CIENCIA 7(2), 192-200, 1999
Maracaibo, Venezuela

Elimination of baseline deformation from infrared spectra prior to protids determination

Jean-Max Payet, Maya Cesari, Claude Rouch, Michel Pabion and Frédéric Cadet*

Laboratoire de Biochimie, Faculté des Sciences, Université de La Réunion
15 avenue René Cassin. BP 7151, 97715 Saint-Denis Messag. Cedex 9, La Réunion, France-Dom

Recibido: 25-09-98 Aceptado: 16-05-99

Abstract

We have recently showed by using Principal Component Analysis and Principal Component Regression on Mid-infrared spectra of biological samples that their protids concentrations could be determined with a good precision (CADET F. ***Spectrosc Lett*** 29(5):919-936, 1996). However the precision of the results can be improved; one way is to correct baseline deformations. Baseline is often assumed to be low-degreed polynomials.

In this paper, the terms with weak *n* values were subtracted in order to improve the precision of the quantitative determination of protids from the Legendre polynomial functions that were obtained from the decomposition of the spectra. The filtering parameters range from *n* =0 to *n*= 8, by increments of 2. The mean of the difference between reference and predicted values are 0 before correction and -0.01 after correction, while standard deviation values are 0.12 and 0.11 before and after correction for n=6 or n=8.

Key words: Baseline deformation; Legendre polynomials; mid-infrared; protids.

Eliminación de la deformación de la línea de base de espectros infrarrojos previa a la determinación de prótidos

Resumen

Recientemente hemos mostrado a través del Análisis de Componentes Principales en los espectros del infrarrojo medio de muestras biológicas que las concentraciones de prótidos pueden ser determinadas con buena precisión (CADET F. ***Spectrosc Lett*** 29(5):919-936, 1996). Sin embargo, la precisión de dichos resultados puede ser mejorada; una forma es corregir las deformaciones de la línea de base. Se asume usualmente que la línea de base puede ser representada por un polinomio de orden bajo.

En este trabajo, los términos con bajos valores de *n* fueron sustraídos de funciones polinomiales de Legendre las cuales fueron obtenidas de la descomposición de los espectros. Esto se hizo para mejorar la precisión en la determinación de prótidos. Los parámetros de filtro abarcan desde n=0 hasta n=8, a través de incrementos en 2. El valor promedio de las diferencias entre las referencias y los valores predichos fueron 0 antes de la corrección y -0,01 después de la

* To whom correspondence should be addressed. Fax: + 262 93 82 37. E-mail: cadet@univ-reunion.fr

correción, mientras que los valores de la desviación estándar fueron 0,12 y 0,11 antes y después de la corrección para n=6 y n=8.

Palabras clave: Deformación de línea base; espectros infrarrojos; polinomios de Legendre; prótidos.

Introduction

One of the major problems encountered in near and mid-infrared spectroscopy is baseline deformations (1-3). These deformations are dependant of the apparatus used; the variations in the spectra of the same sample measured successively several times would be due to the non-repeatability of the incident ray. Mid-infrared (MIR) spectroscopy is used for the determination of polypeptides and proteins secondary structures (4, 5). The most representative band occurs between 1700 and 1600 cm^{-1}. Two other bands appear between 1600 and 1200 cm^{-1}. These bands are designated as amide I, II and III respectively.

Near Infrared spectroscopy is one of the most widely used method for the quantitative analysis of major biochemical constituents (water, proteins, lipids and sugars) in food industry (6-8). However, few cases of applications of MIR spectroscopy to food analysis has been reported for cereals and other cereal products (9) or for the quantitative determination in milk and milk products (10).

In the sugar industry, in addition to the sucrose content, the measurement of alpha-amino acids shows to be useful. Despite the fact that α-NH_2 concentration is relatively weak in sugar cane juices, we have recently showed that (11) by using Principal Component Analysis (PCA) and Principal Component Regression (PCR) on the mid-infrared spectra, the concentrations of the α-NH_2 (amino-acids, peptids, proteins) could be predicted and hence, the protid concentration were determined with good precision in the biological samples. However the precision could be improved; one way of improving the precision of the results is to correct baseline deformations.

