

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/230766027>

Identification of Chemical Constituents by Multivariate Near-Infrared Spectral Imaging

ARTICLE *in* ANALYTICAL CHEMISTRY · MARCH 1992

Impact Factor: 5.64 · DOI: 10.1021/ac00030a017

CITATIONS

24

READS

15

4 AUTHORS, INCLUDING:



Dominique Bertrand

data_frame

189 PUBLICATIONS 2,496 CITATIONS

[SEE PROFILE](#)



Marie-Françoise Devaux

French National Institute for Agricultural R...

95 PUBLICATIONS 2,036 CITATIONS

[SEE PROFILE](#)

Identification of Chemical Constituents by Multivariate Near-Infrared Spectral Imaging

Paul Robert,* Dominique Bertrand, Marie Françoise Devaux, and Alain Sire

Institut National de la Recherche Agronomique, Centre de Nantes, B.P. 527 rue de la Géraudière, 44026 Nantes Cedex 03, France

A near-infrared imaging spectroscopic system was tested to identify three main components of wheat (bran, gluten, starch). The system described below permitted the recording of images between 900 and 1900 nm by steps of 50 nm. A direct examination of images showed that starch was correctly identified at 1550 nm. However, such a direct study of images was not sufficient to characterize all the constituents. The images, therefore, were linearly combined by applying discriminant analyses. The more relevant wavelengths (950, 1450, 1500 nm) were determined using a stepwise discriminant analysis, and a mapping of the chemical constituents was obtained by applying a canonical discriminant analysis. In the segmented image, the percentages of well-classified pixels were 92 for bran, 95 for gluten and 99 for starch.

INTRODUCTION

An expansion in image analysis applications is occurring within the agriculture and food industries with the result that image analysis can be used for the characterization of food products. Images are often studied for detecting or enhancing geometrical structures. For instance, Goodman and Rao used image analysis to classify rice kernels according to their physical dimensions.¹ Numerous authors used the extraction of size and shape parameters in order to discriminate cereal grains.²⁻⁴ Besides the assessment of size and shape parameters, color measurements can be achieved to characterize various agricultural and food products. Neuman et al.^{5,6} described a procedure for color analysis of individual wheat grains. The results obtained indicated that grain color measurement might facilitate the identification of grains in the wheat-grading context.

The study of color or intensity of the points (pixels) in an image can be a way to obtain chemical information. Newman⁷⁻¹⁰ showed that image analysis makes it possible to estimate the fat and lean contents in meat and meat products. In the case of minced meat, the fat was differentiated from lean using an ultraviolet light. In such conditions, fat fluoresces white in the visible range. In the UV region, the movement of α amylase and the cell-wall breakdown during barley malting were visualized by taking advantage of the fluorescent effect of Calcofluor.^{11,12} Computerized image analysis was investigated by Heyne et al.¹³ to determine the protein quality of simulated pizza crusts. They concluded that the method recommended could be used to control oven temperature and to prevent heat-processing damage of proteins.

The relationship between an image and the chemical properties of a sample is generally complex. For example, the most common problems occur when several chemical constituents are present in a sample. Problems such as matrix effects, chemical interferences, and/or inadequate variations in pixel intensities between constituents could arise, making accurate chemical identification impossible. However, if images at multiple wavelengths are recorded, as was suggested by Geladi et al.,¹⁴ the power of discrimination increases. This

approach is commonly applied to satellite images for the identification of vegetation.¹⁵ In the food domain, Munk et al.¹⁶ used intrinsic autofluorescence properties of the different botanical components in cereal seeds. By selecting the wavelengths of excitation and emission, the authors obtained images that discriminated endosperm, aleurone, pericarp, and testa. Imaging spectroscopic studies were performed in the near-near-infrared region (700–1100 nm) by Taylor and McClure¹⁷ using a CCD (charge coupled device) video camera and narrowband interference filters. The combination of images recorded at different wavelengths made it possible to measure the distribution of chlorophyll and moisture in plant materials and to detect healthy and unhealthy leaf tissues.

