






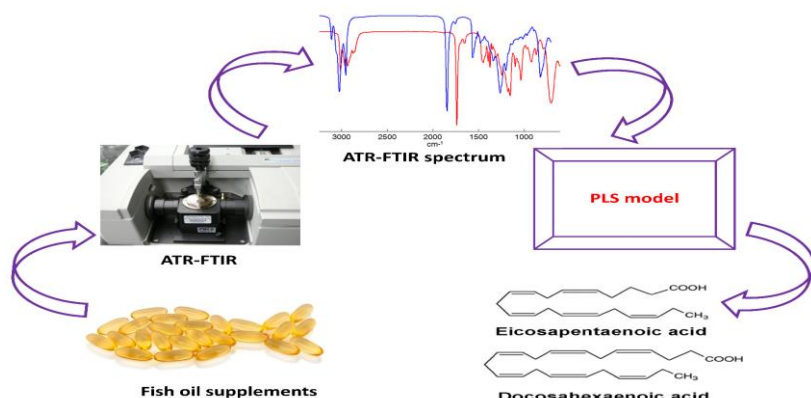
Full paper | <http://dx.doi.org/10.17807/orbital.v16i219869>

# Partial Least Squares Regression Method to Predict Docosahexaenoic and Eicosapentaenoic Acids in Fish Oil Supplements

Thiago Inácio Barros Lopes\* <sup>a</sup>, Tainara Andrade do Nascimento <sup>b</sup>, Elba de Souza Pereira <sup>b</sup>, Samuel Leite de Oliveira <sup>c</sup>, Diego Galvan <sup>d</sup>, and Glaucia Braz Alcantara <sup>b</sup>

Fish oil dietary supplements have been linked to various health benefits due to the high concentration of omega-3 polyunsaturated fatty acid ( $\omega$ -3 PUFA). The potential use of attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy with partial least squares regression (PLSR) was assessed to determine the docosahexaenoic acid (DHA), eicosapentaenoic acid (EPA), and  $\omega$ -3 PUFA in commercial fish oil capsules taking as a reference <sup>1</sup>H NMR spectroscopy values. Comparing the results achieved by interval PLS (iPLS), synergy interval PLS (siPLS), and backward interval PLS (biPLS) algorithms, it was found that siPLS provided the best results. The proposed method predicted DHA with a coefficient of determination ( $R^2$ ) of 0.990, root mean square error of cross-validation (RMSECV) of 0.625%, and root mean square error of prediction (RMSEP) of 1.941. EPA ( $R^2$ =0.976, RMSECV=0.789%, and RMSEP=2.795%) and  $\omega$ -3 PUFA ( $R^2$ =0.978, RMSECV=2.667%, and RMSEP=3.980%). The results indicated that ATR-FTIR and siPLS provided a robust method that could be employed in the analysis and quality control of fish oil supplement capsules. This method has the advantage of being simple, fast, and non-destructive for quantitative analysis.

## Graphical abstract



## Keywords

biPLS  
DHA  
EPA  
Infrared spectroscopy  
iPLS  
siPLS

## Article history

Received 13 Dec 2023  
Accepted 22 May 2024  
Available online 24 Jun 2024

Handling Editor: Adilson Beatriz

## 1. Introduction

<sup>a</sup> Instituto Federal de Educação, Ciência e Tecnologia de Mato Grosso do Sul (IFMS), Aquidauana, MS, Brazil, CEP 79200-000. <sup>b</sup> Institute of Chemistry, Federal University of Mato Grosso do Sul (UFMS), CP 549, zip code 79074-460, Campo Grande, Mato Grosso do Sul, Brazil. <sup>c</sup> Institute of Physics, Federal University of Mato Grosso do Sul (UFMS), CP 549, zip code 79070-900, Campo Grande, Mato Grosso do Sul, Brazil. <sup>d</sup> Department of Chemistry, Federal University of Santa Catarina (UFSC), CP 476, zip code 88040-900, Florianópolis, Santa Catarina, Brazil. \*Corresponding author. E-mail: [inacio\\_thiago@hotmail.com](mailto:inacio_thiago@hotmail.com)

Omega-3 polyunsaturated fatty acids ( $\omega$ -3 PUFA) are essential nutrients that provide beneficial health through their molecular, cellular, and physiological actions [1]. Docosahexaenoic (DHA) and eicosapentaenoic (EPA) acids are essential, meaning that the body does not synthesize them and must be achieved from dietary sources. In many cases, supplementation is necessary to provide a satisfactory ingestion of these fatty acids (FA) [2]. EPA and DHA are almost exclusively found in seafood products. Fish oil is the main source of  $\omega$ -3 PUFA in supplementation form; but other sources, such as microalgae, have received interest. The concentration of  $\omega$ -3 PUFA in fish oil is highly variable, depending on the fish type and season. Additionally, the content of specific FA (EPA and DHA) can be increased through transesterification, fractionation, and/or enzymatic processes [3]. During the transesterification process, the glycerol backbone of triacylglyceride (TG) is removed from all FA. Then, the PUFA as free FA is esterified again as an ethyl ester (EE), while some shorter chain FA are taken out.

Dietary supplements accounted for over 65% of the total  $\omega$ -3 PUFA ingredients in 2014, while functional foods accounted for over 11% of the total. The market size was USD 1.88 billion in 2014, and it is expected to grow to USD 3.79 billion by 2022 [4]. The high cost of fish oil makes it susceptible to adulteration and mislabeling. To improve profits, unscrupulous manufacturers replace high-priced oils with cheaper ones (such as vegetable oil), decreasing the nutritional value and causing health damage to the consumer. Therefore, there is a demand to develop rapid, effective, and low-cost methods for the quantification of EPA and DHA in fish oil capsules.

