



Cite this: *Chem. Soc. Rev.*, 2014, 43, 8200

# Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials

Marena Manley

Near-infrared (NIR) spectroscopy has come of age and is now prominent among major analytical technologies after the NIR region was discovered in 1800, revived and developed in the early 1950s and put into practice in the 1970s. Since its first use in the cereal industry, it has become the quality control method of choice for many more applications due to the advancement in instrumentation, computing power and multivariate data analysis. NIR spectroscopy is also increasingly used during basic research performed to better understand complex biological systems, *e.g.* by means of studying characteristic water absorption bands. The shorter NIR wavelengths (800–2500 nm), compared to those in the mid-infrared (MIR) range (2500–15 000 nm) enable increased penetration depth and subsequent non-destructive, non-invasive, chemical-free, rapid analysis possibilities for a wide range of biological materials. A disadvantage of NIR spectroscopy is its reliance on reference methods and model development using chemometrics. NIR measurements and predictions are, however, considered more reproducible than the usually more accurate and precise reference methods. The advantages of NIR spectroscopy contribute to it now often being favoured over other spectroscopic (colourimetry and MIR) and analytical methods, using chemicals and producing chemical waste, such as gas chromatography (GC) and high performance liquid chromatography (HPLC). This tutorial review intends to provide a brief overview of the basic theoretical principles and most investigated applications of NIR spectroscopy. In addition, it considers the recent development, principles and applications of NIR hyperspectral imaging. NIR hyperspectral imaging provides NIR spectral data as a set of images, each representing a narrow wavelength range or spectral band. The advantage compared to NIR spectroscopy is that, due to the additional spatial dimension provided by this technology, the images can be analysed and visualised as chemical images providing identification as well as localisation of chemical compounds in non-homogenous samples.

Received 5th February 2014

DOI: 10.1039/c4cs00062e

[www.rsc.org/csr](http://www.rsc.org/csr)

### Key learning points

- (1) Principles of and difference between NIR spectroscopy and NIR hyperspectral imaging.
- (2) Interpretation and visualisation of NIR spectra and images.
- (3) Multivariate data and image analysis for quantitative and qualitative analyses.
- (4) Food and non-food applications of NIR spectroscopy and NIR hyperspectral imaging.

## Introduction

The first non-visible region in the absorption spectrum, *i.e.* near-infrared (NIR), was discovered in 1800 by Frederick William Herschel, a professional musician and astronomer.<sup>1</sup> The NIR region was, however, not considered significant until 150 years later. During this time, analytical techniques that

could provide more unambiguous results were favoured over NIR spectroscopy, especially in terms of the explanation of molecular structures (*e.g.* mid-infrared (MIR) spectroscopy). The principles of chemical or gravimetric methods such as Kjeldahl for protein determination and oven drying for moisture analysis, respectively, was also more clearly understood at the time. The subsequent revival of the NIR region and development of NIR technology (since 1949) have been documented by Karl Norris,<sup>2</sup> known to have pioneered NIR spectroscopy development. The greatest impact on the progression of NIR technology

Department of Food Science, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. E-mail: [mman@sun.ac.za](mailto:mman@sun.ac.za)



was, however, in the early 1970s, when Phil Williams started using NIR spectroscopy to measure protein and moisture contents as a basis for trading wheat. The development of NIR technology, from its initial discovery in 1800 until 2003, highlighting the most important citations dealing with NIR technology during this time, has been extensively reviewed by McClure.<sup>3</sup>

In spite of being a secondary method (*i.e.* requiring reference values for the purpose of calibration model development), NIR spectroscopy is now considered equally significant among other major analytical technologies. NIR spectroscopy is, in contrast to most other analytical (*e.g.* gas and high performance liquid chromatography) and conventional chemical (*e.g.* Kjeldahl, Soxhlet) methods, rapid, chemical-free, easy to use (once calibrations have been developed) and non-destructive. Although the accuracy of the NIR method depends to a great extent on the accuracy and precision of the reference method, NIR measurements and predictions are considered more reproducible.

NIR spectroscopy is applied as a tool during process analytical technology (PAT) and quality control (QC) as the method of choice in various fields, *i.e.* agriculture,<sup>4</sup> food,<sup>5</sup> bioactives,<sup>6</sup> pharmaceuticals,<sup>7</sup> petrochemicals,<sup>8</sup> textiles,<sup>9</sup> cosmetics,<sup>10</sup> medical applications<sup>11</sup> and chemicals such as polymers.<sup>12</sup> NIR spectroscopy is also increasingly used in aquaphotomics,<sup>13</sup> which has been introduced as a new approach to describe and visualise the interaction of water with solvents with visible and near-infrared (vis-NIR) light absorbance patterns.

Similar to NIR spectroscopy, imaging technology, is not new. The term 'hyperspectral imaging' was first used by Goetz *et al.*<sup>14</sup> for remote sensing (*i.e.* the observation of a target by a device without physical contact) applications.<sup>14,15</sup> It was only by the late 1990s that this technology became available for applications in food and agriculture, when it was being applied in association with NIR spectroscopy.<sup>15</sup> It is known that NIR spectroscopy only provides a mean spectrum (average measurement) of a sample, irrespective of the area of the sample scanned. As the spectra

collected are averaged to provide a single spectrum, the information on spatial distribution of constituents within the sample is thus lost. The development of NIR hyperspectral imaging, which combines NIR spectroscopy with digital imaging, enables both spatial (localisation) and spectral (identification) information to be obtained simultaneously. Hyperspectral images thus have the potential of describing distribution of constituents within a sample. The use of NIR hyperspectral imaging has been and is still being investigated extensively to determine quality and safety of agricultural and food products.<sup>15</sup> Other fields of interest and research areas where NIR hyperspectral imaging is increasingly applied include pharmaceuticals,<sup>16</sup> medical applications,<sup>17</sup> archaeology<sup>18</sup> and palaeontology.<sup>19</sup>

This *tutorial review* will focus on NIR spectroscopy and NIR hyperspectral imaging analysis of biological materials. The first section will introduce the basic principles of these two techniques, followed by an overview of multivariate data and image analysis techniques for both quantitative and qualitative analysis. The last section will review applications within the respective fields.

## Fundamental principles of near-infrared spectroscopy

Spectra in the NIR region result from energy absorption by organic molecules, and comprise overtones and combinations of overtones originating from fundamental bond vibrations (stretching or bending) occurring in the mid-infrared (MIR) region of the spectrum. The features in NIR spectra of organic compounds are thus orders of magnitude weaker than those in the MIR. Overtones can be found by dividing the wavelengths in the infrared region by approximately 2, 3 or 4 and provide the advantage of a dilution series. NIR spectra can thus be collected directly on samples without the need for dilution, enabling direct analysis of solid samples. Ease of sample preparation and presentation results in NIR spectroscopy often being used in favour of MIR spectroscopy.

The NIR region extends from 800 to 2500 nm (12 500 to 4000 cm<sup>-1</sup>; 120 to 375 THz), between the visible from 380 to 780 nm (26 316 to 12 820 cm<sup>-1</sup>; 385–790 THz) and MIR from 2500 to 15 000 nm (4000 to 400 cm<sup>-1</sup>; 30 to 120 THz) regions. NIR spectra contain information about the major X–H chemical bonds, *i.e.* C–H, O–H and N–H. All molecules containing hydrogen will have a measurable NIR spectrum, resulting in a large range of organic materials to be suitable for NIR analysis.

Due to the overtone and combination modes and large numbers of possible vibrations, NIR spectra are very complex, consisting of many overlapping peaks (referred to as 'multicollinearity'), which result in broad bands. This makes it difficult to interpret NIR spectra visually, assign specific features to specific chemical components or extract information contained in the spectra easily. It was, however, realised early on that, with the use of appropriate regression techniques, relationships between absorption values at specific wavelengths and reference values of the constituent to be predicted could be established. Specific chemical



**Marena Manley**

*Marena Manley was born in 1961 and grew up in the Northern Cape, South Africa. She obtained her BSc in Food Science degree at Stellenbosch University and subsequently her Honours and MSc degrees at the University of Pretoria. She received her PhD (in near-infrared spectroscopy as applied to wheat hardness) in 1995 from the University of Plymouth, UK. After 18 months in industry, she joined Stellenbosch University in 1997 as a lecturer*

*where she is currently appointed as Professor. Her research interests involve the application of near-infrared (NIR) spectroscopy, NIR hyperspectral imaging and lately also X-ray micro-computed tomography to study whole grain endosperm texture and other cereal defects.*



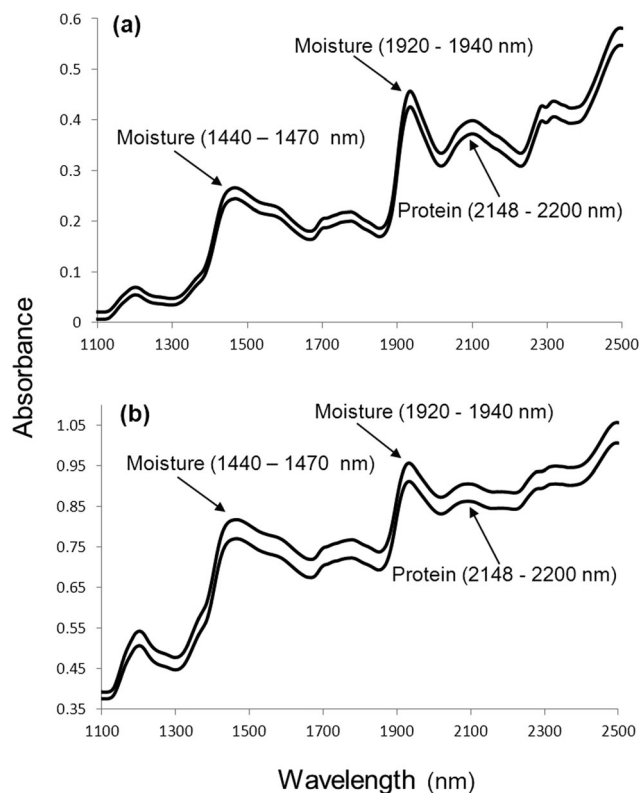


Fig. 1 Typical NIR spectra of a biological material, in this case (a) ground and (b) intact whole wheat with moisture (1440 to 1470 nm and 1920 to 1940 nm) and protein (2148 to 2200 nm) absorption bands indicated.

constituents are usually identified by a spectral band or more than one wavelength. Towards the end of the 1960s, Norris<sup>2</sup> proposed the use of multiple linear regression (MLR) to analyse NIR spectra, which resulted in NIR spectroscopy drawing attention of researchers as a practical non-destructive quantitative analytical technique. In the 1970s, this type of data analysis method used for spectral analysis became known as chemometrics. With the invention of the computer and its subsequent development, chemometrics has developed into a research field in its own right, which has affected the analysis of NIR spectral data significantly.

