Development of mathematical model for predicting rice gel consistency by near infrared spectroscopy

Liu Jianxue[1], Zhang Yuxian[1], Dong Tieyou[1], Wang Jianhua[1], Wu Shouyi[2]

(1. *Food and Bio-engineering College, Henan University of Sciences and Technology, Luoyang* 471003, *China;*

2. *Biology and Enviroment College, Jiangsu University, Zhenjiang* 212013, *China*)

Abstract: Gel consistency(GC) is one of the most important cooking and eating characteristics of rice. The testing method of rice GC by near infrared spectroscopy (NIRS) was developed, and the prediction models of rice GC was set up based on the partial least squares(PLS) methods. Sixty spectrums of rice samples were calibrated for their GC values. It shows that the correlation coefficient between the PLS evaluation and chemical method is 0.95, and the standard error of calibration (SEC) is 0.66. To validate the calibration, an independent set of 41 rice samples of the same breed was used. The correlation coefficient is 0.92, and the standard error of validation is 0.78. The result shows that the infrared spectroscopy technique can be used to test the rice GC rapidly.

Key words: rice; gel consistency; partial least squares method; near infrared spectroscopy; mathematical model

0　Introduction

More and more attention is being paid not only to the nutrition quality of foodstuffs, but also to the taste characteristics, so the gel consistency (GC), as one of the most important cooking and eating characteristics of rice, is becoming the main concern. If the GC value of rice is too high, the cooked rice will be too soft, and *vice versa*. Therefore it is important to breed new varieties of rice with medium GC level, right amount of amylose content and high protein content. For a long time, the measurement of the physical and chemical evaluating indexes of rice quality is costing and time consuming. Over the past 20 years, Fourier transform near infrared spectroscopy (NIRS) has been used to quantitatively predict the concentration of various constituents in food and agricultural products. Recent studies have demonstrated that NIRS models had the sufficient accuracy for breeding programs and quality monitoring, and can be developed to predict amylose content[1, 2] and protein content[3] in milled rice based on the principle of artificial neural network algorithm. In this paper, a quick testing method for measuring rice GC with NIRS is represented.

The gel consistency represents some characteristics of the complicated components of rice itself. The starch (combined with amylose and amylopectin) will become viscosity when heated in water. It is difficult for the starch with high amylose content to be gelatinized, but the gel will be very stable if it is gelatinized. The higher the temperature is, the more severely the amylum molecules vibrate. As a result, the molecule bond will break, and there will be more points and positions in the amylum molecules combining with H-bond in water. The pasty, viscosity and gel consistency of starch depend not only on temperature, but also on other commonly coexisting components (constituents) such as sugar, protein, fat, water. It is possible to take the advantage of NIRS to measure the GC of rice.

It is on the interferometer that the NIRS can be used to identify and analyze some components quantitatively. After a near infrared beam irradiates the sample through the interferometer, the interferential curve was gained, and the change from the interferential curve to the near infrared spectrum can only take place through the Fourier digital transformation by computer. Every substance such as starch, protein, fat has its own proper infrared absorbance, and the measuring of rice GC is just based on the characteristic of this proper absorbance.

1　Partial least squares method

PLS modeling is a powerful new multivariate full scanning spectrum method with the signal data press

functions of other full spectrum methods such as PCR (Principal Component Regression) and CLS (Classical Least Squares) [4, 5]. It has been presumed that the information of interest about a response variable is mainly contained in the directions of the predictor space which have both large variations and high correlations with the response variable. Thus in matrix notation, the calibration spectra can be represented for PLS regression by using the spectral decomposition notation as follows

$$A = HP^T + E$$

Where A is the $m \times n$ matrix of calibration spectra; P^T is the $h \times n$ matrix with the rows of P being the new PLS basis set of h full spectrum vectors, often called loading vectors or loading maxrix. H is an $m \times h$ matrix of intensities (or scores) in the new coordinate system of the h PLS loading vectors for the m sample spectra. E is now the $m \times n$ matrix of spectral residuals not fit by the optimal PLS model.

