CrossMark

ORIGINAL PAPER

Assessment of Interval PLS (*i*PLS) Calibration for the Determination of Peroxide Value in Edible Oils

Xianghe Meng[1] · Qin Ye[1] · Xiaohua Nie[1] · Qiuyue Pan[2] · Lianzhou Jiang[3]

Abstract Partial least squares (PLS) is a commonly used (and sometimes misused) chemometric technique for calibrating Fourier transform infrared spectroscopy, and allows the analysis of a variety of quality parameters associated with edible oils. Peroxide value (PV) is a typical parameter of interest; however, developing a robust, optimal, and reliable calibration method can be a daunting task. This paper examines and compares the use of interval PLS as a tool to develop a PLS PV calibration method for a single-bounce attenuated total reflectance accessory relative to full spectrum PLS and experienced PLS, making use of correlation, variance, and pure component spectra. Using mixtures of fresh and oxidized oil covering a PV range of 1–20 meq/kg, backward interval PLS could systematically produce quality calibrations without the need to resort to experienced PLS. The experienced PLS requires a degree of spectral knowledge as well as diligent and tedious spectral examination, including largely unstructured iterative calibrations and cross-validations to improve calibration performance. The backward interval PLS is also better than the full spectrum PLS in terms of model performance. In addition, the general model developed could account for the errors caused by oil types.

Introduction

The oxidative state of edible oils is an important quality variable as they are susceptible to oxidation during processing, storage, and over time. The peroxide value (PV) is commonly used as a measure of the initial primary oxidation products formed, which in turn deteriorate into secondary oxidative products such as ketones, aldehydes, free fatty acids (FFA), and alcohols; these compounds can cause off-flavors, as well as affecting the color and texture, and can lead to a decrease in the nutritional value of the oil [1]. As an important and practical quality indicator of the oxidative status of edible oils, PV determination is a common analytical method sanctioned by the key international organizations in this field [e.g., American Oil Chemists' Society (AOCS), Association of Official Analytical Chemists, and International Standards Organization]. PV is determined by measuring the amount of iodine that forms due to the reaction between peroxides (formed in fat or oil) and iodide ions under acidic conditions. Acetic acid neutralizes the base produced and prevents the production of hypoiodite, which would interfere with the determination of iodine during titration with sodium thiosulfate, with soluble starch as an endpoint indicator.

Although methodologically simple, manual titrimetric procedures tend to be time consuming, reagent intensive, and environmentally problematic [2], and spectroscopic methods are a preferred alternative due to the speed of analysis, potential for automation, and reduction of reagents used. Due to these potential advantages, infrared spectroscopy has emerged as a quantitative analytical tool

✉ Xianghe Meng
  mengxh@zjut.edu.cn

✉ Lianzhou Jiang

1 Ocean College, Zhejiang University of Technology,
   No. 18, Chaowang Road, Xiacheng District,
   310018 Hangzhou, Zhejiang Province, China

2 College of Food Science and Nutritional Engineering, China
   Agricultural University, 100083 Beijing, China

3 School of Food Science, North East Agricultural University,
   150030 Harbin, China

for a variety of lipid quality parameters including FFA, iodine value, saponification number, *trans* content, and PV. Mid-Fourier transform infrared (FTIR) spectroscopy is often the method of choice due to its ability to provide detailed spectral functional group information related to the makeup of lipids, typically CH, OH, OOH, *cis*, *trans*, and ester bonds, etc. Thus, based on variations related to these primary bands, one can establish quantitative relationships between absorbance changes and the parameter of interest (e.g., ester linkage with saponification number) and relate these changes to a well-established primary analytical method. As noted earlier, spectroscopists with an appreciation of which spectral regions are relevant to the problem at hand can work toward developing a method based on this information using a structured experiential approach. Thus, a key strength of mid-infrared (MIR) spectroscopy is that it provides direct guidance as to where to look for spectral changes that are expected to be related to the measurement of interest. Although the MIR spectrum provides guidance to the experienced spectroscopist as to where to measure, it is rare that a simple direct absorption measurement will allow the accurate development or determination of a particular characteristic. PV determination is a classic example of this, as the –OOH absorption is confounded by other OH absorptions (e.g., alcohols and FFA), as well as the variability induced by hydrogen bonding, making partial least squares (PLS) a necessity to select the relevant data. van de Voort *et al*. [3]. developed an experiential PLS-based mid-FTIR method for the quantitative determination of the PV of vegetable oils using *tert*-butyl hydroperoxide as a model and the 3750–3150 cm^{-1} spectral region, which can be used to predict the PV of oils with a coefficient of variation (CV) of ~±5 %. Subsequently, an alternative method based on the stoichiometric reaction of triphenylphosphine with hydroperoxides to produce a spectrally distinct and strong IR signal was developed, thus negating the need for PLS and facilitating simple peak height and area measurement [4]. Although this approach works well where it is applicable, depending on the complexity of the matrices encountered, PLS can still be an important adjunct to further enhance the accuracy and universality of the calibration. This has been demonstrated in the analysis of lubricants for acid and base numbers, where a stoichiometric reaction provides a product with a strong and distinct IR signal, but the calibration is enhanced by PLS to account for matrix variability. It would clearly be of great benefit if the development of a PLS calibration could be simplified to the point of allowing automation, and with a reasonable guarantee that it is likely to produce the best possible calibration. This is the potential promise of the software produced by Nørgaard *et al*. though interval PLS (*i*PLS) [5].

