International Journal of u- and e- Service, Science and Technology
Vol.9, No. 12 (2016), pp.87-96
http://dx.doi.org/10.14257/ijunesst.2016.9.12.08

Research on MW-IPLS in Wavelength Selection based on NIR of Rice Moisture

Weizheng Shen[1], Jingjing Wang[2], Fengzhu Hu[3] and Qingming Kong[4*]

School of Electronic Engineering and Information Northeast Agricultural University, Harbin, 150030, China
[4]wzshen@neau.edu.cn

Abstract

The moisture content of rice is of great significance for the eating quality and food safety. Therefore, it is very necessary to establish a rapid, stable, reliable and high prediction accuracy quantitative analysis model that can be used in on-line detection. In this study, the quantitative analysis technique of near infrared diffuse reflectance spectroscopy was used to detect the moisture content in rice. We combine the algorithm (MW-IPLS) based on MWPLS (Moving Window Partial Least Squares) with IPLS (Interval Partial Least Squares) to optimize the characteristic wavelength. Then we establish partial least squares regression in the preferred characteristic wavelength range. The experimental results show that the model of quantitative analysis using the MW-IPLS algorithm to optimize the characteristic wavelength is optimal comparing with the whole spectrum and single method such as MWPLS and IPLS. The numbers of Factors, R^2P, RMSECV and RMSEP are 6, 0.8597, 0.2523 and 0.2753 respectively. Therefore, using the WM-IPLS algorithm to optimize the characteristic wavelength can reduce the processing capacity of the data and make the model more concise. In addition, it also provides a new method for the analysis of near infrared spectral characteristic wavelength selection.

Keywords: *Near Infrared Spectroscopy, Moving Window Partial Least Squares, Interval Partial Least Squares*

1. Introduction

Rice is one of the most important food crops in China. The sown area of rice crops accounts for about 1/4 of the total area. Production accounts for about 1/2 of the total grain output which accounts for more than half of the commercial grain. With the continuous improvement of people's living standards, people's demand on the eating quality of rice is getting higher and higher. However, in recent years, the fact that there are some sharpshooters who are posing old rice as the new rice into the marketing has serious impact on the normal circulation of rice and the public food safety. The demand to relevant departments and enterprises to fast and accurate detection technology in rice quality become increasingly strong. In the quality assessment of rice, moisture is an important evaluation index. In the rice storage process, rice will become fissuring and aging because the bran powder layer of the rice surface absorb moisture from the air in the rice have no protection of husk peels and contact with the storage environment directly. The change of moisture content in rice not only affects the edible quality of rice, but also relates to the food safety of the public.

The detection of the moisture content of rice mainly depends on four kinds of detection methods introduced by determination of moisture in foods in GB 5009.3-2010. Among them, the most commonly used method is the direct drying method. However, these

[*] Corresponding Author

ISSN: 2005-4246 IJUNESST

methods are cumbersome procedure. More importantly, it is not conducive to online testing and we can't achieve real-time detection and real-time feedback to better guide the purpose of production. Therefore, a quick and nondestructive method of the establishment of rice in the moisture has important significance to product high quality rice. Near Infrared Spectroscopy as a new direction and a new method for rapid detection has been widely used in food, juice, feed, agricultural products quality and other analytical detection in foreign countries [1-8]. Scholars of domestic and foreign have demonstrated its feasibility in the rice component detection through a large number of experimental studies [9-12]. But when full-spectrum band or a single method of analysis is used in the modeling process, the spectrum containing a large amount of redundant information will affect the performance of the model. This study presents the WP-IPLS band selection method to establish near infrared calibration model of rice water initially which provides a reference for the study of the method of near-infrared spectral characteristic wavelength selection by using Antaris II near infrared spectrometer and 109 samples of rice in Heilongjiang area as samples.

2. Materials and Methods

2.1. Samples Collection, Preparation and Calibration

The 109 rice samples for experiments came from around the Heilongjiang province covering the province in different regions, different climate, different soil types and different varieties. Firstly, we made the samples into powder by the mill processing 60s until more than 95% passing the sieve. Then we put the samples in a sealed bag. Finally, the samples were stored at room temperature in the dark.

