Curve fitting applied to near infrared deconvolution of wheat functional proteins in flour

Ian J. Wesley
CSIRO Plant Industry, PO Box 7, North Ryde, NSW 1670, Australia.

Brian G. Osborne
BRI Australia Ltd, PO Box 7, North Ryde, NSW 1670, Australia.

Robert S. Anderssen
CSIRO Mathematical and Information Sciences, GPO Box 664, Canberra, ACT 2601, Australia.

John H. Skerritt
CSIRO Plant Industry, GPO Box 1600, Canberra, ACT 2601, Australia.

Introduction

A traditional method for assessing the quality of a flour for baking purposes is to perform a number of rheological measurements on the dough (Farinograph Water Absorption, Extensibility and Mixograph) and then to compare the results. The data thereby collected is used to assess the likely behaviour of the flour in a large scale industrial processing, and as selection criteria in plant breeding programmes. Crucially, in plant breeding, because quite large amounts of sample are required for some of the measurements they cannot be performed on the earliest generation material, which comes from small plots.

The literature on near infrared (NIR) calibrations contains details about a variety of dough quality parameters. However, the performance of the calibrations is highly dependent on the error of the laboratory reference method which, in many cases, can be as much as ±10%. When combined with the large diversity of growing conditions and potential wheat varieties in Australian breeding programmes, the possibility of developing stable Australia-wide calibrations is greatly diminished. A further complication is that many of the dough quality parameters cannot be easily related to spectroscopically active components of the flours.

The protein fraction of a flour can be divided into three broad categories—gliadins, glutenins and water soluble proteins. Of these, the ratio of glutenin to gliadin in the protein fraction has been shown to have a significant impact on the performance of the flour as it is processed. The primary method for determining the gliadin/glutenin ratio is SE-HPLC analysis of the flour protein extract. This technique has an error of approximately 0.2% and the usual range of values for both the gliadin and glutenin content in the flour is 3–9%. Although gliadin and glutenin are spectrally dissimilar, the differences are small and are masked by the spectrum of the starch in a flour spectrum. Consequently, calibration by

Near Infrared Spectroscopy: Proceedings of the 9th International Conference

traditional methodologies (partial least squares , simple regression) does not yield calibrations which are stable across sites, seasons and varieties.

Faced with these problems, the earlier curve-fitting procedure of Hrushcka and Norris[1] has been investigated as a possible alternative. Theoretical spectral curves which represented pure glutenin, gliadin and starch were derived using least squares analysis. These curves were fitted to a small section of the sample spectral curve using least squares to gain an estimation of the gliadin and glutenin content. The results obtained are independent of season and site and are transferable between instruments of a similar design with minimal effort.

Materials and methods

Flour samples

Wheat samples were obtained from a doubled haploid population, derived from a cross between the wheat varieties Halberd and Cranbrook. This population was developed as the two parents represent extremes in dough extensibility. Halberd is a South Australian wheat variety with low extensibility and Cranbrook, a Western Australian variety possessing high extensibility. The advantage of a doubled haploid population is that all the lines are pure and homozygous, a situation which normally takes 7–8 generations to approach in a breeding programme. Halberd and Cranbrook have different high and low molecular weight glutenin alleles at each of the six loci so there are 64 possible different glutenin subunit combinations. Consequently, a large range in glutenin and gliadin content, and therefore in dough processing properties, can be expected. Of the 170 lines produced, 60 of these different combinations are represented in this material. The 170 lines, together with the parents, were grown in a replicated trial at Roseworthy in South Australia in 1996. Straight run flours were prepared by using a Bühler Laboratory mill MLU 202 (Bühler Bros., Uzwil, Switzerland). Further test sets were obtained from Agrifood Technology (Werribee, Victoria, Australia) and NSW Department of Agrticulture (Wagga Wagga, NSW, Australia).