In spite of NIR spectra comprising overlapping peaks and broad spectral bands, visual spectral interpretation remains vital before any data analysis is performed. Fig. 1 illustrates moisture (O–H stretch first overtone; 1440 to 1470 nm and combination of O–H stretch and O–H deformation, O–H bend second overtone; 1920 to 1940 nm) and protein (N–H bend second overtone, combination of C–H stretch and C=O stretch; combination of C–O stretch, N–H in-plane bend and C–N stretch; 2148 to 2200 nm) absorption bands for ground and whole wheat. Osborne *et al.*<sup>20</sup> contributed significantly to the interpretation of spectra with a detailed list of molecular bonds (related to chemical substances) and the corresponding wavelengths in the NIR region where these bonds absorb. Fig. 2 shows spectra of the ground powder of an herbal tea (honeybush), ground black pepper and olive oil depicting absorption of O–H (moisture; in

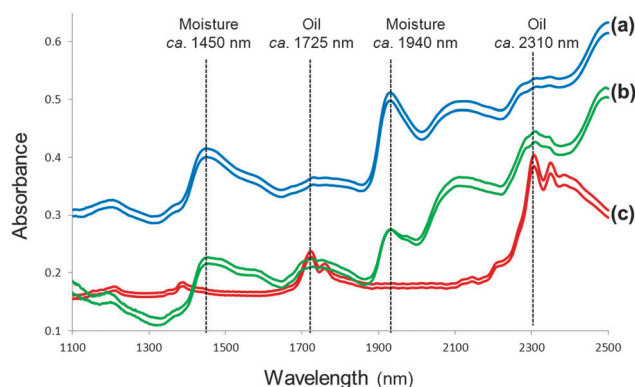


Fig. 2 Spectra of (a) an herbal tea, (b) ground black pepper and (c) olive oil with absorption of O–H (moisture) and C–H (oil) molecular bonds indicated.

the tea and black pepper) and C–H (oil; in the olive oil and to some extent in the black pepper) molecular bonds. Extracting information from NIR spectra, however, remains a challenge. Reliable data analysis can only be performed once it has been ensured that the originally collected spectra are of good quality with a high signal-to-noise (S/N) ratio.

NIR spectra from at least 100 or ideally more samples should be collected for calibration model development. Before any measurements are made, it is important to optimise the preparation and presentation of the sample to the instrument depending on the application *e.g.* milling, drying, freeze drying. Sample size, orientation (*e.g.* in the case of single grains) and environmental conditions should also be optimised. To ensure a high S/N ratio 64 or more scans should be recorded per sample. Esteve Agelet and Hurburgh<sup>21</sup> emphasised good practices to be followed during sample and spectra collection. NIR spectroscopy being a secondary method requires accurate reference analysis. This might require duplicate analysis that would enable determination of the standard error of the laboratory (SEL), a useful validation statistic to determine the accuracy of the NIR model standard error of prediction (SEP) in comparison to the reference method.

An often overlooked advantage of NIR spectroscopy is that a number of predictions can be made from a single collected spectrum if the sample preparation was the same when the calibration models were developed for these properties. The measurement of *e.g.* moisture content and chemical composition (*e.g.* protein, fat, active or bioactive components) is thus possible from a single spectrum. Especially taking the specificity of traditional wet chemistry methods into consideration, only one property can be measured at a time. These methods are usually destructive; thus, each property or constituent is measured on a different sample. In the case of NIR spectroscopy, all properties can be measured on the exact same sample.

## Fundamental principles of near-infrared hyperspectral imaging

The main advantage of NIR hyperspectral imaging is that it facilitates visualisation of the distribution of different chemical





components in a sample. Due to a spectrum collected at each pixel in the image, it is most suitable for analysis of heterogeneous samples. Whereas the initial imaging systems (staring) required the samples to be stationary, newer systems (pushbroom) collect images from moving samples, enabling online analysis. The main disadvantages of NIR hyperspectral imaging include it being costly, especially when wavelengths up to 2500 nm are required. Wavelengths from 1100 to 2500 nm requires the more expensive indium gallium arsenide (InGaAs)-based or mercury cadmium telluride (HgCdTe)-based array detectors whereas for wavelengths up to 1100 nm the lower cost silicon-based detectors can be used. Data collection and analysis requires sensitive detectors and fast computers, respectively and substantial data storage capacity is required due to the size of the hyperspectral images. As for NIR spectroscopy the challenge of extraction of only useful information from the large data sets and the complexity of the spectra remains.

The pixels of a digital colour image comprise a combination of primary colours. An RGB image will thus have red, green and blue channels (Fig. 3). A greyscale image has just one channel. An NIR hyperspectral image, obtained when NIR spectroscopy is combined with greyscale digital imaging, comprises single channel images. Each of these greyscale images represents an individual wavelength, and is stacked consecutively to form a hyperspectral image. NIR hyperspectral images are acquired at wavelengths in the NIR region.<sup>15</sup> The collected image data is arranged into a three-way data matrix (or hypercube). The first two axes ( $x$  and  $y$ ) of the matrix are the vertical and horizontal pixel coordinates (spatial dimension), while the third ( $z$ ) axis represents the spectral dimension (wavelengths). The obtained hypercube with its spatial and wavelength dimensions contains an NIR spectrum for each pixel in the image (Fig. 4). Each pixel within an NIR hyperspectral image thus represents a single spectrum, in principle different to its neighbour. Due to the added spatial dimension, spectral (chemical or physical) information is obtained for each pixel in the image. NIR hyperspectral imaging is therefore highly suitable for analysis of samples of heterogeneous nature. From a hyperspectral image, the distribution of constituents (that absorb in the NIR wavelength region),

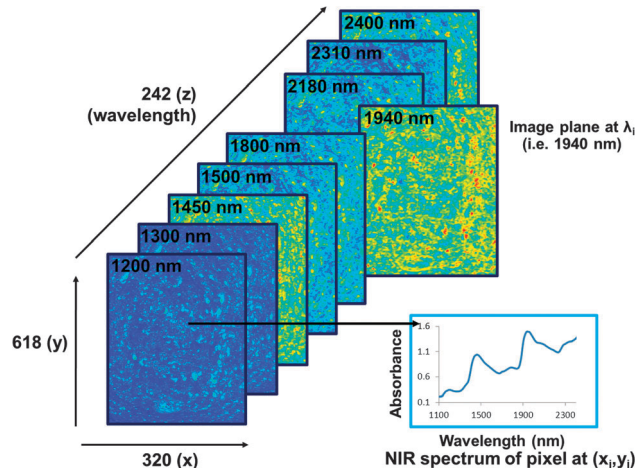


Fig. 4 Illustration of an NIR hyperspectral imaging hypercube comprising wavelength ( $z$ ) and spatial ( $x$  and  $y$ ) dimensions. The spectrum of each pixel can be visualised as well as the image plane at each respective wavelength. Image planes are shown at a few selected wavelengths. This image was acquired from a slice of fresh bread with the broad moisture absorption bands clear in the spectrum of one of the pixels and the variation due to moisture illustrated in the image plane at ca. 1450 and 1940 nm. Different colours indicate different chemical absorptions or components.

as reflected in the spectra at each pixel, within a sample can be determined and visualised.

Hyperspectral images can be acquired using either of two configurations,<sup>15</sup> *i.e.* the staring imager and the pushbroom or linescan system. With the staring imager, whole images are acquired consecutively, one wavelength at a time using either a liquid crystal tunable filter (LCTF) or an acousto-optic tunable filter (AOTF). A disadvantage, though, is that, during the time required to record the wavelengths one by one, changes in the sample can take place. Collecting images with the staring imager takes a few minutes, the samples must be stationary and although high spatial resolution is possible, the images will have a lower spectral resolution. With the linescan or pushbroom system, all spectral information is acquired simultaneously. This is done line by line and requires the sample to move relative to the instrument. Linescan images provide a good compromise between spatial and spectral resolution and an image can be collected within a few seconds. Lately most systems use the much faster pushbroom configuration that simulates movement of samples along a conveyor belt. The pushbroom configuration is thus ideally suited for on-line quality control. It is also possible to collect images point by point (whiskbroom imager) which results in high spectral but low spatial resolution. These systems are, however, not suitable for real-time analysis as it takes more than an hour to collect an image. As for NIR spectroscopy, good imaging practices should be followed as suitably reviewed by Boldrini *et al.*<sup>22</sup>

It is possible to obtain useful information from the image even before any data analysis is applied. Apart from the differences or similarities between spectra at each pixel, differences between image planes at respective wavelengths (Fig. 4) can also be determined. This enables visualisation of chemical (or physical)

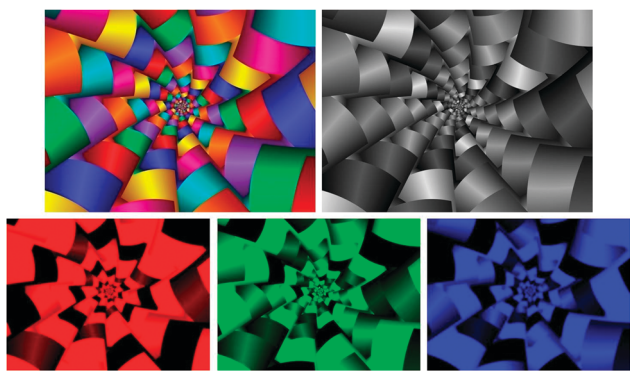


Fig. 3 A colour digital image with a greyscale image (top), consisting of only one channel. A colour image is a combination of the primary colours: three channels in red, green and blue creating an illusion of colour (bottom).



