

# Influence of Yearly Variability of Agricultural Products on Calibration Process: A Triticale Example

B. Igne,<sup>1,2</sup> L. R. Gibson,<sup>3</sup> G. R. Rippke,<sup>1</sup> and C. R. Hurburgh, Jr<sup>1</sup>

## ABSTRACT

Cereal Chem. 84(6):576–581

Many elements can influence the calibration process for near-infrared spectroscopy: variability within the population of interest, environmental effects, hardware instability, and others. This study evaluated two techniques to develop prediction models for triticale grain moisture and protein. The spectral addition strategy added new samples to the calibration set year after year. The spectral adaptation strategy selected only the spectral variability needed to successfully apply the model to new material. Triticale was a good study material because it was undergoing genetic change through the study period and is very responsive to climate vari-

ability. The two calibration techniques were significantly different from each other in terms of precision and accuracy. Spectral adaptation was the best technique with a relative predictive determinant (RPD) of 4.33 and a bias of 0.17% versus RPD of 3.50 and a bias of -0.52% for spectral addition. These results are contradictory to common practices that tend to add to the calibration set a maximum variability over as much time as possible. For highly variable matrixes, a constant adaptation rather than expansion of the calibration pool may be more appropriate.

With the application of near-infrared spectroscopy (NIRS) to the characterization of organic compounds, statisticians and chemometricians have developed techniques to deal with multivariate data containing high colinearity. Techniques such as partial least squares regression or principal component regression (Martens and Jensen 1983; Adams 1995) give very good results for applications involving linear relationships among spectral data and reference chemical data. Many authors have published successful applications of these algorithms (Rippke et al 1996; Kays et al 2000; Burns and Ciurczak 2001; Siesler et al 2002; Delwiche and Reeves 2004).

When dealing with agricultural products, calibration developers face an issue less prevalent with synthetic products: the environmentally or culturally created yearly variability of biochemical properties. This variability can be caused by the environmental effects (climate, soil, plant diseases, and production practices) and by genetic changes. The consequences are an increase of outliers or the appearance within the subpopulations calibration set. An accurate prediction model requires that this progressive variability be included in the calibration set. This often causes a decrease of prediction performance (both precision and accuracy). The problem can be accentuated if rapid genetic changes are occurring. It is never possible to be totally sure that the next year is fully represented or measured accurately and precisely until after the fact.

This study 1) demonstrated the effect of year-to-year variability on the calibration process, 2) evaluated spectral pretreatment techniques to control this variability, 3) compared validation strategies, and 4) compared two calibration processes in their ability to produce accurate and precise results in the following year. A five-year calibration effort with triticale was used as a test case.

## MATERIALS AND METHODS

### Samples

Triticale (*xTriticosecale* Wittmack) is a cereal resulting from the intergeneric crossing of wheat and rye. More information about triticale and its potential uses can be found in Igne et al (2006). Triticale grain samples used in this study were harvested from

crop years 2002 to 2006. Variety trials managed by the Department of Agronomy at Iowa State University at five Iowa locations were sampled (Gibson et al 2004a,b). Both winter and spring cultivars were grown in 2002, 2003, and 2004, while in 2005 and 2006, only winter lines were cultivated. Winter samples of 2002, 2003, 2004, and 2006 included both commercial cultivars and breeding lines from the University of Nebraska. No breeding lines were included in 2005. Spring triticale samples included only commercially available varieties. Lines changed annually; some were included in all crop years, but most were not. A total of 578 and 504 triticale samples, for moisture and protein, respectively, were selected from 2,994 available samples (17.4% spring lines, 82.6% winter lines). Samples were selected on their estimated moisture and protein contents using commercially available neural-network-based wheat moisture and protein calibrations (models WBMO 0024 for moisture and WBPR0028 for protein) (Foss North America, Eden Prairie, MN) to obtain a uniform distribution within each crop year. No samples were selected for moisture content in 2002.