The aim of the present work was to design an imaging spectroscopic system that operated in the near-infrared range. Near-infrared spectroscopy is commonly used to predict composition or quality of numerous agricultural commodities¹⁸ and therefore could appropriately be combined with image analysis in the study of foods. A compelling challenge tackled in this paper was to study images with a spectral approach: a sample was not characterized by a single image, but by a set of images recorded at different wavelengths. Such a set of images cannot be studied by classical image processing. Geladi and Esbensen¹⁹ thus suggested applying multivariate statistical processing to extract chemical information from the set of images.

In this work, the efficacy of an imaging spectroscopic system was tested for the identification of three major constituents of cereals. As cereal processing consists of tissue fractionation, it might be of interest to study the *in situ* distribution of the constituents in grains. Three major components of wheat were chosen: starch, gluten, and bran. Images were recorded from 900 to 1900 nm in 50-nm intervals producing images at 21 wavelengths. Discriminant analyses were applied to identify the three components.

MATERIALS AND METHODS

Imaging Spectroscopic System. The imaging spectroscopic system previously described by Robert et al.²⁰ included an infrared radiation source, a monochromator, a sample compartment, and a near-infrared tube camera. The slits of the monochromator were adjusted to obtain a nominal bandwidth of 25 nm. The incident and observation angles of the light in the sample compartment were set to 30° and 10°, respectively, in relation to the normal of the sample. The near-infrared camera was fitted with a Nikon 55-mm photographic lens and with a filter (J 820 MTO) that eliminated residual visible light. The images were digitized with a resolution of 512 × 512 pixels (ISM 169 digitizing board, Matrox, 8 bits/pixel). Intensity values of the pixels were coded on 256 gray levels, where 0 corresponded to black and 255 to white. The digitized images were recorded with an IBM-compatible microcomputer. Hard copies of images were obtained by using a Theta Scan printer (D-Scan, CH-5504).

Samples. Pressed pellets of 12-mm diameter were prepared from starch, gluten and bran of wheat. The pellets were loaded in a closed cup of 35-mm diameter and surrounded with a ceramic powder (Al_2O_3 , 85%; SiO_2 , 15%, Pechiney) that did

not show an absorption band in the near-infrared range and reflected most of the incident light. The recorded images subsequently showed the three components of wheat on the same view.

Data Acquisition. As was shown previously,²⁰ noise increased and sensitivity decreased for the highest wavelength values in the imaging spectroscopic system. It was shown that the noise could be reduced by averaging several images. The optimal number of scans for noise reduction ranged from 10 to 16. However, in the present work, a large number of scans resulted in a time-consuming data acquisition. Therefore, the sample was scanned eight times in order to adequately reduce noise and to increase the speed of image acquisition. A "final" image corresponded to the addition of eight "elementary images". Furthermore, the "final" images were compressed from 512×512 to 256×256 pixels by averaging four neighboring pixels. The compression was needed for reducing the computing duration of the discriminant analyses applied to every pixel. The gray level of each resulting pixel was the average of 32 values (8 "elementary" images \times 4 neighboring pixels).

Mathematical Treatments. Each pixel in an image is characterized by its spatial coordinates x and y and by its gray-level value $G_{x,y}$. For a set of images recorded at different wavelengths l_1, l_2, \dots, l_n , the gray-level values of a given pixel can be represented by a vector $G_{x,y}(l_1, l_2, \dots, l_n)$. In this case, the dimension n of each vector was 21. In the present study, the set of data available corresponded to 65 536 (256×256) pixels or vectors $G_{x,y}(l_1, l_2, \dots, l_{21})$. These vectors could be compared to near-infrared spectra containing n points.

In near-infrared spectroscopy, statistical processing is applied to spectra for the determination of chemical constituents.²¹ This same data treatment could be usefully performed on spectroscopic images. However, the number of vectors $G_{x,y}(l_1, l_2, \dots, l_{21})$ was too large for the usual statistical treatments. The reduction of the number of vectors can be accomplished by selecting subsets of pixels. A first set of vectors was visually chosen for calibration: 30 pixels were selected for each chemical constituent (bran, gluten, starch, and ceramic). A second set was set up for verification (20 pixels for each chemical constituent). All the other pixels were considered as unknown.