This paper examines and compares the results of full PLS models to that of experiential PLS and *i*PLS

[including backward *i*PLS (*Bi*PLS)] for the determination of the PV of edible oils to determine if high quality calibrations can be achieved using *i*PLS techniques relative to the AOCS standard method. In addition, the effect of the oil type on the performance of the models is also explored.

Materials and Methods

Oil Samples

Refined soybean oil ($n = 12$), peanut oil ($n = 10$), rapeseed oil ($n = 13$), and sunflower seed oil ($n = 10$) were purchased from local suppliers, and used frying oil (UFO) that had been used for frying chicken was obtained from food vendors near the Zhejiang University of Technology(Hangzhou, China). Before being mixed with refined oil, simple refinement procedures such as filtration and deodorization were carried out on the UFO. Three hundred and thirty-five binary blends were prepared by mixing normal refined edible oil with the UFO in the range of 0.1–100 % (w/w). To minimize oxidation and hydrolysis, all samples were stored at 4 °C in amber glass bottles after preparation.

Spectra Acquisition and Pretreatment

Infrared spectra were recorded from 4000 to 400 cm^{-1} using a Bruker Tensor 27 FTIR spectrometer (Billerica, MA, USA) equipped with a single-bounce attenuated total reflectance (ATR) accessory. All spectra were recorded at 4 cm^{-1} with 32 scans (~2 min of scanning). Prior to and after loading the sample on the ATR crystal, the crystal was wiped clean with a non-abrasive tissue soaked in *n*-hexane and allowed to dry. Each sample was independently scanned in triplicate, from which a mean spectrum was then calculated and used for spectral analysis and compared to the PV results for the oil blends obtained using the AOCS method. The calibration methods were developed by PLS regression and *Bi*PLS regression using TQ Analyst software from Nicolet (Madison, WI, USA). By appropriate pretreatment of the infrared spectra, a variety of non-target factors on the spectra could be weakened and even eliminated, and the spectra information could then be enhanced. It is also helpful to lay a certain foundation for the establishment of the calibration model and the forecast of the unknown composition or the nature of the samples. The pretreatment methods used in this study are as follows: standard normalization of vector (SNV), multiplicative scatter, first derivative (FD), second derivative (SD), and Savitzky–Golay smoothing.

Outlier Detection

The original spectral dataset of 335 samples was initially examined by applying classical diagnostics for multivariate outlier detection in the PLS models based on the identification of samples that had excessive Mahalanobis distances. When these leveraged values exceeded three times the mean leverage value in the calibration set, they were classed as outliers and removed. A total of five samples out of 335 were classed as outliers. Samples presenting residual standard deviations twice the mean residual variance in the calibration set were also flagged as abnormal observations [6].

