SCIENCE CHINA
Information Sciences

• RESEARCH PAPERS •
Special Focus

March 2011 Vol. 54 No. 3: 598–608
doi: 10.1007/s11432-011-4198-7

Nondestructive determination of nutritional information in oilseed rape leaves using visible/near infrared spectroscopy and multivariate calibrations

LIU Fei, NIE PengCheng, HUANG Min, KONG WenWen & HE Yong*

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029*, China*

Received May 10, 2010; accepted December 20, 2010

Abstract Nitrogen, phosphorus and potassium content are the three most important nutritional parameters for growing oilseed rape. We investigated visible and near infrared (Vis/NIR) spectroscopy combined with chemometrics for the fast and nondestructive determination of nutritional information in oilseed rape leaves. A total of 154 leaf samples were collected, with 104 randomly selected as the calibration set, and the remaining 50 samples used as the validation set. The performance of eight different preprocessing methods was compared in partial least squares (PLS) models. Some effective wavelengths selected by a successive projections algorithm (SPA) were also used to develop linear SPA-PLS, nonlinear back propagation neural network (BPNN), and nonlinear least squares-support vector machine (LS-SVM) models to determine nutritional information. The best prediction models were DOSC-PLS for nitrogen with r=0.9743 and RMSEP=0.1459, DOSC-SPA-BPNN for phosphorus with r=0.7054 and RMSEP=0.0594, and DOSC-SPA-BPNN for potassium with r=0.9380 and RMSEP=0.1788. The prediction precision for nitrogen and potassium determinations was acceptable for further practical applications, whereas further studies are needed to improve the prediction precision for phosphorus. The results indicated that Vis/NIR spectroscopy is feasible for nondestructive determination of nutritional information in oilseed rape leaves. It also provided an alternative technique for detecting other growth information about oilseed rape leaves.

Keywords visible and near infrared spectroscopy, oilseed rape, nutritional information, successive projections algorithm, direct orthogonal signal correction, chemometrics

Citation Liu F, Nie P C, Huang M, et al. Nondestructive determination of nutritional information in oilseed rape leaves using visible/near infrared spectroscopy and multivariate calibrations. Sci China Inf Sci, 2011, 54: 598–608, doi: 10.1007/s11432-011-4198-7

1 Introduction

Oilseed rape (*Brassica napus* L.) is the third most important oilseed crop worldwide. Oilseed rape is widely planted in many areas, and rape oil accounts for about 35% of all edible oil consumed in China [1]. Rapeseed is one of the major sources of edible oils for humans and residual feed for domesticated animals [2, 3]. Obtaining growth information about oilseed rape is beneficial to improve rapeseed yield and quality. Nitrogen, phosphorus and potassium are the most important nutritional factors effecting oilseed

*Corresponding author (email: yhe@zju.edu.cn)

rape growth. Nutritional information affecting growth may be useful to ensure optimal fertilization, as plants would grow well with favorable yield under adequate fertilization. Surplus fertilization causes wasted resources and environmental pollution. Traditional methods for gaining such nutritional information are quite complex. Examples include the Kjeldahl and Dumas combustion methods for nitrogen detection [4], Mo-Sb colorimetry for phosphorus detection [5] and flame photometry for potassium detection [6]. Such methods are costly, time consuming, laborious and unsuitable for rapid detection. Furthermore, these methods are also destructive determination methods unsuitable for field monitoring during the various growing stages of oilseed rape. Hence, a fast nondestructive method is necessary to determine the nutritional information of oilseed rape.

Visible and near infrared (Vis/NIR) spectroscopy has been widely applied in agriculture, food, medicine and other industrial fields requiring fast, nondestructive, low cost and reliable detection methods for both quantitative and qualitative analysis [7]. For oilseed rape growth detection, NIR spectroscopy has been applied to detect chlorophyll in rape leaves [8], nitrogen content of oilseed rape leaf and canopy [9–17], leaf area index [18], amino acid presence [19] and acetolactate synthase (ALS) activity and protein content in oilseed rape leaves [20, 21]. Many reports were focused on the determination of nitrogen content using all wavelengths across the spectrum [9–17]; however, a variable selection process was proposed for nitrogen detection. Furthermore, none considered the fast and nondestructive determination of phosphorus and potassium in oilseed rape using Vis/NIR spectroscopy.

The objectives of the current study were: 1) to study the feasibility of using Vis/NIR spectroscopy (400–1000 nm) to determine nutritional information (nitrogen, phosphorus and potassium) in oilseed rape leaves; 2) to evaluate the performance of different spectral data preprocessing methods, including Savitzky-Golay (SG) smoothing, standard normal variate, multiplicative scatter correction, first-derivative, second-derivative, de-trending and direct orthogonal signal correction (DOSC); and 3) to achieve the best calibration model after comparing the linear partial least squares (PLS), nonlinear back propagation neural network (BPNN) and least squares-support vector machine (LS-SVM) models, using effective wavelengths selected by successive projections algorithm (SPA).

