Research Article

Rapid Analysis of Geographical Origins and Age of *Torreya grandis* Seeds by NIR Spectroscopy and Pattern Recognition Methods

Lu Xu,[1] Si-Min Yan,[1] Chen-Bo Cai,[2] Wei Zhong,[2] and Xiao-Ping Yu[1]

[1] *Zhejiang Provincial Key Laboratory of Biometrology and Inspection & Quarantine, College of Life Sciences, China Jiliang University, Hangzhou 310018, China*
[2] *Department of Chemistry and Life Science, Chuxiong Normal University, Chuxiong 675000, China*

Correspondence should be addressed to Xiao-Ping Yu; yxp@cjlu.edu.cn

This paper develops a rapid method for discriminating the geographical origins and age of roasted *Torreya grandis* seeds by near infrared (NIR) spectroscopic analysis and pattern recognition. 337 samples were collected from three main producing areas and produced in the last two years. The objective of geographical origins analysis is to discriminate the seeds from Fengqiao with a protected geographical indication (PGI) from those of another two provinces. Age classification is aimed to detect the old seeds produced in the last year from the freshly produced ones. Partial least squares discriminant analysis (PLSDA) was used to develop classification models, and the influence of data preprocessing methods on classification performance was also investigated. Taking second-order derivatives of the raw spectra proves to be the most proper and effective preprocessing method, which can remove baselines and backgrounds and reduce model complexity. With second derivative spectra, the sensitivity and specificity were 0.939 and 0.871 for age discrimination, respectively. Perfect classification was obtained, and both sensitivity and specificity were 1 for discrimination of geographical origins.

1. Introduction

Chinese *Torreya* (*T. grandis*), a species of conifer in the Cephalotaxaceae family, is a rare cash crop tree endemic to southern and eastern China [1, 2]. The plant has been extensively used to cure cough, rheumatism, and intestine verminosis as a potent folk medicine. Various chemical components have been identified in *T. grandis* with an extensive spectrum of biological and medical activities, including anti-inflammatory, antihelmintic, antitussive, carminative, laxative, antifungal, antibacterial, and antitumor activities [3–7]. Its seeds are rich in proteins, fatty acids, carbohydrates, calcium, phosphorus, and iron [8]. Due to the high nutritional value, *T. grandis* seeds are served as a high-quality nut, and cakes, biscuits, and candies made from the seed kernels are also very popular. The seeds have an oil content of 54.62%–61.47% [8], and the oil is bright yellow with pleasant fruit flavors. The oil contains a very high level of unsaturated fatty acids, which accounts for 79.41% of the total fatty acids

[8]. *T. grandis* seed oil is valued as a healthy and functional edible oil, and investigations have demonstrated its effects in modifying lipid metabolism, lowering triacylglycerol and cholesterol levels, and preventing atherosclerosis [9–13].

T. grandis seeds are sold in the form of a roasted nut, and the oil is often extracted by expressing the roasted seeds. In China, *T. grandis* is mainly distributed in southern and eastern provinces, among which the most important producing area is Zhuji in Zhejiang province with an annual yield of about 850,000 kilograms, accounting for a major share of the domestic market. *T. grandis* from Zhuji is awarded a Protected Geographical Indication (PGI) and is known nationwide for its extra-high quality. Although two neighboring provinces, Anhui and Jiangxi, also have a considerable yield of *T. grandis* seeds of similar appearance, their mouth feel and taste are much inferior to those of Zhuji. Another quality factor is the age of seeds. Because the seeds are rich in unsaturated fatty acids and proteins, which tend to be oxidized and spoiled, the old *T. grandis* seeds have

a degraded mouth feel and taste. Oil extracted from the old seeds would be of much lower grade. Unfortunately, for economical reasons, the shelf-life of produced *T. grandis* seeds is sometimes tampered by some producers or old seeds are introduced into fresh ones. Therefore, to ensure consumer interests, effective and rapid methods are required to detect old *T. grandis* seeds from fresh ones and discriminate seeds of different geographical origins.

Some researches have been dedicated to investigations of the chemical compositions of *T. grandis* seeds influenced by different factors, such as the geographical origins and processing conditions [14–16]. All the studies are based on chemical analysis methods, which are important to understand the biological and pharmaceutical properties of *T. grandis* seeds but usually lack a comprehensive view of chemical compositions. Actually, it is difficult to perform a thorough chemical analysis of *T. grandis* seeds and characterize the quality parameters by the levels of a few chemical components. Recently, spectroscopy coupled with chemometric methods has been widely applied in food analysis [17–19]. The rationale behind such techniques is that chemical compositions of complicated samples can be characterized by multivariate spectral signals; then useful information concerning food quality can be extracted by chemometric methods. Near infrared (NIR) spectroscopy is one of the most widely used spectroscopic techniques in food quality evaluation and has some advantages over traditional chemical analysis methods, including lower sample preparation requirements, reduced analysis time and cost, the feasibility for multicomponent analysis, and the potential use for online analysis [20, 21].

For quality evaluation of *T. grandis* seeds, this paper aims at developing a rapid analysis method to discriminate the age and geographical origins of *T. grandis* seeds by NIR spectroscopy and chemometrics. Partial least squares discriminant analysis (PLSDA) [22] was used as the pattern recognition method. The influence of different data preprocessing methods on classification performance was also studied.

2. Materials and Methods

2.1. Collection of Samples.
For geographical origins analysis, 240 freshly roasted (the nuts were produced in 2011) seeds were collected from three main producing areas, namely, Fengqiao in Zhejiang (110), Anhui (58), and Jiangxi (72). Although *T. grandis* is distributed in many southern and eastern provinces in China, just the above three provinces have a considerable growing area and output. Therefore, the objective of geographical origins analysis is to discriminate the (fresh) seeds from Fengqiao with PGI and those from the other two provinces. Age analysis is focused on the most important products from Fengqiao. Therefore, another 97 "old" Fengqiao objects produced in 2010 were also collected. All the objects were prepared by roasting with sand bath and were stored in a cool place with integral endocarp and packaging before spectroscopic analysis. The detailed information concerning samples is shown in Table 1.

