Identifying camellia oil adulteration with selected vegetable oils by characteristic near-infrared spectral regions

Xuan Chu*, Wei Wang*,‡, Chunyang Li†,§, Xin Zhao* and Hongzhe Jiang*

*College of Engineering, China Agricultural University
Beijing 100083, P. R. China

†Institute of Food Science and Technology
Jiangsu Academy of Agricultural Sciences
Nanjing 210014, P. R. China
‡playerwxw@cau.edu.cn
§lichunyang968@126.com

In this paper, a methodology based on characteristic spectral bands of near infrared spectroscopy (1000–2500 nm) and multivariate analysis was proposed to identify camellia oil adulteration with vegetable oils. Sunflower, peanut and corn oils were selected to conduct the test. Pure camellia oil and that adulterated with varying concentrations (1–10% with the gradient of 1%, 10–40% with the gradient of 5%, 40–100% with the gradient of 10%) of each type of the three vegetable oils were prepared, respectively. For each type of adulterated oil, full-spectrum partial least squares partial least squares (PLS) models and synergy interval partial least squares (SI-PLS) models were developed. Parameters of these models were optimized simultaneously by cross-validation. The SI-PLS models were proved to be better than the full-spectrum PLS models. In SI-PLS models, the correlation coefficients of predition set (Rp) were 0.9992, 0.9998 and 0.9999 for adulteration with sunflower oil, peanut oil and corn oil seperately; the corresponding root mean square errors of prediction set (RMSEP) were 1.23, 0.66 and 0.37. Furthermore, a new generic PLS model was built based on the characteristic spectral regions selected from the intervals of the three SI-PLS models to identify the oil adulterants, regardless of the adultrated oil types. The model achieved with Rp= 0.9988 and RMSEP = 1.52. These results indicated that the characteristic near infrared spectral regions could determine the level of adulteration in the camellia oil.

Keywords: Camellia oil; adulteration detection; characteristic near infrared spectral regions; partial least squares; synergy interval partial least squares.

‡,§Corresponding authors.

1. Introduction

Camellia oil is extracted from the seed of camellia (Camellia oleifera Abel) which is one of the four major woody oil plants in the world.[1] The oil was mainly produced in southern China and southeast Asia. Camellia oil is titled as "eastern olive oil", with the similarity to olive oil in physicochemical properties and acid profiles.[2] According to some reports, camellia oil could help to meliorate gastric mucosal injury, lower blood pressure and prevent heart disease.[3,4] As a result of the higher nutritional and medical values, the demand for camellia oil is increasing and it is generally 3–5 times more expensive than other edible oil in China.[5] Therefore, adulterating them with other cheaper oils could be very lucrative. Once adulterated, its unique quality and nutritional characteristics will be destroyed,[6] and this will definitely damage the interests of consumers. The camellia oil annual yield in China is approximately 300,000 tons and increasing.[7] Thus, monitoring the authenticity will help to standardize market order and protect consumers' rights.

To guarantee camellia oil quality, some qualitative chemical detection methods are used. The adulterated oil can be qualitatively identifuied by the discolouration, crystallization or opacitas effects after some reagents being added to.[8] However, considering the diversification of the adulterants, it is very difficult to detect the possible adulterants one by one. Quantitative detection methods, i.e., gas chromatography (GC) and liquid chromatography (LC), can determine the content of specific constituents in adulterated camellia oil accurately.[9,10] These detection methods are time consuming, complicated and traditionally require toxic chemical reagents to preprocess samples. Thus, seeking a real-time and nondestructive identification technique to detect adulteration is important.

Near-infrared (NIR) spectroscopy technology comprises of weak and broad overtones and combination bands of fundamental vibrational transitions associated mainly with many functional groups. Thus, it can be used to discriminate sets of similar biological materials.[11] Since it has rapid, nondestructive, pollution free and low request of environmental control features, it is widely used in quality detection of edible oils and fats.[12–15]

For detecting adulterated camellia oil by NIR spectroscopy, the adulterated samples were normally prepared by mixing pure oil with some cheaper edible oil manually. Chemometrics methods can highlight the slight differences both in the position and absorbance of the bands of different kinds of oils.[16] Li et al.[17] used NIR spectroscopy (1000–2500 nm) with four pattern recognition methods to authenticate pure crude and commercial camellia oil. Each of the four chemometrics method could achieve a classification rate with 98.3% or higher. Hussain et al.[18] built a model that could identify the adulterated camellia oil with an accuracy of 100% by NIR spectra with PLS method. The NIR spectra were obtain by dipping a fiber into the samples prepared through adding soybean oil into camellia oil with the mass percentage of 5–25%. Yuan et al.[6] explored the feasibility of NIR transmittance spectroscopy for the rapid determination of soybean oil and maize oil adulterated in binary and ternary system camellia oils. The results showed that this technology could be able to classify camellia oil, soybean oil, maize oil, and adulterated oil with discriminant equations patterntechniques at 95% confidence level. These good performances displayed that NIR spectroscopy could be successfully used as a rapid, simple, and nondestructive method to discriminate pure camellia oil. However, due to it was local specialty, few reports about the detection of camellia oil adulteration by NIR were found. Some adulteration detection for olive oil could be treated as references,[19–21] as the similarities of camellia oil with olive oil in physical and chemical constants.

The researches about the identification of adulterated camellia oil by NIR spectroscopy technology were mainly based on the full spectral region. Although the results were satisfied, those models always lead to complex calculations as they contained a large number of variables. Many studies have shown that simpler and more accurate calibration models may be achieved by selecting the most informative spectral variables instead of using the full spectrum, as the effect of unwanted spectral regions that are unrelated to oil properties could be eliminated.[22] In application, using the characteristic spectral variables can simplify the design of detection instrument. In recent years, many researchers selected appropriate variables and constructed analysis models to quantify compositions content in oils. Pereira et al.[23] used only one spectral region to determine quality parameters in vegetable oils. Yang et al.[24] adopted variables selection algorithms to measure acid value (AV) in peanut oils. Moreira et al.[25] investigated a variable combination method to monitor twelve physical and chemical parameters

of vegetable oils. These results shown that it was possible to determine the quality of oils with a low number of selected wavelengths. However, there were few researches about finding informative spectral variables to detect adulterated camellia oil.