2 Materials and methods

2.1 Sample collection and spectral acquisition

Oilseed rape (*B. napus*, cv. ZS758) is a leading cultivar and was used in this experiment. Field trials were conducted at the farm of Zhejiang University, Hangzhou, China (30° 10'N, 120° 12'E). In a rotational cropping system after the harvest of rice (*Oryza sativa* L.), rapeseeds were sown in early October, 2007. Approximately one month later, the rape seedlings were transplanted to a silt-loam soil, with initial parameters of 0.18% total nitrogen and 1.96% organic matter at a density of 97500 plants/ha. A total of 31 plots (1 reference plot with normal fertilizer, 30 experimental plots) were prepared, and each plot was 2.0 m long and 1.3 m wide. The experiment was prepared in a quadratic orthogonal regression design for fertilizing nitrogen, phosphorus and potassium in the soil. Different amounts of fertilizers were applied in the plot soil to obtain representative leaf samples for a stable and robust calibration model. Herein, the experimental levels refer to the upper and lower levels of applied fertilizers. Table 1 shows the quadratic orthogonal regression design for fertilizing nitrogen, phosphorus and potassium in the soil.

The oilseed rape leaves were collected on two occasions during the seedling stage. For the first sampling (January 15, 2008), two plants from each plot were collected and the functional leaves of each plant were considered as one leaf sample. A total of 62 leaf samples were collected during the first time. For the second sampling (March 21, 2008), three plants from each experimental plot (3×30) and two plants of the reference plot (2×1) were collected, and the functional leaves of each plant were considered as one leaf sample. A total of 92 leaf samples were collected. A single leaf sample actually included several leaves. A total of 154 leaf samples were collected for further analysis. There was no standard method to separate calibration and validation sets. A random selection method with a 2:1 ratio for calibration and validation samples was applied in this experiment. Thus, 104 leaf samples were randomly selected as the

Table 1 Quadratic orthogonal regression design for fertilizing nitrogen, phosphorus and potassium in the soil

Plot No.	Nitrogen	Phosphorus	Potassium
1	1[a]	1	1
2	1	1	−1
3	1	−1	1
4	1	−1	−1
5	−1[b]	1	1
6	−1	1	−1
7	−1	−1	1
8	−1	−1	−1
9	−1.21541[c]	0	0
10	1.21541[d]	0	0
11	0[e]	−1.21541	0
12	0	1.21541	0
13	0	0	−1.21541
14	0	0	1.21541
15	0	0	0

Note. a: 1 means the upper level for fertilizer; b: −1 means lower level for fertilizer; c: −1.21541 means the minimum level for fertilizer; d: 1.21541 means maximum level for fertilizer; e: 0 means normal level for fertilizer.

calibration set, and the remaining 50 samples were used as the validation set. No single sample was used for both calibration and validation.

2.2 Spectral acquisition

A handheld FieldSpecPro FR-Spectroradiometer (325–1075 nm)/A110070 (Analytical Spectral Devices, Boulder, CO USA) was used for the scanning of rapeseed leaf samples. The light source consisted of a Lowell pro-lam interior light source assembly/128930 with Lowell pro-lam 14.5V Bulb/128690 tungsten halogen bulb that could be used in both the visible and NIR regions. The field-of-view (FOV) of the spectroradiometer was 25°. The spectroradiometer was placed at a height of approximately 100 mm and an angle of 45° away from the center of sample. The light source was placed at a height of approximately 250 mm, 45° angle away from the sample. Reflectance spectra from 325 to 1075 nm were measured with an average of 30 scans used for each spectrum. Three spectra were collected for each sample and the average spectrum of these three measurements within the 400–1000 nm wavelength region was used in later analysis. There were 601 wavelength variables from 400–1000 nm with 1 nm intervals for further spectral analysis.

2.3 Nutritional information determination

The reference method for nitrogen content was the Dumas combustion method using Rapid N Cube (Elementar Analysensysteme, Hanau, Germany). Samples were weighed by a four decimal point balance, Sartorius BS224S (Sartorius AG, Goettingen, Germany). After complete combustion, reduction, purification and detection, the nitrogen content of oilseed rape leaves was obtained through the Rapid N Software V 3.4.0 (Elementar Analysensysteme, Hunau, Germany). The reference method for phosphorus content was the Mo-Sb colorimetry method, using an AQ2 Automated Discrete Analyzer (Seal Analytical, UK). The reference method for potassium content was the flame photometer FP-640 (Shanghai Precision & Scientific Instrument Co. Ltd, Shanghai, China). Phosphorus and potassium content were determined by the Hangzhou Centre of Inspection and Testing for Agricultural Products, Ministry of Agriculture, China.

