FAST DISCRIMINATION OF NANFENG MANDARIN VARIETIES BASED ON NEAR INFRARED SPECTROSCOPY TECHNIQUE

Huamao Zhou[1], Chao Zhou[1], Honghui Rao[1], Yande Liu[2,*]

[1] College of Engineering, Jiangxi Agricultural University, Nanchang 330045, China
[2] School of Mechtronics Engineering, East China Jiaotong University, Nanchang,330045, China
*Corresponding author, Address: School of Mechtronics Engineering, East China Jiaotong University, Nanchang, 330013, P. R. China, Tel:+86-791-7046173, Fax:+86-791-7046173, Email:zhm_77@163.com

Abstract. The potential of visible and near infrared (Vis/NIR) spectroscopy was investigated for discriminating the varieties of Nanfeng mandarin fruit nondestructively. The spectra were collected by a spectrophotometer in the wavelength range of 600–1040 nm. Relationship between the spectra and Nafeng mandarin varieties was established using principal component analysis (PCA), supervised independent modeling of class analogy (SIMCA) and backward propagation neural network (BPNN). By comparison the best result was obtained by BPNN with recognition rate of 97.5%. The results suggested Vis/NIR spectroscopy combination with BPNN was a new approach to discriminate of the varieties of Nanfeng mandarin fruit nondestructively.

Keywords: Vis/NIR spectroscopy, Nafeng mandarin fruit, varieties, SIMCA, BPNN

1   Introduction

Visible and near infrared (Vis/NIR) spectroscopy combination with chemometric techniques provides a powerful tool for monitoring the variation of processes for quality control purpose. The fact that samples requires virtually no treatment, and the expeditiousness with which spectra can be obtained, have made NIRS a preferred choice over traditional techniques involving lengthy sample conditioning procedures.

The characteristics of the absorption bands in Vis/NIR spectra are dictated both by the chemical composition of the sample and by physical properties such as color, firmness, size and defect. As a result, the Vis/NIR technique can be used to discrimination both internal and external quality, however, it is difficult to detect accurately and stably. In order to obtain accurate and simple calibration models, one can be used mathematical pretreatments to minimize unwanted spectral contributions in determining the quality of the samples.

These appealing features of the Vis/NIR technique have been fostered its use in a variety of fields including the meat, fruits and vegetables, grain and grain products, dairy products, oils, fish and fish products industries (Huang et al. 2008).

Nanfeng mandarin is popular for its high internal and external qualities with respect to sweetness, vitamin C, color, shape. But the quality is different to different varieties, and Vis/NIR has been applied to discrimination varieties, such as Chinese bayberry (Li et al. 2007), fruit vinegars (Liu et al. 2008), finishing oil (Blanco et al. 2002), tea (Chen et al. 2006) and transgenic tomatoes (Xie et al. 2007), which are important to quality control and customer choice.

The objectives of this work are to evaluate the use of Vis/NIR in discriminating the varieties of Nanfeng mandarin fruit including Daguoxi, Xiaoguoxi, Guihuadixi and Zaoshuxi, and the methods of principal component analysis (PCA), supervised independent modeling of class analogy (SIMCA) and principal component analysis-back propagation neural network (PCA-BPNN) are attempt to discriminate varieties accurately.

2 Materials and methods

2.1 Nanfeng mandarin fruit

A total 320 Nanfeng mandarin samples were harvested from a local orchard in Jiangxi province. There were four varieties of Daguoxi, Xiaoguoxi, Guihuadixi and Zaoshuxi. The varieties of Daguoxi, Xiaoguoxi and Guihuadixi are cultivated in Nanfeng, Jiangxi. Zaoshuxi is cultivated in Xunwu, Jiangxi. By comparison, Daguoxi is heavier than other varieties, fruit index of Xiaoguoxi is different from others, and the maturity of Zaoshuxi is earlier. Guihuadixi is a variety mutated from Xiaoguoxi. The Statistics of physical and chemical properties of four varieties are listed in Table 1. Base on these features, it is possible to discriminate the four varieties using destructive method, and NIRS is a rapid, accurate and stable tool to discriminate varieties nondestructively. Then 240 samples were selected to establish or train models including 60 samples of every variety, and others were for prediction.

They were placed in airtight polyethylene bags and stored in an ice filled refrigerator to keep at cold temperature (4±1℃). All fruit samples were allowed to equilibrate to room temperature (20℃) before NIRS measurement.