Stepwise discriminant analysis was performed on the calibration and verification sets to determine the more relevant wavelengths for the identification of the four chemical constituents. Canonical discriminant analysis was applied to the same sets of vectors using the selected wavelengths. This method permitted the assessment of discriminant factors that were linear combinations of the initial wavelengths. Similarity maps of the pixels could be drawn by plotting the scores corresponding to two chosen factors.

The unknown pixels were classified in the four chemical groups by assessing their scores related to the discriminant factors. A segmented image was obtained: an arbitrary gray-level value was assigned to each chemical group, and the pixels were given the gray-level value of their group.

Software that allowed the averaging and the compression of images, the interactive selection of pixels for calibration and verification, and the statistical treatments were developed in the laboratory. The algorithms of stepwise discriminant analysis and of canonical discriminant analysis were derived from Romeder²² and from Foucart,²³ respectively.

RESULTS AND DISCUSSION

Near-Infrared Images. In near-infrared spectroscopy, most of the absorption bands observed between 900 and 1900 nm arise from overtones of stretching vibrations involving C-H, O-H, and N-H functional groups. Images of wheat components recorded at 1200 and 1550 nm are shown Figure

Table I. Stepwise Discriminant Analysis^a

wavelength (nm)	950	1500	1450	1000	1700	1600	900
ceramic	56.7	96.7	83.3	83.3	90.0	90.0	93.3
bran	26.7	50.0	83.3	90.0	86.7	86.7	93.3
gluten	56.7	63.3	86.7	86.7	93.3	90.0	93.3
starch	26.7	100.0	100.0	100.0	100.0	100.0	100.0

^a Percentages of calibration pixels correctly classified in their qualitative groups. The wavelengths were introduced according to their relevance.

Table II. Stepwise Discriminant Analysis^a

wavelength (nm)	950	1500	1450	1000	1700	1600	900
ceramic	60.0	100.0	90.0	75.0	80.0	90.0	95.0
bran	25.0	45.0	100.0	90.0	90.0	80.0	100.0
gluten	65.5	60.0	90.0	90.0	90.0	90.0	90.0
starch	25.0	100.0	100.0	100.0	100.0	100.0	100.0

^a Percentages of verification pixels correctly classified in their qualitative groups. The wavelengths were introduced according to their relevance.

1. The pixel values of the background (ceramic) were arbitrarily set to zero (black) in order to make the visualization easier. While all the three components were white at 1200 nm, starch was gray at 1550 nm. The white color observed at 1200 nm indicated that bran, starch, and gluten absorbed the near-infrared light very slightly. This wavelength can be assigned to the second overtone of C-H bonds. At 1550 nm, the mean gray levels of starch, gluten, and bran were 50, 85, and 90, respectively. These results showed that starch absorbed more than bran and gluten at 1550 nm. Law and Tkachuk²⁴ assigned the wavelength at 1550 nm to wheat starch, and more precisely to intramolecularly hydrogen bonded O-H groups. The direct observation of the image recorded at 1550 nm revealed the abilities of the system to identify chemical constituents. However, such a direct examination would generally not allow an identification of the chemical constituents given that in the near-infrared range the absorption bands are wide and overlap each other.