2.4 Spectral preprocessing and variable selection

Preprocessing methods were applied to remove the spectral baseline shift, noise and light scatter influence to achieve better performance [22]. For comparison, various preprocessing methods were investigated,

including SG smoothing, standard normal variate (SNV), multiplicative scatter correction (MSC), first-derivative (1-Der), second-derivative (2-Der), de-trending and DOSC. SG smoothing, SNV and MSC could be used for de-noising, light scatter correction and light path length correction [23, 24]. Derivatives were applied to correct the baseline shift [22]. De-trending attempted to remove nonlinear trends in spectral data [25]. DOSC was a transform of OSC, which could reduce the main variance sources such as temperature effects, time influences and instrumental differences in spectral data [26, 27]. The performance was determined by the prediction results in the calibration stage. DOSC was performed by Matlab v. 7.0 (The Math Works, Natick, MA, USA). Other preprocessing methods were implemented by "The Unscrambler V 9.8" (CAMO AS, Oslo, Norway).

The aforementioned preprocessing methods did not reduce the dimension of input variables (spectral data). Hence, some variable selection methods were necessary to reduce the dimension and complexity, such as SPA. SPA is a newly developed informative variable selection method, which could obtain several relevant variables with least collinearity and redundancies. Herein, the selected relevant variables, named effective wavelengths, were extracted for further applications. Details of the SPA algorithm can be found in the literature [28, 29]. SPA was implemented by Matlab software.

2.5 Calibration models

During calibration, linear and nonlinear calibration methods could be used to handle the relationship between spectral data (X-variables) and chemical responses (Y-variables). PLS analysis was generally used for linear calibrations [30]. BPNN and LS-SVM were investigated for nonlinear calibrations.

For PLS models, the input X-variables (spectral data) were extracted into new eigenvectors (latent variables, LVs) to present the most relevant information of the original spectra. During the calibration stage, a full cross-validation process was used to ensure a stable and robust PLS model. The prediction performance was evaluated by the samples in the validation set.

The basic error BPNN is one of the most popular neural network topologies. The model structure of BPNN models was input layer-hidden layer-output layer. The selected effective wavelengths by SPA were used as the input layer. The nutritional parameters were considered as the output layer. The network output expresses the similarity that an object corresponds with a training pattern. Along with every process of a training pattern and adjustment of the weight factors, the difference between the desired and calculated network output value, defined as the network output error, will gradually become less and less until it meets the desired value. The sigmoid function was used as the transfer function. The sigmoid function is a simpler transfer function and could give a quick feedback result during the computation process. It is also the standard used transfer function for BPNN models. Details for the BPNN algorithm can be found in the literature [31].

LS-SVM employed a set of linear equations using support vectors instead of quadratic programming problems to reduce the complexity of optimization processes. The details of LS-SVM can be found in previous studies [21, 32, 33]. In LS-SVM models, the input variables were settled by effective wavelengths selected by SPA. The radials basis function (RBF) kernel was recommended as the kernel function because it can reduce the computational complexity, handle nonlinear correlations and give a good performance under general smoothness assumptions. The model parameters of $gamma(\gamma)$ and $sigma^2$ (σ^2) were settled by a two-step grid search technique with leave-one-out cross-validation. The free LS-SVM toolbox (LS-SVM V1.5, Suykens, Leuven, Belgium) was applied to develop LS-SVM models. The LS-SVM algorithm could be performed by Matlab.

The prediction capability was mainly assessed by the following indices: correlation coefficients of calibration (R) and validation (r), root mean squares error (RMSE) of calibration (RMSEC) and validation (RMSEP). The RMSE was calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}},$$

where n is the number of samples, and y_i and $\hat{y}_i$ are the reference and predicted values of the ith sample, respectively.

