

# Comparison of Infrared Spectroscopy and Nuclear Magnetic Resonance Techniques in Tandem with Multivariable Selection for Rapid Determination of $\omega$ -3 Polyunsaturated Fatty Acids in Fish Oil

Di Wu · Xiaojing Chen · Fang Cao · Da-Wen Sun ·  
Yong He · Yanhui Jiang

Received: 29 March 2013 / Accepted: 17 June 2013 / Published online: 3 July 2013  
© Springer Science+Business Media New York 2013

**Abstract** This study investigated the potential of using four spectroscopic techniques including visible–short-wave near infrared, long-wave near infrared (LNIR), mid-infrared, and nuclear magnetic resonance (NMR) spectroscopy in tandem with multivariable selection and calibration for rapid determination of three important  $\omega$ -3 polyunsaturated fatty acids (PUFA), namely eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA), and docosapentaenoic acid (DPA) in fish oil. Quantitative models were established between the spectral data and reference PUFA contents of samples based on partial least squares regression (PLSR) algorithm. Successive projections algorithm (SPA) and uninformative variable elimination (UVE) were used to select the most important variables for prediction. The average decrements of 23.20 % for root mean square error

of cross-validation (RMSECV) and 64.90 % for the absolute value between root mean square error of calibration and RMSECV (AV\_RMSE) in all 12 cases achieved after over 90 % variables were eliminated. UVE was also found to be helpful to improve the efficiency of SPA's variable selection in 8/12 cases. The best predictions for EPA, DHA, and DPA were all achieved by NMR spectroscopy (determination coefficients of cross-validation ( $r_{CV}^2$ ) of 0.970, 0.982, and 0.983 and the RMSECV of 11.48, 4.73, and 0.77 mg/g for the EPA, DHA, and DPA predictions, respectively). LNIR spectra also did good predictions similar to NMR. The results demonstrated that the laborious and time-consuming gas chromatography method could be replaced by spectroscopic techniques in tandem with PLSR modeling and variable selection in order to provide a rapid and reliable inspection of PUFA in fish oil.

Di Wu and Xiaojing Chen contribute equally to this paper.

D. Wu · F. Cao · Y. He (✉)  
College of Biosystems Engineering and Food Science,  
Zhejiang University, Hangzhou 310058, China  
e-mail: yhe@zju.edu.cn

D. Wu · D.-W. Sun  
Food Refrigeration and Computerised Food Technology (FRCFT),  
School of Biosystems Engineering, University College Dublin,  
National University of Ireland, Agriculture & Food Science Centre,  
Belfield, Dublin 4, Ireland

X. Chen  
College of Physics and Electronic Engineering Information,  
Wenzhou University, Chashan University Town, Wenzhou,  
Zhejiang Province 325035, China

Y. Jiang (✉)  
School of Management, Hunan University,  
Changsha 410082, China  
e-mail: huihui0733@163.com

**Keywords** Fish oil · Polyunsaturated fatty acids · Nuclear magnetic resonance · Visible and near infrared spectroscopy · Mid-infrared spectroscopy · Variable selection

## Introduction

As derived from the fish tissues, fish oil contains a large amount of long-chain  $\omega$ -3 (n-3) polyunsaturated fatty acids (PUFA). It was reported that about 20 % of the total fatty acids in fish oil are PUFA (Cozzolino et al. 2005). PUFA have been proved to have protective properties against cancer (Berquin et al. 2007), cardiovascular disease (Wang et al. 2006), Parkinson's disease (Bousquet et al. 2008), cognitive decline (Morris et al. 2005), rheumatoid arthritis (Rennie et al. 2003), lupus (Wright et al. 2008), depression (Nemets et al. 2006), and psoriasis (Wolters 2005). PUFA also have beneficial

effects on the development of infant's nervous system (Jensen et al. 2010), visual system (Institute of Medicine 2005), and immune system (Calder et al. 2010). Due to low dietary intakes of PUFA in daily diet, fish oil has become one of the popular dietary supplements of PUFA for consumers. The market of fish oil for providing rich PUFA has rapidly expanded in recent years.

The quality of PUFA in fish oil should be strictly controlled. Currently, inaccurate listing of PUFA levels is a core quality problem for fish oil supplements. For some fish oil products, the amounts of PUFA shown on the packages do not match the actual contents in the fish oil. This mislabeling seriously infringe on the rights and interests of consumers. However, it is very difficult for consumers to tell the factual contents of PUFA in fish oil barely by using their naked eyes to observe the fish oil. Traditionally, fatty acids in foods and biological samples are customarily analyzed by gas chromatography (GC) of their methyl ester derivatives or directly as the ethyl esters (Curtis 2007). However, GC methods always take a very long time to analyze and need a lot of sample preparation. There are considerable managing steps involved in GC processes, which may cause oxidation of lipids, give variable results depending on the columns used, and pose difficulty in correctly identifying each fatty acid in the chromatograms (Igarashi et al. 2000). In addition, besides the inspection of PUFA in final products, it is needed for the determination of fatty acids in fish oil during the process in an industrial operation, preferably in real time to monitor the progress of the enrichment and to determine the optimal end point (Azizian et al. 2010); however, GC methods are not suitable for rapid determination of fish oil's quality. The lack of rapid tools for determining PUFA content in fish oil has been a main obstacle for the development of quality control for fish oil and promoting the development of better sensing techniques for such objectives.

Nowadays, some spectroscopic techniques, such as visible and near infrared (Vis–NIR) spectroscopy, mid-infrared (MIR) spectroscopy, and nuclear magnetic resonance (NMR) spectroscopy, have been successfully proved to be efficient and advanced tools for rapid and nondestructive determination of food quality. Visible spectra mainly contain the pigment information, and near infrared spectra arise from overtones and combination bands of C–H, O–H, N–H, and S–H stretching and bending vibrations. Mid-infrared spectra provide more information of frequencies and intensities, which are richer and stronger than near infrared spectra as bands in the mid-infrared spectra are fundamental bands of near infrared spectral bands (Wu et al. 2008). NMR spectroscopy measures the resonant frequencies of certain atomic nuclei present in different chemical surroundings, which provide physical and chemical properties of molecules where the atoms are contained. Different from the time-consuming GC measurement, spectroscopic techniques have the advantages of simplicity, rapid, and minimal sample preparation, so that

they agree with the green chemistry principle. PUFA have been rapidly determined using near infrared (Azizian et al. 2010), mid-infrared (Flåtten et al. 2005), Raman (Olsen et al. 2007), or NMR (Sacchi et al. 1993; Miyake et al. 1998; Tyl et al. 2008) spectroscopic techniques. Besides, endeavors of using near infrared spectroscopy (NIR) have been reported for the quality evaluation of fish oil, such as iodine value and saponification value (Endo et al. 2005), and moisture (Cozzolino et al. 2005). NMR spectroscopy also has many successful applications in the quality assessment of oil products (Guillén and Ruiz 2003; Masoum et al. 2007).

The main components of PUFA in fish oil include eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA), and docosapentaenoic acid (DPA). All of these acids are omega-3 fatty acids, in which the first double bond is located at the third carbon from the methyl end. Clinical investigations show that EPA can treat several psychiatric and neurodegenerative diseases due to its anti-inflammatory and neuroprotective effects (Song and Zhao 2007). Studies have also suggested that EPA may be of benefit in depression (Huan et al. 2004; Martins 2009). DHA is the predominant fatty acid of membrane phospholipids in the brain grey matter and in the retina of mammals (Guesnet and Alessandri 2011). Early DHA intervention has beneficial effects on cognition in age-related cognitive decline (Yurko-Mauro et al. 2010). In addition, DHA was found to have the ability of inhibiting growth of human colon cancer cells (Schønberg et al. 2006) and was shown to enhance the efficacy of chemotherapy in prostate cancer cells (Shaikh et al. 2008). On the other hand, available data suggest that DPA has beneficial health effects (Kaur et al. 2011). Therefore, it is our interest to investigate the feasibility of using spectroscopic techniques for determining specific contents of EPA, DHA, and DPA in fish oil. To the best of our knowledge, this is the first study on measuring these important PUFA contents in fish oil using visible–short-wave near infrared spectroscopy (Vis–SNIR), long-wave near infrared spectroscopy (LNIR), MIR, and NMR spectroscopy in tandem with multivariate selection and calibration and comparing the performances of these spectroscopic techniques.

The overall objective of this study was to compare Vis–SNIR, LNIR, MIR, and NMR spectroscopy in tandem with multivariate selection for the rapid and reliable assessment of EPA, DHA, and DPA in fish oil. The specific objectives of the current work were to (1) acquire spectral data of tested fish oil samples in Vis–SNIR (346–1,050 nm), LNIR (870–2,534 nm), MIR (4,000–400  $\text{cm}^{-1}$ ), and NMR (0–7.4 ppm) regions; (2) select important variables that were most correlated to the assessments of EPA, DHA, and DPA in fish oil by using successive projections algorithm (SPA) and uninformative variable elimination (UVE) algorithm; (3) establish multivariate calibration models using partial least squares regression (PLSR) based on the whole range spectra or spectra at the selected variables; (4) compare the

prediction abilities of four spectroscopic techniques Vis–SNIR, LNIR, MIR, and NMR spectra; and (5) evaluate the improvement ability of considering UVE before the SPA calculation for the variable selection.

