Available online at www.sciencedirect.com
ScienceDirect
Food Chemistry 106 (2008) 386–389
Food Chemistry
www.elsevier.com/locate/foodchem

Analytical, Nutritional and Clinical Methods

# Discrimination of European wheat varieties using near infrared reflectance spectroscopy

Carlos Miralbés *

*Harinera La Meta S.A., Pol. Ind. "El Segre" C/Industria Parc.204-R, 25199, Spain*

Received 9 October 2006; received in revised form 14 May 2007; accepted 28 May 2007

**Abstract**

Near infrared reflectance (NIR) spectroscopy combined with chemometrics was used to discriminate wheat varieties. A total of 249 samples of different wheat varieties from the 2003–2004 harvest were used to develop the best discriminant equation, by applying various scatters and mathematical treatments in the range of 400–2500 nm. Wheat varieties from Spain were 'Sarina', 'Bolero', 'Berdún', 'Soisson', 'Chamorro', 'Artur Nick', 'Berdun', 'Marius', 'Anza', 'Kalifa', and wheat varieties from France were 'Galibier' and 'Quality'. The equation developed with the highest accuracy had an applied scatter of weighted multiplicative scatter correction, a math treatment of 2, 15, 8 (order of derivative, gap data points over which the derivative was taken, number of data points used in performing average smoothing). The percentage of correctly identified varieties was 99.5% for the calibration sample set and 94% for the validation sample set. The results demonstrated the usefulness of NIRS combined with chemometrics as a rapid method for discrimination of European wheat varieties. Although the application of the discriminant equation developed for the 2003–2004 harvest yielded a high rate, further test measurements are necessary to evaluate the robustness of the equation.


*Keywords:* NIRS; Discrimination; Wheat; Varieties

## 1. Introduction

In the milling industry, the wheat delivery stage is one of the fundamental points in obtaining homogeneous flours, an incorrect classification of varieties will lead to low homogeneity in the batches produced. Millers have a limited number of silos in which to store the grain. Generally, wheat is classified according to variety; and within the same variety, if the parameters are differentiable, such as protein and gluten, it can be sub-classified into other silos. These classifications by variety are done by visual evaluation, requiring a long, dedicated training process. Today, classification is done by trained personnel that examines the size, shape, colour and other physical aspects of the grain. Wheat delivery is always aided by a lab that provides more detailed information on certain physiochemical properties of wheat (alveograph parameters, wet gluten, dry gluten, falling number) that cannot be evaluated at delivery time because the methods are time consuming. It has been years since NIR technology arrived to the milling industry, allowing to obtain rapid, accurate data for quantitative determination of humidity and protein (Delwiche, 1995, 1998). Experiments have been carried out to predict other parameters, such as dry gluten, wet gluten and rheological properties, with satisfactory results (Delwiche, Graybosch, & Peterson, 1998; Miralbés, 2003).

In Spain, as in other countries in the European Economic Community, the classification of varieties or classes of wheat is done based on the value of alveographic parameters. Generally, the object is to preserve the varietal character; if however, there are no silos available, various varieties or classes of wheat with the same alveographic value can be classified in the same silo.

Often the primary goal of analytical measurement tasks is not to find good estimates of continuous reference values

* Tel.: +34 973216139; fax: +34 973205643.
*E-mail address:* cmiralbes@lameta.es

0308-8146/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.foodchem.2007.05.090

but rather to determine whether a sample belongs to one of a number of categories or subgroups. Variety identification by electrophoresis is not applicable to routine control in industry. Current methods for wheat variety identification are sophisticated and time consuming (electrophoretic identification of cereal varieties by acid-PAGE, variety identification by HPLC, electrophoretic identification of cereal varieties by SDS-PAGE). A rapid screening method is necessary mainly for the segregation of wheat in grain silos according to its variety. Researchers have used discriminant analysis using NIRS for classification of wheat classes, however in some cases the accuracy decreases when the model was applied to samples grown during years that are not included in the model's calibration set (Delwiche & Norris, 1993; Delwiche, Chen, & Hruschka, 1995; Baker, Herrman, & Loughin, 1999).

The objective of this study was to develop discriminant analyses through NIRS and its application in routine analyses in the milling industry at the wheat reception stage in order to permit rapid and accurate varietal classification.