Table 1. Statistics of physics and chemical properties of four varieties of Nanfeng mandarins

Varities	Producing Area	Number	Physics properties								
			Weight(g)			Fruit Index			Color(ΔE)		
			Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.	Min.
Daguoxi	Nanfeng, Jiangxi	80	45.08	64.26	30.39	0.68	0.76	0.58	78.69	80.97	72.00
Xiguoxi	Nanfeng, Jiangxi	80	23.16	33.47	15.68	1.17	1.51	0.81	78.09	81.05	74.52
Guihuadixi	Nanfeng, Jiangxi	80	27.35	38.23	20.77	0.70	0.78	0.60	78.70	80.63	69.01
Zaoshuxi	Xunwu, Jiangxi	80	23.59	32.90	17.16	0.68	0.75	0.59	76.54	78.87	74.27
Varities	Producing Area	Number	Chemical properties								
			Souble Solid Content (^{O}Brix)			Titration Acidity (%)			Vitamin C (mg/100g)		
			Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.	Min.
Daguoxi	Nanfeng, Jiangxi	80	13.28	17.00	9.60	0.42	0.74	0.22	13.05	28.29	4.09
Xiguoxi	Nanfeng, Jiangxi	80	14.46	19.90	11.80	0.56	0.68	0.31	18.00	25.42	7.53
Guihuadixi	Nanfeng, Jiangxi	80	13.59	15.20	10.40	0.50	0.68	0.32	17.91	26.01	10.30
Zaoshuxi	Xunwu, Jiangxi	80	18.00	19.80	15.60	0.51	0.64	0.39	15.39	22.78	11.94

Fruit Index is defined as the ratio between horizontal and vertical diameter of fruit.

$$\Delta E = \sqrt{(L^2 + a^2 + b^2)}$$ (CIE Lab 1976)

2.2   System set up and transmission measurements
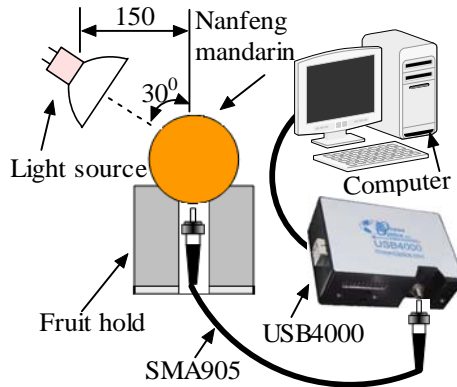
Fig. 1. The schematic of Nanfeng mandarin NIRS acquisition device

The system was consisted of a tungsten halogen lamp (24V/50W), CCD spectrometer (USB4000), an optical fiber (SMA905), USB data line and PC. The CCD spectrometer was 3648 pixel photodiode array. The wavelength range of the spectrometer was 400-1040nm with a 0.2nm sampling interval. The angle was about

30° between the lamp axis and the vertical line. The horizontal distance was about 150mm between the center of lamp and fruit hold.

Vis/NIR spectra were collected and transformed by OOIBase32 software (Oceanoptics INC., USA), from three positions marked with a circle beforehand on each mandarin around equatorial position. The data was analyzed by 'Unscrambler v9.5' software (CAMO AS, Trondheim, Norway) and Matlab 7.0 software (Mathworks, Natick, USA).

3   Results and Discussion

3.1   PCA Classification

PCA was performed to examine qualitative differences between the four kinds of Nanfeng mandarins by the raw spectra of 600-1040nm. Fig. 2 shows the three dimensional (3D) principal component score plot using first three score vectors, PC1, PC2, PC3 derived from raw spectra of samples. The initial three factors, which account for the most spectral variations 98% (77%, 19%, 2% for the first three principal components, PC1, PC2, PC3, respectively) related to the characteristic and indicated as positive or negative, are used to make differentiation clearer. From this figure, we can find that the samples are divided into two groups. One group is Xiaoguixi Nanfeng mandarin, others are Daguoxi, Guihuadixi and Zaoshuxi Nanfeng mandarins which overlapped each other. It can be assumed that the samples can be differentiated by using PCA in term of varieties. PCA gives very important information about the basic data structure regarding a potential capability of separation of objects. The result suggests that four varieties could not be separated clearly by PCA method. So SIMCA and PCA-BPNN were attempted to improve separation.
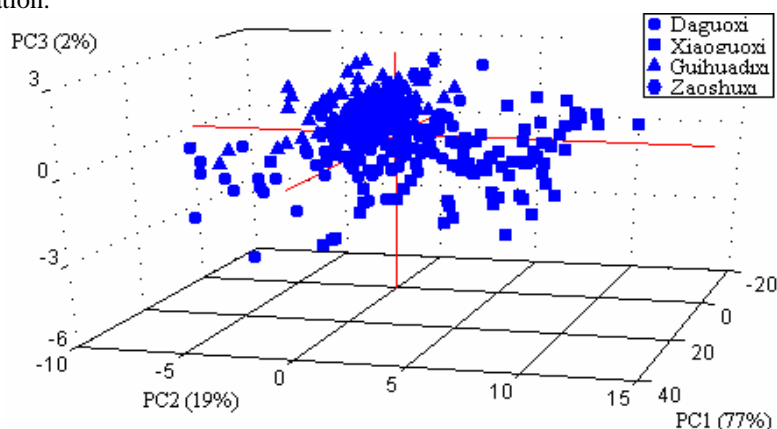

