengths.
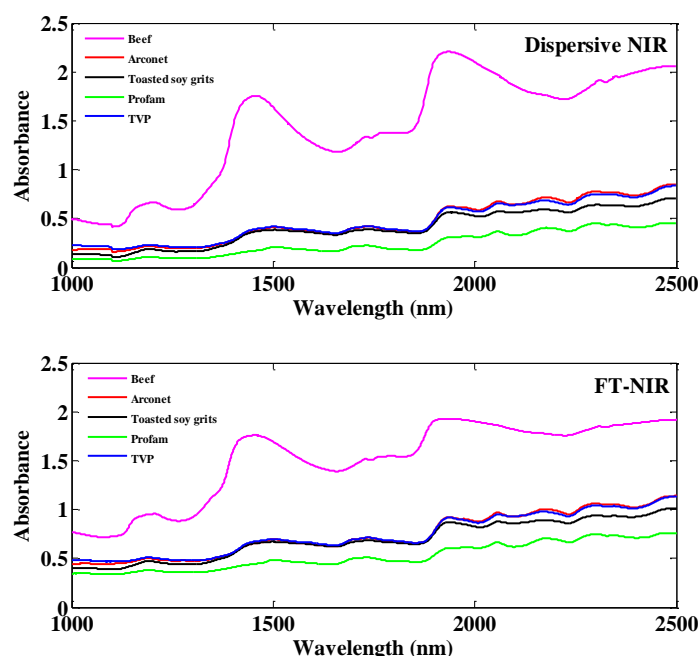
Figure 2. Dispersive NIR and FT-NIR spectra of ground beef (0%) and soy samples (100%) in the region of 1000–2500 nm.

3.2. Quantitative Analysis with PLS Regression

To quantify the soy protein contents in the mixtures, four models were developed corresponding to the four different soy products for each of two NIR spectra. The number of latent variables (LV) used in each model was chosen to optimize the model performance and minimize the model errors. The results of R_c^2, R_{cv}^2, RMSEC and RMSECV of different models developed by the dispersive and FT-NIR spectral data were compared in Tables 1 and 2. Overall, the RMSEC are all low (<0.021) and all the predicted results had a high correlation of the calibration data ($R_c^2 = 0.99$). As expected, PLS regression with dispersive spectral data for toasted soy grits had the lowest R_{cv}^2 value (0.91) and the highest RMSECV value (0.1164). The reason is that toasted soy grits mixtures showed the lowest correlation with the dispersive spectra (Figure 1a). Generally, the FT-NIR system gave better results for the four soy types than the dispersive system by showing a higher R_{cv}^2 (0.98–0.99) and lower RMSECV (0.013–0.057). Also, as Figure 1b indicated, the results of toasted soy grits and TVP showed slightly worse values than the other two soy types.

Table 1. Results of PLS regression based on dispersive spectra, PLS: partial least squares; RMSEC and RMSECV: root mean square error of calibration and cross-validation; TVP: textured vegetable protein; Toasted: toasted soy grits. LV: latent variables.

Soy Types	LV	R_c^2	RMSEC	R_{cv}^2	RMSECV
Arconet	6	0.99	0.0050	0.99	0.0135
Toasted	9	0.99	0.0035	0.91	0.1164
Profam	2	0.99	0.0204	0.98	0.0249
TVP	3	0.99	0.0130	0.99	0.0204

Table 2. Results of PLS regression based on FT-NIR spectral data, FT-NIR: Fourier transform near-infrared spectroscopy.

| Soy Types	LV	R_c^2	RMSEC	R_{cv}^2	RMSECV
Arconet	8	0.99	0.0047	0.99	0.0173
Toasted	10	0.99	0.0013	0.99	0.0228
Profam	9	0.99	0.0023	0.99	0.0131
TVP	6	0.99	0.0068	0.98	0.0570