Baseline is often assumed to be low-degreed polynomials (12, 13). So, in the present work, from the Legendre polynomial functions that were obtained from the decomposition of the spectra (14), the terms with weak *n* values were subtracted in order to improve the precision of the quantitative determination of protids.

Material and Methods

Biological samples

After pulverization in a desintegrator, 100 g of sugar-cane was pressed for two and a half minutes at 250 bars in a hydraulic press in order to obtain the raw juices. The raw juice was filtered instantaneously, via a highly porous plastic filter. An Attenuated Total Reflectance cell was filled with this juice. The reference protid contents was measured colorimetrically, via the α-amino group measurement, according to the nynhydrine method from the International Commission for Uniform Methods of Sugar Analysis (15). The calibration set was constituted of 20 biological samples while the verification set was composed of 15 samples.

Mathematical treatment

Mathematical treatments were performed on a Compaq personal computer with software written in “C” language and developed in our laboratory. Multidimensional statistical analyses, such as principal component analyses (PCA), describe variation in multidimensional data by few synthetic variables. These synthetic variables are linear combination of all the original variables and have the advantage of having no correlation with each other. Simpler descriptions of data sets are thus obtained with minimal loss of information. These treatments were used for morphological analysis of spectra (16) and for graphical representation of spectra similarity (17).

PCA was applied to the spectra from 800 to 1250 cm^{-1} (with 235 data points used as principal variables). Spectra were centered prior to PCA according to:

$$X_{ij} = A_{ij} - A_j - A_i + A$$

where X_{ij} = centered data; A_{ij} = spectral data (log 1/R) of spectrum i and wavenumber j; A_j = mean value of spectral data at wavenumber j for every spectrum; A_i = mean value of spectral of spectrum i for every wavenumber; and A = average mean of all spectral data in the collection.

Principal component regression (PCR) was used to establish a prediction equation. PCR is basically a multilinear regression applied to scores assessed by PCA (18, 19). Interest in the introduction of scores according to their predictive ability had already been shown (20, 21).

Concentrations are predicted according to:

$$C_{n \cdot 1} = X_{n \cdot \mathrm{k}} \cdot V_{k \cdot p} \cdot R_{p \cdot 1}$$

where C is the column vector of predicted concentrations, X is the centered matrix of spectral data, V is the matrix of latent vectors of PCA, and R is the column vector of the regression coefficients of the prediction equations. n, k, p are respectively the number of samples; number of wavenumbers; number of significant principal components. The dot product $V \cdot R$ is a vector, the components of which may be interpreted in terms of absorption bands. Plotting the components against the corresponding wavenumbers gives a spectral pattern. Peaks correspond to absorption bands which are characteristic of the measured chemical constituents. Hollows indicate that when the concentration increases, the corresponding absorption bands will decrease (22-24).

Mid-FTIR spectra

Mid-Fourier Transform Infrared (Mid-FTIR) spectra were collected on a Michelson-100 Fourier Transform spectrophotometer. Attenuated Total Reflectance spectra were obtained with a Specac Overhead Attenuated Total Reflectance system. The crystal of the reflectance element is made from zinc selenide, a material that is quite inert to water. This quite rapidly cleaned between samples by being sprayed with water and then dried with filter paper.

The data were recorded from 1180 to 1900 cm^{-1} in 4 cm^{-1} increments at log(1/R), in which R is the ratio of the reflected intensity for the background to that of the sample. The combination of four scans resulted in an average spectrum.