Construction of Models

The basic approaches used to develop the global PV models included classic full spectrum PLS regression models and models based on experiential region selection PLS guided by correlation and variance spectra; these were then compared to *i*PLS and *Bi*PLS. First, the full spectra for the classic PLS and related spectra regions for experienced PLS were selected separately for model construction with PV as a response variable using Turboquant (Thermo Electron Corp., Waltham, MA, USA). Then, *Bi*PLS was applied to automatically select suitable spectral regions to determine whether similar or improved regression models for the prediction of PV could be obtained. In addition, individual submodels based on each oil type determined by *Bi*PLS were also developed. The relative quality of the results of the models were compared using the root mean square of the cross-validation (RMSECV) of the calibration set as well as the root mean square of prediction (RMSEP) in the validation set. Wavelength region selection was carried out by experienced PLS and *Bi*PLS. Experienced PLS is similar to *i*PLS, i.e., the dataset is split into a number of intervals and the PLS models are calculated for each interval and present the RMSECV for each interval. What differs in *i*PLS is that the spectrum is split not into equidistant intervals but by spectral assignments based on experiences. The *Bi*PLS algorithm used here was developed by Nørgaard, Saudland, Wagner, Nielsen, Munck, and Engelsen [4]. The principle of this algorithm is to split the spectra into a given number of equidistant intervals; afterwards, the PLS models are calculated with each interval left out, i.e., if one chooses 20 intervals, then each model is based on 19 intervals, leaving out one interval at a time. The first interval left out is the one that gives the best performing model with respect to RMSECV when left out. This procedure is continued until one interval remains. The regions with the lowest RMSECV are chosen.

Altogether, 335 samples were analyzed, including 12 soybean, ten peanut, 13 rapeseed, ten sunflower, one UFO, and 290 binary blends of the four refined oils with UFO. Around three-quarters of the samples were used to build the calibration, and the remaining quarter were used for validation. Each type of oil was also modeled separately, and all the oil samples, including the soy, rapeseed, sunflower, and corn oil, were simulated as a general model to evaluate the robustness of the models developed.

External Validation

To further validate the effectiveness of the *Bi*PLS model developed for PV determination in practical samples, nine types of UFO from different sources were collected as the external validation set and each sample was analyzed in triplicate; the results are expressed as the means ± standard deviations.

Statistical Analysis

The standard PV of the samples were determined by the official AOCS methods Cd 8–53. Spectra processing, classic PLS, and experienced PLS regression data were analyzed using TQ Analyst, and *Bi*PLS regression was performed using iToolbox, which was developed by Nørgaard *et al*. [4].

Results and Discussion

Chemical Data and Spectral Information

The sample set contained 335 oils including refined UFO, soybean oil, rapeseed oil, peanut oil, sunflower oil and their blends with UFO in proportions ranging from 0.10 to 100 %. The reference PV ranged from 1.33 to 19.62 meq/kg, and the descriptive statistics (mean, standard deviation, and range) for the PV of the 335 oil samples are shown in Table 1.

The spectra of all the oils had the most intense absorption bands between 3100 and 2800 cm^{-1}, including a small band at ~3008 cm^{-1} due to trans CH stretching, a max at 2954 cm^{-1} due to CH$_3$ asymmetrical stretching, and bands at 2922 and 2853 cm^{-1} attributed to symmetrical and asymmetrical stretching of CH$_2$ [7], respectively. The strong peak at ~1745 cm^{-1} is attributed to the C=O stretching vibration of the triglycerides, the weak peak near 1654 cm^{-1} is attributed to the stretching vibration of the C=C group of *cis*-olefins, 1400–1300 cm^{-1} is due to the blending vibrations of aliphatic CH$_2$ and CH$_3$, while 1300–1000 cm^{-1} is attributable to the stretching vibration of C–O ester groups and CH$_2$ wag. Below 1000 cm^{-1}, the band near 723 cm^{-1} is attributable to the overlapping of the (CH$_2$)$_n$ rocking vibration and the out-of-plane vibration of

Table 1 Chemical characterization of 335 oil samples used for building PLS models

