Discrimination of Bread-Baking Quality of Wheats According to Their Variety by Near-Infrared Reflectance Spectroscopy

MARIE FRANÇOISE DEVAUX,[1] DOMINIQUE BERTRAND,[1] and GUY MARTIN[2]

ABSTRACT

Cereal Chem. 63(2):151–154

Assessment of the bread-baking quality of French soft wheat by near-infrared reflectance (NIR) spectroscopy, using standard mathematical treatments and reference tests, has had limited success. New mathematical treatments of NIR spectra are reported that assign a technological value to wheat varieties to differentiate potential bread-baking quality categories. The seven most cultivated soft wheat varieties in France were chosen for this study. Two preliminary mathematical treatments were applied to reduce the number of spectral data (351 wavelengths) without loss of information. These were fast Fourier transform and principal component analysis of the Fourier coefficients. By these methods, the log (1/R) spectral values were transformed to 10 independent synthetic variables that were subsequently treated by multiple discriminant analysis. The first mathematical treatment was made to discriminate and recognize the seven varieties, and 74% of unknown samples were identified (79 samples for calibration and 42 for prediction). The two varieties that were principally confused (Fidel and Talent) were of the same bread-baking quality. The second mathematical treatment concerned three bread-baking categories: good, unsuitable, and irregular. Correct category assignments were made for 95% of unknown samples. Advantages of those mathematical methods are discussed.

Industrial production requires a rapid quality assessment of raw materials. There is a great need in the baking industry for a method to rapidly estimate the bread-baking quality of wheat. This quality is often estimated by laboratory screening tests, such as the Zeleny index, the falling number of Hagberg, and the Chopin alveograph (Godon 1984). These analyses are relatively time consuming, and agreement between laboratory tests and baking experiments (AFNOR 1980) is often poor (Godon 1984).

Near-infrared reflectance (NIR) spectroscopy has been widely developed as a fast analytical method. Until now, the prediction of bread-baking quality of wheat with NIR has met with limited success (Osborne 1984). A possible reason is the questionable suitability of some manual reference quality tests. NIR analysis requires calibration with samples of known quality and, therefore, its accuracy depends on the accuracy of the reference methods. Another way to predict baking quality by NIR would be to differentiate wheats without reference tests, but taking into account only their recognized baking quality. In France, the Office National Interprofessionnel des Céréales (ONIC) classifies soft wheat varieties into three categories: 1) regularly good baking quality, 2) regularly unsuitable baking quality, and 3) irregular baking quality.

The objective of this study was to develop mathematical treatments of NIR spectra to identify different French soft wheat varieties and quality categories. Multilinear regressions, which are usually employed to set up prediction equations (Osborne 1984), are unsuitable for direct discrimination. It is necessary to use other methods, principally multiple discriminant analysis (MDA) (Davies 1971, Foucart 1982). Usual statistical methods of discrimination applied with microcomputers are unable to handle more than approximately 50 variables. Moreover, MDA is not applicable to data that are highly correlated. Digitized NIR spectra, obtained with reference spectrometers, may include as many as 700 points, and spectral data of different wavelengths are generally correlated (Bertrand 1985).

Fast Fourier transform (FFT) has been shown to efficiently reduce the number of spectral data points with very slight loss of accuracy (McClure et al 1984). Another mathematical treatment, principal component analysis (PCA), allows the creation of new synthetic variables that are not correlated (Davies 1971, Foucart 1982). In the present study, these two treatments were applied sucessively; resulting data were then treated by an MDA program.

All of these mathematical treatments were tested on a collection of wheat samples from the seven main soft wheat varieties cultivated in France, representing about 70% of production. Two discrimination tests were performed, one for wheat variety identification and the other for prediction of baking quality categories as defined by the ONIC.

MATERIALS AND METHODS

Wheat Collection

The wheat collection came from the annual quality survey of the ONIC and the Institut Technique des Céréales et des Fourrages (ITCF). The number of collected samples of each variety was roughly proportional to its crop area. One hundred and twenty-one samples of seven varieties (Fidel, Talent, Hardi, Arminda, Festival, Camp Remy, and Top) from the 1984 crop year were evaluated. These samples were divided into two groups, one group of 79 samples for calibration and another of 42 samples as a prediction set. The wheat samples were ground with a Cyclotec laboratory mill (screen 0.5 mm).