2.2. Spectrum Acquisition

In the experiment, the rice samples were scanned by the near infrared spectrometer of Thermo Antaris II. Because the rice samples were solid powders, so we used diffuse reflectance spectral scanning. The resolution of spectral scanning was 4 cm^{-1} and the spectral range was 4000-12000 cm^{-1} corresponding to wavelength range 1000-2500nm which was a total of 519 wavelength points. We scanned the surface by integrating sphere, 11.9W/7V halogen lamps, SabIR optical fiber detector treating the air as the contrast object. We scanned 64 times and took the average. Figure 1 was the original spectrum experiment 109 samples. The horizontal axis was the wave number from 4000 cm^{-1} to 12000 cm^{-1} and the ordinate was the absorbance of the sample.

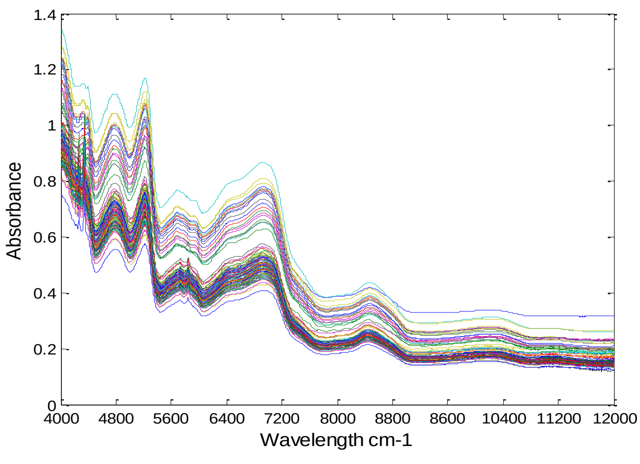


Figure 1. The Original Spectrum of Rice

2.3. Standard Chemical Calibration

Chemical analysis of moisture content in the sample was determined by the direct drying method in the standard GB 5009.3-2010. Firstly, we took clean aluminum flat weighing bottle and put them at 101 to 105 DEGC in the drying box. The bottle cover should be inclined to be supported on the side of the bottle. The bottle was removed after heating for 1 h and put them in the cooling of the dryer for 0.5 h. And weigh and repeat dry until a constant weight that the mass difference between before and after does not exceed 2 mg. Secondly the mixed sample was quickly milled to a particle of less than 2 mm. Next weigh 2 g (accurate to 0.0001 g) into the weighing bottle in case that the thickness of the sample was not more than 5 mm. After covering with lid for precision weighing, put them in the drying box. The bottle cover should be inclined to be supported on the side of the bottle. The bottle was removed after drying for 4 h and weighed after putting them in the cooling of the dryer for 0.5 h. Then put them in the drying box for 1h again and weigh after putting them in the cooling of the dryer for 0.5 h. And repeat the operation until a constant weight that the mass difference between before and after does not exceed 2 mg. Finally, each sample was measured with 10 parallel samples, and the average value of each sample was measured. Table 1, shows the statistical results of the water content of rice.

Table 1. Statistics of Moisture Content in Rice

Sample classification	Number	Min (%)	Max (%)	Average (%)	Standard Deviation
Calibration set	79	8.37	12.40	10.71	0.79
Validation set	30	8.67	12.15	10.69	0.78

3. Results and Discussion

3.1. Spectra Denoising

In addition to containing samples of their own chemical information, the spectrum also contains other unrelated information and noise. Therefore pretreatment method is very necessary to establish the model in the chemometrics method. In this study, smoothing method was used to remove noise spectra of samples. We selected five different smoothing windows from 7 point to 15point to verify. The smoothing processed to achieve the best denosing results when selecting the window size of 15 points. Figure 2, was the spectrum after the removal of the noise with a 15 points smoothing.