Results and Discussion

The three major bands that are characteristic of proteins in MIR are located between 1200 and 1710 cm^{-1} (25): amide I, C=O stretch (1710-1580 cm^{-1}); amide II, N-H bending (1580-1500 cm^{-1}) and amide III, C-N stretch (1300-1200 cm^{-1}). The Mid-FTIR spectrum of a sample of raw sugar cane juice is shown in Figure 1. The absorption band of water (1500-1700 cm^{-1}) is stronger than that of the protids between 1500 and 1180 cm^{-1}; this band is superimposed to the bands characteristic of amides I and II. In this region protids are characterised by two peak centered at 1650 cm^{-1} and 1540 cm^{-1} and by a hollow at 1600 cm^{-1} (26). Hence, in order to extract the spectral information that correspond to protids, the collected spectra of all the biological samples from the calibration set were entered into a principal component analysis (PCA). The collection of spectra is modelised by PCA into a sum of characteristic signals which form a spectral

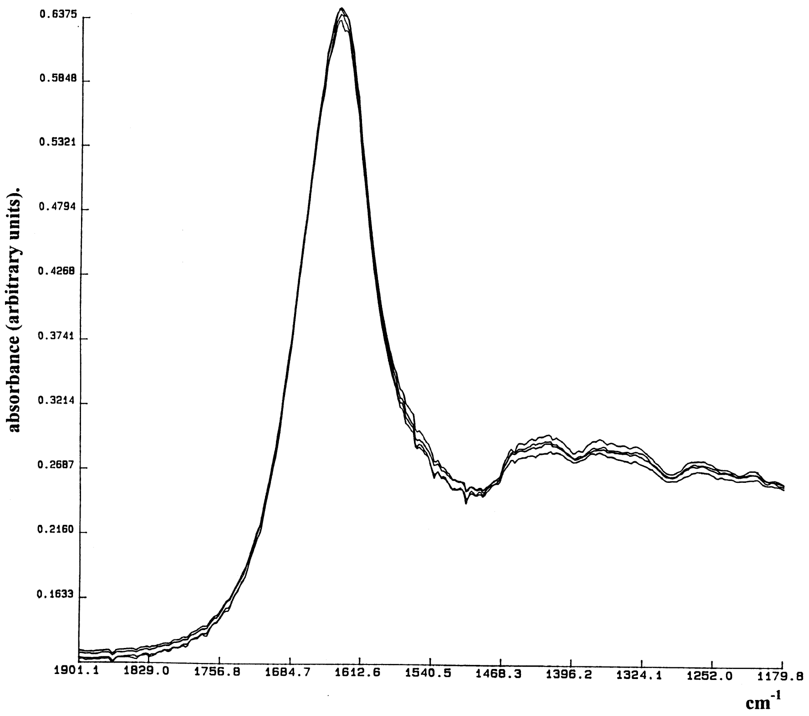

Figure 1. Mid-FTIR of a Biological sample in the 1180-1900 cm^{-1} range before correction.

pattern (16, 17). This spectral representation of the principal component of PCA features characteristic absorption bands of biochemical constituants in a sample. This procedure was applied before and after the determination of the baseline deformation using decomposition Legendre polynomials.

Baseline correction was hence perfomed by subtracting low-degree terms from the polynomial functions obtained by the decomposition of spectra. The visual comparison of two spectra between α_1 cm^{-1} and α_2 cm^{-1} mainly consist in comparing the areas beneath the spectra. A scalar product can hence be associated with the integral between α_1, and α_2. Lipkus (14) has introduced the use of an orthonormed polynomial Legendre system whose scalar product is given by:

$$\langle f,g\rangle \quad T\cdot {}_{\alpha 1}^{\alpha 2} f\cdot g \qquad [1]$$

If the orthonormed system (α_1, α_2) is considered (P_i, $i \in N$), a spectrum $S(\alpha)$ can be decomposed into:

$$S(\alpha) \quad \overset{\infty}{\underset{i\ \ 0}{}}\ \beta_i\cdot P_i \qquad [2]$$

With the elimination of the first terms, it becomes:

$$S_0(\alpha) \quad S(\alpha) \quad \overset{n}{\underset{i\ \ 0}{}}\ \beta_i\cdot P_i\,[3]$$

where $S_0(\alpha)$ represents the corrected spectrum and *n* an integer.

