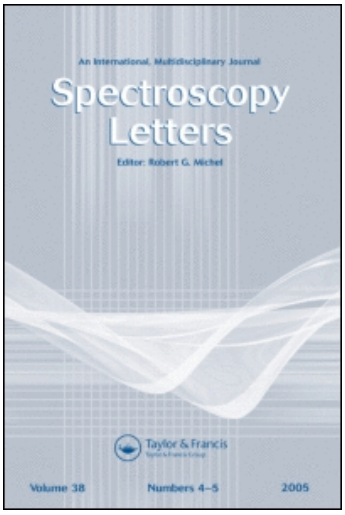

Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Application of Legendre Polynomials Correction Before Component Quantification of Biological Mid-infrared Spectra

Frédéric Cadet[a]
[a] Laboratoire de Biochimie, Faculté des Sciences, Université de la Réunion, Réunion, France-Dom

To cite this Article Cadet, Frédéric(1996) 'Application of Legendre Polynomials Correction Before Component Quantification of Biological Mid-infrared Spectra', Spectroscopy Letters, 29: 5, 937 — 951
To link to this Article: DOI: 10.1080/00387019608001622
URL: http://dx.doi.org/10.1080/00387019608001622

SPECTROSCOPY LETTERS, 29(5), 937-951 (1996)

APPLICATION OF LEGENDRE POLYNOMIALS CORRECTION BEFORE COMPONENT QUANTIFICATION OF BIOLOGICAL MID-INFRARED SPECTRA

Key words: Legendre polynomials, biological samples, Mid-FTIR, PCA, PCR

Frédéric CADET

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.

Quantitative determination from infrared spectra still faces problems that are not completely resolved to date. One of these problems is the deformation of baselines.
We have tested the Legendre baseline correction method by subtracting from a spectrum the low-degreed terms from a

Legendre polynomial function which is obtained by the decomposition of the spectrum.
As opposed to the filtered spectra with n=0 which remained very similar to the original spectra, the filtered spectra, as from n=2, exhibit a great amount of distorsion. This result suggests that it is as from n=2, that the terms that are meaningful to the signal (not just to baseline), are to be substracted.
Mathematical methods such as Principal Component Analysis and Principal Component Regression are used to analyse quantitatively components from biological samples. The thus predicted sucrose content values for values of n that range between 0 and 8 (n={0, 2, 4, 6, 8}) show that the best results are obtained for n=4. The mean and standard deviation values of the difference between reference and predicted values varies between $8.5x10^{-2}$ and $6.6x10^{-2}$ and between 0.224 and 0.195 respectively ; these values are minimum when the n=4 term is subtracted from the spectra.
This processing hence, sensibly improves the quantification of components from Mid-FTIR spectra of biological solutions. This procedure would be interesting in cases where the precision of quantitative measures is essential.

INTRODUCTION

Spectroscopic methods such as infrared, Raman specroscopy and Nuclear Magnetic Resonance (NMR) can easily be used for routine analysis in food and agricultural industries.

Quantitative determination from infrared spectra still faces problems that are not completely solved to date. One of these problems is the deformation of baselines (Powell, 1984, Hruschka, 1987; Cadet and Offmann, 1995). Till now, the origin of such a phenomenon is not completely understood. These deformations are linked to physical problems that are

dependant of the apparatus used ; the variations in the spectra of a same sample measured successively several times would be due to the non-repeatability of the incident ray.
Near infrared spectroscopy is the most widely used technique in food and agricultural industry for the quantitative determination of major biochemical constituants (Osborne and Fearn, 1986). A greater diversity of the products are being analysed (Osborn 1981, Williams and Norris, 1987).

MIR spectroscopy has been used in only few cases for the analysis of food and agricultural products (complex spectra, strong absorption due to water, weak penetration of incident rays) but with the use of powerful micro-computers and with the advent of new techniques such as ATR cells (Attenuated Total Reflectance), MIR spectroscopy has considerably developped (Depecker *et* al., 1985; Crocombe *et* al., 1987; Cadet *et* al, 1991). The interest of using Principal Component Analysi (PCA) for the study of infrared spectra is already established (Gillete and Koenig, 1982; Bertrand *et* al., 1984; Cowe and Nocol., 1985). We have proposed in 1991 the combination of multidimensionnal statistical analysis with Mid-FTIR spectroscopy for the prediction of sucrose content in biolgical samples.
In this present work, the Legendre baseline correction method is tested ; from the Legendre polynomial functions that were obtained from the decomposition of the spectra, the terms with weak n values were subtracted. This method is motivated by the fact that baseline is often assumed to be low-degreed polynomials (Abbink Spaink *et* al., 1986, Liu and Koenig, 1987). The Mid-FTIR spectra obtained are processed before PCA and PCR. The aim of this work is to improve the precision of the quantitative determination of sucrose.