Thus, the main goal of this paper was to identify camellia oil adulteration with selected vegetable oils by a low number of characteristic near infrared spectral variables. To achieve this goal, the specific objectives were (i) to determine the amount of each kind of adulterants in camellia oil by full spectra; (ii) to find some characteristic spectral regions to detect each type of adulteration in camellia oil separately; (iii) to build a simplified and generic adulterant identification model by the selected characteristic regions, which could quantitatively detect adulterations in camellia oil regardless of the adultrant oil types.

2. Materials and Methods

2.1. *Sample preparation*

Sunflower oil, peanut oil and corn oil were widely used edible vegetable oils in China, and they were often used as adulterants in camellia oil because they were cheap and easily got. Thus, these three kinds of vegetable oils were chosen in this study. From the view of common brand and their sales, we just chosed one type of each oil and purchased from the local market. The adulteranted oils were prepared by manually mixing camellia oil with each type of the vegetable oils with different volume proportions (1-100% V/V) respectively, from 1% to 10% V/V in 1% V/V steps, from 10% to 40% V/V in 5% V/V steps and from 40% to 100% V/V steps (a total of 22 volume proportion for each type of the cheaper vegetable oil). The blended oil injected into a quartz cells ($10 \times 10 \times 45$ mm) to the hight of 30 mm was taken as a sample. For each volume proportion, 9 samples were prepared. In total, 9 samples of pure camellia oil and 594 samples (198 samples for each of the three cheaper vegetable oil) of adulteranted camellia oils were ready, and the sample variance was 927.93 for each type of the adulterants.

2.2. *Spectra acquisition*

NIR spectra were collected by a near-infrared spectrophotometer (AvaSpec-NIR256, AVANTES, Apeldoom,

Holland) along with its supporting software (Ava-Soft 7.0, AVANTES, Apeldoom, Holland). In this work, the FT-NIR spectrometer covered the NIR spectral region from 940 nm to 2574 nm with a resolution of 6.4 nm and it was calibrated by the reference information acquired from whiteboard and dark current. The oil samples were put on a custom-made quartz cell holder, and their absorption spectra were subsequently collected by averaging 100 scans with an integration time of 0.0319 ms.

The raw spectra of all the pure and adulterated camellia oil samples were shown in Fig. 1. For the raw spectra, only the spectra ranging from 1000 nm to 2300 nm were reserved. The wavelengths including 940 nm to 998 nm and 2294 nm to 2574 nm were defined as noisy spectral regions due to the less sensitivity of the spectrometer. As shown in Fig. 1, there are four characteristic absorbance bands on the curves. In the NIR region (between 750 nm and 2500 nm), vibration and combination overtones of the fundamental O–H, C–H, S–H and N–H structures are the main recordable phenomena.[11] Based on previous studies, band 1 ranging from 1111 nm to 1265 nm is related to the second overtone of C–H stretching vibrations of methyl and methylene groups; band 2 ranging from 1342 nm to 1492 nm is related to the C–H stretching vibrations; band 3 ranging from 1639 nm to 1887 nm is associated with the first over tone of CH_2 stretching vibrations of methyl and methylene groups; and band 4 locating between 2083 nm and 2222 nm is associated with C–H stretching vibration with other vibrational modes.[26] For different samples, the entire spectra looked closely similar both on the spectral pattern and the absorbance values. Chemometrics methods were utilized to highlight the small discrepancies of chemical composition between pure camellia
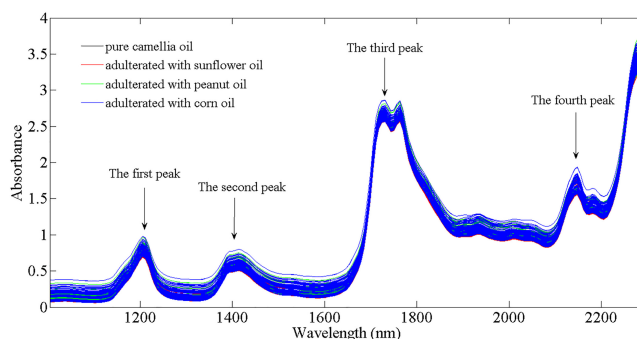

Fig. 1. Original spectra of the pure and adulterated camellia oil samples.

oil and those adulterated with other cheaper vegetable oils.

2.3. Primary chemometrics methods

In this work, chemometric methods were performed with MATLAB 2013b Software (MathWorks, Natick, United states). NIR spectroscopy, which contains hundreds or thousands of wavelengths, is based on molecular overtone and combinations of stretchings and vibrations. Thus it is difficult to assign any special features to some specific chemical components. The complexity of NIR spectra prompts the need for multivariate analysis.[27] Partial least square regression (PLS) is one of the most used regression techniques for multivariate analysis, and it allows a sophisticated statistical approach that is feasible for ordinary multiple regression.[28,29] Synergy interval PLS (SI-PLS) is an expansion of the PLS, it spilts the full spectra into a number of intervals and operates PLS algorithm on the selected spectral region.[30] This method can help to remove some useless information to make the calibration developments more convenient. In this study, PLS and SI-PLS were selected as identification models establishment methods. The performance of these models were evaluated by the following statistical indexes: the root mean square error of cross-validation (RMSECV), the root mean square error of prediction (RMSEP) and the correlation coefficient (R) of calibration (Rc) and prediction set (Rp).[31]

2.3.1. Partial least squares regression

PLS aimed at finding a mathematical relationship between full spectrum information of samples and the adulteration level of samples in this paper.[32] In PLS, both predictors (matrix X) and responses (matrix Y) were transferred into a set of independent factors (latent variables, LVs) which could describe the maximum covariance between spectral information and response variables. The optimal number of LVs used in the establishment of PLS model was determined by cross-validation.[33]

