

A Comparative Study of Fourier Transform Raman and NIR Spectroscopic Methods for Assessment of Protein and Apparent Amylose in Rice

Miryeong Sohn,^{1,2} David S. Himmelsbach,¹ and Franklin E. Barton, II¹

ABSTRACT

Cereal Chem. 81(4):429–433

Fourier-transform Raman (FT-Raman) spectroscopy and near-infrared (NIR) reflectance spectroscopy were used to compare calibration models for determining rice cooking quality parameters such as apparent amylose and protein. Samples from two seasons were used in each calibration set. The laboratory values ranged from 4.89 to 12.48% for protein and from 0.2 to 25.7% for amylose. The data for both FT-Raman and NIR were preprocessed with orthogonal signal correction (OSC) for standardization. For both spectroscopic methods, five models were optimized by partial least squares regression (PLSR) and by Martens' uncertainty regression (MUR), including no processing, smoothing, normalization, first derivative (D1), and second derivative (D2).

Based solely on standard error of cross-validation (SECV), the FT-Raman method was superior to the NIR method for protein. For amylose, the FT-Raman and NIR methods resulted in similar calibration statistics with a high precision, with the FT-Raman requiring fewer factors. The best FT-Raman models were generated from OSC preprocessing with MUR for protein (SECV 0.15%, five factors) and from OSC without MUR for amylose (SECV 0.70%, seven factors). The best NIR models were obtained with D2 transform of OSC spectra for protein (SECV 0.22%, four factors) and with OSC spectra for amylose (SECV 0.57%, 11 factors).

The protein and amylose content of rice are important factors related to rice quality (Champagne et al 1997, 1998; Windham et al 1997), and several spectroscopic analyses have been reported for the determination of the two the constituents in rice, nondestructively. The fact that near-infrared (NIR) spectroscopy could be used to measure compositional and quality factors in rice has been previously reported (Satake 1990; Villareal et al 1994; Delwiche et al 1995, 1996; Barton et al 1998, 2000). The Raman spectrum is more spectrally rich and offers better structural information than the NIR spectrum. In addition, Raman is more selective at measuring nonpolar bond vibrations in the presence of water due to the Raman effect being sensitive to the polarizability of bonds as opposed to polarity. Fourier-transform systems have advantages over dispersive systems due to better wavelength accuracy, higher throughput, and (in the detector-noise limited regime at longer wavelengths) also providing the multiplex advantage (Chase 1986; Birch and Chunnillal 1999; Weesner and Longmire 2001). Recently, FT-Raman has been used to evaluate rice quality as a nondestructive and rapid spectroscopic method. Barton et al (2000) and Himmelsbach et al (2001) noted that near-infrared FT-Raman (NIR/FT-Raman) spectroscopy has the potential to assess rice quality by protein and amylose. The models obtained by this method had low standard error prediction (SEP) for both constituents: SEP 0.138% using six factors for protein; SEP 1.05% using eight factors for apparent amylose. The previous data were only from one-year samples with a narrow range for protein and an uneven distribution of amylose content, and no further results have been reported for updating FT-Raman models. In a previous study (Sohn et al 2004), we reported NIR calibration models for protein and amylose analysis and the best PLS algorithms through the use of derivative processing, where samples from two years with a greater range and distribution of the sample values were used.

The objective of this study was to develop robust FT-Raman models for rice qualities using samples grown in two seasons and to compare the precision with the NIR models. In addition, the

efficacy of various chemometric techniques were tested to see whether they offered improved results over the standard PLS1 algorithms for both spectroscopic methods.

To allow an equitable comparison, we used exactly the same set of calibration samples and the same preprocessing of the data for FT-Raman and NIR.

MATERIALS AND METHODS

Samples and Reference Analysis

A total of 214 rice flour samples were prepared from crop years 1996 ($n = 90$) and 1999 ($n = 124$). Protein ($N \times 5.95$) was determined by the method of combustion using a Leco model FP-2000 nitrogen analyzer in duplicate assays on a 0.5-g assay of ground rice (Approved Method 46-30, AACC 2000). Apparent amylose was determined by the method of Juliano (1971). All replicates were averaged for use in the data analysis.