## Materials and Methods

### Sample Preparation and Spectral Measurements

Fish oil from Amway Nutrilite®, By-Health®, and General Nutrition Centers (GNC)® were purchased from the local markets, resulting in 28 sets of fish oil samples obtained. Four samples were obtained from four batches with different production time for each set in the NMR analysis, and five samples were obtained from five batches with different production time for each set in other spectroscopic analysis. Therefore, there were 140 (5 samples per set×28 sample sets), 140 (5 samples per set×28 sample sets), 140 (5 samples per set×28 sample sets), and 112 (4 samples per set×28 sample sets) samples obtained for Vis–SNIR, LNIR, MIR, and NMR spectral measurements, respectively. A USB4000 miniature fiber optic spectrometer (Ocean Optics, Inc., USA) was used for the spectral scanning in the wavelength range of 346–1050 nm that covers the Vis–SNIR spectral wavelengths. The measurement of LNIR, which covers the wavelength range of 870–2,534 nm, was achieved by using an NIR 256–2.5 spectrometer (Ocean Optics, Inc., USA). Both spectrometers were equipped with a light source, SMA-terminated optical fibers, and a cuvette holder (DH2000, P400-2-VIS/NIR, and CUV-UV Holder for 1-cm cuvettes, respectively, Ocean Optics, Inc., USA). Optical fibers were used for connecting both source and detector to the cuvette holder. The quartz cell with a 1-cm path length was filled with oil sample and was inserted into the cuvette holder for the transmittance spectral measurement. An FTIR spectrometer Nicolet iS10 (Thermo Fisher Scientific Inc., Waltham, MA, USA) attached with a ZnSe attenuated total reflection accessory was used for the MIR acquisition in the wave number range of 4,000–400  $\text{cm}^{-1}$ . The NMR spectra of samples in the chemical shift from 0 to 7.4 ppm were collected by a Bruker Avance 500 MHz NMR spectrometer (Bruker, Karlsruhe, Germany) operating at 500.17 MHz for the proton nucleus at 298 K. A full spectrum of each sample was recorded with reference to tetramethylsilane. An exponential window function with a line-broadening factor of 1 Hz was applied to the free induction decay before the Fourier transformation. The  $^1\text{H}$ -NMR spectra were phased and baseline corrected using TopSpin 2.1 (Bruker, Karlsruhe, Germany) and were automatically reduced by using the AMIX (version 2.5, Bruker GmbH, Karlsruhe, Germany).

### GC Analysis

GC was used to detect the reference contents of EPA, DHA, and DPA in fish oil samples. The fish oils were saponified and the fatty acids were esterified (with a 30 % solution of  $\text{BF}_3$  in methanol) in accordance with the AOAC method (AOAC Method 963.22 2000). The methyl esters were analyzed in a gas chromatography (Agilent 6890 N; Agilent Technologies Inc., USA) with a flame ionization detector and an Agilent DB-1701 capillary column (30 m×0.32 mm×0.25- $\mu\text{m}$ ; Agilent Technologies Inc., USA). A 1- $\mu\text{L}$  aliquot of treated sample was injected with nitrogen (99.995 % purity) as a carrier gas at a flow rate of 3.0 mL/min. The column temperature was maintained at 180 °C for 2 min after injection then programmed at a heating rate of 10 °C/min to 240 °C (hold 10 min). The injector port and the detector temperatures were set at 250 °C. The flow rates of air and hydrogen were 450 and 45 mL/min, respectively. Identification of the fatty acids was carried out using the retention times relative to commercial methyl esters standards (EPA, DHA, DPA) (Sigma, USA). Calculations were based on previous analysis of standard mixtures and individual correction coefficients.

### Spectral Preprocessing

Spectral preprocessing is an integral part of chemometric analysis, which is usually performed prior to model calibration. The goal of the preprocessing is to reduce the effects of length variation of light path and light scattering and to enlarge the hidden information in the original spectral data. It should be noticed that spectral preprocessing is not obligatory and should be considered only when it can really improve the model's predicting ability. In this work, seven widely used preprocessing algorithms were considered, which were Savitzky–Golay smoothing (SGS), standard normal variate (SNV), multiplicative scatter correction (MSC), Savitzky–Golay first derivative (1st Der), Savitzky–Golay second derivative (2nd Der), SGS combined with 1st Der, and SGS combined with 2nd Der. Savitzky–Golay smoothing is an averaging algorithm that fits a polynomial equation to the data points to reduce noise from the spectral data without reducing the number of spectral variables. As a row-oriented transformation, SNV centers and scales each individual spectrum to remove the spectral scatter. MSC is designed to correct scatter effects in spectral data and also has functions for treating other similar effects, such as path length, offset shifts, and interference. Derivatives are accomplished to correct baseline effects in spectra and resolve nearby peaks. The prediction abilities of these preprocessing algorithms were compared based on the PLSR calibration with the full range of spectra.

## Multivariate Calibration

One of the advantages in using spectroscopic techniques is the wealth of information that reside in the spectra, which requests the use of intriguing chemometric tools. Such massive amounts of data contain many uninformative signals and need a strategy of multivariate chemometrics to calibrate a quantitative/qualitative model by extracting meaningful information efficiently. In this study, such strategy was implemented by using PLSR algorithm to extract meaningful information from the massive amounts of spectral data, which were then used to establish a quantitative relationship with the concentrations of EPA, DHA, and DPA in corresponding fish oil. Before the calculation of PLSR, the spectrum of each sample was arranged together to form a matrix consisting of a spectral matrix ( $X$ ) as response variables and a column vector ( $Y$ ) as the dependent variable. PLSR was then aimed to establish a quantitative relationship between the spectral matrix ( $X$ ) and the column vector ( $Y$ ). The column vector in this work was the reference values of one of the PUFAs (EPA, DHA, or DPA) obtained from the traditional GC measurement.

PLSR has been widely used as a classical multivariate calibration technique for spectral analysis with predictive purpose. Known as a bilinear factor method, PLSR attempts to find multidimensional direction in the spectral matrix ( $X$ ) that explains the maximum multidimensional variance direction in the column vector ( $Y$ ). In PLSR calculation, both the spectra (response variables) and concentration (dependent variables) information are decomposed simultaneously, ending in a set of orthogonal factors (latent variables, LVs) projected. Different from principal component regression, dependent variables are actively considered in the decomposition process of PLSR, which ensures that the first few LVs are most related to predict dependent variables. Typically, the optimal LV set should be used in the models to obtain efficient and robust models and avoid overfitting. Predicted residual error sum of squares (PRESS) is commonly used as a standard to determine the optimal LV set, in which the PRESS values of the first few LVs are calculated and the LVs with the lowest value of PRESS are defined as the optimal LV set. The PRESS is determined using the following equation in full cross-validation:

$$\text{PRESS} = \sum (Y_{\text{Val}} - Y_{\text{Ref}})^2 \quad (1)$$

where  $Y_{\text{Val}}$  is the calculated values of EPA, DHA, or DPA in fish oil samples estimated by full cross-validation, and  $Y_{\text{Ref}}$  are the reference values of samples.

## Variable Elimination

Variable selection is a critical step in the spectra analysis (Balabin and Safieva 2011; Balabin and Smirnov 2011).

Because of possessing hundreds or thousands of spectral variables, spectral data are characterized as high dimensionality, which can lead to convergence instability. On the other hand, congruent spectral variables often exhibit similar spectral information, which is a well-known problem of multicollinearity. The correlated variables are considered as redundancy. In addition, as useless or irrelevant information for the model calibration might be contained in some spectral variables, these variables should be eliminated to avoid worsening the predictive ability of the model. Therefore, the selection of informative variables is preferable for reducing the spectral high dimensionality and multicollinearity, thus simplifying the structure of the calibration model and improving the model's predictive capability. In the variable reduction scheme, only those variables that are informative for the model calibration and have less multicollinearity are selected, and redundant, uninformative, and multicollinear variables are not considered in the model calibration and thus eliminated.

In this paper, two novel variable selection algorithms were investigated, namely SPA and UVE. SPA is a variable selection algorithm proposed by Araujo et al. (2001). It was designed to solve the collinearity problems of spectral variables by selecting variables with minimal redundancy. There are two steps in SPA. The first step consists of a sequence of projection operations carried out in the columns of the spectral matrix, which constructs candidate subsets of variables. The second step consists of evaluating candidate subsets of variables according to the prediction performance of the calibrated model. Details of SPA description are shown in the literature (Wu et al. 2012). Although SPA could select a subset of variables with a minimum of collinearity, some of these variables might be uninformative and could worsen the model's performance in some cases. On the other hand, UVE is a PLSR-based variable selection algorithm proposed by Centner et al. (1996), which tries to eliminate the variables that have no more information for modeling than noise. In UVE process, the stability of each variable is calculated by dividing the mean of the PLSR regression coefficients by the standard deviation of the regression coefficients of the variable. With the help of using an artificial random variable matrix as a reference, these spectral variables that are less important in the model than the random variables are eliminated. However, there would be still dozens or hundreds of variables selected by UVE and these variables might have a problem of multicollinearity. In order to complement advantages of both methods and overcome their shortcomings, it is suggested to conduct SPA on informative variables selected by UVE, which is called UVE-SPA variable selection strategy. It is dedicated that the variables selected in this work were wavelength (Vis-SNIR and LNR spectroscopy), wave number (MIR spectroscopy), or chemical shift (NMR spectroscopy) variables, rather than latent variables in PLSR models.



## Model Evaluation

In the spectral analysis for quantitative regression, the calibrated spectral model must be evaluated for its validity by the validation process. According to some other works (Marini et al. 2003; Cozzolino et al. 2007), cross-validation is preferred as the validation method when the number of samples is limited (Martens and Dardenne 1998). Many works working on spectral analysis used cross-validation instead of splitting samples into a calibration set and a prediction set (ElMasry et al. 2011; Papadopoulou et al. 2011; Rivero et al. 2012; Kimiya et al. 2013), which show that the cross-validation is a commonly used and accepted validation method. Since the samples from the same sample set that had similar properties should not be validated in the leave-one-out way, segmented cross-validation was applied instead of leave-one-out cross-validation for the validation purpose. In segmented cross-validation, one segment is kept out of the calibration at a time. By choosing different segments, the predictions can be made on all segments for the validation procedure. Although leave-one-out cross-validation could be too optimistic sometimes, the cross-validation with several samples kept out for validation at a time was proved to be able to give a more realistic idea of the predictive power of the model than leave-one-out cross-validation (Sileoni et al. 2013). In this study, there were 28 segments generated to represent the 28 sample sets of fish oil for the segment cross-validation in the PLSR process. After the PLSR models were calibrated, their accuracy and predictive ability were evaluated in terms of determination coefficient of calibration ( $r_c^2$ ) and root mean square error of calibration (RMSEC) for the calibration procedure and determination coefficient of cross-validation ( $r_{cv}^2$ ), root mean square error of cross-validation (RMSECV), and residual predictive deviation (RPD) for the validation procedure. Generally, a good model should have high determination coefficients ( $r_c^2$ ) and ( $r_{cv}^2$ ) and RPD as well as low root mean square error (RMSEC and RMSECV). Besides, a robust model with less overfitting should have a small absolute value between RMSEC and RMSECV (AV\_RMSE). The achievements of preprocessing algorithms and multivariate calibration were conducted using Unscrambler V9.7 (CAMO Process AS, Oslo, Norway), while the calculation of SPA and UVE was performed based on MATLAB 7.12.0 (MathWorks, Natick, USA).

