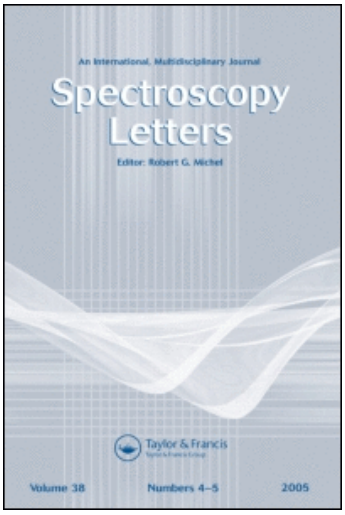

Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Extraction of Characteristic Bands of Sugars by Multidimensional Analysis of Their Infrared Spectra.

FrÉDÉRic CADET[a]; Bernard OFFMANN[a]
[a] Laboratoire de Biochimie, Faculté des Sciences, Université de la Réunion, Réunion, France-Dom

To cite this Article FrÉDÉRic CADET and Bernard OFFMANN(1996) 'Extraction of Characteristic Bands of Sugars by Multidimensional Analysis of Their Infrared Spectra.', Spectroscopy Letters, 29: 3, 523 — 536
To link to this Article: DOI: 10.1080/00387019608006668
URL: http://dx.doi.org/10.1080/00387019608006668

SPECTROSCOPY LETTERS, 29(3), 523-536 (1996)

EXTRACTION OF CHARACTERISTIC BANDS OF SUGARS BY MULTIDIMENSIONAL ANALYSIS OF THEIR INFRARED SPECTRA.

Key words : Mid-FTIR, PCR, ATR, specific sugar IR bands.

Frédéric CADET* and Bernard OFFMANN

Laboratoire de Biochimie, Faculté des Sciences. Université de la Réunion, 15 avenue René Cassin. BP 7151, 97715 Saint-Denis Messag Cedex 9, Réunion, France-Dom. Fax : Intl (262) 93 81 66.

ABSTRACT.

Collected Mid-IR Attenuated Total Reflectance (ATR) spectra of various sugars were assessed by multidimensional statistical analysis. Through Principal Component Analysis (PCA) of collected spectra of various pure 10% sugar solutions and from the spectroscopic representation of the factorial axes, characteristic frequencies of monosaccharides and oligosaccharides were directly and automatically obtained within a few seconds. Monosugars are characterised by a hollow at 998 cm^{-1} and by a single unique major peak (1049 cm^{-1}) in the 1075 - 1030 cm^{-1} region while oligosaccharides showed three characteristic bands in the same region, the major peak is shifted to 1033 cm^{-1}. Owing to the glycosidic link vibrational motion, oligosaccharides are also characterised by a band at 998 cm^{-1}. Spectroscopic representation of the axes issued of data from both sugar families (mixture of mono and oligosaccharides) is an average of the two individual spectral patterns. Predictive measures of concentrations of sugar solutions were performed by principal component

* to whom correspondence should be addressed.

regression (PCR) of the factorial coordinates and prediction equations were obtained. The predicted concentrations of standard 10% pure sugar solutions averaged 10.069% and 9.909% for monosugars and oligosugars respectively and a concentration of 10.015% from the mixed set of data was obtained. The ability of these factorial coordinates to predict quantitative variable are good with correlation coefficients ranging from 97.4% to 99.9%.

INTRODUCTION.

Infrared spectroscopy is a very useful technique for the identification and characterization of molecules. Furthermore Near Infrared Reflectance is widely used as a tool for quantitative analysis of many food components (Norris., 1978, Osborne *et al.*, 1982). The recent advent of Mid Infrared Fourier transform spectroscopy (F.t.i.r) has provided valuable information through data processing of digitized spectra (Koenig.; 1981) and has allowed the study of biomolecules in aqueous solutions.

However complex spectra make interpretation difficult. Through multidimensional statistical analysis these problems may be overcome (Bertrand *et al.*, 1988, Devaux *et al.*, 1988, Cadet *et al.*, 1991).