Spectroscopy

Raman spectra were collected using a 950 FT-Raman spectrometer (Nicolet Analytical Instruments, Madison, WI) interfaced with a PC running Thermo Nicolet OMNIC (v. 5.2) software. Samples were placed in a spinning cup (5 cm, o.d.) and rotated during scanning. A laser light source of 1,064 nm at 500 mW was used with a CaF₂ beam splitter and a Ge detector cooled by liquid N₂. The spectral data were scanned in the 0–3,600 cm⁻¹ Raman shift (Stokes region). The data obtained were the result of 128 scans at a resolution of 16 cm⁻¹. Triplicate spectra were collected on each sample and then averaged to give a single spectral file for each sample. Spectral data were preprocessed by white light correction against a KBr background and subsequently truncated to 230–3,100 cm⁻¹. NIR spectra were collected using a scanning monochromator (model 6500, NIRSystems, Silver Spring, MD) operated by WINISI software. The reflectance spectral data were scanned in the 1,100–2,498 range at 2-nm intervals. The NIR spectra used were the average of data from three separately packed spinning cups.

Data Processing and Chemometric Analyses

The FT-Raman data were preprocessed with multipoint baseline correction and normalization in GRAMS/32 (v. 4.1, Galactic Industries Corp., Salem, NH) before chemometric analysis as described by Himmelsbach et al (2001). Preprocessing of the spectral data and chemometric analyses for FT-Raman and NIR were performed with Unscrambler software (v. 8.0, CAMO, Trondheim, Norway).

¹ USDA-Agricultural Research Service, Richard B. Russell Agricultural Research Center, P.O. Box 5677, Athens, GA 30605.

² Corresponding author. Phone: (706) 546-3374. Fax: (706) 546-3607. E-mail: msohn@qaru.ars.usda.gov

Preprocessing included smoothing with a three-point moving average (SM), mean normalization (MN), and Savitzky-Golay derivative methods (SG) with 1st and 2nd derivatives (Savitzky and Golay 1964). Application of orthogonal signal correction (OSC) was performed with Unscrambler Accessory Pack for Spectroscopy. The number of orthogonal components and iterations used were 2 and 10, respectively. Five-point quartic fit and five-point quadratic fit were used as derivative conditions for 1st (D1) and

2nd (D2) derivative models for protein, and five-point quadratic fit and 17-point quadratic fit were used for D1 and D2 amylose models. In all cases, Raman and NIR spectra were mean-centered.

Calibrations were developed using the partial least squares (PLS1) (Martens and Martens 1986; Martens and Naes 1989) and Martens' uncertainty regression (MUR) programs in the Unscrambler software package. The MU technique eliminates those spectral variables that do not contribute to the partial least squares regression (PLSR)

