

Chemometric Localization Approach to NIR Measurement of Apparent Amylose Content of Ground Wheat

I. J. Wesley,^{1,2} B. G. Osborne,¹ R. S. Anderssen,³ S. R. Delwiche,⁴ and R. A. Graybosch⁵

ABSTRACT

Cereal Chem. 80(4):462–467

The development of new wheat cultivars that target specific end-uses, such as low or zero amylose contents of partially waxy and waxy wheats, has become a modern focus of wheat breeding. But for efficient and cost-effective breeding, inexpensive and high-throughput quality testing procedures, such as near infrared (NIR) spectroscopy, are required. The genetic nature of a set of wheat lines, which included waxy to nonwaxy cultivars, results in a bimodal distribution of amylose contents that presents some special challenges for the formulation of stable NIR calibrations for this property. The obvious and intuitive solution lies in the use of some form of localization procedure and we explored three localization algorithms in comparison with the default partial least squares. Localization with respect to the waxy (zero amylose) cultivars resulted in a modified partial least squares calibration with a standard error of prediction of 0.16%. The results establish unambiguously that there are advantages in performing a suitable localization to achieve a

reliable NIR calibration and prediction. The accuracy of the method can also be enhanced by application of an appropriate resampling strategy. In addition, there are advantages in performing a suitable localization to achieve a reliable NIR calibration-prediction. It resolves the issue of how to utilize the bimodal distribution of apparent amylose values. The best results are obtained when the localization is performed simultaneously with respect to the sample property under investigation and the NIR spectra. The key problem with the measurement of amylose is the laboratory reference method which, in reality, only measures the apparent amylose content of the wheat. As a direct consequence, the measurements of amylose have such a large error that traditional calibration-prediction procedures generate unacceptable results. To resolve this difficulty, a statistically based resampling strategy is proposed as a method of identifying samples where there is a large error in the reference measurement.

Near infrared (NIR) spectroscopy is a popular method for screening quality characteristics such as protein and moisture contents and hardness of material in wheat (*Triticum aestivum* L.) breeding programs. The popularity of the NIR method, arising from the speed of analysis together with small sample size requirements, has led to intensive research to extend its range of applications in wheat breeding. This has resulted in the development of NIR calibrations for flour extraction, yellow pigment, and water absorption (Crosbie et al 2001) and protein functionality (Delwiche et al 1998; Pawlinsky and Williams 1998). In addition, Wesley et al (1999, 2001) have reported how NIR measurements can be inverted to recover the proportional presence of protein components such as glutenin and gliadin by first estimating the structure of NIR spectra of the individual components from an unmixing experiment. However, very little progress has been made in estimating the proportional presence of carbohydrate components in wheat.

Properties such as flour paste viscosity and the quality of end products such as noodles are highly influenced by the carbohydrate composition of the grain, in particular the amylose and amylopectin fractions (Epstein et al 2002). Several authors (Villareal et al 1994; Delwiche et al 1995, 1996; Bao et al 2001;) have reported on the application of NIR to the measurement of amylose in rice. However, the transfer of this technology to wheat, which generally exhibits much less variability in the amylose-to-amylopectin ratio, is potentially more challenging.

The amount of amylose present in wheat is controlled by the expression of three structural genes, located at separate loci, that each encode a different form of granule-bound starch synthase (GBSS). Wild-type wheat possesses all three isoforms, but breeders' lines may possess a null allele at one, two, or all three loci. Generally, the amylose content increases (and the amylopectin-to-amylose ratio

decreases) with the number of active GBSS isoforms (Graybosch et al 1998). Therefore, a set of samples representing all possible isoforms, as described by Delwiche and Graybosch (2002), provided an ideal opportunity to study the utility of NIR to predict the amount of amylose in wheat. However, as a consequence of the genetics, the amylose content of the lines in this sample population separates into a bimodal distribution consisting of the essentially zero amylose content of the triple null "waxy" cultivars and the amylose values of the "partial waxy" and "nonwaxy" lines. The primary aim of this study was to report an NIR calibration procedure that results in the determination of amylose in wheat with sufficient accuracy for use in breeding programs.

From a more generic NIR perspective, these data represent an opportunity for the careful testing of recent developments of NIR methodology. 1) To explore the strengths and limitations of Modified Partial Least Squares (MPLS) in analysis of Delwiche and Graybosch (2002) data because these data are representative of situations where the values of the property predicted are bimodal (or have a wide range of values) when the corresponding NIR spectra are closely clustered. 2) To assess the performance of LOCAL (Shenk et al 1997, 1999; Berzaghi et al 2000) and CARNAC (Davies et al 1988) on this data. 3) To assess the use of resampling strategies to reduce the effect of poor quality reference data.

MATERIALS AND METHODS

The samples and data collection protocols below are extensively described in Delwiche and Graybosch (2002).