### Reference Chemistry Analysis

Approved Method 44-15A (moisture determination by air oven) (AACC International 2000) was used to determine the moisture content of the whole-grain samples. Measurements were taken within three days after the spectral acquisition. Protein concentration was measured by AACC Approved Method 46-30 (crude protein determination by combustion). An analyzer (CHN-2000, Leco Corporation, St. Joseph, MI) was used to perform the crude protein determination. The protein content was calculated by multiplying the nitrogen value by 5.7 (the same multiplier as used for wheat). Concentration values were then converted to a 12% moisture basis using a second oven moisture of the ground material taken immediately before combustion analysis. Reference data statistics are provided in the Table I.

### Spectral Acquisition

Near-infrared spectra were acquired on whole seeds (Infratec grain analyzer 1241, Foss North America, Eden Prairie, MN). This transmittance instrument scans a spectral range of 850–1048 nm at 2-nm increments. A path length of 18 mm was used. This is the instrument used by the United States Department of Agriculture in official inspections (GIPSA 2006).

Spectral acquisition and chemical analysis for selected samples of each year were performed shortly after each harvest season. The calibration set contained samples scanned in 2003, 2004, 2005, and 2006 for moisture and 2002, 2003, 2004, 2005, and 2006 for protein.

<sup>1</sup> Department of Agricultural and Biosystems Engineering, 1551 Food Sciences Bldg, Iowa State University, Ames, IA 50011-1061.

<sup>2</sup> Corresponding author. Phone: (515) 294-6358. Fax: (515) 294-6383. E-mail address: igneb@iastate.edu

<sup>3</sup> Department of Agronomy, 1126C Agronomy Hall, Iowa State University, Ames, IA 50011-1010.

## Data Analysis

The development of triticale prediction models was studied by using two techniques: spectral addition and spectral adaptation. The spectral addition system collected the cumulative yearly variability; samples of the years 2004, 2005, and 2006 were successively added to 2003 samples for the creation of the moisture prediction models. Samples of the years 2003, 2004, 2005, and 2006 were successively added to the spectra of the year 2002 for protein models. This is the system usually employed by near-infrared users to develop calibration models. The spectral adaptation system measured the impact of the variability brought by a specific year in terms of climate and genetic variability. Fifteen year combinations, including successive or nonsuccessive years (2002 to 2005 samples) were used as the calibration set to predict 2006 samples. This system was applied to the protein calibration process only. Calibrations were developed with both winter and spring lines and with winter lines only to evaluate the varietal effects.

Both processes were designed to evaluate the yearly impact of samples on prediction model performances and on the calibration process. Samples for each year were added or removed from the calibration set as a group. Chemometrics software such as WinISI (Infrasoft International, State College, PA), Unscrambler (CAMO Software, Woodbridge, NJ), or PLS\_Toolbox (Eigenvector Research, Wenatchee, WA) provide features to select samples necessary or unnecessary to the calibration set. In this study, we were more interested in the impact of adding or removing samples on the basis of their year of harvest to study their effect on the calibration set rather than studying the impact of some samples.

## Data Pretreatment

Three spectral pretreatment techniques were compared: second-derivative using Savitzky-Golay algorithm (15-point window and third-order polynomial), standard normal variate (SNV), and multiplicative scatter correction (MSC). For all pretreatments, spectral and reference data were standardized (autoscaled) to have zero mean and unity standard deviation (Naes et al 2002).

## Calibration Procedure

Prediction models were developed using partial least squares regression (PLS) on MATLAB v. 7.0.4 (The MathWorks, Natick, MA) with the PLS\_Toolbox v. 3.5.4 (Eigenvector Research, Wenatchee, WA). The principle of PLS is to compress the spectral data information (X) related to the reference chemical composition of interest (y) to a set of latent variables (new X variables). The optimal number of latent variables is found by minimizing the standard error of cross-validation, and limiting over-fitting. Outliers were identified and removed during the calibration process by selecting samples with unusual residual values or high Hotelling's T<sup>2</sup> values (Cogdill et al 2005).