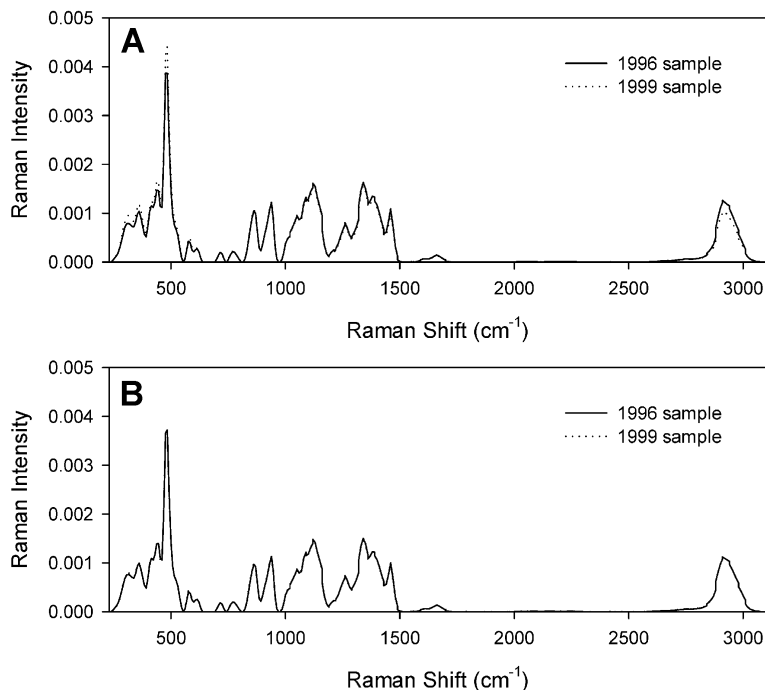


Fig. 1. NIR-FT/Raman spectra of rice flour samples from 1996 and 1999 with the same constituents. **A**, Original spectra preprocessed with baseline correction and mean normalization. **B**, Orthogonal signal correction (OSC) spectra with baseline correction and mean normalization.

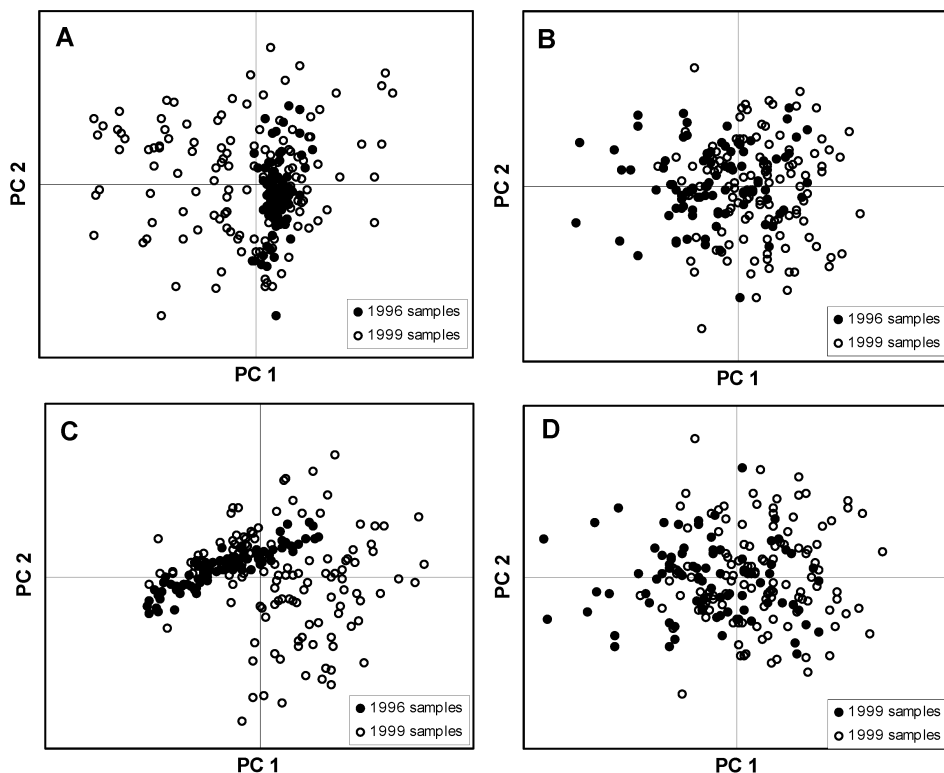


Fig. 2. Score plots of spectra with or without orthogonal signal correction (OSC) of two years data for FT-Raman modeling. **A**, Amylose without OSC. **B**, Amylose with OSC. **C**, Protein without OSC. **D**, Protein with OSC.

model to simplify the final model and make it more reliable (Martens and Martens 2000; Westad and Martens 2000). Evaluation of the model performance was tested by random cross-validation, with 20 segments and 10–11 samples per segment, and no outliers were removed. The number of factors used in this study was suggested from the PLS analysis software.

RESULTS AND DISCUSSION

The FT-Raman spectra for sample sets from two years were collected under different instrumental conditions due to changes in the instrument laser and detector, thus a standardization was required to reduce spectral differences between instruments (Dardenne 2002). In this study, we used OSC as a preprocessing method for the standardization of the data from two years. The OSC removes extraneous variance from x -data through correlation between x -variance and y -chemical data, and it would be expected to minimize the spectral difference between the data from two years.

