4

Spectroscopic Technique: Fourier Transform Near-infrared (FT-NIR) Spectroscopy

Vincent Baeten, Marena Manley, Juan Antonio Fernández Pierna, Gerard Downey and Pierre Dardenne

Introduction 117
Theory and instrumentation 119
New trends in chemometrics as applied to NIR spectroscopic data 126
Authentication by FT-NIR 130
Authentication by FT-NIR microscopy 140
Conclusions 142
References 142

Introduction

Correct and defensible labeling is of prime importance for the consumer and the producer of authentic food products. For the consumer, labeling and its control are essential in terms of the identification of the food product. He or she may pay more in order to get a food product with well-defined attributes, such as species or geographic origin, or more subjective features, as in products labeled home made, organic or fair trade. Food products are not only valued for their appearance, taste and nutritional value (whatever their major and minor compound composition), but also for their tremendous symbolic power. For the producer, labeling and its control are crucial in terms of economic strategy. Indeed, producers selling food with specific quality attributes invest more in the production of their food products, as they expect a substantial return. This return may be obtained by charging a higher price and/or the development of a loyal customer base.

Modern Techniques for Food Authentication
ISBN: 978-0-12-374085-4

All actors in the food chain need to have analytical tools at their disposal to verify the nature of high-value foods in particular, and to protect their brands. Ideally, these tools should permit rapid, non-destructive and inexpensive control at any point of the food chain, and should be part of the traceability strategy for the food product. For both food producers and consumers, confirmation that a food product is the one expected (i.e. the real thing) is crucial because this is the basis of mutual trust. Such confirmation requires looking beyond the mere superficial surface appearance or the composition of the products (Van der Reyden, 1996; Downey *et al*., 2006).

Among the panoply of methods for the assessment of a food product's authenticity, several vibrational spectroscopic techniques have recently been proposed. Several reports have proposed methods based on ultraviolet (UV), near-infrared (NIR), mid-infrared (MIR) and Raman spectroscopy. *Vibrational spectroscopy* techniques have been used for many years as favored tools for the study of the molecular structure of organic matter. On the other hand, for many decades several methods based on UV, NIR, MIR or Raman spectroscopy have been proposed and widely used as methods of choice for forensic studies (e.g. authentication of hair, fibers, paint, drugs and poisons) and assessment of the authenticity of art works (Brettel *et al*., 2005).

Spectroscopic methods, based mainly on NIR techniques, are often presented as new approaches for at-line, on-line and in-line control of authentication of food products. These techniques are already routinely used in the industry to control both raw materials and finished products for specific production standards as a common authenticity issue. This means that tedious reference methods only need to be used if deviations from these quality standards occur during production (Müller and Steinhart, 2007). Demonstration of the potential of vibrational spectroscopy techniques for the assessment of value-added claims like geographic origin, species discrimination, detection and quantification of adulteration and the assessment and discrimination of process type or brands have been reported since the beginning of the 1990s (Dennis, 1998).

The main limitation of the spectroscopic approach is the fact that it needs large datasets in order to calibrate any given instrument, and only few publications have dealt with the interpretation of the spectroscopic features related to specific authenticity issues. The main challenge therefore facing the spectroscopists is to extract the information in such a way that it can be used in qualitative and quantitative analysis. NIR spectra can contain up to thousands of absorbance values at defined wavelengths (i.e. variables), and the challenge is to characterize the spectral data set and isolate the variables that can be correlated with the information of interest (i.e. authenticity issue) (Baeten and Aparicio, 2000). In order to achieve this goal, a wide range of chemometric tools is at the disposal of analysts, who have to select the appropriate one according to their specific aims and the characteristics of the dataset. Among the many methods proposed for authentication of food products, spectroscopic methods seem to be the preferred ones to flag suspicious samples before, during and after the production of a food product. The real future challenge for the spectroscopic techniques will be the demonstration of their daily use in the industry and the marketplace for food product authentication.

The growing interest in spectroscopic techniques for developing methods and strategies to assess the authenticity of products may be gleaned from the number

Table 4.1 Selection of EU-funded projects including research for the development and validation of spectroscopic methods

Acronym	Food product	Authenticity issue	Spectroscopic method	Web site
CO-EXTRA	GMO	Transgenic/non-transgenic	NIR imaging	http://www.coextra.eu/
FEEDFAT	Animal fats	Discrimination	MIR	http://www.ub.edu/feedfat/
TRACE	Olive oil, honey, cereal, meat	Geographic origin	NIR, MIR, Raman	www.trace.eu.org
MEDEO	Olive oil	Adulteration	MIR, Raman	http://www.cica.es/aliens/igmedeo/
SAFEED-PAP-PAP-PAP	Feed, processed animal protein	Adulteration	NIR microscopy, NIR imaging	http://safeedpap.feedsafety.org/
STRATFEED	Feed	Adulteration	NIR, NIR microscopy	http://www.stratfeed.cra.wallonie.be
TYPIC	Dry cured ham, wine	Geographic origin, brand	NIR, MIR, Raman	http://www.typic.org/

Source: www.trace.eu.org

of European projects involving these techniques and financed by the European Commission. Table 4.1 summarizes some of the European projects which include research for the development and validation of spectroscopic methods. An updated list can be found on the website of European project TRACE (Tracing the origin of food; http://www.trace.eu.org/library/links.php), which is an integrated project financed in the EU 6th Framework Programme. This particular project aims to improve the health and well-being of European citizens by delivering integrated traceability systems that will enhance consumer confidence in the authenticity of food.

This chapter is complementary to Chapter 3, which was dedicated to near-infrared spectroscopy. The focus in this chapter is specifically on Fourier transform near-infrared spectroscopy (FT-NIR) and microscopy (FT-NIRM), and their applications for the authentication of agro-food products.

Theory and instrumentation

Regarding the history of near-infrared spectroscopy, a turning point was the work of Sir Frederick William Herschel, reported in 1800. Herschel discovered that the sun's energy was not limited to what we can see. He demonstrated this by projecting a rainbow onto a bench using glass prisms which are transparent to short-wave NIR radiation. He positioned a series of blackened bulb thermometers on a bench and measured the relative heat in the different parts of the rainbow. The temperature increased by moving from the blue to the red. Herschel's scientific insight meant that he did not stop the temperature measurement when he reached the end of the visible red color region of the dispersed light, but continued to observe temperature when he placed a thermometer beyond that point. This work was a key milestone in the discovery of the electromagnetic spectrum (Davies, 1998; Pasquini, 2003).

Near-infrared spectroscopy: few elements of theory

The electromagnetic spectrum is usually divided into several regions, from high to low energy, including, among others, γ-rays, X-rays, ultraviolet (UV), visible (VIS),

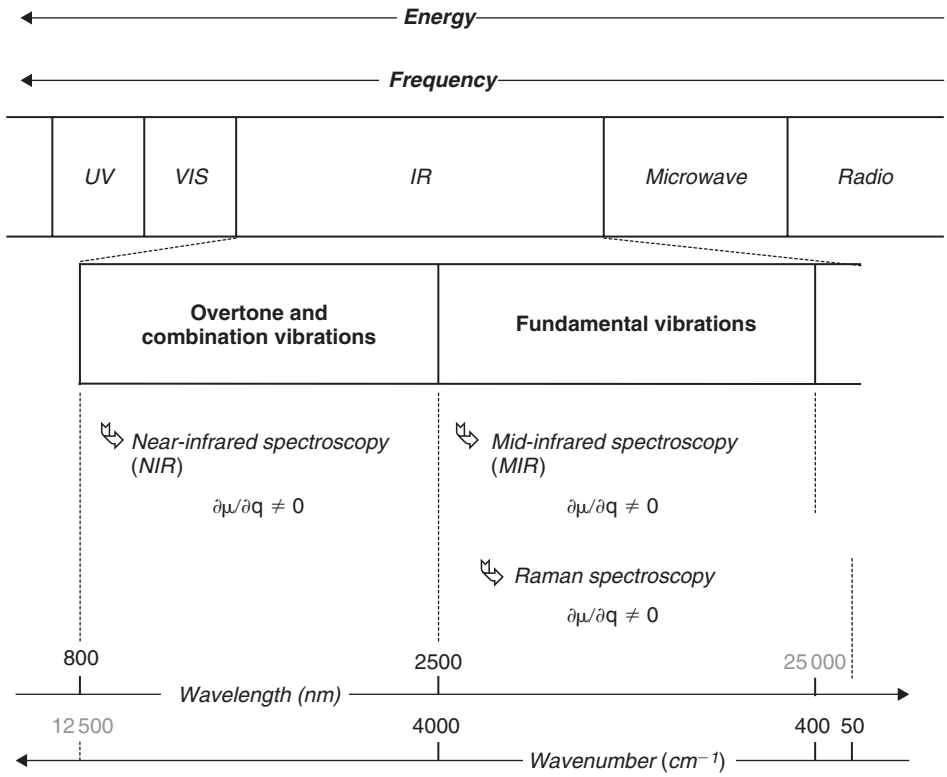

Figure 4.1 Features of the NIR and MIR regions of the electromagnetic spectrum (Source: Baeten and Dardenne, 2002).

infrared (IR), microwaves and radio waves. Specific atomic or molecular transition corresponding to different energies is characteristic of each region of the electromagnetic spectrum. Moreover, the infrared region is divided into near-infrared (NIR), mid-infrared (MIR) and far-infrared regions. Figure 4.1 shows the regions of the electromagnetic spectrum covering the spectroscopic techniques discussed in Chapters 2–6 (namely NIR, MIR and Raman spectroscopy). The figure includes the energetic transition involved as well as the corresponding wavelengths and wavenumber ranges (Baeten and Dardenne, 2002). To summarize, near-infrared spectroscopy (NIR) is a vibrational spectroscopy that concerns photon energy ($h\nu$) in the energy range of 2.65×10^{-19} J, which corresponds to the wavelength range of 750–2500 nm (and to the wavenumber range of 13 300–4000 cm^{-1}).