Analysis of sugar solutions by Infrared spectroscopy have been extensively reported (Kuhn., 1950, Barker *et al.*, 1956, Back *et al.*, 1983). Some workers have tried to find characteristic wavelengths to predict sugar content (Henry., 1985, Mills et al., 1986). This method based on isobestic points determination, needs a careful and lengthy visual examination of a series of sugar spectra.

Recently we have reported to use a combination of infrared spectroscopy and multidimensional statistical analysis for the analysis of sucrose solutions (Cadet *et al.*, 1991) and for enzymatic studies (Cadet *et al.*, 1995).

The present study was undertaken in a scope of developping a quick and reliable method for the characterization and quantitative analysis of sugar families by using Mid-Infrared Attenuated Total Reflectance spectra potentialities combined with computer based multidimensional statistical analysis. No visual examinations of individual spectra is hence required by this method.

MATERIALS AND METHOD.

All D-oses used were of highest grade available and were purchased from **Sigma**.

Mid-Infrared Attenuated Total Reflectance 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 ATR system. The crystal of the reflectance element is made from zinc selenide, a material that is quite inert to water; it is quite rapidly cleaned between samples by being sprayed with water and then dried with filter paper.
The data were recorded from 800 to 1250 cm-1 in 4-cm increments at log(1/R), in which R is the ratio of the reflected intensity for the background to that of the sample. Although the ATR experiment does involve the reflection of the radiation within a crystal, the interaction of the radiation with the sample is the transmittance of radiation through the sample; this depth of penetration is wavelength dependent, but it is passing through a finite layer of the sample. For this reason, plots can read according to absorbance (or transmittance). The combination of four scans resulted in an average spectrum. The intensity the spectra was low; the highest peaks had log(1/R) values less than 0.6 on baseline spectra.

Mathematical treatments.

Mathematical treatments were performed on a **Compaq** personal computer with software written in "C" language and developped 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 (le Nouvel, 1981) and for graphical representation of spectra similarity (Devaux *et al.*, 1988).

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 wavelength j; A_j = mean value of spectral data at wavelength j for

every spectrum; A_i = mean value of spectral of spectrum i for every wavelength; 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 (Lefebvre., 1983, Lebart *et al.*, 1977). Interest in the introduction of scores according to their predictive ability had already been shown (Dagnelie, 1975, Bertrand *et al.*, 1987).

Concentrations are predicted according to :

$$\mathbf{C}_{n.l} = \mathbf{X}_{n.k} \cdot \mathbf{V}_{k.p} \cdot \mathbf{R}_{p.l}$$

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; the number of wavelengths; the number of significant principal components. The dot product **V . R** is a vector, the components of which may be interpreted in terms of absorption bands. Plotting the components against the corresponding wavelengths 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 (Bertrand *et al.*, 1988).

Spectra of pure 10% (g/100 ml) sugar solutions were grouped in three calibration sets according to the class of sugar they belong ; the monosaccharide family, the oligosaccharide family and a third calibration set (M/OCALIB) grouped data from these two families (table 1).
Two verification sets were constituted from spectra of pure 10% (g/100 ml) sugar solutions according to the class of sugars defined above (table 1); the monosaccharide and the oligosaccharide verification sets (table 1).

Two verification sets were constituted from spectra of pure 10% (g/100 ml) sugar solutions according to the class of sugars defined above (table 1); the monosaccharide and the oligosaccharide verification sets (table 1).

RESULTS AND DISCUSSION.

Mid-IR ATR spectra of various pure sugar solutions at a concentration of 10% (g/100 ml) were collected and are shown in figures 1a to 1e. For

Table 1. Sugars constituting the calibration sets and verification sets. M/OCALIB was a mixture of the two calibration families.

Calibration sets		Verification sets	
Monosaccharide family	Oligosaccharide family	Monosaccharide family	Oligosaccharide family
mannose	maltose	xylitol	lactose
xylose	melilobiose	galactose	cellobiose
sorbitol	sucrose	fructose	
adonitol	raffinose	mannitol	
arabinose	water		
glucose			
sorbose			
water			

each of the three calibration sets, the collected spectra of the various sugars were introduced in a principal component analysis (PCA). A collection of spectra forms a hypervolume in an *n*-dimensional space. The description of this volume using the spectral data may not be the most simple mathematical way to characterise the set of spectra. As infrared data are highly correlated, there is a large redundancy of information which can be eliminated by the creation of a more suitable system of axes. PCA allows the creation of such a system of axes not correlated with each other, called 'principal components', which are linear combination of the original ones. The relationship between chemical and spectral data is established by multilinear regression between chemical values, sugar concentration in the present case, as the predictive variable and spectral data, principal component, as the predictive one.