Samples

Partial waxy experimental wheats were identified from the F3 generation of populations derived from the following crosses: MT8713/NE87612, NE90476/Ike, NE90616/Ike, and SD88137/Ike. Genotypes were identified by electrophoretic analysis of the waxy (granule-bound starch synthase) proteins. The F4 generation was grown at Berthoud, CO. The F5 generation, along with the control cultivars Redland (wild-type), TAM-202 (partial waxy, single null), and Ike (partial waxy, double null) were grown in 4-row plots at Lincoln and Sidney, NE. Waxy wheat lines derived from the cross Kanto107 × BaiHuo were grown in 1998 in Brawley, CA. Waxy lines were derived from single F2 waxy grain identified by staining samples with I₂KI.

¹ BRI Australia Ltd, PO Box 7, North Ryde, NSW 1670, Australia.

² Corresponding author. E-mail: i.wesley@bri.com.au. Phone: +612 9888 9600. Fax: +612 9888 5821.

³ CSIRO Mathematical & Information Sciences, GPO Box 664, Canberra, ACT 2601, Australia.

⁴ USDA/ARS, Beltsville Agricultural Research Center, Beltsville, MD 20705-2350.

⁵ USDA/ARS at the University of Nebraska, Lincoln, NE.

Apparent Amylose Determination

Apparent amylose contents of the wheat samples were measured by the iodine binding method of Knutson and Grove (1994), as modified by Delwiche and Graybosch (2002). The reference method has a standard deviation of repeatability of 0.33% (low amylose) and 0.95% (high amylose). A number of samples have apparent amylose contents of 3–9%. These samples should be waxy and therefore have little or no amylose present. The reasons for the apparent presence of amylose are discussed elsewhere (Delwiche and Graybosch 2002) and, because of the uncertainty over their provenance, they were excluded from the analysis presented here.

NIR Spectroscopy

Wheat samples were ground on a laboratory-scale cyclone grinder fitted with a 0.5-mm screen (Udy Corp., Ft. Collins, CO) before NIR measurement. Spectra were measured in reflectance mode using a spectrophotometer equipped with a spinning sample module (model 6500, Foss NIRSystems Inc., Silver Spring, MD). Spectral data were recorded from 1100 nm to 2498 nm at 2-nm intervals and saved as the average of 32 scans for each sample. The spectrum of each sample was measured in duplicate (with repacking) and the duplicate spectra averaged before further data processing.

Data Processing

Data processing was performed using WinISI v. 1.04 (Infrasoft International LLC, Silver Spring, MD), Grams32 v. 5.05 (Thermo Galactic, Salem, NH), and Excel (Microsoft Corp., Seattle, WA). The spectral data was split into a calibration set of 141 (17 low amylose, 124 high amylose) samples and a validation set of 46 (5 low amylose, 41 high amylose) samples. The assignment of samples was grouped to ensure that both high and low amylose samples were included in both sets. Otherwise, assignment to one or other of the sets was at random. Master Curve Deconvolution (MCD) was used to derive reliable spectra of amylose and amylopectin. Two different localization methods (LOCAL and CARNAC) were investigated and compared with MPLS calibration, a proprietary implementation of partial least squares where the calibration data is renormalized during the calculation of the PLS factors (Shenk and Westerhaus 1991).

Master Curve Deconvolution (MCD)

The spectra described above and the laboratory data were used in the unmixing process described in Wesley et al (1999, 2001) to create master curves for amylose and amylopectin.

Hierarchical MPLS

A global MPLS calibration was used to determine whether the unknown belongs to the high or low amylose group, and then the appropriate localized calibration was utilized to make the final

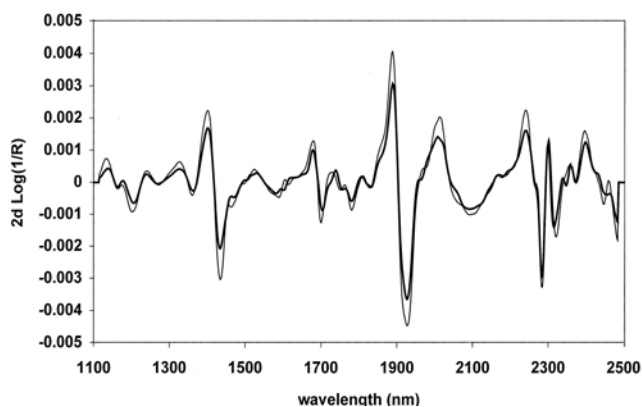


Fig. 1. Master curves for amylose (thin line) and amylopectin (thick line) unmixed from the spectra of ground wheat.

prediction. The calibrations were performed using WinISI software on the second derivative spectra determined with a 16-nm gap and a 16-nm smooth. Full cross-validation was performed and no outliers were eliminated.

LOCAL

LOCAL (Shenk et al 1997, 1999; Berzaghi et al 2000) is a proprietary algorithm in WinISI software (Infrasoft International LLC, Silver Spring, MD). The method involves calculating the correlation coefficient between the target spectrum and each spectrum in the library and selecting library spectra that have a high correlation with the target spectrum. The selected spectra are then used to develop a calibration by MPLS and the target spectrum predicted. In the current study, the processing options (for example, spectral data pretreatment, number of PLS terms) were determined by trial and error to arrive at the best algorithm for a specific property of a particular sample type.

