Application of normal probability plots and outer product analysis to infrared spectra of sugar beets

J. Maalouly[a], L. J. Eveleigh[a], P. Mery[b], D. N. Rutledge[a], C. J. Ducauze[a]

[a]INAP-G, Chimie Analytique, 16 rue Claude Bernard, 75005 Paris, France

[b]ARTB, 43-45 rue de Naples, 75008 Paris, France

Introduction

The reference method for the measurement of sucrose in sugar beets is polarimetry. It's a destructive method that uses lead acetate - a toxic and polluting salt – as a clarifying agent.

Infrared Spectroscopy can also be used to determine the percentage of sugar in beets. It is an instrumental technique that requires only little sample preparation and it can also be used to determine other quality parameters.

Based on reference values, regression techniques such as Partial Least Squares (PLS) are used to develop mathematical models to predict the sugar content of beets.

Before establishing the model, it is preferable to eliminate uninformative parts of the Near Infrared spectra, in order to improve the model performance. Keeping these parts may lead to noise overfitting; several different methods can be used to determine regions of interest in a spectrum.

Objective

NIR and MIR spectra contain informations about the constituents of sugar beets and more precisely about sucrose. However, the interpretation of bands and peaks in both domains is not straightforward especially in the presence of water.Outer Product Analysis (OPA) combines Near Infrared (NIR) and Mid Infrared (MIR) spectra of sugar beets to facilitate the interpretation of both signals. It can reveal how signals vary simultaneously as a function of some property such as concentration.[1]

Normal Probability Plots of the intensities of the data set of NIR spectra of beets highlight non-linear plots of points corresponding to zones that contain interesting groupings of samples.

Materials and methods

Samples

Before their transformation beets are cut out in three parts (the top, the middle & the root), and each part is grated separately. They are analysed in the minute that follows their grating. Two sets of signals (MIR & NIR) are acquired for the same samples, so that they can later be combined. Sugar content varies between 13 and 20%.

Apparatus

Two Bruker Fourier transform Infrared Spectrometers are utilised for this study. A Vector22 apparatus is used to acquire the NIR spectra by diffuse reflection with an integrating sphere detector. Spectral range is 12000 cm^{-1} to 3500 cm^{-1}. The MIR spectra are acquired from 4000 cm^{-1}

to 650 cm^{-1} with a Vector33 apparatus, a ZnSe Attenuated Total Reflection (ATR) and a DTGS detector.

The background is the air spectra. For NIR spectra the resolution is 8 cm^{-1} and 32 scans / spectrum. For MIR spectra the resolution is 4 cm^{-1} and 64 scans / spectrum. Three spectra are done for each sample.

Chemometric Analysis

Outer Product Analysis

Outer Product Analysis (OPA) starts by calculating the products of the intensities in the two signal domains (NIR & MIR) for each sample (figure 1). All the intensities of one domain are multiplied by all the intensities in the other domain, resulting in a data matrix containing all possible combinations of the intensities in the two domains. The Outer Product of two signal-vectors of lengths r and c for the n samples gives n (r rows by c columns) matrices which are then unfolded to give n (r x c)- long row-vectors.

This procedure corresponds to a mutual weighting of the two signals:
-    If the two signals intensities are high simultaneously, the product is higher;
-    If the two signals intensities are low simultaneously, the product is lower;
-    If one is high and the other low, the product tends to an intermediate value.

After analysis of the set of n (r x c)-long row-vectors, each vector of calculated statistical parameters is folded back to give a (r row by c columns) matrix, which may be easily examined to detect the relations between the two domains.[1, 2]
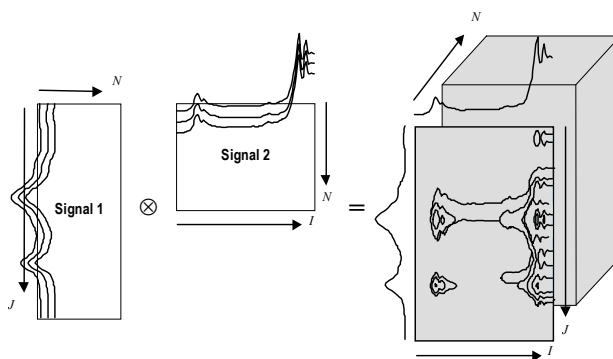