2.3.2. Synergy interval PLS regression

In order to remove useless information and simplify calculation, synergy interval PLS (SI-PLS) was used to select effective spectral intervals data to calibrate PLS model. The principle of this algorithm is to split the spectra into a series of smaller equidistant regions and develop all possible PLS models in all possible combinations of two, three or four intervals. The combination of the intervals and the number of LVs were also chosen by the cross-validation.[34,35]

3. Results and Discussion

3.1. Identification models for single oil adulteration type

3.1.1. Full-spectrum PLS models

The spectral data ranging from 1000–2300 nm were used to bulid identification model. Before calibration, the spectral data usually need to be preprocessed to reduce some unwanted information, such as shifts in baseline, non-linearities and random noise.[36] In this work, standard normal variate (SNV), savitzky–golay smoothing (SG), and first and second derivative were used respectively. In each type of oil adulterants, the samples were randomly divided into calibration and prediction set with a 2:1 division. The calibration set was used to build the identification model for each kind of adulteration, and prediction set was used to verify the predictive ability of the model. The PLS models based on raw spectral data and preprocessed spectra were built respectively, and their results, e.g., Rc, RMSECV, Rp and RMSEP were listed in Table 1.

The results in Table 1 indicated that it could be feasible to identify the level of adulterations in camellia oil by the full-spectrum PLS model. Considering the similarities of the PLS model results between raw spectra and corrected spectra, it was unnecessary to adopt any pretreatment method. So the raw spectra data were selected in the subsequent analysis. For the three types of oil adulteration, Rc and Rp are all over 0.99; and both RMSECV and RMSEP are around 1% V/V. The results of the three models also show a good predictive ability. It indicated that NIR spectroscopy coupled with PLS method has a good ability to detect adulteration content.

3.1.2. SI-PLS models

Although the PLS models gave good results, these models were based on full spectra which included

Table 1. Results of PLS models built by raw and pretreated spectra spectra.

Adulteration type	Prepressing method	The number of LVs	Calibration set		Prediction set	
			Rc	RMSECV	Rp	RMSEP
Sunflower oil	**Original spectra**	**11**	**0.9993**	**1.2**	**0.9990**	**1.27**
	SNV	7	0.9991	1.3	0.9987	1.47
	S_G smooth	8	0.9993	1.2	0.9990	1.29
	The first dervative	6	0.9993	1.15	0.9991	1.24
	The second dervative	7	0.9993	1.16	0.9991	1.25
Peanut oil	**Original spectra**	**12**	**0.9995**	**0.91**	**0.9995**	**1.10**
	SNV	13	0.9996	0.82	0.9996	1.01
	S_G smooth	13	0.9997	0.7	0.9996	0.93
	The first dervative	9	0.9997	0.69	0.9997	0.91
	The second dervative	8	0.9999	0.73	0.9996	0.97
Corn oil	**Original spectra**	**11**	**0.9999**	**0.47**	**0.9999**	**0.52**
	SNV	12	0.9999	0.46	0.9998	0.52
	S_G smooth	9	0.9999	0.47	0.9999	0.52
	The first dervative	6	0.9999	0.45	0.9999	0.5
	The second dervative	10	0.9999	0.37	0.9999	0.41

uninformative regions. To reduce calculation and improve the performance of the PLS model, it was necessary to optimize the selection of spectral.[37] For each kind of the adulterants, SI-PLS was adopted for the establishment of a new detection model. The calibration and prediction sets for SI-PLS modeling were also randomly divided with a 2:1 division. During calibrating of SI-PLS models, three parameters, namely, the number of subintervals division, the combination of subintervals, and the number of LVs of SI-PLS model could all influence the performance of SI-PLS.[38] The full spectra were equally-divided into 6, 7, 8, . . . ,17 subintervals. Total 12 subintervals division cases were got. In each case, SI-PLS model was developed based on two, three or four subintervals and different number of LVs, respectively. These two parameters were determined according to the lowest RMSECV. The optimum model was determined by trading off Rc, Rp, RMSECV, and RMSEP. The optimum model for each case was summarized in Table 2, and the optimum models for each adulteration were prominent with the bold black font.

For the sunflower oil adulteration type, the optimum SI-PLS model was obtained with the combination of 4 intervals from 11 intervals and 8 LVs. The combinations of intervals selected were [4 6 8 9], which corresponded to the four spectral ranges, i.e., 1374.0–1484.4 nm, 1607.7–1717.9 nm, 1.840.7–1950.2 nm, 1956.6–2065.5 nm. Figure 2(a) shows all the divided intervals and the combinations of selected subintervals. The spectral intervals containing the second characteristic band in the original spectra which corresponded to methyl and methylene groups as mentioned before. The values of RMSECV, Rc, RMSEP and Rp were 1.02, 0.9995, 1.23 and 0.9992, respectively.

As shown in Table 2, the optimum SI-PLS model for detecting camellia oil adulterated with peanut oil was established by the combination of 4 intervals from 10 intervals and the 13 LVs. All the intervals were shown in Fig. 2(b), and the optimum combinations of subintervals were [2 5 7 8] which were marked in gray shadow. These subintervals corresponded to the spectral sections of 1140.5–1263.7 nm, 1529.8–1653.1 nm, 1789.0–1911.6 nm and 1918.0–2039.9 nm. The subintervals contained the first characteristic band in the original spectra. The values of RMSECV, Rc, RMSEP and Rp were 0.53, 0.9999, 0.66 and 0.9998.

Table 2 showed the optimum SI-PLS identification model for camellia oil adulterated with corn oil. The optimal number of LVs was 15. The optimum spectral subintervals were [2 6 8], which ranged from 1140.5–1263.7 nm, 1659.6–1782.6 nm, and 1918.0–2039.9 nm. Figure 2(c) showed all the divided 10 intervals and the three optimum spectral subintervals for the SI-PLS model. Those spectral intervals involved the first and third characteristic bands in the original spectra. Among them, the first characteristic band was the same as the first subintervals selected by peanut identification model.

Table 2.   Results of SI-PLS models in each case.