Weighted CARNAC

The original implementation of CARNAC (Davies et al 1988) used Fourier analysis to compress the spectral library to a library of Fourier coefficients. The Fourier coefficients were then weighted according to the property of interest by multiplying by the regression coefficients obtained from a simple multiple linear regression performed with the Fourier coefficients to predict the property of interest using the library. An alternative approach to achieving the same end is to use Principal Component Scores instead of Fourier coefficients. The version of CARNAC proposed below is a weighted version of the original algorithm as originally described by Davies et al (1988) and understood from Davies (2002) and was programmed into an Excel spreadsheet. For each target sample, the sum of squares of differences, between the target sample's weighted principal component scores and each of the library samples' weighted principal component scores, were calculated. The library samples, where this value was below a predefined cut-off limit, were retained for further processing. A weight function was calculated for each of the samples retained in the library as the reciprocal of the product of the regression error for the retained sample and the sum of squares of differences between the weighted principal component scores for target and retained sample. This weighting provides a mechanism for only selecting library samples where the reference and spectral data are reliable. The assumption is that a low regression error for the library sample implies that the spectral and reference data are closely linked. A high regression error would suggest a larger uncertainty in the reliability of the library data. Finally, the corresponding weighted average of the laboratory reference values for the retained library samples was calculated to become the final estimate of the property of the target sample.

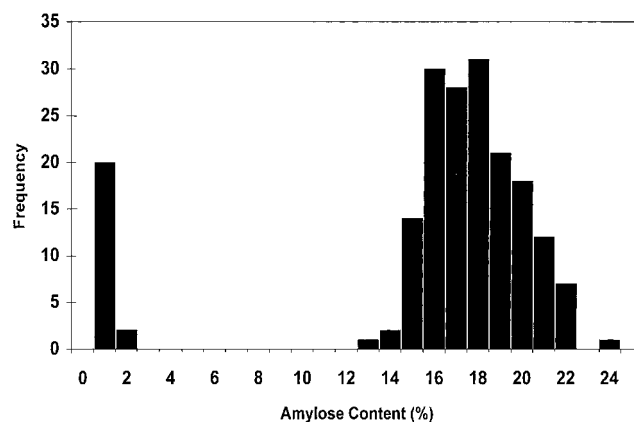


Fig. 2. Distribution of amylose contents.

Statistically Based Resampling (Bootstrapping)

When confronted with a situation where the parameter that determines the optimal statistical methodology that should be applied to given data is unknown, the statistician performs resampling on the given data (as if they were the original population from which the given data have been sampled) to determine an estimate of this parameter which is then utilized as if it were the true value. This idea has its roots in the history of statistics (Efron and Tibshirani 1993) and is known today as resampling or bootstrapping. Resampling is not to be confused with cross-validation where one is aiming to estimate the prediction error of some statistical methodology. The way that the resampling is performed depends heavily on the nature of the problem under investigation. For the present problem, resampling was investigated for the 165 high-amylose content samples only (i.e., those samples possessing at least one active GBSS gene). Twenty percent of the samples were selected at random (using the WinISI random selection procedure) and removed from the set. Fifty percent of the remaining samples were selected at random using a spreadsheet (Fearn 2002) and used with WinISI software to create a calibration equation for the amylose and protein contents using MPLS with full cross-validation and no outlier elimination. The 50% selected were returned to the remaining samples and another 50% selected at random and another set of calibrations determined. This was repeated until 31 calibrations had been developed. The 31 calibrations were used to predict the amylose and protein contents for each of the samples in the 20% pool and the results for these 31 predictions were averaged. The whole cycle was repeated for another 20% of the 165 samples where the selection was restricted to samples not previously selected until all samples had been processed in this manner.

RESULTS AND DISCUSSION

A major limitation in the NIR measurement of amylose in wheat kernels results from the iodine-binding reference method. In fact, the method refers to apparent amylose content because of the nonspecificity of the iodine binding reaction because it is unable to distinguish between the long linear chains of α -(1 \rightarrow 4) glucans, that occur in the α -(1 \rightarrow 6) branched amylopectin and the linear amylose molecules. The presence of lipids in the starch and operator dependence also make significant contribution to the poor performance of the technique (Regina et al 1997).