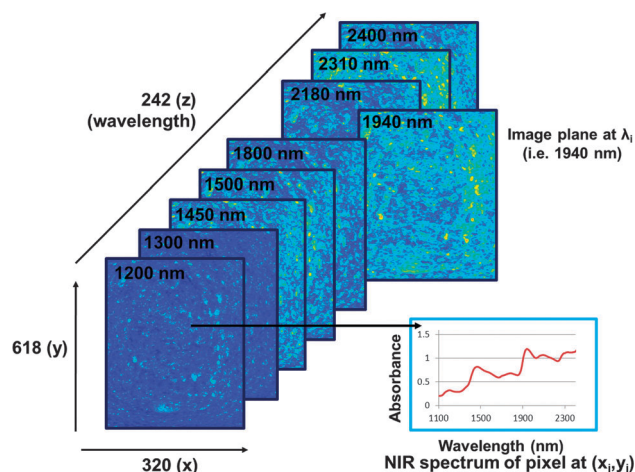


Fig. 5 The NIR hyperspectral imaging hypercube from the same slice of bread as in Fig. 4, but after it has been dried for 1 h at 60 °C. The decrease in absorption at the moisture bands can be seen in the spectrum of one of the pixels when compared to that in Fig. 4 (see also Fig. 6). Similarly the reduced variation due to moisture is clear in the image planes at ca. 1450 and 1940 nm. Different colours indicate different chemical absorptions or components.

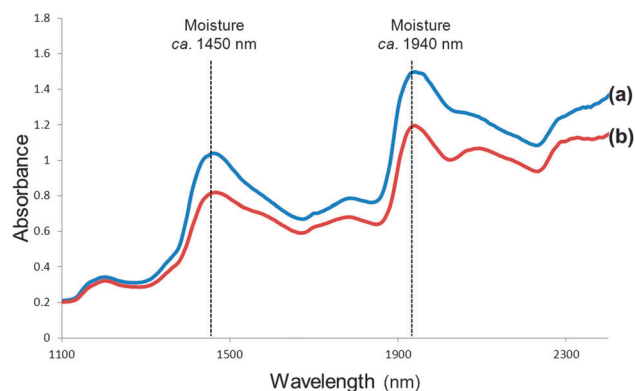


Fig. 6 Average spectra of small areas of the image of the slice of bread (a) fresh and (b) after being dried for 1 h at 60 °C, illustrating reduced absorption at the respective moisture bands (1440 to 1470 nm and 1920 to 1940 nm).

information and potential identification of the chemical component of interest. Fig. 5 shows image planes (of the hypercube) at selected wavelengths for the same slice of bread than that in Fig. 4 after it has been dried for one hour at 60 °C. The loss of moisture can clearly be seen in the image planes at ca. 1450 and 1940 nm when compared to the same image planes in Fig. 4. Fig. 6 shows average spectra of the images of the slice of bread before and after drying illustrating the reduced absorption at the respective moisture bands (1440 to 1470 nm and 1920 to 1940 nm).

## Multivariate data analysis

NIR spectroscopy data is multivariate in nature due to a large number of data points (one at each wavelength) being collected for each sample during spectral collection. NIR data analysis is also complicated by the overlapping peaks. Apart from chemical

information, physical properties are also reflected in the spectra. Differences between samples result in only small spectral differences. Multivariate data analysis or chemometrics is thus required to extract suitable information from the spectra that would correlate with the measured property (e.g. protein, fat, moisture) under investigation. Mathematical procedures are used to remove unwanted information (such as spectral noise or effect of particle size) without losing important or required information. During the development of NIR models independent variables (measured absorbance values at all wavelengths) *i.e.* the X-matrix are correlated with the Y-matrix (concentration values of samples needed to be predicted in future). Thus due to the complexity of NIR spectra and little spectral differences between different samples, multivariate data analysis (spectral pre-processing followed by model development) is essential for effective use of NIR spectroscopy as an analysis technique.

## Spectral pre-processing

Pre-processing or pre-treatment of spectral data are often required to reduce noise or unwanted background information and increases the signal from the chemical information.<sup>23</sup> The noise levels detected by the instrument thus have to be kept to a minimum to compensate for this lower sensitivity and to ensure a high S/N ratio. Noise can be introduced during sample preparation and presentation, conditions under which the spectra are collected and by instrument drift during scanning. The weakly absorbing bands of NIR spectroscopy are much more significantly affected than those regions that produce strong absorptions. Conditions of spectral collection should thus be carefully controlled. A common practice to remove noise is to collect and average multiple scans. To reduce the noise levels significantly usually requires 64 or more repeated scans. If not adequate, smoothing techniques may be used.

The application of spectral pre-processing methods improve the subsequent data analysis (exploratory analysis, calibration and classification model development) and may be scatter correction methods that also adjust baseline shifts and derivatives.<sup>23</sup> Derivatives always apply a smoothing step before calculating the derivative. The most common pre-processing methods, as reviewed by Rinnan *et al.*,<sup>23</sup> include the moving-average method (Savitzky–Golay), normalisation, derivatives (Savitzky–Golay), multiplicative scatter correction (MSC) and standard normal variate (SNV).<sup>21</sup> MSC and SNV are two methods well known for their ability to correct for spectral distortions due to multiplicative scattering,<sup>23</sup> commonly noticed when samples consist of particles differing in size. Different particle sizes cause scattering which results in additive variation of the spectra baseline intensity (baseline slope) as the wavelengths increase. Derivatives (such as Savitzky–Golay) can also correct for this.<sup>23</sup> In addition, derivatives could also solve the most common problem of NIR spectroscopy, *i.e.* overlapping peaks.

Mean centring of spectra is a pre-processing technique mostly used with principal component analysis (PCA).<sup>21</sup> It entails the calculation of the average spectrum of the data set with subsequent subtraction of this average from each spectrum. Another benefit of mean centring is that it reduces the number



of variables to be used; the final model then becomes much less complex. PCA is an explorative data analysis technique usually employed to view data for inconsistencies and outliers before regression techniques are applied.

The selection of the most appropriate spectral pre-processing and regression methods is usually done through trial and error. Calibration and prediction statistics are evaluated to select the most accurate calibration model with an optimum number of components.

### Exploratory NIR spectroscopy data analysis

Unsupervised methods, *e.g.* PCA, are mostly used as investigative tools in the early stages of data analysis to determine possible relationships between samples. PCA is used to describe the majority of the variation in multivariate data sets. It compresses the data by constructing new variables. Scatter plots of *e.g.* the first two new variables can be used to visualise the relationship between samples in the multivariate space.<sup>24</sup> PCA is applied to spectral data only and no prior knowledge on the chemical composition of the sample is required. Investigating the PCA scores plots, samples deviating from others, either due to concentration (mistakes when performing reference method or copying data) or spectral (samples belonging to a different population or instrument not functioning properly) differences can *e.g.* be identified for further investigation.

### Quantitative NIR spectroscopy multivariate data analysis

The application of NIR spectroscopy as an analytical technique would not be possible without the use of multivariate data analysis (chemometrics). Chemometrics is firstly used to resolve overlapping peaks bands and broad spectral bands in NIR spectra. This enables the technique to be used for quantitative measurement after appropriate and reliable calibration models have been developed.<sup>20</sup> Fig. 1 illustrates typical broad bands found in NIR spectra of ground and intact whole wheat. NIR absorption bands occurring at longer wavelengths are usually better resolved than absorption bands at the shorter wavelengths. Overlapping peaks in an NIR spectrum limits the use of only a single wavelength. Using all spectral variables (wavelengths) will result in improved models, but it requires multivariate regression techniques such as principal component regression (PCR) or partial least squares (PLS) regression. These two regression methods are based on the principle that only a few linear components (combinations) within the spectral data can be used in the regression equation. These components are determined using appropriate mathematical calculations. In this way only the most relevant information from the spectral data set is used for the regression or calibration model development.

The need for calibration model development means that NIR spectroscopy is a secondary method and that the accuracy of the method depends to a large extent on the accuracy and repeatability of the reference method. The first step in quantitative NIR calibration model development thus involves acquiring a set of calibration or training samples with known reference values (chemical constituents, physical characteristics or other indirect properties) covering the range of variation expected in

unknown samples to be analysed in future.<sup>20</sup> To develop a calibration model, a mathematical relationship must be established between the NIR spectra and the respective reference values previously determined by an independent analytical method for each sample. This can be done on either raw or pre-processed spectra. Unless scattering properties are contributing to the property measured, pre-processed spectral data are most commonly used.

The property to be measured during NIR calibration model development should be either of organic nature (*e.g.* moisture or protein that will absorb in the NIR region through direct measurement), be correlated with a physical characteristic (*e.g.* particle size), or it should be a compound that does not absorb in the NIR region but can be measured through co-variation or in other words indirect measurement (*e.g.* salt content). The aim of model development is to fit the NIR spectral data and reference values to a straight line and to compare it statistically to a theoretically perfect line through the origin at 45° to both axes.<sup>20</sup> This calibration model, after being adequately validated on an independent validation set, can then be used to predict the properties or constituents in unknown samples on the basis of their NIR spectra. Regression methods commonly used are multiple linear regression (MLR), which utilises only selected wavelengths, PCR and PLS regression. MLR is the easiest way to perform an inverse multivariate calibration based on least square fitting of the reference to the spectral data. Although it is usually applied when only a limited number of discrete wavelengths are available, it can also be applied to full spectrum data sets. It is most often applied according to the stepwise forward method where the first wavelength to be selected will be the one with the highest correlation. The regression then finds the next wavelength, which will increase the correlation (*e.g.* coefficient of determination ( $R^2$ )) and reduce the error (*e.g.* standard error of prediction (SEP)). The process stops when addition of another wavelength has no effect or starts to reduce the correlation and increase the error. PCR or PLS regression both use the whole spectrum to calculate linear combinations (components) for regression modeling. More detail on the principles of these methods can be found in Osborne *et al.*<sup>20</sup> and Næs *et al.*<sup>24</sup>

During development of multivariate calibrations, a crucial aspect to consider is the correct selection of calibration and validation samples.<sup>21,25</sup> Validation samples should have no influence on the calibration procedure or selection of best calibration model; it should thus not be used to select the optimum number of components. Validation sets should be collected from experiments different to that of the calibration set. *E.g.* it should be agricultural samples from a new harvest season or chemical samples from a new batch. If a validation set is not selected to be completely independent, the predictive performance of a calibration, validated in this manner, is likely to be overestimated, as could be the case when using cross-validation. During cross-validation a single sample or groups of samples are consecutively removed from the calibration set and used as validation samples during a number of prediction iterations and the standard error of cross-validation (SECV) or





root mean standard error of cross-validation (RMSECV; corrected for bias) reported. For efficient validation of a calibration model, an entirely independent validation set should be used.<sup>25</sup> Finally, an important aspect when developing NIR calibration models, is the correct reporting of calibration and prediction statistics for efficient interpretation of the repeatability and accuracy of the developed calibration model.<sup>25</sup> Prediction statistics that are important to report include standard error of prediction (SEP) or root mean standard error of prediction (RMSEP; corrected for bias) and coefficient of determination ( $R^2$ ). For interpretation of accuracy of prediction models it is also advisable to consider the standard error of laboratory (SEL) as indication of reproducibility of the reference method. The RPD which is the ratio of the standard error in prediction to the standard deviation (of the validation samples) is also advisable to use to illustrate suitability of prediction models. It attempts to scale the error in prediction with the standard deviation of the property. RPD values greater than 3 are useful for screening, values greater than 5 can be used for quality control, and values greater than 8 for any application.