Principal component regression allows the establishment of prediction equations. In our case as the number of samples is relatively limited three prediction equations the were achieved using the highest number of axes so as to obtain the best prediction equations. Spectral patterns were assessed in order to identify the spectral regions playing a decisive part in the predictions. The patterns were linear combinations of eigenvectors extracted from PCA. These vectors give a spectroscopic representation of the factorial axes and are powerful tools for interpretation of IR spectra. It is very difficult to find characteristic wavelengths from MIR spectra. In a family of spectra, the spectral pattern corresponding to the vectors allows direct identification of characteristic absorption bands that are the most representative of the family.

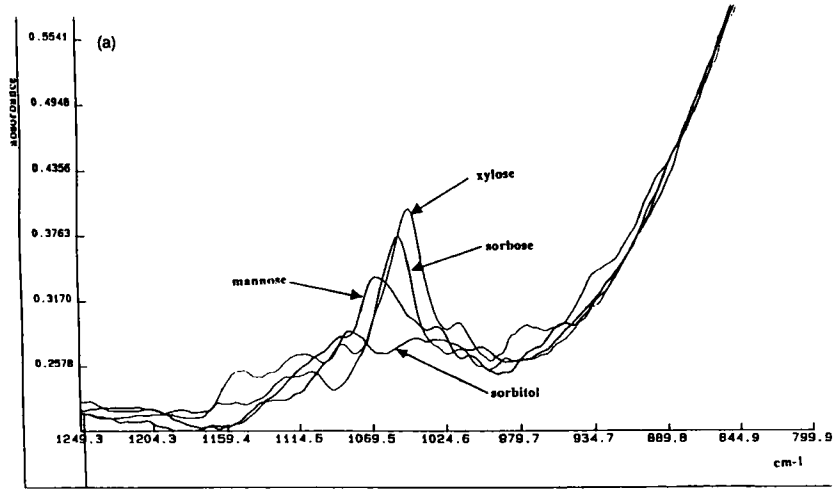

Figure 1a. Mid-Infrared ATR spectra of pure 10% (g/100 ml) solutions of mannose, xylose, sorbitol and sorbose.

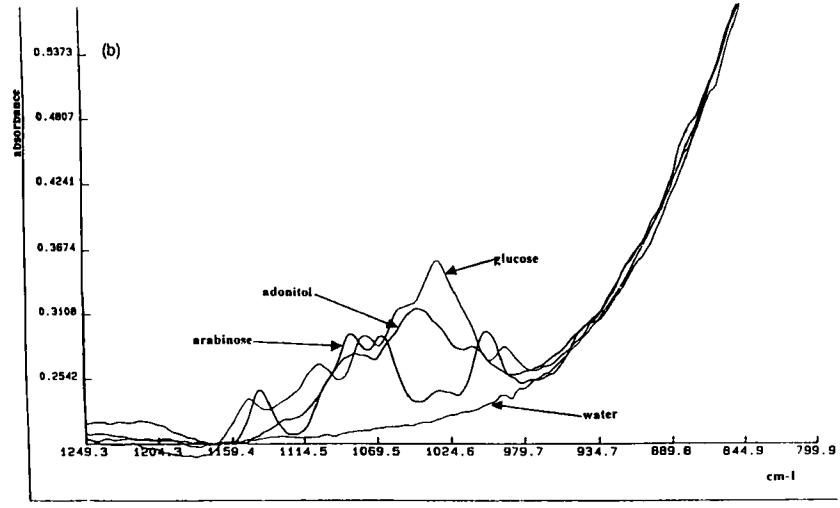

Figure 1b. Mid-Infrared ATR spectra of pure 10% (g/100 ml) solutions of adonitol, glucose, arabinose and of pure water.