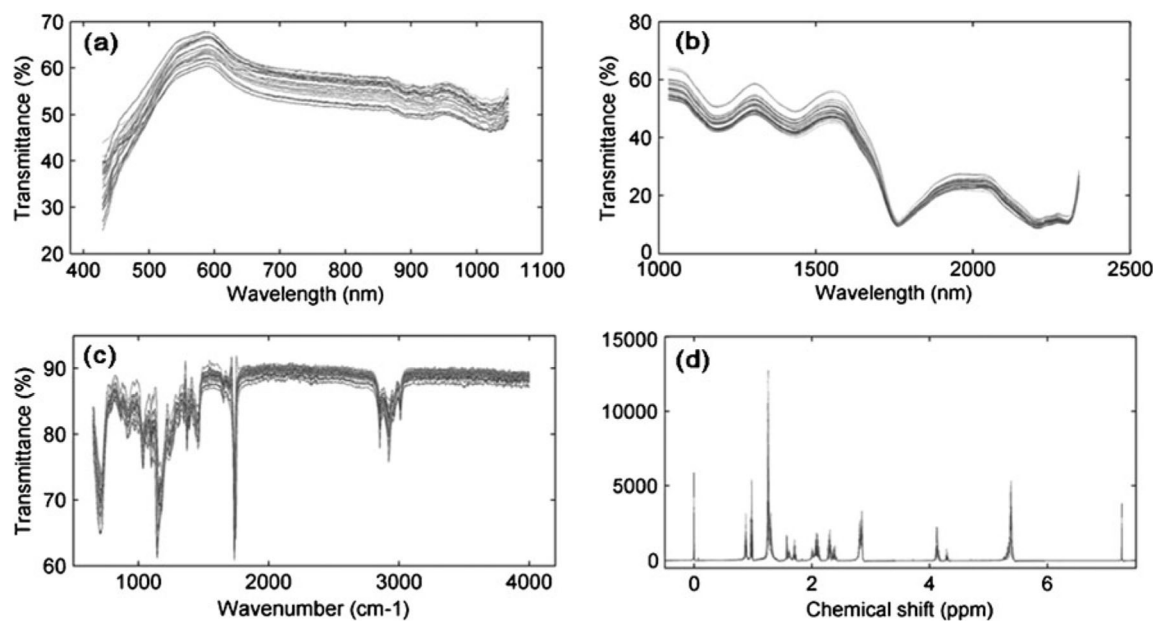
## Results and Discussion

### Spectra of Fish oil

Typical spectral profiles of the tested fish oil samples in Vis–SNIR, LNIR, MIR, and NMR ranges are shown in Fig. 1. Generally, the shapes of the original spectra were quite homogeneous for different samples in all four sub-images.

Due to spectra's overtone and combination absorptions of molecules, the spectral transmittance of fish oil shows quite even profiles in visible and near infrared spectral regions (Fig. 1a, b) with only some broadband peaks. In the visible spectral region (429–700 nm, Fig. 1a), the spectral pattern had high transmittance in red regions (about 550–650 nm) and low transmittance in blue region (about 430–500 nm), which explained why fish oil gets its golden color. There were some transmittance peaks over the entire near infrared spectral region (700–2337 nm, Fig. 1a, b). According to the literatures of Osborne and Fearn (1986), two weak peaks found at around 900 and 1020 nm in Fig. 1a were assigned, respectively, to C–H stretching third overtone of CH<sub>3</sub> and a combination of C–H stretching first overtone and C–H deformation second overtone of CH<sub>3</sub>. Medium peak between 1,100 and 1,300 nm was due to C–H stretching second overtone of CH, CH<sub>2</sub>, CH<sub>3</sub>, and HC=CH. Another medium peak between 1,300 and 1,550 nm was caused by a combination of C–H stretching first overtone and C–H deformation first overtone in CH, CH<sub>2</sub>, and CH<sub>3</sub>. The high transmittance peak at 1,760 nm was assigned to C–H stretching first overtone of CH<sub>2</sub> and CH<sub>3</sub>. There were two high transmittance peaks after 2,000 nm. In specific, the one at around 2,200 nm was caused by a combination of CH<sub>2</sub> stretching and C=C stretching of HC=CH, and the one at around 2,300 nm was due to a combination of C–H stretching and C–H deformation in CH<sub>3</sub> and CH<sub>2</sub>.

In the MIR region as shown in Fig. 1c, the transmittance peak at around 700 cm<sup>-1</sup> was assigned to the rocking and bending out vibration of  $-(CH_2)_n-$  and  $-HC=CH-$  (*cis*). Two medium peaks were found at around 1,100 and 1,030 cm<sup>-1</sup>, which were due to  $-C-O$  stretching. The one at around 1,150 cm<sup>-1</sup> was assigned to stretching and bending vibration of  $-C-O$  and  $-CH_2$ , and  $-C-H$  (CH<sub>3</sub>) bending (sym) vibration was found at around 1,370 cm<sup>-1</sup>. The transmittance peaks at around 1,460, 1,650, and 1,740 cm<sup>-1</sup> were assigned to  $-C-H$  (CH<sub>2</sub>, CH<sub>3</sub>) bending (scissoring),  $-C=C-$  (*cis*) stretching, and  $-C=O$  (ester) stretching vibrations, respectively. There were also some peaks between 2,800 and 3,100 cm<sup>-1</sup>, in which the one at around 3,010 cm<sup>-1</sup> was assigned to  $=C-H$  (*cis*) stretching, the one at around 2,920 cm<sup>-1</sup> was assigned to  $-C-H$  (CH<sub>2</sub>) asym stretching, and the one at around 2,855 cm<sup>-1</sup> was assigned to  $-C-H$  (CH<sub>2</sub>) sym stretching. In the NMR spectra as shown in Fig. 1d, the peak between 5.20 and 5.50 ppm was assigned to  $-CHOCOR$  of glyceryl group and  $-HC=CH-$  of acyl group; the peak between 4.10 and 4.30 ppm was assigned to  $-CH_2OCOR$  of glyceryl group; the peak between 2.75 and 2.90 ppm was assigned to  $=HC-CH_2-CH=$  of acyl group; the peak between 2.25 and 2.40 ppm was assigned to  $-OCO-CH_2-$  of acyl group; the peak between 1.95 and 2.15 ppm was assigned to  $-CH_2-CH=CH-$  of acyl group (except for  $-CH_2-$  of DHA acyl group in  $\beta$  position); the peak between



**Fig. 1** Typical spectral profiles of the tested fish oil samples. **a** Visible and short-wave near infrared spectra. **b** Long-wave near infrared spectra. **c** Mid-infrared spectra. **d** NMR spectra

1.67 and 1.74 ppm was assigned to  $-\text{OCO}-\text{CH}_2-\text{CH}_2-$  of EPA acyl group; the peak between 1.55 and 1.65 ppm was assigned to  $-\text{OCO}-\text{CH}_2-\text{CH}_2-$  of acyl group (except for DHA and EPA acyl groups); the peak between 1.20 and 1.40 ppm was assigned to  $-(\text{CH}_2)_n-$  of acyl group; the peak between 0.94 and 1.00 ppm was assigned to  $-\text{CH}_3$  of polyunsaturated  $\omega$ -3 acyl group; and the peak between 0.86 and 0.93 ppm was assigned to  $-\text{CH}_3$  of saturated, monounsaturated  $\omega$ -9 and  $\omega$ -7, and diunsaturated  $\omega$ -6 acyl groups. The assignments of wave number for MIR and chemical shift for NMR were carried out according to the literature of Guillen et al. (2008).

As a conclusion from analyzing Fig. 1, many chemical molecules and functional groups of constituents in fish oil had their spectral information that was important to determine PUFA contents of fish oil in a quantitative manner. However, when the spectra of more samples were considered, their profiles had a similar pattern. It is difficult to directly observe spectra for the PUFA determination when more samples were considered, as their spectral curves would be overlapped and there were no feature peak-related PUFA contents directly. Thus, chemometrics were employed for the data mining and analysis purpose, resulting in calibrated quantitative models. In addition, because the noise-to-signal ratios of the spectral detectors in the spectroscopic systems were rather low in some parts of spectral data, the spectral matrix with only the spectral range of 429–1,048 nm for Vis–SNIR spectral matrix (3,250 variables  $\times$  140 samples), the range of 1,031–2,337 nm for LNIR spectral matrix (201 variables  $\times$  140 samples), and the ranges of 650–1,494, 1,687–1,782, and 2,820–3,036  $\text{cm}^{-1}$

for MIR spectral matrix (2,400 variables  $\times$  140 samples) were used for further developing the calibration models. For NMR spectra, the chemical shifts with values less than 40 were eliminated as there was no useful information but only noise, resulting in the NMR spectral matrix with the size of 3,636 variables  $\times$  112 samples. In addition, the reference values of PUFAs obtained using GC analysis are shown in Table 1. The correlation coefficients of EPA and DHA, EPA and DPA, and DHA and DPA were 0.66, 0.71, and 0.83.

#### EPA Analysis

Calibration of multivariate models was first executed using the full range spectra of samples. A classic linear calibration method named PLSR was applied for the purpose of model calibration based on the spectral information of samples and their corresponding reference PUFA contents determined by GC method. After the model was established, the segment cross-validation was carried out in order to estimate the predictive capability of the established model and to evaluate the model's overfitting level.