In order to compare the actual soy contents and predicted soy contents by cross-validated models, Figure 3 reveals the quantitative ability of the two spectra visually. Figure 3a,b presents the correlation diagram between actual soy contents and NIR-predicted soy contents for different types of soy products using models developed by dispersive NIR and FT-NIR spectra, respectively. As shown, Figure 3b presents excellent prediction ability regardless of the soy product. Although the correlation coefficients for the two kinds of NIR models were both close to one, the FT-NIR spectra apparently worked better in the quantification of soy products content in ground beef.
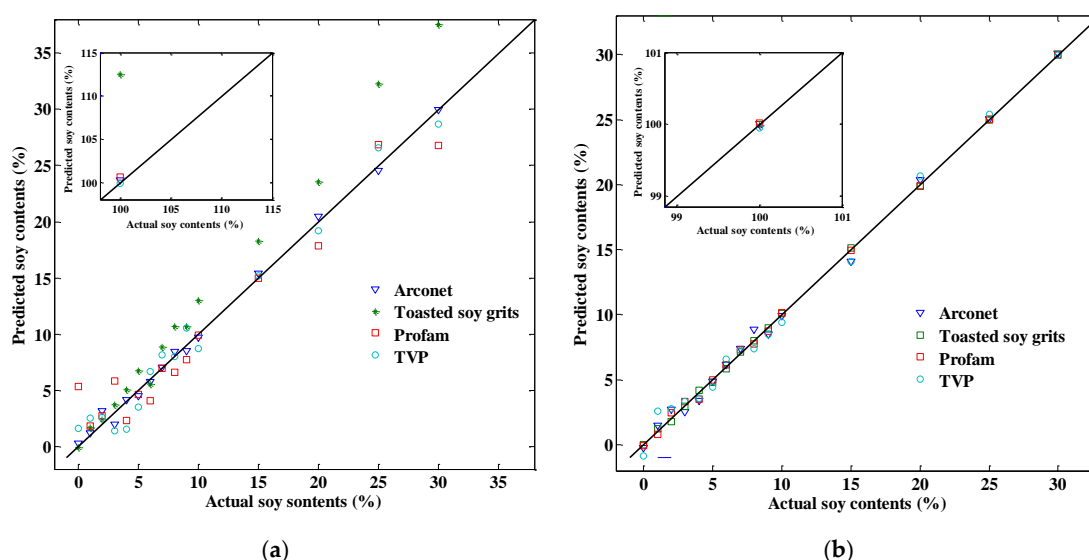
Figure 3. Predicted versus actual soy contents in ground beef using PLS regression model. (**a**) Model developed by dispersive NIR spectra; (**b**) Model developed by FT-NIR spectra, PLS: partial least squares.

3.3. Principal Component Analyses

Firstly, the standard normal variate (SNV) preprocessing method was performed to centralize the data matrix before principal component analysis (PCA) was conducted. Then, the high level class (consisting of samples with higher than 10% soy contents) and the low level class (the rest of the samples) were divided. A two-dimensional PCA score plot of all 64 samples (Figure 4) revealed that the samples could be segregated by high (in red) and low mixture content (in black). For dispersive NIR data (Figure 4a), the first two principal components could account for 98.90% of the data variance, among which PC_1 and PC_2 held 93.32% and 4.58% individually. There seemed to be a trend (from negative to positive on the PC_1 axis) describing classes of mixed soy content, with samples belonging to high soy content on the left and 0% on the right. In addition, Figure 4 shows that TVP and the Profam beef-soy mixture are located in the upper and middle region while Arconet and toasted soy grits are in the middle and lower region. On the basis of this observation, the first factor (PC_1) may contain information that describes the high and low soy content level in beef, whereas the second factor (PC_2) contains information differentiating soy product types. As for FT-NIR data (Figure 4b),