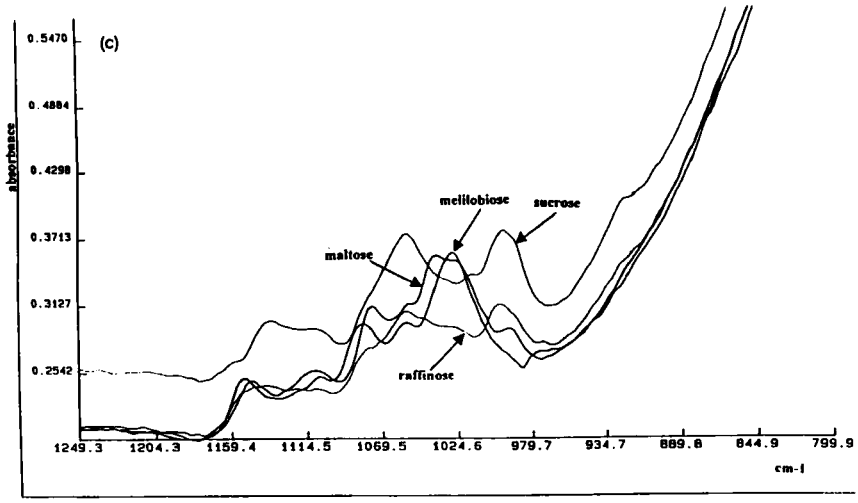

Figure 1c. Mid-Infrared ATR spectra of pure 10% (g/100 ml) solutions of maltose, raffinose, melilobiose and sucrose.

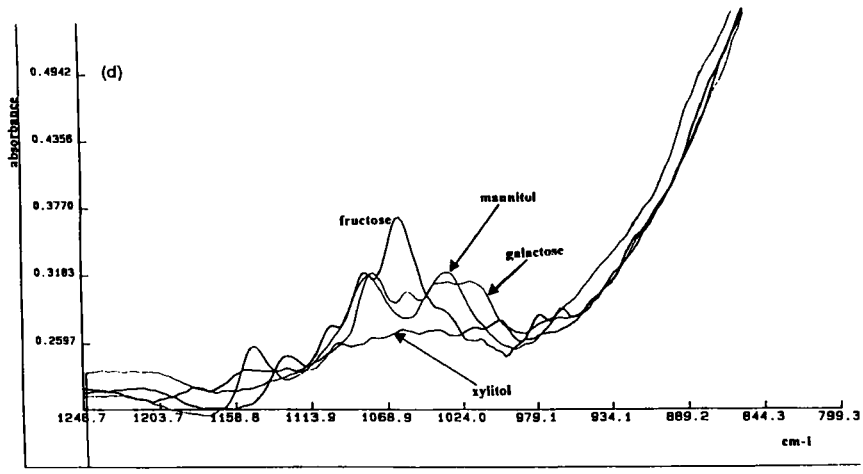

Figure 1d. Mid-Infrared ATR spectra of pure 10% (g/100 ml) solutions of mannitol, fructose, galactose and xylitol.

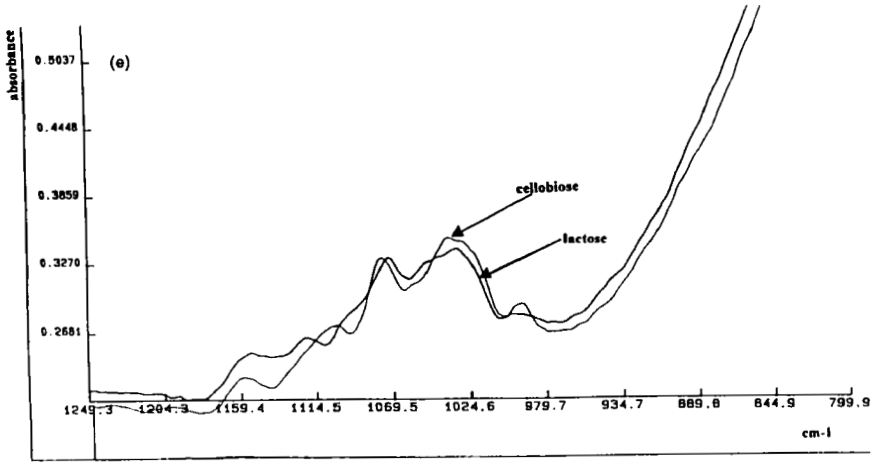

Figure 1e. Mid-Infrared ATR spectra of pure 10% (g/100 ml) solutions of lactose and cellobiose.