Master Curve Deconvolution (MCD)

In the development of an NIR method for measuring some property, it must be remembered that the best results will be obtained from properties where there is a clear and unambiguous linkage between changes in the property and changes in the spectrum. For example, in wheat, the major components (protein, starch, and water) each have spectra that are very different, therefore it is quite easy to determine the proportional presence of these components. It has been previously shown (Wesley et al 1999, 2001) that measuring the composition of the protein components (glutenin and gliadin) independent of the total protein content requires a more sophisticated algorithm that decouples the two properties (composition and content). The basis of this methodology is to first determine master curves for the components of interest by

NIR spectrum unmixing (Martens and Naes 1991; Wesley et al 1999) and then deconvoluting the target spectrum with the master curves. The unmixing of NIR spectra of carefully controlled mixtures of amylose and amylopectin, or from real samples where the amylose and amylopectin content are known, allows one to obtain representative spectra for amylose and amylopectin with an accuracy not previously achieved. For example, the curves given in Williams and Norris (2001) are unreliable because the particle size effects have not been removed in a manner that is achieved automatically with unmixing. In particular, care must be taken to ensure thorough mixing and to ensure that the moisture content is carefully controlled by appropriate equilibration and storage of the mixtures. As starch is hygroscopic, this is no easy matter. The master curves determined in the current study for amylose and amylopectin are shown in Fig. 1. Globally, as one would expect because of the chemical similarity of amylose and amylopectin, they are very similar.

The principal requirement of MCD is that the individual component master curves are sufficiently linearly independent to allow the successful deconvolution of the target spectrum. As can be seen from Fig. 1, this is only true for the amylose and amylopectin spectra on disjointed subsets of wavelengths. Though their utilization represents an alternative strategy for implementing MCD, it is not pursued here. Otherwise, appropriate independent data is required (such as Raman spectra) before MCD could achieve such a separation with the full spectrum or the development of a segmented MCD that only works with a disjoint subset of spectral intervals on which there are clear differences between the amylose and amylopectin spectra. In many ways, this represents a generalization of the methodology developed for the calibration-prediction performed by the early filter instruments where each single filter wavelength is replaced by a local interval of wavelengths.

Hierarchical MPLS

Partial least squares (PLS) techniques have proven very popular when dealing with complex NIR calibration problems. MPLS appears to be one of the better PLS-calibration methodologies. PLS techniques work best on a unimodal property distribution (Fearn 1992). As noted above, the genetic nature of the sample set examined here automatically gives rise to a bimodal population with respect to amylose content (Fig. 2).

However, a hierarchical approach to MPLS calibration-prediction can be used to take into account the bimodal distribution of the data. The basic steps are 1) a global MPLS calibration is used to determine whether the unknown belongs to the high or low amylose group; and 2) the appropriate localized calibration is utilized to make the final prediction.

A better prediction of the low amylose values was obtained if the calibration was based on the set of spectra with low amylose values. Less accurate estimates were obtained when using the full range of values.

This represents direct proof that even MPLS requires appropriate prelocalization to take full advantage of its inherent stabilization behavior (de Jong and Kiers 1992). Because the set of high amylose values dominates the calibration, the prediction of the high amylose values by MPLS from the high amylose lines of the sample set does not show such dramatic improvement (Table I).

TABLE I
Comparison of Standard Errors of Prediction for Different Calibration-Prediction Strategies
Using Partitioned Calibration and Validation Set

Calibration-Prediction Range	MPLS	LOCAL	CARNAC	Resampling
Low calibration - low prediction	0.21	0.28	0.9	-
Full calibration - low prediction	0.69	1.19	0.90	-
High calibration - high prediction	1.41	1.42	2.0	1.1
Full calibration - high prediction	1.50	1.41	2.0	-
Full calibration - full prediction	1.49	1.38	1.90	-

LOCAL

The situation described above seems to be ideally suited to some version of localization. A number of variants have been proposed (Locally Weighted Regression [Naes et al 1990; Wang et al 1994], LOCAL [Shenk et al 1997; Shenk et al 1999; Berzaghi et al 2000] and CARNAC [Davies et al 1988]) and some are implemented in commercially available software. While the operational details of the three methods differ, all utilize the same basic algorithm. The conceptualization is quite intuitive and natural. The NIR spectrum for a new sample is measured and then, from the library of spectra, those spectra that are similar to the spectrum of the new sample are chosen. The estimate of the required property for the sample is then determined by applying a calibration-prediction based on the subset of chosen spectra.

As Fig. 3 clearly shows, LOCAL picks, for a test sample which is known to have a high amylose content, a subset of spectra that have wide protein values and amylose values covering a wide range of the amylose and protein values in the library. Interestingly, in this situation, because of the strong coupling between the NIR wheat spectra and protein content, when compared with that for amylose, the localization of the protein values is considerably stronger than is seen for the amylose values. Figure 3 is indicative of how LOCAL performs more generally. LOCAL gives a coarse localization with respect to the laboratory values of all the properties under consideration because localization is only performed with respect to the NIR spectra. Consequently, it finds a wide range of property values that will often contain the value of the target sample.

As is clear from the SEP values listed in Table I, LOCAL performs as well as MPLS in all the categories tabulated there. However, this is not surprising because LOCAL, after performing its library localization, uses MPLS to perform the calibration-and-prediction.