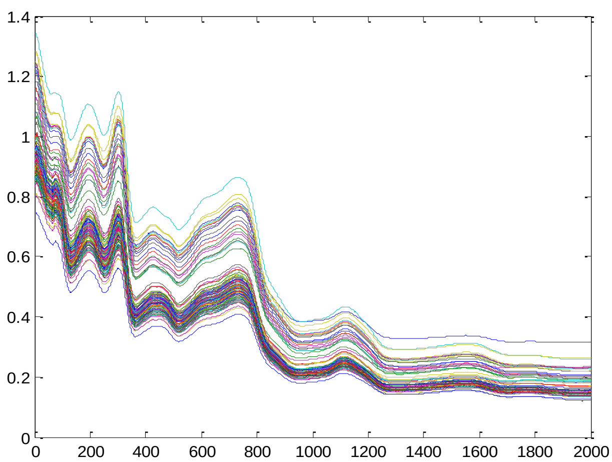
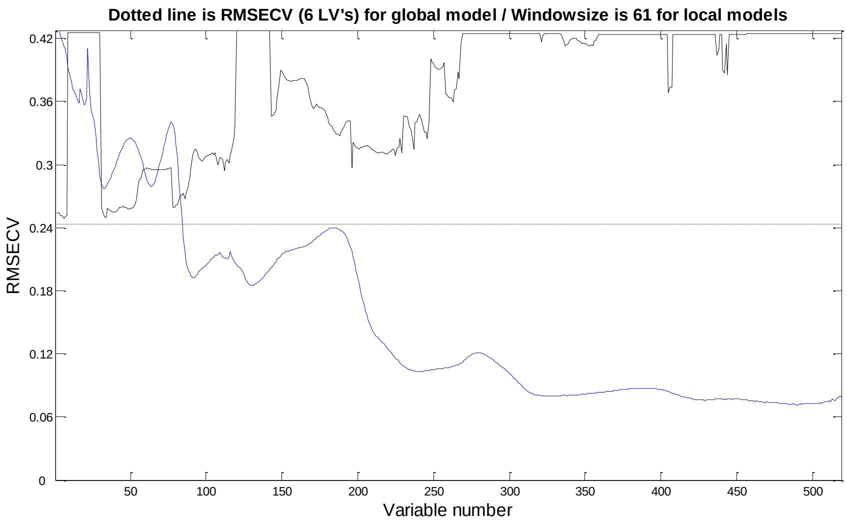


Figure 2. The Spectrum after Removing Noise

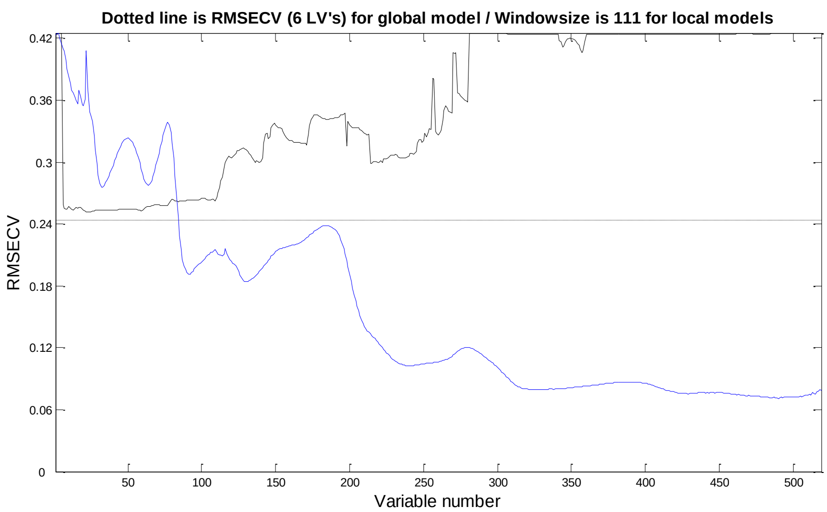
3.2. Moving Window Partial Least Squares Band Selection

We used MWPLS band selection method for the calibration model after smoothing processing. The MWPLS band selection is carried out for the smoothing model. MWPLS is a band selection method based on partial least squares algorithm [13]. The prediction accuracy of PLS model can be improved significantly established by their preferred spectrum bands [14-18]. Its basic principle is to move a window along a continuous spectrum axis. A model is established and a series of residual sum of squares (PRESS or SSR) is obtained corresponding to different window (moving wavelength point) and the main factor by using mutual authentication method each moving one wavelength point. Whether moving window PLS method or interval PLS method, the width of the window selection is very important.