Traditionally, chromatographic methods are employed in FA determinations, but the technique requires complicated, time-consuming protocols, long run-times, and destructive analysis. Alternatively, hydrogen nuclear magnetic resonance ( $^1\text{H}$  NMR) has been employed in fish oil analysis, reducing the analysis time and sample preparation [5]. However, NMR applicability is limited by the high equipment cost and necessity of a deuterated solvent. Alternatively, attenuated total reflectance-Fourier transform infrared (ATR-FTIR) spectroscopy and other spectroscopy equipment have opened the door to novel methods for oil analyses, providing fast measurements with minimal sample preparation. Shapaval et al. [6] applied FTIR to quantify PUFA in *Mucor* fungi. Lopes et al. [7] used IR spectroscopy combined with chemometrics to discriminate the types of FA esterification (TG or EE forms). Cebi et al. [8] provide a hand-held and portable FTIR applications for on-site food quality control from farm to fork. Stanley et al. [2] used Raman spectroscopy with partial least squares regression to quantify PUFA and other fatty acid classes directly through capsules.

Partial Least Squares Regression (PLSR), or only PLS, is a "projection method" based on projecting data into subspaces of reduced dimensions, in which the model is built in a single step, in which  $X$  and  $y$  are decomposed simultaneously. PLSR is commonly used in analytical chemistry to construct linear regression models, being an excellent option for modeling multicollinear data sets, as occurs with spectroscopic techniques. The prediction performance and robustness of PLSR models can be improved by variable selection [1]. Many algorithms have been developed and applied to the optimal variable selection, such as genetic algorithm, moving windows based methods, and interval PLS. Papers have reported that interval methods, such as interval PLS (iPLS), synergy interval PLS (siPLS), and backward interval PLS (biPLS), could provide better results than PLS in oil analyses

[9].

This work aimed to develop a rapid, effective, non-destructive, and cost-effective method based on ATR-FTIR with PLSR for monitoring and quantifying DHA, EPA, and  $\omega$ -3 PUFA in capsules of fish oil supplements. All steps of the PLSR, such as spectral pre-processing, variable selection, and the number of latent variables, were systematically studied.

## 2. Material and Methods

### Sample acquisition

Fourteen commercial fish oil supplement samples were purchased from a local market (Campo Grande, MS, Brazil). All products were within shelf life when analyzed. Nineteen binary fish oil blends were prepared in the laboratory by carefully weighing the fish oils.

### $^1\text{H}$ NMR spectra acquisition and data processing

For  $^1\text{H}$  NMR analysis, 50 mg of oil was weighed and dissolved in deuterated chloroform (500  $\mu\text{L}$ ,  $\text{CDCl}_3$ ) with tetramethylsilane (1%, TMS) and transferred to 5 mm NMR tubes. Experiments were carried out without spinning at  $25^\circ\text{C}$  on a Bruker DPX300 7.05 T spectrometer (São Paulo, Brazil) equipped with a 5 mm dual probe operating at 300.1317 MHz for hydrogen. The  $^1\text{H}$  NMR spectra acquisition parameters were: data acquisition, 2.73 s; pulse length, 12.5  $\mu\text{s}$ ; relaxation delay, 1.0 s (repetition time, 3.73 s); data points, 32K; spectral width, 20 ppm; dummy scans, 2; scans, 64; pulse angle,  $30^\circ$ ; temperature,  $25^\circ\text{C}$ ; and frequency offset, 5.56 ppm.

Spectral processing was undertaken with a Topspin<sup>®</sup> 3.5 pl7 (Bruker BioSpin) platform. All spectra were submitted to apodization (0.1 Hz line broadening) and zero-filled to 64K data points before Fourier transformation. Automatic phase (optimizing the total intensity of the spectrum) and baseline corrections (fifth-order polynomial fitted on the noise regions of the spectrum) were performed with manual corrections applied when necessary. The chemical shifts were expressed in parts per million (ppm) relative to the TMS signal (0.000 ppm).

### Assignment and integration

Assignment of the  $^1\text{H}$  NMR spectra was performed based on the literature data [5]. The proposed assignments were confirmed by two-dimensional NMR and  $^{13}\text{C}$  NMR spectra. The peak areas were achieved by integration of the interesting signal using Topspin<sup>®</sup> 3.5 pl7 (Bruker BioSpin).

### DHA, EPA, and $\omega$ -3 PUFA quantification by $^1\text{H}$ NMR

Reference values for DHA, EPA, and  $\omega$ -3 PUFA were obtained from  $^1\text{H}$  NMR spectra. The relative proportion of DHA was calculated according to Equation 1:

$$DHA\% = \frac{I_1/2}{(I_H + I_1/2)} * 100$$

Equation 1:

where  $I_1$  (ca. 2.37 ppm) represents the signal areas from the DHA  $\alpha$ -methylene protons, which are upfield from the  $\alpha$ -methylene protons from other FA  $I_H$  (ca. 2.31 ppm) due to the inductively withdrawing nature and anisotropic effects from the carbonyl group and the C4 double bond.