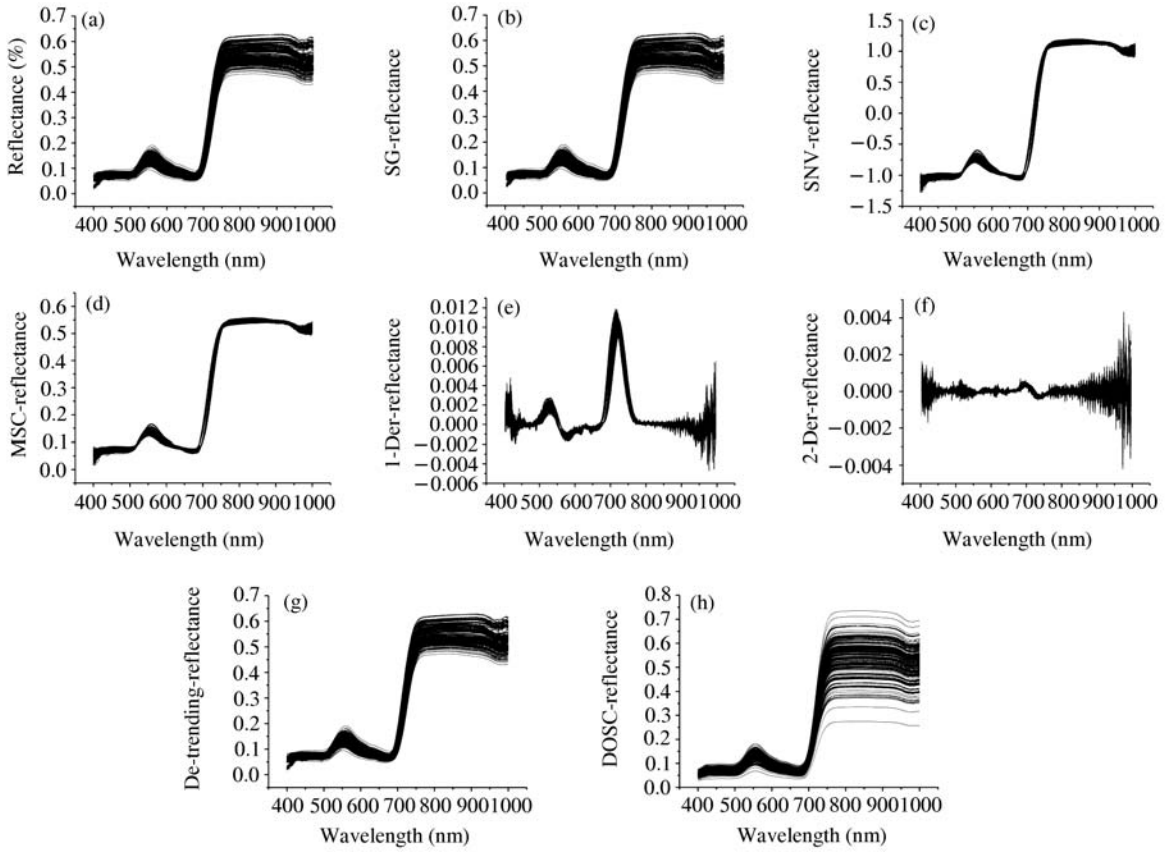

Figure 1 The (a) raw reflectance spectra, and preprocessed reflectance spectra of oilseed rape leaves by (b) SG, (c) SNV, (d) MSC, (e) 1-Der, (f) 2-Der, (g) De-trending and (h) DOSC.

3 Results and discussion

3.1 Spectral features of oilseed rape leaves

The raw and preprocessed reflectance spectra of oilseed rape leaves are shown in Figure 1. The raw reflectance spectra of oilseed rape leaves are shown in Figure 1(a), and reflectance values were elevated at around 550 nm and 750–1000 nm. Spectra preprocessed by SG, SNV and MSC in Figure 1(b)–(d) retained the main features of the raw spectra (Figure 1(a)). The spectral curve in DOSC spectra (Figure 1(h)) enhanced the variation in the 750–1000 nm region, which may have been caused by different contents of nutritional parameters.

3.2 Statistical values of nutritional information

Statistics of nutritional information in oilseed rape leaves are shown in Table 2. The values of nitrogen, phosphorus and potassium content in the calibration set covered a larger range than those in the validation set, which was helpful to develop stable and general models.

3.3 PLS models

Different PLS models were developed using the aforementioned preprocessing methods. The cross-validation process was applied to avoid the over-fitting problem and to evaluate model performance during the calibration stage. Different latent variables were selected for the optimal PLS models. The samples in the validation set were used to assess the prediction performance of the developed models. According to the aforementioned evaluation standards, such as r and RMSEP, the optimal preprocessing method could be achieved. The prediction results in calibration and validation sets by PLS models are

Table 2 Statistical values of nitrogen, phosphorus and potassium content in the calibration and validation sets

Nutritional parameters	Data set	Sample number	Minimum value of nutritional content (%)	Maximum value of nutritional content (%)	Mean value of nutritional content (%)	Standard deviation of nutritional content
Nitrogen	Calibration	104	2.961	6.777	5.033	0.705
	Validation	50	3.524	6.197	5.013	0.634
	All	154	2.961	6.777	5.027	0.698
Phosphorus	Calibration	104	0.500	1.600	0.698	0.138
	Validation	50	0.550	0.890	0.682	0.082
	All	154	0.50	1.600	0.693	0.123
Potassium	Calibration	104	0.790	3.280	1.869	0.562
	Validation	50	0.860	2.880	1.843	0.517
	All	154	0.790	3.280	1.861	0.546

shown in Table 3. As stated in Table 3, the optimal performance in the validation set was achieved by DOSC spectra with r=0.9743 and RMSEP=0.1459 for nitrogen, r=0.6971 and RMSEP=0.0632 for phosphorus, and r=0.9316 and RMSEP=0.865 for potassium.

The result for nitrogen and potassium was acceptable, considering the prediction performance within validation set. However, the results for phosphorus were insufficient for practical applications. The basis for acceptable precision was made from experiential conclusions. If prediction results had a correlation coefficient higher than 0.8, prediction precision was deemed acceptable and sufficient for further practical applications. If the correlation coefficient was lower than 0.8, more work was deemed necessary to improve the prediction performance for quantitative analysis. Furthermore, 601 variables were employed as inputs in the PLS models, and there may have been collinearity and redundancy among these variables. To develop a more parsimonious model, a relevant variable selection procedure should be performed, and the successive projections algorithm is recommended for this purpose.

3.4 Effective wavelength selection by SPA

To allow a comprehensive comparison, the spectra obtained from the optimal two preprocessing methods were used for SPA calculation. Hence, the spectra after MSC and DOSC were used for nitrogen, those after 1-Der and DOSC were used for phosphorus, and those after 1-Der and DOSC were used for potassium.