MATERIAL AND METHODS

Biological samples. Sampling of sugar of sugarcane by coring is used. The average of core is about 7000 g. After

pulverization, a subsample of approximately 1000 g is removed. A hydraulic press is used to extract juice from the samples obtained from the coring and from the disintegrator. The subsample is pressed for two and a half minutes at 250 bars. The raw juices that are obtained and that contains impurities and fibres are filtered instantenueously via a highly porous plastic filter when the ATR cells are filled.

Corrected polarization

The reference sucrose content values were determined by corrected polarimetric measures (Cadet *et al* 1991).

The calibration and verification sets are each constituted of a family of 19 raw sugar cane juices. The reference sucrose content values ranged between 16.870 and 20.390 % g/100 ml (mean=18.55, SD=1.06) for the calibration set. For the verification set, the sucrose concentration varied between 16.592 and 20.808 % (mean=18.36, SD=1.136).

Mathematical treatment

The softwares used for the mathematical treatments were developped written in "C" language in our laboratory. Principal Component Analysis (PCA), Principal Component Regression (PCR) and the "spectra pattern" notion used in this paper were extensively described previously (Cadet *et* al, 1991, Cadet *et* al., 1995).

RESULTS AND DISCUSSION

Several hypothesis were made to explain and correct baseline deformations observed during the recording of infrared spectra. Baselines can be described by low-degreed polynomial functions : (i) a perfect baseline is given by the equation y=0; (ii) a shift in the spectrum without internal deformation is

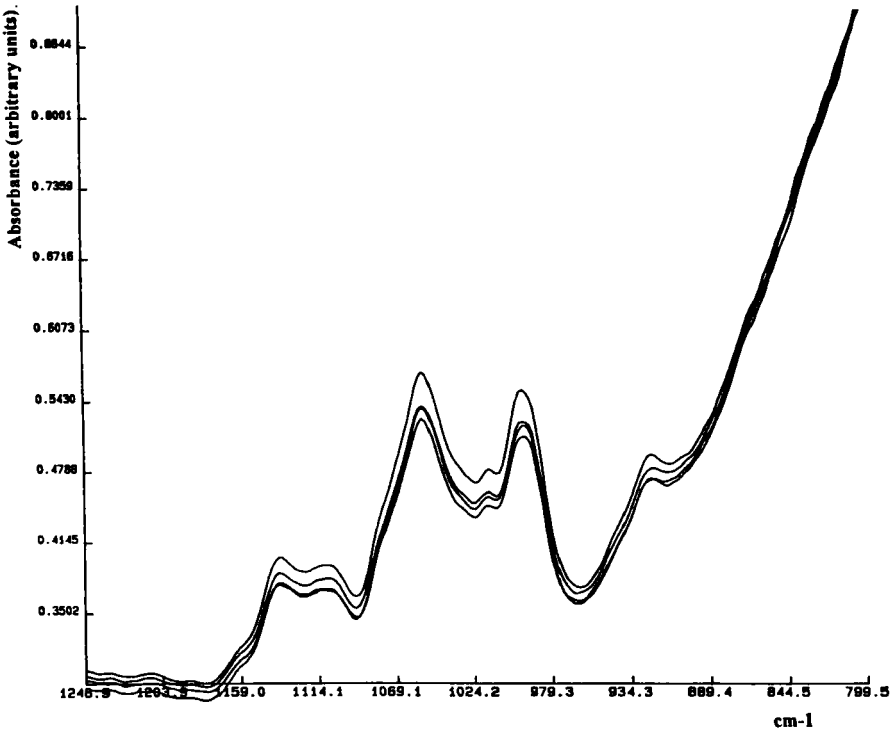

Figure 1. Mid-FTIR (800-1250 cm-1) spectra of raw sugar cane juices before correction.

observed along the absorption axis and is given by y=a ($a \neq 0$); (iii) a shift without deformation along a fixed axis is given by y = ax+b ; (iv) spectral deformations concentrated around a given wavelength can be described by an equation : y = ax^2+bx+c...