2.2. NIR Spectroscopy Measurement.
Because the seed is not contacted closely with the endocarp and the seed coat would

TABLE 1: Detailed information of the *T. grandis* seeds analyzed.

Producing areas	Batch size	Types[a]
Zhongjialing (Fengqiao)	23	Fresh
Xikeng (Fengqiao)	20	Fresh
Dujiakeng (Fengqiao)	22	Fresh
Lixuan (Fengqiao)	21	Fresh
Waixuan (Fengqiao)	24	Fresh
Xikeng (Fengqiao)	23	Old
Dujiakeng (Fengqiao)	30	Old
Zhongjialing (Fengqiao)	24	Old
Waixuan (Fengqiao)	20	Old
Huangshan (Anhui)	22	Fresh
Xiuning (Anhui)	17	Fresh
Yixian (Anhui)	19	Fresh
Wuyuan (Jiangxi)	31	Fresh
Yushan (Jiangxi)	22	Fresh
Guangfeng (Jiangxi)	19	Fresh

[a] Old: produced in 2010; fresh: produced in 2011.

introduce unwanted variations in the measured spectra, NIR spectra were measured with the intact and bare kernel with seed coat removed. The NIR spectra were collected in the diffuse reflectance mode by using a Bruker-TENSOR37 FTIR spectrometer (Bruker Optics, Ettlingen, Germany). A fiber bundle was used to illuminate the sample and collect the scattered light. The fiber probe was placed directly to contact with equatorial region of the seed kernel, because the internal composition changes are more easily explored in the equatorial region rather than the two sides. Particularly, the radius of curvature in the two ends is too small to have close contact with the probe. To account for the differences in the internal composition, the diffuse reflectance spectrum was obtained by averaging three measurements carried out round the equatorial region of a seed. Each spectrum was the average of 64 scanning spectra, and more scans did not improve the signal quality significantly. The range of the raw spectra was from 12,000 to 4000 cm^{-1}, and the data were measured with an interval of 7.714 cm^{-1}, so each raw spectrum has 1037 individual data points. The temperature was kept around 25°C and the humidity was kept at a steady level in the laboratory.

2.3. Outlier Diagnosis.
All the data analysis was performed on MATLAB 7.0.1 (MathWorks, Sherborn, MA, USA). Outliers are the abnormal objects that deviate from the global behavior of all other observations. For classification models, outliers in the training set would lead to bias or even breakdown of the models and outliers in the test set would cause misleading results for evaluation of model performance. Considering the high-dimensional nature of NIR spectra and to overcome the masking effects caused by multiple outliers, robust statistical methods with dimension reduction strategies are needed to analyze the data. Therefore, robust principal component analysis (ROBPCA) [23] was used to detect the outliers. Given the number of significant principal components (PCs), ROBPCA computes the robust latent variables and the corresponding orthogonal distance (OD) and score distance (SD). With a given model significance level, ROBPCA classifies each object

into one of the four groups: regular points (with small SD and small OD), good PCA-leverage points (with large SD and small OD), orthogonal outliers (with small SD and large OD), and bad PCA-leverage points (with large SD and large OD).

2.4. Preprocessing and Data Splitting. When the measured spectra are contaminated with significant noises, unwanted variations such as backgrounds and baselines, and other undesirable factors, data preprocessing is required to reduce their negative influence on classification models. Considering the lack of sufficient prior information concerning the measured spectra, different options were investigated for data preprocessing. Smoothing can depress random noise present in the signal and enhance the signal-to-noise ratio (SNR). The S-G polynomial fitting algorithm [24] was used for smoothing for its effectiveness and simplicity. Taking second derivatives can enhance spectral differences and resolution and remove most of the baselines and backgrounds. In order to avoid the degradation of SNR by enhancing noise, the derivative spectra were also computed by S-G, the polynomial fitting algorithm, rather than by direct differencing. Standard normal variate transformation (SNV) [25] was to reduce scattering effects in the spectra and correct the interference caused by possible variations in optical path.

The DUPLEX algorithm [26] was applied to split the measured spectral data into a representative training set and prediction set. DUPLEX firstly selects the two objects with the largest Euclidian distance and puts them in the training set and then selects the two objects with the largest distance among the remaining samples and puts them in the training set. The above procedure is repeated until enough test objects are obtained and all the remaining objects are put into the training set. By alternatively selecting the objects for the training set and test set, DUPLEX gives data in the test set with a distribution almost equal to that of the training set.

2.5. PLSDA. For PLSDA, the training set of NIR spectra can be arranged in an $N \times p$ matrix $\mathbf{X}$ containing the absorbance measurements at p wavelengths for N samples. For multiclass problems, N denotes the total size of all the B different classes (in this paper, $B = 2$). A response matrix $\mathbf{Y}$ ($N \times B$) can be constructed containing the category variables of each sample in $\mathbf{X}$, where each row vector in $\mathbf{Y}$ indicates the class of an object. If a sample belongs to class i ($i = 1 : B$), the ith element of its response variable is assigned a value of 1 and otherwise 0. Then B partial least squares (PLS) regression models are developed to fit each column of $\mathbf{Y}$ using all the predictors in $\mathbf{X}$. A future object is classified into class j ($j = 1 : B$) when the jth element of its predicted response vector is the nearest to 1. In this paper, because the models deal with two-class objects, one response variable of PLS regression was used which has 1 for one class and -1 for the other class. A cutoff value of 0 was used to assign a new object according to its predicted response value.