#### Calibration Based on Full Range Spectra and Analysis of Spectral Preprocessing

At the calibration stage of using full range spectra, besides the use of original spectral data, the preprocessed spectra of different preprocessing algorithms (SGS, SNV, MSC, 1st Der, 2nd Der, SGS-1st Der, and SGS-2nd Der) were also considered in order to determine the best input spectral data for the PLSR calibration and variable selection. When Vis–

**Table 1** Statistics of eicosapentaenoic acid, docosahexaenoic acid, and docosapentaenoic acid concentration (in milligrams per gram) of the fish oil samples

PUFA	Maximum	Minimum	Mean	Standard deviation	Range
EPA	353.22	104.87	214.29	64.44	248.35
DHA	223.33	100.25	173.89	33.79	123.08
DPA	32.6	10.95	21.57	5.71	21.65

*PUFA* polyunsaturated fatty acids, *EPA* eicosapentaenoic acid, *DHA* docosahexaenoic acid, *DPA* docosapentaenoic acid

SNIR spectra were considered, the PLSR model established using original spectral data had a good result with  $r_{CV}^2$  of 0.852 and RMSECV of 25.69 mg/g. When the seven spectral preprocessing algorithms (SGS, SNV, MSC, 1st Der, and 2nd Der) were considered respectively, the best performance was obtained by the MSC-PLSR model with  $r_{CV}^2$  of 0.884 and RMSECV of 22.79 mg/g. Since there was no much improvement by considering the preprocessing, the original Vis–SNIR spectra were used in further calculation. When LNIR spectra were considered, although the original spectra only obtained the PLSR model with  $r_{CV}^2$  of 0.773 and RMSECV of 31.83 mg/g, the performances of seven PLSR models with spectral preprocessing were all better than that of the PLSR model with the original spectra. The best prediction result with  $r_{CV}^2$  of 0.920 and RMSECV of 18.89 mg/g

was obtained when 2nd Der was applied. For both the MIR and NMR spectra, the best results (shown in Table 2) were achieved when the original spectra were used as the inputs of PLSR model. The prediction became worse when the spectral preprocessing was considered for both MIR and NMR spectra.

#### *SPA Calculation Based on the Full Range Spectra*

The selection of the most important variables is critical to analyze the spectral data, which leads to reducing the high dimensionality and multicollinearity of spectra and improving the model's predictive ability in terms of accuracy and robustness. In the “Calibration Based on Full Range Spectra and Analysis of Spectral Preprocessing” section, the full

**Table 2** Predictive results of eicosapentaenoic acid (EPA) content in fish oil using infrared spectroscopy and nuclear magnetic resonance with chemometrics (available range of EPA is between 104.87 and 353.22 mg/g)

	Sample number	Variable selection method	Variable number	Number of latent variables	Calibration		Prediction		
					$r_C^2$	RMSEC (mg/g)	$r_{CV}^2$	RMSECV (mg/g)	RPD
Vis–SNIR	140	Full range	3,250	9	0.986	7.54	0.852	25.69	2.508
Vis–SNIR	140	SPA	8	7	0.912	19.16	0.865	24.53	2.627
Vis–SNIR	140	UVE	1,509	8	0.986	7.67	0.942	16.07	4.010
Vis–SNIR	140	UVE-SPA	12	8	0.957	13.30	0.891	22.10	2.916
LNIR	140	Full range	201	9	0.991	6.06	0.920	18.89	3.411
LNIR	140	SPA	16	10	0.982	8.66	0.932	17.43	3.697
LNIR	140	UVE	123	8	0.992	5.74	0.943	15.96	4.038
LNIR	140	UVE-SPA	10	5	0.976	9.91	0.944	15.77	4.086
MIR	140	Full range	2,400	8	0.987	0.73	0.918	19.15	3.365
MIR	140	SPA	8	6	0.926	17.55	0.894	21.76	2.961
MIR	140	UVE	551	7	0.990	6.44	0.974	10.84	5.945
MIR	140	UVE-SPA	7	6	0.930	17.06	0.903	20.82	3.095
NMR	114	Full range	3,636	8	0.985	7.90	0.959	13.56	4.752
NMR	114	SPA	10	5	0.969	11.41	0.954	14.29	4.509
NMR	114	UVE	1,826	4	0.977	9.80	0.967	12.19	5.286
NMR	114	UVE-SPA	8	5	0.976	9.86	0.970	11.48	5.613

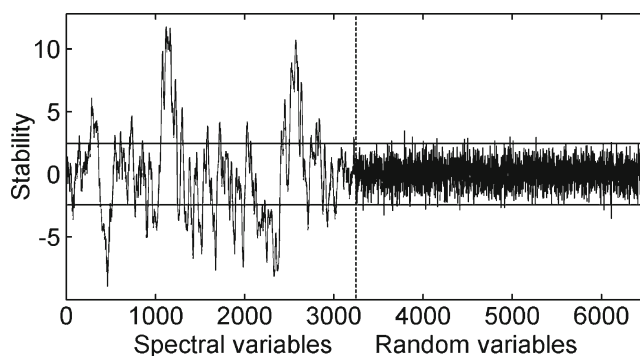
Vis–SNIR visible–short-wave near infrared spectroscopy, *LNIR* long-wave near infrared spectroscopy, *MIR* mid-infrared spectroscopy, *NMR* nuclear magnetic resonance, *SPA* successive projections algorithm, *UVE* uninformative variable elimination, *UVE-SPA* SPA calculation on informative variables selected by UVE,  $r_C^2$  determination coefficient of calibration, *RMSEC* root mean square error of calibration,  $r_{CV}^2$  determination coefficient of cross-validation, *RMSECV* root mean square error of cross-validation, *RPD* residual predictive deviation

range spectra were delivered to the model calibration and no consideration was given to select the important variables contributed for the EPA prediction. In the current section and Sections 3.2.3 and 3.2.4, three variable selection strategies of SPA, UVE, and UVE-SPA were considered to select the important variables.

In the SPA calculation, after comparing the root mean square error (RMSE) values of different candidate subsets of variables that were obtained by a sequence of projection operations, eight variables which had the lowest RMSE were selected from the full Vis–SNIR spectra for the EPA analysis. The selected variables then formed new reduced spectral matrix by selecting the spectral data only at the selected variables, which were then set as the inputs of PLSR model instead of the full range spectra (Table 2). For the Vis–SNIR spectral analysis, the SPA-PLSR model with only eight input variables (0.25 % of variables in the full range) had a similar result with  $r_{CV}^2$  of 0.865 and RMSECV of 24.53 mg/g compared with the result of the PLSR model with the full range spectra (F-PLSR). For the LNIR, MIR, and NMR spectral analysis, 16, 8, and 10 variables were selected by SPA. The input variables of corresponding SPA-PLSR models decreased by 92.04, 99.67, and 99.72 %, respectively, and the corresponding RMSECV decreased by 7.73 %, increased by 13.63 %, and increased by 5.38 %. In addition, the AV\_RMSE of SPA-PLSR models was only 29.59, 68.36, 22.86, and 50.88 % of those of F-PLSR models for Vis–SNIR, LNIR, MIR, and NMR spectral analysis, respectively, showing that the consideration of SPA could reduce the multicollinearity of inputs of the PLSR models and improve the models' robustness. This is agreed with the conclusion of Wu et al. (2013).

#### UVE Calculation Based on the Full Range Spectra

In the process of UVE calculation, the importance of both spectral variables and random variables was quantitatively measured according to their stability, which is shown in Fig. 2 (take Vis–SNIR spectra as an example). Spectra variables were at the left part of the sub-image, while random variables were at the right part. Two horizontal lines were the lower and upper cutoffs used for the variable elimination. Variables that had stability within the cutoff lines were treated as uninformative ones and were eliminated. The remaining variables were then used as the inputs of PLSR models, whose results are shown in Table 2. In general, consideration of UVE significantly improved the predictive ability of PLSR models for the Vis–SNIR and MIR spectral analysis. Their RMSECV decreased by 37.45 and 43.39 % compared with the corresponding F-PLSR models. Meanwhile, the RMSECV decreased by 15.51 and 10.10 % for LNIR and NMR spectral analysis. The results indicated that variables with no more information for EPA prediction than



**Fig. 2** Stability of each variable calculated in UVE process (take Vis–SNIR spectra as an example). Two horizontal lines indicate the lower and upper cutoff

noise were eliminated by the UVE analysis. Compared with the results of SPA-PLSR models, UVE-PLSR models had better prediction capabilities. However, it was noticed that although uninformative variables were eliminated by UVE analysis, there were still hundreds of variables remained because UVE can only select informative variables that might have a problem of multicollinearity. In order to obtain the most important variables that were informative and had no multicollinearity shortcoming, SPA was further applied to select the important variables based on the informative variables selected by UVE.

#### SPA Calculation Based on the UVE-Selected Spectra

Similar to the SPA calculation based on the full range spectra, the best candidate subset of variables was determined by RMSE screen plot in the SPA calculation based on the UVE-selected spectra, resulting in 18, 14, 13, and 13 variables selected from Vis–SNIR, LNIR, MIR, and NMR spectra, respectively. Table 2 presents the main statistics achieved in the calibration and validation of the PLSR models developed by using the important variables selected by UVE-SPA. The variables selected by UVE-SPA decreased by 99.20, 91.87, 98.73, and 99.56 % compared with the variables selected by UVE for Vis–SNIR, LNIR, MIR, and NMR spectra, respectively, showing that the calculation of SPA after UVE could significantly reduce the input variables of UVE-PLSR models. Meanwhile, the UVE-SPA-PLSR models could maintain the accuracy of the UVE-PLSR models for LNIR and NMR spectra, but the RMSECV of UVE-SPA-PLSR models for Vis–SNIR and MIR spectra increased by 37.52 and 92.07 % compared with those of the UVE-PLSR models. On the other hand, the consideration of UVE before SPA improved the efficiency of the SPA calculation for all the spectroscopic techniques analyzed in this study except for Vis–SNIR spectra. The improvement of SPA calculation by conducting SPA after UVE analysis is agreed with the results of other works (Chen et al. 2010; Wu