## 2. Materials and methods

### 2.1. Methods

Moisture, protein and wet gluten were determined according to the approved AACC methods (AACC, 2000). Deformation energy (W) was determined by alveograph test according to the approved AACC method (AACC, 2000).

### 2.2. Wheat samples

A total of 249 samples of different wheat varieties ('Sarina', 'Bolero', 'Berdún', 'Soisson', 'Chamorro', 'Artur Nick', 'Berdun', 'Marius', 'Anza', 'Kalifa') from different areas of Spain and different varieties of wheat ('Galibier' and 'Quality') from different areas of France were used from the 2003–2004 harvest. A fifty out of 249 samples were used for the validation. All the samples coming from unique lots were used to establish the validation equations.

The hardness of the varieties were characterized according to the texture of the grain as either soft or hard. The colour was determined visually.

### 2.3. NIRS hardware

A near infrared spectrophotometer (Foss NIRSystem 6500) was used to collect reflectance spectra of whole wheat samples. The instrument was equipped with a rectangular sample transport cell (100 mm long, 61 mm width, and 12 mm depth). The samples were scanned 32 times to create average spectra in the reflectance mode over the range of 400–2500 nm at 2 nm intervals. With a sample of one kilo of each variety, six re-packs between scans were performed, where the same grain was not used during the re-packing. The sample quantity was between 20 and 22 g. Spectra were obtained by averaging six scans. All spectra were recorded as log $(1/R)$, where $R$ is the relative reflectance. The sample set used in the study was split into a calibration set containing 199 samples and a validation sample set comprising 50 samples.

### 2.4. Discriminant equation and validation

A WINISI III (ver 1.50e) software was used for spectral data analysis and development a chemometric models. Discriminant analysis with Partial Least Squares (PLS2) were performed using WINISI software(Foss NIRSystems). When a PLS2 algorithm is applied, the results of the spectral decomposition give one set of scores and one set of eigenvectors for calibration. The PLS2 procedure used in discriminant does not result in a single set of regression coefficients. It results in two vectors, a weight vector and a loading vector, per factor. Prior to calibration, Principal Components Analysis (PCA) was performed to examine the qualitative difference between varieties and to remove outliers with a standardized Mahalanobis (GH) distance greater than 2.5.

By examination of raw reflectance spectra of wheat samples no significant spectral differences were observed. However, when spectra were averaged and derivatized, reasonable spectral differences were observed between varieties.

In each of the discriminant equations developed, five cross-validations were applied with 15 terms, varying the spectral pre-treatments and scatters corrections, and an uncertainty factor of 1. With full cross-validation, each sample is removed one at a time from the sample set, a new calibration performed and a predicted scored calculated for the sample is removed. This procedure is repeated until all samples have been removed from the sample set once.

## 3. Results and analysis

Physical–chemical characteristics of the varieties of wheat used for the calibration and validation of varieties from the 2003–2004 harvest are shown in Table 1. As can be seen, there was a large variability in the protein content, wet gluten, and energy deformation. These varieties studied are used by the Spanish mill industry for the production of cookies, bread, and pastries.

### 3.1. Spectral analysis of varieties

Differences in size, shape, colour and chemical composition, but mainly in kernel texture (hard and soft wheat) cause differences in the spectra characteristics of wheat varieties. When corrected for scatter by weighted multiplicative scatter correction (WMSC), followed by a second derivative, several absorbance peaks arise corresponding to major chemical constituents in wheat (Fig. 1). Only three outliers were eliminated.

