We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,800 Open access books available
116,000 International authors and editors
120M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected. For more information visit www.intechopen.com
Chapter 4

Carbohydrate Analysis by NIRS-Chemometrics

Mercedes G. López, Ana Sarahí García-González and Elena Franco-Robles

Abstract
Near-infrared spectroscopy (NIRS) is a high-throughput, low-cost, solvent-free, and nondestructive analytical tool. Chemometrics is the science that employs statistical and mathematical methods to explain near-infrared spectra; it has been proven that when they are coupled, their effectiveness highly improved the carbohydrate characterization. This chapter focuses on the fundamentals of near-infrared spectroscopy in the study of carbohydrates, as well as the application of partial least squares regression (PLSR) and principal component analysis (PCA), as the most useful chemometric techniques involved in carbohydrate analysis. The theoretical aspects and practical applications starting from simple to complex carbohydrates mixtures are covered. Indeed, the contributions from different fields extend the implementation of near-infrared spectroscopy from industrial quality control to scientific research.

Keywords: near-infrared spectroscopy, chemometrics, carbohydrates, polysaccharides, partial least squares regression, principal component analysis

1. Introduction

In a vibrational spectroscopy, near-infrared spectroscopy (NIRS) covers the transition from the visible spectral range to the mid-infrared region. The NIR spectral region ranges from 800 to 2500 nm (12,500–4000 cm$^{-1}$) with absorptions representing overtones and combinations mainly associated with –CH, –OH, –NH, and –SH functional groups [1]. NIR spectroscopy in combination with chemometric analyses can provide unique information in a wide field of applications from life sciences to environmental issues. It is more frequently used in
the agricultural field [2–5], in particular, on the elucidation of nonstructural carbohydrates (NSCs) of plants. NSCs are products of the photosynthesis, providing substrates for growth and metabolism and can be stored by the plant playing a central role in the plant response to the environment [6, 7]. This type of carbohydrates is classified into monosaccharides (glucose and fructose), disaccharides (sucrose), polysaccharides (starch and fructans), oligosaccharides (raffinose), and sugar alcohols (inositol, sorbitol, and mannitol) [8, 9].

NIR spectroscopy is widely used to follow the chemical, physical, technological, or physiological processes that affect the structure and composition of carbohydrates found in many different organisms [10]. The success of this technique relies on the rapid and nondestructive analysis of the sample without the use of chemicals [11]. In addition, the data can be analyzed with chemometric methods. In this regard, partial least squares regression (PLSR) and principal component analysis (PCA) are two of the most recognized statistical methods that can be used to build NIR-chemometric models. PLSR is a well-established method for multivariate modeling and calibration [12]. Meanwhile, PCA analyzes data tables representing observations described by several dependent variables, which are, in general, intercorrelated [13].

The objective of this chapter is to give a comprehensive overview of NIR spectroscopy for analyzing carbohydrates, such as glucose, fructose, sucrose, and fructans. In addition, we describe NIR spectroscopy and multivariate methods used to identify, classify, and quantify carbohydrates in plant tissues. Furthermore, we present the main applications of NIR chemometrics on carbohydrate analyses.

2. NIR spectra: characteristic bands of oligosaccharide and polysaccharide

The term “near” in NIR relies on the position of the electromagnetic energy lying next to or near the visible energy range. Molecular vibrations in the middle infrared (MIR) range cover absorptions bands between 2500 and 25,000 nm (4000 and 400 cm⁻¹) representing the most intense and simplest bands in the whole infrared range, whereas NIR bands arise in the interval between 800 and 2500 nm (12,500 and 4000 cm⁻¹) covering absorptions corresponding to overtones and combinations of fundamental vibrations [14]. NIR spectroscopy is concerned with both electronic and vibrational transitions [1]. Bands due to electronic transitions are observed in the NIR region and in general are presented as weak bands. Moreover, bands arising from overtones and combination modes are so-called forbidden transitions. Starting from the diatomic molecule as the simplest vibrating system, described by the harmonic and anharmonic oscillator, the study of more complex substances is referred to as polyatomic molecules [14].

