

Analytical, Nutritional and Clinical Methods

Calibration model optimization for rice cooking characteristics by near infrared reflectance spectroscopy (NIRS)

J.G. Wu, C.H. Shi *

Agronomy Department, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310029, PR China

Received 24 August 2005; received in revised form 18 July 2006; accepted 19 July 2006

Abstract

The aim of this study was to evaluate the potential of near-infrared reflectance spectroscopy (NIRS) as a rapid method to estimate the cooking characteristics of rice. A total of 586 samples from rice breeding lines from 1999 to 2002, which had high variation for agronomy, location and year, were scanned by NIRS for calibration optimization by chemometric methods. Two subsets of 212 samples from one year and 400 samples from three years were employed to find suitable sample status for extending the NIRS utilities. There were three algorithms of PCA, PL1 and PL2, in which the first one was only based on the sample spectra variant and the remaining two based on the variant of both spectra and chemical characteristics to describe the relationship between any two neighboring samples. According to the results of calibration and validation by the three algorithms used, the suitable calibration samples could be chosen by the cutoff of neighborhood distance (NH) of 0.35, 0.4 and 0.45, respectively. For the cooking characteristics, the combination of SNV+D/^{1,4,4,1} was the best pretreatment and the accuracy models were obtained with low SECV and high 1 – VR of amylose content (1.42% and 0.95%), gel consistency (9.49 and 0.76 mm) and alkali spread value (0.86 and 0.79 grade). The models developed using brown rice and milled rice were superior to those using intact rice grains, but slightly poorer to those using their corresponding flour samples. Therefore, on-line monitoring of rice quality could be conducted in rice processing at milling stages. Due to the fewer sample mass destroyed, the brown rice (3 g) and brown rice flour (3 g or 0.5 g), by which the NIRS models were successfully developed for cooking characteristic analyses, could be introduced into quality evaluation of germplasm and intermediate lines selection in breeding projects.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Rice; Cooking characteristic; Near-infrared reflectance spectroscopy (NIRS); Chemometrics

1. Introduction

Rice is one of the most important staple foods in the world, specifically in Asian countries. For improving the rice quality, it is necessary to test rice quality characteristics in the breeding projects and product process. Near infrared reflectance spectroscopy (NIRS), which possess the merits of minimal sample preparation, high speed and multi-characteristics testing of analysis, is being widely applied to estimate rice quality characteristics (Batten, 1998; Wu & Shi, 2003). By using milled rice

flour, accurate NIRS calibration models for amylose content (AC), the most important cooking characteristics, have been developed (Bao, Cai, & Corke, 2001; Bean et al., 1990; Delwiche, Bean, Miller, Webb, & Williams, 1995; Shu, Wu, Xia, Gao, & McClung, 1999a; Sohn, Barton, McClung, & Champagne, 2004). The NIRS models of this characteristic have also been successfully obtained when using other samples status such as milled whole-grain (Delwiche, McKenzie, & Webb, 1996; Windham et al., 1997), brown rice flour (Shu, Wu, Xia, Gao, & McClung, 1999b) and single grain (Wu & Shi, 2004). NIRS has been used to estimate rice nutrient characteristics, such as protein content of milled whole-grain sample (Barton, Himmelsbach, McClung, & Champagne, 2002; Delwiche et al., 1996) and amino acid content of

* Corresponding author. Tel./fax: +86 571 86971117.

E-mail address: chhshi@zju.edu.cn (C.H. Shi).

milled rice flour (Barton, Windham, Chanpagne, & Lyon, 1998; Wu, Shi, & Zhang, 2002). Furthermore, NIRS model of some other cooking-related characteristics, for example, texture, gelatinization temperature, gel consistency and RVA-related characteristics from various samples have been well developed (Bao et al., 2001; Delwiche et al., 1996; Meullenet, Mauromoustakos, Horner, & Marks, 2002; Windham et al., 1997).

The NIRS technique can estimate amounts of rice quality characteristic due to good relationship between reference data and spectra data obtained. But the NIRS models with different accuracy potential can be developed by using different sample geometries such as large bulk sample cups, spinning cup whole kernel, and spinning cup ground (Barton II et al., 1998). Meanwhile, sample status, another main factor affecting the determination by NIRS, must be chosen in order to increase the accuracy of NIRS model. Sample selection of calibration set and chemometric methods are other two important aspects for developing accurate calibration model (Shenk & Westerhaus, 1991).