Parameters	Oil type	Number of Samples			Mean	Min	Max
		Calibration Set	Validation set	Total			
PV(meq/kg)	Soybean oil	62	21	83	4.54	1.65	12.97
	Rapeseed oil	62	23	85	6.60	3.05	14.54
	Sunflower oil	62	20	82	11.04	8.27	19.29
	Peanut oil	62	23	85	7.48	4.24	18.48
	All oil samples	248	87	335	9.25	1.33	19.62
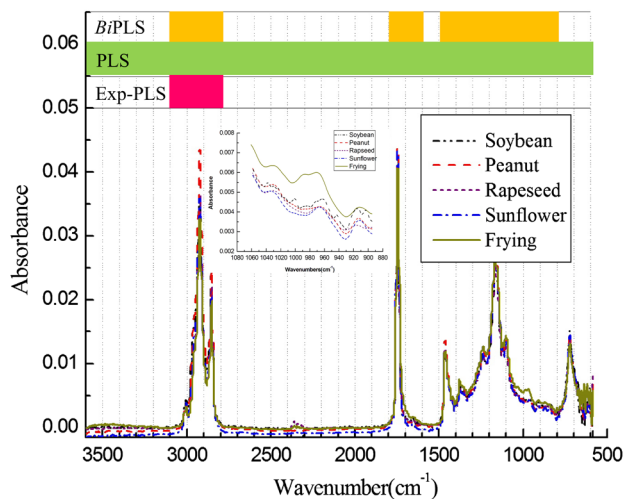
Fig. 1 Typical FTIR spectra of soy, peanut, rapeseed, sunflower and the used frying oils and spectral regions selected for PV model construction by *Bi*PLS

Table 2 Cross-validation errors of the PLS full spectra model with or without data pre-processing to model the PV of 335 oil samples

Mathematical pretreatment	LV	PV	
		R^2	RMSECV
Original spectra	12	0.82429	8.50
MSC	11	0.83768	4.05
Standard normalization of vector, SNV	10	0.9724	1.14
First derivative, FD	10	0.9688	1.14
Second derivative, SD	6	0.7928	3.05
SNV+SD	10	0.8982	2.70
SNV+FD	8	0.9521	1.76
SNV+smoothing 7	6	0.9070	1.98
SNV+smoothing 9	6	0.9175	1.84
SNV+Smoothing 11	9	0.9555	1.51

cis-disubstituted olefins. Different oils have slight differences in the peak position and band absorbance due to the differences in their triglyceride composition [8].

Exposure of edible oils to high temperatures, light and oxygen environments, and oxidative reactions may result in the loss of the ester bond and modification of the double bonds, as well as the formation of primary and second oxidation products and *trans* fat. Spectrally, bands involved in oxidation include –OOH (3444 cm^{-1}), HC= (3008 cm^{-1}), C=C (1654 cm^{-1}), C–O (1000–1300 cm^{-1}), and *trans* HC= (968 cm^{-1}), but are not limited to these bands; of these, –OOH is thought to be the more obvious determinant region in terms of PV. The inset in Fig. 1 highlights the C–O band at 1030 cm^{-1} and trans HC= (968 cm^{-1}) of the UFO, and visually differentiates the oxidized oil from the un-oxidized refined edible oils.

Spectral Pre-Processing and Outlier Elimination

Table 2 shows the performance of the different PLS regression models developed (without variable selection) using the MIR spectra of 335 oil samples and the respective reference values determined for the PV, corresponding

to the different pre-processing methods tested for signal correction.

In all cases, a high number of PLS components [ten latent variables (LV)] were needed to develop calibration models capable of tackling the interference and redundancy of the MIR signals with any degree of predictive accuracy. From the results, most of the pre-processing methods were superior to the raw data in terms of improving model reliability. The best model performance results (R^2 0.9724, RMSECV 1.14) were obtained when SNV pre-treated spectra were used for PLS model construction. The FD was fairly similar to those from studies by Yildiz *et al*. [9] and Pizarro *et al*. [10], who found that FD pretreatment was one of the more useful preprocessing methods in the development of near-infrared spectroscopy and Fourier transform near-infrared spectroscopy (FT-NIR) models for PV determination; any differences are likely due to the quality of the raw spectra and/or the mode of spectral acquisition. Combining SNV with smoothing, FD, or SD did not further improve model performance, and thus, only SNV pretreated spectra were used to optimize all subsequent models.

To exclude the Y-outliers (Abs), which are spectroscopic measurements representing uncontrolled error, the data was examined for outliers using the Chauvenet test. This test uses the spectral and PV information for each component

Table 3 Results of conventional PLS using full spectra, experienced PLS using the concerned spectral regions as well as *Bi*PLS using optimized regions for PV determination using standard normalization of vector data