the PCA of this sample set also resulted in two PCs which accounted for 97.92% of the data variance, among which 92.92% was for PC_1 and 5.00% was for PC_2. The information contained in the two PCs was similar to the dispersive data, namely PC_1 was relevant for the soy mixture content in the mixtures and PC_2 was related to the soy type.
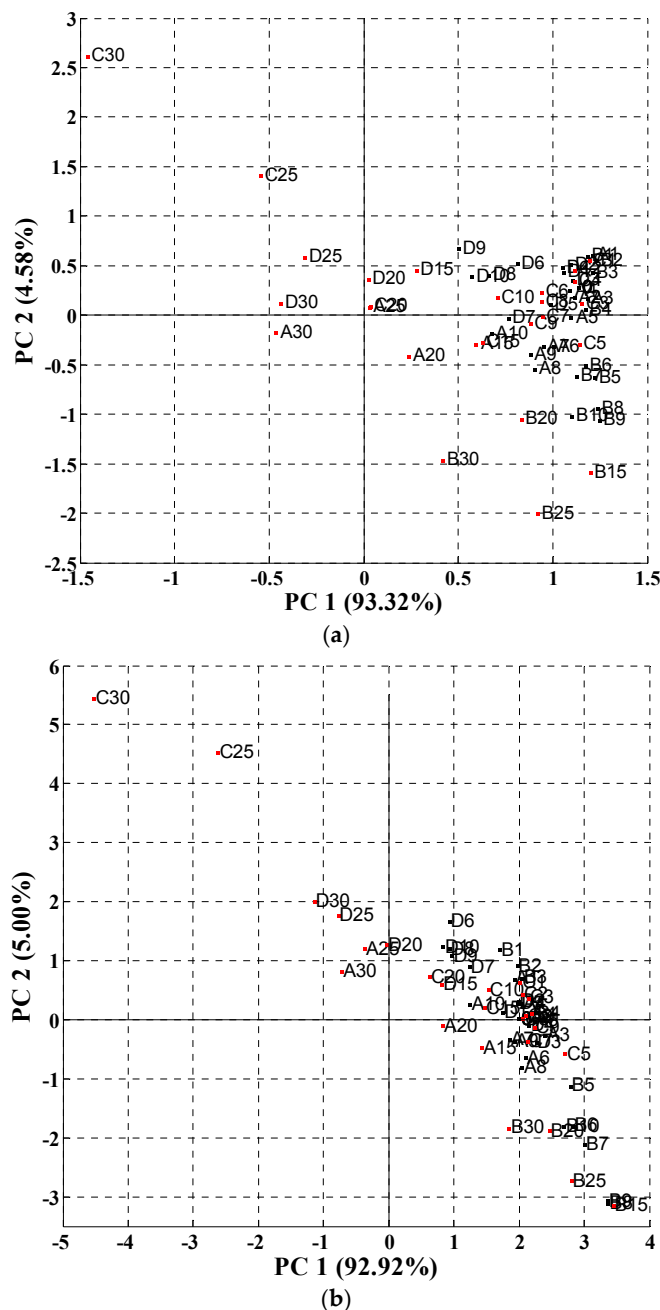
Figure 4. Score plots of the first two PCs of PCA. Samples are named according to the soy types (A = Arconet, B = toasted soy grits, C = Profam, D = TVP, 16 samples for each soy type) and soy contents (unit: %, from 0, i.e., pure beef, to 100, i.e., pure soy products). (**a**) PCA for dispersive spectral data; (**b**) PCA for FT-NIR data, PC: principal component; PCA: principal component analysis.

The loading charts were often used in the PCA analysis, and loading refers to a weighted importance of each variable for the particular principal component for demonstrating which variables were the most useful in the data to explain the data variance in the samples. Figure 5a shows the

loading chart of first two PCs based on dispersive NIR data while Figure 5b is the loading chart based on FT-NIR data.
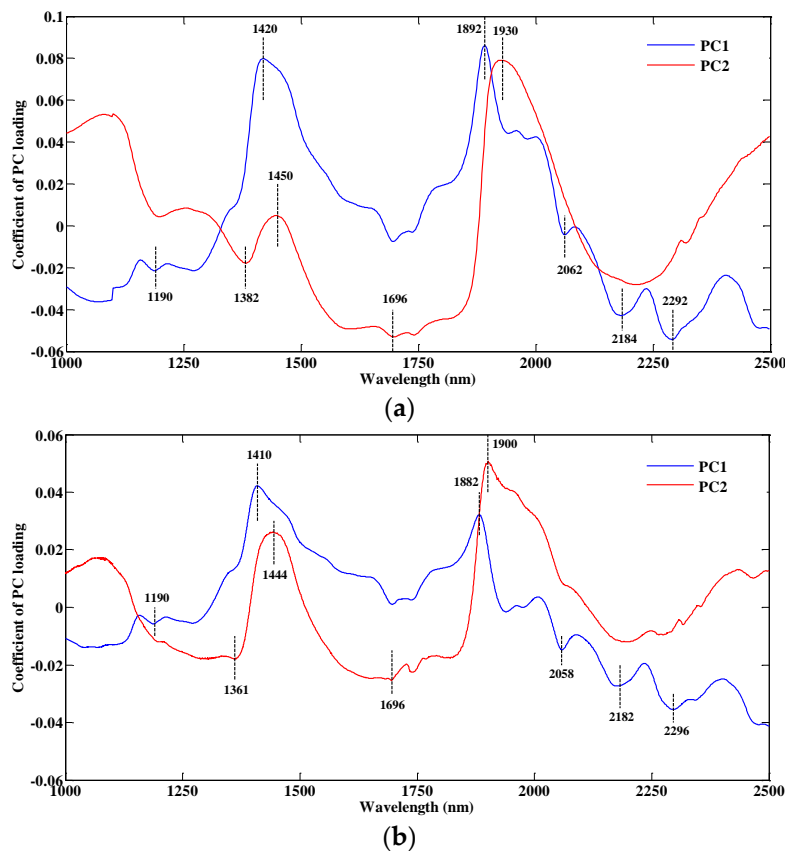
Figure 5. Characteristics of loading line plots based on PCA. (**a**) First three PC loading lines based on dispersive NIR data; (**b**) First two PC loading lines based on FT-NIR data.