It is not trivial to select the number of PLSDA components or the model complexity. Too many latent variables would result in an unnecessarily complicated model with degraded and instable predictions; selecting too few components would lose useful data information and underfit the

data. Therefore, in this paper, an improved cross-validation algorithm, Monte Carlo cross-validation (MCCV) [27], was used to estimate the number of latent variables. By multiple splitting of training data and having a higher percent of training objects for prediction, MCCV can reduce the risk of selecting too many latent variables. The mean percentage error of MCCV (MPEMCCV) was used as the criterion for selecting PLS components:

$$\text{MPEMCCV} = \frac{\sum_{i=1}^{k} E_i}{\sum_{i=1}^{k} L_i}, \tag{1}$$

where K is the times of data splitting, L_i is the number of prediction objects, and E_i is the number of misclassified for the ith splitting. To evaluate the performance of classification models, sensitivity (Sens.) and specificity (Spec.) of prediction were used as follows:

$$\begin{aligned} \text{Sens.} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{Spec.} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \end{aligned} \tag{2}$$

where TP, FN, TN, and FP denote the numbers of true positives, false negatives, true negatives, and false positives, respectively.

3. Results and Discussions

3.1. NIR Spectra. Some of the raw NIR spectra of *T. grandis* seed kernels are shown in Figure 1. Seen from Figure 1, the spectra of three groups of seeds share very similar absorbance patterns in the range of 4000–12000 cm^{-1}. The wide bands in 10000–11000 cm^{-1} are the overlapping of the second overtones of –OH (H$_2$O) stretching ($\sim$10400 cm^{-1}) and the third overtones of –CH stretching ($\sim$11000 cm^{-1}) in various groups. The absorbance in 8200–8600 cm^{-1} is mainly caused by the second overtones of C–H stretching in various groups. The wide bands in 6100–7000 cm^{-1} might be the overlapping of the first overtones of O–H (H$_2$O) stretching and N–H stretching. The peaks around 5700 cm^{-1} (two peaks at 5685 cm^{-1} and 5809 cm^{-1}) can be attributed to the first overtones of C–H stretching in various groups. The peaks around 5160 cm^{-1} are the combination of asymmetric stretching and bending vibration of H$_2$O. Peaks in 5160 cm^{-1} and 6100–7000 cm^{-1} reflect the variations of water contents in different seeds. The absorbance in 4600–4800 cm^{-1} is the combination of O–H bending and C–O stretching. This interval, as well as the peaks around 4300 cm^{-1} (two peaks at 4266 cm^{-1} and 4335 cm^{-1}) caused by C–H stretching and C–H deformation, is very stable and carries much chemical information. Accurate assignments of peaks were difficult due to low resolution and significant baselines; therefore, chemometric methods are necessary to extract the useful information from spectra for classification. The spectral interval of 9000–12000 cm^{-1} carries little chemical information but contaminated with significant noise, so this interval was excluded for further data analysis.

Figure 2 demonstrates the principal component analysis (PCA) plot of raw NIR spectra (4000–9000 cm^{-1}) of three
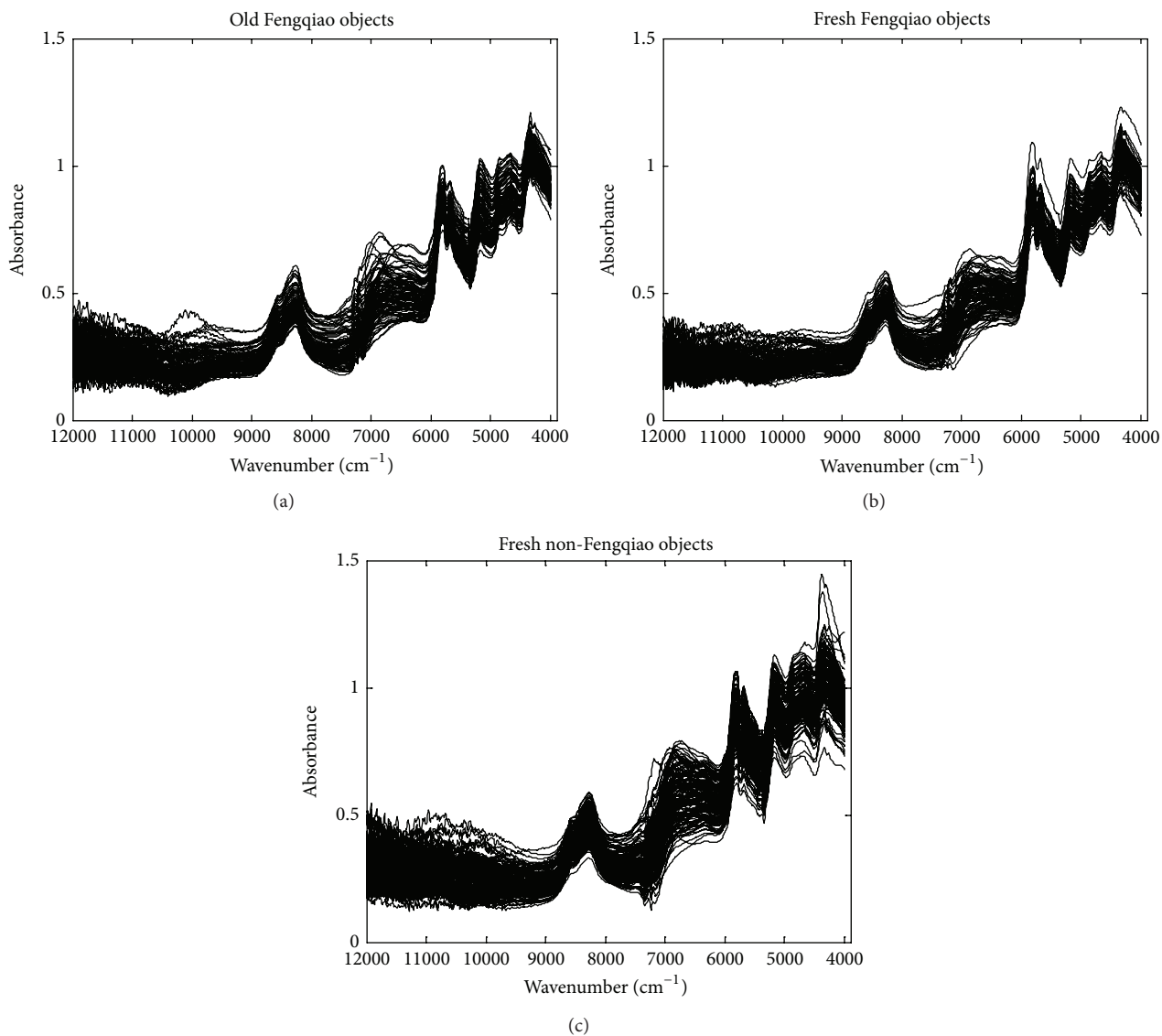
(a)