Reasons have been identified that could result in decreased accuracy of NIR calibration models.<sup>34</sup> (1) A narrow range in the variability of the reference values (*i.e.* low SD) is known to impact negatively on NIR predictability. In a quality control environment, it is difficult to obtain samples with a wide range of variability and this problem is thus not easy to solve. (2) Analytical differences exist when using *e.g.* the Kjeldahl method to measure nitrogen content. This could affect crude protein predictions. Similarly, if there are large errors or if poor reproducibility is observed for the reference method it would reduce NIR prediction accuracy. (3) If NIR spectra are collected from intact samples (preferred for commercial on-line measurement) it could result in reduced accuracy due to the heterogeneity of the samples. Although not ideal, chemical composition can be predicted more accurately on homogeneously milled or minced than on intact samples and even more so if a finer sieve size or grind is used. (4) Determination of minerals remains a challenge. Similar to salt (NaCl), pure minerals or inorganic compounds do not absorb in the NIR region. Measurement of *e.g.* ash content is thus possible due to associations of the mineral content with the organic fraction of the sample or by forming salts that modify the spectra, most likely the water bands.

### Qualitative NIR spectroscopy multivariate data analysis

Qualitative analysis by NIR spectroscopy is usually required to discriminate between different classes of a commodity (*e.g.* a given food ingredient), confirm the authenticity of another (*e.g.* a pure olive oil) or detect an adulterated food.<sup>26</sup> Solving these type of problems requires the comparison of the authentic material with the unknown sample. Using NIR spectroscopy involves the collection of typical NIR spectral signatures of the commodity (often a food) under investigation. Due to variation in biological materials, one typical spectrum does not exist for each class.<sup>26</sup> Variation is caused by different varieties, geographical area and seasons. A library of representative spectra (ideally containing all possible variation) thus needs to be collected

to be compared with the spectrum of the unknown sample. The application of NIR spectroscopy for the confirmation of authenticity of foods and food ingredients and discrimination between different classes of foods has been appropriately reviewed by Downey<sup>26</sup> and Manley *et al.*<sup>27</sup>

Qualitative multivariate data analysis techniques or pattern recognition methods, compare NIR spectra and search for similarities or differences within the spectra.<sup>23,26</sup> The aim is to develop classification models that would give as many correct classifications as possible. The first step is always to determine the number of classes to be considered and the specific requirements that a sample has to fulfil in order to be assigned to a certain class.

Two different approaches can be used during qualitative applications of NIR spectroscopy, *i.e.* unsupervised and supervised.<sup>27</sup> When using supervised methods, the classes of the sample set used for classification model development (*i.e.* the training set) are known beforehand whereas, in unsupervised methods, there is no information available about the class structure. Unsupervised methods, *e.g.* PCA, are mostly used as investigative tools in the early stages of data analysis to determine possible relationships between samples. Supervised methods commonly used are soft independent modelling of class analogy (SIMCA), linear discriminant analysis (LDA), multiple discriminant analysis (MDA), factorial discriminant analysis (FDA), PLS discriminant analysis (PLS-DA), canonical variate analysis (CVA), artificial neural networks (ANNs) and k-nearest neighbour (k-NN) analysis. These methods, as applicable for authentication studies, have been summarised by Manley *et al.*<sup>27</sup> Detailed explanations of these and other qualitative (classification) techniques such as support vector machine (SVM) classification can be found in Næs *et al.*<sup>24</sup>

## Multivariate image analysis

A single hyperspectral image can consist of up to 200 000 spectra. Multivariate image analysis (MIA) techniques are required to handle such large data sets.<sup>28</sup> The input data for MIA is usually a hypercube, but it can also be a mosaic (number of combined hypercubes).<sup>28</sup> Once a hypercube has been selected or a mosaic constructed, a number of MIA techniques may be applied. These techniques are usually applied in a specific sequence and repeated a number of times, with changes in *e.g.* pre-processing techniques, until the optimum regression or classification model has been developed. The image analysis sequence usually starts with cleaning of the image, which involves removal of unwanted background and correction of shading effects. This is followed by exploratory analysis (*e.g.* PCA) before regression (*e.g.* PLS) or classification (*e.g.* PLS-DA) models are developed. Because of the huge amount of available data, model development results can successfully be visualised by means of plots (*e.g.* PCA scores plots) and images (*e.g.* PCA scores images or PLS-DA classification images). If nonlinear regression modelling needs to be addressed, artificial neural networks (ANN) may be considered.



### Quantitative NIR multivariate image analysis

MIA may be applied to raw images or pre-processed images. Pre-processing techniques are used to reduce noise (thus increase S/N ratio) and to remove any irrelevant information. These pre-processing methods have been described earlier<sup>23</sup> and were reviewed in more detail for imaging applications by Nicolai *et al.*<sup>29</sup> Images always contain errors which should be efficiently removed before MIA is applied. Errors may be due to shape of the sample, causing shading effects, camera and optical related errors, with dead pixels present in the detector and background, all of which are unrelated to the property of interest for which a model needs to be developed. It is thus essential to spend adequate time to remove these errors and unwanted information (including background) before any modelling is started. Background can be removed by identifying a threshold after the subtraction of the low reflectance image value from that of the high reflectance image. PCA scores plots (the background and other effects such as shading would form clusters in the scores plot) and scores images can also be used interactively by means of brushing.<sup>28</sup> By selecting the clusters in the scores plot the associated scores would be highlighted in the scores images, enabling visual confirmation of the cluster to be removed *e.g.* background. This data (spectra at respective pixels) can then be removed from the data set.

The same principles as for NIR spectroscopy are employed when the regression techniques MLR, PCR and PLS are applied to hyperspectral images.<sup>28,29</sup> However, in contrast to NIR spectroscopy, the number of samples (or spectra) used for image regression models is much larger (*ca.* 200 000) than the number of variables (*ca.* 240). In the case of NIR spectroscopy these two are almost the same (*ca.* 200). The large number of available spectra enables representative selection of calibration (training) and validation (test) sets. The main advantage of images is that all samples (spectra) have spatial coordinates.<sup>28</sup> This makes construction of classification and prediction images possible. These images can be visually inspected and interpreted. A disadvantage of multivariate image regression, though, is that in principle, the reference values for all the calibration samples (thus each spectrum at every pixel) should be known. Determination of these values at each pixel position in an image, using traditional methods (wet chemistry) is not feasible. An average reference value (obtained from the whole imaged sample) is usually used. This limitation needs to be considered when NIR hyperspectral imaging is used for quantitative measurement and improved methods might need to be developed in future.

The success of MIA depends mostly on the quality of the spectral data, image cleaning and data pre-processing. Due to the vast amount of data available, the assessment of regression models or prediction results can be visualised as histograms and concentrations or heat maps (graphical representation of data as colours).<sup>28</sup> MIA thus provides a powerful tool for increasing the evaluation and understanding of sample constituent concentrations and their distribution or spatial variation throughout the sample matrix.

### Qualitative NIR multivariate image analysis

Due to the added spatial dimension and its applicability to heterogeneous samples, NIR hyperspectral imaging is most suitable for authentication, discrimination and classification applications. Qualitative modelling requires discrimination between two classes of samples (or groups/clusters of pixels in the case of images).<sup>28</sup> PLS may be applied as a discriminative technique for qualitative modelling, *i.e.* PLS discriminate analysis (PLS-DA).<sup>27</sup> The difference between the two techniques is that the wet chemistry reference values are replaced with dummy variables, *e.g.*  $-1$  and  $+1$ , each referring to either of the two classes (selected groups of pixels). This is a supervised clustering method, because it is known in advance which pixels are  $-1$  and which are  $+1$ . When using  $-1$  and  $+1$  as dummy variables,  $0$  may be chosen as a cut-off for class membership. It is essential to use a completely independent test set to evaluate the prediction ability of the classification model.

Clusters can also be obtained with PCA, an unsupervised method, using distances between the samples (pixels) in the multivariate space. PCA is well suited for hyperspectral images as it can handle many spectra (pixels) at a time and can also be used for classification.<sup>28</sup> The benefit of applying PCA, which is also a data reduction technique, is that it reduces the data set comprising 100 000s spectra to a smaller number of latent variables for further usage and/or interpretation. Classification results can be visualised as principal component (PC) scores images, PC scores plots, classification plots and classification images. Fig. 7 illustrates how multivariate image analysis enables visualisation of results. The PC scores image enables visualisation of similarity in samples by means of a heat map. In this case similar colours refer to similar score values which can be interpreted as characteristics, *e.g.* similar chemical composition. This allows one to distinguish between, in this case, hard (H) and soft (S) maize kernels. The PC scores plot shows clusters illustrating similarity between spectra (pixels) based on distances in the multivariate space. These clusters can be assigned dummy variables and can be shown as a classification plot and subsequently be projected onto the scores image to form a classification image. In this case, the clusters in the PCA scores plot obtained were due to differences in endosperm texture of whole maize kernels. Relevant information in imaging data is often only being observed in lower-order principal components and not in *e.g.* principal component one. It is thus important to evaluate these components also.