During the SPA, the maximum number of selected variables was set at 30, and cross-validation was applied in the selection process. Different wavelengths were selected by SPA (shown in Table 4). As shown in Table 4, selected effective wavelengths were ranked in the order of importance. For example, 638 nm was the most important variable within all selected effective wavelengths for MSC spectra. These effective wavelengths were applied as inputs of the BPNN and LS-SVM models, with these models then employed to determine the nutritional information.

3.5 PLS, BPNN and LS-SVM models

After the SPA process, the selected effective wavelengths were used as input variables to develop the PLS, BPNN and LS-SVM models.

In the SPA-PLS models, cross-validation was also applied and different LVs were extracted for a robust performance. The prediction results for nitrogen, phosphorus and potassium in calibration and validation sets are shown in Table 5. The optimal prediction results were obtained by DOSC-SPA-PLS for nitrogen (r=0.9436 and RMSEP=0.1477), phosphorus (r=0.6940 and RMSEP=0.0635) and potassium (r=0.9307 and RMSEP=0.1875). The results indicated that DOSC is a powerful preprocessing method for improving prediction performance in PLS models.

In the SPA-BPNN models, the transfer function was a sigmoid function. Three-layer BPNN models were achieved for nitrogen, phosphorus and potassium content determination. The nodes of the input layer, the hidden layers and the output layer were of the structure 15-8-1 for nitrogen, 4-3-1 for phosphorus, and 6-4-1 for potassium. The dynamic parameter was set as 0.6 which was determined after several trials

Table 3 Prediction results for nitrogen, phosphorus and potassium by PLS using different preprocessing methods

Nutritional	Preprocessing	LVs	Calibration		Validation	
			R	RMSEC	r	RMSEP
Nitrogen	Raw	8	0.9083	0.3008	0.7372	0.4422
	SG	8	0.9022	0.3101	0.7348	0.4434
	SNV	6	0.9133	0.2928	0.7516	0.4290
	MSC	6	0.9134	0.2927	0.7536	0.4272
	De-trending	9	0.9339	0.2571	0.7259	0.4698
	1-Der	6	0.9094	0.2990	0.7455	0.4400
	2-Der	6	0.8604	0.3665	0.6534	0.5043
	DOSC	1	0.9786	0.1480	0.9743	0.1459
Phosphorus	Raw	1	0.2955	0.1311	0.2219	0.0843
	SG	1	0.2957	0.1311	0.2225	0.0843
	SNV	3	0.3940	0.1261	0.1953	0.0902
	MSC	3	0.3940	0.1261	0.1953	0.0902
	De-trending	1	0.3180	0.1301	0.2159	0.0828
	1-Der	3	0.4860	0.1199	0.4091	0.0866
	2-Der	1	0.3348	0.1293	0.2716	0.0820
	DOSC	1	0.7784	0.0861	0.6971	0.0632
Potassium	Raw	5	0.6315	0.4339	0.6808	0.3760
	SG	5	0.6309	0.4341	0.6887	0.3721
	SNV	14	0.8771	0.2688	0.6777	0.4020
	MSC	14	0.8767	0.2692	0.6834	0.3975
	De-trending	2	0.5791	0.4562	0.6676	0.3819
	1-Der	3	0.6648	0.4180	0.7212	0.3546
	2-Der	9	0.8838	0.2618	0.6673	0.4246
	DOSC	1	0.9221	0.2165	0.9316	0.1865

Table 4 Selected effective wavelengths (EWs) by SPA

Nutritional	Preprocessing	No.	Selected EWs (nm)
Nitrogen	MSC	15	638, 414, 465, 431, 970, 954, 959, 747, 424, 514, 915, 450, 673, 940, 935
	DOSC	1	958
Phosphorus	1-Der	4	502, 952, 955, 783
	DOSC	1	958
Potassium	1-Der	6	579, 950, 481, 405, 453, 462
	DOSC	1	627

in the region of 0.6–0.9. The least training speed and the parameter of *sigmoid* were set as default values of 0.1 and 0.9, respectively. The residual error was set at 0.00001. The times of training was set as 3000. The prediction results for nitrogen, phosphorus and potassium in calibration and validation sets are shown in Table 5. The optimal prediction results were obtained by DOSC-SPA-BPNN for nitrogen (r=0.9682 and RMSEP=0.3716), phosphorus (r=0.7054 and RMSEP=0.0594) and potassium (r=0.9380 and RMSEP=0.1788). These results indicated that the DOSC-SPA process was better than those of MSC-SPA and 1-Der-SPA, which was consistent with those obtained from the SPA-PLS models.