The NIR region can be divided into three regions. Region I spans from 800 to 1200 nm (12,500–8500 cm⁻¹), also known as the “the short-wave NIR region (SWNIR),” “near-NIR region (NNIR),” or “the Herschel region,” represents bands resulting from electronic transitions, overtones, and combinations modes. Region II ranges from 1200 to 1800 nm (8500–5500 cm⁻¹) and covers first overtones of XH (X = C, O, N), stretching vibrations and various types of combination modes. Finally, Region III (1800–2500 nm or 5500–4000 cm⁻¹) is a combination mode region. Many applications utilize Regions II and III [1].
Absorptions due to different functional groups, especially –CH, –OH, and –NH, are displayed as molecular overtones and combination vibrations at specific wavebands [15, 16]. NIR spectral data are influenced by a particle size (e.g., ground or powder) and need to be properly calibrated [17]. In Table 1, the characteristic bands of oligosaccharide and polysaccharide are listed.

NIRS has been used as a fingerprint technique for all kinds of samples (liquids, solids, and semisolids), independently of their nature, relatively simple substances or pure compounds, most times they show broad and overlapping bands, it is impossible to correctly assign the specifically vibrations, and cannot be used for structural determination of a sample [18].

<table>
<thead>
<tr>
<th>Carbohydrate type</th>
<th>Waveband</th>
<th>Wavenumber</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>nm</td>
<td>cm⁻¹</td>
</tr>
<tr>
<td>Glucose</td>
<td>OH stretch 1st overtone</td>
<td>2340, 2255, 2150, 2085, 1902, 1730, 1590, 1520, 1385, 1195</td>
<td>[19]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4274, 4435, 4651, 4796, 5258, 5780, 6289, 6579, 7220, 8368</td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td>OH stretch/OH bend</td>
<td>1688</td>
<td>5924</td>
</tr>
<tr>
<td>Sucrose</td>
<td>OH stretch 1st overtone</td>
<td>1433</td>
<td>6978</td>
</tr>
<tr>
<td>Sucrose/glucose/fructose</td>
<td>OH combination</td>
<td>1928</td>
<td>5186</td>
</tr>
<tr>
<td>Sucrose/glucose/fructose</td>
<td>OH stretch/OH stretch/CO stretch</td>
<td>2123–2200</td>
<td>4710–4717</td>
</tr>
<tr>
<td>Crystalline sucrose</td>
<td>OH stretch 1st overtone</td>
<td>1443–1440</td>
<td>6930–6944</td>
</tr>
<tr>
<td>Polysaccharides</td>
<td>CH stretch/CH deformation combination</td>
<td>2328</td>
<td>4295</td>
</tr>
<tr>
<td>Polysaccharides</td>
<td>OH stretch/OH stretch/CO stretch</td>
<td>2274, 2271–2270</td>
<td>4398, 4403–4405</td>
</tr>
<tr>
<td>Polysaccharides</td>
<td>OH combination</td>
<td>2090</td>
<td>4785</td>
</tr>
<tr>
<td>Polysaccharides</td>
<td>OH stretch/OH bend</td>
<td>1920</td>
<td>5208</td>
</tr>
<tr>
<td>Polysaccharides</td>
<td>CH combination/CH 1st overtone</td>
<td>2328, 2270, 2078, 1920, 1857–1853</td>
<td>4295, 4405, 4813, 5208, 6300–6317</td>
</tr>
<tr>
<td>Polysaccharides</td>
<td>OH stretch 1st overtone</td>
<td>1437–1389</td>
<td>6960–7200</td>
</tr>
</tbody>
</table>

Table 1. Characteristic bands of oligosaccharide and polysaccharide.

3. Multivariate data analysis by NIRS

NIR spectra are characterized for their complexity and difficulty to be interpreted. For these reasons, multivariate methods from chemometrics are required to understand NIR spectra.

Chemometrics comprise the development and use of mathematical and statistical methods for applications in chemistry. As a discipline, the aim of chemometrics is to provide methods to extract relevant chemical information out of measured chemical data in order to represent and display this information.
Figure 1 shows a general scheme for multivariate techniques, including the two different chemometric groups that are frequently employed in the NIR spectra analysis: the qualitative (classification) methods and the quantitative (regression) methods. As a first step, before choosing any method, usually NIR spectra are preprocessed with mathematical treatments, such as baseline correction, normalizations, derivatives, and smoothing, in order to enhance the relevant information and reduce the influence of side information contained in the spectra. The classification methods are used to group or separate the samples according to their spectra. The regression methods correlate the spectrum to quantifiable properties of the samples.