PLS method	Spectra region (cm^{-1})	LV	Calibration set				Validation set	
			N.S.	R^2	RMSEC	RMSECV	N.S	RMSEP
Conventional PLS	3800–600 Full spectra	10	258	0.9847	0.690	1.02	72	0.943
Experienced PLS	3800–3100 –OO–H	3	258	0.8048	2.35	2.50	72	2.77
	3100–2800 H–C=, CH$_2$, CH$_3$	7	258	0.9756	0.868	1.01	72	1.04
	2800–2700	6	258	0.9163	1.58	1.68	72	1.91
	2700–2000	6	258	0.9141	1.60	1.74	72	2.05
	2000–1800	7	258	0.9202	1.55	1.73	72	1.96
	1800–1600 C=C, C=O	7	258	0.9405	1.34	1.50	72	1.52
	1600–1500	7	258	0.8993	1.73	1.91	72	2.21
	1500–1400	8	258	0.9408	1.34	1.48	72	1.43
	1400–1300	6	258	0.9405	1.34	1.44	72	1.52
	1300–1000 C–OO	6	258	0.9563	1.16	1.26	72	1.20
	1000–900 trans H–C=	3	258	0.9486	1.25	1.43	72	1.50
	900–800	5	258	0.8930	1.78	1.85	72	2.19
	800–700	9	258	0.9389	1.36	1.53	72	1.78
	700–600	5	258	0.8738	1.92	2.05	72	2.82
*Bi*PLS	3100–2800, 1800–1600, 1500–800	10	258	0.9886	0.605	0.86	72	0.713

NS numbers of samples

to determine Mahalanobis distance (H statistic), which was calculated from the principal component analysis scores. Leverage and studentized residual analysis were also used to assess the differences between individual sample spectra and the average spectrum of the set. On the basis of these analyses (not shown), of the 335 samples, five samples were flagged and removed. After eliminating these outliers, the RMSECV for PV decreased from 1.14 to 1.02, while R^2 increased from 0.9724 to 0.9847 (see Table 3).

Conventional and Experienced FA Model

Initially, regardless of oil type, all the samples were used to build general models for PV determination using conventional full spectrum FA as well as by experienced FA using specific spectrum regions; the results are compared in Table 3. The full spectrum produced a reasonable calibration model (RMSECV 1.02, RESEP 0.943), with a 0.9847 coefficient of determination using ten LV. In an attempt to further improve this full spectrum model, the spectrum was divided into eight subintervals based on the main spectral assignments related to changes expected as a result of oxidation. Models based on specific isolated spectral regions reduced the number of LV to 3–7; however, most models resulted in high RMSEP values. The best experienced FA model gave a performance similar to the full spectrum FA model (RMSECV 1.01, RMSEP 1.04, R^2 0.9756). Although experienced FA simplified the PV model, with only 1/10 of the spectrum required for model construction,

there was no significant improvement in terms of model performance relative to the classic FA, indicating that spectral knowledge is helpful in developing quality calibrations. It is noteworthy that the 3800–3100 cm^{-1} region did not perform well in the experienced FA model (R^2 0.8048, RMSECV 2.5) relative to the findings of van de Voort *et al*. [3]. The FA model for PV determination developed by van de Voort *et al*. was based on the band at 3750–3150 cm^{-1}, which reflects the absorption of hydroperoxide; this model was found to be better than the chemical method in terms of the CV. However, their model was built with a transmission spectrum, which is more sensitive than the ATR spectra used in this study; it is likely that the hydroperoxide absorptions obtained by ATR account for the poor performance of the OH region.

*Bi*PLS Models

General Models

To select the spectral intervals by *Bi*PLS, the full spectrum (excluding the 2700–2000 and 4000–3100 cm^{-1} bands) was divided into 18 equidistant subintervals, with 100 cm^{-1} per interval. PLS models were calculated with a different interval left out each time, the first interval deleted by the process was interval 18 (3100–3000 cm^{-1}), which produced the largest improvement in terms of RMSECV; this result indicates that the importance of this interval is minimal. This iterative procedure was continued until only