In order to explain the properties of electromagnetic radiation, it is necessary to refer to the classical theory describing electromagnetic radiation as a wave and the quantum theory stating that electromagnetic radiation is a stream of energetic particles. The *classical theory* says that the properties of light can be explained on the basis of an electric field associated with a perpendicular magnetic field of high frequency. This field moves in the direction of the light. The electric and magnetic fields interact with organic matter to give rise to a spectrum. The movement of the

radiation has the properties of a sine wave described by the equation $Y = A \sin \omega t$, where Y is the displacement with an amplitude A, ω is the angular velocity (rad s^{-1}) and t is the time in seconds. The frequency υ expressed as cycles per second (s^{-1} or Hz) corresponds to the number of times ($\omega/2\pi$) that the pattern is repeated in 1 second. The distance covered in one complete cycle, known as the wavelength λ, is an additional property of the wave describing the radiation. As the classical theory does not explain all the properties of the electromagnetic radiation and its interaction with matter, and fails to account for phenomena associated with the absorption of the energy, it was necessary to complement this theory.

The *quantum theory* views electromagnetic radiation as a stream of discrete particles; Planck (1925) was the first to put forward the hypothesis that the electromagnetic wave was not continuous but composed of corpuscular units called quanta. The energy of a quantum of radiation is defined (for a specific molecule all the energy levels are allowed) and characterized by its frequency (Osborne and Fearn, 1986; Lachenal, 1998a, 1998b).

As discussed before, NIR spectroscopy involves radiation in the 780–2500 nm (wavenumber range 12 800–4000 cm^{-1}) region, with energy higher than in MIR. Traditionally, NIR spectra are expressed as absorbance versus wavelength (expressed in nm). Figure 4.2 shows the NIR spectra of several agro-food products. Each spectrum was collected with a FT-NIR instrument in 40 s with a resolution of 16 cm^{-1} and is the average of 64 scans.

The advantages and drawbacks of the methods based on NIR spectroscopy are various, and can be split between analytical, spectroscopic and instrumental features. The *analytical* advantages include speed; no sample preparation; no requirement for chemical reagents; being non-destructive; the possibility to perform qualitative and quantitative analysis, and to handle almost all kind of samples irrespective of their size or shape; the relatively low cost per analysis; and the opportunity to perform direct, non-invasive and *in situ* analysis. The main analytical limitations are the need to calibrate the spectrometer, usually requiring usually hundreds or thousands of spectra with reference values and the use of chemometrics, as well as the limited number of available validated methods according to international standards.

From a *spectroscopic* point of view, NIR spectroscopy has the advantages of providing spectra with a high intensity and high resolution, and a precise spectral frequency measurement; being fluorescence-free; and ease of sample presentation. Regarding limitations, this technique is characterized by poorly-resolved spectra; the absence of information from non-polar groups; a lack of structural selectivity and of sensitivity; and spectra influenced by temperature changes.

The *instrumental* advantages of NIR spectroscopy include the marketing of push-button instrumentation; the possibility to work with aqueous samples; its suitability for at-, on- and in-line process control; its compatibility with long fiber-optics; its suitability for process monitoring; and the existence of hyphenated techniques such as NIR microscopes and imaging instruments. The main instrumental drawback is the fact that there are no officially accepted and agreed standards by all manufacturing companies for sample presentation and software to handle and exploit data treatment. A more detailed description, as well as references discussing these different

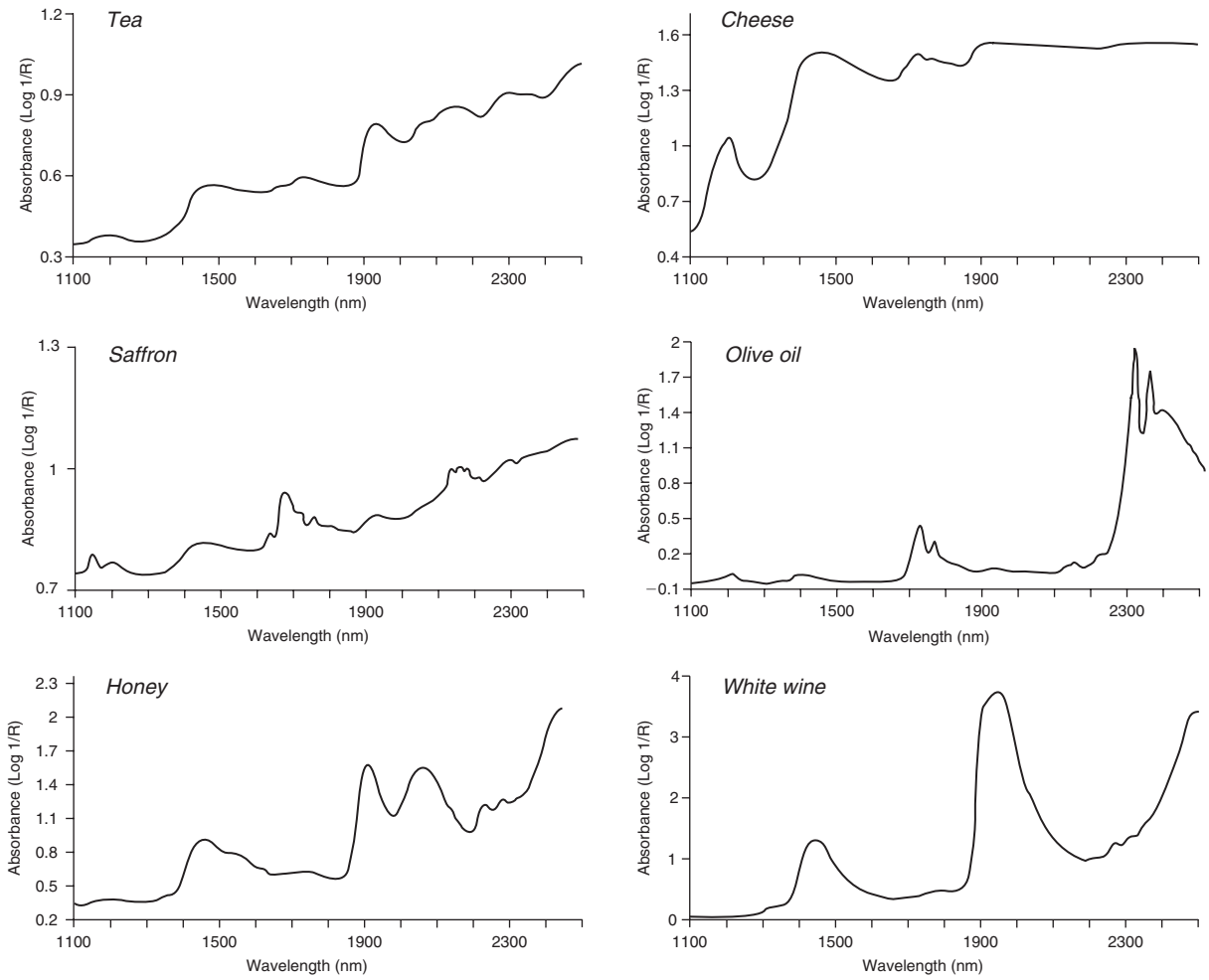

Figure 4.2 Representative NIR spectra of various agro food products (tea, saffron, honey, cheese, olive oil, white wine). Spectra have been collected with a FT-NIR spectrometer (MPA, Brüker, resolution 16 cm^{-1}, scans 60, analytical time 40 s) (Source: CRA-W).

advantages and drawbacks not only for NIR but also for MIR and Raman spectroscopy, can be found in the publication by Baeten and Dardenne (2002).

Fourier transform near-infrared spectroscopy: instrumentation features

The evolution of instrumentation in NIR spectroscopy is very rapid, particularly in process analytical chemistry, which is critical for pharmaceutical, chemical and agro-food industries. A general trend is that the analyses are moving closer to the points of sampling by means of fiber-optics allowing real-time and continuous analysis. An additional development is that of dedicated instruments combining the instrument, the interface between the instrument and the sample, as well as software integrating

data acquisition, chemometrics and data archiving for specific applications (e.g. NIR wine analysis, NIR feed or food analysis) (Ciurczak, 1991).