Characterisation of sugar families.

Spectral patterns of the principal components are shown in figure 2 (axis 2 for monosaccharides, axis 1 for oligosaccarides and axis 1 for M/OCALIB with correlation coefficients of 64%, 93% and 61% respectively). The monosaccharide and oligosaccharide families have distinct spectral patterns.

Oligosaccharides are characterized by a peak at 998 cm^{-1} owing to the presence of glycosidic links. This band was reported to disappear during enzymatic hydrolysis of sucrose and was assigned to the vibrational motion of the disaccharide link (Cadet *et al.*, 1995). On the contrary a hollow is observed with the monosugars family at this frequency. In the 1075-1030 cm^{-1} region only a major peak (1049 cm^{-1}) is observed with the monosaccharide family while three distinct peaks are obtained with the oligosaccharide family spectral pattern with the major peak being centered at 1033 cm^{-1}. This shift in the frequencies of the major peaks can be explained by the difference in the chemical environment surrounding the atoms concerned with this vibrational band. Similar shifts are observed with the minor peaks that are common to both families ; the 1113-1108 cm^{-1} bands and the C-O and C-C stretching bands at 1152-1145 cm^{-1} (Back et al., 1984).

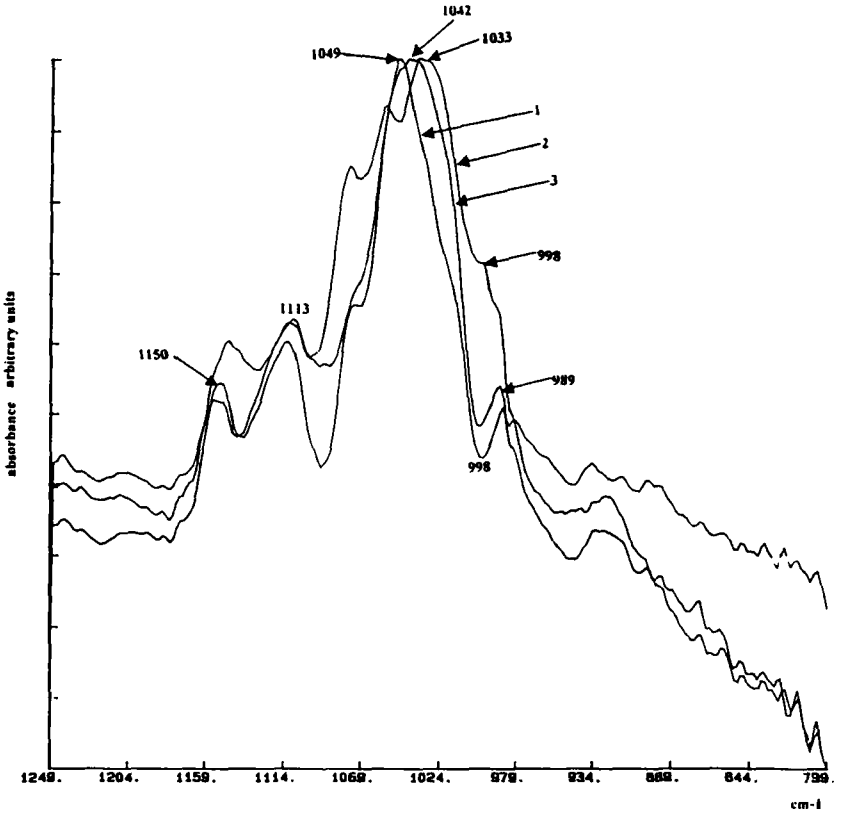

Figure 2. Spectroscopic representation of principal component extracted from the monosaccharides (1), oligosaccharides(2) and the monosaccharide/oligosaccharide (M/OCALIB) calibration sets (3).