Spectra can be expanded in terms of a set of orthonormal polynomials derived from the Legendre polynomials, and leading terms of the expansion (which contains most of the baseline variation), can therefore be removed.

The calibration set is constituted of a collection of 20 biological samples. The α-NH_2 content of the calibration set ranged from 0.34% to 1.12% (g/100 mL) with a mean and standard deviation (SD) values of 0.58% and 0.22% respectively. The verification set is constituted of 15 samples. The α-NH_2 content ranged from 0.33% to 1.05% (g/100 mL) with a mean and standard deviation values of 0.58 and 0.21%.

The morphological consequences of applying Legendre correction to the spectra for the following values of the filtering parameters *n* from 0 to 8, by increments of 2, are illustrated in Figures 2 to 6.

Principal Component Regression on the calibration set scores as assessed by PCA were carried out in order to establish prediction equations that linked spectral data to protids (α-NH_2) content. We have previously showed that when the 10 (as assessed by PCA) most correlated axes to protids chemical values are used, the correlation coefficient is above 0.99 (11). Only the first 10 axes were used for the prediction equation.

Predicted protids concentration for n=0 to 8 are given in Table 1. The mean of the difference between reference and predicted values are 0 before correction and -0.01 after correction, while standard deviation values are 0.12 and 0.11 before and after correction with n=6 or n=8 (Table 1).

Conclusion

Baseline correction was perfomed by subtracting low-degree terms from the polynomial functions obtained by the Legendre decomposition of spectra. This procedure slightly improves the values of protids content that were determined before. When the precision of the quantitative measures is important, this procedure can show to be interesting.

Acknowledgements

This work was supported by a grant from the Ministère de la Recherche et de la Technologie and the Conseil Régional de la Réunion.

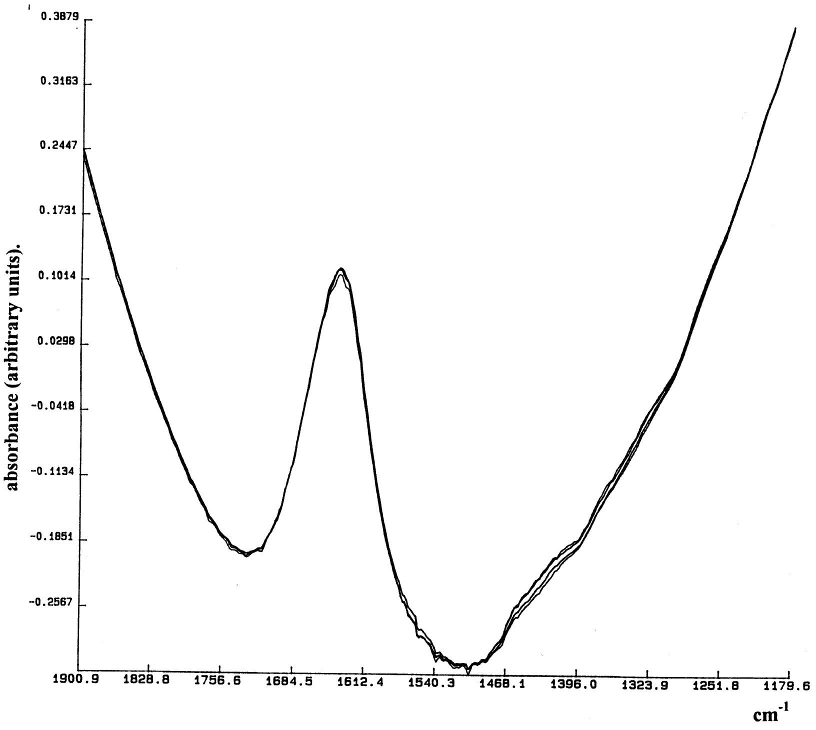

Figure 2. Mid-FTIR (1180-1900 cm^{-1}) spectra of protids after Legendre baseline correction, n=0.