Figure 1A shows the FT-Raman spectra of one 1996 sample and one 1999 sample that have identical protein and amylose contents, and Fig. 1B shows the OSC spectra of the same samples. Raman intensity differences were observed between the two groups of samples, with the largest differences at 230–630 cm^{-1} and at 2,920 cm^{-1} and smaller differences at 1,000–1,500 cm^{-1} . After standardization, however, the spectra were matched exactly, removing the difference between the two groups.

Table I shows the statistical results of PLSR and MUR models developed for determining protein and amylose in rice using FT-Raman spectroscopy. “None” means the spectra preprocessed with baseline correction and normalization and no chemometric pro-

cessing. The OSC data produced results clearly better than the nonchemometrically preprocessed data for both constituents. The coefficient of determination (R^2) and standard error of cross-validation (SECV) were 0.992, 0.15% for protein, and 0.991, 0.70% for amylose, respectively. This result is related to the score plots of the samples shown in Fig. 2. First, two principal components (PC) from PLS were used for the two-dimensional scatter plots of scores. The plots of data from two years without OSC gave different patterns in modeling for both constituents. The 1996 samples were clustered together, whereas the 1999 samples showed more variability. The samples from two years showed similar plots after OSC processing, indicating that the differences due to instrumental conditions or other variables were reduced for the data from two years.

For the protein PLS model, SM and derivative (either D1 or D2) processing were not helpful in improving performance. Additional processing with MN of OSC spectra also degraded the model. Further processing of the data with MU slightly improved the results, primarily because the region at 1,800–2,700 cm^{-1} contained no significant spectral response. This gave a model with R^2 0.993 and SECV 0.15%. This result is graphically depicted in the top portion of Fig. 3 in a plot of actual versus predicted value.

For the amylose PLS model, the OSC produced calibration statistics superior to those obtained with non-OSC processed data; the number of factors and SECV were both reduced. No further processing appeared to improve the model as with protein. The MUR did not improve the model compared with statistics of the PLS alone. The best FT-Raman model for amylose resulted in SECV 0.70% and R^2 0.991 using seven factors (Fig. 3). The FT-Raman result obtained with data from two years was similar to that obtained with data from one year for protein (SEP 0.14%, R^2

TABLE I
Calibration Statistics of PLSR and MUR Models for Protein and Amylose Developed Using NIR-FT/Raman Spectroscopy

Constituent	Preprocessing ^a	PLSR				MUR			
		No. of Factors	R^2	SECV (%)	Bias (%)	No. of Factors	R^2	SECV (%)	Bias (%)
Protein	None	6	0.950	0.38	0.000	5	0.948	0.38	-0.001
	OSC	5	0.992	0.15	0.000	5	0.993	0.15	0.000
	OSC+SM	5	0.991	0.16	0.001	5	0.991	0.16	-0.001
	OSC+MN	3	0.417	1.12	0.007	2	0.451	1.09	-0.001
	OSC+SG-D1	5	0.985	0.21	-0.000	4	0.980	0.24	0.000
	OSC+SG-D2	5	0.977	0.25	-0.003	4	0.977	0.26	0.001
Amylose	None	10	0.906	2.30	0.018	8	0.916	2.18	-0.025
	OSC	7	0.991	0.70	-0.017	6	0.990	0.75	-0.000
	OSC+SM	6	0.986	0.88	-0.005	6	0.988	0.84	-0.000
	OSC+MN	4	0.886	2.52	0.007	4	0.889	2.48	0.011
	OSC+SG-D1	5	0.961	1.49	0.012	6	0.969	1.34	-0.037
	OSC+SG-D2	6	0.959	1.53	-0.017	5	0.950	1.69	-0.005

^a OSC, orthogonal signal correction; SM, 3-point moving average smoothing; MN, mean normalization; SG-D1, Savitzky-Golay 1st derivative (5-point quartic fit for protein and 5-point quadratic fit for amylose); SG-D2, Savitzky-Golay 2nd derivative (5-point quadratic fit for protein and 17-point quadratic fit for amylose).