Figure 1. Principle of the OPA

Results of the Outer Product Analysis

This method is applied to the NIR and MIR spectra. The two sets of spectra are sorted in order of increasing sugar content. For the MIR, the first derivative of the spectra is used; this pre-treatment emphasises pronounced but small features over a broad background.[3] NIR spectra are normalised.
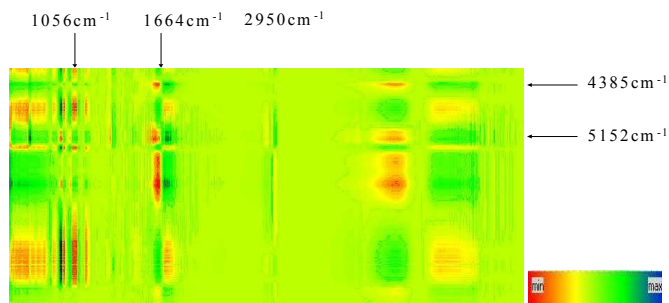
Figure 2. Colour view of the B coefficients

Figure 2 shows the B coefficients of the PLS regression (4 Latent Variables) between the Outer Product matrix and sugar content.

This colour view highlights the presence of peaks relative to the information that varies simultaneously in both signals. For instance, wavelength at $5152cm^{-1}$ corresponds to O-H bend second overtone of H_2O.[4] MIR b-coefficient profile at this wavelength confirms that water is the most absorbing component in this case (Figure 3), and it also shows the presence of CO_2.
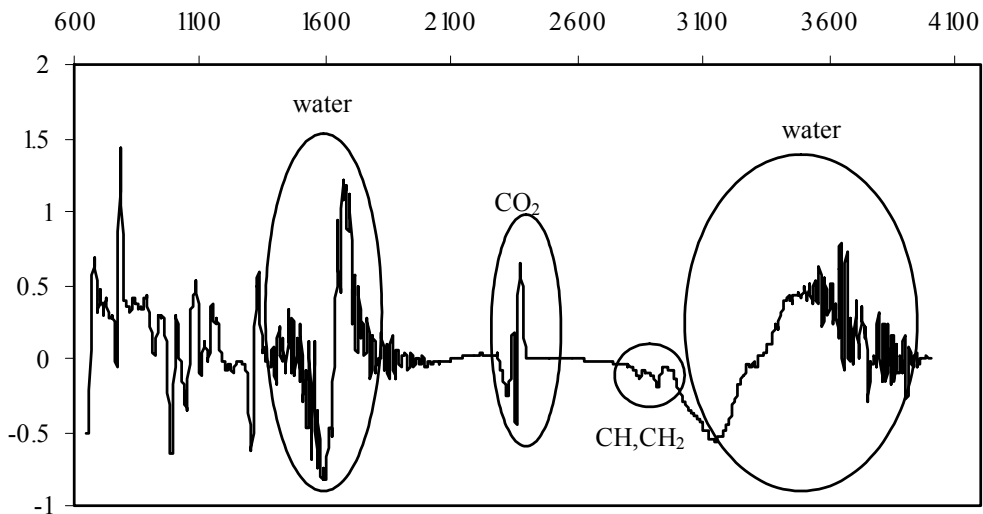

Figure 3- MIR b- coefficient profile at 5152cm⁻¹

However it is also possible to notice the presence of peaks in the $2900-2955cm^{-1}$ due to CH and CH_2 vibrations attributed to sugar (CH) and probably to proteins (CH_2). This feature was almost invisible in the original spectra.