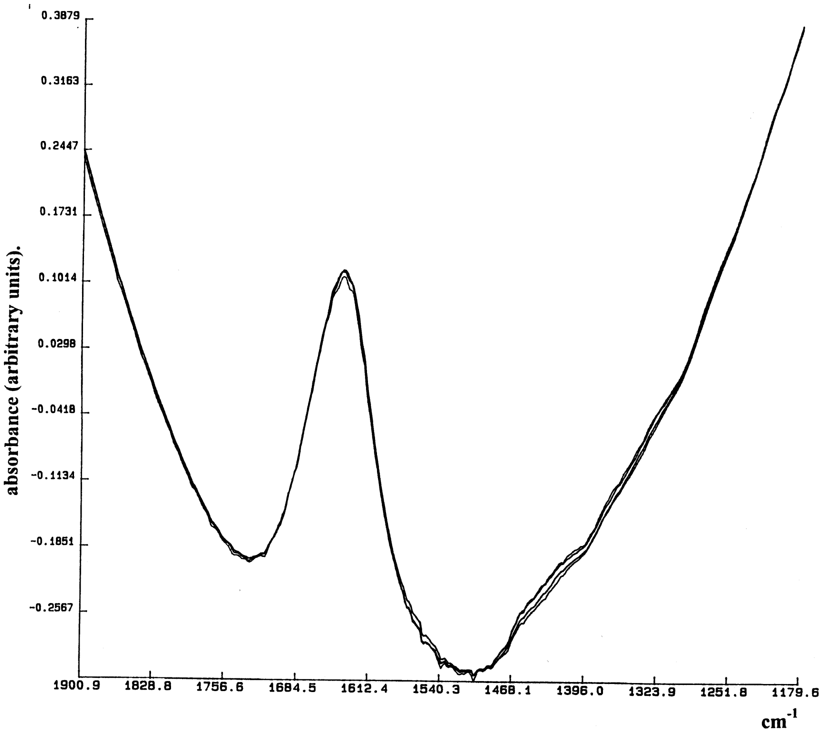

Figure 3. Mid-FTIR (1180-1900 cm^{-1}) spectra of protids after Legendre baseline correction, n=2.

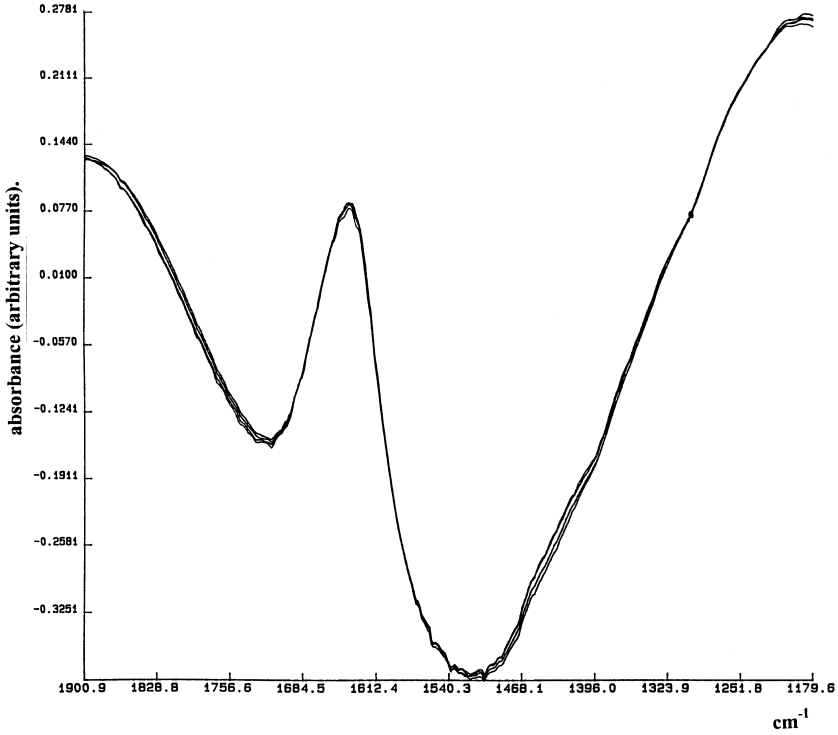

Figure 4. Mid-FTIR (1180-1900 cm^{-1}) spectra of protids after Legendre baseline correction, n=4.