Table 4 Results of *Bi*PLS models for PV determination using SNV data

Interval number used in model	Deleted interval no.	Calibration set					Validation set	
		N.S.	LV	R^2	RMSEC	RMSECV	N.S.	RMSEP
18	–	258	7	0.9847	0.69	1.02	72	0.943
17	18	258	10	0.9775	0.660	0.95	72	0.977
16	17	258	10	0.9887	0.659	0.889	72	0.753
15	5	258	10	0.9890	0.652	0.857	72	0.751
14	6	258	10	0.9877	0.646	0.849	72	0.746
13	4	258	10	0.9888	0.653	0.866	72	0.736
12	9	258	10	0.9886	0.605	0.86	72	0.713
11	1	258	10	0.9884	0.658	0.882	72	0.712
10	12	258	10	0.9883	0.661	0.876	72	0.712
9	10	258	10	0.9881	0.668	0.881	72	0.716
8	14	258	10	0.9873	0.688	0.871	72	0.738
7	3	258	10	0.9863	0.716	0.899	72	0.738
6	15	258	9	0.9844	0.786	0.946	72	0.757
5	11	258	10	0.9821	0.805	0.992	72	0.85
4	16	258	8	0.9744	0.897	1	72	0.938
3	13	258	7	0.9706	0.975	1.03	72	1.01
2	8	258	8	0.9707	0.971	1.05	72	1.09
1	2	258	8	0.9559	1.73	1.32	72	1.22

one interval remained (interval 7). The optimized process and results are illustrated in Table 4.

As can be seen in Table 4, the RMSECV and RMSEP first decreased steadily until interval 9, and then increased as additional intervals were left out. In terms of RMSECV and RMSEP, the best PLS model performance was obtained using the 12th interval. These "optimal" spectral regions are combinations of 3100–2800, 1800–1600, and 1500–800 cm^{-1}, and amongst those listed in Table 3 and highlighted in Fig. 1, the number of LV, R^2, RMSEC, RMSECV, and RMSEP were 10, 0.9886, 0.605, 0.86, and 0.71, respectively. This optimized general model performance was substantially superior to that of the conventional full spectrum model and the experienced PLS model developed in this study. It was also much better than the FT-NIR/PLS models developed by Hong *et al*. for PV determination in edible oils [11], Yildiz *et al*. in soybean oil [9], and Wu *et al*. in waste oil [12]. These results suggest that there is a significant advantage in terms of calibrating using the *Bi*PLS algorithm. After selection of the spectral regions, the *Bi*PLS general model was even better than specific models for PV determination in olive oil by the application of stepwise orthogonalization of predictors to mid -FTIR [10]. The number of LV, RMSEC, RMSECV, and RMSEP from Pizarro *et al*.'s model were 10, 0.7, 0.9, and 0.8, respectively. The PV predictions from the *Bi*PLS general model correlate well with the chemical PV, as shown in Fig. 2.
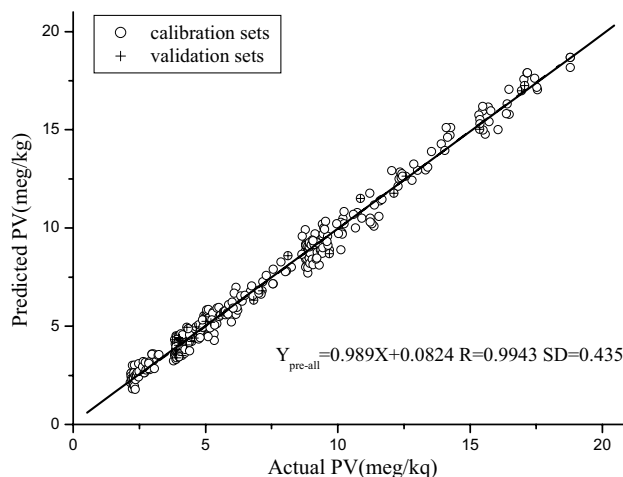

Fig. 2 The linear correlation between PV prediction in single soy, peanut, rapeseed, sunflower oil by general FTIR/*Bi*PLS model and their actual values from AOCS method Cd 8-53

The regression equation is highly linear, with a correlation coefficient of 0.994, a regression coefficient (0.9689) near 1.0, and an intercept (0.0824) near zero, indicating good prediction performance from the *Bi*PLS model.

Variations in triglyceride composition among the oil types were included here as part of the model, and it is likely that the PV model is oil-type dependent, which could account for the better results obtained for virgin

Table 5 The external validation of the FTIR/*Bi*PLS general model with nine kinds of used frying oils