(b)

(c)

Figure 1: Raw NIR spectra of *T. grandis* seed kernels from Fengqiao (old, $n = 97$), Fengqiao (fresh, $n = 110$), and other areas (fresh, $n = 130$).
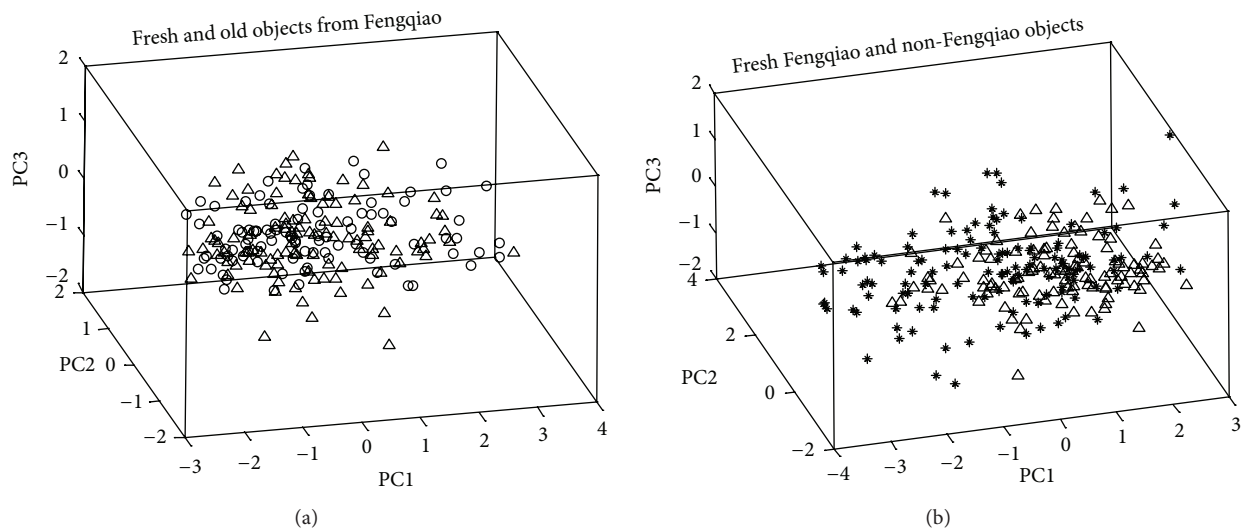

(a)

(b)

Figure 2: PCA plot of raw NIR spectra (4000–9000 cm^{-1}) of different seed kernels; Fengqiao (old, $n = 97$, ○), Fengqiao (fresh, $n = 110$, △), and other areas (fresh, $n = 130$, ∗).
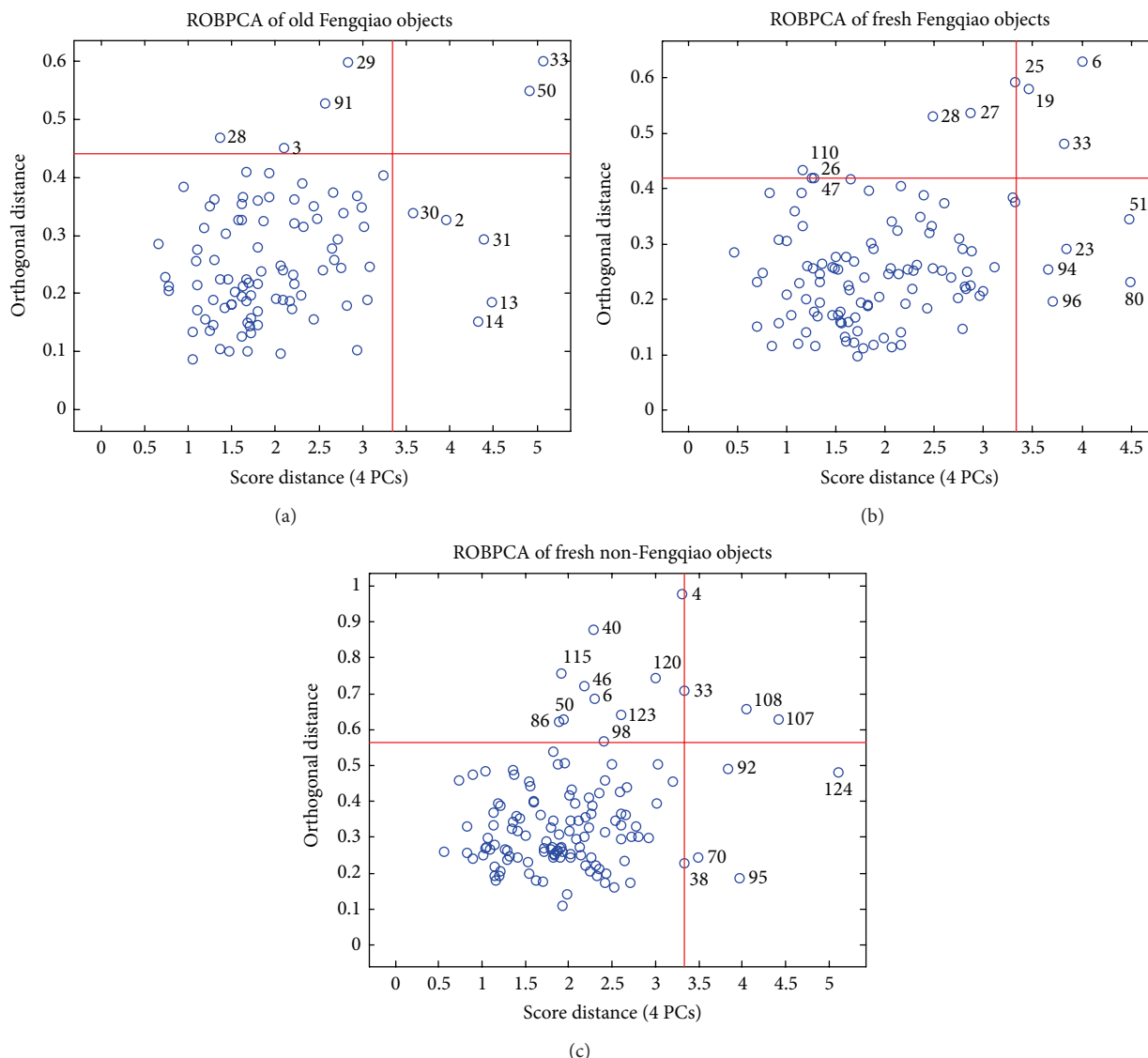
FIGURE 3: ROBPCA outlier diagnosis plots of the three classes of seed kernels based on raw spectra (4000–9000 cm^{-1}).