The relative proportion of EPA was calculated according

to Equation 2:

$$EPA\% = \frac{I_F}{(I_F + I_E + \frac{I_I}{2})} * 100$$

Equation 2:

where  $I_F$  (ca. 1.75 ppm) represents the signal areas from the EPA  $\beta$ -methylene group,  $I_E$  (ca. 1.65 ppm) is the signal from other  $\beta$ -methylene groups, and  $I_I$  represents the signal areas from the DHA  $\alpha$ -methylene group.

The relative proportion of  $\omega$ -3 PUFA was calculated according to Equation 3:

$$\omega-3\text{ PUFA}\% = \frac{I_B}{(I_A + I_B)} * 100$$

Equation 3:

where  $I_B$  (0.95 ppm) represents the signal areas from the terminal methyl groups of  $\omega$ -3 PUFA and  $I_A$  (0.85 ppm) represents the signal areas from the terminal methyl groups of other FAs. The  $\omega$ -3 PUFA differs from other FA because of the anisotropic effect from the pi electrons in the double bond. The inductively withdrawing nature of the  $sp^2$  carbons moves the methyl proton of  $\omega$ -3 PUFA downfield by 0.10 ppm relative to those of non- $\omega$ -3 PUFA [5].

#### ATR-FTIR measurements

Following clean-up and background correction, the samples were analyzed as films. The supplement capsules were drilled, and one drop of the fish oil was directly placed in the ATR accessory. The ATR-FTIR spectra were obtained as an average of 16 scans in the region between 4000 and 600  $\text{cm}^{-1}$  with a spectral resolution of 4  $\text{cm}^{-1}$ . Measurements were performed on a PerkinElmer Spectrum 100 FTIR spectrometer equipped with a GeSe (germanium-selenium) ATR accessory using PerkinElmer Spectrum software version 10.03.07. All samples were tested in triplicate.

#### Partial least square regression models

The original spectral dataset with 33 samples (in triplicate) was examined for outlier detection using Mahalanobis distances and residual standard deviations. Samples with leveraged values exceeded three times the mean leverage value and/or mean residual variance were classed as outliers. Data analysis was performed using iToolbox algorithms in MATLAB software (MATLAB v7.0, The Math Works, MA, USA) downloaded from <http://www.models.kvl.dk/>. The samples

were sorted by DHA, EPA, and  $\omega$ -3 PUFA values and an independent test set was constructed, providing the following two data sets: calibration set (23 samples) and prediction set (10 samples). The calibration and validation set showed similar mean and standard deviation of concentration values. The performance of the final PLSR model was evaluated according to the root mean square error of cross-validation (RMSECV) and prediction (RMSEP). RMSECV was the parameter governing the variable selection for all tested methods. The RMSECV value was achieved by cross-validation using three segments (syst123). The RMSECV was calculated as follows with Equation 4:

$$\text{Equation 4: } RMSECV = \sqrt{\frac{\sum_{i=1}^{n_{cv}} (y_i - \hat{y}_i)^2}{n_{cv}}}$$

where  $y_i$  is the reference value achieved by  $^1\text{H}$  NMR,  $\hat{y}_i$  denotes the estimated values for sample  $i$  when the model was constructed with the sample removed, and  $n_{cv}$  is the number of samples in the cross-validation set. The complexity of PLSR models was determined by plotting the RMSECV against the number of latent variables and selecting the number with a minimum value of RMSECV.

The RMSEP was calculated for the test set as follows with Equation 5:

$$\text{Equation 5: } RMSEP = \sqrt{\frac{\sum_{i=1}^{n_v} (y_i - \hat{y}_i)^2}{n_v}}$$

where  $y_i$  is the reference value achieved by  $^1\text{H}$  NMR,  $\hat{y}_i$  denotes the estimated values from the PLSR model, and  $n_v$  is the number of samples in the test set. RMSEP estimates the prediction error used to reveal problems such as overfitting.

The coefficient of determination between the predicted and the measured value was calculated for the calibration, cross-validation, and test set as follows (Equation 6):

$$\text{Equation 6: } R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where  $\bar{y}$  is the mean of the reference measurement results for all samples in the set.

### 3. Results and Discussion

A representative ATR-FTIR spectrum is displayed in Figure 1A. The assignment was performed by comparing with the literature data [10].