Accurate NIRS models are based on sample's representation of the product in calibration sets. Many aspects such as physical status, chemical constitution, origin (location or year) would affect the sample's representativeness. It is important to develop new approaches on available methods to select a representative set of samples for calibration. Abrams, Shenk, Westerhaus, and Barton (1987) used the method of a statistical clustering procedure to select samples and found that the calibration based on clusters are not consistently better than those based on random sampling. Another approach developed was useful to select calibration samples in a program called SUBSET, which used r^2 as a criterion for eliminating redundant samples (Windham, Mertens, & Barton, 1989). This approach relied on the correct r^2 limit setting, which is dependent on the spectral variation of the population and must be set by trial and error. Shenk and Westerhaus (1991) developed a powerful approach to select samples based on their spectra by eliminating samples with extreme spectra and those with similar spectra. They used the program CENTER to define population with maximum standardized H distance (GH) of 3.0 and then use the program SELECT to choose samples with minimum standardized H (NH) distance of 0.6 from their nearest neighbors. These procedures could assemble a comprehensive set and provide an acceptable number of samples for calibration. But the NH distances are relative to the population broadness and spectra similarity of samples. Therefore, the suitable NH distance and number of samples should be optimized for the individual raw populations when used for calibration. Based on original sample population with high variable characteristics for year and location, the goals of this study were to find the suitable sample number in calibration set, optimize the sample status and compare the calibration effects by using various chemometric methods when NIRS technique was employed to estimate rice cooking characteristics.

2. Experimental

2.1. Rice materials

A total of 586 advanced lines of *indica* rice from the breeding projects within four years (1999–2002) were collected as original sample population. The rice samples were from eight locations in Zhejiang Province, with high variation for agronomy, location and year (Table 1). By using milled rice flour, all the rice accessions collected were grouped as sample set 1 and were used to optimize chemometrical methods for the calibration of rice cooking characteristics. Sample set 2 composed of 212 samples from 2001 and sample set 3 consisted of 400 samples from 2000 to 2002. These were employed to find suitable sample status on NIRS calibration analysis. Rice sample seeds were dehusked in an electrical dehusker (model B-76, China) and milled by sample miller (model JB-20, China) to give brown rice sample and milled rice sample. These two kinds of rice samples were further ground using model 3010-019 cyclone grinder (Fort Collins, Colorado, USA) equipped with 0.5 mm screen, yielding the other two flour samples, brown rice flour and milled rice flour, respectively. The milled rice flour was analyzed for reference and NIRS, and the others only for scanning by NIRS.

2.2. Determination of cooking characteristics

Three important cooking characteristics including amylose content (AC), gel consistency (GC) and alkali spread value (ASV) were analyzed. AC was determined according to the protocol of Williams, Wu, Tsai, and Bates (1958) with modifications by Juliano (1971). GC and ASV were measured by the methods of Cagampang, Peterz, and Juliano (1973) and Little, Hilder, and Dawson (1958), respectively. Duplicate assays were performed for all characteristics.

Table 1
Collection of rice samples and calibration sets

Location	Sample number			
	1999	2000	2001	2002
Hangzhou	56	31	69	34
Fuyang	49	33	68	23
Wenzhou	8	10	10	–
Zhoushan	10	12	10	–
Shaoxin	25	5	10	–
Ningpo	3	–	5	–
Taizhou	10	10	10	–
Jiaxin	15	20	20	–
Jinhua	10	10	10	–
Total	186	131	212	57
Set 1			586	
Set 2				212
Set 3				400

–, no sample collected.

2.3. Spectrum collection

All the samples were scanned on a NIRSystems model 5000 monochromator (NIRSystems, Inc., Silver Spring, MD, USA). Either 3 g or 0.5 g rice sample was loaded in a small ring cup of 36 mm or 18 mm inner diameter respectively, and reflectance spectra ($\log 1/R$) from 1100 to 2498 nm were recorded at 2 nm intervals. Each sample was subsequently scanned 32 times and average spectrum was collected.

2.4. Population definition and sample selection

In order to avoid calibration model disturbed by abnormal or extreme samples, population definition was introduced to this project. The CENTER algorithm (Shenk & Westerhaus, 1991) was used to compute Mahalanobis distances (so-called Global H distance, GH) The scatter of spectra were first corrected by a standard normal variate and detrending (Barnes, Dhanoa, & Lister, 1989) and derivatized with a 1, 4, 4, 1 math treatment. In the procedure of principal components analysis, three loading types of PCA, PL1 and PL2 were used when the variations were limited to the spectra only, combined spectra with a single reference or with all reference values, respectively. The population boundary was defined using $GH = 3$ and the outliers must be removed. In order to select a suitable and representative samples for calibration set, the other algorithm SELECT (Shenk & Westerhaus, 1991) was employed and the pretreatment method for spectra used as in the CENTER program. Then Mahalanobis distances between all pairs of spectra (neighbor H distance, NH) were calculated. The cutoff of NH was optimized to select the representative samples for best calibration effects.