TABLE II
Calibration Statistics of PLSR and MUR Models for Protein and Amylose Developed Using NIR Spectroscopy

Constituent	Preprocessing ^a	PLSR				MUR			
		No. of Factors	R^2	SECV (%)	Bias (%)	No. of Factors	R^2	SECV (%)	Bias (%)
Protein	None	6	0.976	0.26	-0.000	4	0.973	0.29	0.000
	OSC	5	0.981	0.24	-0.001	4	0.976	0.26	-0.002
	OSC+SM	5	0.980	0.24	-0.002	3	0.975	0.27	-0.000
	OSC+MN	2	0.493	1.08	0.011	2	0.500	1.07	0.004
	OSC+SG-D1	4	0.979	0.25	-0.000	4	0.979	0.25	-0.000
	OSC+SG-D2	5	0.983	0.23	0.000	4	0.983	0.22	-0.002
Amylose	None	14	0.976	1.13	-0.000	13	0.974	1.18	0.011
	OSC	11	0.994	0.57	-0.000	8	0.980	1.04	0.002
	OSC+SM	11	0.994	0.58	0.000	8	0.980	1.04	-0.007
	OSC+MN	13	0.977	1.12	-0.002	10	0.967	1.34	-0.015
	OSC+SG-D1	6	0.975	1.15	0.007	5	0.972	1.24	0.006
	OSC+SG-D2	6	0.974	1.18	0.000	5	0.973	1.20	-0.001

^a OSC, orthogonal signal correction; SM, 3-point moving average smoothing; MN, mean normalization; SG-D1, Savitzky-Golay 1st derivative (5-point quartic fit for protein and 5-point quadratic fit for amylose); SG-D2, Savitzky-Golay 2nd derivative (5-point quadratic fit for protein and 17-point quadratic fit for amylose).

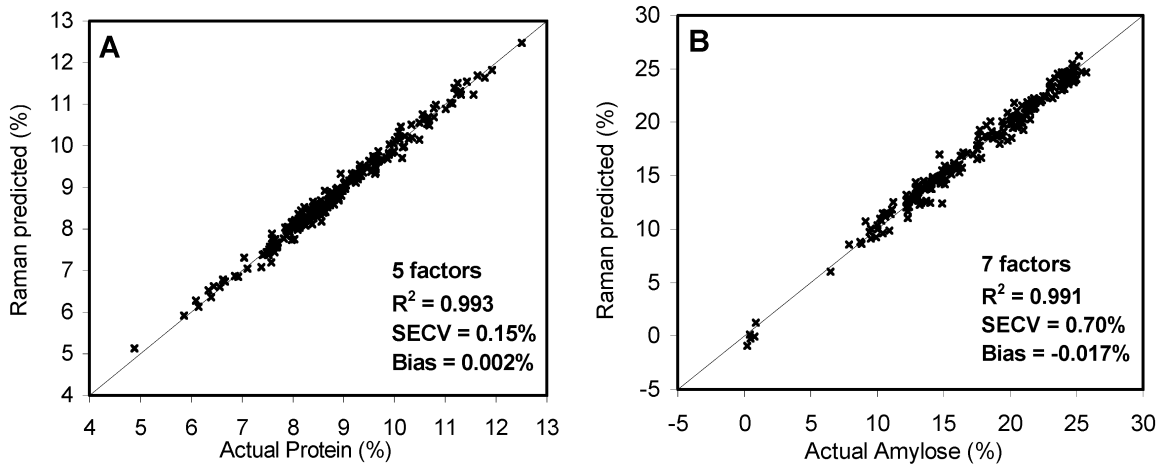


Fig. 3. Plots of actual versus predicted values of best FT-Raman models for protein (A) and amylose (B).