As for the PC_1 loading line in Figure 5a, two key wavelengths are shown: 1420 nm, which is assigned to the O-H first overtone of hydrocarbon, and 1892 nm, which is attributed to the C=O stretch second overtone of fatty acids. The small peak at 1190 nm band likely corresponds to the overtones of the CH-stretching mode from lipid molecules [8]. In addition, some other significant peaks were also found at wavelengths such as 2062, 2184 and 2292 nm. The band assignments may all be relevant to the protein; 2062 nm corresponds to the N-H stretch combination of protein, 2184 nm is attributed to the N-H bend second overtone of protein and 2292 nm likely corresponds to the C-H bend second overtone of protein. In the PC_2 loading line, the peak of 1450 nm could correspond to the O-H stretch first overtone of water [8]. Moreover, the wavelength of 1930 nm is attributed to a stretching and bending combination for the band of water [12]. The wavelengths of 1382 and 1696 nm are assigned to the C-H combination of $-CH_2$ and the C-H stretch first overtone of $-CH_3$, respectively. Hence, the characteristic wavelengths selected in the PC_1 loading line were likely relevant to the lipid and protein which might help to analyze the mixed soy content in meat. In addition the wavelengths in the PC_2 loading line were mainly related to water and hydrocarbon in samples.

As a comparison, Figure 5b indicates the important wavelengths found in the PC loading lines based on the FT-NIR spectral data. The general trend of the PC_1 and PC_2 loading curves is similar to the PC loading lines in Figure 5a. In addition, the two peaks of 1410 and 1882 nm correspond to the C-H combination and the -OH stretch of water, respectively. As expected, the bands of 1190, 2058, 2182, and 2296 nm are also seen in the PC_1 loading line. The peak at 1361 nm could be assigned to the C-H combination of $-CH_3$, and 1900 nm corresponds to the C=O stretch second overtone of fatty acids.

The information of the other two wavelengths of 1444 and 1696 nm represented is consistent with the peaks of 1450 and 1696 nm in the PC_2 loading line in Figure 5a.

3.4. Qualitative Analysis of Soy Product Types

The possibility to differentiate soy protein types in ground beef was also evaluated by a classification approach using support vector machine (SVM). In SVM, the classification decision function is fixed based on the structural risk minimum mistake instead of the minimum of the misclassification on the training set to avoid an over-fitting problem [13]. In our work, SVM was selected because it is more suitable to grasp the nonlinear relationship between the data and features when the sample size is small. Samples of each type in various contents were all used in the calibration set. As for validation, eight samples for each soy type were randomly selected in the duplicate test to create a set of total 32 samples. The classified results from comparing the two models developed by dispersive and FT-NIR spectra are shown in Tables 3 and 4. For dispersive spectral data, the evaluated model offered a very high discriminating ability, showing a total accuracy of 95% and 84.38% for calibration and validation. In the calibration set, only toasted soy grits was wrongly discriminated at low contents, which might due to it having the lowest correlation with the dispersive NIR spectra among the four types. The classification model developed by FT-NIR spectra data gave a result of a total of 83.33% and 78.13% recognition accuracy for the calibration and validation set, respectively. However, the low soy content ($\leq$5%) classification performance was quite unsatisfactory, as there were incorrectly classified samples for Arconet, toasted soy grits, and Profam. Although better results were obtained by dispersive NIR data than by FT-NIR, low mixed soy protein contents ($\leq$5%) cannot be detected using either method.

Table 3. Classified results of four soy categories using SVM with dispersive NIRS models, Ar: Arconet; To: toasted soy grits; Pr: Profam; TV: TVP; SVM: support vector machine; NIRS: near-infrared spectroscopy.

Predicted vs. Actual	Calibration												Validation			
	$\leq$5% Content				5%–10% Content				10%–100% Content							
	Ar	To	Pr	TV	Ar	To	Pr	TV	Ar	To	Pr	TV	Ar	To	Pr	TV
Ar	5	0	0	0	5	0	0	0	5	0	0	0	7	0	1	0
To	1	2	2	0	0	5	0	0	0	5	0	0	1	5	2	0
Pr	0	0	5	0	0	0	5	0	0	0	5	0	0	0	7	1
TV	0	0	0	5	0	0	0	5	0	0	0	5	0	0	0	8
Accuracy						95%									84.38%	

Table 4. Classified results of four soy categories using SVM with FT-NIRS models.