In this study, we chose the moving window size of 61-111 point to use WMPLS characteristic wavelength selection method and then accorded the value of RMSECV to determine the best features bands. Because the data was more, not all the data of moving window was listed. Figure 3, only was the characteristic wavelength selection results of window size of 61 points and 111 points.



a. The Characteristic Wavelength Selection Results of Window Size of 61 Points



b.The Characteristic Wavelength Selection Results of Window Size of 111 Points

Figure 3. MWPLS Characteristic Wavelength Optimization Result

Table 2, showed the model calibration results established by selecting the optimal band in different window sizes. As can be seen from the table, the value of RMSECV became gradually smaller with the gradual increase of the window size. The modeling result was optimal when the size of the moving window was 111 and the wavelength range was between 4108 cm^{-1} and 4911 cm^{-1}. At this point, the value of RMSECV was 0.2569, and the value of R^2 was 0.8858.

Table 2. Modeling Results of MWPLS in Different Window Size

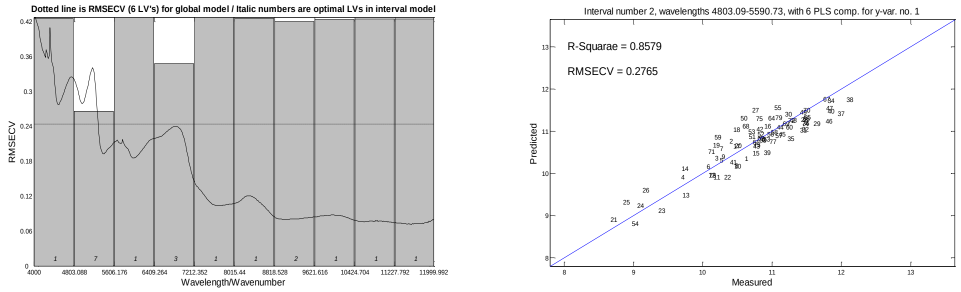
Window Size	Factors	Interval(cm^{-1})	RMSECV	R^2
61	7	4000 - 4108	0.2884	0.8460
71	8	4385 - 4693	0.2838	0.8492
81	6	5103 - 5589	0.2779	0.8556
91	7	4123 - 4470	0.2731	0.8653
101	6	4815 - 5362	0.2682	0.8716
111	6	4108 - 4911	0.2569	0.8858

3.3. Interval Partial Least Squares Band Selection

IPLS band selection method is to divide the whole band spectral region into a number of equal width intervals and perform PLS model based on the spectral range of each interval. We selected the minimum value of RMESCV as the optimal interval by comparing the RMSECV of model established each interval.

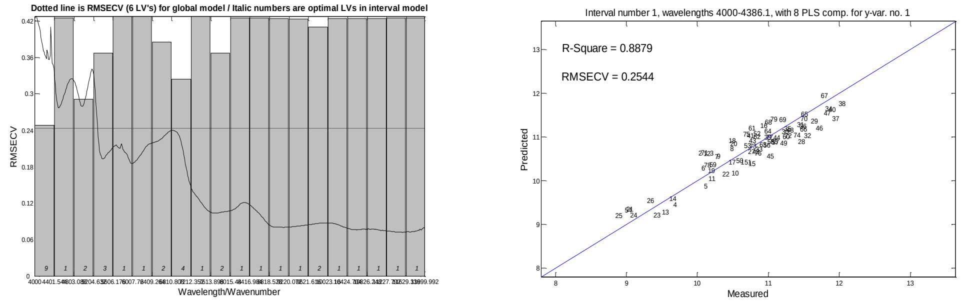
Firstly, the IPLS band selection method was used to establish cross-validation models based on PLSR treating 10, 20 and 30 points intervals respectively as input variables. Figure 4.a, denoted the validation set root mean square error obtained by the model 10 points interval established. By comparing the RMSECV of each band, we can see it is better at second and fourth band. Therefor we selected the second band that has the lowest RMSECV to establish the model. The wavelength range was between 4803 cm-1-5590 cm-1. Figure 4.b, showed the model validation results. Figure 5.a, denoted the validation set root mean square error obtained by the model 20 points interval established. We can see the value of RMSECV is better at first, third, eighth and tenth band. Therefor we selected the first band that has the lowest RMSECV to establish the model. The wavelength range was between 4803 cm-1-5590 cm-1. Figure 5.b, showed the model validation results. Figure 6.a, denoted the validation set root mean square error obtained by the model 30 points interval established. Therefor we selected the fourth band that has the lowest RMSECV to establish the model. The wavelength range was between 4833 cm-1-5096 cm-1. Figure 6.b, shows the model validation results.