HPLC procedure

Different wheat storage protein extracts were subjected to SE-HPLC, using a Beckman System Gold HPLC (Beckman Instruments, Inc., Fullerton, CA, USA), configured with two 126 Pumps, a 166 Detector and a 507E Autosampler. Samples included total protein, soluble polymeric protein and insoluble polymeric protein, extracted using the Batey *et al.*[2] and Gupta *et al.*[3] procedures.

Preparation of glutenin and gliadin samples

Three varieties of wheat flour were used—Sunstate, Eradu and Cadoux. Gluten, starch and water soluble material were separated from the flours by hand washing, followed by freeze drying. Glutenin- and gliadin-rich fractions were isolated from each gluten by precipitation, using dilute hydrochloric acid (0.2 M) at pH 5.3 for gliadin, and pH 3.9 for glutenin, homogenised, centrifuged and the pH of the fractions adjusted to 5.8 with NaOH (0.2 M) solution.[4] Finally, the samples were freeze-dried and ground.

NIR spectra

NIR spectra recorded using an NIRSystems 6500 scanning monochromator spectrometer, fitted with a sample transport accessory (Foss NIRSystems Inc., Silver Spring, MD, USA). The wavelength range was 400–2500 nm, with a 2 nm gap between datapoints. Data was exported to GRAMS5 software (Galactic Industries Corporation, Salem, NH, USA) for subsequent processing.

Model development

If measurements of y, S_{C1} and S_{C2} are available, at a series of points x_1, …, x_n, the model can be solved by least squares regression to recover estimates of a and b. The least squares estimate of a and b is given by (the normal equations)

$$y(x)=aS_{C1}(x)+bS_{C2}(x)$$

$$\begin{bmatrix} a \\ b \end{bmatrix}=(A^T A)^{-1} A^T y$$

where

$$A=\begin{bmatrix} S_{C1}(x_1) & S_{C2}(x_1) \\ \vdots & \vdots \\ S_{C1}(x_n) & S_{C2}(x_n) \end{bmatrix} \text{ and } y=\begin{bmatrix} y(x_1) \\ \vdots \\ y(x_n) \end{bmatrix}$$

Since S_{C1} and S_{C2} have a similar shape, and are only observed experimentally, the solution of the normal equations is unstable—we cannot guarantee that a and b will be positive. Thus, it is necessary to regularise the model. In our approach, Ridge Regression has been applied to the normal equations.

$$\begin{bmatrix} a \\ b \end{bmatrix}=(A^T A+\alpha \mathrm{I})^{-1} A^T y$$

The effect is to force the estimates of a and b to be similar and thereby become positive.

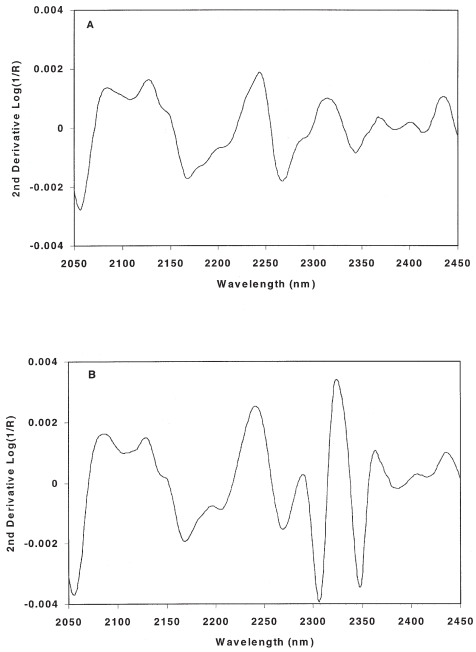

Figure 1. Derived curves for pure (a) gliadin and (b) glutenin.