Predicted vs. Actual	Calibration												Validation			
	$\leq$5% Content				5%–10% Content				10%–100% Content							
	Ar	To	Pr	TV	Ar	To	Pr	TV	Ar	To	Pr	TV	Ar	To	Pr	TV
Ar	0	0	4	1	5	0	0	0	5	0	0	0	6	0	2	0
To	0	1	3	1	0	5	0	0	0	5	0	0	2	4	2	0
Pr	0	0	4	1	0	0	5	0	0	0	5	0	0	0	7	1
TV	0	0	0	5	0	0	0	5	0	0	0	5	0	0	0	8
Accuracy						83.33%									78.13%	

4. Conclusions

This study is a preliminary investigation of using dispersive NIR and FT-NIR spectroscopy to predict the contents and types of dry soy protein products added to ground beef. The high R_{cv}^2 (>0.91) and low RMSECV (<0.12) of PLS regression indicated that both NIR spectroscopy methods are

able to quantify the soy content in ground beef. The SVM classification approach was developed to identify mixed soy categories and showed a total accuracy of 95% and 83.33% for models developed by dispersive NIR and FT-NIR spectra, respectively. However, when the soy content is less than 5%, neither method can precisely classify the products (<85%). The overall result indicates that it is feasible to use a spectrographic method to predict soy products in ground meat. However, it should be also emphasized that this work involved soy contents only beyond 1% and more development is needed for using NIR methods to predict less than 1% of soy contents in ground beef products.

References

1. Ballin, N.Z. Authentication of meat and meat products. *Meat Sci.* **2010**, *86*, 577–587. [CrossRef] [PubMed]
2. Belloque, J.; García, M.C.; Torre, M.; Marina, M.L. Analysis of soyabean proteins in meat products: A review. *Crit. Rev. Food. Sci.* **2002**, *42*, 507–532. [CrossRef] [PubMed]
3. Waggle, D.H.; Decker, C.D.; Kolar, C.W. Soya products in meat, poultry and seafood. *J. Am. Oil Chem. Soc.* **1981**, *58*, 341–343. [CrossRef]
4. Friedman, M.; Brandon, D.L. Nutritional and health benefits of soy proteins. *J. Agric. Food Chem.* **2001**, *49*, 1069–1086. [CrossRef] [PubMed]
5. Sohn, M.R.; Zhuang, H.; Lawrence, K.; De Haseth, J. NIR spectroscopy for determining soy contents in processed meat products. In Proceedings of the 14th International Conference on NIR Spectroscopy, Bankgkok, Thailand, 7–16 November 2009; pp. 7–12.
6. De Marchi, M. On-line prediction of beef quality traits using near infrared spectroscopy. *Meat Sci.* **2013**, *94*, 455–460. [CrossRef] [PubMed]
7. Hovingbolink, A.H.; Vedder, H.W.; Merks, J.W.M.; de Klein, W.J.H.; Reimert, H.G.M.; Frankhuizen, R.; van den Broek, W.H.A.M. Perspective of NIRS measurements early post mortem for prediction of pork quality. *Meat Sci.* **2005**, *69*, 417–423. [CrossRef] [PubMed]
8. Bowker, B.; Hawkins, S.; Zhuang, H. Measurement of water-holding capacity in raw and freeze-dried broiler breast meat with visible and near-infrared spectroscopy. *Poult. Sci.* **2014**, *93*, 1834–1841. [CrossRef] [PubMed]
9. Gayo, J.; Hale, S.A.; Blanchard, S.M. Quantitative analysis and detection of adulteration in crab meat using visible and near-infrared spectroscopy. *J. Agric. Food Chem.* **2006**, *54*, 1130–1136. [CrossRef] [PubMed]
10. Cozzolino, D.; Restaino, E.; La Manna, A.; Fernandez, E.; Fassio, A. Usefulness of near infrared reflectance (NIR) spectroscopy and chemometrics to discriminate between fishmeal, meat meal and soya meal samples. *Cienc. E Investig. Agrar.* **2009**, *36*, 209–214. [CrossRef]
11. Schmutzler, M.; Beganovic, A.; Böhler, G.; Huck, C.W. Methods for detection of pork adulteration in veal product based on FT-NIR spectroscopy for laboratory, industrial and on-site analysis. *Food Control* **2015**, *57*, 258–267. [CrossRef]
12. Osborne, B.G.; Fearn, T.; Hindle, P.H. *Practical near Infrared Spectroscopy with Applications in Food and Beverage Analysis*, 2nd ed.; Wiley: New York, NY, USA, 1993; p. 227.
13. Brereton, R.G.; Lloyd, G.R. Support vector machines for classification and regression. *Analyst* **2010**, *135*, 230–267. [CrossRef] [PubMed]