Chemometric Tools for NIRS and NIR Hyperspectral Imaging

Laura M. DALE1, 2, André THEWIS1, Ioan ROTAR3, Juan A. FERNANDEZ PIERNA4, Christelle BOUDRY1, Roxana M. VIDICAN5, Vincent BAETEN6

1) Animal Science Unit, Gembloux Agro-Bio Tech, University of Liege, 2, Passage des Déportés, 5030 Gembloux, Belgium; dale lm@yahoo.com; dale.laura@student.ulg.ac.be; athewis@ulg.ac.be; christelle.boudry@ulg.ac.be.
2) Department of Grassland and Forage Crops, University of Agricultural Science and Veterinary Medicine Cluj Napoca, 3-5, Calea Manastur, 400372, Cluj Napoca; rotarioan52@yahoo.fr; roxanavidican@yahoo.com;
3) Walloon Agricultural Research Centre, Valorisation of Agricultural Products Department, 24 Chaussée de Namur, 5030 Gembloux, Belgium; fernandez@cra.wallonie.be; baeten@cra.wallonie.be.

Abstract. Nowadays in agriculture, new analytical tools based on spectroscopic technologies are developed. Near Infrared Spectroscopy (NIRS) is a well known technology in the agricultural sector allowing the acquisition of chemical information from the samples with a large number of advantages, such as: easy to use tool, fast and simultaneous analysis of several components, non-polluting, non-invasive and non-destructive technology, and possibility of online or field implementation. Recently, NIRS system was combined with imaging technologies creating the Near Infrared Hyperspectral Imaging system (NIR-HSI). This technology provides simultaneously spectral and spatial information from an object. The main differences between NIR-HSI and NIRS is that many spectra can be recorded simultaneously from a large area of an object with the former while with NIRS only one spectrum was recorded for analysis on a small area. In this work, both technologies are presented with special focus on the main spectrum and images analysis methods. Several qualitative and quantitative applications of NIRS and NIR-HSI in agricultural products are listed. Developments of NIRS and NIR-HSI will enhance progress in the field of agriculture by providing high quality and safe agricultural products, better plant and grain selection techniques or compound feed industry’s productivity among others.

Keywords: NIRS, NIR-HSI, non-destructive methods, chemometric tools, agriculture applications

INTRODUCTION

Near Infrared Spectroscopy (NIRS) is a well known technology in the agricultural sector allowing the acquisition of chemical information from the samples (Rodriguez-Otero, et al., 1997). By NIRS, C-H, N-H and O-H bonds are induced to vibrate. This principle is used to identify and quantify components. NIRS allows the acquisition of the reflectance spectra of opaque milled or intact materials. NIRS is characterized by acquisition of a typical NIR spectrum which can be considered as the spectral signature or spectral fingerprint of the material. However because of NIRS low sensibility different functional groups may be detected in overtones and combinations bands (Burns and Margoshes, 1992). It is an adequate technique for the analysis of major components (chemical composition, microorganism detection and quantification) in agriculture with minimum sample preparation.

NIRS is a well known technology in the agricultural sector since the scientific works regarding soybean moisture of Norris in the 1960’s (Hart et al., 1962). In the last years, new areas of work based on the NIRS technology have been developed combining NIR systems with other techniques such as imaging or chemometric tools.
and a microscope to create the NIR microscopy (NIRM) (Baeten et al., 2012). More recently, NIR was combined with imaging technologies creating the Near Infrared Hyperspectral Imaging System (NIR-HSI) (Fernández Pierna et al., 2004; Fernández Pierna et al., 2012). ElMasry and Sun (2010) explained that «hyper» in hyperspectral means «over» or «too many» and it refers to the large number of wavelength bands measured. While, the «spectroscopy» word means «seeing», the «spectrometry» word means «measuring» and the hyperspectral word means «many bands». NIR-HSI system provides spectral and spatial informations from an object that forms a three dimensional “hypercube”, which can be used to extract physical and chemical information from an object. The images provide sufficient information to identify and distinguish each spectrum as a unique material. NIR-HSI has different instrumentation approaches: point (staring) scan, push-broom (line) scan or plane (global) scan (Fernández Pierna et al., 2009).

Fifteen years ago, only experts in remote sensing spectral images had access to hyperspectral systems or software tools to explore such images. In the last decade, hyperspectral image analysis was developed into one of the most powerful and fastest growing technology in remote sensing (Kavitha et al., 2012).

The objectives of this paper are to describe the advantages and disadvantages of NIRS and NIR-HSI and to display how a spectrum or an image can be analysed. Finally, some applications made by the NIRS and NIR-HSI in agriculture are listed.