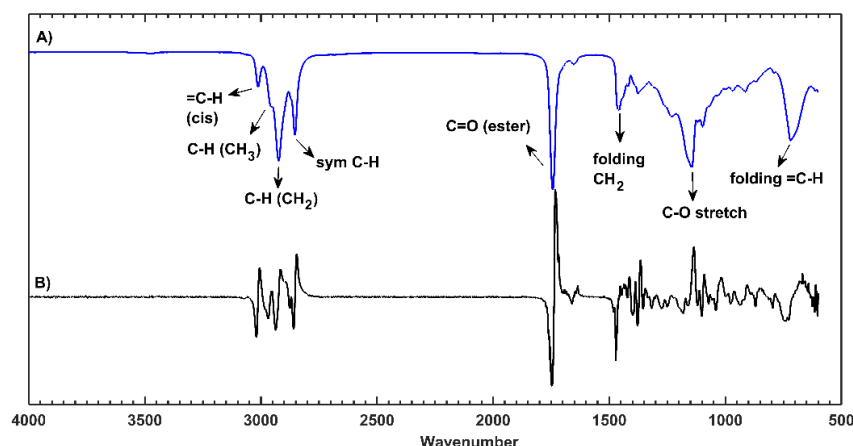


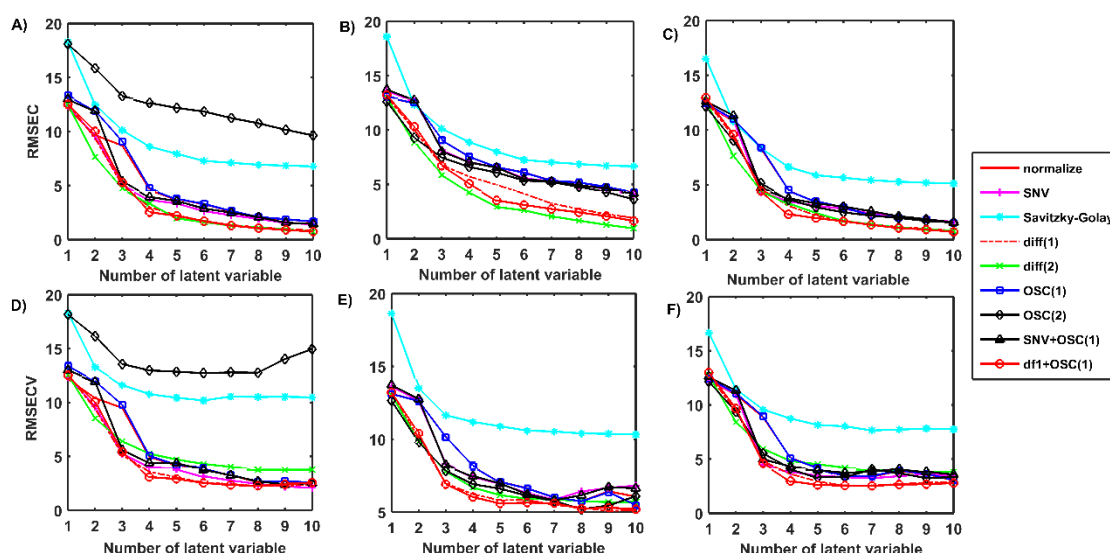
Fig. 1. Representative ATR-FTIR spectrum from fish oil A) before and B) after the first derivative and orthogonal signal correction. [Color figure online].

## Pre-treatment methods

The ATR-FTIR spectra can be affected by spectral noise, baseline drift, and other deviations. In these cases, spectral pre-treatment is mandatory to improve the accuracy of PLSR [11]. Pre-processing is an essential step in data analysis in which raw data is transformed into “clean data”, from which unwanted variations such as instrumental and experimental are removed. Choosing an optimal pre-processing method depends on many latent variables and the goal of data analysis. In this sense, several pre-treatment methods have been described in the literature. Standard normal variation (SNV) and multiplicative scatter correction (MSC) are frequently employed to correct the baseline and path length interferences or to reduce the scattering caused by small particles. Once a sample is analyzed as a film, controlling the path length is difficult or even impossible. These normalization methods effectively minimize or eliminate variations due to path length and baseline offsets, thus improving the linearity of the relationship between the constituents and the spectral values. SNV is preferable since it corrects each spectrum individually at each wavelength, i.e., it is a row-oriented correction method. Each spectrum is mean-centered and divided by its standard deviation so that the new spectra are centered at zero, and their standard deviations are one, with a common scale for all spectra. The

Savitzky–Golay filter is often employed for smoothing the data, increasing the signal-to-noise ratio and removing narrow spikes in the spectra. The Savitzky–Golay second derivative transformation with third-order polynomial was evaluated for smoothing and differentiation of the spectra to reduce random noise and enhance spectral resolution. First and second derivatives are routinely performed to extract relevant information from the spectra, resolve peak overlap, enhance the peak resolution, and eliminate the baseline drift. However, the derivative process increases the signal-to-noise ratio, making spectral interpretation difficult [12]. While orthogonal signal correction (OSC) is a spectral pre-processing technique that removes variation from a data matrix  $\mathbf{X}$  that is orthogonal (not correlated) to the matrix  $\mathbf{Y}$  response.

These pre-treatments and some combinations were evaluated to check the quality of the PLSR models. Full PLSR models were constructed, and cross-validation was employed to select the optimum pre-treatment method [13]. The cross-validation step is preferred for small datasets, as, in these situations, aggregation can improve model stability. However, it is worth highlighting that such results may be slightly less optimistic for more realistic scenarios that include a more expressive set of data. **Figure 2** shows the relationship between the latent variables and RMSECV values.



**Fig. 2.** Effect of pre-treatment on PLSR models: RMSEC from **A)** DHA, **B)** EPA, and **C)**  $\omega$ -3 PUFA and RMSECV from **D)** DHA, **E)** EPA, and **F)**  $\omega$ -3 PUFA. [Color figure online].

A systematic trend between the RMSEC and RMSECV values is expected in a robust model. The findings show that the combination of the first derivative and OSC with one factor provides the best results for the prediction of DHA, EPA, and  $\omega$ -3 PUFA. The derivative process resolves peak overlap, enhances peak resolution, and eliminates baseline drift (**Figure 1B**); but derivative processing increases the noise. The OSC allows removing the noise (eliminating the  $\mathbf{X}$  information not correlated to the  $\mathbf{Y}$  response). In this sense, OSC could be a perfect association with the derivative method.