Fig. 2. Three dimensional score plot of the first three principal components for four varieties of Nanfeng mandarins

3.2   SIMCA Classification

When building SIMCA model, the four classification models are built by PCA analysis, respectively, including Daguoxi, Xiaoguoxi, Guihuadixi and Zaoshuxi Nanfeng mandarins. The samples of test set was selected randomly, the assignment of the samples can be seen for Table 2. The appropriate number of principal component factors would be concerned in building SIMCA model. In our study the number of PCs is selected by root mean square error of cross validation, and the appropriate number of PCs is 3 for four varieties, respectively (Fig. 3). At 99% confidence level, the best results of identification are achieved, that Zaoshuixi and Guihuadixi could not be discriminated well, So PCA-BPNN method is applied to improved separation.

Table 2. The results of classification by SIMCA

varieties	Sample numbers	PCs	The identification rate in test set			
			Daguoxi	Xiaoguoxi	Guihuadixi	Zaoshuxi
Daguoxi	20	3	18(90%)	0	0	0
Xiaoguoxi	20	3	0	15(75%)	0	0
Guihuadixi	20	3	0	0	19(95%)	0
Zaoshuxi	20	3	0	0	6(30%)	20(100%)
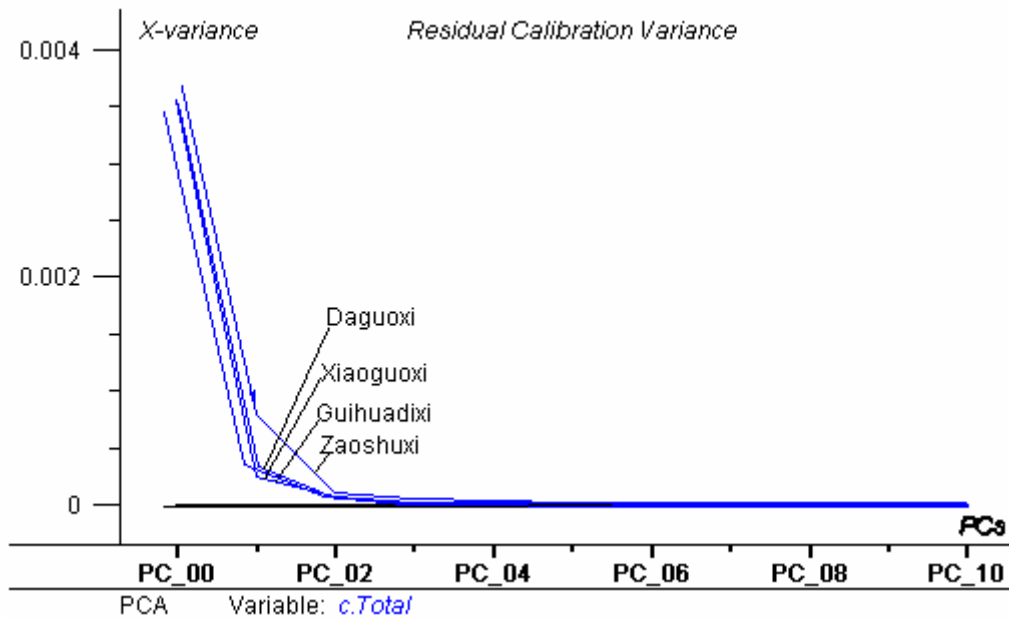
**Fig. 3.** The relationship between RMSECV and principal in model

3.3   PCA-BP Classification

As show in Table 3, the first seven PCs are considered to explain the original variables. And these variables are not interacting from each other. The first seven PCs, which can explain the 99.881% of variables, were selected as the input vectors to build the PCA-BPNN model. The output vectors of these samples were assumed to be 1, 2, 3 and 4. And 1 was denoted as Daguoxi mandarin, 2 was denoted as Xiaoguoxi mandarin, 3 was denoted as Guihuadixi mandarin and 4 was denoted as Zaoshuxi mandarin. Adjusting by many times, the optimal nodes of the hidden layer of neural network came up to 9. Therefore the seven PCs as the input neurons were connected with 9 hidden neurons randomly; the node of output neuron layer was defined as a vector of [1, 2, 3, 4]. The transfer function of hidden layer was tansig function. The transfer function of output layer was purelin function. The train function was trainlm, the threshold value was 0.08. The whole samples were separated randomly into train and test set (240:80) randomly.