NIR instruments can be classified into three groups. The first group includes sequential instruments in which absorbance measurements for the respective wavelengths are collected sequentially in time. In this group we find all the instruments using monochromators, filters or other devices allowing the sequential selection of wavelengths. The second group consists of multichannel instruments having several detectors that separately record the absorbance values at several wavelengths, such as diode array instruments. The third type of spectrometer regroups the multiplex instruments in which the detector simultaneously collects information at several frequencies. Fourier transform instruments are the most common of this type of instrument (Bertrand and Baeten, 2006). Sequential and multichannel instruments are extensively described in Chapter 3 of this book. In the following paragraphs, attention is paid to multiplex instruments and more specifically to FT-NIR instruments.

Multiplex instruments based on the use of interferometers are FT-NIR spectrometers combining most of the best features in terms of wavelength precision, accuracy, high signal-to-noise ratio and high scan speed. These instruments have gained more and more importance in the last decade. They have a light source emitting in the NIR range and directing radiation to the interferometer. For example, if radiation with wavelength λ is sent to the interferometer, the radiation λ is sent to a beam splitter that reflects approximately half of the incident radiation and transmits the other half. The reflected part of the radiation encounters a stationary mirror while the transmitted part is sent to a second mirror; both parts of the radiation are recombined at the level of the beam splitter and directed to the sample. Because the second mirror is moving, the pathways to and from the movable mirror are variable as a function of mirror position. At different mirror positions, a difference (also referred to as retardation, δ) in path-length produces interference – both constructive and destructive interferences can occur. Constructive interference will occur when the retardation of the two mirrors is equal to $\delta = n\lambda$ (where n is an integer); the interference will be destructive when the retardation is equal to $n\lambda/2$ (where n is odd) (Williams and Norris, 2001; King *et al*., 2004). The most common interferometer is the Michelson interferometer (Williams and Norris, 2001), which includes a beam splitter, stationary and moving mirrors, and a laser to follow the position of the moving mirror. Figure 4.3 shows the schema of the Michelson interferometer. Data accumulated during the motion of the moving mirror, information in the time domain, is transformed into information in the frequency domain through application of the fast Fourier transformation.

Advantages of Fourier transform infrared spectrometers are numerous. First, interferometers allow radiation at a range of wavelengths to be produced near-simultaneously, decreasing the time require to acquire a full spectrum. This advantage is known as the *multiplex or Fellgett advantage*. They also have a throughput advantage, as interferometers have no entrance or exit slits and all the NIR radiation passes through, is emitted or reflected from the sample and reaches the detector at once. This advantage is called the *Jacquinot advantage* (Williams and Norris, 2001). Fourier transform-based instruments also offer excellent resolution and wavelength

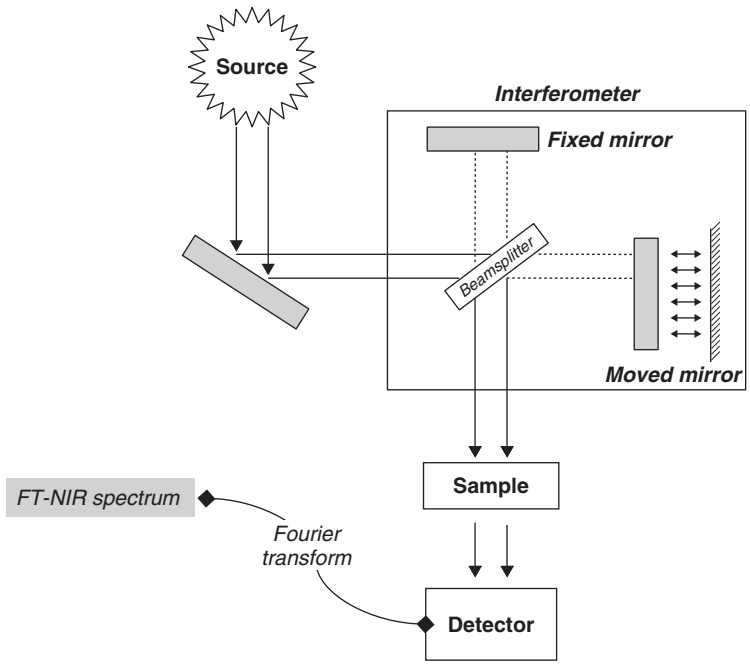

Figure 4.3 Schema of a typical Michelson interferometer including the beam splitter, the stationary mirror and the moving mirror.

reproduction; this is possible through the use of a laser in a path parallel to the NIR path to allow verification of the moving mirror position (Williams and Norris, 2001; King *et al*., 2003).

FT-NIR allows instrument manufacturers to develop new types of instrumentation, such as the introduction of hyphenated techniques combining microscopy and FT-NIR spectroscopy. Fourier transform near-infrared microscopes allow analysts to switch from macroscopic to microscopic analysis. Infrared microscopy is routinely used as a standard laboratory procedure in forensic analysis. Actually, commercial NIR microscopes allow spectra to be collected from extremely small sample areas (5 µm × 5 µm). Such instruments include a camera and a viewing system for magnifying the visible light image of the sample to be analyzed, allowing the identification and the isolation of one point or a series of points of interest. The device allows the collection of spectra at a large number of sample points from an inhomogeneous surface (e.g. particles from food or feed meal, slices of salami) and produces a NIR map or NIR cube. Figure 4.4 presents the schema of a typical NIR microscope, while Figure 4.5 shows a picture of a NIR microscope (Perkin-Elmer photo image instrument) as well as the spectra collected from a salami sausage sliced (the spectra were collected from pieces of meat and fat). Infrared maps are obtained by the procedure of mapping that allows the automatic and sequential collection of near-infrared spectra of a series of points. The incorporation of multichannel detectors in recent NIR microscopes has made this kind of instrument more powerful because of the

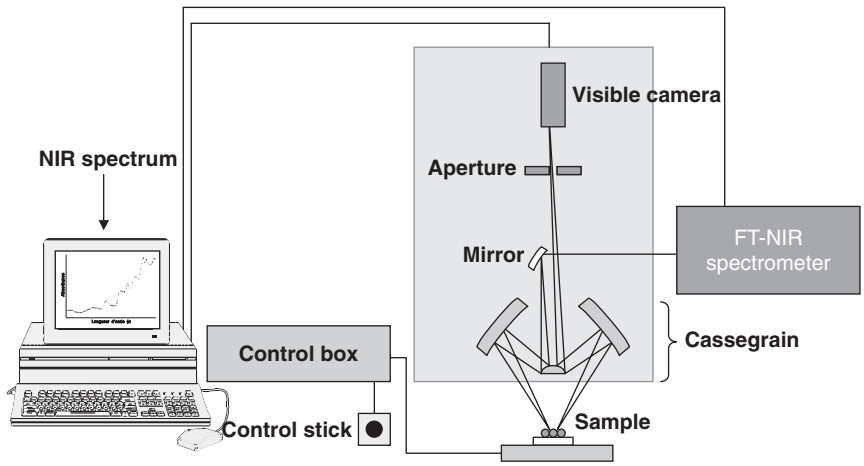

Figure 4.4 Schema of a typical near infrared microscope.

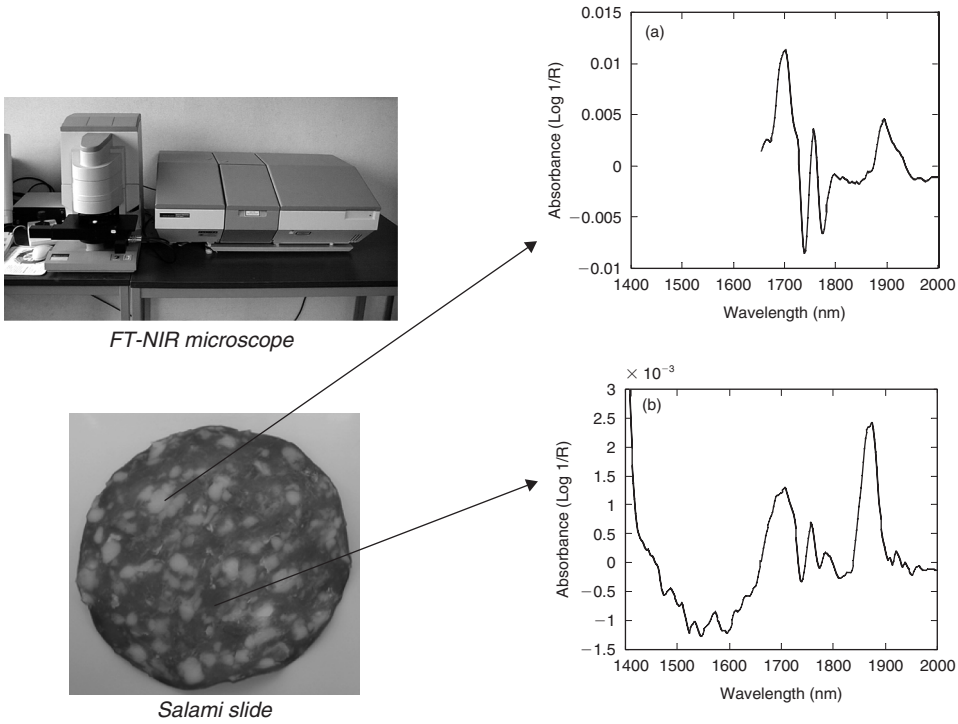

Figure 4.5 Example of a NIR microscope (Perkin-Elmer photo image instrument) and spectra collected from a salami sausage slide: (a) spectrum of meat area; (b) spectrum of fat area.