## Validation Procedure

Three validation strategies were evaluated across pretreatments: cross-validation by leave-one-out, random withholding of 25% of the calibration set for validation only, and next crop year samples. The first two are common techniques in NIRS (Williams and Norris 2001; Naes et al 2002). The third is used by the Iowa State Uni-

versity Grain Quality Laboratory for long-term calibration studies (Rippke et al 1995). Calibration performance was evaluated in terms of precision, accuracy, and model fit. The standard error of prediction (SEP) or standard deviation of differences and the relative predictive determinant (RPD) were used to evaluate the precision (Williams and Norris 2001). The accuracy was determined by the bias (average of differences). Bias is a good indicator of similarity between validation samples and the calibration set. Finally, the fit of the model was indicated by the coefficient of determination ( $r^2$ ) on the validation samples.

In the spectral adaptation strategy, RPD was used to assess the precision of the different models because we were predicting a fixed validation set. RPD was preferred to SEP because it is calculated by dividing the standard deviation of the validation data by the SEP. RPD shows the ability of the calibration data to predict a specific validation set based on its range and variability (a large standard deviation means that the validation set is more likely to be representative of the variability that the calibration can encounter in routine analysis). It also allows an easy comparison between models using the scale provided by Williams and Norris (2001) while SEP does not.

## RESULTS AND DISCUSSION

### Spectral Addition

Triticale moisture and protein prediction models were developed by successively adding crop years' samples to the original calibration set, with the preprocessing techniques applied. For moisture, the four pretreatments gave equal SEP values for cross-validation and for the validation by 25% of the calibration set ( $\alpha = 0.05$ ). For the next year's sample validation method, autoscaling alone consistently outperformed the other three pretreatments ( $\alpha = 0.05$ ) (data not shown). Fig. 1A and Table II present the validation results for autoscaling. Cross-validation and 25% calibration set validation did not present significant differences in SEP and biases for all years. The prediction of next year sample moisture was not consistent from year to year; SEP and bias were significantly different from the other two techniques ( $\alpha = 0.05$ ). Only the prediction of 2005 samples by 2003 and 2004 samples gave the same results as with the validation methods on same year samples. This could mean that 2005 samples presented a variability already present in the calibration pool for the prediction of moisture content.

Second-derivative was the best pretreatment for protein. For this pretreatment, SEP for validation by next year samples were significantly lower than those obtained with SNV and MSC ( $\alpha = 0.05$ ). SEP for cross-validation by leave-one-out and validation using 25% of the calibration set were significantly lower than with autoscaling alone ( $\alpha = 0.05$ ) (data not shown). Contrary to moisture, there were no statistically significant differences among the three validation techniques (Fig. 1B and Table III). In 2002 and 2004, the cross-validation by leave-one-out gave lower SEP, while validation using 25% of the calibration set was better in 2003 and 2005. Validation by next year samples gave surprisingly similar SEP to other validation methods using same year samples in 2003 (leave-one-out), 2005, and 2006 (25% of the calibration set). For 2004, validation by next year samples gave the lowest SEP.

TABLE I  
Reference Data Statistics for Triticale Samples Used for Calibration and Model Validation

	Moisture				Protein (12% mb)				
	2003	2004	2005	2006	2002	2003	2004	2005	2006
Observations (n)	118	94	98	268	126	93	94	98	93
Minimum (%)	10.62	9.51	8.94	11.63	9.96	8.28	9.89	11.04	9.29
Maximum (%)	13.55	11.45	10.71	14.72	17.91	20.89	16.43	16.25	14.50
Mean (%)	11.59	10.40	9.80	12.61	12.84	13.16	12.89	13.48	11.32
SD (%)	0.50	0.49	0.35	0.68	1.83	2.12	1.67	1.18	1.06

This situation shows instability of prediction models related to the introduction of new calibration samples with characteristics sometimes similar to or different from those already present. While validation on same year sample techniques presented biases of zero or very close to zero, biases obtained with next year sample validation methods were significantly larger. We note that in 2005 no bias was observed for all three validation techniques. This confirms that even if SEP are equivalent, bias is an important factor for trust in the model.

In moisture, validation techniques on same year sample results were not representative of either accuracy or precision of the calibration applied to the next years' material. In protein, these same techniques were a very good indication of the calibration precision on next year samples but did not accurately provide an indication of accuracy.