et al. 2010; Wu et al. 2011). The numbers of the variables selected by UVE-SPA decreased by six, one, and two than those selected by only SPA for LNIR, MIR, and NMR spectra, respectively, meanwhile their RMSECV decreased by 9.52, 4.32, and 19.66 %. The AV\_RMSE also decreased by 33.18, 10.69, and 43.75 % for these three kinds of spectroscopy techniques, showing that their UVE-SPA-PLSR models were more robust than their SPA-PLSR models. For the Vis-SNIR spectra, although its UVE-SPA-PLSR model had a lower RMSECV and higher  $r_{CV}^2$ , there were 12 variables used, while its SPA model only had eight variables. Considering the UVE-SPA model had a larger AV\_RMSE, the important variables for Vis-SNIR spectra in EPA analysis were determined as those selected by SPA. In addition, by comparing the results of the F-PLSR models and the corresponding best PLSR models with variable selection (BV-PLSR model), namely the SPA-PLSR model for Vis-SNIR spectra and the UVE-SPA-PLSR models for LNIR, MIR, and NMR spectra, it was found that the RMSECV values of the BV-PLSR models decreased by 4.52, 16.52, and 15.34 % for Vis-SNIR, LNIR, and NMR spectra, respectively, while that of the MIR model increased by 8.72 %. On the other hand, the AV\_RMSE values of the BV-PLSR model decreased by 70.41, 54.33, 79.59, and 71.38 %. In general, the BV-PLSR models had similar performances of the F-PLSR models, but had lower AV\_RMSE than the F-PLSR models in the EPA analysis. The selected important variables of the BV-PLSR models for Vis-SNIR, LNIR, MIR, and NMR spectroscopy in EPA prediction are shown in Table 3. In Vis-SNIR analysis, 935 nm was assigned to C–H stretching third overtone of CH<sub>2</sub>, 1,028 nm was assigned to a combination of C–H stretching first overtone and C–H deformation second overtone of CH<sub>3</sub>, and 1,045 nm was

assigned to a combination of C–H stretching first overtone and C–H deformation first overtone of CH<sub>2</sub>. In LNIR analysis, 1,178 nm was assigned to C–H stretching second overtone of HC=CH, 1,204 nm was assigned to C–H stretching second overtone of CH<sub>2</sub>, 1,376 nm was assigned to a combination of C–H stretching first overtone and C–H deformation of CH<sub>3</sub>, 1,880 nm was assigned to C=O stretching second overtone of –COOH, 1,932 nm was assigned to a combination of O–H stretching and O–H of H<sub>2</sub>O, and 2,100 nm was assigned to a combination of O–H stretching and O–H of –OH. In MIR analysis, 658 cm<sup>-1</sup> was assigned to the rocking and bending out vibration of –(CH<sub>2</sub>)<sub>n</sub>– and –HC=CH– (*cis*), 1,737 and 1,750 cm<sup>-1</sup> were assigned to –C=O (ester) stretching vibrations, 2,929 cm<sup>-1</sup> was assigned to –C–H (CH<sub>2</sub>) asym stretching, and 2,997 cm<sup>-1</sup> was assigned to =C–H (*cis*) stretching. In NMR analysis, the variables were mainly assigned to –CH<sub>3</sub>, –(CH<sub>2</sub>)<sub>n</sub>–, –CH<sub>2</sub>–CH=CH–, and –OCO–CH<sub>2</sub> of acyl group.

### DHA Analysis

Similar to the EPA analysis, the investigation of four spectroscopic techniques for the DHA prediction started at the calibration based on the full range spectra with the analysis of different preprocessing algorithms. The original spectra were the most suitable data for the analysis of Vis-SNIR, MIR, and NMR spectra, and the spectra preprocessed by SNV had the best prediction result for LNIR spectra. It was noticed that when full range spectra were considered, the DHA content in fish oil could not be well determined by using Vis-SNIR or MIR spectra (Table 4). Variable selection was then conducted to see if the performances of four kinds of spectra could be further improved. At the first step, SPA

**Table 3** Selected important variables of the BV-PLSR models for Vis-SNIR, LNIR, MIR, and NMR spectroscopy in EPA, DHA, and DPA prediction

	Spectroscopy	Important variables of the BV-PLSR model
EPA	Vis-SNIR (nm)	434; 470; 587; 667; 807; 935; 1,028; and 1,045
EPA	LNIR (nm)	1,111; 1,178; 1,204; 1,376; 1,475; 1,626; 1,815; 1,880; 1,932; and 2,100
EPA	MIR (cm <sup>-1</sup> )	658; 841; 1,312; 1,737; 1,750; 2,929; and 2,997
EPA	NMR (ppm)	0.906, 0.992, 1.356, 2.115, 2.352, 2.377, 2.394, and 4.137
DHA	Vis-SNIR (nm)	489, 520, 587, 667, 893, 935, and 1,025
DHA	LNIR (nm)	1,031; 1,151; 1,290; 1,310; 1,396; 1,416; 1,495; 1,587; 1,757; 1,899; 2,093; 2,266; and 2,317
DHA	MIR (cm <sup>-1</sup> )	662; 697; 721; 761; 836; 855; 867; 1,026; 1,057; 1,147; 1,370; 1,717; 1,741; and 2,970
DHA	NMR (ppm)	0.873, 0.912, 0.921, 1.641, 1.963, 2.272, 2.396, 2.764, 4.134, and 4.138
DPA	Vis-SNIR (nm)	444; 470; 513; 587; 892; 935; 1,005; and 1,037
DPA	LNIR (nm)	1,078; 1,184; 1,310; 1,370; 1,403; 1,482; 1,514; 1,534; 1,880; 2,228; 2,254; and 2,279
DPA	MIR (cm <sup>-1</sup> )	658; 661; 703; 763; 835; 1,320; 1,735; 1,741; and 3,035
DPA	NMR (ppm)	0.921, 0.926, 1.295, 1.962, 2.054, 2.386, 2.769, 2.863, 4.137, and 4.146

For full spellings of the abbreviations, see Table 2  
BV-PLSR best PLSR model with variable selection

**Table 4** Predictive results of docosahexaenoic acid (DHA) of fish oil using infrared spectroscopy and nuclear magnetic resonance with chemometrics (available range of DHA is between 100.25 and 223.33 mg)

	Sample number	Variable selection method	Variable number	Number of latent variables	Calibration		Prediction		
					$r_c^2$	RMSEC (mg/g)	$r_{cv}^2$	RMSECV (mg/g)	RPD
Vis–SNIR	140	Full range	3,250	8	0.964	6.37	0.748	17.61	1.919
Vis–SNIR	140	SPA	7	5	0.886	11.42	0.823	14.76	2.289
Vis–SNIR	140	UVE	1,335	7	0.974	5.48	0.903	10.93	3.091
Vis–SNIR	140	UVE-SPA	7	5	0.907	10.30	0.854	13.40	2.522
LNIR	140	Full range	201	11	0.975	5.32	0.869	12.68	2.665
LNIR	140	SPA	13	12	0.965	6.36	0.908	10.61	3.185
LNIR	140	UVE	125	7	0.952	7.42	0.872	12.55	2.692
LNIR	140	UVE-SPA	13	11	0.969	5.97	0.906	10.75	3.143
MIR	140	Full range	2,400	8	0.965	6.35	0.770	16.81	2.010
MIR	140	SPA	15	6	0.89	11.18	0.816	15.03	2.248
MIR	140	UVE	501	8	0.980	4.72	0.940	8.56	3.947
MIR	140	UVE-SPA	14	7	0.937	8.45	0.898	11.22	3.012
NMR	114	Full range	3,636	12	0.991	3.28	0.957	7.24	4.667
NMR	114	SPA	16	14	0.986	4.02	0.972	5.87	5.756
NMR	114	UVE	1,968	9	0.988	3.67	0.974	5.67	5.959
NMR	114	UVE-SPA	10	7	0.987	3.88	0.982	4.73	7.144

For full spellings of the abbreviations, see Table 2

was carried out for selecting important variables based on the full range spectra. There were 7, 13, 15, and 16 variables selected by SPA for Vis–SNIR, LNIR, MIR, and NMR spectra, respectively. New reduced spectral matrixes were then formed based on these variables to establish PLSR models (Table 4). Compared with the F-PLSR models, the application of SPA improved the corresponding PLSR models with the decrements of RMSECV by 16.18, 16.32, 10.59, and 18.92 % for Vis–SNIR, LNIR, MIR, and NMR spectra, respectively. The AV\_RMSE was also decreased by about 60 % when SPA was considered. All of these were achieved when over 90 % variables were deleted.

When UVE was used to eliminate uninformative variables, the RMSECV of Vis–SNIR, MIR, and NMR models decreased by 37.93, 49.08, and 21.69 % and the AV\_RMSE decreased by 51.51, 63.29, and 49.49 % of those of the corresponding F-PLSR models. Although the RMSECV of LNIR spectral model only decreased by 1.03 %, its AV\_RMSE decreased by 30.30 %, and 37.81 % variables were eliminated at the same time. In general, UVE selected informative variables from the full range spectra for DHA prediction, which were then used for further SPA calculation. As a result, there were 7, 13, 14, and 10 variables selected by UVE-SPA for Vis–SNIR, LNIR, MIR, and NMR spectra, respectively. The prediction results of the UVE-SPA-PLSR models for the DHA prediction are shown in Table 4. Compared with the results of UVE-PLSR models, the RMSECV

of UVE-SPA-PLSR models decreased by 14.34 and 16.58 % for LNIR and NMR spectra, respectively, and increased by 22.60 and 31.07 % for Vis–SNIR and MIR spectra, respectively. Therefore, further calculation of SPA on UVE could maintain the models' prediction ability while dramatically reducing the numbers of input variables. On the other hand, the calculation of UVE before SPA was proved to be an effective way to improve the selection of SPA for all the spectroscopic techniques except for LNIR spectra. Especially for the NMR spectra, when only ten variables were selected by UVE-SPA compared with 16 variables selected by SPA, the RMSECV of UVE-SPA-PLSR model decreased by 19.42 % of that of the SPA-PLSR model. The UVE-SPA model for MIR spectra also had a decrement of 25.35 % of that of the corresponding SPA-PLSR model. In addition, the AV\_RMSE of the UVE-SPA models for NMR and MIR spectra decreased respectively by 28.05 and 54.05 % of those of SPA models. In the comparison between the BV-PLSR models (SPA-PLSR model for LNIR spectra and UVE-SPA-PLSR models for Vis–SNIR, MIR, and NMR spectra) and the F-PLSR models, it was found that the BV-PLSR models had better performances with the RMSECV decrements of 23.91, 16.32, 33.25, and 34.67 % for Vis–SNIR, LNIR, MIR, and NMR spectra, respectively. On the other hand, the AV\_RMSE values of the BV-PLSR models decreased by 72.42, 42.26, 73.52, and 78.54 %. The above results demonstrated that the variable selection improved the