2.5. Calibration and validation procedures

All procedures were carried out with WinISI II (version 1.04) software (Infrasoft International, Port Matila, PA, USA). By using modified partial least squares (MPLS) regression method (Shenk & Westerhaus, 1991), calibrations models for cooking characteristics were developed, previously using the spectra pretreatment ss in the CENTER process. Cross validation for the MPLS calibration was performed for avoiding overfit. The statistic 1 minus variance ratio ($1 - VR$) and standard error of cross validation (SECV) were calculated as an estimate of the coefficient of determination and validation error, respectively. The similar two parameters in calibration, the coefficient of determination (RSQ) and standard error of calibration (SEC), were given for evaluating the calibration models.

3. Results

3.1. Rice population definition and sample selection

The spectra of milled rice flour in set 1 were used to optimize the sample selection and calibration parameters. According to the spectra data, Global H distances (GH) were calculated by using CENTER program, and the population was defined through a cutoff of $GH = 3$, resulting in the new population of 584 sample. There existed spectral similarity between sample spectra in the population data, which was unsuitable to develop robust calibration models. The sample spectra in the population observed were very dense in the global space calculated by the first three principle components. Because of model accuracy being influenced by redundant samples, the calibration set should be selected. In defined population, Neighbor H distances (NH) between every two spectra were computed by the SELECT program. Calibration set was constructed by a

Table 2
Standard deviation and sample number of cooking characteristics in calibration set derived using the three algorithms of principle component analysis

NH distance	AC			GC	ASV
	PCA (spectra)	PL1 (spectra + AC)	PL2 (spectra + all characteristics)	PCA (spectra)	PCA (spectra)
1.2	6.08 (43)	7.19 (62)	7.18 (66)	17.43 (43)	1.83 (43)
1.1	6.51 (49)	6.67 (65)	6.37 (82)	18.87 (48)	1.99 (47)
1.0	6.25 (68)	6.60 (80)	6.39 (95)	18.30 (67)	2.01 (67)
0.9	6.25 (68)	6.94 (94)	6.67 (116)	18.31 (67)	2.01 (67)
0.8	6.34 (87)	6.60 (123)	6.47 (146)	18.90 (87)	1.97 (86)
0.7	6.61 (103)	6.24 (161)	6.32 (186)	19.45 (101)	1.86 (99)
0.6	6.59 (140)	6.20 (211)	6.11 (234)	19.90 (137)	1.93 (135)
0.55	6.41 s(158)	6.16 (234)	6.17 (265)	19.18 (155)	1.90 (152)
0.5	6.30 (182)	6.11 (264)	6.16 (304)	18.75 (179)	1.91 (175)
0.45	6.20 (221)	6.11 (311)	6.01 (348)	18.91 (218)	1.89 (214)
0.4	6.24 (255)	6.10 (348)	6.02 (386)	19.45 (252)	1.91 (247)
0.35	6.29 (290)	6.04 (401)	6.06 (438)	19.45 (287)	1.90 (282)
0.3	6.12 (333)	5.98 (446)	5.90 (493)	19.13 (330)	1.92 (325)
0.2	5.97 (460)	5.93 (548)	5.89 (569)	18.89 (457)	1.92 (452)
0.1	5.90 (571)	5.88 (584)	5.88 (583)	18.84 (568)	1.88 (561)

Note: selected sample number are bracketed.

cutoff of $NH = 0.35$, resulting in the selected samples dispersed in the global space.

Table 2 lists a series of calibration sets selected by a cutoff of various NH distances. Before NH distance calculation, the principle component analysis was performed on the defined population. The different algorithms could affect the principle component number and further affect NH distance. Therefore, three algorithms such as PCA, PL1 and PL2 were applied to principle component analysis, where PCA was based on the variant of spectra only, PL1 based on the combination variant of spectra and AC, PL2 based on that of spectra and all three references. When the smaller NH value was used, the larger number of samples was selected by using all three algorithms, and vice versa. The standard deviation (SD) of AC, GC and ASV were fluctuated when PCA algorithm was used. The SDs of AC from PL1 to PL2 algorithms decreased fluctuant when NH distance declined from 1.2 to 0.1. When the same NH distance used, the sample number selected from PL1 algorithm was larger than that from PCA, while it was smaller than that from PL2. Those indicated that the calculated variant of population was enlarged by combining spectral data and references' data.