To guarantee the best results in the calibration-prediction of properties of materials from NIR spectra, it is not only necessary to localize the NIR spectra in the library with respect to the NIR

spectrum of the sample being tested but also to localize the choice of spectra with respect to some estimate of the value of the property of the sample being tested. Though an iterative MPLS strategy could be used for achieving this goal, other strategies are possible including qualitative expedients. The Delwiche and Graybosch (2002) investigation is an example of such a qualitative localization in that the data come from a limited number of related crosses. This then explains why MPLS gives a good prediction (as illustrated in Fig. 4) without first performing a localization with respect to the approximate amylose content of the sample being tested.

The critical step in the process is the definition of similar as it determines how selection of samples for the calibration-prediction process is to be performed.

Weighted CARNAC

In the SEP measures list in Table I, the results for the weighted CARNAC (Fig. 5) are marginally worse than those obtained using LOCAL and MPLS. This is due to the fact that the localization performed by CARNAC does not take the property values directly into account and the weighting of the library in CARNAC is dependent on how well the weighting coefficients relate to the property of interest. Another possible reason CARNAC may not be functioning in this application is because there is an insufficient number of low amylose samples in the library set. This would mean that a good selection of similar samples cannot be made and the relationship in the samples that are chosen is poor.

Figure 6 shows the samples selected by CARNAC in the protein-amylose space for the same sample as used in Fig. 3. When compared with Fig. 3, it is clear that CARNAC performs no better or no worse a property localization than LOCAL. This is confirmed in Fig. 7 by examining the full test set. The rectangular boxes show the range of amylose values (for the lines in the selected test set population) for both CARNAC and LOCAL, where the extreme left-hand sample in each graph is the test sample used

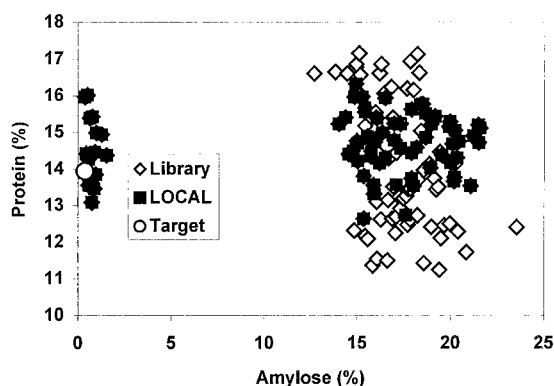
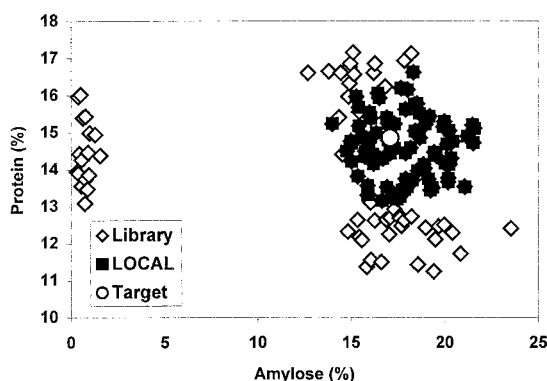


Fig. 3. Samples selected by LOCAL for two target samples.

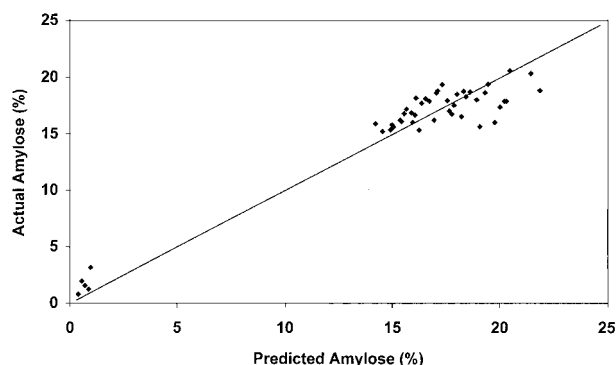


Fig. 4. Prediction of amylose using modified partial least squares regression across the whole data set.

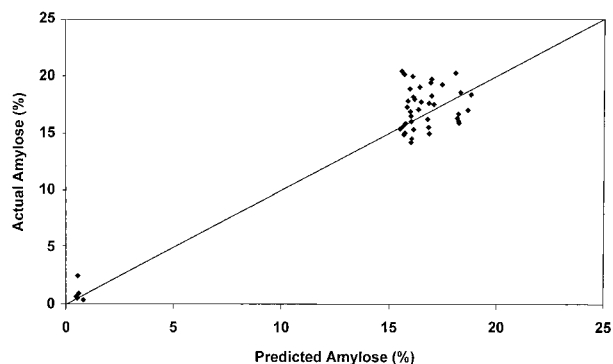


Fig. 5. Actual vs. predicted amylose using a weighted CARNAC on a weighted principal component score library.

in Figs. 3 and 6. It is clear that, while CARNAC always performs an appropriate localization with respect to amylose content, LOCAL often selects samples with high and low amylose values. In fact, for the waxy samples, LOCAL always selects samples with both high and low amylose values. For each of the test samples, the amylose range selected by LOCAL always contained the amylose value of the target sample. On the other hand, the range selected by CARNAC did not always include the amylose value of the target sample. (CARNAC cannot make a prediction if the number of samples selected from the library is <5.) Samples where the target amylose content is outside the range of the samples selected by CARNAC raises the issue of the validity of the reference values (both in the library and of the target sample) and of the linkage between the reference value and the NIR spectral data.