The component concentrations matrix C is given by

$$C = UQ^T + F$$

Where Q^T is an $l \times n$ matrix; U is the $m \times h$ matrix of the unknown calibration coefficients relating the l component concentrations to the spectrum intensities, and F is the $m \times l$ vector of random concentrations errors or residuals that are not fit by the model. B is a diagonal matrix, which was gained by relating with U and H: $U = HB$. In the process of iteration, a weight W has been introduced, the iteration can run as follows

$$W = U^T A / (U^T U)$$
$$W = W / \|W\|$$
$$t = AW / (W^T W)$$
$$q^T = t^T C / (t^T t)$$
$$U = Cq / (q^T q)$$

Where t is the rows of H, and q is the rows of Q, $\|W\|$ is the Euclid norm of W. While the value of $\|t_{new} - t_{old}\| / \|t_{new}\|$ is less than the threshold, the first column of loading matrix is stored, then using the calculated matrices of E and F as modified A and C. The above process was repeated to calculate the second column of loading matrix until the desired number of loading vectors has been calculated. From the PLS model

$$C = AB + F$$

Following operation can get the regressive coefficient matrix B

$$B = W (P^T W)^{-1} Q^T$$

Since model error is presumed to be the error in the component concentrations, this algorithm minimizes the squared errors in concentrations during calibration.

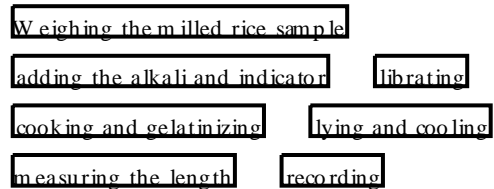## 2 Materials and experiment

2. 1 Materials

The Chinese varieties of short round-grain, long-grain rice, glutinous rice and some hybrid rice, were selected to comprise the calibration and validation samples. The gel consistency of rice was evaluated by testing the flowage length of mushy rice. The rice can be classified into three grades according to the GC, viz. low (GC 40 cm), medium (40 cm < GC 60 cm) and high (GC> 60 cm) GC rice. The GC ranges of above samples are 36 100 cm, 31 100 cm, 51 100 cm and 34 100 cm, respectively. The moisture contents of samples were controlled at 9. 8%.

For each sample, 100 g of grain was ground by a polish rice grinder for 2. 5 min. The milled rice was ground for analysis in the Udy cyclone mill (Udy Fort Collins, Co. , USA) with 0. 2 mm mesh screens. Finally it was packaged in the kraft bag, and kept together with standard samples under room temperature for 2 days to balance the moisture content.

2. 2 Chemical analyses

When the milled rice sample was put into the dilute alkali (0. 2 mol/L KOH), heated for 15 min, and then cooled, the paste was formed. The cooled gel in test tables flowed slowly when it was laid on a flat glass table. The GC was obtained throughout the testing of the flowage length. In this paper, the GC was measured according to reference [8]. The process of measuring the GC is shown as follows:

Weighing the milled rice sample

adding the alkali and indicator   librating

cooking and gelatinizing   lying and cooling

measuring the length   recording

2. 3 Spectroscopic analyses

A Fourier transform infrared spectrometer (Model Vector-33, Bruker Co. , German) was used to take reflectance readings over a wavelength range of 800 2500 nm. Reflectance reading was taken every 4 nm increment. Sixty-four repetitive scans were accumulated in computer memory and averaged, transformed to $\log(1/R)$, and then stored to computer file, forming one spectrum per sample. Each sample was measured 3 times under the same condition.

The prediction model of the GC of rice was developed by using PLS algorithm based on the NIR

spectra. Samples were placed into calibration or validation sets by being sorted according to GC values. Every three samples were included in the calibration set, the following two samples make up the validation set. Distribution of GC values for varieties used in the calibration and validation sets are shown in Fig. 1. PLS analysis was performed using calibration set spectra with the software package OPUS, the number of factors selected to optimize the model is accomplished by calculating the PRESS (prediction residual error sum of squares) through a cross validation procedure. Upon completion of the PLS analysis, a diagnostic function of the programs determined whether calibration samples were concentration or spectral outliers. Spectral outliers may occur due to the measurement errors by the instrument or special impurities in the sample not found in other samples. A concentration outlier occurs when the known value differs from the predicted by at least $3 \times$ SEC (standard error of cross validation). Outliers were subsequently removed from the calibration set to optimize the model.
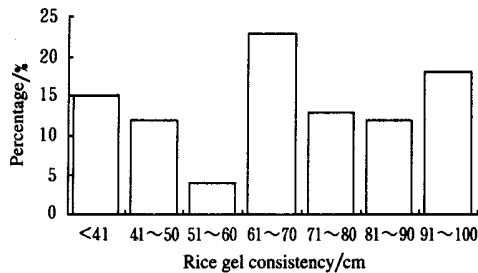

Fig. 1   Distribution of gel consistency values for rice used in the calibration and validation sets

The calibration model and the validation set to which the model was applied to predict GC was evaluated using several statistics as follows: the coefficient of correlation between predicted GC and reference values, the standard error of cross validation from the calibration set (SEC), and standard error of prediction from the validation set (SEP), the bias or mean difference between predicted GC and reference values and the ratio of the SEP to the standard deviation of reference GC values.