ADVANTAGES AND DISADVANTAGES

Major advantages of spectroscopic techniques are: the ease of use, repeatability and reproducibility, reasonable start-up cost, non-polluting, non-invasive and non-destructive analyses and the possibility of online or directly in the field implementation. By NIR-HSI many spectra can be recorded simultaneously for one sample instead of a unique, average spectrum obtained when using classical NIRS (Fig. 1).

Fig.1. Acquisition of spectrum/spectra by (i) conventional NIRS system and (ii) Laboratory scale NIR-HSI system (source CRA-W, Gembloux, Belgium)

Legend: (i) NIRS system (Forage Crops Lab, USAMV Cluj, Romania); (i1) Sample support for NIRS system; (i2) Typical spectrum of NIRS system; (ii) Laboratory scale NIR-HSI system (source CRA-W, Gembloux, Belgium); (ii1) Photography of sample; (ii2) Hyperspectral image of sample; (ii3) Typical spectra of Laboratory scale NIR-HSI system

Hyperspectral images provide more information i. e. spectral and spatial information, than could be obtained when NIRS technology was used. One spectral image regroups numerous spectra. So each pixel of spectral image corresponds to one spectrum of the target sample. The spectrum or spectral signature can be used to characterize, classify or even identify any given material (Shaw and Manolakis, 2002).
The disadvantages of NIRS and NIR-HSI are: the relatively high price of instruments; the requirement of huge hardware speed; the necessity of calibration models for standardization; the possible presence of pixels (spectrum) in the image which do not contain any chemical information like for example bad, dead, noise, blinking or drifting pixels (Chang, 2000; ElMasry and Sun, 2010).

**SPECTRA ANALYSES**

Qualitative and quantitative analyses by NIRS usually require the application of calibration algorithms based on physico-chemical measurements. In order to get efficient qualitative and quantitative information from data coming from NIRS and NIR-HSI instrumentation, chemometric tools are necessary (Roggo et al., 2005). Chemometrics is the science of extracting relevant information from measurements made in chemical systems, using mathematical and statistical procedures (Massart et al., 1988).

The building of calibration models starts with spectra pre-processing treatment: after collection of spectra, it is necessary to perform a pre-treatment to remove high- or low-frequency interferences. Different types of pre-processing treatments are: polynomial baseline correction, Savitzky - Golay derivative, Standard Normal Variate (SNV), mean-centering and unit variance normalization among others (Gowen et al., 2007; ElMasry and Sun, 2010).

The most frequent pre-processing treatment used in practical are:
- SNV transformation removes the slope variation from spectra caused by scatter and variation of particle size (Candolfi et al., 1999).
- Derivative conversion, unimportant baseline signal from samples are removed by taking the derivative of the measured responses with respect to the variable number (index) or other relevant axis scale (wavelength, wavenumbers, etc.) (Wise et al., 2006). First derivative is usually used to remove any offset from the sample and de-emphasizing lower-frequency signals (Wise et al., 2006) while second derivative will accentuate the higher-frequency to enhance selectivity (Wise et al., 2006).

After spectra pre-processing treatment the calibration algorithms can be applied for classification and quantification. In order to analyse data, many multivariate analytical tools are used, such as: principal component analysis (PCA), principal component regression (PCR), multi-linear regression (MLR), partial least squares regression (PLS), modified partial least squares regression (MPLS), partial least squares discriminant analysis (PLS-DA), linear discriminant analysis (LDA), support vector machines (SVM), artificial neural networks (ANN), soft independent modeling of class analogy (SIMCA) baseline shift (BLS), spectral information and divergence (SID) (Chang, 2000; Fernandez Pierna et al., 2006; Gowen et al., 2007; Gómez - Sanchis et al., 2012).

The most frequent mathematical algorithms used in practice are:
- PCA: used for data compression and information extraction (Wise et al., 2006);
- MLR: allowed to establish a link between a reduced number of wavelengths (or wavenumber) and a property of the samples (Roggo et al., 2005) and to find a single factor that best correlates predictor (X) variables with predicted (Y) ones (Wise et al., 2006);
- PLS: used to establish a linear link between two matrices, the spectral data X and the reference values Y (Roggo et al., 2005); in other words, it attempted to find factors for both capture variance and also to achieve correlation (Wise et al., 2006), while PLS-DA is performed to discriminate between classes (McGovernin et al., 2011);
- ANN: it can be applied for pattern recognition, classification or clustering and quantitative modelling (Dolmatova et al., 1997);
- SVM: used for solving problems of nonlinear classification/regression, model estimation and density estimation (Fernandez Pierna et al., 2006).
The spectral pre-processing treatments and calibration algorithms are extensively reviewed in the literature (Dolmatova et al., 1997; Roggo et al., 2005; Fernandez Pierna et al., 2006; Wise et al., 2006; ElMasry and Sun, 2010; Yao and Lewis, 2010; Fernandez Pierna et al., 2011).