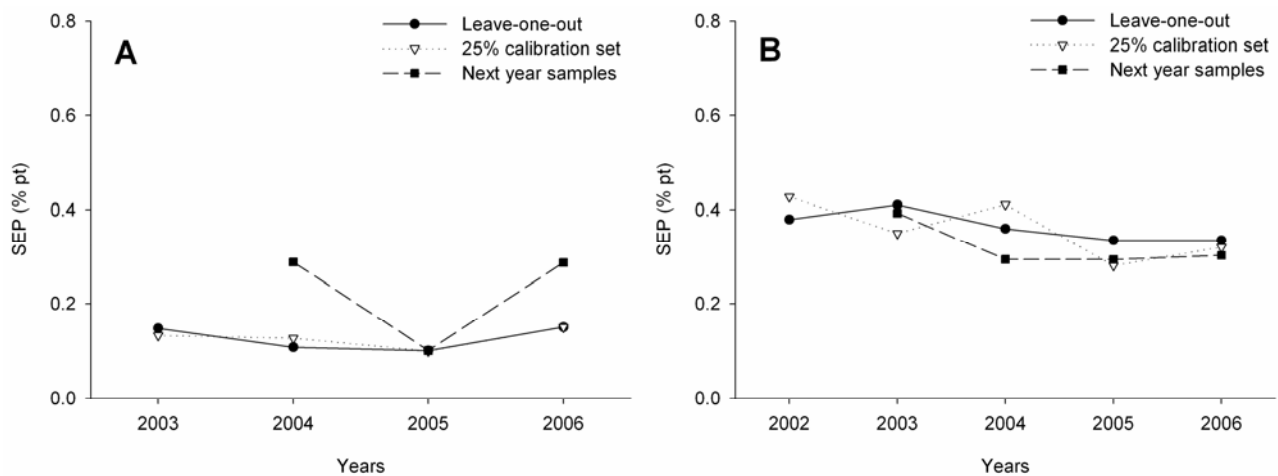
This situation is important because calibration models are often validated with techniques using samples presenting a similar variability to those present in the calibration set, then are applied to material with a variability not included into the calibration set. A good calibration management program would correct the bias problem quickly. However, the precision problem is a more important issue that can only be fixed by adding appropriate samples in the calibration set or by reorganizing it. This situation could be observed with triticale, a crop that was annually changing because active breeding improvements were made in addition to different

climatic conditions from year to year. A similar problem could appear when a calibration has not been updated for some period, then is exposed to new variability in the population (such as from breeding or environment changes).

### Spectral Adaptation

The spectral adaptation was an iterative search to determine what years would be most appropriate to predict a given new year. Second-derivative was used as the pretreatment method because it gave the best results in the spectral addition phase for the prediction of protein content. Results are presented in Fig. 2.

Precision increased as calibration combinations approached the target year for both winter/spring line and winter line only calibrations. New genetic properties not present in the early years were present in the validation year. Extra or older samples added extra noise to the models. These results are contrary to the commonly used spectral addition technique based on historical data bases developed over many years to include a maximum of variability into the model. The combination 2003-2004 gave the highest RPD of 4.47, while 2005 alone gave 4.34. This could mean that the sample properties in 2006 were more like those in 2003 and 2004 than those in 2005 alone. There may have been overall environmental effects not removed by second-derivative pretreatment or greater genetic similarities between 2003 and 2004 with 2006 than between 2005 and 2006.



**Fig. 1.** Progressive validation results from the spectral addition process. **A**, Moisture models with autoscaling as the only pretreatment. **B**, Protein models with second-derivative and autoscaling for spectral preprocessing. SEP, standard error of prediction.