Baseline correction was hence perfomed by subtracting low-degree terms from the polynomial functions obtained by the decomposition of spectra.

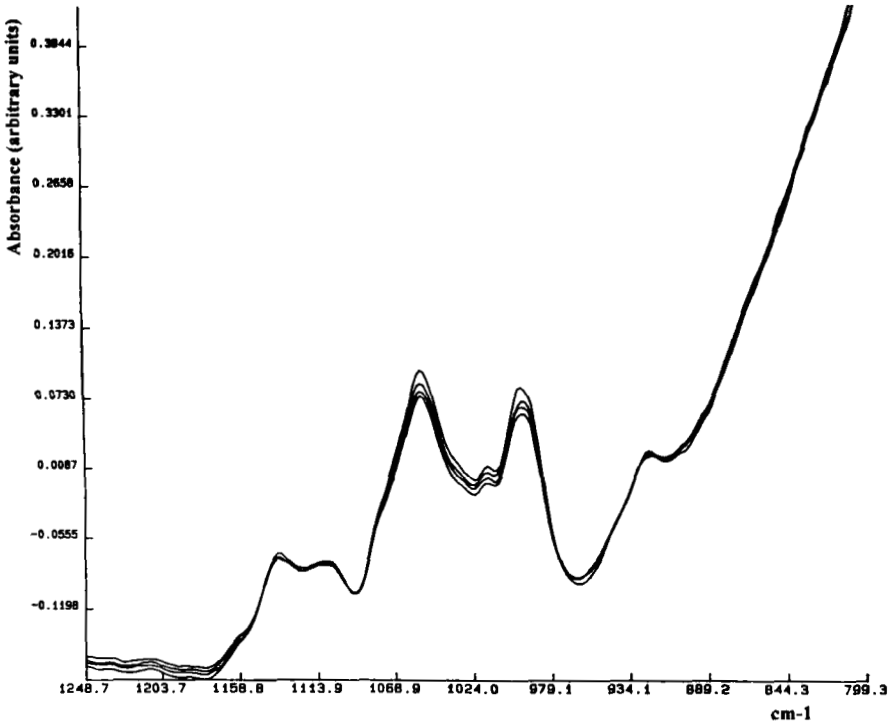

Figure 2. Mid-FTIR (800-1250 cm-1) spectra of raw sugar cane juices after Legendre baseline correction, n=0.

Decomposition Legendre polynomials.

The decomposition of spectra consist in establishing a basis for the whole continuous functions and which would be constituted of polynomial functions. The question is how can the correlation between two spectra be caracterised by baseline deformations ? The visual comparaison of two spectra between a_1 cm^{-1} and a_2 cm^{-1} mainly consist in comparing the areas beneath the spectra. A scalar product can hence be associated with the integral between a_1, and a_2.