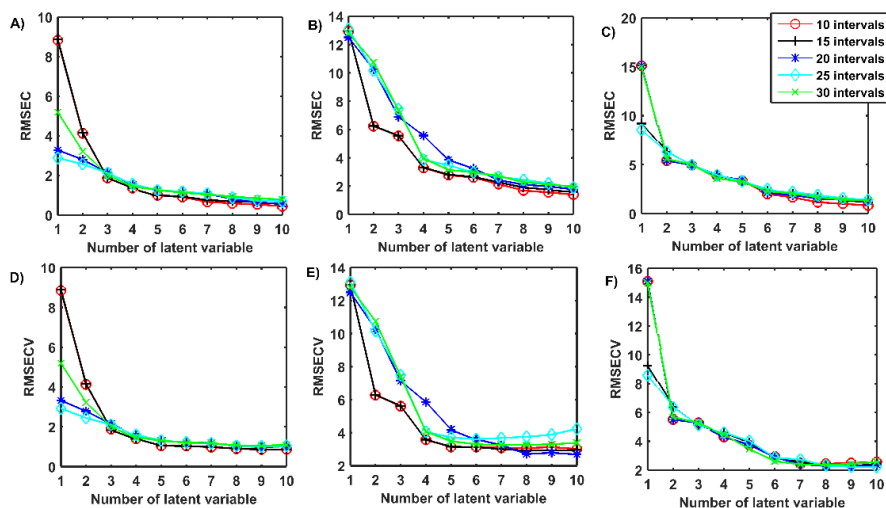
## Interval PLS

Variable selection methods based on intervals split the spectral data into a given number of intervals. Then, PLSR models were calculated in each interval once, giving superior prediction to the global (full-spectrum) model. The superiority

is caused by greater information carried by the selected interval correlated to the response matrix and by the exclusion of noise and interferences carried by other variables not correlated to the response. The best PLS model is chosen to take into account the validation parameter of RMSECV [14]. There is a minimal probability of hitting the optimal interval with the equidistant subdivisions, so an optimal interval might be found by carrying out small adjustments in the interval limits [14].

**Figure 3** shows the relation between the latent variables and RMSEC/RMSECV values achieved in each iPLS model. The ATR-FTIR spectra were divided into 10, 15, 20, 25, and 30 equidistant intervals. There is a significant effect on RMSECV due to the number of intervals, mainly for a low number of latent variables.



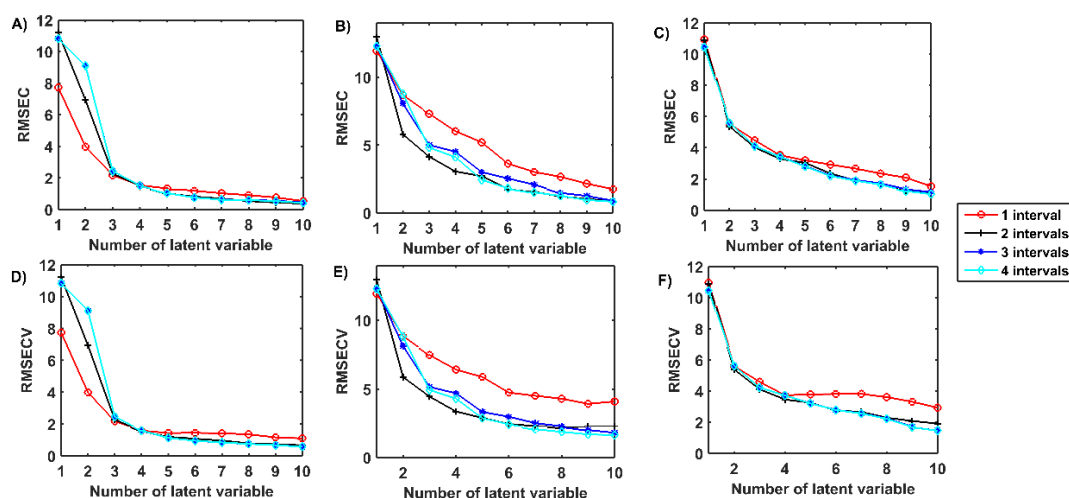


**Fig. 3.** Effect of the number of equidistant intervals on RMSEC from iPLS models of **A)** DHA, **B)** EPA, and **C)**  $\omega$ -3 PUFA and RMSECV of **D)** DHA, **E)** EPA, and **F)**  $\omega$ -3 PUFA. [Color figure online].

In general, 10 and 15 equidistant intervals provided better results for models with more than four latent variables, with fifteen equidistant intervals chosen for the following steps. The 1504–1279  $\text{cm}^{-1}$  region for DHA, the 1504–1279  $\text{cm}^{-1}$  region for EPA, and the 3092–2866  $\text{cm}^{-1}$  region for  $\omega$ -3 PUFA yielded the best results using the iPLS model.

#### Synergy interval PLS

The basic principle of the siPLS algorithm is similar to iPLS. First, the spectra are split into a given number of intervals. Next, PLSR regression models are performed for all possible combinations of two, three, or four intervals [15]. The subinterval combination improves the model performance, avoiding loose relevant information regions compared to iPLS. **Figure 4** shows the relationship between the number of latent variables and RMSEC/RMSECV to each siPLS model in different combinations.



**Fig. 4.** Effect of the number of combined intervals on siPLS models. RMSEC from **A)** DHA, **B)** EPA, and **C)**  $\omega$ -3 PUFA and RMSECV from **D)** DHA, **E)** EPA, and **F)**  $\omega$ -3 PUFA. [Color figure online].