3.2. NH distance optimization by calibration and validation

Three cooking characteristics of the calibration sets as their NH distances calculated by PCA algorithm were calibrated and validated (Fig. 1). The determination coefficients (RSQ and 1-VR) of calibration and validation fluctuated slightly for all characteristics, whereas the standard errors (SEC and SECV) fluctuated dramatically. With the decline of NH distance, the SEC and SECV of the AC tended to decrease gradually and they decreased significantly at the point of the $NH = 0.35$ (Fig. 1, top). Therefore, the selection by this point resulted in a 290-sample population, which was suitable to calibrate the AC. For the GC and ASV, the SEC increased fluctuant with the decline of NH from 1.2 to 0.1, while the SECV decreased. From the point of $NH = 0.55$, the curves of SEC and SECV for these two characteristics began to fluctuate faintly (Fig. 1, middle and bottom). The populations for the GC and ASV selected at this point was only 155 and 152 samples, which were considered enough as calibration set. More than two thirds of population samples were redundant for these two characteristics. Regarded as the AC, the calibration set can be selected at the point of $NH = 0.35$.

When the principal components were calculated by using the PL1 and PL2 methods, the different NIRS models of AC could be developed (Fig. 2). The RSQ and 1-VR of the AC models developed by these two methods showed that their value gradually increased with the decline of the NH value (i.e. increment samples selected). The SEC and SECV of the models developed by the PL1 exhibited the same rule, which fluctuant decreased when the smaller NH values were used. From the point of $NH = 0.4$, they

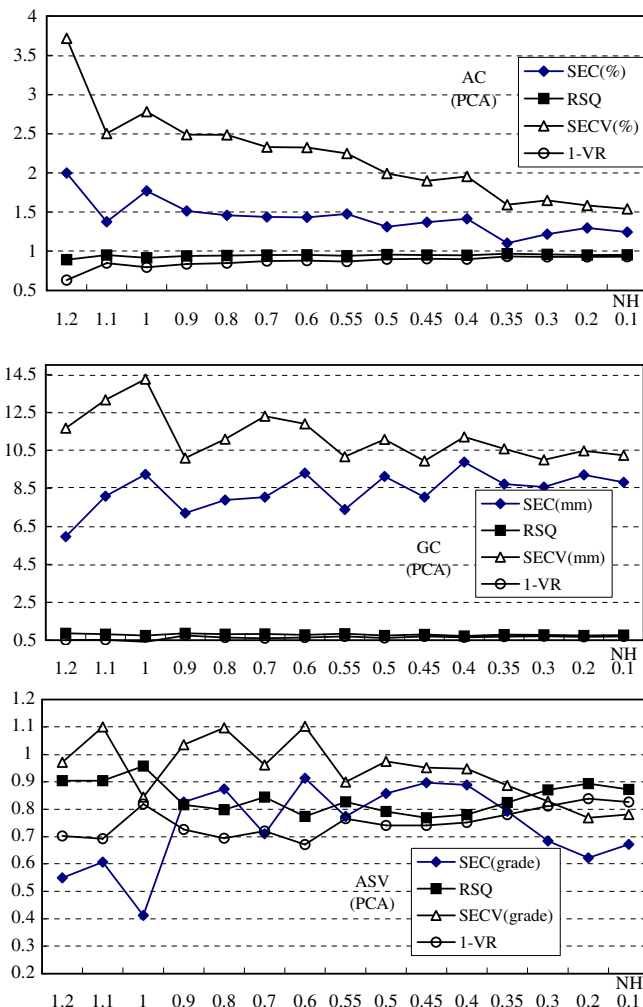


Fig. 1. Calibration and validation effects on different subset from SELECT program. top, middle and bottom were the effects on amylose content (AC), gel consistency (GC) and alkali spread value (ASV), respectively. PLS method applied to scatter corrected spectroscopy with SNV+D and treated by 1,4,4,1.

decreased dramatically (Fig. 2, top). Therefore, the samples selected by this point were sufficient for NIRS calibration. As for the SEC and SECV by the PL2 method, they showed different rule, in which the former fluctuant decreased and the latter did significantly (Fig. 2, bottom). At the point of the $NH = 0.45$, the curve trend of these two errors became plain. So the samples selected by this point were sufficient for calibration model. When the PL1 and PL2 were used, the suitable points defined by the $NH = 0.4$ and $NH = 0.45$ resulted in the same 348 samples selected for calibration set, which was larger than the sample number (290) designated by the PCA method. Thus, three algorithms for principal components analysis, which resulted in the various calibration set, could be used for this population. But only the PCA algorithm can be used to define the calibration set in order to select the representative samples and thereafter reduce the chemical analysis work, when the new characteristics were used to develop the NIRS models.