Used frying oil	AOCS method value	FTIR value	RSDa (%)
1 (frying chips)	12.26 ± 0.37	12.49 ± 0.28	1.90
2 (frying chicken chops)	11.07 ± 0.29	10.59 ± 0.12	4.36
3 (frying doughnuts)	12.75 ± 0.22	12.35 ± 0.08	3.16
4 (deep frying dough sticks)	10.08 ± 0.07	9.75 ± 0.24	3.25
5 (frying chicken wings)	11.74 ± 0.24	11.58 ± 0.21	1.40
6 (frying chips)	7.89 ± 0.35	8.07 ± 0.17	2.33
7 (frying stinky tofu)	7.38 ± 0.25	7.50 ± 0.09	1.62
8 (frying stinky tofu)	11.37 ± 0.28	11.14 ± 0.15	2.02
9 (frying chips)	4.94 ± 0.44	5.21 ± 0.19	3.56

$Y_{Sub\_Sunflower} = 0.995X + 0.0521 \ R = 0.9982 \ SD = 0.177$
$Y_{Sub\_Peanut} = 0.995X + 0.0369 \ R = 0.9993 \ SD = 0.149$
$Y_{Sub\_Rapeseed} = 1.00X + 0.0168 \ R = 0.9975 \ SD = 0.288$
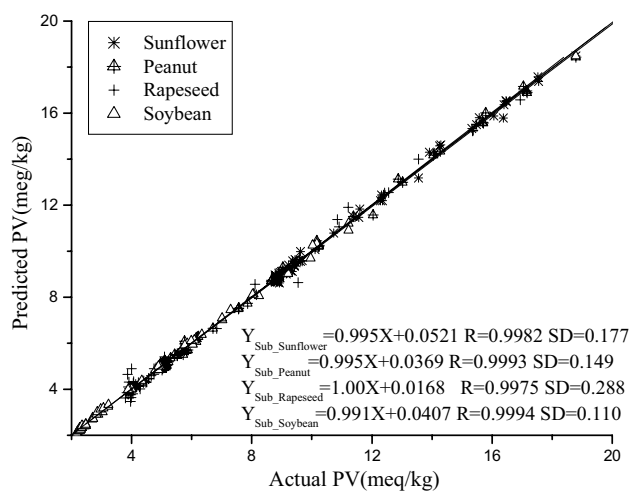$Y_{Sub\_Soybean} = 0.991X + 0.0407 \ R = 0.9994 \ SD = 0.110$

Fig. 3 The linear relationship between submodel prediction of PV in four types of oil (soybean, peanut, rapeseed, sunflower) and actual values from AOCS method Cd 8-53

walnut oil reported by Liang *et al.* [13]. These authors developed a MIR/PLS calibration transmission model for PV determination using 18 virgin walnut oils, and validated the model with ten additional samples. It is likely that only minor variation in the triglycerides composition within a single oil type resulted in significantly lower RMSEC (0.4838) and RMSEP (0.3545) values. However, as no cross-validation was performed and the path length of the transmission cell was not provided, the efficacy and accuracy of their model is in question and cannot be reproduced.

In order to verify the stability of the general model, nine different kinds of UFO were collected and used for external validation. The obtained results are presented in Table 5, which shows that the FTIR model gave lower relative standard deviations (<5 %) for accuracy. The good agreement between the AOCS values and FTIR prediction indicates the remarkable prediction performance of the general *Bi*PLS model developed in the present study.

Submodels

To further analyse whether the oil type affects the PV model performance, submodels of the individual oils were constructed using the *Bi*PLS algorithm. These submodel performances are presented in Table 6.

The highest RMSECV and RMSEP in these analyses were obtained for rapeseed oil, with values of 0.344 and 0.187, respectively, about 1.5 and three times that of soy, respectively. However, the average performance of the individual models was about four times better than that of the combined oil model in terms of RMSECV and RMSEP. Additionally, the number of factors required was significantly lower, clearly indicating that the oil type strongly affects the results and adds to the difficulty of constructing the model. Thus, in developing the PV models, working with a specific oil type leads to a significant improvement in the accuracy of the results. Figure 3 illustrates the performance of individual oil *Bi*PLS models on validation samples, which correlate well with the actual PV. All the

Table 6 Calibration and validation of the submodels for PV determination in single type of oil corresponding to combined spectral regions of 3100–2800, 1800–1600 and 1500–800 cm^{-1} obtained by *Bi*PLS

Model	LV	Calibration set				Validation set	
		N.S	R^2	RMSEC	RMSECV	N.S	RMSEP
General model*	10	258	0.9886	0.605	0.86	72	0.713
Submodel_soybean oil	4	62	0.9996	0.0749	0.105	21	0.123
Submodel_rapeseed oil	4	62	0.9987	0.216	0.344	20	0.187
Submodel_sunflower oil	6	62	0.9993	0.107	0.175	20	0.160
Submodel_peanut oil	7	62	0.9999	0.0426	0.148	20	0.149

* The general model performance here was used for comparison

NS number of samples

regression equations were highly significant, with correlation coefficients >0.997.