groups of seed kernels. For classification data of age and geographical origins, the first three principal components (PCs) explain 91.3% and 89.7% of the total variances, respectively. Seen from Figure 2, the projections of the samples in both classes onto the 3-PC subspace are very dispersive, and in each case, the two classes to be discriminated are highly overlapped. Obviously, the data have a contorted and complex structure, and three PCs are insufficient to discriminate them from eachother. Moreover, seen from Figure 2, there are some extremely distributed or abnormal points in the PCA space, so outlier detection and data preprocessing are required to reveal the real structure of data.

3.2. Outlier Detection, Data Splitting, and Preprocessing. Outlier detection was performed on the raw spectra (4000–9000 cm^{-1}) of each group by ROBPCA, where the fraction of outliers the algorithm can resist was set to be 0.10. Following the rule of thumb, the number of PCs was selected to

account for at least 95% of the total data variance. The ROBPCA diagnosis plots of the three classes of seeds are shown in Figure 3. OD is a measure of the distance from the sample to the reconstructed subspace, and SD describes the sample dispersion in the class projected onto the model space. Therefore, both orthogonal outliers and bad PCA-leverage points should be excluded from the training set. Because each class of seeds was collected from different producing subareas, there might be considerable difference in the chemical compositions. Therefore, good PCA-leverage points should be reserved to represent the within-group spectral variations. According to Figure 3, objects 6, 9, and 13 were detected as outliers for old Fengqiao seeds (objects 3, 28, 29, 33, 50, and 91), fresh Fengqiao (objects 6, 19, 25, 26, 27, 28, 33, 47, and 110), and fresh non-Fengqiao seeds (objects 4, 6, 33, 40, 46, 50, 86, 98, 107, 108, 115, 120, and 123), respectively. The above detected outliers were excluded from further data analysis.
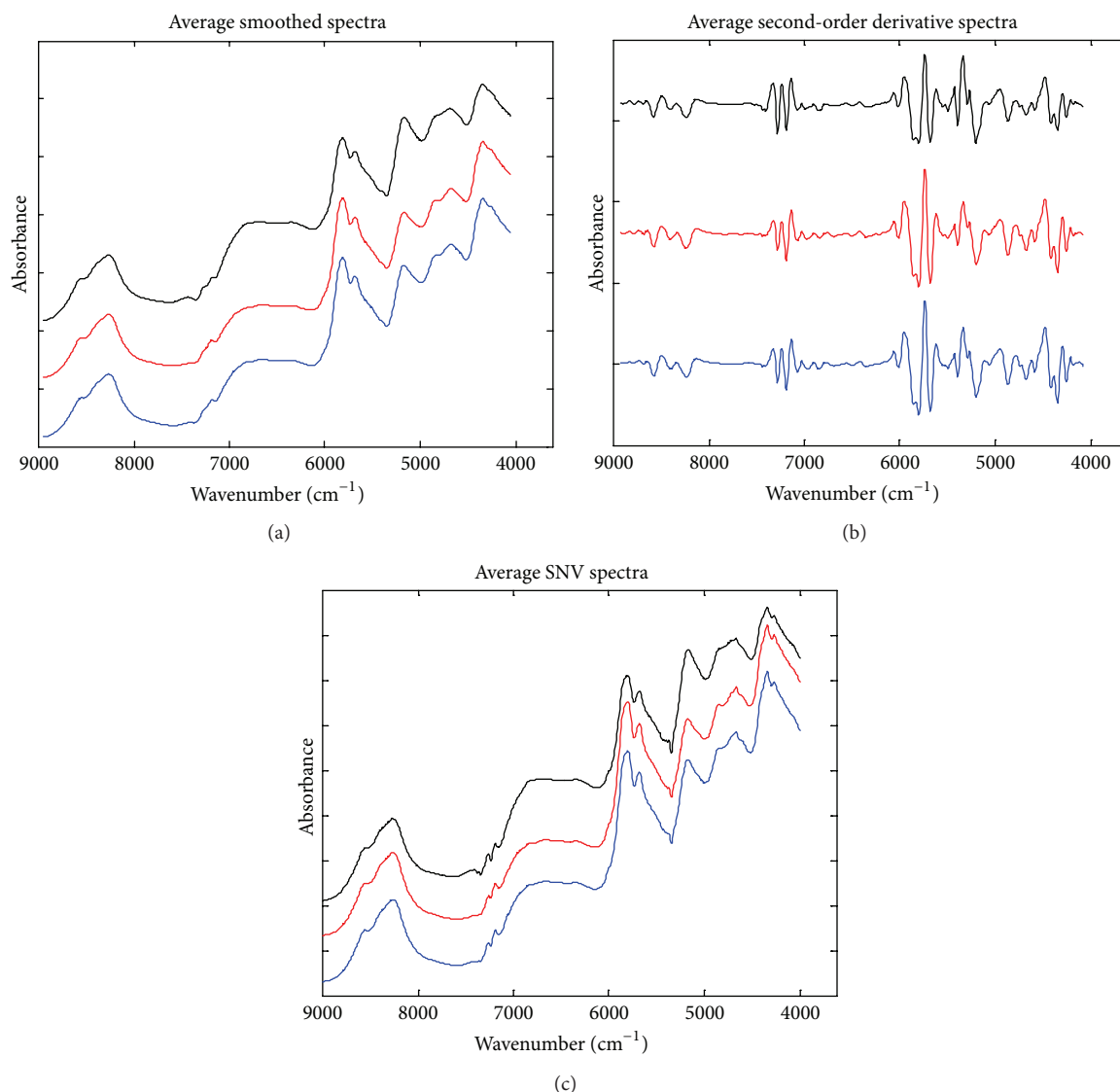
FIGURE 4: Average preprocessed spectra of different classes of *T. grandis* seeds by smoothing, taking second derivatives, and SNV transformation; old Fengqiao (blue), fresh Fengqiao (red), and fresh non-Fengqiao (black).