simultaneous acquisition of spectral data from several points. A multichannel detector includes several photoelectric detector elements and permits the simultaneous recording of reflected or transmitted light from a number of points. Depending on the instrument type, two kinds of camera can be distinguished; one-dimensional

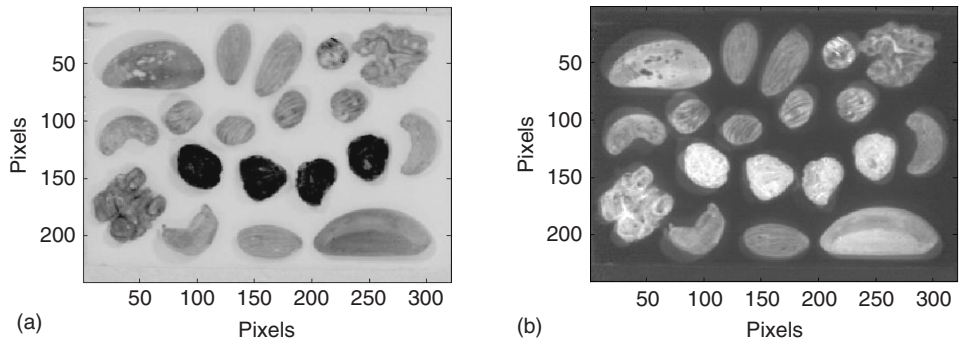

Figure 4.6 NIR images at 1420 nm and first principal component (PC1) image of a mix of dried fruits.

(i.e. a line scan camera) and two-dimensional (Baeten and Dardenne, 2002). Figure 4.6 presents the NIR images at two different wavelengths of a mixture of dried fruits.

New trends in chemometrics as applied to NIR spectroscopic data

In this section, the aim is not to give a summary of all the possible chemometric solutions for the handling, transformation and exploitation of NIR data. The focus will mainly be on two chemometric tools which, in recent years, have proved to be adequate to solve specific problems, such as co-linearity and non-linearity, associated with spectroscopic data: these are *artificial neural networks (ANNs)* and *support vector machines* (*SVM*). These chemometric methods are applied in order to automate the extraction of information from NIR data and to reduce the need for constant expert analysis of data. Chapters 3 and 16 give a complete overview of the mathematical techniques commonly used.

Artificial neural networks (ANNs) for authentication using spectroscopic data

An artificial neural network (ANN) is an information-processing system designed to mimic functions of the human brain, i.e. it is based on generalizations of human cognition. Among the many applications of ANNs, classification is perhaps the most interesting for data mining. In this case, the network is trained to classify certain patterns into specific groups and is then used to classify novel patterns which have never been presented to the network before. Many applications demonstrate the suitability of ANNs for classification (authentication) as well as modeling tasks. ANNs are well-known in the area of biometry for fingerprint, face or eye identification, as well as for handwritten or signature authentication. They have also been widely-used in spectroscopy, and several papers deal with the problem of authentication of various products.

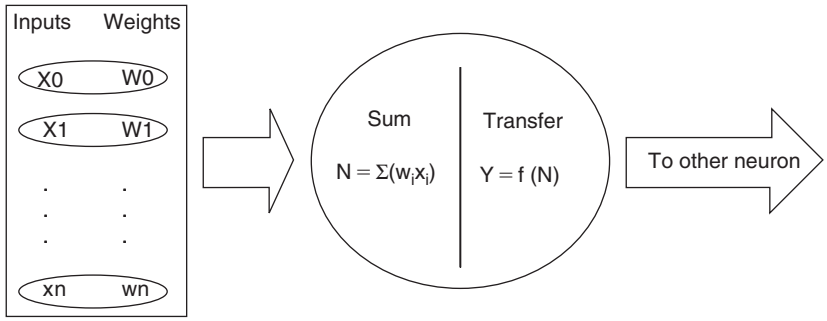

Figure 4.7 ANN – representation of a simple neuron in which all the connection weights for each input are summed, resulting in a unique complex function each time the neural network is trained with a set of inputs and outputs (w = weight).

In the *Handbook of Expert Systems in Manufacturing*, VerDuin (1991) describes how neural networks are used during the manufacture of pharmaceuticals, chemicals, rubber, plastics, metals, ceramics and foods.

The structure of an ANN consists of many simple elements called neurons. Neurons are connected via synapses (connection links) that modulate signals passing through them; each synapse has an associated weight w. The net input N is the function of all transmitted signals x_i and their corresponding weights w_i in a neuron: $N = \zeta(w_i x_i)$ (weighted sum of inputs). Each neuron applies an activation (transfer) function to its net input N in order to provide an output signal for each neuron. The output of this function is the output (activation) of another neuron connected as an input to other neurons (Figure 4.7).

A neural network is characterized by its architecture or the pattern of connections between the neurons. Neurons are arranged in several layers: an input layer that receives the inputs, a hidden layer(s) which transforms the input representation into a new "hidden" representation, and an output layer, the units of which send the predicted values out (i.e. the class label). Input data are signals x_i of the input layer, and initial weights are random values. Then an activation (transfer) function is applied, which determines the output. Normally a hyperbolic tangent function is chosen as the transfer function.

An important point is the learning algorithm or the method of determining the weights on the connections. Before using a network for prediction, it must be trained with known data. This is necessary to ensure that the network provides useful results. The most commonly-used learning algorithm is based on the "back-propagation of errors". While learning, the network compares its output with observed (known) output values of learning data. The effectiveness is usually determined in terms of the root mean square (RMS) error between the actual and the desired outputs averaged over the learning data. After comparison, the network changes weights backwards from output layer to input layer with respect to the output error.

The advantage of the ANN is its capacity for adaptation, i.e. its auto-organization and learning procedure, as well as good generalization ability. However, as explained by Despagne *et al*. (1998), this flexibility can become a pitfall because the number of

weights in an ANN is such that the training data will be rapidly overfitted when the number of samples is too low.

Support vector machines (SVM) for authentication using spectroscopic data

Support vector machines are a relatively new learning method used for binary classification. SVMs are classifiers which have demonstrated high generalization capabilities in many different tasks, including authentication. Several papers have reported on object recognition problems such as face or fingerprint authentication systems (Zhou *et al*., 2007) as well as authentication in food and feed products (Fernández Ocana *et al*., 2004; Fernández Pierna *et al*., 2005a).

The main idea of SVM is to find a decision boundary or hyperplane that separates the data perfectly into two (or more) classes. However, since the data are often not linearly separable, SVM needs to map the data from the initial (wavelength) space to a new higher-dimensional feature space in which the data can be linearly separable. Fortunately, SVM introduces the notion of the kernel trick; the advantage is that this higher-dimensional feature space does not need to be dealt with directly. As long as some necessary conditions are met, some mathematical functions are available to be considered to perform the mapping to the higher space. Linear, polynomial or radial basis Gaussian (RBF) functions are the most widely used kernels. The main difficulty when using SVM is to determine the optimal model, i.e. the optimization of the two parameters presented in SVM – the penalty C that has to be added in order to take into account those samples that cannot be separated, and the width of the Gaussian function σ in the case that a kernel is used. Figure 4.8a shows the discrimination model (or hyperplane) between olive oil and hazelnut oil when a linear kernel is applied to spectroscopic data. Figures 4.8b and 4.8c show the model with two different combinations of C and σ. The best generalization is found when using a linear kernel, or for RBF with $C = 100$ and $\sigma = 0.25$ (Figure 4.8b), i.e. when the kernel is chosen to be close to linearity and when too many objects are misclassified. It can be seen that even if some points are misclassified, the generalization for prediction is expected to be good. However, while better discrimination can be obtained using, for instance, $C = 100$ and $\sigma = 0.035$, the generalization to new samples will not be as good as for the previous model (Figure 4.8c). For this reason, the selection of the parameters for a SVM model is a very important point depending on the aim of the user; in most cases it is better to assure a good prediction for the majority of the data than to have some misclassification errors.

Capron *et al*. (2007) studied the authentication of wines from third countries using the content of 63 different chemical parameters. For this, a database containing more than 1000 samples of authentic and commercial wines from Hungary, the Czech Republic, Romania and South Africa was created. The aim of the study was to evaluate whether it is possible to determine the country of origin of a wine based on its chemical content. Multivariate tools such as partial least squares (PLS) regression, classification and regression trees (CART) and SVM were applied. One SVM model authenticated the genuine wine samples with a success rate of more than 94%.