Adulteration type	The number of intervals	Selected intervals	The number of LVs	Calibration set		Prediction set	
				Rc	RMSECV	Rp	RMSEP
Sunflower oil	6	[2 3 5]	9	0.9994	1.13	0.9991	1.17
	7	[2 3 4 6]	12	0.9994	1.04	0.9990	1.32
	8	[3 4 6 7]	9	0.9994	1.06	0.9992	1.13
	9	[2 5 7]	10	0.9994	1.08	0.9991	1.23
	10	[5 7 8]	6	0.9994	1.1	0.9989	1.47
	11	**[4 6 8 9]**	**8**	**0.9995**	**1.02**	**0.9992**	**1.23**
	12	[3 5 6 9]	7	0.9994	1.07	0.9991	1.24
	13	[7 9 11]	8	0.9994	1.05	0.9992	1.15
	14	[4 7 10 12]	8	0.9995	1.03	0.9991	1.22
	15	[10 11 13]	6	0.9994	1.08	0.9993	1.13
	16	[6 8 11 14]	8	0.9995	1.03	0.9989	1.43
	17	[8 9 12 14]	6	0.9995	1.02	0.9990	1.26
Peanut oil	6	[1 2 3]	9	0.9998	0.56	0.9998	0.75
	7	[1 2]	6	0.9997	0.81	0.9996	1.61
	8	[1 2 3 4]	10	0.9998	0.58	0.9998	0.73
	9	[1 2 5 8]	11	0.9998	0.63	0.9998	0.58
	10	**[2 5 7 8]**	**13**	**0.9999**	**0.53**	**0.9998**	**0.66**
	11	[2 5]	10	0.9997	0.76	0.9996	1.01
	12	[2 4 5 6]	12	0.9998	0.54	0.9998	0.80
	13	[2 5 7 11]	13	0.9998	0.58	0.9998	0.75
	14	[2 3 5 7]	12	0.9998	0.58	0.9997	0.82
	15	[2 3 10]	7	0.9997	0.69	0.9997	0.86
	16	[3 8]	13	0.9995	0.96	0.9995	0.99
	17	[2 3 11]	7	0.9998	0.68	0.9997	0.88
Corn oil	6	[1 3 4]	14	0.9999	0.34	0.9999	0.38
	7	[1 2 4]	14	0.9999	0.34	0.9999	0.49
	8	[1 2 4 5]	15	0.9999	0.34	0.9999	0.41
	9	[2 4 5 8]	15	0.9999	0.3	0.9999	0.80
	10	**[2 6 8]**	**15**	**0.9999**	**0.36**	**0.9999**	**0.37**
	11	[2 7 10]	13	0.9999	0.35	0.9999	0.45
	12	[2 7]	10	0.9999	0.44	0.9999	0.47
	13	[2 8 12]	11	0.9999	0.34	0.9999	0.45
	14	[2 7 9 12]	15	0.9999	0.33	0.9999	0.43
	15	[2 3 9 13]	11	0.9999	0.36	0.9999	0.39
	16	[2 3 9 14]	15	0.9999	0.34	0.9999	0.50
	17	[3 5 10 15]	15	0.9999	0.34	0.9999	0.46

Table 2 also displayed that the results for identifying corn oil in camellia oil were Rc = 0.9999 and RMSECV = 0.36 in calibration set, and Rp = 0.9999 and RMSEP = 0.37 in prediction set.

In general, the correlation coefficients (R) of the SI-PLS models for each adulteration type were all over 0.9992, and the corresponding root mean square errors were all around 1% V/V. The RMSECV and Rc were all close to the corresponding RMSEP and Rp, respectively. These results indicated that the SI-PLS models were stable and robust. In contrast with PLS, SI-PLS showed its superiority in detecting the three kinds of vegetable

oils in camellia oil, respectively. All the three SI-PLS models had higher correlation coefficient (R) and lower RMSECV and RMSEP comparing to those in the full-spectrum PLS models. The situation could be explained by the following reasons. The PLS models were obtained from full-spectral region which contained many collinear or irrelevant variables. These variables would weaken the performance of PLS model inevitably. While, SI-PLS models searched all possible subinterval combination firstly and then reasonably selects several spectral subintervals.[39] By removing those redundant variables and retaining interesting information
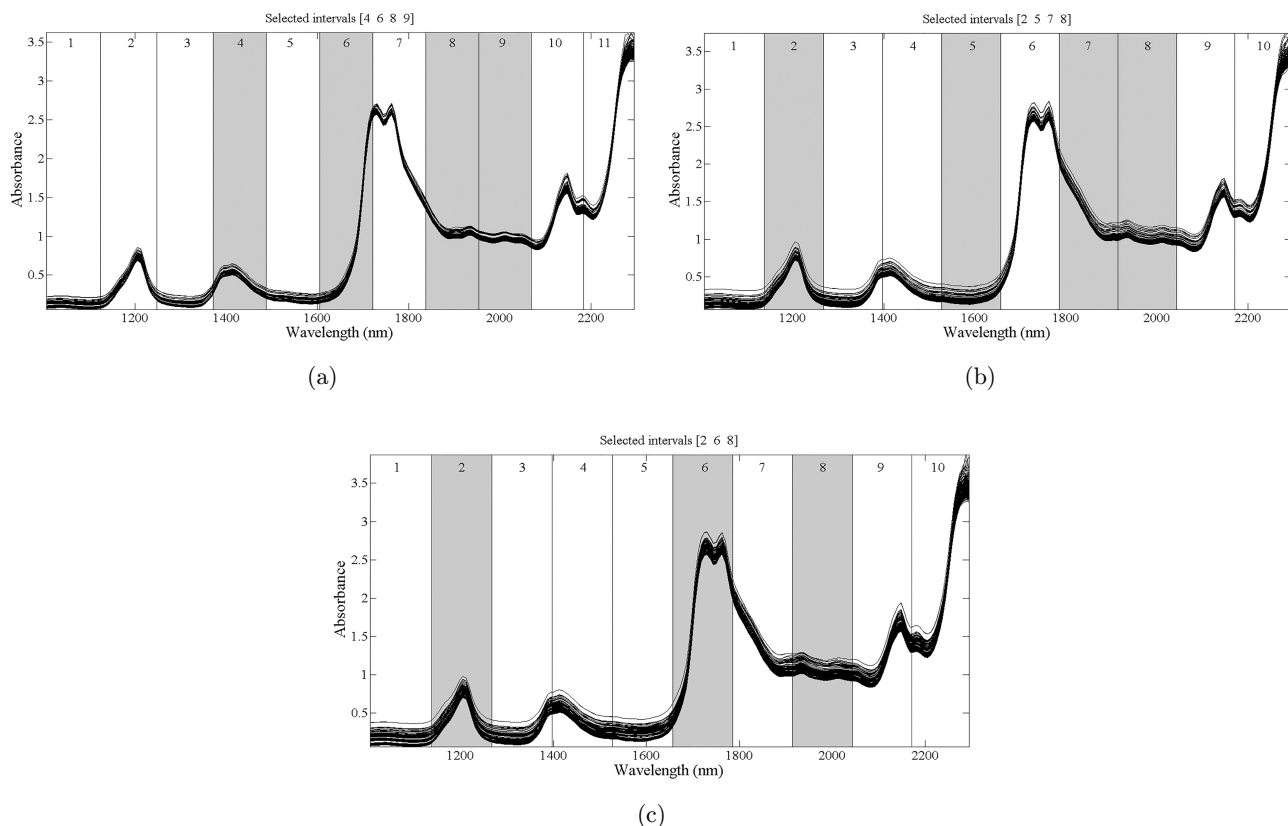
(a)