Figure 4 demonstrates the average preprocessed spectra of *T. grandis* seed kernels. By comparison of the raw spectra with smoothed spectra, although smoothed spectra can slightly improve the SNR, it might lose some useful high-frequency information in the raw data. The second-order derivative spectra can remove most of the baselines and enhance some detailed information and peak resolution. SNV spectra can remove some spectral variations while enhancing others. The effects of data preprocessing should be evaluated by classification performance.

Because the distributions of three groups of objects are not the same, the DUPLEX method was performed on each of the groups separately. For the old Fengqiao seeds, fresh Fengqiao seeds, and fresh non-Fengqiao seeds, 60, 68, and 78 objects were selected by DUPLEX for training and 31, 33, and 39 objects were left for prediction. For different classification objectives, the data were combined to form training and test sets. For discrimination of geographical origins and age, 146 (68/78, Fengqiao/non-Fengqiao) and 128 (68/60, fresh/old) objects were used for training; 72 (33 + 39, Fengqiao/non-Fengqiao) and 64 (33/31, fresh/old) objects were used for prediction, respectively.

3.3. Model Optimization and Validation. Monte Carlo cross-validation (MCCV) was used to estimate the number of PLSDA latent variables. The mean percentage error of MCCV (MPEMCCV) with different model complexity was computed, and the number of latent variables was determined to obtain the lowest MPEMCCV value. Because 3 PCs can explain about 90% of the total data variances, the largest number of PLS components was set to be 10 to avoid overfitting of classification models.

The sensitivity and specificity of prediction were used to evaluate the model performance and the influence of

TABLE 2: Discrimination of age and geographical origins of *T. grandis* seeds by PLSDA.

Preprocessing	Age classification			Geographical origins classification		
	Sensitivity[a]	Specificity[b]	Lv[c]	Sensitivity	Specificity	Lv
Raw data	0.727 (24/33)	0.839 (26/31)	10	0.879 (29/33)	0.923 (36/39)	10
Smoothing	0.636 (21/33)	0.807 (25/31)	10	0.879 (29/33)	0.923 (36/39)	10
2nd derivative	0.939 (31/33)	0.871 (27/31)	7	1.000 (33/33)	1.000 (39/39)	6
SNV	0.788 (26/33)	0.774 (24/31)	10	0.909 (30/33)	0.897 (35/39)	10

[a] Numbers in parentheses indicate TP/(TP + FN).
[b] Numbers in parentheses indicate TN/(TN + FP).
[c] Numbers of PLSDA latent variables.

(a)

(b)
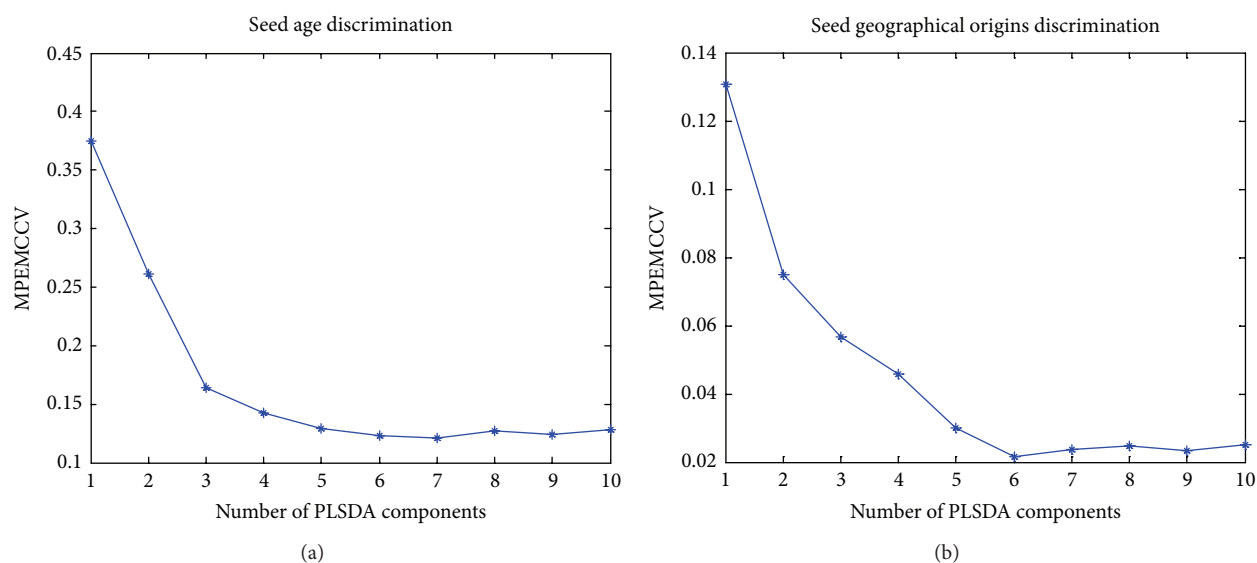
FIGURE 5: The selection of PLSDA components by MCCV for discrimination of seed age and geographical origins with second derivative spectra (4000–9000 cm^{-1}).