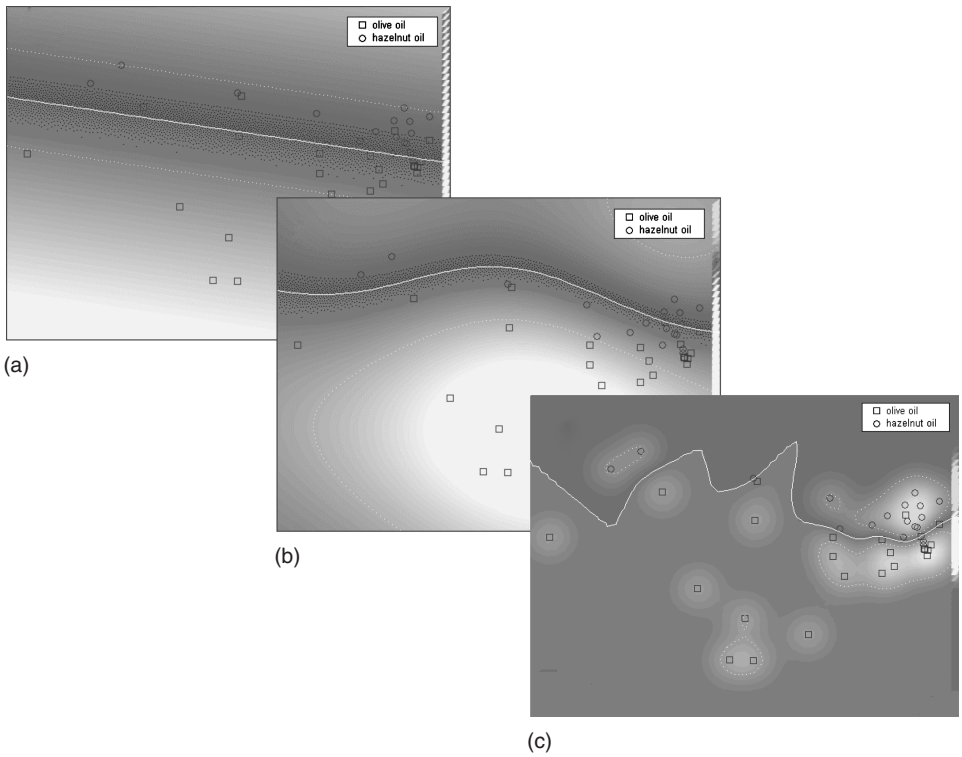

Figure 4.8 (a) SVM model using a linear kernel; (b) SVM model using a RBF kernel with C = 100 and $\sigma = 0.25$; (c) SVM model using a RBF kernel with C = 100 and $\sigma = 0.035$.

For the discrimination of commercial samples, the SVM model determined the country of origin of a wine with a correct classification rate of more than 90%.

Cogdill and Dardenne (2004) presented SVM in a familiar way for people working in chemometrics and NIR spectroscopy. Datasets of NIR spectra and reference values were compiled for apples, meat and corn, and were used for regression analysis. Each consisted of spectra from a typical NIR analyzer, and each sample type had multiple analytes. A second example consisted of animal feed spectra in order to apply discriminant analysis with the objective of detecting meat and bonemeal contamination of ruminant feed. In their study, they concluded that SVM gave a better predictive performance compared to other techniques such as discriminant PLS or ANN, and that SVM methodology has a place in NIR spectroscopy and chemometrics.

Fernández Pierna *et al*. (2005a) studied the development of a new system to detect meat and bone meal (MBM) in compound feed, which will be used to enforce legislation concerning feedstuffs enacted after the European mad cow crisis. For this, data obtained by a NIR imaging spectroscopy system have been analyzed using PLS, ANN and SVM. Although all three chemometric methods were able to model the data effectively, SVM was found to perform substantially better than PLS and ANN, exhibiting a high correct classification rate (>93%) and a much lower rate of false-positive identification (<0.4%). Subsequently, a classification of starches according

to the type of chemical modification was performed by applying different supervised discrimination methods to their associated IR data (Fernández Pierna *et al*., 2005b). Representative samples of each group were available for which the relevant characteristics (chemical modification) were known. SVM showed a correct classification rate of more than 95% when performing leave-one-out cross-validation, and more than 80% for an independent test sample set.

SVM is one of the few computationally-efficient approaches with a well-defined theory which explain its accuracy and robustness. The main advantages of SVM are its ability to minimize the generalization error and to apply non-linear classifiers by mapping the input space to a high-dimensional feature space where linear classification can be performed. Thanks to the kernel trick, SVM can be deployed by using different kernel functions, making SVM independent of the dimensionality of this feature space.

Authentication by FT-NIR

In this section, papers presenting the application of FT-NIR spectroscopic techniques for the authentication of agro-food products are reviewed. Emphasis is put on the aim of the study, the features of the spectroscopic analysis, the chemometric tools used and the results achieved.

Geographic origin assessment

Geographic origin of cheese

The research teams of Bosset and Dufour (Pillonel *et al*., 2003; Karoui *et al*., 2005; Karoui and Debaerdemaker, 2007) studied the determination of the geographic origin of cheese using spectroscopic methods. They developed rapid, economical, non-destructive and multi-parametric methods for the geographic origin assessment of Emmental cheese and, more generally, the geographic origin of European hard cheese. Each region produces a cheese with typical features, such as the ripening time that can vary from 6 weeks to several months. The originality of the approach used by these authors lies in the fact that they simultaneously investigated several spectroscopic methods (NIR, MIR and fluorescence) and how they complement one another. NIR spectra of cheese present several absorption bands characteristic of overtones and combinations of C–H, N–H and O–H bonds (see Figure 4.2 for an example of cheese NIR spectrum). The spectra are mainly influenced by the O–H groups of water absorption bands (1470 and 1940 nm) and the C–H_2 groups of lipids and proteins (2173, 2350 and 2380 nm).

The FT-NIR instrument used (NIRLab N-200, Büchi Labortechnik AG, Flawil, Switzerland) was initially manufactured as an inspection tool to establish and authenticate chemical stock in factories. The NIRLab spectrometer uses a refractive wedge, which oscillates back and forth in the radiation beam to produce a difference in light path based on thickness and refractive index. It uses polarizers on either side of the dual refracting prisms. The moving prism has large amplitude of oscillation.

The consequence is that the oscillation takes place rather slowly and influences the speed of the motion. However, large amplitude allows precise repositioning (Williams and Norris, 2001).

In the first study involving FT-NIR for cheese authentication, Pillonel *et al*. (2003) investigated the potential of NIR and MIR to discriminate between the different geographic origins of Emmental cheeses. They investigated 20 cheese samples from France (Savoie and Bretagne regions), Germany (Allgaü region), Austria (Voralberg region), Finland (Middle region) and Switzerland. In this study, about 150 g of grated cheese was placed in a glass Petri dish and measured by diffuse reflection. Spectra consisted of the means of 64 co-added scans recorded from 1000 to 2500 nm with a spectral resolution of 1.25 nm (2 cm^{-1}). The authors selected spectral regions in order to eliminate zones with low signal-to-noise ratio or with no significant spectral information. To explore and exploit the information included in the spectra, principal component analysis (PCA) and linear discriminant analysis (LDA) were applied. PCA was chosen in order to reduce the number of variables, since principal component scores were used as input for the LDA technique. For LDA, the stepwise backward procedure was used. In addition, the authors used the jackknife classification test to evaluate the robustness of the discriminant functions. The goal of the application of these multivariate statistics was to asses the feasibility of NIR to address the authenticity issue. They focused initially on the discrimination between cheese samples from Switzerland and those coming from the other countries, and afterwards between samples from all the regions combined. LDA allowed a total classification of the cheese samples investigated, since 100% correct classification was obtained for the Emmental cheeses produced in the six European regions. Based on the median normalized distance calculated from the Switzerland group, they determined that the Finland samples were always the easiest group to discriminate. The study concluded the promising potential of the NIR technique. However, with only 20 samples, the models suffered from over-fitting and were consequently not very robust.

The work of Pillonel *et al*. (2003) was followed by a study including the analysis of a larger number of samples and involving NIR, MIR and front-face fluorescence spectroscopy (Karoui *et al*., 2005). In this study, 91 Emmental cheeses from different European countries (Austria, Finland, Germany, France and Switzerland) were investigated using a Büchi NIRLab N-200. The analyzed samples ranged in age from 12 to 16 weeks, and reflected the normal ageing time of commercial cheeses. Spectra were normalized by reducing the area under each spectrum to 1. PCA, factoral discriminant analysis (FDA) and common component and specific weights analysis (CCSWA) were applied. The aim of FDA was to predict membership of an individual cheese to the defined groups; CCSWA was used in order to describe the several spectroscopic data sets obtained on the analyzed samples. CCSWA deals with the total variance in data sets.

After performing FDA, the cheeses were discriminated according to their geographical origin. The first discriminant factor (63% of the total variability) allowed separation of the cheeses from Switzerland and Finland from those from Austria, France and Germany. Using all the discriminant factors, 100% of the cheeses from Austria were correctly classified, followed by 94.7%, 83.3%, 76.9% and 66.7% correct

classification for cheeses from Switzerland, France, Germany and Finland, respectively. As FDA was applied on the first 20 PCs of the PCA, it may be that the data were over-fitted, which would have increased the rate of correct classification. Therefore, the authors concluded that NIR allowed a fairly good recognition of the geographic origin of Emmental cheese. The CCSWA applied on all the spectroscopic and physicochemical data permits the conclusion that the spectral data obtained by infrared (NIR and MIR) and fluorescence spectroscopic methods were independent, and that the two first common components were related to different phenomena observed. The conclusion of the study stated that infrared spectroscopy in combination with chemometrics can be applied to characterize the geographical origin of various dairy products; front-face fluorescence spectroscopy in combination with chemometrics may be used for the identification of cheeses made from either raw or pasteurized milk.