(b)

(c)

Fig. 2. The intervals used for the three SI-PLS models: (a) the intervals used for sunflower oil identification model, (b) the intervals used for peanut oil identification model, and (c) the intervals used for corn oil identification model.

as much as possible, SI-PLS models could achieve better performance than full-spectrum PLS models.[37] The corresponding functional groups of the spectral ranges further explained why SI-PLS models had a better performance.

In this study, one spectral region was selected as public region if this region was used in at least two SI-PLS models. The analysis about subintervals implied that there were 3 public regions, which corresponded to the following band ranges, i.e., 1140.5–1263.7 nm, 1607.7–1717.9 nm, and 1840.7–2039.9 nm. The position of the public regions on the original spectra was shown in Fig. 3. The public regions of spectral ranges were also corresponded to some vibrations of chemical bonds in the characteristic matters of different kinds of vegetable oils. The first public region (1140.5–1263.7 nm) contained the region of 1111–1265 nm which was attributed to second overtone of C–H stretching vibrations in methyl and methylene groups.[40] The first and second public regions were both in the region of 1075–1639 nm which was related to the second overtone of C–H stretching in monounsaturated

fatty acids and polyunsaturated fatty acids.[41] In the third public region, it included the spectra region of 1955–2035 nm which is associated with phenolic compounds.[42] And it crossed with the region of 1990–2175 nm which was related to the C–C and C–H stretch combination tones of unsaturated fatty acids.[41] Thus, these public regions were mainly related to unsaturated fatty acids and phenolic that were the main differences between camellia oil and other cheaper vegetable oils. Therefore, it was
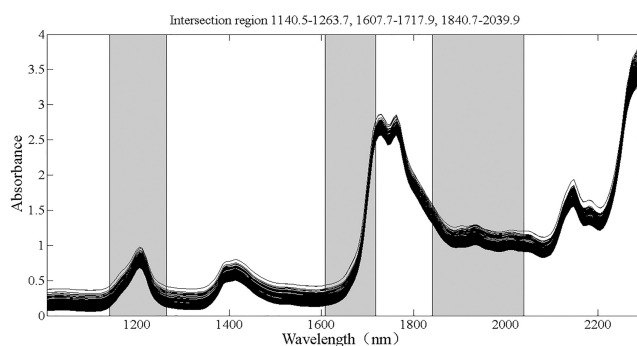

Fig. 3. The position of the public regions on the original spectra.

inferred that those public regions played an important role in molding for each adulteration detection.

3.2. *Generic PLS model by optimum characteristic spectral regions*

There were good results for identifying each type of adulterantion by using characteristic spectral regions. Public regions were then selected from the characteristics spectral regions respectively used in detecting camellia oil adulterated with sunflower, peanut and corn oil. Those public regions were selected as optimum characteristic spectral regions to built a new generic adulterant identification model which could quantitatively detect adulterations in camellia oil regardless of the adultrated oil types. All the 603 samples containing pure camellia oil and adulterated camellia oil were randomly divided into
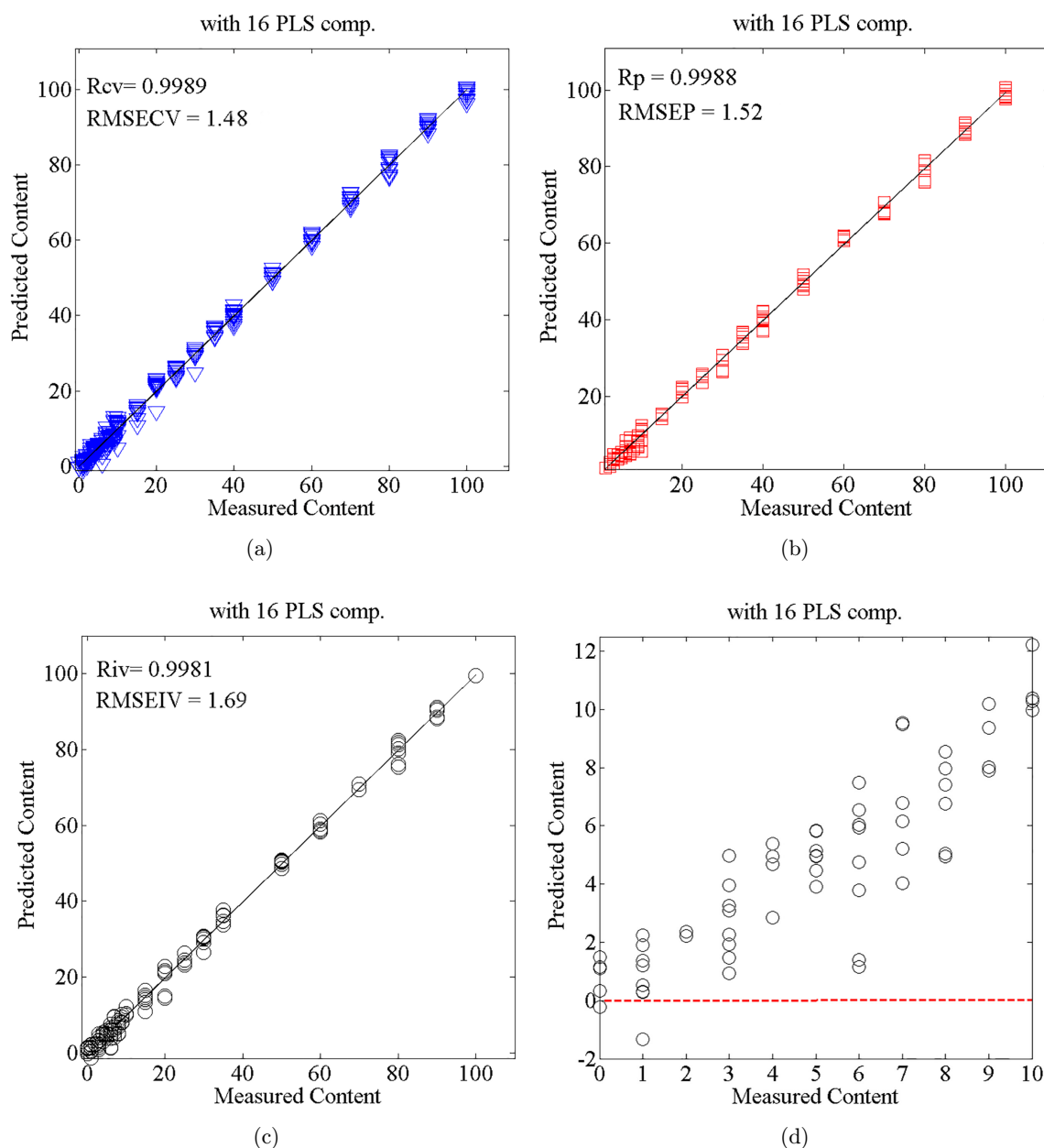