Conclusion

This paper compared the algorithms of full spectrum PLS, experienced PLS, and *Bi*PLS for ATR/FTIR PV determination of edible oils. As expected, the results indicate that beyond simple full spectrum analysis, region selection is critical. Among the three different approaches, *Bi*PLS led to a relatively robust model that does not require in-depth spectroscopic knowledge. The systematic iterative examination of equal intervals led to a workable model with little effort other than the input of the data. It is clear that oil type spectral information appears not to be well modelled, which is likely due to the lack of sensitivity of the single-bounce ATR accessory; this would likely not be as much of an issue if transmission spectra were used. At present, most reported PV models are based on a specific oil type, such as soy, olive, or rapeseed oils, etc. As such, when unknown samples are encountered, one cannot select a specific model, so the accuracy of the prediction cannot match that of when the oil type is known. Therefore, a general model independent of oil type is required for routine analysis. Generally, the prediction performance of calibration models is acceptable assuming that enough oil types are included in the calibration model to ensure the general representativeness of the model. Based on the comparison of the PV models, it would appear that *Bi*PLS is a very useful tool for rapidly determining whether a workable calibration can be developed relative to a primary method. It also provides reasonable assurance that the model will be relatively "optimal" compared with a full spectrum PLS calibration, while requiring little or no spectral knowledge, which would normally be needed to further improve the calibration model.

References

1. Guillén MD, Cabo N (2000) Some of the most significant changes in the Fourier transform infrared spectra of edible oils under oxidative conditions. J Sci Food Agr 80(14):2028–2036
2. Tsiaka T, Christodouleas DC, Calokerinos AC (2013) Development of a chemiluminescent method for the evaluation of total hydroperoxide content of edible oils. Food Res Int 54(2):2069–2074
3. Van de Voort FR, Ismail AA, Sedman J, Dubois J, Nicodemo T (1994) The determination of peroxide value by Fourier transform infrared spectroscopy. J Am Oil Chem Soc 71(9):921–926
4. Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB (2000) Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. Appl Spectrosc 54(3):413–419
5. Ma K, Van de Voort FR, Sedman J, Ismail AA (1997) Stoichiometric determination of hydroperoxides in fats and oils by Fourier transform infrared spectroscopy. J Am Oil Chem Soc 74(8):897–906
6. Martens H, Naes T (1989) Multivariate calibration. Wiley, Chichester
7. Rohman A, Che Man YB (2011) The use of Fourier transform mid infrared (FT-MIR) spectroscopy for detection and quantification of adulteration in virgin coconut oil. Food Chem 129(2):583–588
8. De la Mata P, Dominguez-Vidal A, Bosque-Sendra JM, Ruiz-Medina A, Cuadros-Rodríguez L, Ayora-Cañada MJ (2012) Olive oil assessment in edible oil blends by means of ATR-FTIR and chemometrics. Food Control 23(2):449–455
9. Yildiz G, Wehling RL, Cuppett SL (2003) Comparison of four analytical methods for the determination of peroxide value in oxidized soybean oils. J Am Oil Chem Soc 80(2):103–107
10. Pizarro C, Esteban-Díez I, Rodríguez-Tecedor S, González-Sáiz JM (2013) Determination of the peroxide value in extra virgin olive oils through the application of the stepwise orthogonalisation of predictors to mid-infrared spectra. Food Control 34(1):158–167
11. Hong JH, Yamaokakoseki S, Yasumoto K (1994) Analysis of peroxide values in edible oils by near-infrared spectroscopy. J Jpn Soc Food Sci 41(4):277–280
12. Wu J, Xing S, Xiao H, Xu Y, Zhang X (2012) Rapid determination of peroxide value in simulated waste oil by near infrared spectroscopy. Sensor Lett 10(1–2):1–2
13. Liang P, Chen C, Zhao S, Ge F, Liu D, Liu B, Fan Q, Han B, Xiong X (2013) Application of Fourier transform infrared spectroscopy for the oxidation and peroxide value evaluation in virgin walnut oil. J Spectrosc 2013 Article ID 138728