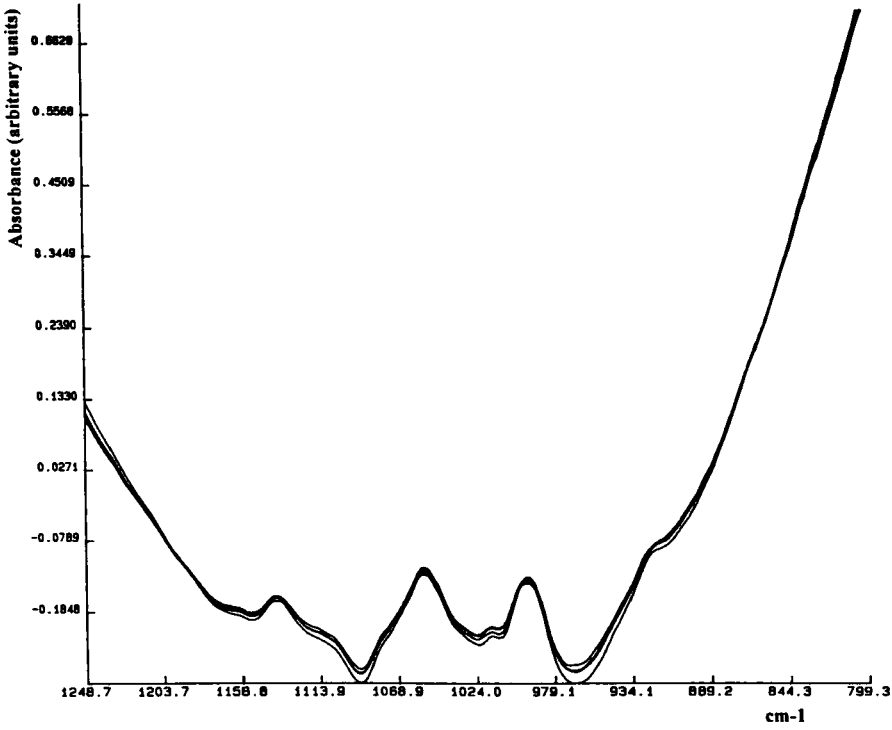

Figure 3. Mid-FTIR (800-1250 cm-1) spectra of raw sugar cane juices after Legendre baseline correction, n=2.

Lipkus (1988) has introduced the use of an orthonormed polynomial Legendre system whose scalar product is given by :

$$\langle f,g\rangle = T \cdot \int_{\alpha 1}^{\alpha 2} f \cdot g \qquad (1)$$

When the orthonormed system (a_1, a_2) is considered : (P_i, i e N) a spectrum S(a) can be decomposed into :

$$S(\alpha) = \sum_{i=0}^{+\infty} \beta_i \bullet P_i \qquad (2)$$

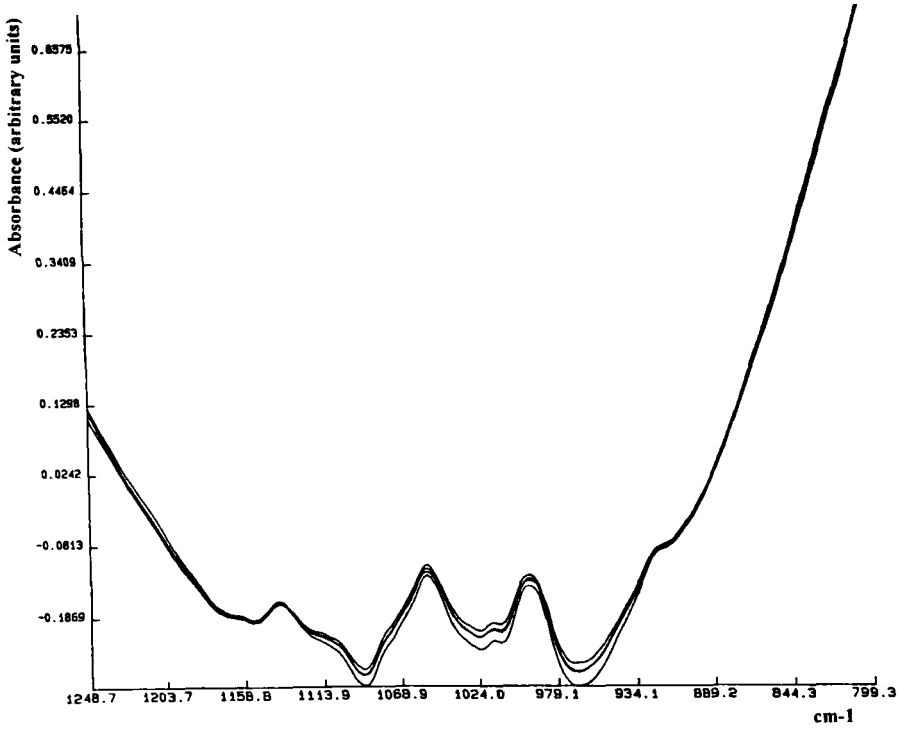

Figure 4. Mid-FTIR (800-1250 cm-1) spectra of raw sugar cane juices after Legendre baseline correction, n=4.

With the elimination of the first terms, it becomes :

$$S_0(\alpha) = S(\alpha) - \sum_{i=0}^{n} \beta_i \bullet P_i \qquad (3)$$

where S0(a) represents the corrected spectrum and n an integer.

It should be noted that : (i) n should be small or else imporant informations contained in the spectrum is eliminated, (ii) a distinction must be made between Legendre filter and baseline correction *sinsu strictum*.