Fig. 4. The scatter plot of the generic PLS model: (a) scatter plot of calibration set, (b) scatter plot of prediction set, (c)scatter plot of independent valiation set, and (d) scatter plot of adulterated camellia oil samples with volume proportions ranged from 0 to 10%V/V in independent validation set.

calibration, prediction and independent validation set with a 3:1:1 division. For each set, they all contained the adulteranted samples in different volume proportions. The optimal number of LVs was 16 which was selected by cross-validation during the establishment of the model. The identification results of the generic PLS model were shown in scatter plots (Fig. 4).

As the spectral regions were reduced in the generic PLS model, the number of LVs in generic PLS model was larger than those in full-spectrum PLS models and SI-PLS models. The scatter plots show that the correlation between actural adulteration level and NIR prediction in the calibration and prediction sets. Here, the value of RMSECV is 1.48, and Rc is 0.9989. In prediciton set, the value of RMSEP is 1.52 and Rp is 0.9988. Relative percent deviation (RPD) was further used to evaluate the model. RPD was introduced as the ratio between the standard deviation of the reference data of the prediction set and the standard error of the prediction set. When the RPD was over 8.1, the model was an excellent calibration. In this study, RPD was 20.1, which indicated that the model was essentially quantitative.[43] The performance of this model was also evaluated by the samples in the independent validation set. The value of the root mean square error of for independent validation (RMSEIV) is 1.69, and correlation coefficient for independent validation set (Riv) is 0.9981. All the parameters indicate the stability of the model.

Compared with those in SI-PLS models, the correlation coefficient was reduced and the root mean square errors was increased in the generic PLS model. This may be caused by the reduction of the spectral variables and the increase of adulterants types. However, compared to SI-PLS models which can identify only one type of adulterants, the generic PLS model can identify the amount of adulterants regardless of their types. As shown in Fig. 4(d), when the adulteration content was 1%V/V, some of the adulteranted camellia oil samples were misjudged as pure camellia oil. It indicated that the detection precisions of adulteration in camellia oil could be as low as 2% V/V. And in real life, in order to pursue more profit, the adulteration level certainly exceeds 2%. Thus, this generic PLS model could be possibly used to detect the content of adulteration in camellia oil.

4. Conclusion

In this study, a method based on characteristic near-infrared spectral regions and multivariate analysis was proposed to identify camellia oil adulteration with three kinds of vegetable oils. The full-spectrum PLS models and SI-PLS models were built for each type of adulterated oil respectively. And a generic PLS model based on the optimum characteristic spectral regions selected by the results of SI-PLS models was further developed, which can detect the content of adulteration in camellia oil regardless of the adulterants type. Following are the conclusions drawn from this study:

(1) Both full near infrared spectral region and characteristic spectral regions could be used to identify the camellia oil adulterated with sunflower, peanut and corn oil. The performance of the identification modles built by SI-PLS is superiority in contrast with PLS models built by full spectral region. In SI-PLS model, the correlation coefficient of predition set (Rp) were 0.9992, 0.9998 and 0.9999 for adulteration with sunflower oil, peanut oil and corn oil seperately; the corresponding root mean square error of prediction set (RMSEP) were 1.23, 0.66 and 0.37.

(2) The optimized characteristic spectral regions (1140.5–1263.7 nm, 1607.7–1717.9 nm, and 1840.7–2039.9 nm) were selected from the public regions of the three SI-PLS models. They were associated unsaturated fatty acids, which represent the difference of camellia oil and other cheap vegetable oil.

(3) The generic PLS model built by the optimized characteristic spectral regions could identify the adulterants in camellia oil regardless of the types of adulterants. The results reached Rp = 0.9988 and RMSEP = 1.52 in prediction set and Riv = 0.9981 RMSEIV = 1.69 in independent validation set.

It should be stressed that this study has only involved a limited kind of camellia oils and cheap vegetable oil. In order to improve the applicability of the model, detection of camellia oil adulterated with different brand and production batch of sunflower oil, peanut oil and corn oil should be carried out gradually, and other kind of cheap vegetable should also be taken as adulterants in the further study.