In the SPA-LS-SVM models, the kernel function was RBF kernel. The model parameters (γ, σ^2) were determined by a two-step grid search and cross-validation process. The search region of (γ, σ^2) was set as (10^{-4}–10^4) for nitrogen, (10^{-3}–10^3) for phosphorus, and (10^{-3}–10^3) for potassium. After computation, the optimal combination of (γ, σ^2) was achieved for different nutritional parameters. The prediction result for nitrogen, phosphorus and potassium in calibration and validation sets are shown in Table 5. The optimal prediction results were obtained by DOSC-SPA-LS-SVM for nitrogen (r=0.9737 and RMSEP=0.1472), phosphorus (r=0.6745 and RMSEP=0.0606) and potassium (r=0.9351 and RMSEP=

Table 5 Prediction results for nitrogen, phosphorus and potassium by PLS, BPNN and LS-SVM models using effective wavelengths (EWs) selected by SPA

Nutritional	Preprocessing	Models	LVs/EWs/(γ,σ^2)	Calibration		Validation	
				R	RMSEC	r	RMSEP
Nitrogen	MSC	SPA-PLS	9/15/-	0.9400	0.2453	0.7247	0.4879
		SPA-BPNN	-/15/-	0.9401	0.3288	0.7219	0.5327
		SPA-LS-SVM	-/15/(9.9×10^4,4.9×10^4)	0.9434	0.2382	0.7450	0.4621
	DOSC	SPA-PLS	1/1/-	0.9787	0.1475	0.9736	0.1477
		SPA-BPNN	-/1/-	0.9593	0.3884	0.9682	0.3716
		SPA-LS-SVM	-/1/(50.3, 78.8)	0.9795	0.1447	0.9737	0.1472
Phosphorus	1-Der	SPA-PLS	2/4/-	0.4084	0.1253	0.4026	0.0849
		SPA-BPNN	-/4/-	0.4330	0.1239	0.4263	0.0860
		SPA-LS-SVM	-/4/(0.4,11.4)	0.4796	0.1219	0.4445	0.0752
	DOSC	SPA-PLS	1/1/-	0.7788	0.0861	0.6940	0.0635
		SPA-BPNN	-/1/-	0.7255	0.0962	0.7054	0.0594
		SPA-LS-SVM	-/1/7.4×10^3, 433.4	0.8417	0.0741	0.6745	0.0606
Potassium	1-Der	SPA-PLS	3/6/-	0.6984	0.4005	0.6432	0.4023
		SPA-BPNN	-/6/-	0.7581	0.3777	0.6362	0.4433
		SPA-LS-SVM	-/6/(10.1,189.8)	0.7449	0.3762	0.6512	0.3964
	DOSC	SPA-PLS	1/1/-	0.9223	0.2162	0.9307	0.1875
		SPA-BPNN	-/1/-	0.9237	0.2149	0.9380	0.1788
		SPA-LS-SVM	-/1/(8.3,27.8)	0.9242	0.2139	0.9351	0.1818

Table 6 The optimal models for nitrogen, phosphorus and potassium from different wavelength regions

Nutritional	Models	No. of wavelengths	Calibration		Validation	
			R	RMSEC	r	RMSEP
Nitrogen	DOSC-PLS	601	0.9786	0.1480	0.9743	0.1459
	MSC-SPA-LS-SVM	15	0.9434	0.2382	0.7450	0.4621
	DOSC-SPA-LS-SVM	1	0.9795	0.1447	0.9737	0.1472
Phosphorus	DOSC-PLS	601	0.7784	0.0861	0.6971	0.0632
	1-Der-SPA-LS-SVM	4	0.4796	0.1219	0.4445	0.0752
	DOSC-SPA-BPNN	1	0.7255	0.0962	0.7054	0.0594
Potassium	DOSC-PLS	601	0.9221	0.2165	0.9316	0.1865
	1-Der-SPA-LS-SVM	6	0.7449	0.3762	0.6512	0.3964
	DOSC-SPA-BPNN	1	0.9237	0.2149	0.9380	0.1788

0.1818). The results indicated that DOSC-SPA was better than MSC-SPA and 1-Der-SPA, which was consistent with the results obtained by the SPA-PLS and SPA-LS-SVM models.

Comparing all PLS, BPNN and LS-SVM models using EWs selected by SPA, the best prediction results were DOSC-SPA-LS-SVM for nitrogen and DOSC-SPA-BPNN for both phosphorus and potassium. The BPNN and LS-SVM models achieved a better performance than PLS models, a reason for which may have been that BPNN and LS-SVM made good use of the latent nonlinear information in the selected effective wavelengths. In contrast, PLS only developed the linear relationship between the selected effective wavelengths and nutritional information. DOSC outperformed MSC and 1-Der, a reason for which may have been that DOSC took the Y-variable into consideration in the calibration set during preprocessing.