**TABLE II**  
Year-by-Year Moisture Calibration Performance with Autoscaling Only as Preprocessing Method for Spectral Addition Strategy to Develop a Triticale Moisture Prediction Model

Type of Validation and Parameters <sup>a</sup>	2003	2003/2004	2003/2004/2005	2003/2004/2005/2006
Leave-one-out				
$r^2$	0.87	0.98	0.98	0.98
SECV (%)	0.15	0.11	0.10	0.15
Bias (%)	0.00	0.00	0.00	0.00
RPD	2.78	7.00	8.12	8.18
25% calibration set				
$r^2$	0.92	0.97	0.98	0.98
SECV (%)	0.13	0.13	0.10	0.15
Bias (%)	0.00	0.01	0.00	0.01
RPD	3.42	6.17	7.72	8.43
Next-year samples (validation year)	—	2004	2005	2006
$r^2$	—	0.74	0.97	0.82
SEP (%)	—	0.29	0.12	0.29
Bias (%)	—	-0.34	0.08	0.07
RPD	—	1.70	6.73	2.36

<sup>a</sup> SECV, standard error of cross-validation; RPD, relative predictive determinant; SEP, standard error of prediction.

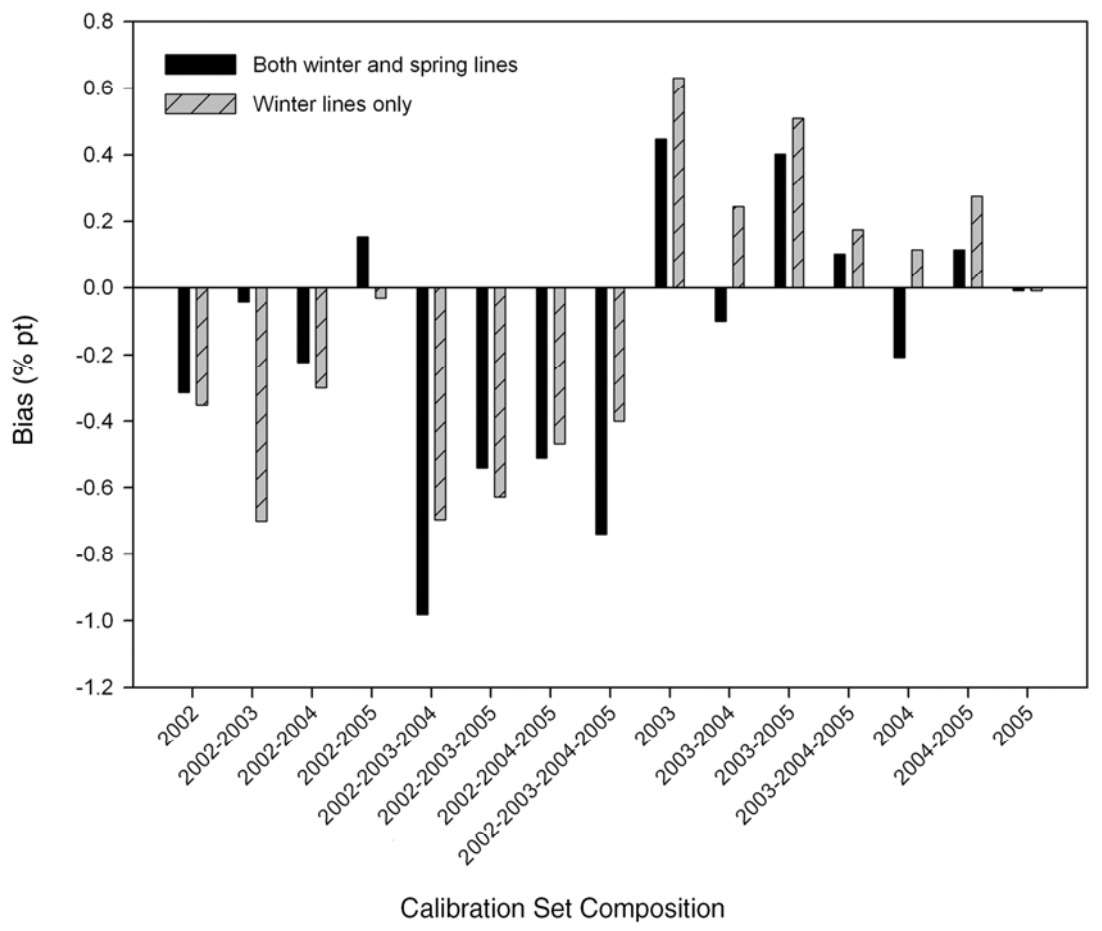
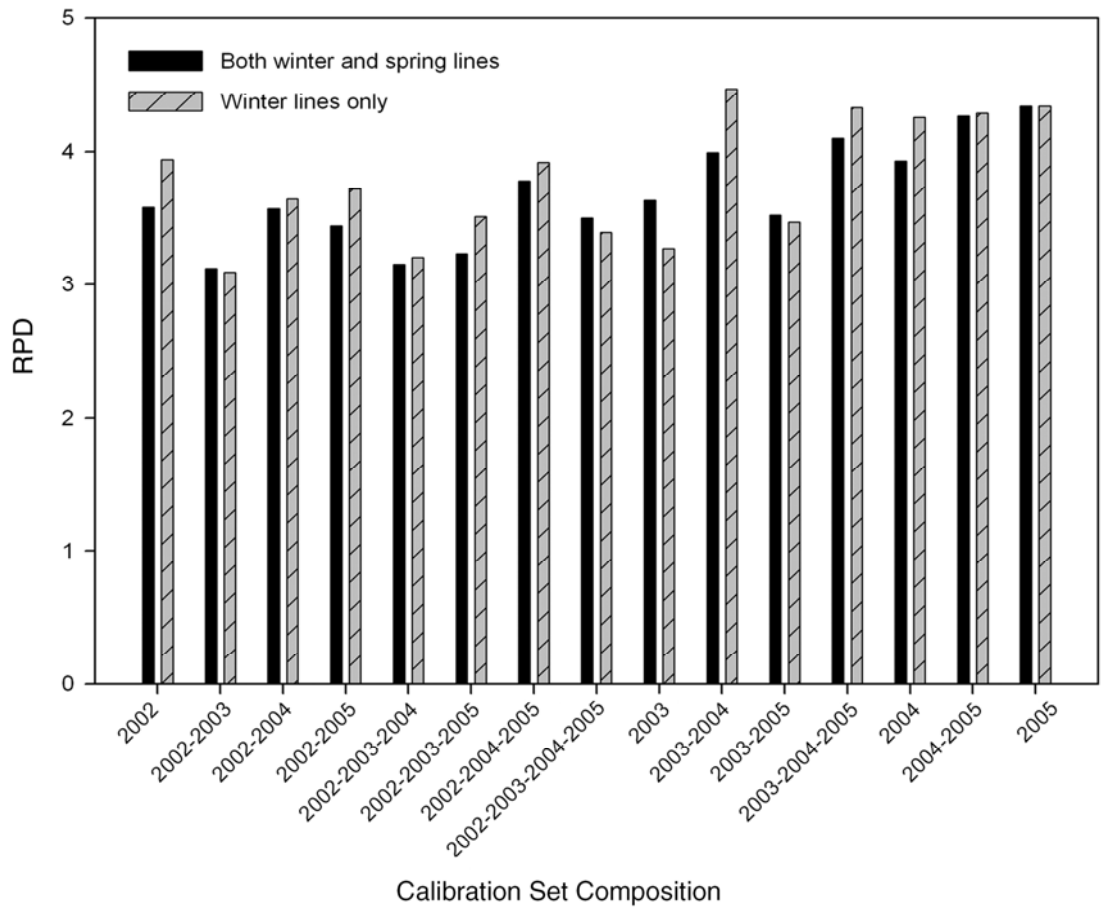


Fig. 2. Precision and accuracy performances for validation on 2006 samples of the spectral adaptation process. RPD, relative predictive determinant.