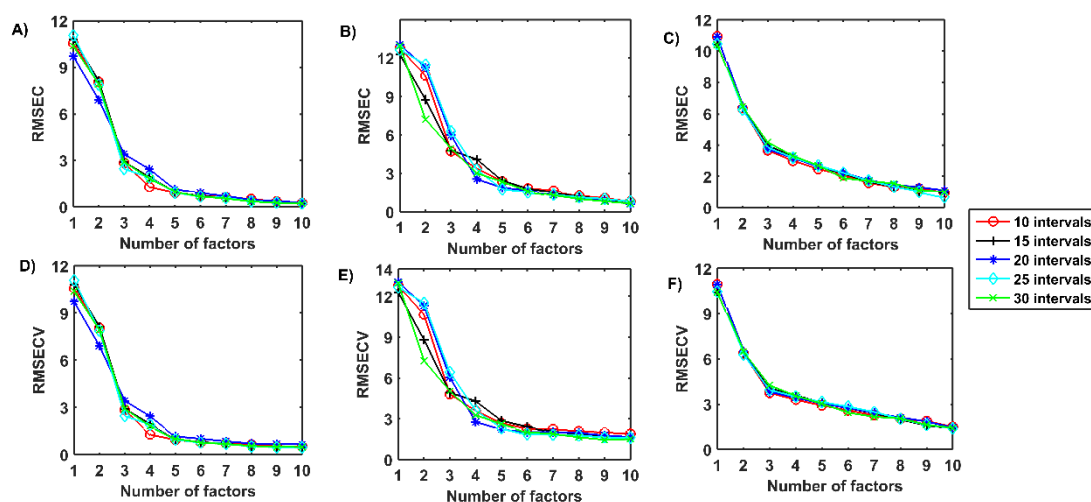
A systematic tendency between RMSEC and RMSECV values indicates a robust model. In general, two intervals show the best results, especially for EPA. The region between 3092–2866 and 1504–1279  $\text{cm}^{-1}$  for DHA, between 3092–2866 and 1730–1505  $\text{cm}^{-1}$  for EPA, and between 3092–2866 and 1957–1731  $\text{cm}^{-1}$  for  $\omega$ -3 PUFA provided the best results using the iPLS model.

#### Backward interval PLS

There are mainly two expansion methods for iPLS, namely forward iPLS (fiPLS) and backward iPLS (biPLS). The principle of the biPLS algorithm is to split the spectra into  $N$  equidistant intervals. Afterward, the PLS models are calculated to  $N-1$

intervals, leaving out one interval at a time. The first omitted interval gives the poorest-performing model with respect to RMSECV [16]. **Figure 5** shows the relationship between the number of latent variables and RMSEC and RMSECV values for different numbers of intervals in the biPLS model.

There is a regular trend between the RMSEC and RMSECV achieved with biPLS, which is to be expected for a robust model. In general, dividing the spectrum into ten intervals provides slightly better results. The regions between 4000–3661 and 1620–6441  $\text{cm}^{-1}$  for DHA; between 3660–3321, 2980–2641, and 2300–941  $\text{cm}^{-1}$  for EPA; and between 3660–3321, 2980–2301, and 1960–941  $\text{cm}^{-1}$  for  $\omega$ -3 PUFA give the best results.



**Fig. 5.** Effect of the number of intervals on biPLS models. RMSEC from **A)** DHA, **B)** EPA, and **C)**  $\omega$ -3 PUFA and RMSECV from **D)** DHA, **E)** EPA, and **F)**  $\omega$ -3 PUFA. [Color figure online].

### Prediction results

**Table 1** presents the number of latent variables and values of RMSEC, RMSECV, RMSEP and  $R^2$  for the best PLSR, iPLS, siPLS, and biPLS models. The data show that the pre-treatment method has a significant effect on the calibration

(RMSEC) and cross-validation (RMSECV) root mean square, but a smaller influence on the prediction (RMSEP), as often observed. The first derivative with the OSC for one factor was the pre-processing used in the following steps, as it provided more promising results.

**Table 1.** PLSR results based on RMSEC, RMSECV, RMSEP,  $R^2$ , number of latent variables, and the selected intervals used in full-spectrum PLSR, iPLS, siPLS, and biPLS models for the prediction of DHA, EPA, and  $\omega$ -3 PUFA.

Model	VL	$R^2_{\text{calibration}}$	$R^2_{\text{cv}}$	$R^2_{\text{predict}}$	RMSEC	RMSECV	RMSEP
<b>DHA</b>							
Raw data	8	0.998	0.997	0.990	1.043	1.188	2.087
Normalize	8	0.998	0.997	0.991	0.950	1.149	2.079
diff(1)OSC(1)	8	1.000	0.998	0.985	0.461	0.967	2.200
iPLS	8	0.999	0.998	0.987	0.697	0.904	2.097
siPLS	9	1.000	0.999	0.989	0.469	0.625	1.941
biPLS	8	0.999	0.999	0.989	0.510	0.633	1.945
<b>EPA</b>							
Raw data	9	0.993	0.984	0.976	2.128	3.300	3.699
Normalize	9	0.994	0.986	0.981	1.996	3.126	3.280
diff(1)OSC(1)	9	0.999	0.993	0.977	0.791	2.234	3.488
iPLS	8	0.995	0.987	0.985	1.899	2.914	3.752
siPLS	8	0.997	0.991	0.962	1.484	2.425	2.595
biPLS	8	0.998	0.994	0.979	1.059	1.959	3.456
<b><math>\omega</math>-3 PUFA</b>							
Raw data	8	0.996	0.991	0.964	1.785	2.723	4.504
Normalize	9	0.998	0.994	0.968	1.378	2.256	4.227
diff(1)OSC(1)	8	0.999	0.987	0.968	1.022	3.239	4.067
iPLS	8	0.996	0.993	0.980	1.768	2.378	3.460
siPLS	8	0.996	0.993	0.978	1.859	2.384	3.520
biPLS	9	0.998	0.996	0.963	1.215	1.890	4.463

**Legend:** LV, number of latent variables; diff(1)OSC(1), first derivative with orthogonal signal correction for one factor.