Successful implementation of NIR calibrations, whether quantitative or qualitative, requires robust calibration models, large datasets including inherent variation, availability of powerful computers, optimised spectral pre-processing methods and suitable regression techniques, such as PLS.<sup>20,24</sup> Variation included in data sets could comprise *e.g.* several geographical areas, varying climatic conditions, seasons and scanning conditions, such as temperature of the sample when collecting the spectra. Another crucial aspect, not considered often enough in any NIR application is that protocols must be put in place to enable and ensure regular maintenance and updates of calibration models. More research





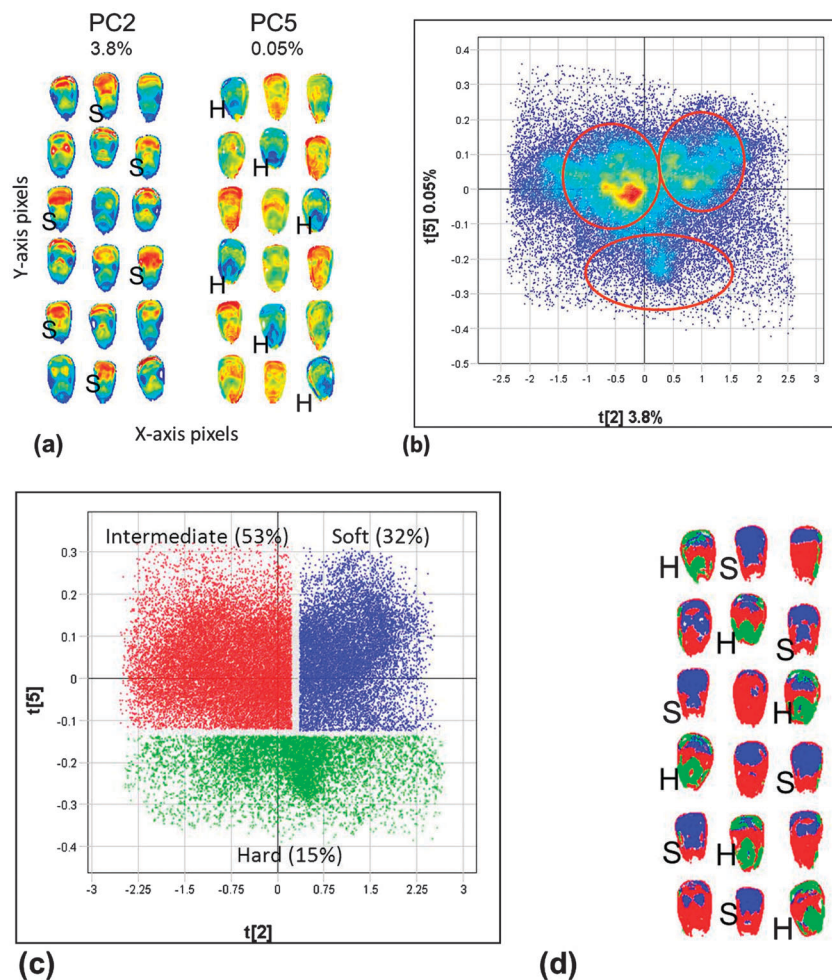


Fig. 7 (a) Scores images of PC2 and PC5 for whole yellow maize kernels enabling visualisation of similarity in chemical composition (similar colours indicate similar chemical composition, in this case similar endosperm texture). (b) Scores plot of PC2 vs. PC5 with three clusters. (c) Classification plot based on clusters identified in the PC scores plot. (d) Classification image after projection of the classes identified in the scores plot onto the scores image.

should also be performed to understand calibration techniques in terms of the physics of NIR light propagation in the sample better.<sup>29</sup>

## Fields of application

The use of shorter wavelengths in the NIR region results in increased penetration depth, compared to other vibrational spectroscopic techniques. This enables direct analysis of solid samples, requiring little to no sample preparation. This, together with the advantages of being chemical-free, rapid, non-destructive and non-invasive made it possible to move NIR spectrophotometers out of the laboratory to the production environment to be used at-line, on-line or in-line for quality control purposes. The recent advancement in instrument development now also enables in-field and on-site analysis with the availability of portable and, more recently, also miniature instrumentation.<sup>30</sup> In principle, NIR spectroscopy applications should only include prediction of properties of organic nature of which

the molecular bonds absorb in the NIR region, *i.e.* direct measurements. A number of applications, however, illustrate that calibration models can also be developed for prediction of physical characteristics of samples. This could include, *e.g.* measurement of properties related to particle size, which enables prediction based on different scattering properties of particles differing in size. Similarly, components with molecular bonds that do not absorb in the NIR region can be measured with NIR spectroscopy. This is possible through co-variation of the non-organic component with an organic component in the sample, *i.e.* indirect measurement. The earliest report of a calibration model based on indirect measurement is that of Hirschfeld<sup>31</sup> who showed that it is possible to measure the concentration of salt (NaCl) dissolved in water. Sodium has no unique NIR absorption band but, because the water bands shift along the wavelength axis proportional to the salt concentration, prediction is possible.

Most applications referred to in this *tutorial review* are based on direct measurement predictions. NIR spectroscopy applications within food and agriculture still dominate with applications



in food safety foremost in recent NIR hyperspectral applications. A brief review of some non-food applications is also included (*e.g.* wood and wood products, soil, medical applications and pharmaceuticals). This tutorial review will be concluded with detection of food adulteration and aquaphotomics.

### Food and agricultural products

The most prominent commercial application of NIR spectroscopy remains in the areas of agricultural raw materials, food ingredients and finished food products. This included a wide range of commodities, *i.e.* meat, fruit and vegetables, dairy, cereals, beverages and tea.<sup>5</sup> More recent applications considered commodities not often investigated in earlier years, *e.g.* cocoa beans, pistachio nuts, hazelnut kernels, honey and transgenic foods (a concern for many consumers).<sup>32</sup> Investigations in the use of NIR spectroscopy on olives and olive oil remained a prominent and popular topic of investigation throughout the years. Typical SEP and RPD (quantitative) and correct classification (qualitative) values will be reported in the sections below for the respective applications.

Since 1998, most NIR hyperspectral imaging applications and thus also review papers mainly focussed on food quality detection. Commodities considered included wheat (*e.g.* preharvest germination) and maize (*e.g.* moisture and oil content), apples (*e.g.* bitter pit and bruise detection) and other fruit (*e.g.* peach, strawberry), cucumber (*e.g.* chilling injury), beef (tenderness), pork (marbling), and fish fillets (fat and moisture content and detection of nematodes and parasites).<sup>15</sup> Feng and Sun<sup>33</sup> advanced on reviews covering mainly food quality, and focussed their review on the application of NIR hyperspectral imaging to food safety assessment. Some of the first NIR hyperspectral imaging food safety applications since 1998 included faecal contamination on fruit, vegetables and chicken carcasses, followed by detection of defects in fruit and vegetables and diseased chicken carcasses, fungal contamination of cereal grains, and parasites on or in fish.<sup>33</sup>

### Meat and meat products

NIR spectroscopy applications on meat considered most often are the prediction of chemical composition such as content of crude protein (SEP = 0.35–1.08%; RPD = 2.62–5.13), intramuscular fat (SEP = 0.18–1.38%; RPD = 7.92–9.17) and moisture (SEP = 0.37–1.00%; RPD = 1.87–7.21), as well as technological properties such as pH (SEP = 0.05–0.16; RPD = 1.08–1.28), colour (SEP = 0.42–4.47; RPD = 0.90–2.16) and water-holding capacity (WHC; SEP = 1.14–2.355; RPD = 0.81–1.27).<sup>34</sup> NIR spectroscopy shows good potential for meat compositional predictions although, based on current prediction statistics, the majority will only be suitable for screening purposes and not adequately accurate for quality control testing (requires RPD > 3).<sup>34</sup> In the late 1990s, NIR predictions of chemical composition of meat included only intramuscular fat and moisture content.<sup>34</sup> These applications have since been extended to include more complex predictions such as ash content (SEP = 0.15–0.23; RPD = 1.26–4.53) showing good potential. Good results obtained for pH predictions were ascribed to an adequately wide range of pH reference data,

the use of a reference method with good repeatability and acquiring spectra from intact meat samples (in contrast to chemical composition predictions when minced meat resulted in better predictions). As would be expected when measuring colour, adding the visible range to the NIR range enabled improved predictions to be obtained. Predicting water-holding capacity and associated properties such as drip loss showed limited potential.

Other applications included the prediction of a number of sensory attributes in meat and meat products (*e.g.* flavour; SEP = 0.20–1.20; RPD = 0.57–1.40) together with its ability to classify meat samples based on quality (60–100% correct classification).<sup>34</sup> Attempting to predict sensory attributes of meat, only beef tenderness could be predicted with reasonable accuracy (SEP = 0.35; RPD = 3.82).<sup>34</sup> The lack of accurate prediction of sensory properties was due to the heterogeneity of intact meat samples, inconsistent sample preparation or presentation to the instrument, inaccurate reference methods and/or the subjectivity of taste panels.

During the late 1990s and early 2000s, the first NIR hyperspectral imaging study on meat, *i.e.* faecal detection on chicken carcasses, were reported.<sup>33</sup> This application has been implemented in a real-time inspection line. Meat quality measurements such as beef tenderness prediction, only followed in the late 2000s.<sup>35</sup> Wavelength ranges at the time included the visible region and only up to about 1100 nm due to the lower cost of silicon-based detectors compared to the more expensive InGaAs-based HgCdTe-based array detectors required for wavelength ranges from 1100 to 2500 nm. As is often the case with investigations using a new technology, these initial studies were only feasibility studies and they leave room for further investigations, especially in terms of validation of the developed methods.