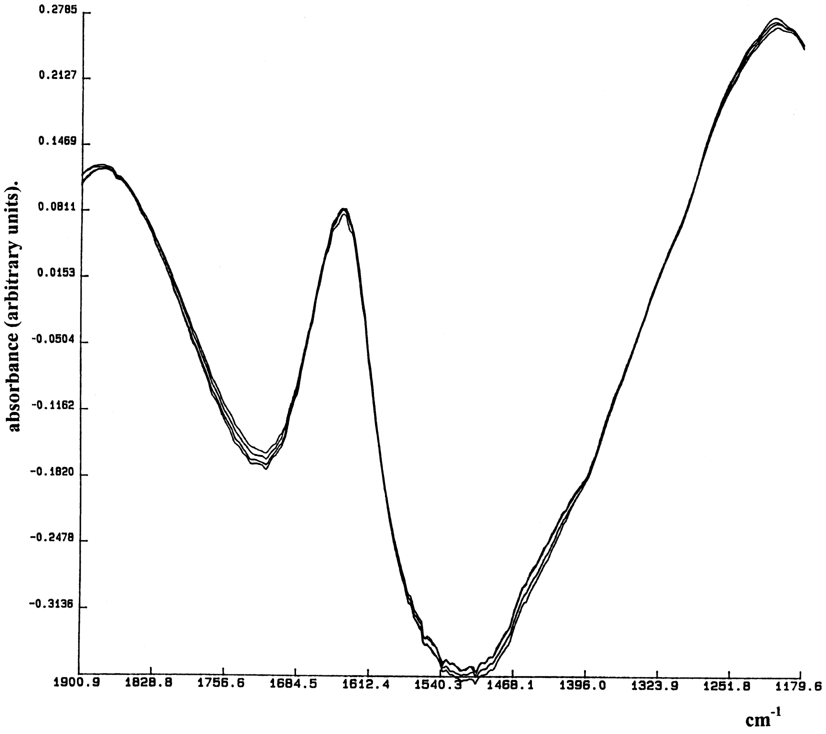

Figure 5. Mid-FTIR (1180-1900 cm^{-1}) spectra of protids after Legendre baseline correction, n=6.