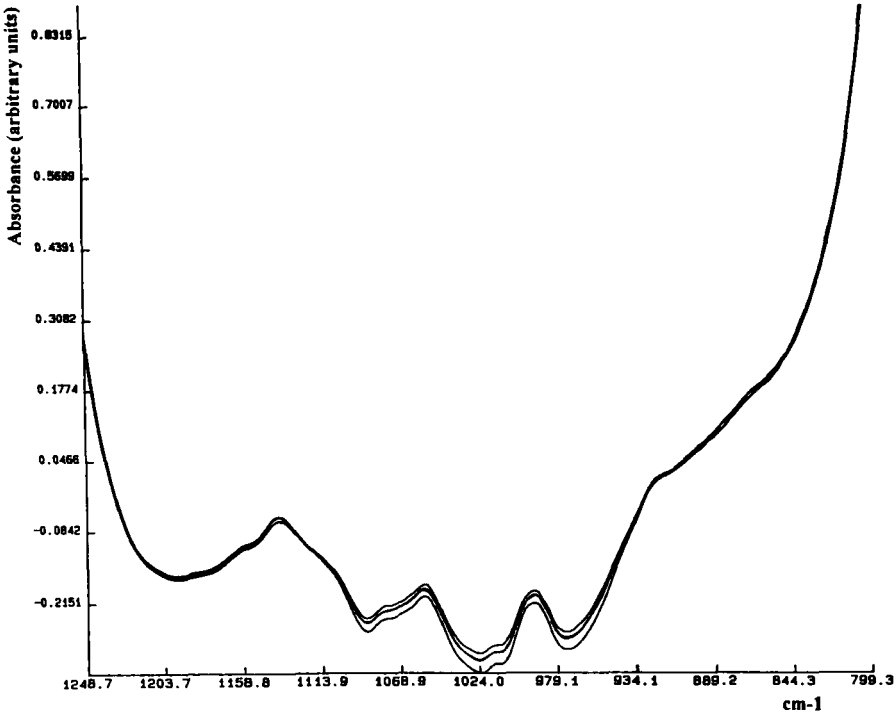

Figure 5. Mid-FTIR (800-1250 cm-1) spectra of raw sugar cane juices after Legendre baseline correction, n=6.

Baseline correction in fact best approach the theoretical baseline. It is an absolute method while Legendre filter simply removes low-degree polynomial terms and is hence a relative method. In fact, the Legendre filter is a normalisation procedure which allows quantitative comparaisons between spectra.

Application

Analysis of a collection of the biological spectra by visual inspection has inherent limitations. If slight differences can be observed in the range 800-1250 cm-1, it is difficult to be certain

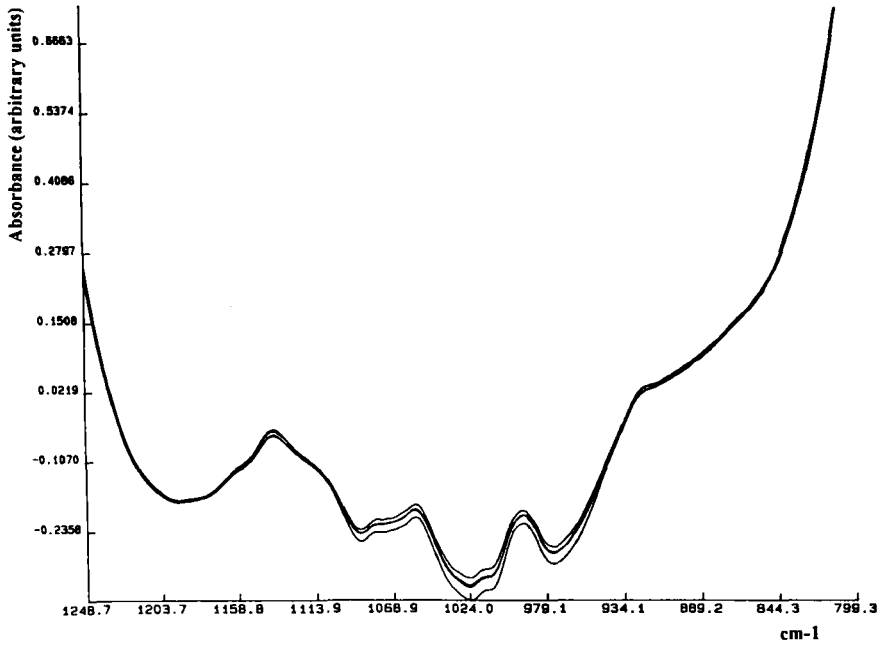

Figure 6. Mid-FTIR (800-1250 cm-1) spectra of raw sugar cane juices after Legendre baseline correction, n=8.

of the origin (concentration, path-length, baseline...) of these differences.