MIR b-coefficient profile at $4385cm^{-1}$ shows[5] that sucrose absorbs at this wavelength (Figure 4). Two bands of water are present in this figure; one of the bands is almost invisible.
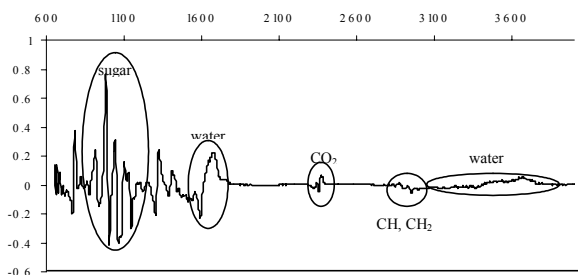
Figure 4. MIR b-coefficient Profile at 4385cm⁻¹

A closer look at the NIR b-coefficient profile at $2952cm^{-1}$ shows the CH $\&CH_2$ vibrations (Figure 5), these peaks are negative. It is also possible to notice the presence of a negative water band. A low intensity sugar band is visible.
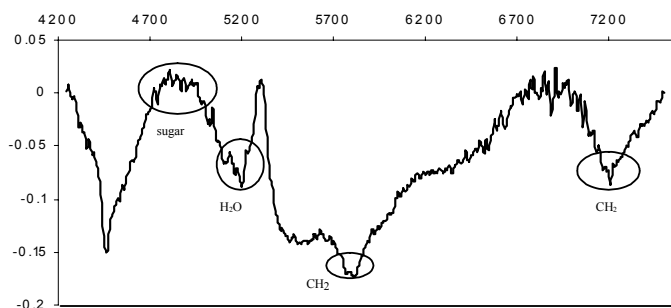

Figure 5. NIR b-coefficient profile at 2952cm⁻¹

Peaks at $1056cm^{-1}$ and $996cm^{-1}$ in figure 4 are related to sugar and the NIR b-coefficient profile at these wavelengths (figure 6) shows the presence of a positive band corresponding to sugar.
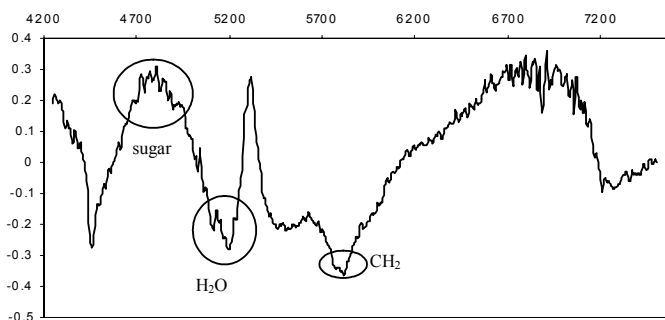

Figure 6. NIR b-coefficient profile at 996cm⁻¹

Figures 5 & 6 show that peaks of sugar and CH_2 are anti-correlated. This surprising result could be due to the contribution of non-sugar constituents, such as proteins, to these CH_2 peaks.

Normal Probability Plots of samples at different wavelengths

The Normal Probability Plot is a graphical technique for assessing whether or not a data set is approximately Normally distributed. The data are plotted against a theoretical normal distribution in such a way that Normally distributed points should form a straight line. Departures from this straight line indicate departures from Normality.

The normal probability plot is formed by:
- Vertical axis: Ordered response values;
- Horizontal axis: Normal order statistic medians.

The observations are plotted as a function of the corresponding normal order statistic medians that are defined as:

$$N (i) = G (U (i)) \qquad (1)$$

Where U(i) are the uniform order statistic medians (defined below) and G is the percent point function of the normal distribution. The percent point of the function is the inverse of the cumulative distribution function (probability that x is less than or equal to some value). That is, given a probability, we want the corresponding x of the cumulative distribution function.[6]

If at a given wavenumber, the infrared signal (ATR or diffuse reflection) does not depend on the sugar content, then one can expect that for a large set of analytical samples (here: beets) the signal should be Normally distributed. If the signal depends on sugar content and if the sugar content itself is not Normally distributed, one would expect a non-Normal distribution for the signal intensities. Thus, a close inspection of the Normal Probability Plots may help to point out relevant wavenumbers to be used during model building.