Overall, the full PLS model provides the worst prediction results due to noisy spectral information that inevitably decreases the model performance. iPLS provides an overview of spectral data to select the informative spectral region and remove noisy regions, but only one interval is selected, and useful spectral information is lost. siPLS overcomes these disadvantages by combining two, three, or four intervals to avoid losing much helpful information in the calibrating model. The siPLS model with two intervals gives good correlation coefficients and the best results for predicting DHA (RMSEC=0.469, RMSECV=0.625 and RMSEP=1.941), EPA (RMSEC=1.484, RMSECV=2.425, and RMSEP=2.595), and  $\omega$ -3 PUFA (RMSEC=1.902, RMSECV=2.667, and RMSEP=3.980).

The relative errors and *F*-values were further employed to determine the siPLS model's predictive ability and verify the reliability of ATR-FTIR for  $\omega$ -3 analysis. The results achieved by the proposed method were very similar to those achieved by  $^1\text{H}$  NMR. The *t*-values for DHA, EPA, and  $\omega$ -3 PUFA were equal to 0.0165, -0.146, and -0.0366, respectively, in the two-sample equal-variance *t*-test. According to the *t*-value distribution table, for a significance level of  $\alpha=0.05$  and a number of samples  $n=99$  (train + test sets), the *t*-critic value was 1.652, showing that there was no significant difference between the reference and prediction values (**Table S1**). The precision between the methods was also similar. The *F*-values for DHA, EPA, and  $\omega$ -3 PUFA were equal to 0.472, 0.440, and 0.457, respectively. According to the *F*-value distribution table, for a significance level  $\alpha=0.05$  and a number of samples  $n=99$  (train + test sets), the *F*-critic was 0.716, showing an equal precision between the reference and the proposed method.

The results achieved by ATR-FTIR demonstrate the capability for determining DHA, EPA, and  $\omega$ -3 PUFA in fish oil supplement capsules. The present method is robust and quantifies selected compounds in both triacylglyceride and ethyl ester forms. However, as anticipated, less optimistic results may be evidenced in more realistic scenarios with a more extensive sample set.

### Spectral interpretation

One of the main benefits of the interval PLSR method is that it allows for objective variable selection even without spectral knowledge. However, the selected intervals could be related to spectral features from a given compound, providing more information regarding the spectral profile. Nevertheless, since FTIR is not a selective spectroscopic method, it is impossible to know if the PLSR models are based on specific signals from DHA and EPA or other compounds correlated with the quantified compound. The two intervals that provide the best prediction results in the siPLS models are presented in **Table S2** and **Figure S1 (Supplementary Material)**.

The 3092–2866  $\text{cm}^{-1}$  region was important for determining all interesting compounds. The region contains signals related to vibrational stretches of unsaturated carbons (3009  $\text{cm}^{-1}$ , =C–H). The position of the double bond system relative to the carbonyl group ( $\Delta$ -position) and the methyl end ( $\omega$ -position) in PUFA has importance for the IR spectra. In this sense, it is possible to quantify  $\omega$ -3 PUFA in the presence of other unsaturated FAs. Although EPA and DHA are both  $\omega$ -3 PUFA, they have the first bond in the  $\Delta 5$  and  $\Delta 4$  positions, respectively. This difference allows for PLSR models to quantify DHA and EPA.

The inclusion of a second interval improved the prediction performance. The second region included in the siPLS model for DHA was the 1504–1279  $\text{cm}^{-1}$  region containing the signal

related to  $\text{CH}_2$  folding (1462  $\text{cm}^{-1}$ ). The methylene-interrupted double bonds in PUFA give rise to a different absorption pattern than the isolated double bonds [17], which can contribute to DHA determination. The second region included in the siPLS model for EPA prediction was the 1730–1505  $\text{cm}^{-1}$  range with the signal related to C=C *cis* stretching (1657  $\text{cm}^{-1}$ ). The second region included in the siPLS model for  $\omega$ -3 PUFA prediction was the 1957–1731  $\text{cm}^{-1}$  region covering the signals related to the carbonyl stretch of triacylglycerol and ethyl ester (1743 and 1734  $\text{cm}^{-1}$ , respectively). The proximity of the double bond to the carbonyl in PUFA can influence the carbonyl signal.

## 4. Conclusions

This should clearly state the main conclusions of the research and give a clear explanation of their importance and relevance. This study demonstrated a fast, simple, and reliable method for quantification of DHA, EPA, and  $\omega$ -3 PUFA in fish oil capsules. ATR-FTIR and PLSR methods have advantages over traditional methods (for example, chromatography), such as a smaller quantity of samples, the analyses can be carried out directly on the oil samples (without preparation), and they are faster and cheaper. The samples analyzed covered a wide range of DHA, EPA, and  $\omega$ -3 PUFA concentrations and included fish oil supplements in triglyceride or ethyl ester forms. Comparing PLSR, iPLS, siPLS, and biPLS algorithms, siPLS showed the best results, providing models with better prediction parameters. The findings achieved by ATR-FTIR and siPLS proved statistically accurate and precise compared to those obtained by  $^1\text{H}$  NMR. The proposed method has the potential to be used in both quality control laboratories and industrial processes of fish oil supplements.