Table 1
Summary of properties of wheat varieties

| Varieties[a] | $n$[b] | CL[c] | HD[d] | $M$[e] | | $P$[f] | | WG[g] | | $W$[h] | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Mean | Range | Mean | Range | Mean | Range | Mean | Range |
| AZ | 36 | Brown | Hard | 12.5 | 11.0–13.1 | 11.5 | 11.1–13.0 | 24.3 | 22.0–25.5 | 102 | 85–108 |
| AR | 10 | Yellow | Soft | 11.5 | 9.1–13.3 | 13.2 | 13.1–14.2 | 27.3 | 26.2–28.5 | 105 | 90–125 |
| BE | 10 | Brown | Soft | 13.2 | 12.5–14.0 | 12.2 | 12.0–12.9 | 23.7 | 23.5–25.5 | 139 | 125–155 |
| BO | 19 | Yellow | Soft | 12.8 | 11.5–13.3 | 12.1 | 11.5–12.5 | 22.5 | 22.0–24.0 | 115 | 90–125 |
| CH | 9 | Yellow | Soft | 11.0 | 9.5–12.3 | 12.0 | 11.0–12.5 | 23.2 | 23.0–24.5 | 65 | 50–80 |
| GA | 12 | Brown | Hard | 13.5 | 12.9–14.7 | 14.8 | 14.7–15.1 | 31.2 | 30.5–33.8 | 350 | 315–405 |
| KA | 11 | Red | Hard | 13.2 | 12.6–13.8 | 13.2 | 12.8–14.1 | 28.2 | 25.4–30.8 | 285 | 260–310 |
| MA | 40 | Red | Soft | 11.9 | 11.1–13.0 | 12.2 | 11.2–13.1 | 23.5 | 22.1–26.0 | 90 | 80–100 |
| QU | 11 | Red | Hard | 13.0 | 12.1–13.9 | 14.9 | 14.7–15.2 | 31.9 | 30.1–33.7 | 355 | 330–395 |
| SA | 46 | Beige | Hard | 12.5 | 11.9–13.2 | 12.1 | 12.0–12.8 | 23.2 | 22.8–24.5 | 121 | 95–126 |
| SO | 45 | Red | Hard | 12.8 | 11.8–12.9 | 12.1 | 11.8–13.1 | 24.1 | 22.9–25.1 | 195 | 160–240 |

[a] AZ, 'Anza'; AR, 'Arthur nick'; BE, 'Berdun'; BO, 'Bolero'; CH, 'Chamorro'; GA, 'Galibier'; KA, 'Kalifa'; MA, 'Marius'; QU, 'Quality'; SA, 'Sarina'; SO, 'Soisson'.
[b] Number of samples for calibration/validation.
[c] Kernel color.
[d] Kernel hardness.
[e] Moisture content (%).
[f] Protein content (%).
[g] Wet gluten (%, 14% mb).
[h] Deformation energy ($10^{-4}$ J).

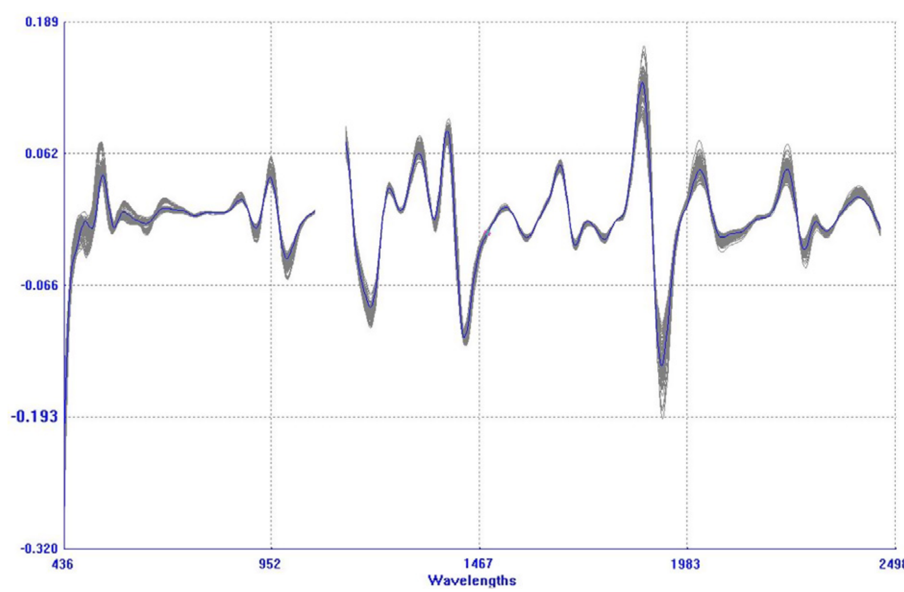


Fig. 1. Second derivative spectra of wheat samples.