3.1. Quantitative analysis

The basic principles used for quantitative analysis are fundamentally invariable for all optical and spectral measurement methods. The principle behind any quantitative analysis is that the desired quantity, property, parameter, or compound can be determined from the signal obtained by an instrument, and this signal differs in a predictable manner for a given experimental system. The magnitude of the signal obtained can be correlated, directly or by mathematical algorithms, to the target characteristic properties of a sample. A common implementation of quantitative analysis is the determination of the concentration of a given analyte. For most applications, an attempt is made to linearize the relationship between the analyte and the instrument response, although this is not essential if a well-defined mathematical relationship can be established. This leads to the generation of a calibration from a characterized standard set (references) with the objective to construct a prediction model for a group of samples (Figure 2) [30].

Many successful NIRS analysis have been performed using PLSR as a quantitative chemometric technique. Its usefulness derives from its potential to analyze data with numerous, noisy, collinear, and even incomplete variables. By establishing a linear relationship between two

![Figure 1. General scheme showing the commonly multivariate techniques employed by NIR spectroscopy.](image-url)
data matrices, the spectral data X and the reference values Y, through a linear multivariate model, the PLSR technique finds out the variables in the X matrix that will best define the Y matrix. In other words, it represents the NIR spectra in the space of wavelengths in order to display directions that will be linear combinations of wavelengths called factors that describe the studied property [31, 32].

3.2. Qualitative analysis

Qualitative analyses are used for the classification of samples in accordance with their NIR spectra. Two general approaches can be used for qualitative classification: the unsupervised and the supervised methods. In the first approach, samples are classified lacking preceding knowledge, except the spectra. On the other hand, supervised methods require a prior knowledge of the sample, for instance, a category membership, generating a classification model with a training set of samples with well-established categories. The obtained model performance is evaluated by relating the classification predictions to the well-known categories of the validation samples [33].

Principal component analysis (PCA) is one of the most popular classification methods utilized in life sciences. PCA is used to visualize the most important information from a given data. One of the most significant advantages of PCA application is the reduction of the number of variables (scores), allowing the representation of a multivariate data table in a small dimensional area. Its purpose is to obtain significant information from the NIR spectra to express it as a set of new orthogonal variables called principal components (PC). The first principal component (PC1) defines the maximum variability scattered within the samples. A second principal component (PC2), uncorrelated and orthogonal to the first principal component,
explains the maximum variability not described by the first component, this behavior continues with the next principal component (PC3), and so on [12]. Thereby, a display pattern of similarity of the variables as points in maps is created.

4. Applications of carbohydrates analysis by NIRS

The near-infrared spectroscopy (NIRS) is a technique that allows the measurement of carbohydrates in a wide variety of samples. Nowadays, NIRS-chemometrics have proven their effectiveness for both qualitative and quantitative carbohydrate analysis. NIRS has several advantages such as allowing the sample remains intact after analysis and giving access to multiple chemical as well as physical properties at the same time [34].

NIR spectroscopy is generally chosen for its high-throughput screening, reduced sample preparation, low cost, and the nondestructive nature toward the analyzed sample [14]. However, establishing a suitable calibration demands a big effort and requires reference values for each sample, which makes it time-consuming and costly at the beginning [35].

In the agrifood sector, the potential of NIRS have been widely investigated, this is a very powerful tool that provides meaningful information about internal and external properties of fruits, such as sugar content, total acidity, pH, soluble solid content, dry matter, firmness, and bruises, to mention some [36]. Moreover, NIRS can be applied to a wide variety of problems such as determination of particle size [38], determination of the best harvesting time [37], and investigation of geographical origin of foods such as apples, meat, and cheese [39].

However, and particularly to specific sugar content, NIRS in combination with PLSR models has been used in sorghum stalks [40] and sweet sorghum (cellulose, lignin, and hemicellulose) [41], fruit juices [42, 43], rice (amylose) [44], whey (lactose) [45], grasses (fructans) [46, 47], maize (nonstructural and water soluble carbohydrates) [48], intact apple fruit to determine fructose, glucose, and sucrose [49], orange [50], apricot [51], sugar beet [52], cherries [53], and other fruits (Table 2). All these studies accorded that the performance of NIR spectroscopy is comparable to the reference chromatographic method, but the former is much faster and easier to carry out.