Results

Pure gliadin and glutenin is difficult to obtain without resorting to quite complex chemistry. There is also the problem of maintaining an internal consistency between the spectra. To overcome these problems, a set of mixtures of the gliadin- and glutenin-rich extractions was made and the least squares analysis used to derive curves which represent the spectral response of the pure material. These curves are shown in Figure 1. The curves were used to fit spectra of the Halberd × Cranbrook test set. The results for glutenin are shown in Figure 2. The method was then applied to two further sets of samples which are more representative of the material typically found in a wheat breeding programme. The results are shown in Figure 3. Figure 4 shows a comparison of the results obtained from flour spectra and whole grain spectra.

Conclusions

We have shown that a least squares ridge regression methodology can be used to extract pure spectral information from mixtures of gliadin, glutenin and starch. This information can be used to rank

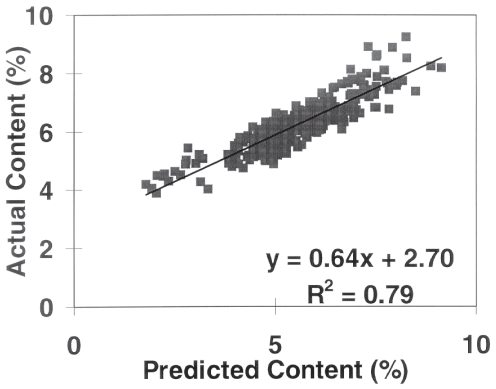

Figure 2. Actual glutenin content vs curvefit-predicted glutenin content for double haploid flour.

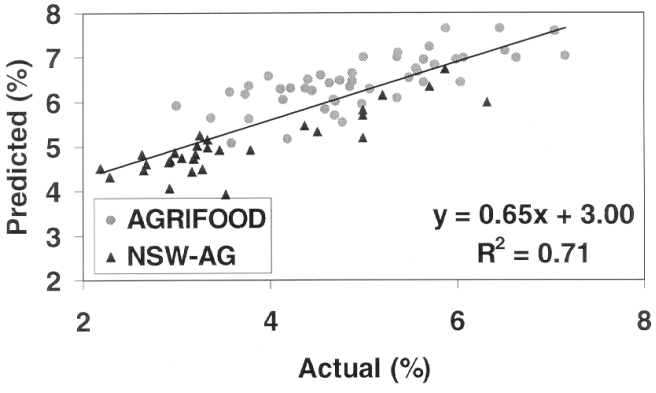

Figure 3. Actual glutenin content vs curfit-predicted glutenin content for flours from breeding programmes.

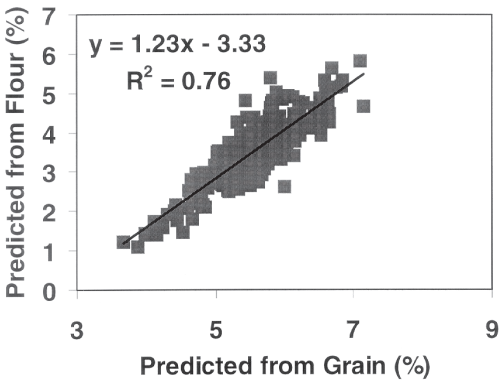

Figure 4. Glutenin content of flour vs glutenin content of whole grain by curvefit methods.

samples according to glutenin content. Measurements on flour and whole grain spectra are highly correlated. The technique can be applied to samples obtained from widely differing environments.

Acknowledgements

This project is funded by the Grains Research and Development Corporation of Australia through the Grain Industries Centre for NIR. Technical assistance was provided by Surjani Uthayakumaran, Sandra Partridge and Oscar Larroque. The input of Ferenc Bekes and Jeff Wood is also acknowledged. Samples were supplied by Geoff Cornish, Helen Allen and Naz Azudin.

References

1. W.R. Hruschka and K.H. Norris, *Appl. Spectrosc.* **36,** 261(1982).
2. I.L. Batey, R.B. Gupta and F. MacRitchie, *Cereal Chem.* **68,** 207 (1991).
3. R.B. Gupta, K. Khan and F. MacRitchie, *J. Cereal Sci.* **18,** 23 (1993).
4. F. MacRitchie, *J. Cereal Sci.* **3,** 221 (1985).