XVII[th] World Congress of the International Commission of Agricultural and Biosystems Engineering (CIGR)

Hosted by the Canadian Society for Bioengineering (CSBE/SCGAB)
Québec City, Canada June 13-17, 2010

COMBINATION OF NEAR INFRARED SPECTROSCOPY AND MULTIVARIATE ANALYSIS TO DETECT VALINE ($C_5H_{11}NO_2$) IN OILSEED RAPE LEAVES

WENWEN KONG[1], FEI LIU, HUI FANG, YONG HE

[1]College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China (Corresponding author: Yong He, Email: yhe@zju.edu.cn)

CSBE101385 – Presented at Section VI: Postharvest Technology and Process Engineering Conference

ABSTRACT Valine ($C_5H_{11}NO_2$) is one of the most important amino acids in oilseed rape and has a close relationship with the influence of herbicide stress during growing stage. The traditional methods for amino acid detection were to use high performance liquid chromatography (HPLC) or amino acid analyzer. These methods were time consuming for preparation and detection, costly, laborious and not convenient for fast determination during each growing status of oilseed rape. Therefore, near infrared (NIR) spectroscopy (1100-2500 nm) combined with multivariate calibrations was investigated to determine valine in oilseed rape leaves under herbicide stress. The calibration set consisted of 186 leave samples, and the validation set consisted of 62 samples. Different preprocessing methods were compared, including multiplicative scatter correction (MSC), first-derivative (1-Der), second-derivative (2-Der), de-trending and direct orthogonal signal correction (DOSC). The optimal performance was determined by the PLS model with correlation coefficients (*r*) and root mean squares error of prediction (RMSEP). The optimal PLS model was achieved by DOSC spectra with *r*=0.9822 and RMSEP=0.0379. Successive projections algorithm (SPA) was recommended for variable selection procedure. After computation, the number of selected variables was 15, 17 and 1 for 1-Der, de-trending and DOSC spectra, respectively. The optimal prediction results was achieved by SPA-LS-SVM model (De-trending spectra) with *r*=0.9864 and RMSEP=0.0327. This result was slightly better than PLS model with DOSC spectra. The results indicated that near infrared spectroscopy combined with SPA-LS-SVM was successfully applied for the determination of valine in oilseed rape leaves under herbicide stress.

Keywords: Near infrared spectroscopy, Oilseed rape, Valine, Successive projections algorithm, Least squares-support vector machine

INTRODUCTION Oilseed rape (Brassica napus L.) is widely grown in China, and it is also one of the most important oil crops. It plays an important role in agricultural production in China. Nowadays, some herbicide would be employed to keep a good environment for the growing of oilseed rape. A novel herbicide, named propyl 4-(2-(4,6-dimethoxypyrimidin -2-yloxy)benzylamino)benzoate (ZJ0273) was an ALS (acetolactate synthase)-inhibiting herbicide, which was powerful for the weed control during the

oilseed rape growing stages. Valine ($C_5H_{11}NO_2$) is one of the most important branch-chain amino acids in oilseed rape, which has a quite close relationship with the influence of herbicide stress during growing stage (Zhao et al., 2009). The traditional methods for amino acid detection were using high performance liquid chromatography (HPLC) or amino acid analyzer. These methods were time consuming for preparation and detection, costly, laborious and not convenient for fast determination during each growing status of oilseed rape.

Near infrared reflectance spectroscopy (NIRS) is a well-establish technique for constituent analysis, because it has many advanced characteristic such as high efficiency, low cost, quick data analysis, better repeatability and simple measuring. It has spread to agriculture, food, cosmetics and other industries for both quantitative and qualitative analysis (Yan et al., 2005). In the application of oilseed rape, NIR spectroscopic techniques had been applied for the determination of chlorophyll of rape leaves (Fang et al., 2007), the determination of acetolactate synthase (ALS) and protein content of oilseed rape leaves using visible/near infrared (400-100 nm) spectra (Liu et al., 2008a; Liu et al., 2009a), total amino acids in oilseed rape leaves (Liu et al., 2009b). However, there were few studies on the noninvasive determination of branch-chain amino acid (valine) using near infrared spectroscopy combined with multivariate calibrations.