Statistically Based Resampling (Bootstrapping)

The failure of CARNAC to perform a satisfactory localization for the subsequent calibration-prediction was not so much the fault of CARNAC as the highly inaccurate amylose laboratory reference values. In fact, they are so uncertain that it is highly likely that even the application of a normally successful localization strategy would fail to yield sufficiently consistent results.

This led naturally to the need to examine how the relatively high accuracy of the NIR spectra could be exploited to obtain more accurate estimates of the amylose laboratory reference values. The fact that the accuracy of the NIR spectral values represents an opportunity to improve uncertain laboratory reference values has been explained elsewhere (Fearn 1996). The possibility of formulating and utilizing an appropriate resampling (bootstrapping) strategy

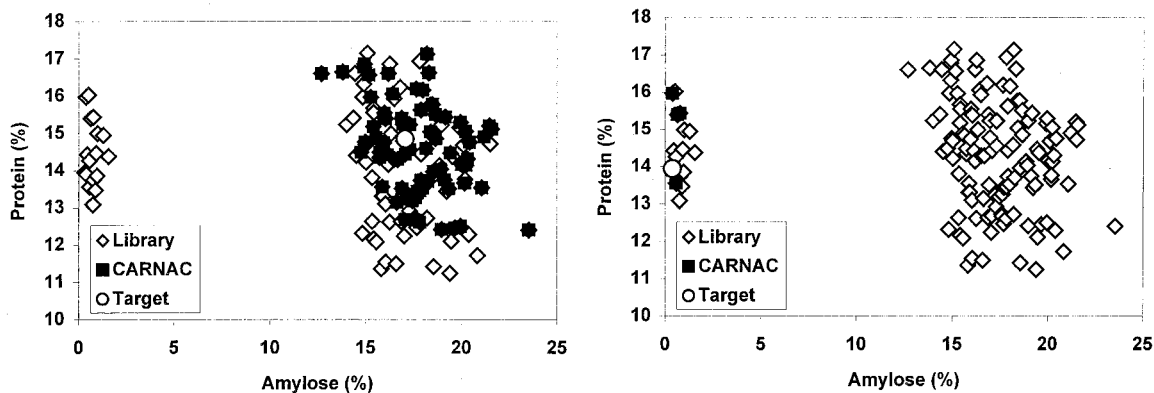


Fig. 6. Samples selected by CARNAC for two target samples.

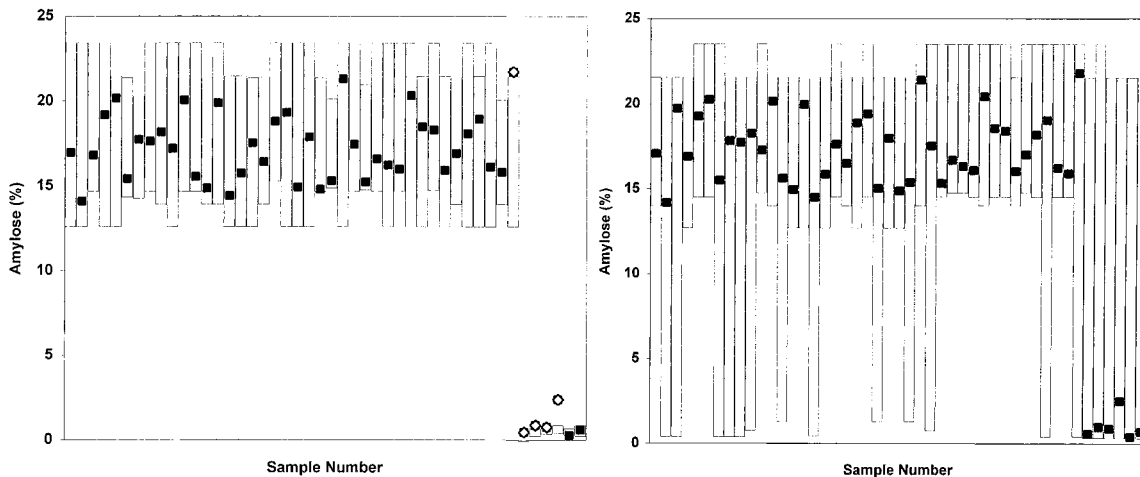


Fig. 7. Range of amylose contents selected by CARNAC (left) and LOCAL (right). Squares represent target values and bars represent range of amylose content of samples selected by two methodologies. Diamonds represent target samples outside the range selected from the calibration set.