It can be seen that the value of RMSECV of the model established by 30 intervals IPLS was obviously increased by comparing the value of the model established by the spectrum of each interval. The higher the interval was, the higher the value of RMSECV was and the bigger the error was. Integrating the models of three intervals, as shown in Table 3, the model was the most optimal when the interval was 20 and wavelengths between 4000 cm-1. In this case, the value of RMSECV was 0.2544 and the value of R2 was 0.8476.



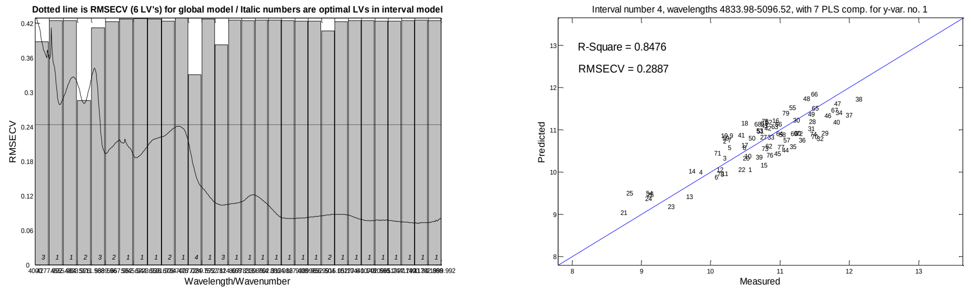
a. The Band Selection Result of 10 Intervals **b. The Optimal Model Results of 10 Intervals**

Figure 4. The Spectrum after 10 Intervals IPLS



a. The Band Selection Result of 20 Intervals **b. The Optimal Model Results of 20 Intervals**

Figure 5. The Spectrum after 20 Intervals IPLS



a. The Band Selection Result of 30 Intervals **b. The Optimal Model Results of 30 Intervals**

Figure 6. The Spectrum after 30 Intervals IPLS

Table 3. Model Validation Results of IPLS in Different Interval

Interval method	Band	Factors	Wavelength(cm^{-1})	RMSECVRMSECV	R^2
10	2	6	4803-5590	0.2765	0.8579
	4	8	6409-7196	0.3501	0.7698
20	1	8	4000-4386	0.2544	0.8879
	3	6	4803-5189	0.2801	0.8536
30	4	7	4833-5096	0.2877	0.8476

3.4. MW-IPLS Band Selection Model Demonstration

The MWPLS and IPLS band selection methods were used for the full spectrum of 519 wavelength points. In MWPLS method, the lowest value of RMSECV was 0.2569 when the moving window size was 111 and the wavelength range was between 4108cm-1 and 4911cm-1.While in IPLS method, the lowest value of RMSECV was 0.2544 when the interval was 20 points and the wavelength range was between 4000cm-1 and 4386cm-1. Therefore, we chose the band as the spectral characteristics comparing with the single method. Table 4, was the prediction model of moisture content established by cross validation method based on PLSR treating the data of full spectrum and after different band selection processing respectively as input variables. As can be seen from the table, the model of quantitative analysis using the MW-IPLS algorithm to optimize the characteristic wavelength is optimal. Figure 7, was the prediction results of the model. It can be seen, the prediction model had a high coefficient of determination and the smallest RMSECV when the main factor of the number was 6. The decision coefficients of the validation set and the prediction set were 0.8902 and 0.8597 respectively. RMSEC and RMSEP were 0.2523 and 0.2753 respectively.