prediction accuracy and robustness of the PLSR model in the DHA analysis. The selected important variables of the BV-PLSR models for Vis–SNIR, LNIR, MIR, and NMR spectroscopy in DHA prediction are shown in Table 3. In Vis–SNIR analysis, 893 and 935 nm were assigned to C–H stretching third overtone of CH<sub>3</sub> and CH<sub>2</sub>, respectively, and 1,025 nm was assigned to a combination of C–H stretching first overtone and C–H deformation second overtone of CH<sub>3</sub>. In LNIR analysis, 1,031 nm was assigned to a combination of C–H stretching first overtone and C–H deformation second overtone of CH<sub>3</sub>, 1,051 nm was assigned to a combination of C–H stretching first overtone and C–H deformation first overtone of CH<sub>2</sub>, 1,396 and 1,416 nm were assigned to a combination of C–H stretching first overtone and C–H deformation of CH<sub>2</sub>, 1,495 nm was assigned to O–H stretching first overtone of –OH, 1,757 nm was assigned to C–H stretching first overtone of CH<sub>2</sub>, 1,899 nm was assigned to C=O stretching second overtone of –COOH, 2,093 nm was assigned to a combination of O–H stretching and O–H deformation of –OH, and 2,266 and 2,317 nm were assigned to a combination of C–H stretching and C–H deformation of CH<sub>3</sub> and CH<sub>2</sub>, respectively. In MIR analysis, 662, 697, 721, and 761 cm<sup>-1</sup> were assigned to the rocking and bending out vibration of –(CH<sub>2</sub>)<sub>n</sub>– and –HC=CH– (*cis*), 1,026 and 1,057 cm<sup>-1</sup> were assigned to –C–O stretching, 1,147 cm<sup>-1</sup> was assigned to stretching and bending vibration of –C–O and –CH<sub>2</sub>, 1,370 cm<sup>-1</sup> was assigned to –C–H (CH<sub>3</sub>) bending (sym) vibration, and 1,717 and 1,741 cm<sup>-1</sup> were assigned to –C=O (ester) stretching vibration. In NMR analysis, the variables were mainly assigned to –CH<sub>3</sub>, –OCO–CH<sub>2</sub>–CH<sub>2</sub>–, –CH<sub>2</sub>–CH=CH–, –OCO–CH<sub>2</sub>–, and =HC–CH<sub>2</sub>–CH= of acyl group.

#### DPA Analysis

Similar to the analysis of the above two PUFA, the original full range spectra and those preprocessed by different preprocessing algorithms of four spectroscopic techniques were used separately to establish PLSR models for the prediction of DPA content in fish oil. It was found that the spectra preprocessed by 2nd Der had the best prediction result for LNIR spectra, while the original spectra were the most suitable data for other three kinds of spectra. When SPA was applied to select the important variables from the full range spectra for establishing PLSR model, the predictive accuracy of Vis–SNIR, MIR, and NMR spectra had been improved with the RMSECV decrements of 22.36, 28.525, and 28.30 %, respectively, while the RMSECV of the LNIR spectral model increased by 4.93 %. On the other hand, the AV\_RMSE values decreased from 34.62 to 61.48 % for four spectroscopic techniques with the involvement of SPA. In addition, there were less than 10 % variables selected by SPA for the establishment of PLSR models.

The other strategy of SPA process for variable selection was to execute it based on the informative variables selected by UVE instead of the full range spectra. After the UVE calculation, there were 1,334, 102, 487, and 1,647 variables selected for Vis–SNIR, LNIR, MIR, and NMR spectra, respectively. The PLSR models established based on these variables selected by UVE had the RMSECV decrements by 21.94, 28.17, 61.41, and 44.65 % for four spectroscopic techniques, respectively, compared with those of F-PLSR models. The above results indicated that the variables selected by UVE were informative and could be used for the further SPA calculation. After the calculation, there were 7, 12, 8, and 10 variables selected by UVE-SPA for Vis–SNIR, LNIR, MIR, and NMR spectra, respectively. The prediction results of the UVE-SPA-PLSR models are shown in Table 5. For the Vis–SNIR spectra, the UVE-SPA-PLSR model had similar results of the SPA-PLSR and UVE-PLSR models. For MIR spectra, the UVE-SPA-PLSR model had a poorer performance than the SPA-PLSR model, in which the RMSECV decreased by 5.63 %. Therefore, the consideration of UVE before SPA had no improvement for the variable selection, and the SPA-PLSR model was the best choice for both Vis–SNIR and MIR spectra in the DPA analysis. On the other hand, the UVE-SPA-PLSR models had better prediction results than the SPA-PLSR models for LNIR and NMR spectra. The RMSECV decreased by 42.95 and 32.46 % and the AV\_RMSE decreased by 41.03 and 18.18 % for LNIR and NMR spectra, respectively. Compared with the results of the F-PLSR models, it was found that the BV-PLSR models had better performances with most variables eliminated. The RMSECV values decreased by 22.36, 40.14, 28.52, and 51.57 % and the AV\_RMSE values decreased by 61.48, 73.56, 34.62, and 66.67 % for Vis–SNIR, LNIR, MIR, and NMR spectra, respectively, showing that the prediction accuracy and robustness of the PLSR model could also be improved by the variable selection in the DPA analysis. The selected important variables of the BV-PLSR models for Vis–SNIR, LNIR, MIR, and NMR spectroscopy in DPA prediction are shown in Table 3. In Vis–SNIR analysis, 892 and 935 nm were assigned to C–H stretching third overtone of CH<sub>3</sub> and CH<sub>2</sub>, respectively, 892 nm was assigned to C–H stretching third overtone of CH<sub>3</sub>, 1,005 nm was assigned to a combination of C–H stretching first overtone and C–H deformation second overtone of CH<sub>3</sub>, and 1,037 nm was assigned to a combination of C–H stretching first overtone and C–H deformation first overtone of CH<sub>2</sub>. In LNIR analysis, 1,078 and 1,184 nm were assigned to C–H stretching second overtone of CH, CH<sub>2</sub>, CH<sub>3</sub>, and HC=CH, 1,310, 1,370, 1,403, 1,482, 1,514, and 1,534 nm were assigned to a combination of C–H stretching first overtone and C–H deformation first overtone in CH, CH<sub>2</sub>, and CH<sub>3</sub>, 1,880 nm was assigned to C=O stretching second overtone of COOH, 2,228 and 2,254 nm were assigned to a combination of CH<sub>2</sub>

**Table 5** Predictive results of docosapentaenoic acid (DPA) of fish oil using infrared spectroscopy and nuclear magnetic resonance with chemometrics (available range of DPA is between 10.95 and 32.60 mg/g)

	Sample number	Variable selection method	Variable number	Number of latent variables	Calibration		Prediction		
					$r_C^2$	RMSEC (mg/g)	$r_{CV}^2$	RMSECV (mg/g)	RPD
Vis–SNIR	140	Full range	3,250	7	0.968	1.02	0.841	2.37	2.409
Vis–SNIR	140	SPA	8	7	0.947	1.32	0.905	1.84	3.103
Vis–SNIR	140	UVE	1,334	6	0.960	1.14	0.903	1.85	3.086
Vis–SNIR	140	UVE-SPA	7	6	0.944	1.36	0.901	1.87	3.053
LNIR	140	Full range	201	8	0.991	0.55	0.943	1.42	4.021
LNIR	140	SPA	16	5	0.964	1.10	0.937	1.49	3.832
LNIR	140	UVE	102	7	0.991	0.55	0.971	1.02	5.598
LNIR	140	UVE-SPA	12	7	0.988	0.62	0.979	0.85	6.718
MIR	140	Full range	2,400	4	0.852	2.20	0.750	2.98	1.916
MIR	140	SPA	9	6	0.920	1.62	0.872	2.13	2.681
MIR	140	UVE	487	10	0.991	0.54	0.962	1.15	4.965
MIR	140	UVE-SPA	8	5	0.899	1.82	0.857	2.25	2.538
NMR	114	Full range	3,636	7	0.967	1.05	0.929	1.59	3.591
NMR	114	SPA	12	7	0.974	0.92	0.964	1.14	5.009
NMR	114	UVE	1,647	8	0.988	0.96	0.978	0.88	6.489
NMR	114	UVE-SPA	10	9	0.989	0.59	0.983	0.77	7.416

For full spellings of the abbreviations, see Table 2

stretching and C= stretching of HC=CH and CH<sub>3</sub>, respectively, and 2,279 nm was assigned to a combination of C–H stretching and C–H deformation in CH<sub>3</sub>. In MIR analysis, 658, 661, 703, and 763 cm<sup>-1</sup> were assigned to the rocking and bending out vibration of  $-(CH_2)_n-$  and  $-HC=CH-$  (*cis*), 1,735 and 1,740 cm<sup>-1</sup> were assigned to  $-C=O$  (ester) stretching vibration, and 3,035 cm<sup>-1</sup> was assigned to  $=C-H$  (*cis*) stretching. In NMR analysis, the variables were mainly assigned to  $-CH_3$ ,  $-(CH_2)_n-$ ,  $-CH_2-CH=CH-$ ,  $-OCO-CH_2-$ , and  $=HC-CH_2-CH=$  of acyl group.

## Discussion

The potential of four spectroscopic techniques of Vis–SNIR, LNIR, MIR, and NMR was investigated for the rapid and reliable assessment of EPA, DHA, and DPA in fish oil. Specifically, in the EPA analysis, the F-PLSR model for NMR spectra had a high  $r_{CV}^2$  of 0.959, which was further improved to 0.970 based on the BV-PLSR model with eight important variables selected by UVE-SPA. The results of LNIR spectra were also acceptable, in which  $r_{CV}^2$  of 0.920 was obtained by the F-PLSR model and  $r_{CV}^2$  of 0.944 by the BV-PLSR model with ten important variables. Although the UVE calculation helped the MIR spectra to obtain a good prediction of EPA with  $r_{CV}^2$  of 0.974, the SPA calculation either based on full range spectra or variables selected by

UVE could only cause the PLSR models having the  $r_{CV}^2$  of around 0.9. Similar situation happened for Vis–SNIR spectra. Although the UVE-PLSR model had  $r_{CV}^2$  of 0.942, either the SPA-PLSR model or the UVE-SPA-PLSR model had  $r_{CV}^2$  lower than 0.9. In the aspect of the robustness of the established BV-PLSR models, NMR spectra had the smallest AV\_RMSE value of 0.162, followed by MIR spectra that had the value of 0.376. The AV\_RMSE values of other two spectra were larger than 0.5.