Table 3. The statistics of PCs and accumulative reliabilities

PC	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Ar(%)	77.585	95.468	98.078	99.670	99.796	99.845	99.881	99.896	99.906	99.915

PC: Principal component; Ar: Accumulative reliabilities.
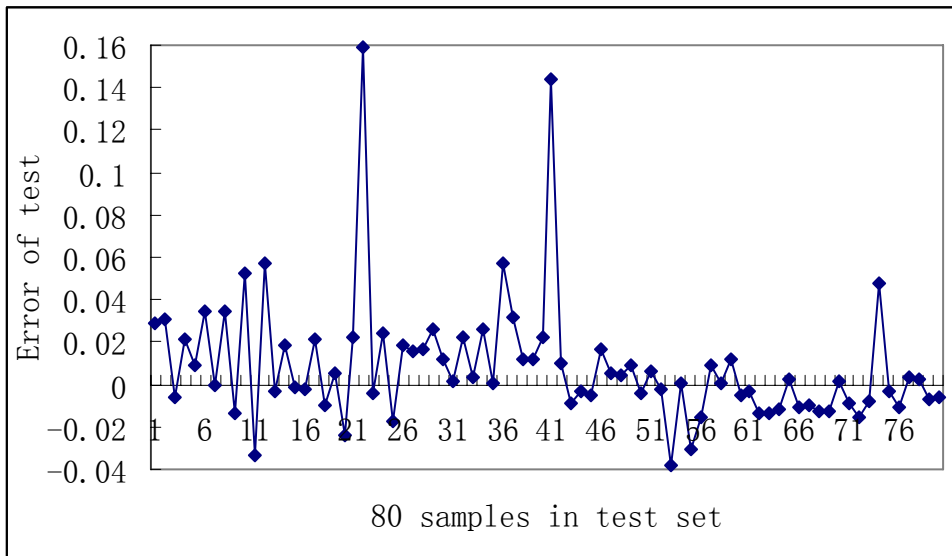
Fig. 4. Error of the model in the test set

Fig. 4 is the plots of the prediction errors of the 80 samples in the test set. The abscissa represents the 80 samples in order. The ordinate represents the error between the test output vectors and standard vectors of these samples. It can be found that the error of almost all samples was be close to zero, expect the NO.7, 15, 35 and 54

samples. The error of the No. 7, 15, 35 and 54 were -0.00017, -0.00118, 0.000126 and 0.000844 respectively. The threshold value was 0.06, so two samples was classified inaccuracy. The total recognition rate is 97.5% for the test set.

4   Conclusions

The contribution of this work is to present a rapid and non-destructive approach for discriminating different varieties of Nanfeng mandarin. In this research, quantitative analysis for varieties of Nanfeng mandarin fruit was carried out. By means of PCA, SIMCA and PCA-BPNN techniques, a relation was established between transmission spectra and varieties of Nanfeng mandarin fruit. The PCA-BPNN shows a better data prediction performance. And the recognition rate of 97.5% achieved. It is proved to be appropriate and successful for species discrimination. This study lays a foundation for further research to develop a portable apparatus in discriminating the different species.

5   Acknowledgement

6   References

1. Huang, H. B., Yu, H. Y., Xu, H. R., et al.. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beerages : A review[J]. Journal of Food Engineering, 2008, 87(3): 303-313.
2. Li, X. L., He, Y., Fang, H.. Non-destructive discrimination of Chinese bayberry varieties using Vis/NIR spectroscopy [J]. Journal of Food Engineering, 2007, 81(2): 357-363.
3. Liu, F., He, Y., Wang, L.. Determination of effective wavelengths for discrimination of fruit vinegars using near infrared spectroscopy and multivariate analysis [J]. Analytica Chimica Acta, 2008, 615(1): 10-17.
4. Blanco, M., Pagès, J.. Classification and quantitation of finishing oils by near infrared spectroscopy [J]. Analytica Chimica Acta, 2002, 463(2): 295-303.
5. Chen, Q. S., Zhao, J. W., Zhang, H. D., et al. Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration [J]. Analytica Chimica Acta, 2006, 572(1): 77-84.
6. Xie, L. J., Ying, Y. B., Ying, T. J., et al. Discrimination of transgenic tomatoes based on visible/near-infrared spectra [J]. Analytica Chimica Acta, 2007, 584(2): 379-384.