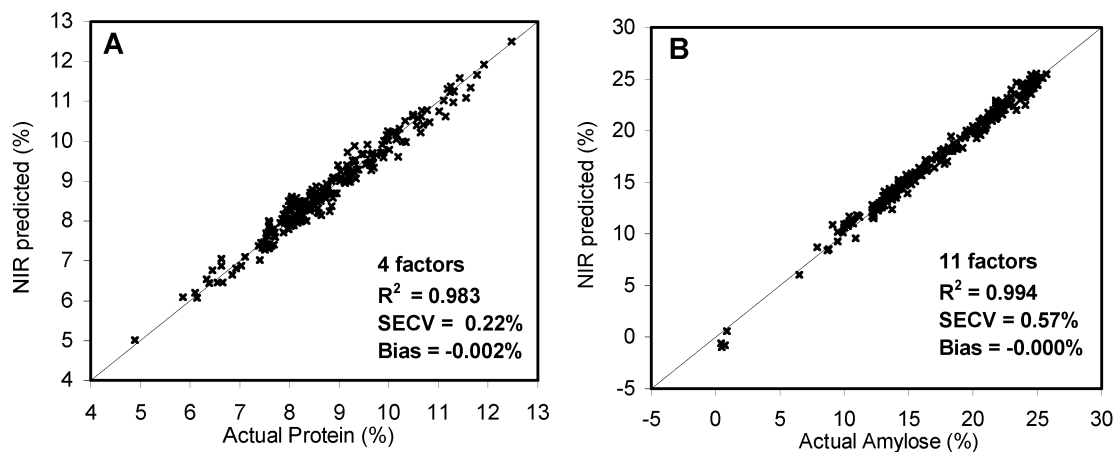


Fig. 4. Plots of actual versus predicted values of best NIR models for protein (A) and amylose (B).

0.992, six factors), but the new amylose model was more robust than the previous one (SEP 1.05%, R^2 0.985, 8 factors) (Himmelsbach et al 2001), even though the expanded sample set was used.

The NIR spectra of the samples from two years were also standardized through the OSC preprocessing like the FT-Raman data, even though there were no instrumental changes, because the spectra were collected over a four-year period. Table II shows results of statistics for NIR models. “None” means no preprocessing of the spectra. For both protein and amylose, the OSC spectra produced better results than the non-OSC processed data, particularly for amylose. The score plots of the spectra with OSC showed the same patterns as the FT-Raman. No more preprocessing was needed for improvement of both protein and amylose models except D2 protein. Second derivatization of the OSC spectra produced a calibration model that was slightly better than that from the OSC spectra, but no big difference.

MUR slightly improved the protein model showing fewer number of factors and error (SECV 0.22% and R^2 0.983) but did not improve the amylose model. The best PLS model resulted in SECV 0.57% and R^2 0.994, respectively. The plots of measured versus predicted values for protein and amylose from the best NIR models are shown in Fig. 4. The protein model using OSC produced a result similar to the previous PLS model developed using multiplicative scatter correction and derivative processing (SECV 0.23% using four factors) (Sohn et al 2004), whereas the amylose model with OSC processing has much higher precision than the previous model (SECV 1.0% using 10 factors).

In this study, the FT-Raman method produced much better results than the NIR method for protein analysis. The FT-Raman

amylose model produced a slightly higher SECV value but required fewer number of the factors (seven PC) than did the NIR (11 PC), thus both NIR and FT-Raman methods produced similar results for amylose analysis.

CONCLUSIONS

The measurement of protein and apparent amylose contents in rice flour could be accomplished with FT-Raman calibration models developed using data from two years with SECV 0.15% and 0.70% for protein and amylose, respectively. These error levels are similar to or lower than those obtained from the NIR analysis. The high precision of the models was achieved from the orthogonal signal correction of the spectra for both constituents and for both instrumental methods.

ACKNOWLEDGMENTS

We would like to thank Anna M. McClung and Elaine T. Champagne for providing samples and reference data. We also wish to thank Judy Davis for acquisition of spectral data.

LITERATURE CITED

- Barton, F. E., II, Windham, W. R., Champagne, E. T., and Lyon, B. G. 1998. Optimal geometries for the development of rice quality spectroscopic chemometric models. *Cereal Chem.* 75:315-319.
- Barton, F. E., II, Himmelsbach, D. S., McClung, A. M., and Champagne, E. T. 2000. Rice quality by spectroscopic analysis: Precision of three spectral regions. *Cereal Chem.* 77:669-672.