On the other hand, NIRS has been applied on food quality evaluation; it is often used to check if fruits or vegetables are green or rotten to detect surface defects. NIRS is also employed to check sugar concentrations, for instance, not only in apples [64], oranges [55, 56], mango [65], kiwifruits [57], sugar beet [54], peaches [66], jujube [67], onion [68], potato tubers [58], Nules Clementine [62], passion fruit [69], but also in fruit juices [43], wine [59], or cakes [60] (Table 2). Additionally, it has been used in breadstuff, dairy products, meat, vegetables, and fish products and in processed food to provide information about overtones and their combinations [70]. Moreover, studies have been performed to demonstrate that NIRS-chemometric analyses are of greater predictive value than mid-infrared data. In Chinese yams, Zhuang et al. [63] analyzed with NIR and MIR spectroscopy, the authors concluded that reasonable results were obtained using both spectral data sets and methods, but that NIR-chemometric data derived better prediction models.
In respect to specific absorption peaks, sugar analyses have been carried out in fruit juices establishing that NIRS can deal with the distortions due to water clusters [20–22, 42].

NIR techniques have also been applied to measure biomass composition, especially on the presence of structural carbohydrates. The National Renewable Energy Laboratory (NREL) reported sorghum composition prediction models for glycan, xylan, lignin, starch, extrac- tives, and ash [71].

NIR spectroscopy is not only useful in laboratory measurements sites but also applicable to online and field studies. The study of 116 syrup samples to compare a portable spectrometer and a benchtop device showed that the reduced wavelength range and reduced resolution of the portable device is sufficient to receive calibrations with $R^2 \geq 0.96$ for standard syrups with comparable standard error of prediction (SEP) values of $1.30$ g/100 g versus $1.19$ g/100 g, $0.94$ g/100 g versus $0.99$ g/100 g, and $2.04$ g/100 g versus $2.46$ g/100 g for glucose, fructose, and sucrose, respectively, to the handheld device [61]. The developed method is suitable to be implemented for quality control in the producing industry as well as in grocery stores.

A relevant novel application of the predictive models, particularly of the direct NIR prediction on diverse parameters on fruit quality was demonstrated. In Ref. [50], the authors com-

<table>
<thead>
<tr>
<th>Sample</th>
<th>Carbohydrate</th>
<th>Analysis</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain sorghum stalks</td>
<td>Sucrose, glucose</td>
<td>PLSR</td>
<td>[41]</td>
</tr>
<tr>
<td>Fruit juices</td>
<td>Glucose, fructose, sucrose</td>
<td>PLSR, PCA</td>
<td>[43, 44]</td>
</tr>
<tr>
<td>Rice</td>
<td>Amylose</td>
<td>mPLSs</td>
<td>[45]</td>
</tr>
<tr>
<td>Whey</td>
<td>Lactose</td>
<td>PLS</td>
<td>[46]</td>
</tr>
<tr>
<td>Grasses</td>
<td>Fructan</td>
<td>PLSR</td>
<td>[47, 48]</td>
</tr>
<tr>
<td>Apple fruit</td>
<td>Glucose, fructose, sucrose</td>
<td>PLS</td>
<td>[50]</td>
</tr>
<tr>
<td>Sugar beet</td>
<td>Sucrose</td>
<td>SEPs</td>
<td>[53, 55]</td>
</tr>
<tr>
<td>Cherries</td>
<td>Total carbohydrates</td>
<td>PLSR</td>
<td>[54]</td>
</tr>
<tr>
<td>Oranges</td>
<td>Glucose, fructose, sucrose</td>
<td>PLSR</td>
<td>[56, 57]</td>
</tr>
<tr>
<td>Kiwifruit</td>
<td>Glucose, fructose, sucrose</td>
<td>PLS</td>
<td>[58]</td>
</tr>
<tr>
<td>Potato</td>
<td>Glucose, fructose</td>
<td>PLSR</td>
<td>[59]</td>
</tr>
<tr>
<td>Wine</td>
<td>Glucose</td>
<td>PLSR, PCR</td>
<td>[60]</td>
</tr>
<tr>
<td>Cakes</td>
<td>Sucrose</td>
<td>MLR</td>
<td>[61]</td>
</tr>
<tr>
<td>Syrup</td>
<td>Glucose, fructose, sucrose</td>
<td>PLSR</td>
<td>[62]</td>
</tr>
<tr>
<td>Nules Clementine</td>
<td>Glucose, fructose, sucrose</td>
<td>PLS, PCR</td>
<td>[63]</td>
</tr>
<tr>
<td>Chinese yams</td>
<td>Total carbohydrates</td>
<td>PLS, PCA, LS-SVM</td>
<td>[64]</td>
</tr>
</tbody>
</table>

mPLSs, various modified partial least square; PLS, partial least square; SEPs, standard errors of prediction; MLR, multiple linear regressions; and LS-SVM, least squares-support vector machine.