MIR spectra are used for this application. Spectra are pre-treated by SNV (Standard Normal Variate). Three groups of beets with different sugar contents were used.
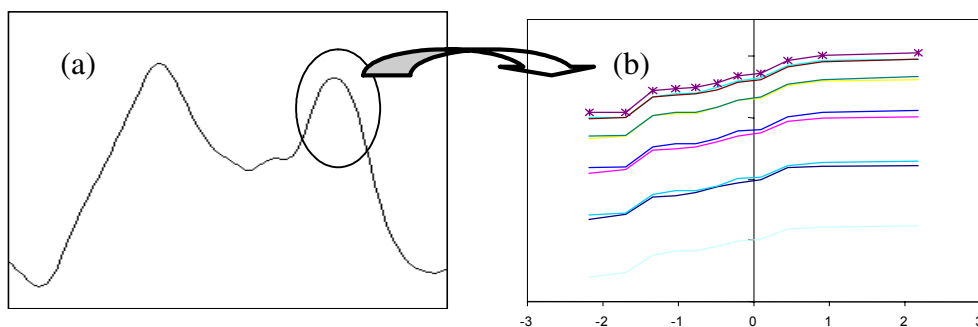

Figure 7. (a) MIR peak corresponding to C-O-C deformation of sugar; (b) Normal Probability Plot in this zone 1005cm^{-1} –987cm^{-1}

Figure 7b shows the presence of more than two groups of samples for the wavenumbers that correspond to the C-O-C peak in figure 7a, characteristic of the sugar.[7] Different slopes characterise each group.
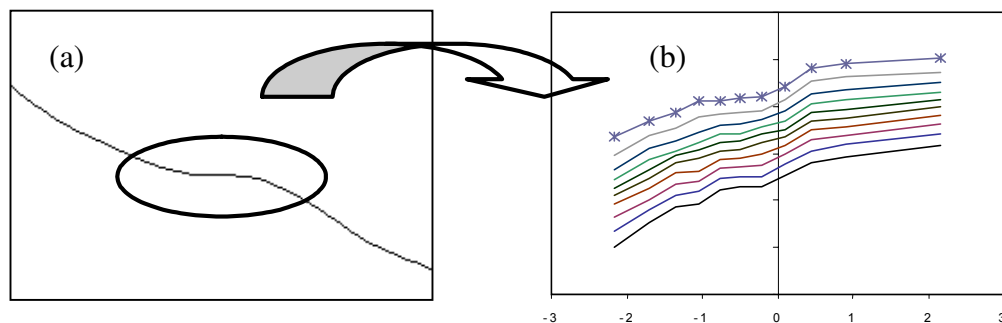
Figure 8. (a) MIR peak corresponding to CH2 vibration; (b) Normal Probability Plot at 2958cm^{-1} – 2940cm^{-1}

Figure 8a shows the presence of a shoulder that characterises the absorbance CH_2 –(CH2 symmetric elongation [7]). The Normal Probability plot at these wavenumbers confirms the presence of different groups of samples (change in the slope).

Conclusion

Infrared spectra contain information about the sugar content of beets. However, the information is not always easy to detect - especially in the presence of water. Outer Product Analysis has proven to be a simple and efficient tool to highlight zones of interest in the spectra.

Normal Probability Plots indicate zones in the spectra where there is a change in the distribution of the samples. They can be useful to show regions where different components absorb, and where there are groupings of samples.

References

1. D.N. Rutledge, A.S. Barros, R. Giangiacomo, *Magnetic Resonance in Food Science*, 179-192, (2001).
2. M.H. Lopes, A.S Barros, C. Pascoal Neto, D. Rutledge, I. Delgadillo, A.M. Gil, *Biopolymers (Biospectroscopy)*, Vol. 62, 268-277 (2001).
3. Spectroscopic Software, Bruker, OPUS NT, Quant 2, 8-10.
4. C. Di Natale, M. Zude-Sasse, A. Macagnano, R. Paolesse, B. Herold, A., D'Amico, *Analytica Chimica Acta*, 459, 107-117, (2002).
5. Y. Roggo, L. Duponchel, B. Noé, J.P. Huvenne, *Near Infrared Spectroscopy*, 10, 137-150, (2002).
6. http://www.itl.nist.gov/div898/handbook/eda/section3/normprpl/htm.
7. D. Bertrand, E. Dufour, La spectroscopie infrarouge et ses applications analytiques, *Editions TEC & DOC*, 171-195.