The aim of this study was to study the feasibility of near infrared spectroscopy to determine the valine in oilseed rape leaves under herbicide stress. In this paper, the performance of six different preprocessing and two different calibrations (PLS and LS-SVM) were compared. The successive projections algorithm (SPA) was applied for relevant variable selection.

MATERIALS AND METHODS

Sample preparation The variety of *Brassica napus*, cv. ZS758 was grown at the farm of Zhejiang University, Hangzhou (30° 10´N, 120° 12´E). The new herbicide ZJ0273 with various concentrations was applied to the oilseed rape leaves including 0, 100, 200 and 500 mg/L. The growing stage was 5-leaf stage and the quantity of ZJ0273 was 500 L/ha. Conventional crop management was used during the growing period. A total of 248 samples were collected by three times, and 80, 80, 88 samples were collected at each time. 186 samples including each ZJ0273 concentration were selected randomly for calibration set, and the remaining 62 samples were used as the validation set. The samples in calibration and validation sets were randomly changed several times to confirm the randomization. No single sample was used in both calibration set and validation set at the same time.

Spectral acquisition and reference method for valine Foss NIRSystems 5000 (Foss NIRSystems, Denmark) within the wavelength region 1100-2500 nm were applied for the reflectance spectral acquisition of all samples. The spectral collection software was WINISI Ⅱ V1.5. The small round cup was used for sample container. A total of 700 data points were scanned with the instrument resolution of 2 nm. Three spectra were stored for each sample, and the averaged spectrum were used as the spectral data for this sample. All spectral data were stored in personal computer for further treatment.

The protocol for amino acid extraction and activity assay was based on the Lisiewska method with some modifications (Lisiewska et al., 2008). The instrument for the determination of valine was HITACHI amino acid analyzer (Model: L-8900, Japan). All detections were carried out in two replications for each sample and the averaged value was used as reference value. The content of amino acids was expressed as mg/100 mg of dry matter (mg/100 mg DW) of treated plants.

Spectral preprocessing and variable selection Different spectral processing methods were compared for the better performance of calibrations. Some preprocessing methods would remove the spectral baseline shift, noise and light scatter influence (Chu et al., 2004). The absorbance spectra were used in the computation by the transformation of log(1/R) (R=reflectance). For comparison, the following preprocessing methods were calculated, including Savitzky-Golay smoothing (SG), standard normal variate (SNV), multiplicative scatter correction (MSC), first-derivative (1-Der), second-derivative (2-Der), de-trending and direct orthogonal signal correction (DOSC). The performance was determined by the prediction results in the latter calibration stage. The pretreatments were implemented by the Unscrambler®9.8 (CAMO AS, Oslo, Norway) and Matlab®7.0 (The Math Works, Natick, MA, USA).

As known, some collinearity and redundancies existed in the full spectral region. Some relevant variable selection method would be needed to select the most informative variables with least collinearity and redundancies. Successive projections algorithm (SPA) was a newly proposed relevant variable selection method (Araújo et al., 2001; Galvão et al., 2008). It could use the projection procedure to select the most relevant variable with least collinearity and redundancies. The selected variable, named effective wavelengths (EWs), could be applied as input for the development of a more parsimonious model.

Multivariate calibration methods In this study, linear and nonlinear calibration methods were both taken into consideration for a better prediction performance comparison. Partial least squares (PLS) analysis was used as the linear calibration, whereas least squares-support vector machine was utilized as the nonlinear calibration.

PLS is the most widely applied calibration method for modeling in the application of NIR spectroscopic techniques. Some new eigenvectors, called latent variables (LVs), were extracted during the computation, and used as the direct inputs of PLS models. Full cross-validation procedure was performed to test the model development stage.