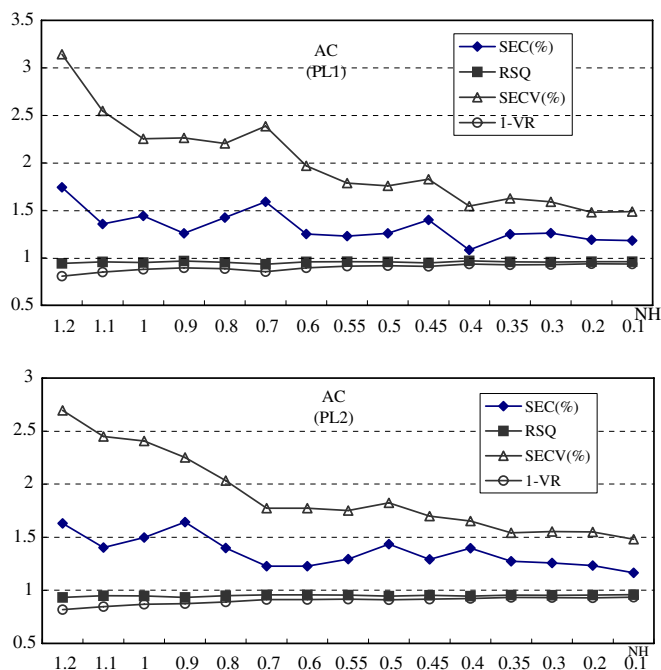


Fig. 2. Calibration and validation effects on amylose content based on principal components determined by PL1 and PL2 methods. Calibration methods were the same as those in Fig. 2.

3.3. Spectra pretreatment methods

By using PCA algorithm and the cutoff by $NH = 0.35$, the achieved calibration set having 290 samples was used to optimize the spectra pretreatment methods. The spectra pretreatment methods of scatter correction with SNV + D, mathematical treatment with derivative and smooth, and the combination of scatter correction and mathematical treatment showed the variant effects on calibration and val-

Table 3
Calibration and validation effects on NIR model of AC, GC and ASV by different spectra pretreatments

Constituent	Pretreatment methods	Calibration		Validation	
		SEC	RSQ	SECV	1 - VR
AC (%)	Control	1.66	0.93	1.91	0.90
	SNV+D	1.49	0.94	1.73	0.92
	1,4,4,1	1.16	0.97	1.43	0.95
	2,4,4,1	1.21	0.96	1.71	0.92
	SNV+D/1,4,4,1	1.18	0.96	1.42	0.95
	SNV+D/2,4,4,1	1.10	0.97	1.59	0.93
GC (mm)	Control	9.90	0.73	10.34	0.70
	SNV+D	9.52	0.75	10.05	0.72
	1,4,4,1	8.89	0.78	10.19	0.71
	2,4,4,1	9.00	0.78	10.93	0.67
	SNV+D/1,4,4,1	8.32	0.81	9.49	0.76
	SNV+D/2,4,4,1	8.72	0.79	10.58	0.69
ASV (grade)	Control	0.83	0.81	0.87	0.79
	SNV+D	0.76	0.84	0.87	0.79
	1,4,4,1	0.77	0.84	0.86	0.79
	2,4,4,1	0.82	0.81	0.93	0.76
	SNV+D/1,4,4,1	0.74	0.85	0.86	0.79
	SNV+D/2,4,4,1	0.79	0.82	0.89	0.78

idation for the AC, GC and ASV (Table 3). For all cooking characteristics, the NIRS models developed using the first derivative were generally superior to those developed using the second derivative for $\log 1/R$ of rice flour spectra. For AC, the effect of the '1,4,4,1' was better than the control and the '2,4,4,1', which was similar to SNV+D. The combination method with SNV+D/'2,4,4,1' showed better effects than their two individuals. Furthermore, the methods of '1,4,4,1' and SNV+D/'1,4,4,1' showed the best effects on the AC models. Compared with the control, the GC model was improved by the pretreatment of SNV+D, or '1,4,4,1', or their combination SNV+D/'1,4,4,1', and the first method was better the second while the last one was best. The GC models pretreated by the '2,4,4,1' and SNV+D/'2,4,4,1' were overfit because they showed lower 1 - VR and higher SECV than the control in the validation. For the ASV, no obvious effect was improved on the model developed by the method of SNV+D, whereas the methods of '1,4,4,1' and the combination SNV+D/'1,4,4,1' showed good effects. Similar to the GC models, the ASV models developed by the pretreatments of the '2,4,4,1' and SNV+D/'2,4,4,1' were also overfit and showed lower prediction accuracy.

In general, the method of SNV+D/'1,4,4,1' was the best pretreatment to the NIRS model development for the cooking characteristics. When this method was used, the models produced the following validation statistics: low SECV and high 1 - VR for AC (1.42% and 0.95%), GC (9.49 and 0.76 mm) and ASV (0.86 and 0.79 grade), respectively.