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 be removed.

Figure 1 shows some spectra before application of Legendre polynomial correction. The 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.

Filtered spectra with n=0 remained very similar to the original spectra, but as from n=2, filtered spectra exhibit a great amonut of distorsion; this is due to the removal from their series expansion of the third term which is a quadratic

Table 1. Reference and predicted sucrose concentration in the verification set before and after Legendre baseline correction, n={0, 2, 4, 6, 8}

Sucrose (g/100 ml)													
	Initial spectra			Corrected spectra									
				n=0		n=2		n=4		n=6		n=8	
Sample number	Reference	Predicted	Deviation	Predicted	Deviation	Predicted	Deviation	Predicted	Deviation	Predicted	Deviation	Predicted	Deviation
1	20.808	21.136	0.328	21.346	0.538	21.143	0.335	21.129	0.321	21.143	0.335	21.142	0.334
2	19.288	19.161	-0.127	19.338	0.05	19.207	-0.081	19.183	-0.105	19.212	-0.076	19.209	-0.079
3	16.592	16.793	0.201	16.862	0.27	16.711	0.119	16.771	0.179	16.736	0.144	16.765	0.173
4	18.927	19.124	0.197	19.187	0.26	19.114	0.187	19.098	0.171	19.113	0.186	19.103	0.176
5	17.266	16.975	-0.291	17.008	-0.258	16.992	-0.274	17.009	-0.257	16.993	-0.273	16.996	-0.27
6	18.035	17.874	-0.161	17.893	-0.142	17.871	-0.164	17.881	-0.154	17.847	-0.188	17.89	-0.145
7	19.066	19.175	0.109	19.116	0.05	19.136	0.07	19.149	0.083	19.171	0.105	19.157	0.091
8	16.863	16.84	-0.023	16.88	0.017	16.788	-0.075	16.847	-0.016	16.83	-0.033	16.844	-0.019
9	18.088	18.279	0.191	18.191	0.103	18.294	0.206	18.279	0.191	18.276	0.188	18.219	0.131
10	19.71	20.024	0.314	19.985	0.275	20.026	0.316	19.994	0.284	20.035	0.325	20.018	0.308
11	19.04	19.243	0.203	19.236	0.196	19.191	0.151	19.24	0.2	19.236	0.196	19.27	0.23
12	18.524	18.915	0.391	18.845	0.321	18.896	0.372	18.901	0.377	18.893	0.369	18.899	0.375
13	17.162	17.021	-0.141	16.999	-0.163	16.998	-0.164	17.01	-0.152	17.017	-0.145	17.006	-0.156
14	19.092	19.519	0.427	19.406	0.314	19.46	0.368	19.478	0.386	19.491	0.399	19.471	0.379
15	19.116	19.049	-0.067	19.023	-0.093	19.068	-0.048	19.07	-0.046	19.073	-0.043	19.021	-0.095
16	19.062	19.112	0.05	19.043	-0.019	19.105	0.043	19.053	-0.009	19.096	0.034	19.059	-0.003
17	17.25	17.15	-0.1	17.064	-0.186	17.106	-0.144	17.154	-0.096	17.122	-0.128	17.114	-0.136
18	18.048	18.257	0.209	18.205	0.157	18.226	0.178	18.226	0.178	18.227	0.179	18.211	0.163
19	16.899	16.799	-0.1	16.64	-0.259	16.758	-0.141	16.821	-0.078	16.799	-0.1	16.818	-0.081
mean			**0.085**		**0.075**		**0.066**		**0.077**		**0.078**		**0.072**
Standard deviation			**0.210**		**0.224**		**0.203**		**0.195**		**0.203**		**0.199**

polynomial (Lipkus, 1988). These results suggests that as from n=2, terms of that are meaningful to the signal not just to baseline, are subtracted.
Predicted sucrose concentration for n=0 to 8 are given in table I.
Even if worst results were expected for n values greater than 2, it is found that the best results are obtained with n=4. This can be explained by the fact that polynomials other than those associated with the baseline are eliminated thus facilitating the quantitative determination from a complex mixture such as sugar cane juices. The mean of the difference between reference and predicted values are 0.085 before correction and 0.077 after correction, while standard deviation values are 0.210 and 0.195 before and after correction respectively with n=4.