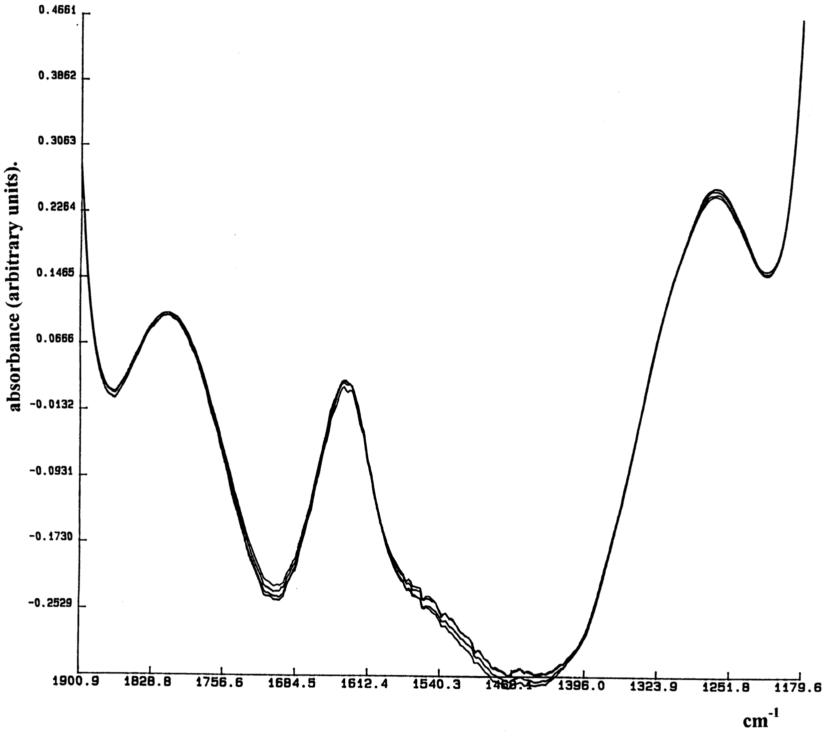

Figure 6. Mid-FTIR (1180-1900 cm^{-1}) spectra of protids after Legendre baseline correction, n=8.

Table 1

Difference between the reference and the predicted α-NH_2 values, before and after decomposition by Legendre polynomials for *n*=0 to 8

Protids														
	Initial spectra			Corrected spectra										
				n=0		n=2		n=4		n=6		n=8		
Sample number	Reference proteins content*	Predicted*	Deviation	Predicted*	Deviation	Predicted*	Deviation	Predicted*	Deviation	Predicted*	Deviation	Predicted*	Deviation	
1	0.820	0.797	-0.023	0.758	-0.062	0.790	-0.030	0.808	-0.012	0.797	-0.023	0.808	-0.012	
2	0.470	0.414	-0.056	0.404	-0.066	0.455	-0.015	0.416	-0.054	0.410	-0.060	0.403	-0.067	
3	0.370	0.608	0.238	0.606	0.236	0.596	0.226	0.573	0.203	0.595	0.225	0.579	0.209	
4	1.050	0.802	-0.248	0.801	-0.249	0.803	-0.247	0.821	-0.229	0.849	-0.201	0.843	-0.207	
5	0.720	0.660	-0.060	0.629	-0.091	0.634	-0.086	0.646	-0.074	0.634	-0.086	0.650	-0.070	
6	0.400	0.406	0.006	0.406	0.006	0.460	0.060	0.436	0.036	0.465	0.065	0.447	0.047	
7	0.800	0.703	-0.097	0.706	-0.094	0.729	-0.071	0.714	-0.086	0.739	-0.061	0.734	-0.066	
8	0.330	0.331	0.001	0.304	-0.026	0.354	0.024	0.328	-0.002	0.319	-0.011	0.312	-0.018	
9	0.740	0.589	-0.151	0.588	-0.152	0.606	-0.134	0.574	-0.166	0.594	-0.146	0.578	-0.162	
10	0.420	0.438	0.018	0.441	0.021	0.480	0.060	0.427	0.007	0.461	0.041	0.435	0.015	
11	0.590	0.577	-0.013	0.594	0.004	0.622	0.032	0.547	-0.043	0.528	-0.062	0.536	-0.054	
12	0.510	0.630	0.120	0.618	0.108	0.599	0.089	0.639	0.129	0.619	0.109	0.625	0.115	
13	0.560	0.667	0.107	0.667	0.107	0.640	0.080	0.637	0.077	0.611	0.051	0.607	0.047	
14	0.560	0.704	0.144	0.715	0.155	0.707	0.147	0.691	0.131	0.688	0.128	0.672	0.112	
15	0.380	0.345	-0.035	0.345	-0.035	0.316	-0.064	0.315	-0.065	0.331	-0.049	0.336	-0.044	
mean			**-0.003**		**-0.009**		**0.005**		**-0.010**		**-0.005**		**-0.010**	
standard deviation			**0.121**		**0.124**		**0.117**		**0.114**		**0.110**		**0.107**	