### Fish and fish products

The most recent review on the use of NIR spectroscopy and NIR hyperspectral imaging, to study fish and fish products, came from the research group of Sun.<sup>36</sup> NIR spectroscopy has mainly been used to determine chemical composition, *i.e.* moisture ( $R^2 = 0.94$ – $0.98$ ; RMSEP/SEC V = 0.27%), fat ( $R^2 = 0.90$ ; RMSEP/SEC V = 0.14–0.67%) and free fatty acids ( $R^2 = 0.96$ ; SEC V = 0.59%) of fish. Because fish is highly perishable, microbial spoilage has been considered using both NIR spectroscopy and NIR hyperspectral imaging.<sup>36</sup> Nematodes could be detected in cod fillets at a detection rate of 58%. The availability of portable and miniature NIR spectroscopy instruments can play a vital role in terms of on-site quality and safety analyses.<sup>30</sup> As was observed for meat, NIR spectroscopy also has limited potential for prediction of sensory properties due to heterogeneity of the fish samples and unavoidable subjectivity of the reference method (taste panel).

### Milk and milk products

Milk, globally an important nutrition source, is mainly consumed in liquid form. Since MIR spectroscopy is widely used for rapid measurement during processing, NIR spectroscopy has been investigated as a complementary method and also for on-site (on-farm) applications due to ease of application.<sup>37</sup> The dairy industry has been using NIR spectroscopy as a routine analysis for over 30 years, with the first applications on milk powders.<sup>37</sup>



NIR technology has traditionally only been used to measure low moisture products. Today, more complex and non-homogeneous products such as cheese, yoghurt and many more, covering almost the entire range of dairy products, are being evaluated for a range of characteristics.<sup>37</sup> Applications on cheese included chemical composition such as dry matter (SEP = 15.4%; RPD = 6.0), fat (SEP = 14.9%; RPD = 3.2) and sodium chloride (SEP = 1.76%; RPD = 2.9). Using perturbations of the water signal in the case of cheese, sodium chloride is routinely predicted with NIR spectroscopy and commercial calibrations are readily available.

The capacity of NIR spectroscopy to predict sensory attributes of cheese such as visual evaluation (presence of holes, SEP = 0.4; RPD = 2.4); texture measurements (hardness, SEP = 0.1; RPD = 3.3; chewiness, SEP = 0.2; RPD = 2.7; creamy, SEP = 0.4; RPD = 1.6); taste (salty, SEP = 0.3; RPD = 1.6; buttery flavour, SEP = 0.3; RPD = 2.1; rancid flavour, SEP = 0.3; RPD = 2.3) and sensations such as pungency (SEP = 0.3; RPD = 2.6) and retronasal sensation (SEP = 0.2; RPD = 2.6) were illustrated by the research group of González-Martín.<sup>38</sup> It is especially this qualitative type of calibration development that has progressed significantly in recent years.<sup>37</sup> A more recent paper by González-Martín *et al.*<sup>39</sup> illustrated the good potential of NIR spectroscopy to predict volatile compounds in milk, *i.e.* 2-nonanone (SEP = 0.087; RPD = 3.4), acetaldehyde (SEP = 0.041; RPD = 2.3), ethanol (SEP = 3.89; RPD = 2.8), 2-heptanone (SEP = 0.17; RPD = 2.8), 2-butanol (SEP = 1.20; RPD = 2.1) and 2-pentanone (SEP = 0.41; RPD = 2.0).

Milk is a challenging matrix to study, since it is a turbid opaque liquid and highly scattering due to the presence of milk fat globules and casein micelles in suspension. It is, however, possible to separate the effects of scatter and absorption.<sup>37</sup> Based on the theory and fundamental principles of scatter, scientists might be able to use the proposed strategy to describe the chemical and physical properties of milk as well as other highly scattering materials better.

What remains to be addressed to improve the effective use of NIR spectroscopy on dairy products are:<sup>37</sup> (1) whether either reflectance or transmission spectroscopy should be used; (2) an optimum wavelength range to be used for dairy analysis; (3) careful consideration of sample selection and preparation for calibration development purposes; and (4) optimum pre-processing techniques to deal with particle size distribution variation between samples (such as milk powders).

The availability of only a few reports on the use of NIR hyperspectral imaging in dairy products is potentially due to the homogeneity of the liquid and powdered milk samples and the difficulty to analyse cheese due to the heat generated by the light source and longer acquisition time (up to a few minutes) of earlier systems.<sup>15</sup> With pushbroom imaging systems (which acquire images within a few seconds) more readily available, the analysis of cheese products should increase.

### Fruit and vegetables

The very early applications in horticulture focussed on dry matter content of onions, soluble solids content (SSC) of apples and water in mushrooms.<sup>29</sup> SSC continued to be evaluated (mostly on apples, RMSEP = 0.3–1.6%) with little reported on vegetables.

This is expected with sweetness (measured as Brix) not as important for vegetables as for fruit. When the water bands in NIR spectra dominate, as in fruit, the concentration of acids is much lower than that of sugars.<sup>29</sup> Indirect measurement of acidity was shown to be possible due to its correlation with sugars, however, it was more difficult to predict ( $R^2 = 0.65$ ; SEP = 0.15%). Similarly, it was possible to measure fruit maturity due to co-variation with sugar content and microstructure of the tissue. The penetration of NIR light into and scattering within fruit and vegetable tissue are affected by the latter.<sup>29</sup> This also makes it possible to measure properties such as stiffness, internal damage as well as sensory attributes. Sorting of fruit based on quality attributes and not only on external appearance, has become a reality due to the availability of NIR spectrophotometers.<sup>29</sup> The short analysis time (seconds) required by diode array instruments also enables on-line sorting of fruit based on quality properties. In spite of NIR spectroscopy being an economical technique (once implemented), it will be costly to replace current systems (based on visual external appearance evaluation of fruit and vegetables). It will thus only become feasible to implement NIR spectroscopy on-line once consumers are willing to pay higher prices for fruit being *e.g.* extra sweet.<sup>29</sup>

In a more recent study, the gross energy of food grade legumes were predicted (SEP = 0.025 kcal g<sup>-1</sup>; RPD = 4.2).<sup>40</sup> The standard error was very low compared to that of the reference method (0.204 kcal g<sup>-1</sup>), *i.e.* the adiabatic bomb calorimeter.

Numerous NIR hyperspectral imaging investigations have been executed on fruit and vegetables during the last decade.<sup>15,29</sup> Due to the penetration depth required, most of the applications were performed in the shorter wavelength ranges of the NIR region (up to 1000 nm); often also including the visible range. This may be ascribed to affordable imaging instrumentation available at the time, which only operates in the shorter wavelengths region. Quality aspects important for fresh fruit and vegetables include measurement of firmness and SSC with detection of early bruising and chilling injury also being important.<sup>15,29</sup> One of the most significant benefits of NIR hyperspectral imaging is that defects such as bruising can be detected and visualised in principal component images or classification plots before they are actually visible on the fruit itself. This enables the opportunity to prevent fruit and vegetable with potentially short shelf-life to enter the supply chain. Safety aspects were addressed by means of detection of faecal contamination on fruit and vegetables (although mostly on apples) and received significant attention since the early 2000s.<sup>33</sup>

### Cereals and cereal products

Since the first application of NIR spectroscopy in the early 1970s, it continued to be researched and applied in this field. Apart from compositional analysis, more complex applications are now being investigated, including analysis in breeding development and genetics, detection of adulteration and presence of weeds and insects in wheat and flour.<sup>5</sup> In a novel approach, reviewed by Woodcock *et al.*,<sup>5</sup> endosperm genes and gene combinations of barley mutants was classified. This allowed the interpretation of physico-chemical and genetic significance of the developed models.



Another unique application investigated changes in dough development during mixing non-invasively.<sup>5</sup> This was done by integrating the second derivative curves of the spectra under the peak at 1125 to 1180 nm. The optimum NIR mixing time was subsequently determined from the curves after plotting the measured areas against mixing time determined earlier. High correlations were found between measurements derived from NIR mixing curves and the measured dough properties when doughs were prepared from flour milled from single variety wheats.<sup>5</sup>

When using NIR hyperspectral imaging to analyse whole cereal grains, a significant advantage is that, although a number of grains can be imaged simultaneously, prediction results from single kernels are obtained. Single kernel analyses with NIR spectroscopy are time-consuming and can be complicated due to the difference in kernel size and alignment when presenting it to the instrument. Elmasry *et al.*<sup>15</sup> reviewed cereal applications covering both quality and safety aspects. The heterogeneous nature of cereal grains, both within and between kernels, makes it highly suitable for image analysis. Both quantitative (*e.g.* moisture, oil and oleic content in maize) and qualitative (*e.g.* classification of wheat classes based on quality, and maize based on kernel hardness) analyses have been performed. Detection of fungal infection in maize has also been considered.

## Wine

The majority of NIR wine applications focus on measurement of wine properties such as alcohol content, sensory and aromatic attributes and fermentation.<sup>5</sup> It was illustrated that NIR calibration models performed better than those developed with spectra collected in the MIR region. The high S/N ratio was given as the main reason for the superior coefficient of determination values of the NIR predictions. An earlier review by Cozzolino *et al.*<sup>41</sup> included the measurement of grape composition with NIR spectroscopy reporting typical prediction errors for total soluble solids (TSS) (SEP = 1.04–2.96° Brix; RPD = 1.33–4.0), total anthocyanins (SEP = 0.05–0.06 mg g<sup>-1</sup>; RPD = 3.8–4.2), acidity (SEP = 1.28 g L<sup>-1</sup>), and pH (SEP = 0.045–0.11; RPD = 2.2–2.8). Wine composition included alcoholic degree (SEP = 0.24%; RPD = 5.7), total acidity (SEP = 0.48 meq L<sup>-1</sup>; RPD = 2.27), pH (SEP = 0.07; RPD = 2.4), glycerol (SEP = 0.72 g L<sup>-1</sup>; RPD = 4.0), reducing sugars (SEP = 0.33 g L<sup>-1</sup>; RPD = 10.3) and total sulphur dioxide (SEP = 23.5 mg L<sup>-1</sup>; RPD = 1.8). The authors reported the suitability of vis-NIR spectroscopy to predict wine quality as judged by both commercial wine quality rankings and wine show scores. Better results were obtained from the commercial wine quality rankings rather than from the sensory data ( $R^2 = 0.84$ ; SEC<sub>V</sub> = 0.97). Some sensory properties (estery, honey, toasty, caramel, perfumed floral and lemon) correlated to some extent ( $R^2 = 0.5$ ), while others were not as successful (passion fruit, sweetness and overall flavour;  $R^2 = 0.30$ ).<sup>41</sup> The study was, however, performed on a limited set of only 40 samples. Although ideally larger samples sets should be used, using a limited set is understandable due to the high cost of performing sensory analysis.