Assessment of calibration performance: NIR calibration models performances can be characterized by several parameters: standard error of calibration (SEC) or standard error of cross validation (SECV). To perform calibration model performance, an independent set of samples is used to get the standard error of prediction (SEP) and the squared coefficient of correlation (RSQ), which are used to describe the NIR analytical error when analyzing samples of unknown quantitative composition (Hartmann and Buning-Pfaue, 1998).

APPLICATIONS IN AGRICULTURE OF NIRS AND NIR-HSI

The first studies based on the NIRS technique were published between 1930 and 1940. By 1990, more than 1000 articles were published using this technique (Burns and Margoshes, 1992). If we look in the agriculture applications more than 1000 are carried out in 2012 (Google Scholar, consulted on 10.05.2012).

Summary of measurement mode, product type, wavelength region used and classification algorithm employed in papers published on NIRS and NIR-HSI applications in agriculture and agro-industries.

<table>
<thead>
<tr>
<th>NIRS system applications</th>
<th>NIR Hyperspectral Imaging applications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Products</strong></td>
<td><strong>Wavelength (nm)</strong></td>
</tr>
<tr>
<td>Forage quality:</td>
<td>1100-2500</td>
</tr>
<tr>
<td>- ash content</td>
<td></td>
</tr>
<tr>
<td>- digestibility</td>
<td></td>
</tr>
<tr>
<td>- protein content</td>
<td></td>
</tr>
<tr>
<td>Compound feedstuffs</td>
<td>400-2500</td>
</tr>
<tr>
<td>Wheat quality - protein content</td>
<td>1100-2500</td>
</tr>
<tr>
<td>Cereals:</td>
<td>1000-2500</td>
</tr>
<tr>
<td>- modifications made by filamentous fungi and yeasts</td>
<td>570-1100</td>
</tr>
<tr>
<td>- damages made by Fusarium culmorum</td>
<td>1100-2400</td>
</tr>
<tr>
<td>Detection of ergot bodies in wheat kernels</td>
<td>1000-2500</td>
</tr>
<tr>
<td>Detection of Fusarium in maize</td>
<td>1000-2500</td>
</tr>
<tr>
<td>Detection of ergot bodies in wheat kernels</td>
<td>1000-1000</td>
</tr>
<tr>
<td>Detection of hollow heart in potatoes</td>
<td>1000-2500</td>
</tr>
<tr>
<td>Detection of meat and bone meal in compound feeds</td>
<td>1000-2500</td>
</tr>
<tr>
<td>Detection of meat and bone meal in compound feeds</td>
<td>1000-2500</td>
</tr>
<tr>
<td>Detection and identification of bacteria in an isolated system</td>
<td>750-2500</td>
</tr>
<tr>
<td>PCA</td>
<td>PLSDA</td>
</tr>
</tbody>
</table>
The first use of NIR in agriculture was for the determination of moisture in soybean (Norris et al., 1976), in the case of HSI system remote sensing the first study concerned the detection and mapping of vegetation and minerals and for NIR - HSI the first case study was the detection of meat and bone meal in compound feeds (Fernandez Pierna et al., 2004). Nowadays NIRS and NIR-HSI are used in all agricultural and agro-industries domains from a large scale to a microscopic level. In table 1, some applications that were carried out in agriculture with NIRS and NIR-HSI systems, are enumerated, in order to illustrate the diversified board area of applications: animal nutrition, plant protection, food and feed quality and safety.

CONCLUSIONS

This short review intents to highlight that NIR spectroscopy and the most recent NIR-HSI systems are extremely reliable, non-destructive and rapid techniques for the prediction of quantitative and qualitative chemical and physical properties. However these techniques need the use of Chemometric tools, preprocessing treatment and assessment of calibration models in order to extract the maximum of information these techniques can provide.

Low-cost NIR-HSI systems will be necessary for future developments. Such as NIR-HSI that could easily identify optimal wavelengths/wavebands according to the different applications. For this reason more robust calibration and validation models are necessary in order to find a model that adequately represents the data. Reliability of the models will encourage more widespread online or on-field utilization of this technology (miniaturization or portable instruments) in agriculture and hence, could improve agricultural productivity and reduce the cost of process monitoring and product inspection.

REFERENCES