NIR Measurements

NIR spectra were recorded on a Technicon Infraalyzer 500 spectrophotometer from 1,100 to 2,500 nm at intervals of 4 nm. Thus the spectra were composed of 351 individual data points. The spectral data were stored in a 128-kilobyte microcomputer on which the following successive treatments were carried out: 1) reduction of data number using FFT algorithm, 2) concentration of the new data into a few synthetic variables using PCA, and 3) discrimination of varieties and bread-baking categories using MDA.

Fast Fourier Transform

The initial N points are transformed into N complex values, which can be expressed as a real and an imaginary part (Oran Brigham 1974, Demars 1981).

$$X(n) = \sum_{k=0}^{N-1} x(k) e^{-2\pi j\, nk/N}$$

$$n = 0, 1, \ldots, N-1$$

where $X(n)$ = complex coefficients,
$x(k)$ = initial data,
N = number of data points, and
$j = \sqrt{-1}$.

McClure (1984) has demonstrated that only the first complex

[1]Institut National de la Recherche Agronomique, Laboratoire de Technologie des Aliments des Animaux, Rue de la Gèraudière, 44072 Nantes Cedex, France.
[2]Institut Technique des Cereales et des Fourrages, 16, Rue Nicolas Fortin, 75013 Paris, France.

values are necessary to retain almost all of the spectral information. Our preliminary experiments showed that the efficiency of the FFT algorithm could be improved by two modifications of the original spectra—the addition of five points at the beginning and at the end of the spectrum, respectively, equal to its first and last values, and the multiplication of this modified spectrum by the Hanning function (Oran Brigham 1974):

$$H(t) = \frac{1}{2} (1 - \cos(2\pi t/T));$$

$$T = 361.$$

As for the FFT algorithm, the number of data points must be an integral power of 2 (128, 256, 512, ...). One hundred and fifty-one artificial variables, equal to zero, were added at the end of the modified spectrum. After Fourier transformation, 80 Fourier coefficients (40 real and 40 imaginary) were kept for the next mathematical treatments. They were also used to test the accuracy of the reconstitution of the spectra.

Principal Component Analysis

Our PCA program was not able to handle more than 50 principal and 30 supplementary variables, so the first 50 Fourier coefficients were chosen as principal variables and the others as supplementary variables. This procedure allowed us to study the quality of representation of the 30 coefficients not involved in the creation of the PCA synthetic variables. Samples from the prediction set were used for principal observation and the others for supplementary observations.

Multiple Discriminant Analysis

MDA was used to determine the variety and the baking quality categories as defined by the ONIC and the ITCF. Quality categories depend on the variety as follows: class 1, wheat of regularly good bread-baking quality, included Hardi, Festival, and Camp Remy; class 2, regularly unsuitable bread-baking quality, included Arminda; and class 3, irregular, included Fidel, Talent, and Top.

The calibration set was used to develop the typical data patterns of the represented groups (variety or quality category); the prediction set contained samples which were tested as if they were unknown. The variables used in MDA were the synthetic variables resulting from PCA treatment.

The MDA program created discriminant variables and gave representations of the groups to be identified. Some nonsignificant discriminant variables were eliminated according to the statistical procedure described by Romeder (1973).

RESULTS AND DISCUSSION

Accuracy of Fast Fourier Transform

The Fourier transform was tested by comparing the spectrum generated from the 80 first Fourier coefficients with the original spectrum. The average difference observed was 0.0014 log (1/R)