Acknowledgments

The authors would like to express their gratitude to Avantes China, for their providing of the spectrometers kindly. This work was supported financially by the China National Science and Technology Support Program (Grant No. 2012BAK08B04) and Gannan Camellia Industry Development and Innovative Center Open Fund (Grant No. YK201610).

References

1. T. Wen, L. Zheng, Z. Gong, L. Li, J. Xie, Q. Ma, "Rapid identification of geographical origin of camellia oil based on near infrared spectroscopy technology," *Trans. Chin. Soc. Agric. Eng.* **32**(16), 293–299 (2016).

2. L. Liu, C. Hu, L. Liu, S. Zhang, K. Chen, D. He, "Rapid detection and separation of olive oil and Camellia oil based on ion mobility spectrometry fingerprints and chemometric models," *Eur. J. Lipid Sci. Technol.* **119**(3), (2017).

3. A. Bumrungpert, P. Pavadhgul, R. W. Kalpravidh, "Camellia oil-enriched diet attenuates oxidative stress and inflammatory markers in hypercholesterolemic subjects," *J. Med. Food* **19**(9), 895–898 (2016).

4. P. S. Tu, Y. T. Tung, W. T. Lee, G. C. Yen, "Protective effect of camellia oil (*Camellia oleifera Abel.*) against ethanol-induced acute oxidative injury of the gastric mucosa in mice," *J. Agric. Food Chem.* **65**, 4932–4941 (2017).

5. D. Zhong, F. Tang, M. Ding, R. Mo, D. Shen, Z. Li, "Simplified determination of organophosphorus pesticides in camellia oil," *Anal. Lett.* **50**(8), 1248–1259 (2017).

6. J. J. Yuan, C. Z. Wang, H. X. Chen, J. Z. Ye, H. Zhou, "Identification and detection of adulterated camellia oleifera abel. oils by near infrared transmittance spectroscopy," *Int. J. Food Prop.* **19**(2), 300–313 (2016).

7. Z. Ni, F. Tang, Y. Liu, D. Shen, R. Mo, "Multielemental analysis of camellia oil by microwave dry ashing and inductively coupled plasma mass spectrometry," *Anal. Lett.* **48**(11), 1777–1786 (2015).

8. General administration of quality supervision, inspection and quarantine of the People's Republic of China (2008a). National standards of the People's Republic of China. Inspection of grain and oils: qualitative test of oils, GB/T 5539–2008 (Standard Press of China, Beijing, 2008).

9. General administration of quality supervision, inspection and quarantine of the People's Republic of China (2003). National standards of the People's Republic of China. Oil-tea camellia oil, GB-11765–2003 (Standard Press of China, Beijing, 2008).

10. H. Jabeur, A. Zribi, J. Makni, A. Rebai, R. Abdelhedi, M. Bouaziz, "Detection of Chemlali extra-virgin olive oil adulteration mixed with soybean oil, corn oil, and sunflower oil by using GC and HPLC," *J. Agric. Food Chem.* **62**(21), 4893–4904 (2014).

11. A. Rohman, A. Nugroho, E. Lukitaningsih, Sudjadi, "Application of vibrational spectroscopy in combination with chemometrics techniques for authentication of herbal medicine," *Appl. Spectrosc. Rev.* **49**(8), 603–613 (2014).

12. T. O. Mendes, R. A. da Rocha, B. L. Porto, M. A. de Oliveira, V. D. C. dos Anjos, M. J. Bell, "Quantification of extra-virgin olive oil adulteration with soybean oil: A comparative study of NIR, MIR, and Raman spectroscopy associated with chemometric approaches," *Food Anal. Methods*, **8**(9), 2339–2346 (2015).

13. S. Li, X. Zhu, J. Zhang, G. Li, D. Su, Y. Shan, "Authentication of pure camellia oil by using near infrared spectroscopy and pattern recognition techniques," *J. Food Sci.* **77**(4), 374–380 (2012).

14. J. Yuan, C. Wang, H. Chen, H. Zhou, J. Ye, "Prediction of fatty acid composition in Camellia oleifera oil by near infrared transmittance spectroscopy (NITS)," *Food Chem.* **138**(2), 1657–1662 (2013).

15. C. A. Nunes, "Vibrational spectroscopy and chemometrics to assess authenticity, adulteration and intrinsic quality parameters of edible oils and fats," *Food Res. Int.* **60**, 255–261 (2014).

16. K. Javidnia, M. Parish, S. Karimi, B. Hemmateenejad, "Discrimination of edible oils and fats by combination of multivariate pattern recognition and FT-IR spectroscopy: A comparative study between different modeling methods," *Spectrochim. Acta, Part A* **104**, 175–181 (2013).

17. S. Li, X. Zhu, J. Zhang, G. Li, D. Su, Y. Shan, "Authentication of pure camellia oil by using near infrared spectroscopy and pattern recognition techniques," *J. Food Sci.* **77**(4), 374–380 (2012).

18. M. N. Hussain, M. F. A. Khir, M. H. Hisham, Z. M. Yusof, "Feasibility study of detecting canola oil adulteration with palm oil using NIR spectroscopy and multivariate analysis," In *2014 Int. Conf. Information, Communication Technology and System (ICTS)*, pp. 111–114 (2014).

19. S. R. Karunathilaka, A. R. F. Kia, C. Srigley, J. K. Chung, M. M. Mossoba, "Nontargeted, rapid screening of extra virgin olive oil products for authenticity using near-infrared spectroscopy in combination with conformity index and multivariate statistical analyses," *J. Food Sci.* **81**(10) (2016).

20. T. O. Mendes, R. A. D. Rocha, B. L. S. Porto, M. A. L. D. Oliveira, V. D. C. D. Anjos, M. J. V. Bell, "Quantification of extra-virgin olive oil adulteration with soybean oil: A comparative study of NIR, MIR, and Raman spectroscopy associated with chemometric approaches," *Food Anal. Methods* **8**(9), 2339–2346 (2015).