Geographic origin of rice wines

Yu *et al*. (2007) have worked on the discrimination of Chinese rice wine of different geographical origins by NIR. Chinese wine is a sweet, golden wine made from glutinous rice and wheat. Although the Shaoxing rice wine is protected by a standard which defines it, it suffers from unfair competition by Chinese rice wine coming from other geographical origins and sold as Shaoxing rice wine. The NIR spectrum of rice wine spectrum is mainly influenced by absorption bands of O–H groups in water, ethanol absorption bands of C–H and O–H groups (2266 and 2305 nm) and sugar absorption bands (1790 nm). A NIR spectrum of sweet white wine is presented in Figure 4.2. Yu and colleagues used a Nexus FT-NIR spectrometer from the Thermo Nicolet Corporation; this instrument is equipped with a Michelson interferometer, an InGaAs detector and a quartz halogen tungsten lamp (50 W) as a broadband light source. The rice wine samples were analyzed in transmission in a 1-mm quartz cell. Air was used as a reference, and spectra were collected from 800 to 2500 nm with 32 co-added scans and a resolution of 10 nm (16 cm^{-1}). The chemometric tools used in this study were PCA and discriminant PLS using leave-one-out cross-validation, and dummy variables were used as reference values.

In their study, Yu and colleagues analyzed 38 bottles of Chinese rice wine samples of two different brands (i.e. Pagoda brand Shaoxing and Fen Lake brand Jiashan); 29 and 9 samples respectively were used for the calibration and validation sets. The NIR spectra of the samples showed some differences in the 1450-nm region, where the absorption intensities of Jiashan rice wine were slightly higher, while Shaoxing rice wine samples had higher absorption intensities in the 2266- and 2305-nm regions. PCA allowed, on the basis of the two first principal components, the discrimination of samples of the two brands from two different geographic origins in China. Analysis of the eigenvector of the two first principal components indicated that the discrimination is based on bands centered near 1410, 1450, 1884, 2064, 2336 and 2370 nm associated with O–H, C=O and C–H groups; these regions are characteristic of water, ethanol and sugar absorption bands. Using PLS regression in order to construct discriminant functions, it has been shown that the wavelength range of 1300–1650 nm gives the best calibration results in comparison to those obtained with

the full spectral range (i.e. 800–2500 nm). Of the samples in the validation set, 100% were correctly classified.

Geographic origin of saffron

The potential of NIR spectroscopy for the assessment of the geographical origin of saffron has been investigated (Zalacain *et al*., 2005). Saffron consists of the dried stigmas of *Crocus sativus* L. The price of this spice depends on its quality and its geographic origin. Zalacain and colleagues analyzed 111 samples of saffron from leading producers in Iran, Greece and Spain. Near-infrared analysis was performed using a Perkin-Elmer FT-NIR instrument equipped with a near-infrared reflectance accessory. The samples were ground and passed through a 0.5-mm sieve before NIR analysis. Approximately 2 g of each powdered saffron sample was placed onto a quartz sample plate and spectra were collected in the 1000–2500 nm range. The authors performed qualitative and quantitative multivariate analysis using principal component regression (PCR) and discriminant analysis (DA). In order to minimize the risk of overfitting, the standard error of validation was used to select the calibration equation. A spectrum of saffron (Figure 4.2) is highly influenced by the spectral profile of crocetin glycosides which constitute the major component of this material. The first stage of the work concerned the prediction of the chemical composition of the samples by NIRS 9 and 13 principal components were used to calibrate the spectrometer for the different parameters. The authors stress the importance of the determination of the moisture content – a very important parameter, as it is fraudulent to sell water at saffron price. The second stage of the study concerned the geographical origin discrimination of the saffron samples studied. The correct identification rates were 100%, 95% and 88% for Iranian, Greek and Spanish samples respectively. The interclass distances showed that the Iranian samples were the most different from the Greek and Spanish, which were very similar.

Variety and species assessment

Discrimination of edible oil and fat sources

Discrimination of oils and fats by chemical and physical techniques has been extensively studied by various authors (Baeten *et al*., 2001a). The interest of this topic is the high price differential of oils and fats coming from different sources. The best example is the olive-oil product, which has added-value compared with other vegetable oils and thus the adulteration of it is economically worthwhile. Several researchers have studied the discrimination of oils and fats by spectroscopic methods, mainly by MIR, Raman and NIR techniques (Bewig *et al*., 1994a, 1994b; Baeten *et al*., 1998a, 1998b, 1998c, 2001a; Hourant *et al*., 2000; Yang *et al*., 2003). Considering FT-NIR spectroscopy, the research team of Van de Voort studied its potential to determine parameters such as peroxide value, *cis* and *trans* fatty acid content, iodine value and saponification number, as well as to discriminate edible oils (Dong *et al*., 1997; Li *et al*., 2000a, 2000b, 2000c). The advantage of NIR techniques in comparison with the other vibrational spectroscopy techniques lies in the ease of sample handling

(disposable vials can be used, thus reducing the time of sample preparation and the avoiding the necessity for detergent to clean the cell), the possibility of producing at-line, on-line and in-line analytical solutions, and the suitability of this technique for remote use through the use of low-cost fiber-optics. An additional advantage of FT-NIR put to the fore by the team of Van de Voort is the ease of maintaining calibration stability with these instruments, in comparison to the challenge that this topic represents with dispersive NIR instruments.

In their study to discriminate edible oil products, Li *et al*. (2000a) reported on the capability of FT-NIR spectroscopy as a practical at-line process control tool for discriminating various formulated oil products. A typical spectrum of an olive oil is presented in Figure 4.2. NIR spectra of oils and fats are mainly characterized by absorption bands in the vicinities of 1720 nm (C–H vibration of –CH3, –CH2 and =C–H groups), 2140 nm (C–H vibration of =C–H groups) and the 2100–2300 nm region (C–H vibration of –CH3 and –CH2 groups). Li and colleagues used a Bomem FT-NIR analyzer equipped with a deuterated triglycine sulphate (DTGS) detector capable of scanning the spectral range between 833 and 5000 nm. The sample holder employed was a heat-controlled device which permitted the use of glass vials of different diameters. These authors used vials of 8 mm in diameter filled with about 0.5–0.7 ml of sample and temperature controlled at 75 ± 0.2°C. The spectra were collected from 833 to 2222 nm at a resolution of 10 nm; 128 scans per sample were co-added. An air background with an empty vial holder was used. Sample spectra were ratioed against the corresponding air background, and subsequently normalized to take into account inherent variation in the vial path-length. The samples used in this study corresponded to four calibration sets of fats and oils (provided by an oil processor) with different iodine value ranges (i.e. group A, 133.3–134.8; group B, 91.3–96.3; group C, 117.1–118.8; and group D, 113.7–117.0) and one validation set of 35 unknown samples. PLS regression was used as a basis for classification by the setting of discrimination criteria based on the output obtained from calibration models: NIR predicted iodine value; the spectral residual and the factor scores were used to discriminate the different groups. The NIR predicted iodine value criterion; the factor scores criterion as well as the combination of the three discrimination criteria gave the best results. The original part of the work was the definition of two sets of limits – (i) the inner region representing zones where a sample is considered to form part of the targeted population and (ii) the intermediate region where the unknown samples are likely to be part of the population – and the outside region where the unknown samples are classified as definitely not belonging to the population. The inner region was defined on the basis of PLS regression results obtained from the calibration set. The intermediate region for each criterion was included in order to be able to adjust the limits according to the deviation accepted by the manufacturer. The authors used the limits of the inner region added or subtracted by the root mean square error obtained from cross-validation of the calibration step in order to set the intermediate region limits for the targeted criterion. Li *et al*. (2000a) concluded, on the basis of the correct prediction of the validation set samples, that their approach allowed discrimination of the samples according to their group of origin. Moreover, the authors underlined the power of the approach to detect unusual samples (or, more

exactly, samples with characteristics not included in the calibration stage, i.e. a blend of samples from different categories) as not coming from one of the pre-defined groups. FT-NIR with the combined predictive and discriminant capabilities of PLS is presented as a powerful and practical analytical quality control tool.

Yang *et al*. (2007) also studied the discriminant possibilities offered by FT-NIR for the discrimination of edible oils. In their work they compared the FT-NIR results with those of FT-MIR and FT-Raman obtained on the same samples. For their near-infrared analysis they used a Nicolet 870 spectrometer including a DTGS detector and a transmission quartz cell which was cleaned with pure chloroform and dried with nitrogen gas. Each spectrum corresponded to 256 co-added scans with a resolution of 10 nm. Spectra were collected between 1250 and 5000 nm using an air spectrum as a background. Samples such as butter, cod-liver oil, lard, canola oil, coconut oil, corn oil, olive oil, peanut oil, safflower oil and soybean oil were studied. A total of 80 samples and 30 samples were used for the calibration and validation sets respectively, although the calibration and validation sets were not fully independent as the samples were issued from the same batch. The authors used linear discriminant analysis (LDA) and canonical variate analysis (CVA) in combination with principal component analysis (PCA) and partial least squares (PLS) as data compression methods. Calculated percentages of correct classification for the validation samples were between 85.6 and 92.2% for the 500–1250 nm region and 84.4–93.3 for the 1540–2500 nm region. The authors concluded that FT-MIR (95.6–98.9% correct classification of validation samples), FT-Raman (85.6–94.4% correct classification of validation samples) and FT-NIR spectroscopy (84.4–93.3% correct classification of validation samples) techniques can be used for rapid classification of edible oils and fats without the need for sample preparation. In this study, the least efficient technique seemed to be the FT-NIR spectroscopic method.