Table 4. Model Validation Results of IPLS in Different Interva

Method	Internal method	Factors	Wavelength(cm^{-1})	RMSEP	R^2
Whole spectrum		6	4000-12000	0.3076	0.8232
MWPLS	111 point	6	4108-4911	0.2777	0.8561
IPLS	20 interval	8	4000-4386	0.2830	0.8503
MW-IPLS		**6**	**4108-4386**	**0.2753**	**0.8597**

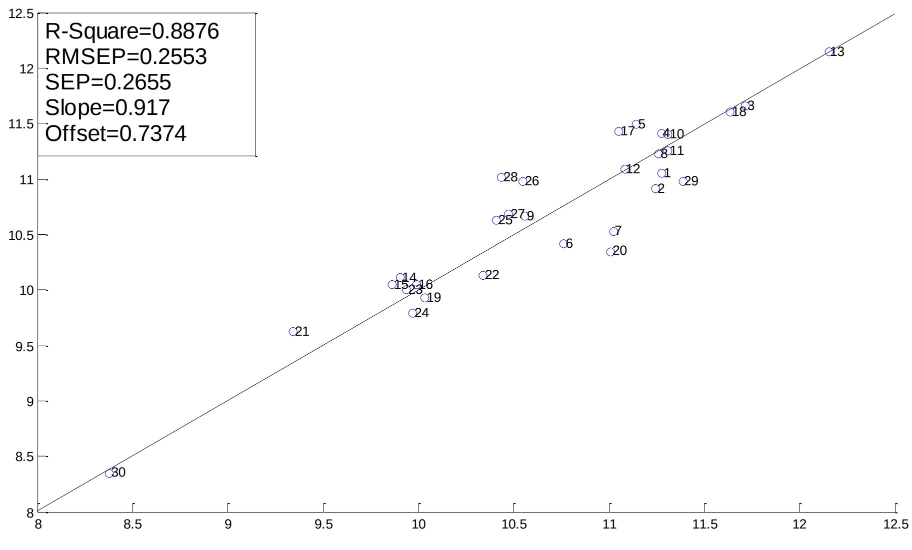


Figure 7. PLSR Model Validation Results

4. Conclusions

We realize the analysis and the establishment of the model in the moisture content of rice by using NIRS and Matlab programming. We find that MW-IPLS is an effective method to optimize the characteristic absorption bands of moisture spectrum by comparing with the full spectrum and a single band selection method. Combined with the IPLS and MWPLS for band selection methods, we can determine the optimal wavelength range between cross that is 4108-4386cm-1. Then we use PLSR to establish the calibration model of rice moisture content whose coefficient of determination R2 reaches 0.8597 and the validation set root mean square error RMSEP is 0.2753. The results show that it is beneficial to establish the higher accuracy and faster data operation calibration model by optimizing spectral characteristic absorption bands of moisture in MW-IPLS to reduce the effective wavelength number. More important is to provide a theoretical basis for the subsequent depth analysis of near infrared spectrum wavelength selection method.

5. Acknowledgements

The authors would like to acknowledge the financial support from the National High Technology Research and Development 863 Program (2013AA102303), the Young talents Project of Northeast Agricultural University(14QC13), the Natural Science Foundation of Heilongjiang Province of China (F201402), The authors also wish to acknowledge the continuous support of the Northeast Agricultural University and National Engineering Research Center for Agricultural Information Technology.

References

[1] F. Liu, Y. He and G. M. Sun, "Determination of Protein Content of Auricularia auricular Using Near Infrared Spectroscopy Combined with Linear and Nonlinear Calibrations, J.Agric.Food Chem, vol. 57, **(2009)**, pp. 4520-4527.

[2] E. Bertone A. Venturello, R. Leardi and F. Geobaldo, "Prediction of the optimum harvest time of Scarlet apples using DR-UV-Vis and NIR spectroscopy", Postharvest Biology and Technology, vol. 69, **(2012)**, pp. 15-23.

[3] E. D. Louw and K. I. Theron, "Robust prediction models for quality parameters in Japanese plums(Prunus salicinaL.) using NIR spectroscopy", Postharvest Biology and Technology, vol. 58, no. 3, **(2010)**, pp. 176-184.