The use of NIR spectroscopy has also great potential to follow the red wine fermentation process by means of ethanol (SEC = 0.15%) and sugar (SEC = 2.6 g L<sup>-1</sup>) contents.<sup>41</sup> A problem identified when monitoring wine fermentation was the change in the sample matrix during the course of fermentation and subsequent analysis.

A need identified,<sup>41</sup> as for many other applications, is the availability of inexpensive portable hand-held instruments, especially for the measurement of the compositional quality of grapes while still on the vine. This has since become a reality with the development of not only portable but especially low-cost miniature instruments.<sup>30</sup> More investigations using miniature instruments are foreseen in the near future.

## Beer

A good quality beer depends on good quality raw materials and continuous information for process control.<sup>42</sup> NIR spectroscopy is ideal for characterisation of the raw materials (barley malt, hop and yeast) and could be applied during process control, to analyse intermediate products and finally the finished product. Most NIR spectroscopy applications till now focussed on determination of barley properties to select the best varieties for producing high-quality malt for high-quality beer production.<sup>42</sup> Studies considered included genotype classification, mycotoxin detection ( $R^2 = 0.993$ ; SEP = 3.097 ppb) and quantitative analysis of intact and ground grain for moisture ( $R^2 = 0.87$ –0.99; SEP = 0.12–0.97%), protein ( $R^2 = 0.71$ –0.99; SEP = 0.09–0.64%) and B-glucan ( $R^2 = 0.59$ –0.79; SEP = 0.15–1.33%), properties that are also important for production of good quality beer. Analysis of intermediate products mostly entails quantitative analysis of wort and in particular extract ( $R^2 = 0.76$ –0.88; SEP = 1.00–2.29%) and free amino nitrogen (FAN) ( $R^2 = 0.51$ –0.74; SEP = 17 mg L<sup>-1</sup>). Analysis of the beer includes determination of real extract (SEP = 0.075–0.28% w/w) and ethanol (SEP/RMSEP = 0.07–0.14%). The majority of these studies were performed on laboratory scale and not in a commercial environment.<sup>41</sup> Process analytical technology (PAT), developing at a fast pace, is playing an ever increasing role in product–process optimisation strategies.<sup>41</sup> There is thus scope for more in-line applications to be considered.

## Non-food near-infrared spectroscopy applications

### Wood and wood products

NIR spectroscopy on wood and wood products has been actively researched during the last 20 years, determining chemical composition, physico-chemical and mechanical properties.<sup>43</sup> The first reported use of NIR spectroscopy on wood pulps was prediction of lignin content, followed by pulp yield and cellulose content of wood measurements.<sup>43</sup> Although NIR spectroscopy is used in the assessment of breeding trial samples, it is still not widely implemented for wood characterisation in commercial environments. This is potentially due to no premium being paid for high-quality wood; quality analyses are thus not essential. Limited use could also be ascribed to the capital cost involved





in obtaining laboratory NIR systems and the need to develop, use and maintain calibration models.<sup>44</sup>

The implementation of suitable, economical portable instruments is thus required as alternatives to laboratory systems.<sup>45</sup> Where suitably accurate NIR models is required for the commercial environment, using NIR spectroscopy as a screening tool in breeding programmes, analytical accuracy might not be necessary and the accuracy obtained with portable systems might be acceptable.

Schwanninger *et al.*,<sup>46</sup> extensively reviewed NIR band assignments for wood, and compiled detailed tables comprising band locations in both wavenumber ( $\text{cm}^{-1}$ ) and wavelength (nm), the component likely to absorb at this band location, the bond vibration, as well as descriptive remarks. Knowledge of the band locations where chemical or functional groups absorb is indispensable for a better understanding of the underlying chemistry behind developed multivariate calibration models.

The highly heterogeneous nature of the wood sample matrix and the importance to know the spatial distribution of wood properties, make wood highly suitable for NIR hyperspectral imaging. With the spatial advantage and ability to visualise NIR hyperspectral image analysis results, more research is required to benefit from this advantage in terms of improved knowledge on the overall heterogeneity of the wood sample matrices. This could lead to a better understanding of the effect of the environment on wood structure.

## Soil

Soil is a fundamental natural source for the production of *e.g.* agricultural produce and food. It is known to be a complex matrix comprising organic and inorganic mineral matter, water and air. One of the difficulties in analysing soils is that no two soils are the same and variations may occur over even short distances. A substantial increase in research on the use of NIR (or vis-NIR) spectroscopy in soil science has been observed during the last 15–20 years.<sup>47</sup> Applications mainly focus on basic soil composition, texture and clay mineralogy. Attention has also been given to nutrient availability and properties such as fertility, structure and microbial activity. The most successful calibrations in soils are those for total ( $R^2 = 0.66\text{--}0.87$ ; RMSEP =  $4.2\text{--}7.9 \text{ mg g}^{-1}$ ) and organic carbon content ( $R^2 = 0.55\text{--}0.92$ ; RMSEP =  $2.5\text{--}29 \text{ mg g}^{-1}$ ) as well as clay content ( $R^2 = 0.56\text{--}0.94$ ; RMSEP =  $1.9\text{--}10.3\%$ ). This is due to clay minerals and soil organic matter both being fundamental constituents of the soil and absorb in the NIR region. Indirect measurements, such as pH, extractable P, K, Fe, Ca, Na and Mg, were found to be highly variable due to the co-variations to constituents that are spectrally active expected to be unstable.<sup>47</sup> Pure metals do not absorb in the NIR region, but can be detected because of co-variation with spectrally active components and they can also be complexed with organic matter. Co-variations upon which indirect calibrations are built may also be very different at different sites, making transferring calibrations geographically difficult. Another reason could be the reference method used, *e.g.* different types of P are measured by different reference methods, which are not always well correlated.<sup>47</sup>

Spectra from field samples are not necessarily worse than spectra from appropriate collected and well-prepared laboratory soil samples.<sup>47</sup> In-field or on-site measurements should be considered more often, with no sampling or sample preparation required. There is also a need for better handling of the variability and complexity of soils and a better understanding of the physical basis for the reflection of light from soils.<sup>47</sup>

## Medical

NIR spectroscopy applied to *in vivo* medical applications dates back to 1977 when Frans Jöbsis reported brain tissue can be measured within the NIR range (700–1000 nm).<sup>11</sup> This enabled real-time, non-invasive analysis of haemoglobin oxygenation.<sup>11</sup> The main advantages of NIR spectroscopy for medical applications are that it does not have any side-effects, can be used in real-time and that it is cost-effective and portable. Currently, the main NIR spectroscopy and NIR hyperspectral imaging applications include pulse oximetry, brain/muscle oximetry, functional brain cortex mapping and optical mammography.<sup>11</sup> Recently, thirteen papers have been published which critically reviewed a number of medical applications of NIR spectroscopy.<sup>11</sup> The review on the early years of medical NIR spectroscopy research and development included an update on the status of current commercial oximeters and relevant applications.<sup>11</sup> Identification of the most relevant clinical application, *i.e.* the evaluation of cerebral oxygenation during adult cardiac surgery and cardiopulmonary bypass, concluded this review. Although many commercial oximeters are available, precision and standardisation need to be improved. A review on the techniques and instrumentation for medical NIR diffuse spectroscopy include the development of optical coherence tomography, application of NIR spectroscopy in pre-term and new-born infants, the evolution of NIR wireless methodology for bladder studies, and functional NIR as a cortical brain imaging technique.<sup>11</sup> The extensive use of NIR spectroscopy to evaluate cerebral and muscle haemodynamic responses during exercise in healthy subjects and athletes has also been reviewed.<sup>11</sup>

## Pharmaceuticals

A common application of NIR spectroscopy is the identification of the active compound or active pharmaceutical ingredient (API) present in a tablet or drug.<sup>7</sup> It may also be used to identify excipients. Due to an inverse relationship between particle size and baseline offset, determination of particle size is also possible. For NIR spectroscopy to be implemented efficiently in the pharmaceutical area as part of PAT, it is required to be used in-line and on-line. One of the most critical stages during production where NIR can be applied is blending (apart from identification of APIs and raw materials). To enable efficient identification of active compounds or excipients, libraries of typically used pharmaceutical materials are usually created. To understand and interpret the spectra better, information about the characteristic bands is required. Band assignments for pharmaceutical ingredients have been addressed in a recent publication reviewed by Jamróiewicz *et al.*<sup>7</sup> The use of NIR hyperspectral imaging is gaining popularity in the field of



pharmaceuticals, where the term “NIR chemical imaging” (NIR-CI) is preferably used. The benefit of the spatial dimension is used here to determine the homogeneity of the distribution of active compounds as well as the contents of these compounds in the tablets.

## Food adulteration

Adulteration of food has been a common practice since ancient times.<sup>26,27</sup> It can be unintentional (addition of foreign substances due to negligence), but is often intentional (addition of foreign substances for economic gain). An earlier practice of adding substances to camouflage bad appearance and taste of rotten food, could be detected visually or with basic scientific instrumentation. Adulteration practices today are, however, more refined and sophisticated, requiring advanced techniques for detection. Most food adulteration investigations using NIR spectroscopy to date have been reported as feasibility studies, performed on a limited number of samples.<sup>26,27</sup> The likely reason being the costs involved to collect suitable and large enough sample databases and evaluation of the models or possibly work performed in-house not published in the public domain.

With the outbreak of the milk powder scandal (addition of melamine) in China in 2008,<sup>33</sup> and the more recent meat adulteration scandal, the detection of adulterants and consideration of appropriate detection methods received renewed attention.<sup>48,49</sup> NIR spectroscopy was considered in favour of Kjeldahl to detect melamine since the Kjeldahl method fails to distinguish between protein-based nitrogen and non-protein nitrogen (derived from small organic molecules such as melamine). The Dumas method also cannot eliminate the negative influence of non-protein nitrogen on the determination of protein levels. The challenge of detecting and quantifying melamine is the very low levels (often present only in ppm). Fu *et al.*,<sup>50</sup> however, claimed that NIR hyperspectral imaging (990–1700 nm) and spectral similarity analyses were effective to detect different concentrations of melamine adulteration (from 0.025 to 1%) in milk powders. They suggested an improvement in the accuracy of these techniques to even lower levels (<0.02% or 200 ppm) by spreading the sample mixtures in a thin layer in larger containers to increase the surface area presented for NIR hyperspectral imaging.