Discriminant Analyses. To select the more relevant images, stepwise discriminant analysis was applied to the vectors $G_{x,y}(l_1, l_2, \dots, l_{21})$. Calibration was performed on 120 pixels divided into four chemical groups, and the results were verified on the verification set of vectors (80 pixels). From the 21 initial wavelengths, 7 were selected in the following order: 950, 1500, 1450, 1000, 1700, 1600, and 900 nm. The percentages of well-classified pixels are given in Tables I and II. In these tables, the rows give the qualitative groups to be identified. The columns indicate the wavelength newly introduced at a given step of discriminant analysis. The data of these tables are the percentages of pixels correctly classified in their qualitative groups. For example, Table I shows that the wavelengths at 950, 1500, and 1450 nm allow a right classification of 86.7% for the pixels associated with gluten. In the verification set (Table II), the first three selected wavelengths were sufficient to correctly classify the pixels with a percentage equal to or greater than 90.0%. The introduction of the wavelengths at 950 and 1500 nm enabled a perfect identification of starch. These wavelengths might be assigned to the first (1500 nm) and second (950 nm) overtones of O-H groups.²⁵ All the verification pixels of bran were identified when the wavelength at 1450 nm was selected, perhaps due to the presence of an absorption band at this wavelength in the spectrum of wheat cellulose.¹⁸ The identification of gluten was partly due to the wavelengths at 1500 (N-H first overtone) and 1700 nm (C-H first overtone). At 1500 nm, both starch and gluten have absorption bands that overlap.

A canonical discriminant analysis was applied to the calibration and verification pixels using the seven previously

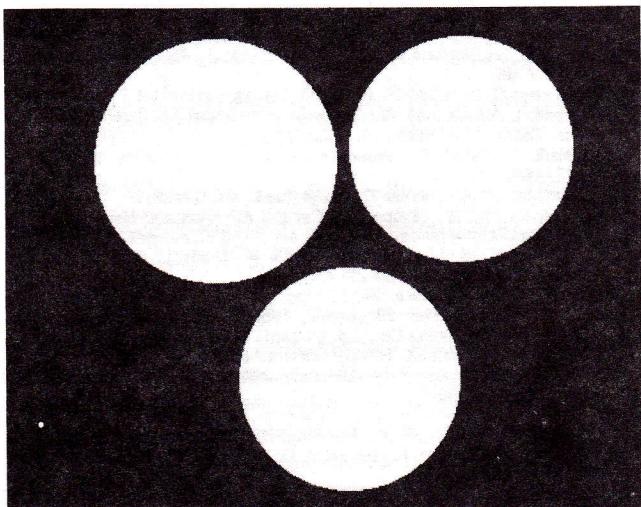


IMAGE RECORDED AT 1200 nm.