LS-SVM was a powerful calibration method to handle linear and nonlinear problems with a good statistical basis (Suykens et al., 1999). The details of LS-SVM could be found in the literatures (Suykens et al., 1999; Liu et al., 2008b). Three problems should be settled before the LS-SVM model development. Firstly, the input variables of LS-SVM were settled by the selected EWs by SPA with different preprocessing methods. Secondly, the radial basis function (RBF) was recommended as kernel function. Thirdly, the model parameters *gamma* (γ) and *sig2* (σ^2) were determined by a two-step grid search technique. All the calculations were performed using Matlab®7.0 (The Math Works, Natick, MA, USA). The free LS-SVM v 1.5 toolbox (Suykens, Leuven, Belgium) was applied with MATLAB software v. 7.0 to develop the LS-SVM models.

The evaluation standards were correlation coefficients (*r*), root mean squares error of prediction (RMSEP), bias, slope and offset. The main indices in this paper were *r* and RMSEP. The good model should be with higher *r* value and lower RMSEP, absolute bias and offset values, and the slope of the regression line should be closer to 1.

RESULTS AND DISCUSSION

Absorbance spectral investigation The typical near infrared absorbance spectra of oilseed rape leaves under herbicide stress was shown in Fig.1. In this work, the spectra with wavelength range from 1100 to 2500 nm were used. It can be found that the trends of all samples were quite similar. It was noticed that there were many absorbance peaks in Fig.1. It might be the influence of chemical compositions with C-H, N-H or O-H bands.

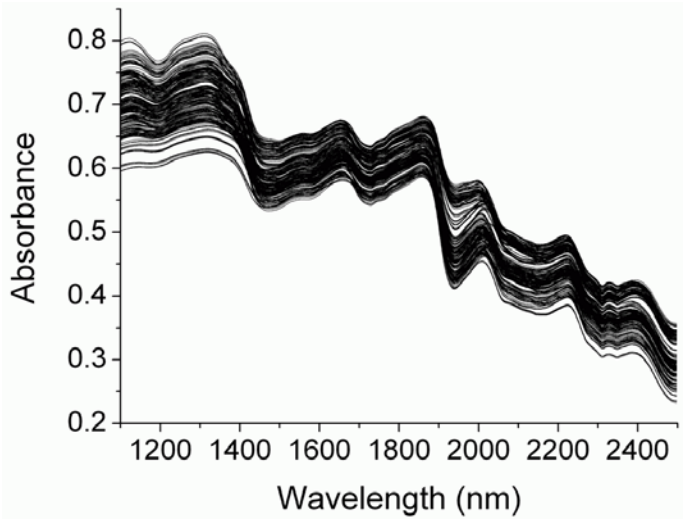

Fig. 1. The raw absorbance spectra of oilseed rape

Table 1 The statistics of valine of oilseed rape leaves (mg/100mg DW)

Set	No.	Range	Mean	S.D.
Cal.	186	0.466-1.441	0.930	0.2002
Val	62	0.473-1.412	0.928	0.2007

Comparison of PLS models using different pre-processing With different pre-processing methods, six PLS calibration models were built with different latent variables. Avoiding overfitting problem and validate the model performance, the cross-validation was applied during the calibration stage. The quality of the calibration model was quantified by the correlation coefficients (*r*) and root mean squares error of prediction (RMSEP). The prediction results by PLS models are shown in Table 2. The *r* was above 0.9 in all models. The optimal PLS model was achieved by DOSC spectra with r = 0.9822 and RMSEP = 0.0379. Then the following PLS models were de-trending and first-derivative (1-Der) spectra based model. These three models were considered as optimal models. These results only considered the prediction performance. But the inputs of these PLS models included more than 700 variables, and there might be much collinearity and redundancy between these wavelengths. Therefore a relevant variable selection procedure was valid for building a more parsimonious model, successive projections algorithm was recommended for such purpose.

Effective wavelengths selection The model with the effective wavelength was better than the model with the full wavelength range. DOSC, de-trending and first-derivative (1-Der) spectra based PLS models were considered as optimal models. Therefore these three preprocessing methods were applied in the SPA. The maximum number of selected variable was set as 30 in the SPA, and cross-validation was also applied in the selection process. After computation, the number of selected variables was 1, 17 and 15 for DOSC, de-trending and 1-Der spectra, respectively. The effective wavelengths which were selected by SPA were shown in Table 3. These wavelengths were ranked in the order of importance. Then a new combinational SPA-LS-SVM calibration was developed using the selected EWs by SPA (Liu et al., 2009c). Then SPA-LS-SVM models would be applied for the determination of valine in oilseed rape leaves under herbicide stress.