[4] M. M. Blanke, "Non-invasive Assessment of Firmness and NIR Sugar (TSS) Measurement in Apple", Pear and Kiwi Fruit.Springer-Verlag Berlin Heidelberg, **(2013)**, pp. 19-24.

[5] I. Jan, A. Rab and M. Sajid, "Storage performance of apple cultivars harvested at different stages of maturity", The Journal of Animal and Plant Sciences, vol. 22, no. 2, **(2012)**, pp. 438-447.

[6] S. R. Delwiche, W. Mekwatanakarn and C. Y. Wang, "Soluble Solids and Simple Sugars Measurement in Intact Mango Using Near Infrared Spectroscopy", HortTechnology, vol. 18, no. 3, **(2008)**, pp. 410-416.

[7] V. Ziosi, M. Noferini, G. Fiori, A. Tadiello, L. Trainotti, G. Casadoro and G. Costa, "A new index based on vis spectroscopy to characterize the progression of ripening in peach fruit", Postharvest biology and technology, vol. 49, no. 3, **(2008)**, pp. 319-329.

[8] E. D. Louw and K. I. Theron, "Robust prediction models for quality parameters in Japanese plums (Prunus salicinaL.) using NIR spectroscopy", Postharvest Biology and Technology, vol. 58, no. 3, **(2010)**, pp. 176-184.

[9] W. Sriswas, V. K. Jindal and W. Thanapsae, "Relationship between sensory textural attributes and near infrared spectra of cooked rice", Near Infrared Spectroscopy, no. 15, **(2007)**, pp. 233-240. [10] I. Elbatawi and G. K. Arafa, "Detecting rice quality as influenced by hulling using visble and near-infrared spectroscopy", Annals Agric.Sci. vol. 53, no. 1 **(2008)**, pp. 77-88.

[10] J. C. Lee, Y. H. Yoon and S. M. Kim, "Development of prediction model for total dietary fiber content in brown rice by fourier transform-near infrared spectroscopy", Food Science and Technology, vol. 38, no. 2 **(2006)**, pp. 165-168.

[11] S. E. Kays, F. E. Barton and W. R. Windham, "Predicting Protein Content by Near Infrared Reflectance Spectroscopy in Diverse Cereal Food Products", J. Near Infrared Spec., vol. 8, no. 1, **(2008)**, pp. 35-43.

[12] M. Sohn, F. E. Barton and A. M. McClung, "Near-Infrared Spectroscopy for Determination of Protein and Amylose in Rice Flour Through Use of Derivatives", Cereal Chemistry, vol. 81, no. 3, **(2004)**, pp. 341-344.

[13] J. Jianhui, R. J. Berry and W. Heinz, "Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data", Anal. Chem., vol. 74, **(2002)**, pp. 3555-3565.

[14] S. Kasemsumran,Y. P. Du and K. Murayama, "Near-infrared spectroscopic determination of human serum albumin,y-globulin,and glucose in a control serum solution with searching combination moving window partial least squares", Analytica Chimica Acta., vol. 512, **(2004)**, pp. 223-230.

[15] Y. P. Du, Y. Z. Liang and J. H. Jiang, "Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares", Analytica Chimica Acta., vol. 501, **(2004)**, pp. 183-191.
[16] T. Davies and K. H. Morris, "NIR spectroscopy", NIR News, vol. 16, no. 7, **(2005)**, pp. 9-12.
[17] N. Kang, S. Kasemsumran and Y. A. Woo, "Optimization of informative spectral regions for the quantification of cholesterol, glucose and urea in control serum solutions using searching combination moving windowpartial least squares regression method with near infrared spectroscopy", Chemometrics and Intelligent Laboratory Systems, vol. 82, **(2006)**, pp. 90-96.
[18] S. Kasemsumran,Y. P. Du and K. Maruo, "Improvement of partial least squares models for in vitro and in vivo glucose quantifications by using near-infrared spectroscopy and searching combination moving window partial least squares", Chem. Int. lab. Syst., vol. 82, **(2006)**, pp. 97-103.

Authors

Weizheng Shen, (1977-), male, Ph.D., professor, mainly engaged in the research and application of information technology in agriculture.