## Aquaphotomics

Water, known as a common component in biological systems, is still not well understood. Due to the complexity of the role it plays in biological systems water has received considerable attention over the years. Aquaphotomics, which is based on vis-NIR spectroscopy and multivariate data analysis, relates water absorption patterns to the respective functions of different biological systems.<sup>13</sup> The aim is to build up a database of water absorption bands (*i.e.* water matrix coordinates) and to identify characteristic water absorption patterns (*i.e.* water molecular structures) that could be used as biological markers.

This could contribute to a better understanding of complex biological systems.

## Conclusion

NIR spectroscopy is well established as a laboratory analysis system. It is however not limited to the laboratory; at-line and in some cases on-line and in-line analysis are becoming increasingly common. There is still scope for improvement of in-line applications. Instrument development, especially miniaturisation, now enables on-site and in-field analysis. The instrument is thus more often taken to the sample rather than bringing the sample to the instrument. Improving accuracy and ensuring stability of these instruments are still required. Further developments in chemometrics will continuously enable more accurate, faster and more robust generic global models. Today, many industries approach NIR spectroscopy as the only viable alternative for quality control.

## References

- 1 F. W. Herschel, *Philos. Trans. R. Soc. London*, 1800, **90**, 255–329.
- 2 K. H. Norris, *J. Near Infrared Spectrosc.*, 1996, **4**, 31–37.
- 3 W. F. McClure, *J. Near Infrared Spectrosc.*, 2003, **11**, 487–518 and papers therein.
- 4 J. S. Shenk, J. J. Workman and M. O. Westerhaus, Application of NIR Spectroscopy to Agricultural Products, in *Handbook of Near-Infrared Analysis*, ed. D. A. Burns and E. W. Ciurczak, Marcel Dekker, Inc., New York, 2nd edn, 2001.
- 5 T. Woodcock, G. Downey and C. P. O'Donnell, *J. Near Infrared Spectrosc.*, 2008, **16**, 1–29 and papers therein.
- 6 C. M. McGovern, J. Weeranantanaphan, G. Downey and M. Manley, *J. Near Infrared Spectrosc.*, 2010, **18**, 87–111.
- 7 M. Jamrógiewicz, *J. Pharm. Biomed. Anal.*, 2012, **66**, 1–10 and papers therein.
- 8 J. Workman Jr, *J. Near Infrared Spectrosc.*, 1996, **4**, 69–74.
- 9 E. Cleve, E. Bach and E. Schollmeyer, *Anal. Chim. Acta*, 2000, **420**, 163–167.
- 10 M. Blanco, M. Alcalá, J. Planells and R. Mulero, *Anal. Bioanal. Chem.*, 2007, **389**, 1577–1583.
- 11 M. Ferrari, K. H. Norris and M. G. Sowa, *J. Near Infrared Spectrosc.*, 2012, **20**, vii–ix and papers therein.
- 12 N. Heigl, C. H. Petter, M. Rainer, M. Najam-ul-Haq, R. M. Vallant, R. Bakry, G. K. Bonn and C. W. Huck, *J. Near Infrared Spectrosc.*, 2007, **15**, 269–282.
- 13 R. Tsenkova, *J. Near Infrared Spectrosc.*, 2009, **17**, 303–314.
- 14 A. F. Goetz, G. Vane, J. E. Solomon and B. N. Rock, *Science*, 1985, **228**, 1147–1153.
- 15 G. Elmasry, M. Kamruzzaman, D.-W. Sun and P. Allen, *Crit. Rev. Food Sci. Nutr.*, 2012, **52**, 999–1023 and papers therein.
- 16 J. Manuel Amigo, J. Cruz, M. Bautista, S. Maspocho, J. Coello and M. Blanco, *TrAC, Trends Anal. Chem.*, 2008, **27**, 696–713.
- 17 M. Sowa, J. R. Friesen, M. Levasseur, B. Schattka, L. Sigurdson and T. Hayakawa, *J. Near Infrared Spectrosc.*, 2012, **20**, 601–615.



- 18 H. Liang, *Appl. Phys. A: Mater. Sci. Process.*, 2012, **106**, 309–323.
- 19 D. B. Thomas, C. M. McGoverin, A. Chinsamy and M. Manley, *J. Near Infrared Spectrosc.*, 2011, **19**, 151–159.
- 20 B. G. Osborne, T. Fearn and P. H. Hindle, *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, Longman Scientific and Technical, Harlow, 2nd edn, 1993.
- 21 L. Esteve Agelet and C. Hurburgh Jr, *Crit. Rev. Anal. Chem.*, 2010, **40**, 246–260.
- 22 B. Boldrini, W. Kessler, K. Rebner and R. Kessler, *J. Near Infrared Spectrosc.*, 2012, **20**, 438–508.
- 23 Å. Rinnan, F. van den Berg and S. Balling Engelsen, *TrAC, Trends Anal. Chem.*, 2009, **28**, 1201–1222.
- 24 T. Næs, T. Isaksson, T. Fearn and T. Davies, *A User-friendly Guide to Multivariate Calibration and Classification*, NIR Publications, Chichester, 2002.
- 25 P. Dardenne, *NIR News*, 2010, **21**, 8–14.
- 26 G. Downey, *J. Near Infrared Spectrosc.*, 1996, **4**, 47–61.
- 27 M. Manley, G. Downey and V. Baeten, Spectroscopic Technique: Near Infrared (NIR) Spectroscopy, in *Modern Techniques for Food Authentication*, ed. D. W. Sun, Elsevier, Oxford, 2008 and papers therein.
- 28 *Techniques and Applications of Hyperspectral Image Analysis*, ed. H. F. Grahn and P. Geladi, John Wiley and Sons, Ltd., 2007.
- 29 B. Nicolaï, K. Beullens, E. Bobeleyn, A. Peirs, W. Saeys, K. Theron and J. Lammertyn, *Postharvest Biol. Technol.*, 2007, **46**, 99–118 and papers therein.
- 30 M. Alcalà, M. Blanco, D. Moyano, N. W. Broad, N. O'Brien, D. Friedrich, F. Pfeifer and H. W. Siesler, *J. Near Infrared Spectrosc.*, 2013, **21**, 445–457.
- 31 T. Hirschfeld, *Appl. Spectrosc.*, 1985, **39**, 740–741.
- 32 A. Alishahi, H. Farahmand, N. Prieto and D. Cozzolino, *Spectrochim. Acta, Part A*, 2010, **75**, 1–7.
- 33 Y. Z. Feng and D.-W. Sun, *Crit. Rev. Food Sci. Nutr.*, 2012, **52**, 1039–1058 and papers therein.
- 34 N. Prieto, R. Roche, P. Lavín, G. Batten and S. Andrés, *Meat Sci.*, 2009, **83**, 175–186 and papers therein.
- 35 G. Elmasry, D. F. Barbin, D.-W. Sun. and P. Allen, *Crit. Rev. Food Sci. Nutr.*, 2012, **52**, 689–711 and papers therein.
- 36 D. Liu, X. A. Zeng and D.-W. Sun, *Appl. Spectrosc. Rev.*, 2013, **48**, 609–628.
- 37 T. M. P. Cattaneo and S. E. Holroyd, *J. Near Infrared Spectrosc.*, 2013, **21**, 302–310 and papers therein.
- 38 M. I. González-Martín, P. Severiano-Pérez, I. Revilla, A. M. Vívar-Quintana, J. M. Hernández-Hierro, C. González-Pérez and I. A. Lobos-Ortega, *Food Chem.*, 2011, **127**, 256–263.
- 39 I. González-Martín, J. M. Hernández-Hierro, C. González-Pérez, I. Revilla, A. Vívar-Quintana and I. Lobos Ortega, *LWT-Food Sci. Technol.*, 2014, **55**, 666–673.
- 40 T. Szigedi, M. Fodor, D. Pérez-Marin and A. Garrido-Varo, *Food Anal. Methods*, 2013, **6**, 1205–1211.
- 41 D. Cozzolino, R. G. Damsbergs, L. Janik, W. U. Cynkar and M. Gishen, *J. Near Infrared Spectrosc.*, 2006, **14**, 279–289.
- 42 V. Sileoni, O. Marconi and G. Perriti, *Crit. Rev. Food Sci. Nutr.*, 2012, DOI: 10.1080/10408398.2012.726659, and papers therein.
- 43 R. Meder, T. Trung and L. Schimleck, *J. Near Infrared Spectrosc.*, 2010, **18**, v–vii.
- 44 R. Meder, J. R. Brawner, G. M. Downes and N. Ebdon, *J. Near Infrared Spectrosc.*, 2011, **19**, 421–429.
- 45 R. Meder and L. Schimleck, *J. Near Infrared Spectrosc.*, 2011, **19**, v.
- 46 M. Schwanninger, J. C. Rodrigues and K. Fackler, *J. Near Infrared Spectrosc.*, 2011, **19**, 287–308.
- 47 B. Stenberg, R. A. Viscarra Rossel, A. Mounem Mouazen and J. Wetterlind, Visible and Near infrared Spectroscopy in Soil Science, in *Advances in Agronomy*, ed. D. L. Sparks, Academic Press, Burlington, 2010, vol. 107, and papers therein.
- 48 E. Domingo, A. A. Tirelli, C. A. Nunes, M. C. Guerreiro and S. M. Pinto, *Food Res. Int.*, 2014, **60**, 131–139.
- 49 M. Kamruzzaman, D.-W. Sun, G. Elmasry and P. Allen, *Talanta*, 2013, **103**, 130–136.
- 50 X. Fu, M. S. Kim, K. Chao, J. Qin, J. Lim, H. Lee, A. Garrido-Varo, D. Pérez-Marin and Y. Ying, *J. Food Eng.*, 2014, **124**, 97–104 and papers therein.