3.6 Model comparison

Comparing all the developed models by PLS, BPNN and LS-SVM using a different number of input variables, the best models were obtained by DOSC-PLS using 601 variables for nitrogen, DOSC-SPA-BPNN using 1 EW for phosphorus, and DOSC-SPA-BPNN using 1 EW for potassium (shown in Table 6). The best prediction results were r=0.9743 and RMSEP=0.1459 for nitrogen, r=0.7054 and RMSEP=0.0594

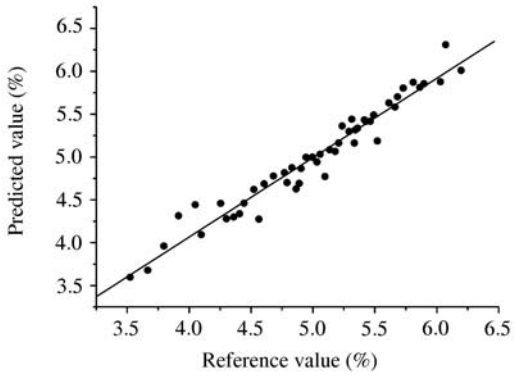

Figure 2 The best prediction results for nitrogen by the DOSC-PLS model.

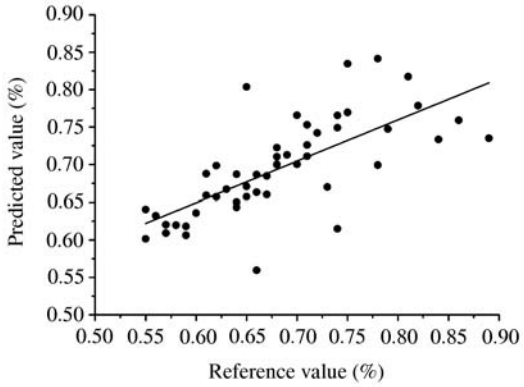

Figure 3 The best prediction results for phosphorus by the DOSC-SPA-BPNN model.

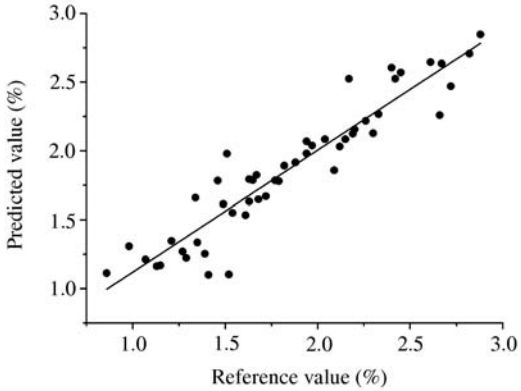

Figure 4 The best prediction results for potassium by the DOSC-SPA-BPNN model.

for phosphorus, and r=0.9380 and RMSEP=0.1788 for potassium. Good prediction results for nitrogen and phosphorus content in this study may have been due to the spectral preprocessing method (DOSC) and informative variable selection method (SPA) used for the multivariate calibration models. DOSC and SPA enhanced the most relevant information and effective wavelengths in the spectral data, which made the calibration model more powerful and robust. The scatter plots in the validation set with best prediction results are shown in Figures 2, 3 and 4 for nitrogen, phosphorus and potassium, respectively. The results for nitrogen and potassium were adequate for further practical applications, whereas more studies should be done to improve the prediction precision for phosphorus.

4 Conclusions

Vis/NIR spectroscopy combined with chemometrics was applied for the nondestructive determination of nutritional information (nitrogen, phosphorus and potassium content) in oilseed rape leaves. The performance of different preprocessing methods were compared in PLS models. Effective wavelengths selected by SPA were used to develop SPA-PLS, SPA-BPNN and SPA-LS-SVM models and determine nutritional information. The best prediction model was DOSC-PLS for nitrogen with r=0.9743 and RMSEP=0.1459, DOSC-SPA-BPNN for phosphorus with r=0.7054 and RMSEP=0.0594, and DOSC-SPA-BPNN for potassium with r=0.9380 and RMSEP=0.1788. The prediction precision for nitrogen and potassium was acceptable for further practical applications, whereas that for phosphorus was insufficient and further work is necessary. The results indicated that Vis/NIR spectroscopy is feasible for the nondestructive determination of nutritional information in oilseed rape leaves. It also provided an alternative technique for detecting other growth information about oilseed rape leaves.

Acknowledgements

This work was supported by the National High-Tech Research & Development Program of China (Grant No. 2007AA10Z210), the National Natural Science Foundation of China (Grant No. 31071332), the Zhejiang Innovation Program (Grant Nos. 2009R50001, YK2008014), the Zhejiang Provincial Natural Science Foundation (Grant No. Z3090295), the Agricultural Science and Technology Achievements Transformation Fund Programs (Grant No. 2009GB23600517) and the Fundamental Research Funds for the Central Universities.