**TABLE III**  
**Year-by-Year Moisture Calibration Performance with Second-Derivative and Autoscaling as Preprocessing Method**  
**for Spectral Addition Strategy to Develop a Triticale Protein Prediction Model**

Type of Validation and Parameters <sup>a</sup>	2002	2002/2003	2002/2003/2004	2002/2003/2004/2005	2002/2003/2004/2005/2006
Leave-one-out					
$r^2$	0.96	0.95	0.96	0.96	0.96
SECV (%)	0.38	0.41	0.36	0.34	0.34
Bias (%)	0.00	0.00	0.00	0.00	0.00
RPD	4.74	4.43	4.77	4.80	4.73
25% calibration set					
$r^2$	0.95	0.96	0.94	0.97	0.96
SECV (%)	0.43	0.35	0.41	0.28	0.32
Bias (%)	-0.06	-0.01	0.01	-0.03	-0.06
RPD	4.34	5.27	4.23	5.73	4.86
Next-year samples (validation year)	–	2003	2004	2005	2006
$r^2$	–	0.39	0.29	0.39	0.92
SEP (%)	–	0.39	0.29	0.12	0.30
Bias (%)	–	-0.80	0.42	0.08	-0.52
RPD	–	5.39	5.65	6.73	3.50

<sup>a</sup> SECV, standard error of cross-validation; RPD, relative predictive determinant; SEP, standard error of prediction.

Calibration set combinations including 2002 crop year samples presented a negative bias (except for the combination 2002-2005). This shows very clearly the impact of this crop year on the calibration. It is noticeable that 2003 crop year samples presented a variability different than 2006, while 2005 crop year samples were very similar to 2006 samples. This conclusion based on biases is different than the one made with RPD, which clearly shows that the evaluation of a calibration needs both precision and accuracy information.

### Calibration Processes Comparison

The calibration that appears the most appropriate for the spectral adaptation process is the one including 2003, 2004, and 2005 crop year samples because it presented a certain history in samples, gave among the highest RPD, and the lowest bias. Fearon (1996) presented a technique originally proposed by Pitman (1939) to compare regression techniques when predicting the same validation set. This method compares the two sets of prediction errors with a *t*-test to determine the significance or not of the difference. The spectral adaptation strategy was significantly more precise. Biases were also significantly different; the spectral adaptation process was more accurate ( $\alpha = 0.05$ ).

### CONCLUSIONS

This study evaluated the impact of adding or removing samples on a year basis on the prediction model performances and the calibration process. We presented two ways of developing a prediction model: a spectral addition approach that follows the common practices of progressive expansion and a spectral adaptation strategy. From our data, the spectral adaptation process was more precise and accurate. We determined that samples from the year 2002 were different in terms of variability and had an impact on both precision and accuracy of the prediction models. This strategy permits selection of the variability to include removal of extra noise from older genetic variability and adaptation of calibrations to the specific needs of users on a yearly basis. The data from triticale, a crop that genetically changes from year to year due to breeding trials, suggest that the best calibration strategy for balancing precision and risk would be to recalibrate with the latest samples only. We could observe that the closer the samples from the validation year, the better the calibrations were for both precision and accuracy. This strategy would require that enough calibration samples be collected each year so that the dataset will be large enough for the calibration algorithm being used. This may

increase costs of calibration but relieves the need to store very long-term libraries of physical samples. In addition, the choice of samples included to predict accurately the next year's samples is hard to anticipate for the chemometrician. For samples with unstable annual variability, the use of spectral adaptation may be at risk if a close control of the calibration performances is not performed. It would be necessary to compare both calibration strategies with crops where the genetics are less variable and where the variation is mainly due to growth environment to measure the usefulness of spectral adaptation strategy for routine analysis.

The use of built-in sample selection techniques from spectroscopic software could be an alternative to remove samples on a year basis because it reduces the risk of removing too many of the potentially useful samples from the calibration. However, it still requires close calibration management to adapt the models to the variability of the new samples to be predicted if needed.

### ACKNOWLEDGMENTS

We thank P. Stephen Baenziger for supplying triticale breeding lines from the University of Nebraska and Jean-Luc Jannink, Ron Skrdla, and George Patrick for management of the variety tests.