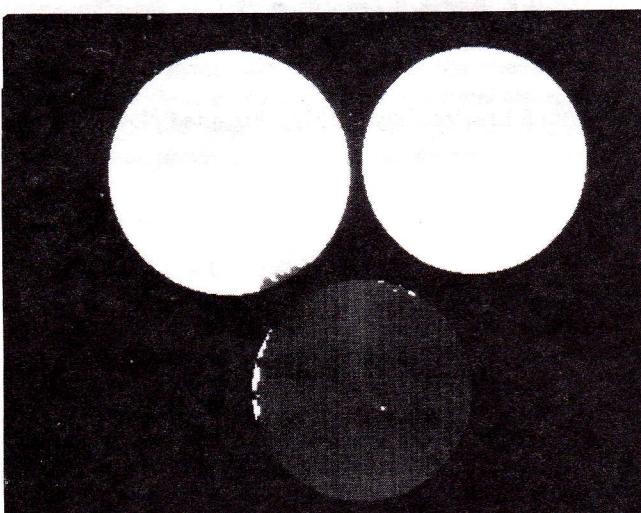
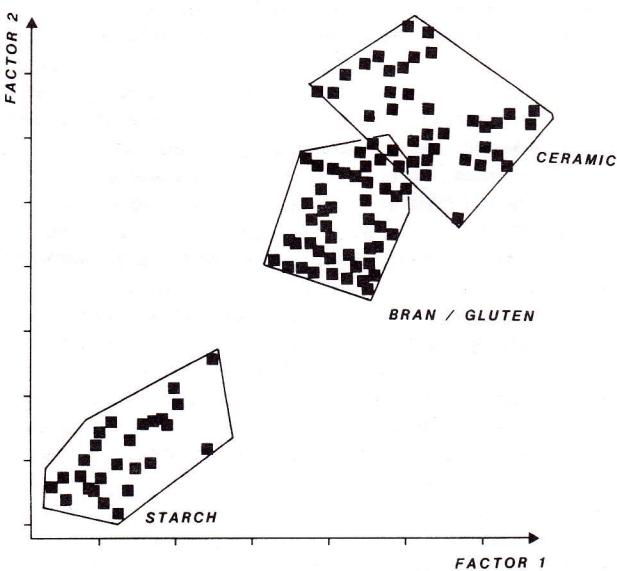
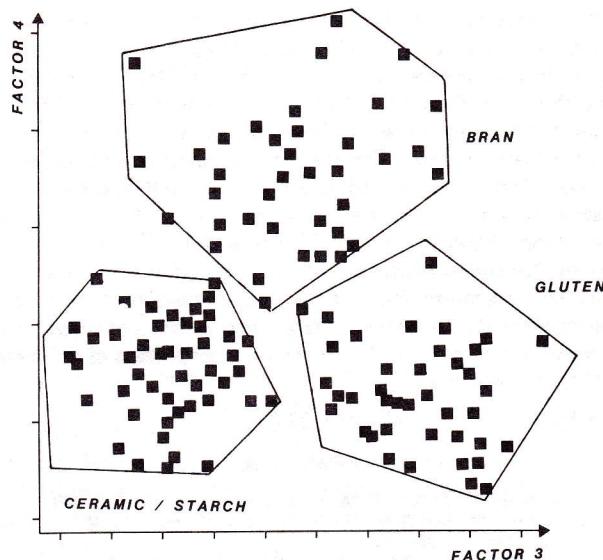
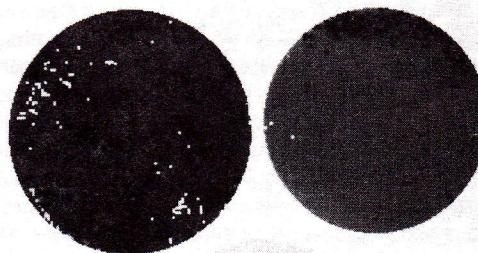


IMAGE RECORDED AT 1550 nm.

Figure 1. Near-infrared images of wheat constituents. Top left, bran; top right, gluten; bottom, starch.**Figure 2.** Canonical discriminant analysis. Similarity map defined by the factors 1 and 2. The calibration and verification pixels were represented by black squares.

selected wavelengths. The first four eigenvalues took into account 32.1%, 29.3%, 24.1% and 14.3%, respectively of the

**Figure 3.** Canonical discriminant analysis. Similarity map defined by the factors 3 and 4. The calibration and verification pixels were represented by black squares.**Figure 4.** Segmented image of wheat constituents. Top left, bran; top right, gluten; bottom, starch.

total variance. Figure 2 describes the similarity map defined by the first two factors. While the pixels of ceramic and starch were correctly identified on this map, those of bran and gluten were overlapped. The discrimination between the bran and gluten pixels could be observed on the map corresponding to the principal factors 3 and 4 (Figure 3). These results indicated that four discriminant factors were needed to identify the constituents involved in this study.

Segmented Image. The unknown pixels were classified into four chemical groups. The corresponding segmented image is given in Figure 4, where the four groups were clearly differentiated from each other and from the ceramic background. However, certain pixels of bran and gluten were misclassified. The percentages of well-classified pixels were 92 for bran, 95 for gluten, and 99 for starch. The results obtained for bran and gluten could be explained by the incident light that was not entirely homogeneous over the field of vision.

CONCLUSION

A video image analysis system operating in the near-infrared range was designed, and a method for mapping chemical constituents was proposed. The pixels of images recorded at several wavelengths were studied with a spectroscopic approach. The spectral differences between pixels were suffi-

ciently large to allow the identification of three natural components of wheat. In a first step, application of stepwise discriminant analysis made it possible to identify the most important wavelengths. Secondly, the images recorded at the chosen wavelengths were linearly combined by canonical discriminant analysis. The pixels were classified in qualitative groups, and a segmented image could be obtained. The segmentation of images was achieved from the spectral information related to chemical composition. Future work includes the characterization of mixtures of raw materials in the food, pharmaceutical, and chemical industries. In addition, the method will be tested on transverse sections of seeds such as wheat grains in order to measure the proportion of different tissue constituents.