Detection and quantification of olive-oil adulteration is also an important challenge in the discrimination of vegetable oils. Several spectroscopic techniques have been tested in order to evaluate their potential in detection of this kind of olive oil fraud (Baeten *et al*., 2001a). Kasemsumran *et al*. (2005) studied the potential of FT-NIR and PLS processing to discriminate and quantify adulterated olive oils. They used a Bruker Vector 22/N FT-NIR spectrometer equipped with an InGaAs detector. Spectra were collected in the transmittance mode from the 833 to 2198 nm region with a resolution of 2.5 nm and 32 co-added scans. The sample temperature was kept at 25 ± 0.2°C. Spectra were treated by means of multiplicative scatter correction (MSC), Savitsky-Golay first derivative and Savitsky-Golay smoothing before applying multivariate analysis. PLS regression analyses were used to calibrate the spectrometer to discriminate between the adulterated and the genuine samples. Four NIR regions were tested separately to build the PLS models. Calibration and validation sets included 200 and 80 mixtures respectively, spiked with one of the four different adulterants studied (i.e. corn oil, hazelnut oil, soya oil or sunflower oil). Olive oil was mixed with the adulterants at different percentages (i.e. 2–50% w/w); a weakness of the study, however, was that the authors selected only one sample for each type of oil considered. PLS models were constructed for each adulterant type and for all considered adulterants. All the discriminant PLS models for classifying the adulterant

types in olive oils gave a correct classification rate higher than 95%, irrespective of the data pre-treatment or the spectral range used. The best prediction results of the PLS calibration models had a root mean square error of prediction (RMSEP) lower than 0.5 for corn-oil, hazelnut-oil and soya-oil adulterants, and lower than 1.1 for sunflower-oil adulterant.

Another interesting study to consider in this review of the FT-NIR methods proposed for oils discrimination is that published by Oliveira *et al*. (2007). In this study, the authors investigated the potential of FT-NIR and FT-Raman spectroscopy for the detection of diesel/biodiesel in vegetable oil. Oliveira and colleagues used a Bruker Equinox 55 FT-NIR instrument equipped with a germanium detector. Spectra were recorded using an immersion transflectance accessory with an optical path-length of 2 mm. The spectral resolution was set at 5 nm, and each spectrum was the result of 16 co-added NIR measurements. PCR, PLS and ANN calibration procedures were tested. One hundred and seventy-five blends corresponding to mixtures of diesel/biodiesel with vegetable oils (i.e. soybean oil, castor oil, palm-tree oil) in the range of 0–5% (w/w) were used in order to calibrate the spectrometer. An independent validation set was used to test the established models. For the construction of the PLS and PCR calibration models, selection of the spectral region to be used has been carried out on the basis of two experimental parameters: (i) the spectral distribution of the standard deviation in the absorbance values for the set of samples used in the calibration stage, and (ii) the spectral distribution of the relative standard deviation in the absorbance for a reduced set of representative samples. The former parameter allows the selection of the spectral region presenting the largest variation that includes most of the variability in the calibration samples, while the latter makes it possible to put to the fore the regions presenting poor signal-to-noise ratios. PCR and PLS calibration models for a quantitative detection of diesel/biodiesel blends adulterated with vegetable oils produces a RMSEP of 0.262 and 0.238 respectively. It is interesting to mention that RMSEP obtained with FT-Raman data was three to six times higher. For the ANN calibration models, the spectral regions used were selected on the basis of the spectral distribution of the standard deviation of the absorbance values of first-derivative spectra of samples used in the calibration. RMSEP of 0.371 and 0.092 for the quantitative determination of vegetable oils in diesel/biodiesel blends were obtained for the FT-NIR and FT-Raman data respectively.

Discrimination of botanical origin of honey

Ruoff and colleagues (2006) studied the botanical origin of honey using FT-NIR. The authors used a Büchi NIRLab N-200. Honey NIR spectra were collected in the 1000–2500 nm spectral range with a resolution of 2.5 nm and 64 co-added scans per sample. Honey samples were heated at 50°C for 9 hours before analysis, and poured into a clean glass Petri dish covered with the transflection plate defining a 0.6-mm path-length. Figure 4.2 presents a typical FT-NIR spectrum of honey. The most important absorption bands are observed in the 1400–2380 nm region, with a water band at 1940 nm and several bands in the 1540–2380 nm region characteristic of C–O and C–C bond vibrations of saccharides. The total of 364 honey samples from 7 years

of production originated predominantly from Switzerland. Samples were classified according to 8 honey types; 185 and 179 samples constituted the group of unifloral and multifloral samples respectively. PCA, PLS and LDA were applied to evaluate the potential of NIR data to discriminate honey samples according to their origin. Using LDA, between 29 and 100% of the unifloral honey samples from the validation set were correctly classified while only 19% of the multifloral were correctly classified. Unifloral samples coming from acacia, fir honeydew and chestnut were the easiest to discriminate by LDA.

Discrimination of pear varieties

Fourier transform near-infrared spectroscopy was also explored as a tool to discriminate samples of different pear varieties (Fu *et al*., 2007). In this study, a Nexus intelligent FT-NIR spectrometer (Thermo Electron Corporation) equipped with an InGaAs detector was used and NIR spectra were collected in the 800–2500 nm range with a resolution of 1.25 nm and a co-added scan number of 64. The authors used a fiber-optic probe to collect diffuse reflectance spectra. For each fruit sample three NIR measurements were made, each at a different location about 120° apart around the equator of the fruit. A total of 240 samples from three varieties of pear were used to study the potential of the FT-NIR technique. Spectra of fruit reveal absorption bands in the vicinity of 970, 1450 and 1940 nm, associated with the O–H vibration of water, and around 1190 and 1790 nm related to C–H vibrations of sugar and other organic matter. Fu and colleagues used discriminant analysis (DA), discriminant partial least squares (DPLS) and probalistic neural network methods (PNN) to calibrate the spectrometer. DA and DPLS models were developed using three spectral regions, i.e. 800–1500 nm, 1500–2500 nm and 800–2500 nm. The PNN method includes four layers, and is a feed-forward network with no back-propagation. The inputs used were spectral data recorded as absorbance at each wavelength. Results obtained in this study indicated the high potential of NIR to classify and discriminate fruits of different varieties. High accuracy and correct classification higher than 99% were obtained whatever the multivariate protocol used (DA, DPLS or PNN).

Detection of castor bean meal

Rodriguez-Saona and colleagues (2000) have proposed the FT-NIR technique to rapidly quantify castor bean meal (CBM) in a selection of flour-based products. CBM contains the extremely potent cytotoxic protein ricin. The authors worked with a Perkin-Elmer Spectrum Identicheck FT-NIR spectrometer, and used capped transparent vials to analyze the samples. Reflectance spectra were collected in the 1000–2500 nm NIR range with a resolution of 2.5 nm and a number of co-added scans equal to 50. Prior to calibration, spectra were centered and baseline corrected using an offset. Then, spectra were transformed using the multiplicative scatter correction pretreatment in order to correct for the scatter effect of the particles. PCA and PLSR multivariate analysis were used to explore the NIR data. PCA analysis allowed discrimination between raw materials (i.e. CBM, soybean meal, powdered meal, tofu, egg yolk and egg white). PLSR models were able to differentiate between CBM

contamination and the addition of other protein-rich products (e.g. corn meal, egg white, defatted soybean) to the matrices. On the basis of this model, quantification of CBM contamination could be determined at level of >0.3% and 0.6% (w/w) in selected wheat flour and blueberry pancake mixes.

Discrimination of leaves from different strawberry varieties

The potential of FT-NIR for the authentication and classification of strawberry leaf varieties has been investigated by López (2002). In this study, a Paragon IdentiCheck FT-NIR system was used, recording spectra from the 1000–2500 nm range at a resolution of 2.5 nm with 32 co-added scans. Five different strawberry varieties were investigated. For each variety, spectra from both sides of the leaves and of the vascular system were recorded. The FT-NIR spectral data were analyzed using PCA and SIMCA methods. While all the varieties show very similar NIR spectra, small differences were still observed. Identification of the materials was possible using interclass distances. The authors concluded that FT-NIR enables the authentication of all strawberry varieties and their origin.