3.4. NIRS models of grain sample status

Set 2 including 212 rice samples were chosen and their spectra for grain sample status of rice grain, brown rice and milled rice were collected (Table 1). For small sample set, only CENTER program was used to define the calibration population based on the whole rice grain spectra, resulting in 210 samples being collected. In this defined population, the means and ranges of the GC, AC and ASV were 64.54 mm and 30–100 mm, 17.36% and 2.0%–27.0%, 4.81 grade and 2.0–7.0 grade, respectively. The effects of calibration and validation for cooking characteristics from three-sample status were obviously variant (Table 4). Using the rice grain spectra, good calibration models for AC and ASV were developed with low SEC and SECV of 3.13 and 3.82, 0.44 and 0.59, respectively. But for GC, calibration model developed using these spectra was poor with high value of SEC and SECV and low value of RSQ and 1 - VR. By using brown rice spectra, the calibration models for GC, AC and ASV were improved dramatically with low SEC and high RSQ of 9.17 and 0.79, 1.84 and 0.92 and 0.32 and 0.91, respectively, when compared with rice grain spectra. Furthermore, their validation results were increased again for these characteristics with lower SECV and higher 1 - VR. When the milled rice spectra were used, SEC and RSQ of GC or ASV were worse than brown rice in calibration procedure.

Table 4
Calibration and validation effects on NIRS models from grain sample status^a

Sample status	Constituent	Calibration		Validation	
		SEC	RSQ	SECV	1 – VR
Rice grain	GC	13.30	0.56	14.97	0.45
	AC	3.13	0.75	3.82	0.63
	ASV	0.44	0.82	0.59	0.69
Brown rice	GC	9.17	0.79	12.08	0.73
	AC	1.84	0.92	2.37	0.86
	ASV	0.32	0.91	0.44	0.84
Milled rice	GC	10.83	0.76	11.73	0.75
	AC	1.67	0.92	2.13	0.90
	ASV	0.31	0.84	0.33	0.88

^a Scatter correction: SNV+D; Mathematical treatment: “1,4,4,1”.

However, all the other parameters for all characteristics by using milled rice were better than by using brown rice. Therefore, calibration models for these characteristics were slightly superior to those using brown rice spectra, resulting in lowest SECV and highest 1 – VR of 11.73 and 0.75, 2.13 and 0.90, and 0.33 and 0.88 for GC, AC and ASV, respectively.

Briefly, the models developed using brown rice spectra and milled rice spectra were superior to those using rice grain spectra. Thus, these two types of sample status were considered as candidates for NIRS analysis.

3.5. Brown rice grain and brown rice flour analyzing

From 2000 to 2002, there were 400 rice samples in set 3 which were collected for NIRS analysis after the sample preparation of brown rice and brown rice flour have been made. By using the CENTER and SELECT programs with PCA calculated method, 372 samples were selected for calibration set whose reference values were not listed. Due to data missing, fewer sample numbers for GC and ASV were 370 and 369, respectively. The variant for all characteristics was good, covering a wide range presented in the common rice breeding. GC, AC and ASV had the corresponding ranges of 30–100 mm, 2.0–27.9% and 2–7 grade and means of 66.46 mm, 17.22% and 4.87 grade. The calibration set was, therefore, suitable to develop the NIRS models for these characteristics for the purpose of screening the breeding intermediate lines because of the requirement sample size used.

Table 5 shows various effects of the calibration models developed from brown rice grain and brown rice flour scanned by using the full cup (sample mass about 3 g) and smaller cup (sample mass about 0.5 g). The models of brown rice grain using full cup had high determination coefficients (RSQ and 1–VR) and low standard error (SEC and SECV) for AC, furthermore, the models for GC and ASV had better calibration effect with high 1 – VR of same value of 0.73, and lower SECV of 11.22 mm

Table 5
Effects of NIRS models in small sample of brown rice and brown rice flour^a

Sample size	Constituent	Calibration		Validation	
		SEC	RSQ	SECV	1 – VR
Brown rice grain (3 g)	GC	9.57	0.74	11.22	0.73
	AC	2.36	0.85	2.74	0.85
	ASV	0.79	0.80	0.91	0.73
Brown rice grain (0.5 g)	GC	11.80	0.63	13.41	0.52
	AC	2.46	0.84	2.85	0.78
	ASV	0.78	0.80	0.87	0.75
Brown rice flour (3 g)	GC	8.03	0.83	9.40	0.76
	AC	1.09	0.97	1.31	0.95
	ASV	0.55	0.90	0.64	0.87
Brown rice flour (0.5 g)	GC	10.08	0.72	10.91	0.67
	AC	1.27	0.95	1.60	0.93
	ASV	0.57	0.89	0.69	0.85

^a Scatter correction: SNV+D; Mathematical treatment: “1,4,4,1”.