### 3.2. Discriminant analysis

Among the various scatter and mathematical treatments used, the best discriminant equation had a WMSC (Weighted Multiplicative Scatter Correction) and a mathematical treatment of 2, 15, 8 (order of derivative, gap data points over which the derivative was taken, number of data points used in performing average smoothing), in which the number of factors was the lowest and the number of incorrectly classified samples was the lowest in the calibration sample set and in the validation sample set (Table 2).

For calibration, there were 36 samples of variety 'Sarina' of which 1 was identified as 'Soisson', and there were 36 samples of variety 'Soisson' of which one was identified as 'Sarina'. For the other varieties all the samples were correctly classified. The percentage of correctly identified varieties was 99.5% for the calibration sample set.

For validation sample set, 'Sarina' and 'Soisson' showed misclassification whereas the other varieties were correctly identified. There were 10 samples of variety 'Sarina' of which 1 was identified as 'Anza', and there were 9 samples of variety 'Soisson' of which 2 was identified as 'Sarina'. The discrepancies were always associated with 'Soisson',

Table 2
Variety classification accuracy of various treatments and scatters

| Scatter[a] | Math treatment[b] | Factor[c] | Misclassified calibration | Misclassified validation |
|---|---|---|---|---|
| NONE | | | | |
| | 0,0,1 | 15 | 5 | 6 |
| | 1,4,4 | 15 | 3 | 5 |
| | 2,15,8 | 14 | 3 | 5 |
| SNV + D | | | | |
| | 0,0,1 | 14 | 3 | 7 |
| | 1,4,4 | 14 | 2 | 6 |
| | 2,15,8 | 15 | 2 | 5 |
| W MSC | | | | |
| | 0,0,1 | 15 | 5 | 7 |
| | 1,4,4 | 15 | 4 | 6 |
| | 2,15,8 | 12 | 2 | 3 |

[a] NONE, Raw spectra; SNV + D, standard normal variate and detrending; W MSC, weighted multiplicative scatter correction.
[b] 1st digit: order of derivative; 2nd digit: gap data points; 3rd digit: smoothing.
[c] Number of factor used for cross-validation.

'Sarina', and 'Anza'. For the validation sample set the percentage of correctly identified varieties was 94%. Taking into account that the spectral distance (Mahalanobis distance) between these groups of varieties is not very far, some overlapping areas can be observed causing classification errors.

The outliers corresponded to the 'Soisson' variety. When the discrimination analysis was performed, including the three outliers, the results of the calibration and validation were the same. This fact was surely due to the small number of outliers in respect to the total population of samples.

## 4. Conclusions

The results obtained showed that the routine application of discriminant equations through the use of near infrared reflectance spectroscopy is a rapid method of varietal identification that does not require qualified personnel in milling industry wheat delivery. In some cases, when individual quality parameters (protein, wet gluten and deformation energy) between wheat varieties are very similar the discriminant equation would act as a preliminary evaluation before applying quantitative calibration equations. This study has worked well for wheat varieties from France and Spain. However, further studies have to be done for mixed varieties or classes of wheat that include a large number of wheat varieties.

## References

AACC (2000). *Approved methods of the American Association of Cereal Chemists* (10th ed.). St. Paul, Minnesota: The Association.

Baker, S., Herrman, T. J., & Loughin, T. (1999). Use of regression and discriminant analyses to develop a quality classification system for hard red winter wheat. *Cereal Chemistry, 76*, 890–893.

Delwiche, S. R. (1995). Single wheat kernel analysis by near-infrared transmittance: Protein content. *Cereal Chemistry, 72*, 11–16.

Delwiche, S. R. (1998). Protein content of single kernels of wheat by near-infrared reflectance spectroscopy. *Journal of Cereal Science, 27*, 241–254.

Delwiche, S. R., Chen, Y.-R., & Hruschka, W. R. (1995). Differentiation of hard red wheat by near-infrared analysis of bulk samples. *Cereal Chemistry, 72*, 243–247.

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

Delwiche, S. R., & Norris, K. H. (1993). Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. *Cereal Chemistry, 70*, 29–35.

Miralbés, C. (2003). Prediction chemical composition and Alveograph parameters on wheat by near-infrared transmittance spectroscopy. *Journal of Agricultural and Food Chemistry, 51*, 6335–6339.