References

1 Li P W, Yang M, Zhang W, et al. Studies on quality of oilseed rape products and its improvement strategy in China. Chin J Oil Crop Sci, 2004, 26: 84–88
2 Fu T D, Yang G S, Tu J X, et al. The present and future of rapeseed production in China. China Oils Fats, 2003, 28: 11–13
3 Li W L, Huang F H, Peng J, et al. Development and feeding study of dehulled double-low rapeseed meal. China Oils Fats, 2006, 31: 51–54
4 Simonne A H, Simonne E H, Eitenmiller R R, et al. Could the Dumas method replace the Kjeldahl digestion for nitrogen and crude protein determinations in foods? J Sci Food Agr, 1997, 73: 39–45
5 Lu C. Comparative study on two methods for determination of total phosphorus in wetland plants. Acta Agr Jiangxi, 2009, 22: 142–144
6 Fox Jr. C L. Stable internal standard flame photometer for potassium and sodium analyses. Anal Chem, 1951, 23: 137–142
7 Yan Y L, Zhao L L, Han D H, et al. The Foundation and Application of Near-infrared Spectroscopy Analysis. Beijing: China Light Industry Press, 2005
8 Fang H, Song H Y, Cao F, et al. Study on the relationship between spectral properties of oilseed rape leaves and their chlorophyll content. Spectrosc Spectr Anal, 2007, 27: 1731–1734
9 Wang Y, Huang J F, Wang F M, et al. Predicting nitrogen concentrations from hyperspectral reflectance at leaf and canopy for rape. Spectrosc Spectr Anal, 2008, 28: 273–277
10 Yi Q X, Huang J F, Wang F M, et al. Monitoring rice nitrogen status using hyperspectral reflectance and artificial neural network. Environ Sci Tech, 2007, 41: 6770–6775
11 Müller K, Böttcher U, Meyer-Schatz F, et al. Analysis of vegetation indices derived from hyperspectral reflection measurements for estimating crop canopy parameters of oilseed rape (Brassica napus L.). Biosyst Eng, 2008, 101: 172–182
12 Zhang Y, Mao H P, Zhang X D, et al. Nitrogen information measurement of canola leaves based on multispectral vision. J Agr Mech Res, 2009, 11: 83–85
13 Zhang X D, Mao H P, Cheng X H. Rape nitrogen content spectral character models based on PCA-SVR method. Trans Chinese Society Agr Mach, 2009, 40: 161–165
14 Zhang X H, Liu S M, He B B. Hyperspectral evaluation of rape nitrogen nutrition using ontinuum-removed method. Trans CSAE, 2008, 24: 151–155
15 Zhang X H, Liu S M, He B B. Analysis on hyperspectral characteristics of rape at different nitrogen leaves. J Beijing Normal Univ (Nat Sci), 2007, 43: 245–249
16 Qiu Z J, Song H Y, He Y, et al. Variation rules of the nitrogen content of the oilseed rape at growth stage using SPAD and visible-NIR. Trans CASE, 2007, 23: 150–154
17 Feng L, Fang H, Zhou W J, et al. Nitrogen stress measurement of canola based on multi-spectral charged coupled device imaging sensor. Spectrosc Spectr Anal, 2006, 26: 1749–1752
18 Huang J F, Wang Y, Wang F M, et al. Red edge characteristics and leaf area index estimation model using hyperspectral data for rape. Trans CSAE, 2006, 22: 22–26
19 Liu F, Zhang F, Fang H, et al. Application of successive projections algorithm for nondestructive determination of total amino acids in oilseed rape leaves. Spectrosc Spectr Anal, 2009, 29: 3079–3083
20 Liu F, Fang H, Zhang F, et al. Nondestructive determination of acetolactate synthase in oilseed rape leaves using visible and near infrared spectroscopy. Chin J Anal Chem, 2009, 37: 67–71
21 Liu F, Zhang F, Jin Z L, et al. Determination of acetolactate synthase activity and protein content of oilseed rape (Brassica napus L.) leaves using visible/near-infrared spectroscopy. Anal Chim Acta, 2008, 629: 56–65
22 Chu X L, Yuan H F, Lu W Z. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique. Prog Chem, 2004, 16: 528–542
23 Savitzky A, Golay M J E. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem, 1964, 36: 1627–1639

24 Dhanoa M S, Lister S J, Sandersona R, et al. The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra. J Near Infrared Spectrosc, 1994, 2: 43–47
25 Barnes R J, Dhanoa M S, Lister S J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl Spectrosc, 1989, 43: 772–777
26 Westerhuis J A, De Jong S, Smilde A K. Direct orthogonal signal correction. Chemom Intell Lab Syst, 2001, 56: 13–25
27 Zhu D Z, Ji B P, Meng C Y, et al. The application of direct orthogonal signal correction for linear and non-linear multivariate calibration. Chemom Intell Lab Syst, 2008, 90: 108–115
28 Araújo M C U, Saldanha T C B, Galvão R K H, et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemom Intell Lab Syst. 2001, 57: 65–73
28 Galvão R K H, Araújo M C U, Fragoso W D, et al. A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm. Chemom Intell Lab Syst, 2008, 92: 83–91
30 Geladi P, Kowalski B R. Partial least-squares regression: a tutorial. Anal Chim Acta, 1986, 185: 1–17
31 He Y, Feng S J, Deng X F, et al. Study on lossless discrimination of varieties of yogurt using the Visible/NIR-spectroscopy. Food Res Int, 2006, 39: 645–650
32 Suykens J A K, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett, 1999, 9: 293–300
33 Suykens J A K, Van Gestel T, De Brabanter J, et al. Least Squares Support Vector Machines. Singapore: World Scientific, 2002