and 0.91 grade, respectively. When using smaller cup, the calibration models for GC and AC possessed lower accuracy and reliability than using full cup because the models showed slightly lower RSQ and 1 – VR values and slightly higher SEC and SECV. But the model for ASV showed the same effect in calibration procedure and the better effect in validation procedure. When the spectra of brown rice flour were used, the calibration effects for all cooking characteristics increased significantly by comparing the same cup types. By using the full cup, the models for all characteristics (GC, AC and ASV) were best with highest RSQ (0.83, 0.97 and 0.90) and 1 – VR (0.76, 0.95 and 0.87) and lowest SEC (8.03 mm, 1.09% and 0.55 grade) and SECV (9.40 mm, 1.31% and 0.64 grade) values (Table 5). Similarly, compared with the full cup, the calibration effects decreased slightly by using the smaller cup. But their SECVs were all smaller than the models developed from brown rice grain scanned using both full cup and smaller cup. In short, the calibration models developed from the spectra of brown rice flour were superior to those of brown rice grain. Except for the brown rice grain with smaller cup, the models developed from these two types of sample status could be used in the cooking characteristics prediction because of their lower SEC and SECV and high RSQ and 1 – VR.

4. Discussion

In this study the calibration results were used as criterion to determine the NH distance and the corresponding number of samples. It was found that the NH distance of 0.35 was a suitable value to select the samples based on the spectra for the calibration of rice cooking characteristics. This rice population might be narrower than the population of forages used by Shenk and Westerhaus (1991). The WinISI II (version 1.04) software provided two more algorithms of PL1 and PL2 to choose samples forming the structure set for calibration. When based on the

combination of spectra and the AC (i.e. PL1 algorithm), and extension of all cooking characteristics (i.e. PL2 algorithm), the suitable NH distances used for sample selection were 0.40 and 0.45 for rice cooking characteristics calibration, respectively.

Through the Pearson correlation analysis, the phenotypic correlation between AC and GC was significantly negative in rice grain due to their high coefficient of -0.698 ($p < 0.01$), but their calibration results were more different. The model with the method used was the best for the NIRS model development for cooking characteristics. When the best pretreatment method of SNV+D/‘1,4,4,1’ was used, the AC model (with SECV, SECV/SD, and RSQ of 1.42%, 0.23% and 0.95%, respectively) was obviously better than that of GC model (with SECV, SECV/SD, and RSQ of 9.49, 0.49 and 0.76 mm, respectively) (Table 3). These might be caused by the determination method of GC. According to the modified protocol of Cagampang et al. (1973) the length of test tube is consistent for only 100 mm, if the length of gel consistency (GC) is larger than 100 mm, the higher part is shifted and considered as equal to 100 mm. Thus, the variant of GC length between samples are eliminated when their GC length are more than 100 mm, while the NIRS spectra exhibited all the information caused by the chemical–physical characteristics in rice grain. In order to improve GC calibration, the length of test tube must be elongated enough to ensure that all GC variant of sample can be tested.

Besides intact rice grain, there are various sample statuses such as brown rice, brown rice flour, milled rice and milled rice flour during the rice processing. The sample status of milled rice or milled rice flour had been employed in most studies on developing NIRS calibration models for rice quality characteristics. In this study, all of these sample statuses have been used and their calibration effects have been compared. In general, the models developed using brown rice and milled rice were nearly equal and superior to those using intact rice grains, but slightly poorer to those using their corresponding flour sample. So the NIRS technique can be used in rice processing to monitor the quality on line at every milling stage. Among these sample statuses, the milled rice grain and milled rice flour need a further process of milling from their brown rice samples, and it will destroy more than 25 g samples from the intact rice grain. Thus these two sample statuses are unsuitable to analyze for single plants, which cannot provide sufficient sample mass for analysis. For the few sample mass (less than 5 g) destroyed, the brown rice and brown rice flour samples should be, hence, emphasized to be introduced into NIRS analysis on rice breeding. Single plants or few sample mass analyses of NIRS can be employed in many research areas such as the intermediate material selection in a breeding project, special quality plants screening in germplasm, mutation library, and genetic material analyses in big populations. In this study, good calibration models of brown rice samples were successfully developed for rice cooking characteristics analyses.

Acknowledgements

The project is supported by the Science and Technology Office of Zhejiang Province (Nos. 2004C22009 and 011102471) and 151 Foundation for the Talents of Zhejiang Province. We also thank Associate Professor Z.W. Zhu at CNRRI for chemical analysis supports and Mr. Kennedy for improving the English of the manuscript.