In the DHA analysis, the BV-PLSR model of Vis–SNIR spectra had the poorest prediction with the RMSECV of 13.40 mg/g. The RMSECV decreased to around 10 mg/g for both the BV-PLSR models of LNIR and MIR spectra. The best prediction with  $r_{CV}^2$  of 0.982 and the RMSECV of 4.73 mg/g was obtained by the BV-PLSR model for NMR spectra. The prediction of DHA using NMR spectra in tandem with multivariable selection and calibration had better prediction than internal standard method and DHA calibration curve method used by Igarashi et al. (2000), in which RMSE values were 14.92 and 17.26 mg/g for above two methods, respectively (calculated from data in Table 3 in the work of Igarashi et al. (2000)). Therefore, the use of multivariable analysis was an efficient way to process NMR data instead of internal standard method and calibration curve method for DHA determination in fish oil. In addition, the BV-PLSR model for NMR spectra also had the smallest AV\_RMSE value of 0.085 showing its good robustness,



while the BV-PLSR models for other three spectra had the AV\_RMSE values all higher than 0.25. In the DPA analysis, the BV-PLSR models of both LNIR and NMR spectra achieved good prediction with  $r_{CV}^2$  of 0.983 and 0.979 and small AV\_RMSE values of 0.018 and 0.022, respectively. On the other hand, the prediction of the BV-PLSR model using either Vis–SNIR or MIR spectra was not very good (the  $r_{CV}^2$  values were 0.905 and 0.872).

The above results indicated that the important PUFA (EPA, DHA, and DPA) contents in fish oil could be determined by spectroscopic techniques in a rapid and reliable way. The NMR spectroscopy was the most suitable one among the four spectroscopic techniques investigated for the determination of EPA, DHA, and DPA contents in fish oil. The RPD values of the BV-PLSR models were 5.613, 7.144, and 7.416, showing that the BV-PLSR models were adequate for quality control and even process control (Williams 2001). The LNIR spectroscopy was also a good choice (RPD=4.086 for EPA, 3.185 for DHA, and 6.718 for DPA). Especially considering that its measurement of using a small spectrometer (dimensions of 153.4×105.2×76.2 mm) was nondestructive, inexpensive, and very convenient to implement. It should be noticed that spectroscopy techniques analyzed with chemometrics are indirect methods that need reference values of target attribute for model calibration. Therefore, the accuracy of spectral models depends on and could not be better than the accuracy of the reference method. The reference method should be conducted carefully to minimize its test error, so that the corresponding established spectral models would be meaningful. In this work, the reference EPA, DHA, and DPA values were measured carefully using GC method. As GC is a standard method to determine fatty acids in foods and biological samples (AOAC Method 963.22 2000), its measured PUFA values could be used as the reference values for establishing spectral models.

Moreover, it was found that there were 11 twelfth cases (four spectroscopic techniques × three PUFA), in which the BV-PLSR models had better predictions (RMSECV decreased by an average of 23.20 % in all 12 cases) than the corresponding F-PLSR models. Meanwhile, the AV\_RMSE of the BV-PLSR models decreased by an average of 64.90 % compared with the corresponding F-PLSR models in all 12 cases. Such improvement was achieved after over 90 % variables of the full range spectra were featured as redundant and uninformative and the remaining variables were used to establish the PLSR model. The above statistical data showed that the variable selection with the help of UVE and SPA was important for improving the prediction accuracy and robustness of the PLSR models for the determination of EPA, DHA, and DPA contents in fish oil. Compared with the work of Azizian et al. (2010) in which the PLSR models were established using near infrared spectra without variable

selection (RMSECV of 3.23 % for EPA and 2.57 % for DHA), the PLSR models established in this work using near infrared spectra with variable selection by UVE-SPA had lower RMSECV of 15.77 mg/g for EPA and 10.75 mg/g for DHA, showing that the variable selection was efficient to increase the prediction accuracy of PUFA in fish oil using spectroscopy technology.

In addition, among the 12 cases researched in this study, there were eight cases in which UVE-SPA was the best strategy of variable selection. The SPA calculation based on the full range spectra was considered as the best for the rest four cases. Although it was found in other cases that the calculation of UVE could absolutely improve the results of SPA for spectral variable selection (Chen et al. 2010; Wu et al. 2010; Wu et al. 2011), the results of this work indicated that the elimination of uninformative variables was helpful for improving the efficiency of SPA for only some spectral data, while for other spectra, SPA could directly select the important variables from the full range spectra. Therefore, it is suggested that for different applications, both variable selection strategies of SPA and UVE-SPA should be analyzed and compared to choose the best one.

## Conclusions

This study was carried out to evaluate the feasibility of using Vis–SNIR, LNIR, MIR, and NMR spectroscopic techniques for rapid prediction of three PUFA (EPA, DHA, and DPA) in fish oil. The results discussed in this paper indicated that the four techniques were reasonably efficient as rapid and convenient tools instead of the traditional GC method for assessing the EPA, DHA, and DPA contents in fish oil. By means of PLSR algorithm, quantitative relationships were correlated between the spectral data and the PUFA contents of fish oil samples. All results showed good performances of the established PLSR models for predicting PUFA by four spectroscopic techniques. Among them, NMR spectra had the best performances for all three PUFA predictions, followed by LNIR, MIR, and Vis–SNIR spectra. Since the full range spectra were characterized as high dimensionality, multicollinearity, and redundancy, variable selection was conducted to locate the most important variables for the prediction. The results indicated that the BV-PLSR models with over 90 % variables eliminated had lower RMSECV with an average decrement of 23.20 % and had lower AV\_RMSE with an average decrement of 64.90 % in all 12 cases compared with the corresponding F-PLSR models. Two variable selection algorithms of UVE and SPA were investigated in this work to see if the efficiency of SPA for variable selection could be improved based on the informative variables selected by UVE. In 67 % cases, UVE-SPA had better prediction results than the direct calculation of

SPA based on the full range spectra, while the UVE calculation had no improvement for the SPA selection in other cases. In view of the quality evaluation of fish oil, the results of this study verified the substantial propensity of the spectroscopic technologies to be excellent alternatives to the time-consuming and conventional GC methods. By implementing spectroscopic techniques as a key component of a rapid inspection system, it could enable the fish oil industry to sort, label, and price fish oil products truthfully according to their PUFA contents.

**Acknowledgments** This study was supported by the Zhejiang Provincial Natural Science Foundation of China under grant Y3110289, Natural Science Foundation of China (31072247), and the Fundamental Research Funds for the Central Universities.

## References

- AOAC Method 963.22. (2000). *Methyl esters of fatty acids in oils and fats, official methods of analysis of the AOAC* (7th ed., pp. 24–26). Gaithersburg: AOAC.
- Araujo, M. C. U., Saldanha, T. C. B., Galvao, R. K. H., Yoneyama, T., Chame, H. C., & Visani, V. (2001). The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems*, 57(2), 65–73.
- Azzian, H., Kramer, J. K. G., Ehler, S., & Curtis, J. M. (2010). Rapid quantitation of fish oil fatty acids and their ethyl esters by FT-NIR models. *European Journal of Lipid Science and Technology*, 112(4), 452–462.
- Balabin, R. M., & Safieva, R. Z. (2011). Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data. *Analytica Chimica Acta*, 689(2), 190–197.
- Balabin, R. M., & Smirnov, S. V. (2011). Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data. *Analytica Chimica Acta*, 692(1–2), 63–72.
- Berquin, I. M., Min, Y., Wu, R., Wu, J., Perry, D., Cline, J. M., Thomas, M. J., Thornburg, T., Kulik, G., & Smith, A. (2007). Modulation of prostate cancer genetic risk by omega-3 and omega-6 fatty acids. *Journal of Clinical Investigation*, 117(7), 1866.
- Bousquet, M., Saint-Pierre, M., Julien, C., Salem, N., Cicchetti, F., & Calon, F. (2008). Beneficial effects of dietary omega-3 polyunsaturated fatty acid on toxin-induced neuronal degeneration in an animal model of Parkinson's disease. *The FASEB Journal*, 22(4), 1213–1225.
- Calder, P. C., Kremmyda, L.-S., Vlachava, M., Noakes, P. S., & Miles, E. A. (2010). Is there a role for fatty acids in early life programming of the immune system? *Proceedings of the Nutrition Society*, 69(03), 373–380.
- Centner, V., Massart, D. L., deNoord, O. E., deJong, S., Vandeginste, B. M., & Sterna, C. (1996). Elimination of uninformative variables for multivariate calibration. *Analytical Chemistry*, 68(21), 3851–3858.
- Chen, X. J., Li, H., Wu, D., Lei, X. X., Zhu, X. O., & Zhang, A. J. (2010). Application of a hybrid variable selection method for the classification of rapeseed oils based on H-1 NMR spectral analysis. *European Food Research and Technology*, 230(6), 981–988.
- Cozzolino, D., Kwiatkowski, M., Waters, E., & Gishen, M. (2007). A feasibility study on the use of visible and short wavelengths in the near-infrared region for the non-destructive measurement of wine composition. *Analytical and Bioanalytical Chemistry*, 387(6), 2289–2295.
- Cozzolino, D., Murray, I., Chree, A., & Scaife, J. (2005). Multivariate determination of free fatty acids and moisture in fish oils by partial least-squares regression and near-infrared spectroscopy. *LWT—Food Science and Technology*, 38(8), 821–828.
- Curtis, J. M. (2007). Analysis of oils and concentrates. In Breivik (Ed.), *Long-chain omega-3 specialty oils* (pp. 219–241). Bridgewater: The Oily Press.
- ElMasry, G., Sun, D.-W., & Allen, P. (2011). Non-destructive determination of water-holding capacity in fresh beef by using NIR hyperspectral imaging. *Food Research International*, 44(9), 2624–2633.
- Endo, Y., Tagiri-Endo, M., & Kimura, K. (2005). Rapid determination of iodine value and saponification value of fish oils by near-infrared spectroscopy. *Journal of Food Science*, 70(2), C127–C131.
- Flåtten, A., Bryhni, E. A., Kohler, A., Egeland, B., & Isaksson, T. (2005). Determination of C22: 5 and C22: 6 marine fatty acids in pork fat with Fourier transform mid-infrared spectroscopy. *Meat Science*, 69(3), 433–440.
- Guesnet, P., & Alessandri, J. M. (2011). Docosahexaenoic acid (DHA) and the developing central nervous system (CNS)—implications for dietary recommendations. *Biochimie*, 93(1), 7–12.
- Guillen, M. D., Carton, I., Goicoechea, E., & Uriarte, P. S. (2008). Characterization of cod liver oil by spectroscopic techniques. New approaches for the determination of compositional parameters, acyl groups, and cholesterol from H-1 nuclear magnetic resonance and Fourier transform infrared spectral data. *Journal of Agricultural and Food Chemistry*, 56(19), 9072–9079.
- Guillén, M. D., & Ruiz, A. (2003). Rapid simultaneous determination by proton NMR of unsaturation and composition of acyl groups in vegetable oils. *European Journal of Lipid Science and Technology*, 105(11), 688–696.
- Huan, M., Hamazaki, K., Sun, Y., Itomura, M., Liu, H., Kang, W., Watanabe, S., Terasawa, K., & Hamazaki, T. (2004). Suicide attempt and n-3 fatty acid levels in red blood cells: a case control study in China. *Biological Psychiatry*, 56(7), 490–496.
- Igarashi, T., Aursand, M., Hirata, Y., Gribbestad, I. S., Wada, S., & Nonaka, M. (2000). Nondestructive quantitative determination of docosahexaenoic acid and n-3 fatty acids in fish oils by high-resolution 1 H nuclear magnetic resonance spectroscopy. *Journal of the American Oil Chemists' Society*, 77(7), 737–748.
- Institute of Medicine, National Academy of Science. (2005). *Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein, and amino acids (macronutrients)*. Washington: The National Academies Press.
- Jensen, C. L., Voigt, R. G., Llorente, A. M., Peters, S. U., Prager, T. C., Zou, Y. L., Rozelle, J. C., Turcich, M. R., Fraley, J. K., & Anderson, R. E. (2010). Effects of early maternal docosahexaenoic acid intake on neuropsychological status and visual acuity at five years of age of breast-fed term infants. *Journal of Pediatrics*, 157(6), 900–905.
- Kaur, G., Cameron-Smith, D., Garg, M., & Sinclair, A. J. (2011). Docosapentaenoic acid (22: 5n-3): a review of its biological effects. *Progress in Lipid Research*, 50(1), 28–34.
- Kimiya, T., Sivertsen, A. H., & Heia, K. (2013). VIS/NIR spectroscopy for non-destructive freshness assessment of Atlantic salmon (*Salmo salar* L.) fillets. *Journal of Food Engineering*, 116(3), 758–764.
- Marini, F., Balestrieri, F., Bucci, R., Magri, A. L., & Marini, D. (2003). Supervised pattern recognition to discriminate the geographical origin of rice bran oils: a first study. *Microchemical Journal*, 74(3), 239–248.
- Martens, H. A., & Dardenne, P. (1998). Validation and verification of regression in small data sets. *Chemometrics and Intelligent Laboratory Systems*, 44(1), 99–121.