3   Results and discussion

Performance statistics for the calibration sets are shown in table 1. An initial PLS analysis of the entire calibration set ($n = 60$) revealed a coefficient of correlation of $R = 0.88$ and SEC of 0.84 using 10 factors in the analysis. Diagnostic functions in the software indicated that 3 of the samples were

concentration outliers and none of the spectral outlier.

Repeating the PLS analysis using the calibration set without the outliers (57) resulted in an improvement of the coefficient of correlation ($R = 0.95$) and a lower SEC (0.66). The plot of GC values determined chemically and predicted by NIRS for the improved calibration model was shown in Fig. 2. The improved calibration model was then used to predict GC for validation samples. The performance statistics were given in table 1. The coefficient of correlation was reduced slightly ($R = 0.92$). While the SEP (0.78) worsened when compared to the SEC (0.66). The correlation plot of the gel consistency values for the validation was shown in Fig. 3.

Table 1   Calibration and validity statistics for gel consistency of rice using partial least squares based on near infrared spectroscopy

Property	Samples number	SEC[1]	SEP[2]	Factors	R
Initial Calibration Set	60	0.84		10	0.88
Without Outlier Calibration Set	57	0.66		10	0.95
Validation Set	41		0.78		0.92

SEC[1] = standard error of calibration.   SEP[2] = standard error of prediction.
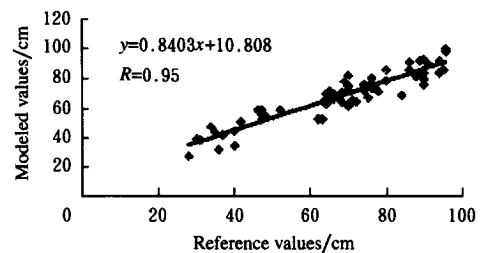

Fig. 2   Comparison between modeled and reference values for gel consistency of calibration samples
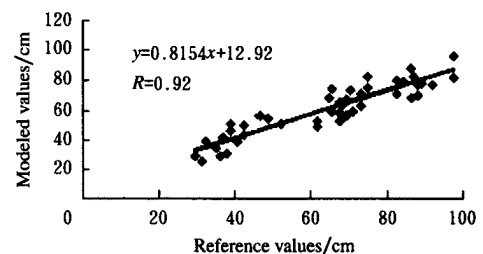

Fig. 3   Comparison between modeled and reference values for gel consistency of validation samples

4   Conclusions

PLS method based on the near infrared reflectance spectroscopy of milled rice samples can be used to establish the prediction model of grain GC characteristics. To determine the rice GC, this

method is faster than the chemical analysis The prediction models are sufficiently accurate in rice breeding programs and quality monitoring, and relieve quality inspection workers of heavy work, save time and reduce cost.

[References]

[1] Delwiche S R, Bean M M, Miller R E, et al. Apparent amylose content of milled rice by near infrared reflectance spectrophotometry[J]. Cereal Chem, 1995, 72: 182- 187.

[2] Liu Jianxue, Wu Shouyi, Fang Ruming. Measurement of rice apparent amylose content by near infrared spectroscopy [J]. Transactions of the Chinese Society of Agricultural Engineering, 2000, 16 (3): 94 - 96 (in Chinese).

[3] Liu Jianxue, Wu Shouyi, Fang Ruming. Rapid measurement of rice protein content by near infrared spectroscopy [J]. Transactions of the Chinese Society of

Agricultural Machinery, 2001, 32 (1): 68 - 70 (in Chinese).

[4] David M, Edward V. Partial least squares methods for spectral analyses 1. relation to other quantitative calibration methods and the extraction of qualitative information[J]. Anal Chem, 1988, 60: 1193- 1202.

[5] Jose Louis Rodriguez-Otero, Maria Hermida, Alberto Cepeda. Determination of fat, protein, and total solids in cheese by near infrared reflectance spectroscopy [J]. Journal of AOAC International, 1995, 78(3): 802- 806.

[6] Hu Weiwang, Xie Bijun. Food Chemistry [M]. Beijing: Science Press, 1991, 59- 64(in Chinese).

[7] Xu Lu. Chemical Metrology[M]. Beijing: Science Press, 1995, 143- 154(in Chinese).

[8] Testing Center of Rice and Products Quality of Agricultural Department. The gel Consistency Testing Regulation, RQTR014- 94[S], 1997, 30- 31.

                          1   1   1   1   2
                   ,   ,   ,   ,
          (1.   ,   471003;   2   ,   212013)

     :   ,   60
               (PLS)   ,   0. 95,
     0. 66;   41   ,   0. 92,
        0. 78   ,
     :   ;   ;   ;   ;