* g / 100 mL

References

1. POWELL J.R, WASACZ F.M., JACOBSEN R.S. ***Appl Spectrosc*** 40(3): 339-345, 1984.
2. HRUSCHKA W.R. ***Near Infrared Technology in the Agricultural and Food Industry,*** Williams P.C. and Norris K.H. (Eds), AACC, Saint Paul, MN (USA), pp. 35-46, 1987.
3. CADET F., OFFMANN B. ***Spectrosc Lett*** 29(4): 591-607, 1996.
4. ELLIOTT A., AMBROSE E.J. ***Nature*** 4206: 921-922, 1950.
5. MIYAZAWA T., BLOUT E.R. ***J Am Chem Soc*** 83: 712-719, 1960.
6. FULLER M.P., GRIFFITHS P.R. ***Anal Chem*** 50(13): 1906-1910, 1978.
7. OSBORNE B.G., FEARN T. ***Near Infrared Spectroscopy in Food Analysis***. Longman Scientific & Technical, Wiley & Sons, NewYork (USA), 1986.
8. WILLIAMS P., NORRIS K. ***Near Infrared Technology in the Agricultural and Food Industry.*** American Association of Cereal Chemists (Ed), Saint Paul, MN, USA, p. 330, 1987.
9. RENARD C., ROBERT P., BERTRAND D., DEVAUX M.F., ABECASSIS J. ***Cereal Chem*** 64(3): 177-181, 1987.
10. CROCOMBE R.A., OLSON N.L., HILLS S.L. ***American Society for Testing and Materials*** 95-130, 1987.
11. CADET F. ***Spectrosc Lett*** 29(5): 919-936, 1996.
12. ABBINK SPAINK H., LUB T.T., OTJES R.P., SMIT H.C. ***Anal Chim Acta*** 183: 141-154, 1986.
13. LIU J., KOENIG J.L. ***Appl Spectrosc*** 41: 447-449, 1987.
14. LIPKUS A.H. ***Appl Spectrosc*** 42(3): 395-400, 1988.
15. SCHNEIDER F. ***Sugar Analysis*** I.C.U.M.S.A. (Ed), Dublin, Ireland, 1985
16. LE NOUVEL J. Etude d'une famille de courbes par des méthodes d'analyse des données. Application à l'analyse morphologique de courbes provenant de données médicales (Thèse de 3ème cycle), Université de Rennes I (France), 1981.
17. DEVAUX M.F., BERTRAND D., ROBERT P.,QANNARI M. ***Appl Spectrosc*** 42(6): 1015-1019, 1988.
18. LEFEBVRE J. ***Introduction aux Analyses Statistiques Multidimensionnelles***, Masson, Paris, 3rd ed., 137-148, 1983.
19. LEBART L., MORINEAU A., TABARD N. **Techniques de la Description Stastitiques** Dunod, Paris, 7-46, 1977.
20. DAGNELIE P. ***Analyse Statistique à Plusieurs Variables,*** Les Presses Agronomiques de Gembloux, Belgium, 185-190, 1975.
21. BERTRAND D., LILA M., FURTOSS V., ROBERT P., DOWNEY G. ***J Sci Food Agric*** 41: 299-307, 1987.
22. BERTRAND D., ROBERT P., DEVAUX M.F., ABECASSIS J. ***Analytical Applications of Spectroscopy,*** C.S. Creaser and A.M.C Davies (Eds), Royal Society of Chemistry (London), 450-455, 1988.
23. CADET F., BERTRAND D., ROBERT P., MAILLOT J., DIEUDONNÉ J., ROUCH C. ***Appli Spectros*** 45(2): 166-172, 1991.
24. CADET F., WONG PIN F., ROUCH C., ROBERT C., BARET, P. ***Biochim Biophys Acta*** 1246: 142-150, 1995.
25. SUSI H., TIMASHEFF S.N, STEVENS L. ***J Biol Chem*** 242: 5460-5466, 1967.
26. YANG W.J., GRIFFITHS P.R., BYLER D.M., SUSI H. ***Appl Spectrosc*** 39(2): 282-287, 1985.