As for the M/OCALIB calibration set (figure 2), the spectral representation of the principal component is an average of the two spectral patterns obtained with the two families separately ; the major peaks were averaged to 1042 cm^{-1}. The 1150 cm^{-1}, 1113 cm^{-1}, 1074 cm^{-1} and 989 cm^{-1} bands that are common to both families are well featured in the M/OCALIB spectral pattern. However the band associated with the glycosidic link (998 cm^{-1}) is completely masked by the hollow at 998 cm^{-1} observed in the spectral pattern of the monosaccharide family. Hence characterization of oligosaccharides by reflectance measurements

Table 2. Influence of the number of regression terms in PCA prediction equation on the predictive aptitude.

	monosaccharide family		oligosaccharide family		M/OCALIB calibration set	
Number of terms	Latest introduced PCA variable	correlation coefficient	Latest introduced PCA variable	correlation coefficient	Latest introduced PCA variable	correlation coefficient
1	2	0,6404	1	0,9359	1	0,6141
2	3	0,8883	4	0,9835	4	0,8587
3	6	0,9643	3	0,9995	7	0,9103
4	1	0,9894	2	1	11	0,9533
5	7	0,9997			8	0,9742

of the glycosidic links in a mixture of oligo and monosaccharides would greatly depend on the relative importance of the oligosaccharide in the sugar solution.

Predictions by using PCR on spectra.

Correlation between the factorial coordinates and the characteristic parameters of samples have been calculated (table 2) ; they give a direct estimation of the factorial coordinate aptitude to predict a quantitative variable. Regressions on these factorial coordinates (PCR) established prediction equations. For the three calibration sets, correlation coefficients were better than 95% when four or more terms are introduced in PCA.

Three prediction equations were established by PCR with the three calibration sets. Spectral data of the verification sets were entered in these three equations in order to predict sugar content. Tables 3a, 3b and 3c show results obtained with reflectance measurements and prediction equations. The overall results are significantly better than those obtained by Henry *et al.*, (1985) and Mills *et al.*, (1986) with the isobestic points method. Hence no visual examination of each spectrum was necessary in order to find characteristic wavelengths that could be used for predicting sugar contents. Mean values obtained are close to the reference 10% value (10.069% and 10.015% for monosaccharides and M/OCALIB respectively) except for the oligosaccharide verification set where only two individuals were studied (mean = 9.741%).

Table 3a. Predicted concentrations of 10 % (g/100 ml) monosugar solutions. Prediction equation was established by PCR on the spectra of the monosaccharide calibration set.

Sugar	reference (%)	predicted (%)
xylitol	10	8,239
galactose	10	9,909
fructose	10	11,070
mannitol	10	11,058
	mean	*10,069*
	SD	*1,336*
	mean error	6,91E-02

Table 3b. Predicted concentrations of 10 % (g/100 ml) oligosaccharide solutions. Prediction equation was established by PCR on the spectra of the oligosaccharide calibration set.

Sugar	reference (%)	predicted (%)
lactose	10	9,078
cellobiose	10	10,404
	mean	*9,741*
	SD	*0,938*
	mean error	-2,59E-01

Table 3c. Predicted concentrations of 10 % (g/100 ml) sugar solutions. Prediction equation was established by PCR on the spectra of the M/OCALIB calibration set.

Sugar	reference (%)	predicted (%)
xylitol	10	7,781
galactose	10	9,307
fructose	10	10,932
mannitol	10	9,199
lactose	10	10,211
cellobiose	10	12,660
	mean	*10,015*
	SD	*1,675*
	mean error	1,50E-02

The worst predicted values are with xylitol (tables 3a and 3c) and with cellobiose (table 3c). Predictions based on the prediction equations established from the corresponding sugar family calibration sets showed better results with galactose (9.909%) and cellobiose (10.404%) than with the equation established from the combined M/OCALIB set. However better predictions are obtained with the latter calibration set for lactose (10.211%), fructose (10.932%) and mannitol (9.199%).