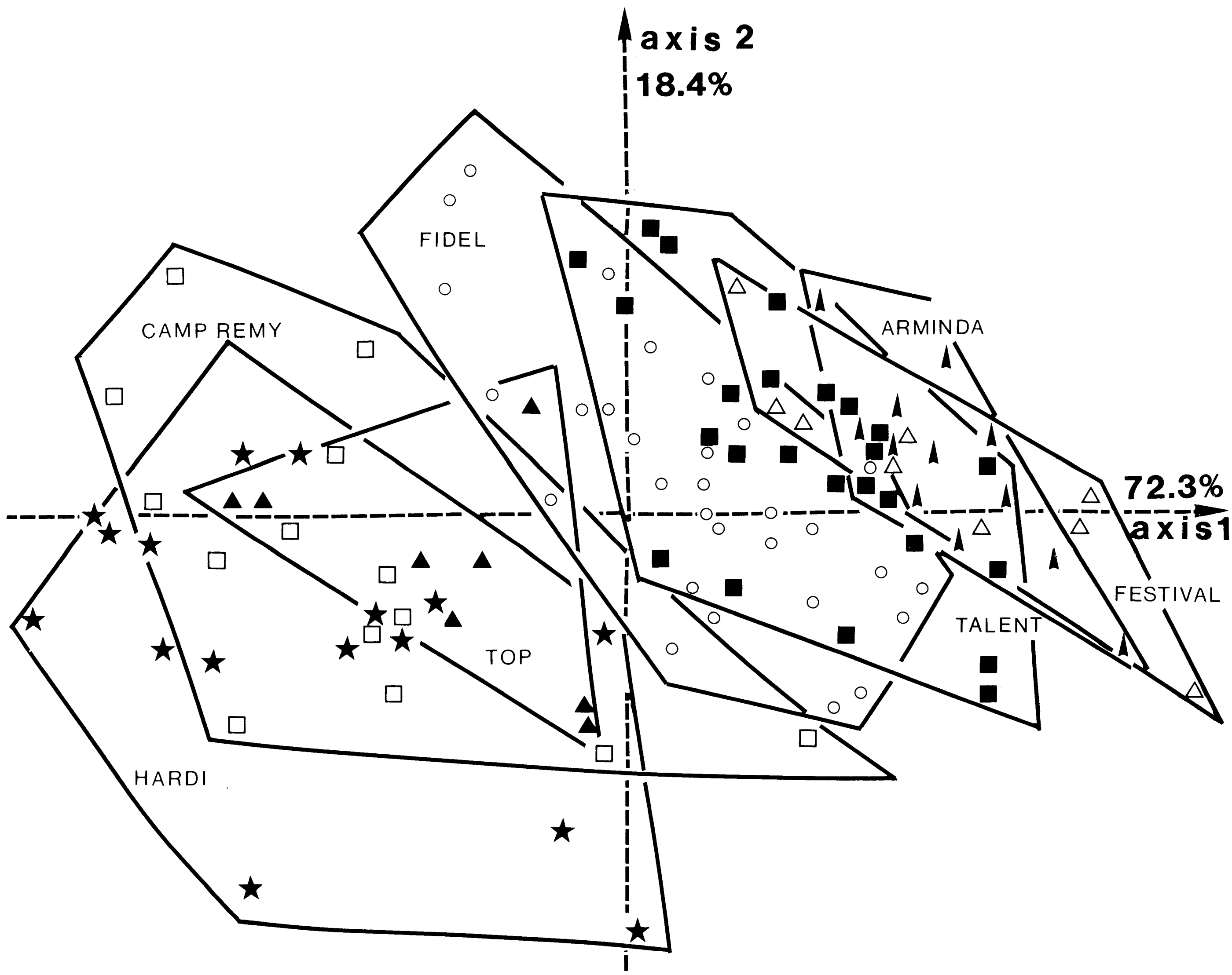

Fig. 1. Principal component analysis of wheat spectra. Representation of the first two variables (axes 1 and 2, respectively).

units, and the standard deviation was 0.0037 log (1/R) units. Our previous experiments showed that the repetition of the NIR record of a same wheat flour gave a maximum standard deviation of about 0.01 log (1/R) units. So, the errors observed in FFT spectral reconstruction were comparatively small.

Principal Component Analysis Treatment

The first 50 Fourier coefficients led to the creation of 10 synthetic variables that concentrate 99.94% of total variance. The next 30 coefficients (supplementary variables) were correctly described. For nine of them, their squared multiple correlation coefficient with PCA variables was more than 0.95. For the other variables, it was more than 0.98. Figure 1 shows the representation of the two first PCA variables (90.7% of total variance). The samples of the same variety were rather close to each other and a differentiation of varieties still appeared without MDA treatment.