- Birch, J. R., and Chunnillal, C. T. 1999. FT techniques for radiometry. Pages 317-318 in: *Fourier Transform Spectroscopy*. K. Itoh and M. Tasumi, eds. Waseda University Press: Tokyo.
- Champagne, E. T., Bett, K. L., Vinyard, B. T., Webb, B. D., McClung, A. M., Barton, F. E., II, Lyon, B. G., Moldenhauer, K., Linscombe, S., and Kohlwey, D. E. 1997. Effects of drying conditions, final moisture content, and degree-of-milling on rice flavor. *Cereal Chem.* 74:566-570.
- Champagne, E. T., Lyon, B. G., Min, B. K., Vinyard, B. T., Bett, K. L., Barton, F. E., II, Webb, B. D., McClung, A. M., Moldenhauer, K. A., Linscombe, S., McKenzie, K. S., and Kohlwey, D. E. 1998. Effects of postharvest processing on rice texture profile analysis. *Cereal Chem.* 75:181-186.
- Chase, B. C. 1986. Fourier transform Raman spectroscopy. *J. Am. Chem. Soc.* 108:7485-7488.
- Dardenne, P. 2002. Calibration transfer in near infrared spectroscopy. *NIR News* 13:3-7.
- Delwiche, S. R., Bean, M. M., Miller, R. E., Webb, B. D., and Williams, P. C. 1995. Apparent amylose content of milled rice by near-infrared reflectance spectrophotometry. *Cereal Chem.* 72:182-187.
- Delwiche, S. R., McKenzie, K. S., and Webb, B. D. 1996. Quality characteristics in rice by near-infrared reflectance analysis of whole-grain milled samples. *Cereal Chem.* 73:257-263.
- Himmelsbach, D. S., Barton, F. E., II, McClung, A. M., and Champagne, E. T. 2001. Protein and apparent amylose contents of milled rice by NIR-FT/Raman spectroscopy. *Cereal Chem.* 78:488-492.
- Juliano, B. O. 1971. A simplified assay for milled rice amylose. *Cereal Sci. Today* 16:334-338, 340, 360.
- Martens, M., and Martens, H. 1986. Partial least squares regression. Pages 292-360 in: *Statistical Procedures in Food Research*. J. R. Piggott, ed. Elsevier Applied Sciences: London.
- Martens, H., and Martens, M. 2000. Modified jack-knife estimation of parameter uncertainty in bilinear modeling by partial least squares regression (PLSR). *Food Qual. Pref.* 11:5-16.
- Martens, H., and Naes, T. 1989. *Multivariate Calibration*. John Wiley and Sons: Chichester, UK.
- Satake, T., 1990. *Modern Rice-Milling Technology*. University of Tokyo Press: Tokyo.
- Savitzky, A., and Golay, M. J. E. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36:1627-1638.
- Sohn, M., Barton, F. E., II, McClung, A. M., and Champagne, E. T. 2004. Near infrared spectroscopy for determination of protein and amylose in rice flour through use of derivatives. *Cereal Chem.* 81:341-344.
- Villareal, C. P., De La Cruz, N. M., and Juliano, B. L. 1994. Rice amylose analysis by near-infrared transmittance spectroscopy. *Cereal Chem.* 71:292-296.
- Weesner, F., and Longmire, M. 2001. Dispersive and Fourier transform Raman. *Spectroscopy* 16:68-77.
- Westad, F., and Martens, H. 2000. Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression. *J. Near Infrared Spectrosc.* 8:117-124.
- Windham, W. R., Lyon, B. G., Champagne, E. T., Barton, F. E., II, Webb, B. D., McClung, A. M., Moldenhauer, K. A., Linscombe, S., and McKenzie, K. S. 1997. Prediction of cooked rice texture quality using near-infrared reflectance analysis of whole grain milled samples. *Cereal Chem.* 74:626-632.

[Received September 25, 2003. Accepted January 13, 2004.]