Conclusion.

The methods and software used allows the combination of the statistical and spectroscopic approaches for the study of collected spectra. Muldimensional statistical analysis points out in few seconds dominant phenomenons and their relative importance from a set data. Sugar classes can be characterized by this approach. Monosaccharides family are essentially characterised by a hollow at 998 cm^{-1} and a major peak at 1049 cm^{-1}. There is a shift in the bands that feature oligosaccharides when compared to monosaccharides. The glycosidic vibrational motion band is also well featured (998 cm^{-1}) as reported in a previous paper. It appears that prediction equations established by the method described in this paper offers a good alternative to the isobestic ponts method for quantitative determination of sugar solutions as successfully described for sucrose content in cane juice (Cadet *et al.*, 1991). Furthermore the overwhole procedure need no more than a few seconds for PCR and prediction and need no visual examination of whole sets of spectra.

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.

REFERENCES.

D.M. Back, P.L. Polavarapu, *Carbohydrate Research*, **121**, 308-311, (1983).

D.M. Back, P.L. Polavarapu, *Carbohydrate Research*, **133**, 163-167, (1984).

S.A. Barker, E.J. Bourne, M. Stacey, D.H. Wiffen, Infrared Spectra of Carbohydrates, In *Methods of Biochemical Analysis* Vol 3 Interscience Pub. Co. New York, (1956).

D. Bertrand, M. Lila, V. Furtoss, P. Robert, and G. Downey, *J. Sci. Food Agric.*, **41**,299-307 (1987).

D. Bertrand., P. Robert., M.F. Devaux., J. Abecassis., '*Assignment of Near Infrared Absorption Bands by Muldimensional Analyses of Spectral Data*,' in Analytical Applications of Spectroscopy, C.S. Creaser and A.M.C Davies, Eds (Royal Society of Chemistry, London), pp 450, (1988).

F. Cadet, D.Bertrand, P. Robert, J. Maillot, J. Dieudonné, C. Rouch, *Applied Spectroscopy*, **45** (2), 166-172, (1991).

F. Cadet, F. Wong Pin, C. Rouch, C. Robert, P. Baret, *Biochimica et Biophysica Acta*, **1246**, 142-150, (1995).

P. Dagnelie, *Analyse Statistique à Plusieurs Variables* (Les Presses Agronomiques de Gembloux, Belgium), 185-190, (1975).

M.F. Devaux., D.Bertrand., P.Robert., M. Qannari, 'Application of multidimensional analysis to the extraction of discriminant spectral patterns from NIR spectra,' *Appl. Spectrosc.*, **42** (6), p 1015-1020.

R.J Henry, *Carbohydrate Research*, **141**, 13-19, (1985).

J.L Koenig, *Acc. Chem. Res.*, **14**, 171-178, (1981).

L.P. Kuhn, Infrared Spectra of Carbohydrates, *Anal. Chem.* **22**, 276-281, (1950).

L. Lebart, A. Morineau, and N. Tabard, *Techniques de la Description Stastitiques* (Dunod, Paris), 7-46, (1977).

J. Lefebvre, *Introduction aux Analyses Statistiques Multidimensionelles* (Masson, Paris), 3rd ed., 137-148, (1983).

J. le Nouvel, *Etude d'une Famille de Courbes par Méthodes d'Analyse de Données. Application à l'Analyse Morphologique de Courbes Provenant de Données Médicales*, Thèse de 3ème Cycle, Université de Rennes I, France, (1981).

B. Mills, E.C. Alyea, F.R. van de Voort, *Spectroscopy Lett.*, **19** (3), 277-291 (1986).

K.H Norris, 'Near-Infrared Reflectance spectroscopy' in *Cereals 78 : Better Nutrition for the World's Millions*, Y. Pomerantz, Ed. (American Association of Cereal Chemists, Manhattan, Mount Edgecombe, Natal, 1978), 59-63.

B.G Osborne, S. Douglas, and T. Fearn, *J. Food Technol.* **17**, 355-360, (1982).

RECEIVED: September 27, 1995
ACCEPTED: November 6, 1995