Multiple Discriminant Analysis

The first 10 PCA variables represented a very large part of total variance, and they were chosen as subjects for MDA treatment. As seven varieties were to be discriminated, six discriminant variables could be created by the calibration set. One variable contributed little to the discrimination; a chi-square test (Romeder 1973) allowed us to delete it. Figure 2 shows the discrimination of variety for the first two MDA variables. Compared to PCA, MDA considerably increased the efficiency of differentiation. Three groups were clearly separated on these axes: 1) Hardi, Camp Remy, and Top; 2) Fidel, Talent, and Festival; and 3) Arminda. It seemed that these groups roughly corresponded to quality categories: Camp Remy, and Hardi on one side, and Fidel and Talent, which are known to have similar bread-baking quality, on the other. Arminda, which is unsuitable for bread baking, stood alone. It must be noted that the other axes, 3, 4, and 5 (not represented), increased the efficiency of the discrimination. A computer classification is given in Table I. Seventy-nine percent of the samples in the calibration set and 74% of the "unknown" samples were correctly reclassified. The principal misidentification was between samples of the same bread-baking category. Only two unknown samples among 42 were incorrectly reclassified in their original bread-baking category.

Direct discrimination of bread-baking categories is represented in Figure 3. The categories were quite well differentiated. Some overlap appeared between categories 1 and 3 (good and irregular). This seems logical because some wheats classified as irregular may be of good baking quality. Category 2 (unsuitable for bread baking) was completely distinguished. The resulting classification is given in Table II. Eighty-seven percent of the samples of the calibration set were correctly classified, and 95% of samples were successfully

TABLE I
Computerized Identification of the Seven Wheat Varieties
(Number of Samples Assigned to Each Variety)

Actual Variety[b]	Variety Identified by MDA[a] — Calibration Set FI	HA	AR	CR	TA	TO	FE	Prediction Set FI	HA	AR	CR	TA	TO	FE
FI	12				8			3		1		7		
HA		10			1				5					
AR			10							4				
CR		3		7					1		4			
TA	3		1		9			1				12		
TO						7							2	
FE			1				7			1				1

[a]Multiple discriminant analysis.
[b]FI, Fidel; HA, Hardi; AR, Arminda; CR, Camp Remy; TA, Talent; TO, Top; FE, Festival.

TABLE II
Computerized Identification of the Bread-Baking Categories
(Number of Samples Assigned to Each Category)

Actual Category[b]	Category Identified by MDA[a] — Calibration Set 1	2	3	Prediction Set 1	2	3
1	24	2	3	11	1	
2		10			4	
3	3	2	35	1		25

[a]Multiple discriminant analysis.
[b]1, regularly good for bread baking; 2, regularly unsuitable for bread baking; and 3, irregular.

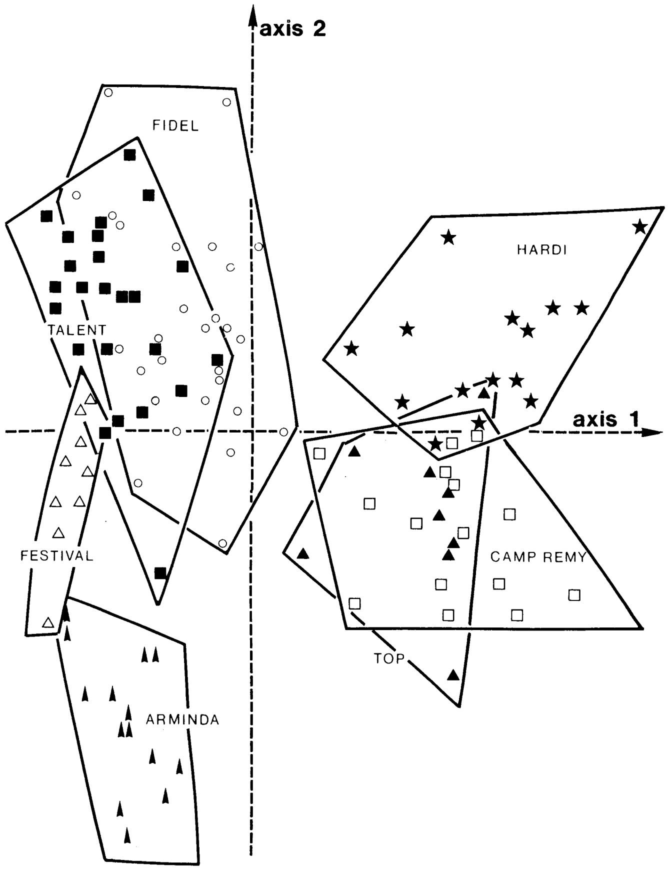

Fig. 2. Multiple discriminant analysis of wheat varieties. Representation of the first two discriminant variables.