Table 2. Samples analyzed by NIR in a carbohydrate study.
pared two commercial portable spectrometers (Vis/NIR spectrometer versus OTF-NIR) for four orange varieties quality: soluble solids content, acidity, titratable acidity, maturity index, flesh firmness, juice volume, fruit weight, rind weight, juice volume to fruit weight ratio, fruit color index, and juice color index, and they found relevant the prediction of maturity index. The Lab spec spectrometer showed better predictive performance than the laminar instrument.

In another study, a Lab spec Pro portable spectrophotometer to conduct an online classification of beef tenderness was also successful [72].

In sugar-flour mixtures, NIR spectroscopy displayed proper results on the characteristic absorption bands of sugars, which are 1200 nm (8333 cm\(^{-1}\)), 1437 nm (6959 cm\(^{-1}\)), 2074 nm (4822 cm\(^{-1}\)), and 2320 nm (4310 cm\(^{-1}\)). However, it was not possible to distinguish various sorts of sugars, for instance, make a difference between the sucrose of the powdered sugar and the numerous carbohydrates present in the flour. Nevertheless, the identification of specific signatures of sugars can be very useful for rapid detection in the industrial sector [73].

Honey represents another class of samples that have proven the effectiveness of a NIR analysis [74]. In a study on Galicia honeys with protected geographical indication (PGI), the samples were processed by different chemometric methods to develop an authentication system specific to this type of honey. In this work, fifteen Galicia certificated PGI honeys were differentiated from fifteen commercial available honeys by PCA, demonstrating that a single and fast chemometric method could be used to indicate the genuineness of Galicia PGI samples. Figure 3A shows the NIR spectra of all the analyzed samples and Figure 3B, illustrates the discrimination of Galicia PGI honeys from the other samples by the PCA plot.

Similarity, the potential use of NIR-PCA analysis to monitor sugar adulteration in onion powders was assessed through a detailed examination of the feasibility of quantification of cornstarch as an adulterating ingredient in onion powders [75]. Spectral analysis of 18 concentrations of starch in 180 onion powders, ranging from 0 to 35%, was conducted. The NIR spectra of the pure and adulterated onion powders (Figure 4A) reveal differences in

Figure 3. (A) NIR spectra of honey samples. (B) Score plot of the honey samples in the space defined by the first two principal components. Adapted from Ref. [74].
the absorption intensities between 1920 (5208 cm$^{-1}$) and 1980 nm (5051 cm$^{-1}$). The absorption bands in these regions correspond to the O–H stretch and O–H band combination and the H–O–H deformation combination, which represents the starch content. The spectral fluctuations from 1400 (7143 cm$^{-1}$) to 1600 nm (6250 cm$^{-1}$) correspond to the first overtone of the hydroxyl group. The precise position of these bands is very sensitive to hydrogen bonding in the starch molecule, causing a difference between genuine and adulterated samples. The application of PCA (Figure 4B) resulted in a data grouping of each of the different concentrations used, working as a discriminative screening tool of authentic and adulterated samples.

Applications of NIRS have been developed also in the nutrition and health fields. NIR and MIR spectroscopy measurements and multivariate calibration methods based on partial least square regression have been used in a determination of fat, proteins, carbohydrates, and energy values in baby food, infant fast food, and canteen menus, with a simple, fast, and good predictive capabilities [70]. Another great diagnostic application is the measurement of blood glucose [1].

Finally, another notable capacity of NIRS was the prediction of carbohydrates concentrations, and distribution, leading to high ratio of performance to deviation (RPD) values, reducing the use of chemicals and working time, confirming that this makes a suitable technique of industry applications [61].

5. Conclusions

The potential of NIR spectroscopy in combination with chemometrics on carbohydrate analysis has been fully demonstrated. NIR is a powerful technique to study carbohydrates composition, type, and levels. This method can be used qualitatively and quantitative to detect, identify, and qualify carbohydrates. These unique capabilities enable the employment of NIR-chemometric in numerous applications: from state-of-the-art scientific experiments to on-line industrial processing control.
Author details

Mercedes G. López*, Ana Sarahí García-González1 and Elena Franco-Robles2

*Address all correspondence to: mlopez@ira.cinvestav.mx

1 Center for Research and Advanced Studies of the National Polytechnic Institute, Irapuato, Mexico
2 University of Guanajuato, Irapuato, Mexico

References