Table 2 The prediction results by PLS models

Preprocessing	LVs	*r*	RMSEP	Bias	Slope	Offset
Raw	6	0.9402	0.0681	0.0033	0.9101	0.0867
MSC	6	0.9455	0.0649	0.0027	0.9022	0.0934
1-Der	8	0.9785	0.0411	-0.0030	0.9584	0.0356
2-Der	6	0.9682	0.0498	-0.0010	0.9351	0.0592
de-trending	11	0.9787	0.0436	0.0150	0.9499	0.0615
DOSC	1	0.9822	0.0379	-0.0045	0.9850	0.0094

Table 3 The effective wavelengths selected by SPA

Preprocessing	No.	EWs (nm)
DOSC	1	1846
de-trending	17	1528,1878,2362,2306,2282,1174,2368,1784,2332,2216,1436, 2250,1142,2386,1401,2052,1106
1-Der	15	2124,1312,2014,2310,2216,2442,2286,1542,2244,2302,1264, 1940,2268,2276,1110

LS-SVM models The SPA-LS-SVM models were built with the selected EWs by SPA, RBF kernel, and combination of (γ, σ^2) determined by two-step grid search technique. According to experience and previous literatures (Liu et al., 2008a; Liu et al., 2008b), the search region of (γ, σ^2) was set as 10^{-3}-10^3.The SPA-LS-SVM models were built using the calibration set. The optimal (γ, σ^2) of DOSC, de-trending and 1-Der spectra based SPA-LS-SVM models were (2.2×10^3, 3.7×10^3), (272.4, 10.6) and (57.0, 52.1), respectively. The prediction results for validation set can be found in Table 4. Comparing with the other two SPA-LS-SVM models (DOSC and 1-Der), SPA-LS-SVM (de-trending) model was optimal. The optimal prediction results was achieved by SPA-LS-SVM model (de-trending) with r = 0.9864 and RMSEP = 0.0327. The scatter plot of reference vs predicted values in validation set are shown in Fig. 2. This result was slightly better than PLS model with DOSC spectra. The results also indicated that NIR spectroscopy combined with SPA-LS-SVM could be applied for the determination of valine in oilseed rape leaves under herbicide stress.

Table 4 The prediction results by SPA-LS-SVM models

Preprocessing	*r*	RMSEP	Bias	Slope	Offset
DOSC	0.9818	0.0378	-0.0045	0.9806	0.0136
de-trending	0.9864	0.0327	-0.0029	0.9742	0.0210

1-Der	0.9839	0.0353	0.0005	0.9684	0.0298

CONCLUSION The above results indicated that near infrared spectroscopy combined with SPA-LS-SVM was successfully applied for the determination of valine in oilseed rape leaves under herbicide stress. In this study, SPA was an efficient method for effective wavelengths selection. The optimal prediction results was achieved by SPA-LS-SVM model (de-trending) with r = 0.9864 and RMSEP = 0.0327. Thus, it could be considered that SPA-LS-SVM model (de-trending) was a useful model for prediction the valine in oilseed rape leaves under herbicide stress based on NIR spectroscopy. This research also provided a convenient, real-time and nondestructive measurement of the valine in oilseed rape leaves under herbicide stress.