CONCLUSION

Legendre correction can be use to facilitate the application of pattern recognition to the field of biological FTIR spectra. Legendre correction suppresses unimportant or misleading differences between spectra. This treatment slightly improves quantitative determination from Mid-FTIR spectra of biological solutions. When the precision of the quantitative measures is important, this procedure can show to be very interesting.

Acknowledgements

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

REFERENCES

- Abbink Spaink H., Lub T.T., Otjes R.P., Smit H.C., (1986) "Baseline correction method for second-harmonic detection with tunable diode lasers", *Anal. Chim. Acta.*, 183, 141-154.

- Bertrand D., Robert P., Tran V., (1984), "Traitements mathématiques des spectres NIR de mélanges," 11ème Congrès de L'Association Internationale de Chimie Céréalière, Vienna, Austria, 6th june.

- Cadet F., Bertrand D., Robert P., Maillot J., Dieudonné J.and Rouch C., (1991) "Quantitative determination of sugar cane sucrose by multidimensional statistical analysis of their Mid-Infrared attenuated total reflectance spectra," *Appli. Spectros.*, 45, 2, 166-172.

- Cadet F., Wong Pin F., Rouch C., Robert C., Baret, P., (1995), "Enzyme kinetics by mid-infrared spectroscopy: b-fructosidase study by a one step assay", *Biochim. Biophys. Acta*, 1246, 142-150.

- Cadet F., Offman B., (1995), "Extraction of characteristic bands of sugars by multidimensional analysis of their infrared spectra", *Spectros. Lett.*, Submitted for publication.

- Cowe I.A. and Mc Nichol J.W., (1985), "The use of principal component in the analysis of near infrared spectra," *Appli. Spectrosc.*, **39**, 2, 257-262.

- Crocombe R.A., Olson N.L., Hills S.L., (1987), "Quantitative fourier transform infrared methods for real complex samples.", American Society for Testing and Materials, 95-130.

- Depecker C., Legrand P., Merlin J.C., Sombret B., (1985), "Contribution de la reflexion diffuse à l'étude de composés biologiques", *Spectrosc. Biol. Mol.*, 69-75.

- Gillette P.C., Koenig J.L., (1982), "Noise reduction via factor analysis in FT-IR spectra," *Appl. Spectrosc.*, **36**, 5, 535-539.

- Hruschka W.R., (1987), "Data analysis : wavelength selection methods." In: "Near Infrared Technology in the agricultural and food industry," Eds Williams P.C. and Norris K.H., AACC, Saint Paul, MN (USA), 35-46.

- Lipkus A.H., (1988), "Application of Legendre polynomials to the elimination of baseline variation in biological FT-IR spectra", *Appl. Spectrosc.*, 42, 3, 395-400.

- Liu J., Koenig J.L., (1987), "A new baseline correcion algorithm using objective criteria", *Appl. Spectrosc.*, 41, 447-449.

- Nyden M.R., (1986), "Computer assisted spectroscopic analysis using orthonormalized reference spectra. Part I : application to mixtures", *Appl. Spectrosc.*, **40**, 6, 868-872.

- Osborne B.G. (1981), "Principles and practice of near infrared (NIR) reflectance analysis", *J. Food Techn.*, **16**, 13-19.

- Osborne B.G. and Fearn T, (1986), "Near Infrared spectroscopy in food analysis," Longman Scientific and Technical, Wiley and Sons, NewYork.

- Powell J.R, Wasacz F.M., Jacobsen R.S., (1984), "An algorithm of the reproductible spectral subtraction of water from FT-IR spectra of proteins in dilute solutions and adsorbed monolayers," *Appl. Spectrosc.*, **40**, 3,339-345.

- Williams P, Norris K., (1987), " Near infrared technology in the agricultural and food industry," American Association of Cereal Chemists Ed, Saint Paul, MN, USA, 330 p.

RECEIVED: January 10, 1996
ACCEPTED: February, 15, 1996