Registry No. Starch, 9005-25-8.

REFERENCES

- (1) Goodman, D. E.; Rao, R. M. *J. Food Sci.* 1984, **49**, 648-649.
- (2) Neuman, M.; Sapirstein, H. D.; Shwedyk, E.; Bushuk, W. *J. Cereal Sci.* 1987, **6**, 125-132.
- (3) Sapirstein, H. D.; Neuman, M.; Wright, E. H.; Shwedyk, E.; Bushuk, W. *J. Cereal Sci.* 1987, **6**, 3-14.
- (4) Symons, S. J.; Fulcher, R. G. *J. Cereal Sci.* 1988, **8**, 211-218.
- (5) Neuman, M. R.; Sapirstein, H. D.; Shwedyk, E.; Bushuk, W. *J. Cereal Sci.* 1989, **10**, 175-182.
- (6) Neuman, M. R.; Sapirstein, H. D.; Shwedyk, E.; Bushuk, W. *J. Cereal Sci.* 1989, **10**, 183-188.
- (7) Newman, P. B. *Meat Sci.* 1984, **10**, 87-100.
- (8) Newman, P. B. *Meat Sci.* 1984, **10**, 161-166.
- (9) Newman, P. B. *Meat Sci.* 1987, **19**, 129-137.
- (10) Newman, P. B. *Meat Sci.* 1987, **19**, 139-150.
- (11) Aastrup, S.; Gibbons, G. C.; Munk, L. *Carlsberg Res. Commun.* 1981, **46**, 77-86.
- (12) Gibbons, G. C. *ASBC J.* 1981, **39**, 55-59.
- (13) Heyne, L.; Unklesbay, N.; Unklesbay, K.; Keller, J. *Can. Inst. Food Sci. Technol. J.* 1985, **18**, 168-173.
- (14) Geladi, P.; Wold, S.; Esbensen, K. *Anal. Chim. Acta* 1986, **191**, 473-480.
- (15) Guilloubez, S. *Agronomie Tropicale* 1985, **40**, 81-88.
- (16) Munk, L.; Feil, C.; Gibbons, G. *Cereals for Food and Beverages*, International Conference on Cereals and Food Beverages, Copenhagen, DK, August 13-17, 1979; Inglett, G. E., Munk, L., Eds.; Academic Press: London, 1980; pp 27-40.
- (17) Taylor, S. K.; McClure, W. F. *Proc. of the 2nd NIRS Conference*, Tsukuba, Japan, May 29-June 2, 1989; Iwamoto, M., Kawano, S., Eds.; Korin Publishing Co., Ltd. (Japan), 1990; pp 393-404.
- (18) Williams, P.; Norris, K. H. *Near infrared technology in the agricultural and food industries*; American Association of Cereal Chemists: St Paul, 1987; p 330.
- (19) Geladi, P.; Esbensen, K. *J. Chemometr.* 1989, **3**, 419-429.
- (20) Robert, P.; Devaux, M. F.; Bertrand, D. *Sci. des Aliments*, in press.
- (21) Downey, G.; Robert, P.; Bertrand, D.; Kelly, P. M. *Appl. Spectrosc.* 1990, **44**, 150-155.
- (22) Romeder, J. M. *Méthodes et programmes d'analyse discriminante*; Dunod: Paris, 1973.
- (23) Foucart, T. *Analyse factorielle - programmation sur micro ordinateurs*; Masson: Paris, 1982.
- (24) Law, D. P.; Tkachuk, R. *Cereal Chem.* 1977, **54**, 256-265.
- (25) Osborne, B. G.; Fearn, T. *Near infrared spectroscopy in food analysis*; Longman Scientific & Technical: Harlow Essex, England, 1986; pp 28-42.

RECEIVED for review July 8, 1991. Accepted November 15, 1991.