preprocessing methods. For discrimination of both age and geographical origins, the fresh Fengqiao seeds were denoted as "positives" and the other class as "negatives." With different preprocessing methods, the prediction results and optimized parameters were listed in Table 2. Seen from Table 2, the models based on the raw data, smoothed spectra, and SNV-transformed spectra all have a very high model complexity and require 10 and more PLS components. Although including more latent variables can improve the classification, the improvements were not significant, and this would lead to instable predictions of future objects. This might be attributed to the nonlinearity caused by the signal baseline and backgrounds. Taking second derivative significantly sharpened the classification models by reducing the baseline and backgrounds. The model complexity of PLSDA based on second derivative was largely reduced. Models with 7 and 6 PLS components obtained the best classification of age and geographical origins, respectively. With second derivative spectra, for age discrimination, the sensitivity and specificity were 0.939 and 0.871, respectively; for discrimination of geographical origins, perfect classification was obtained and both sensitivity and specificity were 1. The selection

of PLS components and prediction results with second-order derivative spectra were demonstrated in Figures 5 and 6, respectively. The comparison of different preprocessing methods demonstrates that the classification performance of a PLSDA model is greatly influenced by backgrounds and baselines in the spectra. Taking second-order derivatives is proved to be the most proper and effective preprocessing method.

4. Conclusions

Rapid discrimination of geographical origins and age of *T. grandis* seeds was developed by using NIR spectroscopy and pattern recognition. The presence of baseline and backgrounds introduces significant nonlinearity factors in the measured NIR spectra; preprocessing by smoothing and SNV can not improve the classification performance. By removing spectral background and baseline, taking second-order S-G derivatives can not only improve classification accuracy but also reduce the complexity of PLSDA. This paper demonstrates NIR combined with chemometrics which provides a useful tool for discriminating age and geographical origins
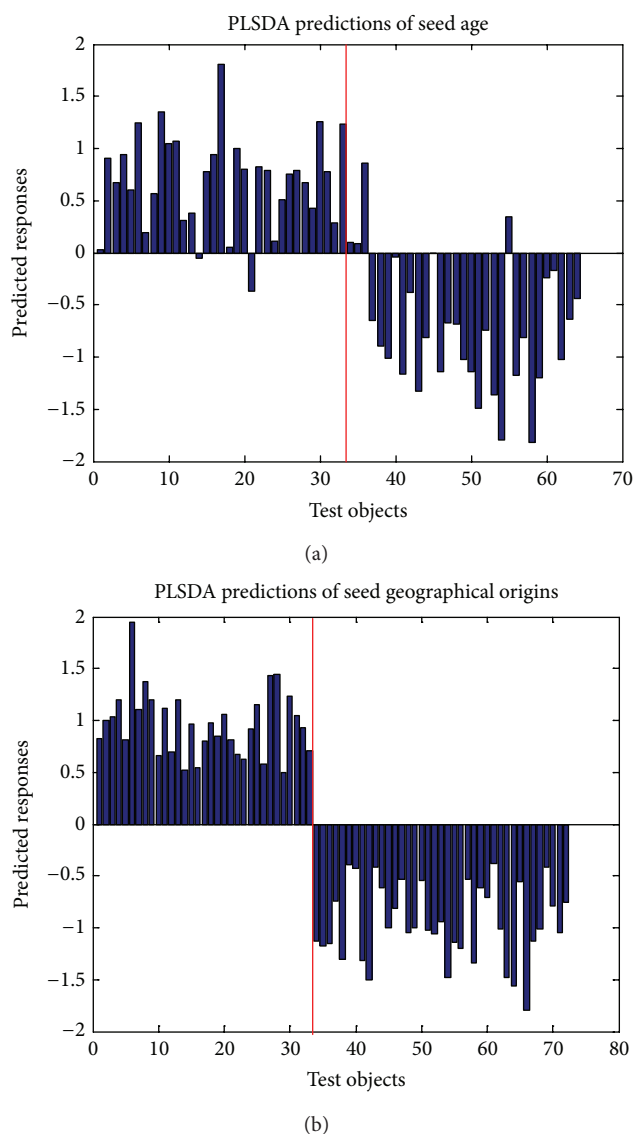
(a)

(b)

FIGURE 6: The predictions of seed age and geographical origins by PLSDA with second derivative spectra; the test objects 1–33 are "positives" (with a response variable of 1 s) and the others are "negatives" (with a response variable of −1 s).

of *T. grandis* seeds. Our future work will be focused on the feasibility of NIR analysis of *T. grandis* oil made of different origins, processing, and raw materials.

Authors' Contribution

L. Xu and S.-M. Yan contributed equally to this work.

Acknowledgments

References

[1] X. Wang, K. Wang, and B. Zhou, "Object-based classification of IKONOS data for endemic *Torreya mapping*," *Procedia Environmental Sciences*, vol. 10, pp. 1887–1891, 2011.

[2] D. C. Hao, P. G. Xiao, B. Huang, G. B. Ge, and L. Yang, "Interspecific relationships and origins of Taxaceae and Cephalotaxaceae revealed by partitioned Bayesian analyses of chloroplast and nuclear DNA sequences," *Plant Systematics and Evolution*, vol. 276, no. 1-2, pp. 89–104, 2008.

[3] B. Chen, X. Cui, X. Zhao et al., "Antioxidative and acute antiinflammatory effects of *Torreya grandis*," *Fitoterapia*, vol. 77, no. 4, pp. 262–267, 2006.