## Supporting Information

Supplementary material is available.

## Acknowledgments

The authors thank Alessandra Ramos Lima and Juliete Rocha de Lima for training and follow-up during the acquisition of ATR-FTIR spectra. The authors also thank the Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul (FUNDECT) (grant numbers TO 0165/12 and 59/300.490/2016) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (grant numbers 481507/2013, 304600/2014-8, and 309636/2017-5). Tainara A. Nascimento was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The authors thank the "Laboratório de Sistemas Embarcados" of the Federal University of Mato Grosso do Sul, for the availability of the computer lab and software Matlab to perform the chemometrics analysis.

## Author Contributions

TIBL worked on conceptualization, methodology, validation, formal analysis, investigation, resources, data curation, writing – original drafting, writing – review & editing, visualization. TAN worked on methodology, formal analysis, writing – review & editing. ESP worked on methodology, validation, formal analysis. SLO worked on resources, writing

– review & editing, funding acquisition. DG writing – review & editing. GBA worked on conceptualization, methodology, resources, writing – original drafting, writing – review & editing, visualization, supervision, project administration, funding acquisition.

## References and Notes

- [1] Li, X. Bi, X.; Wang, S.; Zhang, Z.; Li, F.; Zhao, A. Z. *Front. Immunol.* **2019**, *10*, 2241. [\[Crossref\]](#)
- [2] Stanley, E. G.; Jenkins, B. J.; Walker, C. G.; Koulman, A.; Browning, L.; West, A. L.; Calder, P. C.; Jebb, S. A.; Griffin, J. L. *J. Proteome Res.* **2017**, *16*, 3168. [\[Crossref\]](#)
- [3] Zhang, Z.; Liu, F.; Ma, X.; Huang, H.; Wang, Y. *J. Agric. Food Chem.* **2018**, *66*, 218. [\[Crossref\]](#)
- [4] Davis, J.; EPA/DHA (Omega 3) Ingredients Market Size worth USD 3.79 Billion by 2022: Global Market Insights, Inc. [Internet document]. Global Market Insights, Inc. URL <https://globenewswire.com/news-release/2016/03/30/823963/0/en/EPA-DHA-Omega-3-Ingredients-Market-Size-worth-USD-3-79-Billion-by-2022-Global-Market-Insights-Inc.html> 2016, accessed on November 19, 2023.
- [5] Bratu, A.; Mihalache, M.; Hanganu, A.; Chira, N. A.; Todasca, M. C. *UPB Sci. Bull.* **2013**, *75*, 23.
- [6] Shapaval, V.; Afseth, N. K.; Vogt, G.; Kohler, A. *Microb. Cell Factories* **2014**, *13*, 86. [\[Crossref\]](#)
- [7] Lopes, T. I. B.; Pereira, E. S.; Freitas, D. S.; Oliveira, S. L.; Alcantara, G. B.; *J. Food Sci. Technol.* **2019**, *2*, 1. [\[Crossref\]](#)
- [8] Cebi, N.; Bekiroglu, H.; Erarslan, A.; Rodriguez-Saona, L. *Molecules* **2023**, *28*, 3727. [\[Crossref\]](#)
- [9] Yamashita, G. H.; Anzanello, M. J.; Soares, F.; Rocha, M. K.; Fogliatto, F. S. *Chemom. Intell. Lab. Syst.* **2022**, *231*, 104689. [\[Crossref\]](#)
- [10] Plans, M.; Wenstrup, M. J.; Rodriguez-Saona, L. E. *J. Am. Oil Chem. Soc.* **2015**, *92*, 957. [\[Crossref\]](#)
- [11] Miao, X.; Miao, Y.; Liu, Y.; Tao, S.; Zheng, H.; Wang, J. et al. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2023**, *284*, 121733. [\[Crossref\]](#)
- [12] Varmuza, K.; Filzmoser, P. Introduction to multivariate statistical analysis in chemometrics. 1st ed.; CRC Press: Boca Raton, USA, 2009.
- [13] Westad, F.; Marini, F. *Anal. Chim. Acta* **2015**, *893*, 14. [\[Crossref\]](#)
- [14] Nørgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B.; *Appl. Spectrosc.* **2000**, *54*, 413.
- [15] Leardi, R.; Nørgaard, L. *J. Chemom.* **2004**, *18*, 486. [\[Crossref\]](#)
- [16] Balabin, R. M.; Smirnov, S. V. *Anal. Chim. Acta* **2011**, *692*, 63. [\[Crossref\]](#)
- [17] Lundbom, J.; Kolehmainen, M.; Pulkkinen, L.; Soininen, P.; Tiainen, M.; Schwab, U.; Uusitupa, M.; Tammi, M. *Eur. J. Lipid Sci. Technol.* **2010**, *112*, 1308. [\[Crossref\]](#)

## How to cite this article

Lopes, T. I. B.; Nascimento, T. A.; Pereira, E. S.; de Oliveira, S. L.; Galvan, D.; Alcantara, G. B. *Orbital: Electron. J. Chem.* **2024**, *16*, 79. DOI: <http://dx.doi.org/10.17807/orbital.v16i2.19869>