### LITERATURE CITED

- AACC International. 2000. Approved Methods of the American Association of Cereal Chemists, 10th Ed. Methods 44-15A and 46-30. The Association: St. Paul, MN.
- Adams, M. J. 1995. Chemometrics in Analytical Spectroscopy. R. Soc. Chem.: Cambridge.
- Burns, D. A., and Ciurczak, E. W., eds. 2001. Handbook of Near-Infrared Analysis. 2nd Ed. Marcel Dekker: New York.
- Cogdill, R. P., Anderson, C. A., and Drennen, III, J. K. 2005. Process analytical technology case study. III. Calibration monitoring and transfer. AAPS Pharm. Sci. Technol. 6:284-297.
- Delwiche, S. R., and Reeves, III, J. B. 2004. The effect of spectral pretreatments on the PLS modeling of agricultural products. Near-Infrared Spectrosc. J. 12:177-182.
- Fearon, T. 1996. Comparing standard deviations. NIR News 7:5-6.
- Gibson, L., Jannink, J. L., Skrdla, R., and Patrick, G. 2004a. Winter triticale variety performance in Iowa 2002-2004. Available online at [www.pfi.iastate.edu/OFR/Triticale\\_Winter\\_Ylds\\_2002-04.pdf](http://www.pfi.iastate.edu/OFR/Triticale_Winter_Ylds_2002-04.pdf). Iowa State University: Ames.
- Gibson, L., Jannink, J. L., Skrdla, R., and Patrick, G. 2004b. Spring triticale variety performance in Iowa 2002-2004. Available online at: [http://www.pfi.iastate.edu/OFR/Triticale\\_Spring\\_Ylds\\_2002-04.pdf](http://www.pfi.iastate.edu/OFR/Triticale_Spring_Ylds_2002-04.pdf). Iowa State University: Ames.
- GIPSA. 2006. Active commercially available approved equipment. Approved

- equipment list (Active). Available online at <http://archive.gipsa.usda.gov/reference-library/handbooks/equipment/apprlist-a.pdf>. Grain Inspection, Packers & Stockyards Administration: Washington, DC.
- Igné, B., Gibson, L. R., Rippke, G. R., Schwarte, A., and Hurburgh, Jr., C. R. 2006. Triticale moisture and protein prediction by near-infrared spectroscopy. *Cereal Chem.* 84:328-330.
- Kays, S. E., Barton, F. E., and Windham, W. R., 2000. Predicting proteins content by near-infrared reflectance spectroscopy in diverse cereal food products. *J. Near Infrared Spectros.* 8:35-43.
- Martens, H., and Jensen, S. A. 1983. Two-stage partial least squares regression on latent factors: A new NIR calibration method. Pages 607-647 in: *Proc. 7th World Cereal and Bread Congress, Progress in Cereal Chemistry and Technology*. J. Holas and J. Kratochvil, eds. Amsterdam.
- Naes, T., Isaksson, T., Fearn, T., and Davies, T. 2002. A user-friendly guide to multivariate calibration and classification. NIR Publications: Chichester, UK.
- Pitman, E. J. G. 1939. Tests of hypotheses concerning location and scale parameters. *Biometrika* 31:200-215.
- Rippke, G. R., Hardy, C. L., Hurburgh, Jr., C. R., and Brumm, T. J. 1995. Calibration and field standardization of Tecator Infratec analyzers for corn and soybeans. Pages 122-131 in: *Proc. Int. Conf. Near-Infrared Spectroscopy*, 7th. NIR Publication: Charlton, UK.
- Rippke, G. R., Hardy, C. L., and Hurburgh, Jr., C. R. 1996. Calibration and field standardization of Tecator Infratec analyzers for corn and soybeans. Pages 122-131 in: *Near Infrared Spectroscopy: The Future Waves*. NIR Publications: Chichester, UK.
- Siesler, H. W., Ozaki, Y., Kawata, S., and Heise, H. M., eds. 2002. *Near-Infrared Spectroscopy: Principles, Instruments, Applications*. Wiley-VCH Verlag: Weinheim, Germany.
- Williams, P., and Norris, K., eds. 2001. *Near-Infrared Technology in the Agricultural and Food Industries*. 2nd Ed. AACC International: St. Paul, MN.

[Received March 5, 2007. Accepted July 19, 2007.]