[4] M. K. Saeed, Y. Deng, R. Dai, W. Li, Y. Yu, and Z. Iqbal, "Appraisal of antinociceptive and anti-inflammatory potential of extract and fractions from the leaves of *Torreya grandis* Fort Ex. Lindl," *Journal of Ethnopharmacology*, vol. 127, no. 2, pp. 414–418, 2010.

[5] J. Wang, Y. Huang, M. Fang et al., "Brefeldin A, a cytotoxin produced by *Paecilomyces* sp. and *Aspergillus clavatus* isolated from *Taxus mairei* and *Torreya grandis*," *FEMS Immunology and Medical Microbiology*, vol. 34, no. 1, pp. 51–57, 2002.

[6] B. Galli, F. Gasparrini, V. Lanzotti et al., "Grandione, a new heptacyclic dimeric diterpene from *Torreya grandis* Fort.," *Tetrahedron*, vol. 55, no. 37, pp. 11385–11394, 1999.

[7] Y. Huang, J. Wang, G. Li, Z. Zheng, and W. Su, "Antitumor and antifungal activities in endophytic fungi isolated from pharmaceutical plants *Taxus mairei*, *Cephalataxus fortunei* and *Torreya grandis*," *FEMS Immunology and Medical Microbiology*, vol. 31, no. 2, pp. 163–167, 2001.

[8] Z. J. Li, C. F. Luo, X. J. Cheng, X. J. Feng, and W. W. Yu, "Component analysis and nutrition evaluation of seeds of *Torreya grandis* 'Merrillii'," *Journal of Zhejiang Forestry College*, vol. 22, pp. 540–544, 2005.

[9] R. L. Wolff, F. Pédrono, A. M. Marpeau, W. W. Christie, and F. D. Gunstone, "The seed fatty acid composition and the distribution of Δ5-olefinic acids in the triacylglycerols of some Taxaceae (*Taxus* and *Torreya*)," *Journal of the American Oil Chemists' Society*, vol. 75, no. 11, pp. 1637–1641, 1998.

[10] R. L. Wolff, F. Pédrono, A. M. Marpeau, and F. D. Gunstone, "Seed fatty acid composition and the distribution of Δ5-olefinic acids in the triacylglycerols of some Taxares (*Cephalotaxus* and *Podocarpus*)," *Journal of the American Oil Chemists' Society*, vol. 76, no. 4, pp. 469–473, 1999.

[11] Z. Chen, Z. Chen, L. Hou, Z. Xu, and H. Zheng, "The preventive effect of the oil from the seed of *Torreya grandis* cv. merrillii on experimental atherosclerosis in rats," *Journal of Chinese Medicinal Materials*, vol. 23, no. 9, pp. 551–553, 2000.

[12] Y. Endo, Y. Osada, F. Kimura, and K. Fujimoto, "Effects of Japanese torreya (*Torreya nucifera*) seed oil on lipid metabolism in rats," *Nutrition*, vol. 22, no. 5, pp. 553–558, 2006.

[13] Y. Endo, K. Tsunokake, and I. Ikeda, "Effects of non-methylene-interrupted polyunsaturated fatty acid, sciadonic (All-*cis*-5,11, 14-eicosatrienoic acid) on lipid metabolism in Rats," *Bioscience, Biotechnology and Biochemistry*, vol. 73, no. 3, pp. 577–581, 2009.

[14] L. Song, H. Gao, L. Ge et al., "Effects of different packaging on lipid peroxidation and antioxidant ability of *Torreya grandis* nuts during storage at room temperature," *Scientia Silvae Sinicae*, vol. 45, pp. 49–53, 2009.

[15] F. M. Hu and Z. E. Ding, "Quality studies of three types of Huangshan Chinese Torreya," *Journal of Central South Forestry University*, vol. 23, pp. 1–4, 2003.

[16] L. Ge, H. Gao, H. Chen et al., "Effect of processing on lipid oxidation and antioxidant ability of *Torreya grandis*," *Journal of the Chinese Cereals and Oils Association*, vol. 26, pp. 42–46, 2011.

[17] R. Karoui, G. Downey, and C. Blecker, "Mid-infrared spectroscopy coupled with chemometrics: a tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships—a review," *Chemical Reviews*, vol. 110, no. 10, pp. 6144–6168, 2010.

[18] H. Cen and Y. He, "Theory and application of near infrared reflectance spectroscopy in determination of food quality," *Trends in Food Science and Technology*, vol. 18, no. 2, pp. 72–83, 2007.

[19] R. Hajian and A. Soltaninezhad, "The spectrophotometric multicomponent analysis of a ternary mixture of paracetamol, aspirin, and caffeine by the double divisor-ratio spectra derivative method," *Journal of Spectroscopy*, vol. 2013, Article ID 405210, 7 pages, 2013.

[20] G. Yildiz, R. L. Wehling, and S. L. Cuppett, "Method for determining oxidation of vegetable oils by near-infrared spectroscopy," *Journal of the American Oil Chemists' Society*, vol. 78, no. 5, pp. 495–502, 2001.

[21] G. Knothe, "Analyzing biodiesel: Standards and other methods," *Journal of the American Oil Chemists' Society*, vol. 83, no. 10, pp. 823–833, 2006.

[22] M. Barker and W. Rayens, "Partial least squares for discrimination," *Journal of Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.

[23] M. Hubert, P. Rousseeuw, and T. Verdonck, "Robust PCA for skewed data and its outlier map," *Computational Statistics and Data Analysis*, vol. 53, no. 6, pp. 2264–2274, 2009.

[24] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[25] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Applied Spectroscopy*, vol. 43, no. 5, pp. 772–777, 1989.

[26] R. D. Snee, "Validation of regression models, methods and examples," *Technometrics*, vol. 19, pp. 415–428, 1977.

[27] Q. S. Xu and Y. Z. Liang, "Monte Carlo cross validation," *Chemometrics and Intelligent Laboratory Systems*, vol. 56, pp. 1–11, 2001.