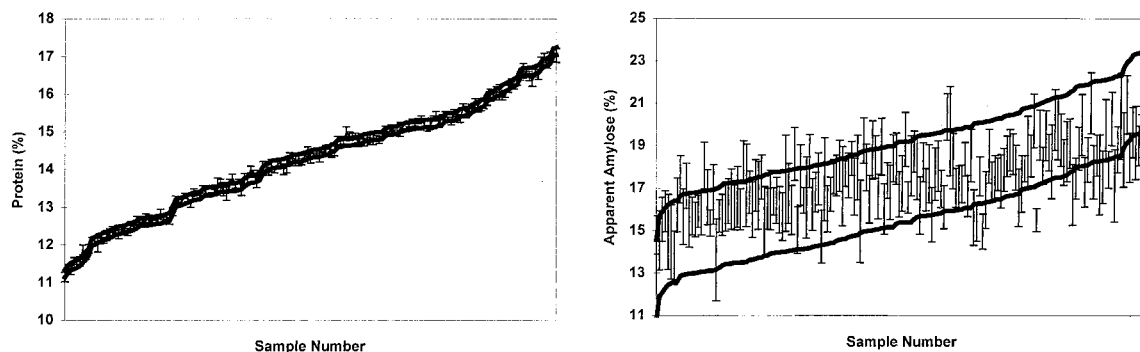


Fig. 8. Resampling results for protein (left) and apparent amylose (right) (samples sorted in increasing laboratory value). Heavy lines represent uncertainty in the laboratory reference value for each sample (2 standard deviations of replicate) and error bars represent uncertainty about NIR prediction (2 standard deviations of replicate of the predicted values).

does not appear to have been explored as outlined above and performed in this study.

The results of the resampling analysis for protein and amylose are shown in Fig. 8. The heavy black lines represent uncertainty in the laboratory reference value for each sample (2 standard deviations of replicate) and the black error bars represent the uncertainty about the NIR prediction (2 standard deviations of replicate of the predicted values).

There is a clear difference between the results for protein and amylose, which reflects the precision of the reference method, and the ability of the NIR calibration to model the relationship between the NIR spectrum and the laboratory value. For protein, the precision of both the reference and NIR methods is high, and there is almost complete overlap of the NIR and laboratory errors, which are small in size. This reflects the well-known ability of NIR to predict protein content of wheat with a high degree of certainty. In contrast, the precision of the reference method for amylose is low (reflected by the broadness of the uncertainty band in Fig. 8) and the repeatability of the NIR prediction is poor (reflected by the large uncertainty bars in Fig. 8). In many cases, there is only partial overlap of the uncertainty ranges; in some cases, the uncertainty ranges do not overlap at all. Samples where there is no overlap of the uncertainty ranges will be poorly modeled by the NIR calibration and can be removed from the modeling process. When the calibration-prediction is performed on the resulting library, it results in a lower standard error for the high-amylose MPLS calibration of 1.1% (compared with 1.4% using all of the samples) (Table I). There were insufficient samples in the low-amylose library to perform the resampling effectively. Considering the importance of localization, it would have been inappropriate to apply the resampling to the whole library and then use it to perform either low- or high-only predictions.

CONCLUSIONS

The results presented here show generically that a suitable localization is required to achieve a successful application of the calibration-prediction to the NIR spectra. In some cases, the localization is obvious and intuitive, such as the use of selected breeding lines. In other cases, the best results are obtained when the localization is performed with respect to the sample property under investigation as well as the NIR spectra. In cases where poor precision of the reference data contributes to the poor performance of a NIR calibration, resampling strategies can be used to identify samples that are possible outliers. Removing these from the modeling procedure improves the performance of the NIR calibration-prediction. These results show how the apparent amylose content of ground wheat can be measured with sufficient accuracy for use in selecting lines for high-, medium-, or low-amylose content in wheat breeding programs.

ACKNOWLEDGMENTS

We thank The Grains Research & Development Corporation for investing in the Australian component of this research.

LITERATURE CITED

Bao, J. S., Cai, Y. Z., and Corke, H. 2001. Prediction of rice starch quality parameters by near-infrared reflectance spectroscopy. *J. Food Sci.* 66:936-939.

Berzaghi, P., Shenk, J. S., and Westerhaus, M. O. 2000. Local prediction with near infrared multi-product databases. *J. Near Infrared Spectrosc.* 8:1-9.

Crosbie, G. B., Wang, Y., Osborne, B. G., Allen, H. M., Palmer, G. A., Black, C., and Mares, D. J. 2001. Collaborative development of NIR quality tests for application in wheat breeding. Pages 562-564 in: Proc. 11th ICC Cereal and Bread Congress. M. Wootton, I. L. Batey, and C. W. Wrigley, eds. RACI: Melbourne, Australia.