21. H. Azizian, M. M. Mossoba, A. R. Fardin-Kia, P. Delmonte, S. R. Karunathilaka, J. K. Kramer, "Novel, rapid identification, and quantification of adulterants in extra virgin olive oil using near-infrared spectroscopy and chemometrics," *Lipids*, **50**(7), 705–718 (2015).

22. M. Vohland, M. Ludwig, S. Thiele-Bruhn, B. Ludwig, "Determination of soil properties with visible to near-and mid-infrared spectroscopy: Effects of spectral variable selection," *Geoderma*, **223**, 88–96 (2014).

23. A. F. C. Pereira, M. J. C. Pontes, F. F. G., Neto, S. R. B. Santos, R. K. H. Galvao, M. C. U. Araujo, "NIR spectrometric determination of quality parameters in vegetable oils using iPLS and variable selection," *Food Res. Int.* **41**(4), 341–348 (2008).

24. M. Yang, Q. Chen, F. Y. Kutsanedzie, X. Yang, Z. Guo, Q. Ouyang, "Portable spectroscopy system determination of acid value in peanut oil based on variables selection algorithms," *Measurement* **103**, 179–185 (2017).

25. S. A. Moreira, J. Sarraguça, D. F. Saraiva, R. Carvalho, J. A. Lopes, "Optimization of NIR spectroscopy based PLSR models for critical properties of vegetable oils used in biodiesel production," *Fuel* **150**, 697–704 (2015).

26. T. Woodcock, G. Downey, C. P. O'Donnell, "Confirmation of declared provenance of European extra virgin olive oil samples by NIR spectroscopy," *J. Agric. Food Chem.* **56**(23), 11520–11525 (2008).

27. B. Wang, B. Peng, "A feasibility study on monitoring residual sugar and alcohol strength in kiwi wine fermentation using a fiber-optic FT-NIR spectrometry and PLS regression," *J. Food Sci.* **82**(2), 358–363 (2017).

28. A. Villar, S. Fernández, E. Gorritxategi, J. I. Ciria, L. A. Fernández, "Optimization of the multivariate calibration of a Vis–NIR sensor for the on-line monitoring of marine diesel engine lubricating oil by variable selection methods," *Chemom. Intell. Lab. Syst.* **130**, 68–75 (2014).

29. M. Yang, Q. Chen, F. Y. Kutsanedzie, X. Yang, Z. Guo, Q. Ouyang, "Portable spectroscopy system determination of acid value in peanut oil based on variables selection algorithms," *Measurement* **103**, 179–185 (2017).

30. E. Teye, X. Huang, "Novel prediction of total fat content in cocoa beans by FT-NIR spectroscopy based on effective spectral selection multivariate regression," *Food Anal. Methods* **8**(4), 945–953 (2015).

31. E. Teye, X. Y. Huang, W. Lei, H. Dai, "Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans," *Food Res. Int.* **55**, 288–293 (2014).

32. S. Kucheryavskiy, C. J. Lomborg, "Monitoring of whey quality with NIR spectroscopy — A feasibility study," *Food Chem.* **176**, 271–277 (2015).

33. M. Khorasani, J. M. Amigo, J. Sonnergaard, P. Olsen, P. Bertelsen, J. Rantanen, "Visualization and prediction of porosity in roller compacted ribbons with near-infrared chemical imaging (NIR-CI)," *J. Pharm. Biomed. Anal.* **109**, 11–17 (2015).

34. L. Nørgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, S. B. Engelsen, "Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy," *Appl. Spectrosc.* **54**(3), 413–419 (2000).

35. L. Huang, J. Zhao, Q. Chen, Y. Zhang, "Rapid detection of total viable count (TVC) in pork meat by hyperspectral imaging," *Food Res. Int.* **54**(1), 821–828 (2013).

36. Y. Shao, Y. Li, L. Jiang, J. Pan, Y. He, X. Dou, "Identification of pesticide varieties by detecting characteristics of Chlorella pyrenoidosa using Visible/Near infrared hyperspectral imaging and Raman microspectroscopy technology," *Water Res.* **104**, 432–440 (2016).

37. C. H. Spiegelman, M. J. McShane, M. J. Goetz, M. Motamedi, Q. L. Yue, G. L. Cote, "Theoretical justification of wavelength selection in PLS calibration: Development of a new algorithm," *Anal. Chem.* **70**(1), 35–44 (1998).

38. H. Jiang, G. Liu, C. Mei, S. Yu, X. Xiao, Y. Ding, "Measurement of process variables in solid-state fermentation of wheat straw using FT-NIR spectroscopy and synergy interval PLS algorithm," *Spectrochim. Acta Part A,* **97**, 277–283 (2012).

39. X. Wang, Y. Bao, G. Liu, G. Li, L. Lin, "Study on the best analysis spectral section of NIR to detect alcohol concentration based on SiPLS," *Procedia Eng.* **29**, 2285–2290 (2012).

40. O. Galtier, O. Abbas, Y. Le Dréau, C. Rebufa, J. Kister, J. Artaud, N. Dupuy, "Comparison of PLS1-DA, PLS2-DA and SIMCA for classification by origin of crude petroleum oils by MIR and virgin olive oils by NIR for different spectral regions," *Vib. Spectrosc.* **55**(1), 132–140 (2011).

41. A. M. Inarejos-García, S. Gómez-Alonso, G. Fregapane, M. D. Salvador, "Evaluation of minor components, sensory characteristics and quality of virgin

olive oil by near infrared (NIR) spectroscopy," *Food Res. Int.* **50**(1), 250–258 (2013).

42. D. Cozzolino, M. J. Kwiatkowski, M. Parker, W. U. Cynkar, R. G. Dambergs, M. Gishen, M. J. Herderich, "Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy," *Anal. Chim. Acta.* **513**(1), 73–80 (2004).

43. A. J. Santos, O. Anjos, R. Simões, J. Rodrigues, H. Pereira, "Kappa number prediction of Acacia melanoxylon unbleached kraft pulps using NIR-PLSR models with narrow interval of variation," *BioResources* **9**(4), 6735–6744 (2014).