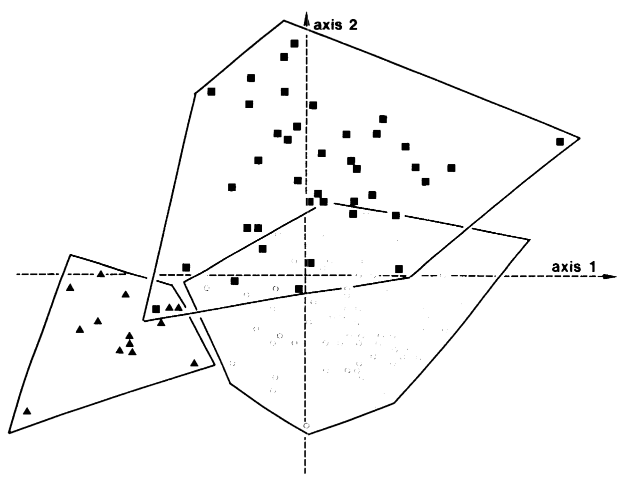

Fig. 3. Multiple discriminant analysis of bread-baking categories. Representation of the two discriminant variables. Symbols: ■, regularly good for bread baking; ▲, regularly unsuitable for bread baking; and O, irregular.

identified in the prediction set, with only two incorrect identifications among 42. This percentage is the same as was observed in the variety discrimination. Therefore, there appears to be correspondence between variety discrimination and prediction of bread-baking category with respect to mathematical treatment of NIR spectra. Through FFT and PCA, the NIR technique was sensitive to these parameters.

CONCLUSION

The proposed method is fast. Calibration needs about 2 hr of computing time, and identification of one sample requires less than 2 min of calculation. In comparison with a method such as electrophoresis (Bushuk 1978), the NIR method of bread quality prediction is much quicker but probably less accurate. However, fast speed of classification is very important in the cereal industry. Data concentration (FFT followed by PCA) was very efficient. Ten synthetic variables were sufficient to describe almost all the spectral information. Any mathematical treatment (MDA, regressions, etc.) calculated from these variables would be less sensitive to NIR background noise and would take into account a large number of original spectral data.

ACKNOWLEDGMENT

We acknowledge with gratitude the help of the Office National Interprofessionnel des Céréales (ONIC), Paris, which provided us with wheat samples of the annual quality survey.

LITERATURE CITED

AFNOR (ASSOCIATION FRANCAISE DE NORMALISATION). 1980. Farine de blé tendre. Essai de panification. Method V-03-713. AFNOR:Paris.

BERTRAND, D., ROBERT, P., and TRAN, V. 1985. Page 93 in: Traitements mathématiques de spectres NIR de mélanges. Proc. 11th ICC Mtg., Vienna.

BUSHUK, W., and ZILLMAN, R. R. 1978. Wheat cultivar identification by gliadin electrophoregrams. I. Apparatus, method and nomenclature. Can. J. Plant Sci. 58:505.

DAVIES, R. G. 1971. Page 291 in: Computer Programming in Quantitative Biology. Academic Press: London.

DEMARS, C. 1981. Transformée de Fourier rapide. Microsystèmes 19:155.

FOUCART, T. 1982. Page 82 in: Analyse factorielle, programmation sur micro ordinateur. Masson: Paris.

GODON, B., and LOISEL, W. 1984. Page 487 in: Guide pratique d'analyse dans les industries des céréales. Lavoisier: Paris.

McCLURE, W. F., ABOUL, H., GIESBRECHT, F. G., and WEEKS, W. W. 1984. Fourier analysis enhances NIR diffuse reflectance spectroscopy. Appl. Spectrosc. 38(3):222.

ORAN BRIGHAM, E. 1974. The Fast Fourier Transform. Prentice Hall: Englewood Cliffs, NJ.

OSBORNE, B. G. 1984. Investigation into the use of near infrared reflectance spectroscopy for the quality assessment of wheat with respect to its potential for bread baking. J. Sci. Food Agric. 35:106.

ROMEDER, S. M. 1973. Page 50 in: Méthodes et programmes d'analyse discriminante. Dunod Bordas: Paris.

[Received June 10, 1985. Accepted October 21, 1985.]