Davies, A. M. C. 2002 The idea behind comparison analysis using restructured near infrared and constituent data (CARNAC). Pages 29-

32 in: Near Infrared Spectroscopy: Proc. 10th Int. Conf. A. M. C. Davies and R. K. Cho, eds. NIR Publications: Chichester, UK.

Davies, A. M. C., Britcher, H. V., Franklin, J. G., Ring, S. M., Grant, A., and McClure, W. F. 1988. The application of Fourier-transformed near-infrared spectra to quantitative analysis by comparison of similarity indices (CARNAC). *Mikrochimica Acta* 1:61-64.

de Jong, S., and Kiers, H. A. L. 1992. Principal covariates regression. I. Theory. *Chemometrics and Intelligent Laboratory Systems* 14:155-164.

Delwiche, S. R., and Graybosch, R. A. 2002. Identification of waxy wheat by near-infrared reflectance spectroscopy. *J. Cereal Sci.* 35:29-38.

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.

Delwiche, S. R., Graybosch, R. A., and Peterson, C. J. 1998. Predicting protein composition, biochemical properties, and dough-handling properties of hard red winter wheat flour by near-infrared reflectance. *Cereal Chem.* 75:412-416.

Efron, B., and Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. Chapman and Hall: New York.

Epstein, J., Morris, C. F., and Huber, K. C. 2002. Instrumental texture of white salted noodles prepared from recombinant inbred lines of wheat differing in the three granule bound starch synthase (waxy) genes. *J. Cereal Sci.* 35:51-63.

Fearn, T. 1992. Flat or natural? A note on the choice of calibration samples. Pages 61-66 in: Proc. Near Infra-Red Spectroscopy: Bridging the Gap between Data Analysis and NIR Applications, K. L. Hildrum, T. Isaksson, T. Naes, and A. Tandberg, eds. Ellis Horwood: Chichester, UK.

Fearn, T. 1996. How accurate can you get? *NIR News* 7:3.

Fearn, T. 2002. Randomisation. *NIR News* 13:4-5.

Graybosch, R. A., Peterson, C. J., Hansen, L. E., Rahman, S., Hill, A., and Skerritt, J. H. 1998. Identification and characterization of U.S. wheats carrying null alleles at the *wx* loci. *Cereal Chem.* 75:162-165.

Knutson, C. A., and Grove, M. J. 1994. Rapid method for estimation of amylose in maize starches. *Cereal Chem.* 71:469-471.

Martens, H., and Naes, T. 1991. *Multivariate Calibration*. John Wiley and Sons: Chichester, UK.

Naes, T., Isaksson, T., and Kowalski, B. R. 1990. Locally weighted regression and scatter-correction of near-infrared reflectance data. *Anal. Chem.* 62:664-673.

Pawlinsky, T., and Williams, P. 1998. Prediction of wheat bread-baking functionality in whole kernels, using near infrared reflectance spectroscopy. *J. Near Infrared Spectrosc.* 6:121-127.

Regina, A., Morell, M. K., Sharp, P. J., Batey, I. L., and Curtin, B. 1997. Comparison of small scale methods for analysing amylose content in wheat. Pages 162-166 in: Proc. 47th Australian Cereal Chem. Conf. A. W. Tarr, A. S. Ross, and C. W. Wrigley, eds. RACI: Melbourne.

Shenk, J. S., and Westerhaus, M. O. 1991. Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop Sci.* 31:469-474.

Shenk, J. S., Westerhaus, M. O., and Berzaghi, P. 1997. Investigation of a local calibration procedure for near infrared instruments. *J. Near Infrared Spectrosc.* 5:223-232.

Shenk, J. S., Berzaghi, P., and Westerhaus, M. O. 1999. LOCAL: A unifying theory and concept for near infrared analysis. Pages 211-214 in: Proc. 9th Int. Conf. on Near Infrared Spectroscopy. NIR Publications: Chichester, UK.

Villareal, C. P., de la Cruz, N. M., and Juliano, B. O. 1994. Rice amylose analysis by near-infrared transmittance spectroscopy. *Cereal Chem.* 71:292-296.

Wang, Z., Isaksson, T., and Kowalski, B. R. 1994. New approach for distance measurement in locally weighted regression. *Anal. Chem.* 66:249-260.

Wesley, I. J., Uthayakumaran, S., Anderssen, R. S., Cornish, G., Bekes, F., Osborne, B. G., and Skerritt, J. H. 1999. A curve fitting approach to the near infrared reflectance measurement of wheat flour proteins which influence dough quality. *J. Near Infrared Spectrosc.* 7:229-240.

Wesley, I. J., Larroque, O., Osborne, B. G., Azudin, N., Allen, H., and Skerritt, J. H. 2001. Measurement of gliadin and glutenin content of flour by NIR spectroscopy. *J. Cereal Sci.* 34:125-134.

Williams, P., and Norris, K. 2001. *Near-Infrared Technology in the Agricultural and Food Industries*. Am. Assoc. Cereal Chem.: St. Paul, MN.

[Received July 3, 2002. Accepted December 19, 2002.]