- Martins, J. G. (2009). EPA but not DHA appears to be responsible for the efficacy of omega-3 long chain polyunsaturated fatty acid supplementation in depression: evidence from a meta-analysis of randomized controlled trials. *Journal of the American College of Nutrition*, 28(5), 525–542.
- Masoum, S., Malabat, C., Jalali-Heravi, M., Guillou, C., Rezzi, S., & Rutledge, D. N. (2007). Application of support vector machines to <sup>1</sup>H NMR data of fish oils: methodology for the confirmation of wild and farmed salmon and their origins. *Analytical and Bioanalytical Chemistry*, 387(4), 1499–1510.
- Miyake, Y., Yokomizo, K., & Matsuzaki, N. (1998). Determination of unsaturated fatty acid composition by high-resolution nuclear magnetic resonance spectroscopy. *Journal of the American Oil Chemists' Society*, 75(9), 1091–1094.
- Morris, M. C., Evans, D. A., Tangney, C. C., Bienias, J. L., & Wilson, R. S. (2005). Fish consumption and cognitive decline with age in a large community study. *Archives of Neurology*, 62(12), 1849.
- Nemets, H., Nemets, B., Apter, A., Bracha, Z., & Belmaker, R. (2006). Omega-3 treatment of childhood depression: a controlled, double-blind pilot study. *The American Journal of Psychiatry*, 163(6), 1098–1100.
- Olsen, E. F., Rukke, E.-O., Flåtten, A., & Isaksson, T. (2007). Quantitative determination of saturated-, monounsaturated- and polyunsaturated fatty acids in pork adipose tissue with non-destructive Raman spectroscopy. *Meat Science*, 76(4), 628–634.
- Osborne, B. G., & Fearn, T. (1986). *Near-infrared spectroscopy in food analysis*. New York: Longman Scientific and Technical.
- Papadopoulou, O., Panagou, E. Z., Tassou, C. C., & Nychas, G. J. E. (2011). Contribution of Fourier transform infrared (FTIR) spectroscopy data on the quantitative determination of minced pork meat spoilage. *Food Research International*, 44(10), 3264–3271.
- Rennie, K., Hughes, J., Lang, R., & Jebb, S. (2003). Nutritional management of rheumatoid arthritis: a review of the evidence. *Journal of Human Nutrition and Dietetics*, 16(2), 97–109.
- Rivero, M. J., Valderrama, X., Haines, D., & Alomar, D. (2012). Prediction of immunoglobulin G content in bovine colostrum by near-infrared spectroscopy. *Journal of Dairy Science*, 95(3), 1410–1418.
- Sacchi, R., Medina, I., Aubourg, S. P., Addeo, F., & Paolillo, L. (1993). Proton nuclear magnetic resonance rapid and structure-specific determination of  $\omega$ -3 polyunsaturated fatty acids in fish lipids. *Journal of the American Oil Chemists' Society*, 70(3), 225–228.
- Schönberg, S. A., Lundemo, A. G., Fladvad, T., Holmgren, K., Bremseth, H., Nilsen, A., Gederaas, O., Tvedt, K. E., Egeberg, K. W., & Krokan, H. E. (2006). Closely related colon cancer cell lines display different sensitivity to polyunsaturated fatty acids, accumulate different lipid classes and downregulate sterol regulatory element-binding protein 1. *FEBS Journal*, 273(12), 2749–2765.
- Shaikh, I. A. A., Brown, I., Schofield, A. C., Wahle, K. W. J., & Heys, S. D. (2008). Docosahexaenoic acid enhances the efficacy of docetaxel in prostate cancer cells by modulation of apoptosis: the role of genes associated with the NF- $\kappa$ B pathway. *Prostate*, 68(15), 1635–1646.
- Sileoni V, Marconi O, Perretti G & Fantozzi P (2013) Evaluation of different validation strategies and long term effects in NIR calibration models. *Food Chemistry*, 141(3), 2639–2648.
- Song, C., & Zhao, S. (2007). Omega-3 fatty acid eicosapentaenoic acid. A new treatment for psychiatric and neurodegenerative diseases: a review of clinical investigations. *Expert Opinion on Investigational Drugs*, 16(10), 1627–1638.
- Tyl, C. E., Brecker, L., & Wagner, K. H. (2008). <sup>1</sup>H NMR spectroscopy as tool to follow changes in the fatty acids of fish oils. *European Journal of Lipid Science and Technology*, 110(2), 141–148.
- Wang, C., Harris, W. S., Chung, M., Lichtenstein, A. H., Balk, E. M., Kupelnick, B., Jordan, H. S., & Lau, J. (2006). n-3 Fatty acids from fish or fish-oil supplements, but not  $\alpha$ -linolenic acid, benefit cardiovascular disease outcomes in primary- and secondary-prevention studies: a systematic review. *American Journal of Clinical Nutrition*, 84(1), 5–17.
- Williams, P. C. (2001). Implementation of near-infrared technology. In Williams & Norris (Eds.), *Near-infrared technology in the agricultural and food industries* (2nd ed., pp. 145–169). Saint Paul: American Association of Cereal Chemists.
- Wolters, M. (2005). Diet and psoriasis: experimental data and clinical evidence. *British Journal of Dermatology*, 153(4), 706–714.
- Wright, S. A., O'Prey, F. M., McHenry, M. T., Leahey, W. J., Devine, A. B., Duffy, E. M., Johnston, D. G., Finch, M. B., Bell, A. L., & McVeigh, G. E. (2008). A randomised interventional trial of  $\omega$ -3 polyunsaturated fatty acids on endothelial function and disease activity in systemic lupus erythematosus. *Annals of the Rheumatic Diseases*, 67(6), 841–848.
- Wu, D., Chen, X., Zhu, X., Guan, X., & Wu, G. (2011). Uninformative variable elimination for improvement of successive projections algorithm on spectral multivariable selection with different calibration algorithms for the rapid and non-destructive determination of protein content in dried laver. *Analytical Methods*, 3(8), 1790–1796.
- Wu, D., He, Y., Feng, S. J., & Sun, D.-W. (2008). Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM. *Journal of Food Engineering*, 84(1), 124–131.
- Wu, D., He, Y., Nie, P. C., Cao, F., & Bao, Y. D. (2010). Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. *Analytica Chimica Acta*, 659(1–2), 229–237.
- Wu, D., Shi, H., Wang, S., He, Y., Bao, Y., & Liu, K. (2012). Rapid prediction of moisture content of dehydrated prawns using online hyperspectral imaging system. *Analytica Chimica Acta*, 726, 57–66.
- Wu, D., Wang, S., Wang, N., Nie, P., He, Y., Sun, D.-W., & Yao, J. (2013). Application of time series-hyperspectral imaging (TS-HSI) for determining water distribution within beef and spectral kinetic analysis during dehydration. *Food and Bioprocess Technology*. doi:10.1007/s11947-012-0928-0.
- Yurko-Mauro, K., McCarthy, D., Rom, D., Nelson, E. B., Ryan, A. S., Blackwell, A., Salem, N., Jr., & Stedman, M. (2010). Beneficial effects of docosahexaenoic acid on cognition in age-related cognitive decline. *Alzheimer's & Dementia*, 6(6), 456–464.