References

- Abrams, S. M., Shenk, J. S., Westerhaus, M. O., & Barton, F. E. II, (1987). Determination of forage quality by near infrared reflectance spectroscopy: efficacy of broad-based calibration equations. *Journal of Dairy Science*, *70*, 806–813.
- Bao, J. S., Cai, Y. Z., & Corke, H. (2001). Prediction of rice starch quality parameters by near-infrared reflectance spectroscopy. *Journal of Food Science*, *66*, 936–939.
- Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and detrending for near infrared diffuse reflectance spectra. *Applied Spectroscopy*, *43*, 772–777.
- Barton, F. E., II, Himmelsbach, D. S., McClung, A. M., & Champagne, E. L. (2002). Two-dimensional vibration spectroscopy of rice quality and cooking. *Cereal Chemistry*, *79*(1), 143–147.
- Barton, F. E., II, Windham, W. R., Champagne, E. T., & Lyon, B. G. (1998). Optimal geometrics for the development of rice quality spectroscopic chemometric models. *Cereal Chemistry*, *75*(3), 315–319.
- Batten, G. D. (1998). Plant analysis using near infrared reflectance spectroscopy: the potential and the limitations. *Australian Journal of Experiment Agriculture*, *38*, 697–706.
- Bean, M. M., Miller, R. E., Juliano, B. O., Norris, K. H., Hruschka, W. R., & Webb, B. D. (1990). NIR and DSC characteristics of rice starch. *Cereal Food World*, *35*, 834.
- Cagampang, G. B., Peterz, C. M., & Juliano, B. O. (1973). A gel consistency test for the eating quality of rice. *Journal of Science of Food Agriculture*, *24*, 1589–1594.
- Delwiche, S. R., Bean, M. M., Miller, R. E., Webb, B. D., & Williams, P. C. (1995). Apparent amylose content of milled rice by near-infrared reflectance spectrophotometry. *Cereal Chemistry*, *72*, 182–187.
- Delwiche, S. R., McKenzie, K. S., & Webb, B. D. (1996). Quality characteristics in rice by near-infrared reflectance analysis of whole-grain milled samples. *Cereal Chemistry*, *73*, 257–263.
- Juliano, B. O. (1971). A simplified assay for milled-rice amylose. *Cereal Science Today*, *16*, 334–340.
- Little, R. R., Hilder, G. B., & Dawson, E. H. (1958). Differential effect of dilute alkali on 25 varieties of milled white rice. *Cereal Chemistry*, *35*, 111–126.
- Meullenet, J. F., Mauromoustakos, A., Horner, T. B., & Marks, B. P. (2002). Prediction of texture of cooked white rice by near infrared reflectance analysis of whole-grain milled samples. *Cereal Chemistry*, *79*(1), 52–57.
- Shenk, J. S., & Westerhaus, M. O. (1991). Population structuring of near infrared spectra and modified partial least squares regression. *Crop Science*, *31*, 1548–1555.
- Shu, Q. Y., Wu, D. X., Xia, Y. W., Gao, M. W., & McClung, A. (1999a). Calibration optimization for rice apparent amylose content by near infrared reflectance spectroscopy (NIRS). *Journal of Zhejiang University Agriculture and Life Science*, *25*, 343–346.
- Shu, Q. Y., Wu, D. X., Xia, Y. W., Gao, M. W., & McClung, A. (1999b). Analysis of grain quality characters in small ground brown rice samples by near infrared reflectance spectroscopy. *Scientia Agronomica Sinica*, *32*(4), 92–97.
- Sohn, M., Barton, F. E., II, McClung, A. M., & Champagne, E. T. (2004). Near-infrared spectroscopy for determination of protein and amylose

- in rice flour through use of Derivatives. *Cereal Chemistry*, 81(3), 341–344.
- Williams, V. R., Wu, W. T., Tsai, H. Y., & Bates, H. G. (1958). Varietal differences in amylose content of rice starch. *Journal of Agricultural and Food Chemistry*, 6, 47–48.
- Windham, W. R., Lyon, B. G., Champagne, E. T., Barton, F. E., II, Webb, B. D., McClung, A. M., et al. (1997). Prediction of cooked rice texture quality using near-infrared reflectance analysis of whole-grain milled samples. *Cereal Chemistry*, 74, 626–632.
- Windham, W. R., Mertens, D. R., & Barton, F. E. II, (1989). Protocol for NIRS calibration: Sample selection and equation development and validation. In G. C. Marten et al. (Eds.), *Near infrared reflectance spectroscopy (NIRS): Analysis of forage quality. Agriculture Handbook* (Vol. 643, pp. 96–103). Washington, DC: USDA-ARS.
- Wu, J. G., & Shi, C. H. (2003). The applications of near-infrared reflectance spectroscopy in plant breeding and germplasm research. *Journal of Plant Genetics and Resource*, 4(1), 68–72.
- Wu, J. G., & Shi, C. H. (2004). Prediction of grain weight, brown rice weight and amylose content in single seeds of rice using near-infrared reflectance spectroscopy. *Field Crops Research*, 87, 13–21.
- Wu, J. G., Shi, C. H., & Zhang, X. M. (2002). Estimating the amino acid composition in the milled rice powder by near-infrared reflectance spectroscopy. *Field Crops Research*, 75(1), 1–7.