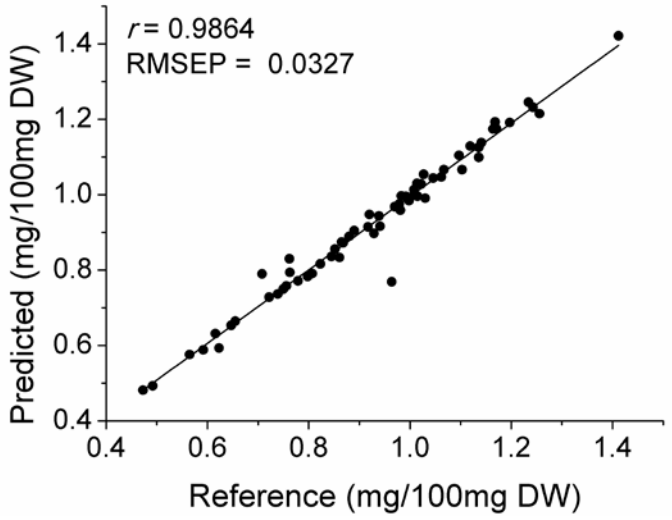

Fig. 2. The scatter plots of reference vs predicted values in validation set by SPA-LS-SVM (de-trending)

Acknowledgements. This work was supported by the 863 National High Technology Research and Development Program of China (2007AA10A210), Natural Science Foundation of China (60802038), Zhejiang Innovation Program for Graduates (YK2008014), Zhejiang Provincial Natural Science Foundation of China (Z3090295), and Agricultural Science and Technology Achievements Transformation Fund Programs (2009GB23600517).

REFERENCES

Araújo, M.C.U., T.C.B. Saldanha, R.K.H. Galvão, T. Yoneyama, H.C. Chame, and V. Visani. 2001. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemom. Intell. Lab. Syst. 57: 65-73.

Chu, X.L., H.F. Yuan, and W.Z. Lu. 2004. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique. Prog Chem. 16: 528-542.

Fang,H., H.Y. Song, F. Cao, Y. He, and Z.J. Qiu. 2007. Study on the relationship between spectral properties of oilseed rape leaves and their chlorophyll content. Spectrosc. Spectr. Anal. 27: 1731-1734.

Galvão, R.K.H., M.C.U. Araújo, W.D. Fragoso, E.C. Silva, G.E. José, S.F.C. Soares, and H.M. Paiva. 2008. A variable elimination method to improve the parsimony of MLR

models using the successive projections algorithm. Chemom. Intell. Lab. Syst. 92: 83-91.

Lisiewska, Z., W. Kmiecik, and A. Korus. 2008. The amino acid composition of kale (*Brassica oleracea* L. var. *acephala*), fresh and after culinary and technological processing. Food Chem. 108: 642-648.

Liu, F., F. Zhang, H. Fang, Z.L. Jin, W.J. Zhou, and Y. He. 2009b. Application of successive projections algorithm for nondestructive determination of total amino acids in oilseed rape leaves. Spectrosc. Spectr. Anal. 29(11): 3079-3083.

Liu, F., F. Zhang, Z.L. Jin, Y. He, H. Fang, Q.F. Ye, and W.J. Zhou. 2008a. Determination of acetolactate synthase activity and protein content of oilseed rape (*Brassica napus* L.) leaves using visible/near infrared spectroscopy. Anal. Chim. Acta 629: 56-65.

Liu, F., H, Fang, F, Zhang, Z.L. Jin, W.J. Zhou, and Y. He. 2009a. Nondestructive determination of acetolactate synthase in oilseed rape leaves using visible and near infrared spectroscopy. Chinese J. Anal. Chem. 37: 67-71.

Liu, F., Y. He, and L. Wang. 2008b. Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy. Anal. Chim. Acta 610: 196-204.

Liu, F., Y.H. Jiang, and Y. He. 2009c. Variable selection in visible/near infrared spectra for linear and nonlinear calibrations: A case study to determine soluble solids content of beer. Anal. Chim. Acta 635: 45-52.

Suykens, J.A.K. and J. Vandewalle. 1999. Least squares support vector machine classifiers. Neural Process. Lett. 9: 293-300.

Yan, Y.L., L.L. Zhao, D.H. Han, and S.M. Yang. 2005. The Foundation and Application of Near Infrared Spectroscopy Analysis. Beijing: China Light Industry Press.

Zhao, D.D., Y.L. Liu, D.W. Lan, A.L. Han, and Q.F. Ye. 2009. Effects of novel herbicide ZJ0273 on acetolactate synthase activity in oilseed rape and crickweed. Journal of Zhejiang University (Agric. & Life Sci.) 35(6): 633-638.