Process type assessment

Discrimination of tea categories

An interesting demonstration of the potential of NIR for the discrimination of food products has appeared in work concerning the identification of tea categories (Zhao *et al*., 2006; Chen *et al*., 2007). The most popular categories of tea are the green and the black teas. Both products involve drying and roasting of the leaves, the black tea requiring an additional step of fermentation. If the fermentation is only partially carried out, the Oolong tea category is obtained. In both studies, a Nicolet Nexus 670 FT-NIR instrument was used with standard quartz cups. Spectra were collected from 909–2632 nm at a resolution of 1.205 nm and with 64 co-added scans. Figure 4.2 presents the spectrum of a green tea sample. Water absorption bands are observed in the vicinities of 1940 and 1430 nm, bands of the carboxylic groups around 1870 nm, and bands of C–H groups around 1385, 1740 and 1720 nm. During the analysis, the authors excluded the absorption bands of water as well as the 909–1111 nm region, which exhibited a high noise level. A total of 150 samples were analyzed, comprising 50 samples of each category (i.e. green tea, black tea, Oolong tea) which were obtained from a total of 9 different geographic origins. Different data pre-processing methods were tested, including standard normal variate transformation, first and second derivatives, and smoothing. Principal component analysis (PCA), support vector machines (SVM) and back-propagation artificial neural network (BP-ANN) methods were used to exploit the variation included in the spectra, and to calibrate the spectrometer for the discrimination of tea categories. PCA analysis showed that the first three principal components were adequate to discriminate between the three categories of teas, although the samples used had considerable differences in their botanical, genetic and agronomical characteristics. For the calibration of the spectrometer with SVM and BP-ANN, the samples were split into calibration and validation sets with a proportion of 3/5 and 2/5 respectively for each category. For the green, black

and Oolong teas, 95%, 100% and 90% respectively of the validation samples were correctly classified using the SVM model developed. The results of the BP-ANN model were 75%, 100% and 80% correct classifications, respectively.

Discrimination and authentication of alcoholic beverages

Pontes and colleagues (2006) proposed a strategy in which FT-NIR and chemometric methods could be used in the classification and verification of adulteration in whiskeys, brandies, rums and vodkas. The idea was that NIR spectroscopy could be used as a screening method, and more time-consuming wet chemistry analytical techniques would then only be applied to samples showing a positive adulteration result in order to confirm the NIR result. They used a Perkin-Elmer Spectrum GX FT-NIR spectrometer, and the sample was placed in a 1-mm path-length quartz flow cell. NIR absorbance values were collected from 1100–2500 nm at a resolution of 1.25 nm and using 64 co-added scans. Analysis was performed on 69 pure and adulterated samples (some of which some were adulterated with ionized water, ethanol or methanol). The sample set was divided into calibration and validation sets comprised of 40 and 29 samples respectively. Second-derivative spectra were calculated with a Savitzky-Golay filter. PCA and SIMCA methods were used for the multivariate analysis of the spectral data. PCA analysis of the calibration set allowed identification of the whiskey and vodka samples using only the two first principal components. The third and fourth PCs allowed the discrimination of brandy and rum samples. SIMCA models were constructed (n = 40) and consequently applied to classify the samples as adulterated or authentic. The prediction ability of each model was evaluated on a test set (n = 29) consisting of laboratory-prepared samples and verified adulterated alcoholic beverages supplied by a regulatory agency. The authors observed that all the samples in the test set were correctly predicted as adulterated or genuine, with a 95% confidence level.

Fourier transform near-infrared spectroscopy was also used to measure the percentage of sugar in grape must, and to discriminate between different must samples in terms of their free amino nitrogen (FAN) values (Manley *et al*., 2001). For the NIR measurements, a Perkin-Elmer Spectrum IdentiCheck spectrometer equipped with a 0.5-mm path-length quartz cell was used. Spectra of 97 samples were collected between 1000 and 2500 nm at a resolution of 2.5 nm and using 16 co-added spectra. The must samples could be classified in terms of their FAN values when SIMCA was applied as a classification method, with correct recognition rates exceeding 80% in all cases. It was also shown by these authors that FT-NIR proved to be a rapid method of discriminating between Chardonnay wine samples (n = 107) in terms of their malolactic fermentation status, using SIMCA, with recognition rates exceeding 88%. Moreover, table wines (n = 200) were also successfully discriminated in terms of their ethyl carbamate content, with recognition rates exceeding 80%. Later, Manley and her collaborators (Manley *et al*., 2003a) illustrated the potential of FT-NIR to classify four different classes of rebate brandy. The brandy samples were analyzed by collecting 16 co-added scans, from 700–2500 nm, also using a Perkin-Elmer Spectrum IdentiCheck spectrometer and presented in a 1-mm path length quartz cell. Using SIMCA, it was possible to discriminate between the hardest and the softest brandy class of one season.

Discrimination of marked age and vintage year of alcoholic beverages

Yu and colleagues (2006) investigated the classification of rice wine with different marked ages based on FT-NIR spectroscopy. They used a Nexus FT-NIR spectrometer (Thermo Nicolet Corporation) to collect the NIR spectra from the 800–2500 nm region. The samples were scanned in a 1-mm optical path-length rectangular quartz cell at 10 nm spectral resolution and using 32 co-added scans. The 69 rice wine samples were from two different brands, of three different marked ages (1, 3 and 5 years) and two vintages (2004 and 2005). PCA and DA analysis were applied. Correct classifications of 100, 94.1 and 100% were obtained for the calibration samples with marked ages of 1, 3 and 5 years respectively. The percentage of correct classification of a validation sample set was 94.4%.

The potential to classify 3-year-old brandy from different seasons (1999, 2000 and 2001) was studied by Manley *et al*. (2003b). In this study, a Perkin-Elmer Spectrum Identicheck 2.0 FT-NIR system equipped with a 1-mm path-length quartz cell was used. Spectra were collected in the 700–2500 nm range, with 16 co-added scans and a spectral resolution of 5 nm. A total of 191 samples of unblended 3-year-old brandy were analyzed. PCA and SIMCA methods were applied to evaluate the discriminant power of NIR data. PCA analysis showed that samples from the 2001 season were the easiest to discriminate, while there was an overlap between samples from the 1999 and 2000 seasons.

Discrimination of the age of cheese

FT-NIR has also shown potential for the discrimination of cheese on the basis of its age. Cattaneo *et al*. (2005) worked with a MPA FT-NIR from Bruker Optics equipped with a fiber-optic probe to study the shelf-life of Crescenza cheese. This cheese represents 40% of the Italian fresh cheese market, and is only produced from pasteurized whole cow's milk. Crescenza cheese suffers from a structural and chemical modification during shelf-life. Spectra were collected from 833–2500 nm at a resolution of 10 nm and with 16 co-added scans. A total of 126 cheese samples from two types of production using different technological processes, and having been stored for 20 days, were used. The NIR spectral data were autoscaled, and second-derivative data were acquired before multivariate analysis. PCA was used for exploratory analysis of the NIR spectra. This method allowed a satisfactory sample distribution that followed the evolution of the cheeses. A clear distinction could be made between fresh (0–6 days' storage) and aged/old Crescenza cheese samples (8–20 days' storage). Applying PCA on FT-NIR spectroscopic data of a reduced range permitted the discrimination between the aged (8–10 days' storage) and older samples (14–20 days' storage). The authors concluded that the main advantage of using a spectroscopic technique was the possibility rapidly to establish a profile for the product associated with its total composition and quality.

Authentication by FT-NIR microscopy

The first study proposing the use of FT-NIR microscopy (NIRM) for authentication purpose was that published by Piraux and Dardenne for feed authentication (Piraux

and Dardenne, 1999). They proposed the use of a new method, based on NIR microscopy, for the detection and quantification of meat and bone meal (MBM) in compound feed in order to comply with the ban following the BSE crisis. Samples were measured using an AutoImage Microscope connected to a Perkin-Elmer FT-NIR, and analyzed using ANNs. With this NIRM instrument, the infrared beam is focused on each particle of a sample using a microscope in order to collect the NIR spectrum (1100–2500 nm). A collection of several hundred spectra is made, which represents the molecular NIR signature of a particle from an ingredient in the compound feed. A predictive discriminant analysis was applied to classify particles into either meat particles or particles of a different nature – i.e. not meat. An artificial neural network (multilayer perceptron network with back-propagation based on the partial least squares scores) was used to discriminate between the respective groups. Their results showed an overall error rate lower than 0.65%, giving an indication of the potential of this technique for the detection of MBM. Additionally, Baeten *et al*. (2001b) used a near-infrared microscope for the detection and quantification of ingredients of animal origin in feedingstuffs. Satisfactory results were obtained for the detection and identification of meat meal, meat and bone meal, bone meals, blood meal, fish meal (muscle chair, bone fish and scale), feather meal, poultry meal, milk powder and egg meal. Moreover, their study proved that 0.5% MBM in a compound feedingstuff can be detected by NIR microscopy.

Gizzi *et al*. (2003) published an overview of the different tests for the detection of animal tissues in feed, including PCR, immunoassay, microscopy and NIR microscopy. In their paper, they showed the main characteristics of NIRM as well as the weaknesses of the method. The main advantages of NIRM are that (i) it is directly based on NIR information; (ii) it can be confirmed by another method (e.g. PCR) that can be used as legal evidence in case of fraud; (iii) NIRM does not require expertise; (iv) it is a non-destructive method; and (v) a single analysis could enable a wide range of feed ingredients to be detected. The weaknesses involve the need to develop sample databases, the limit of detection, and the cost of the equipment. Therefore, these authors concluded that NIRM is the most suitable method for large screening applications in terms of sample output and automation, and that this method is able to achieve limits of detection (LOD) as low as 0.1%. A few years later, Baeten *et al*. (2005) have decreased this LOD by detecting the presence of MBM at concentrations as low as 0.05% mass fraction by the use of NIR microscopy applied to the sediment fraction of the feed. More recently, De la Haba *et al*. (2007) have proposed a method based on NIRM to check the presence of ruminant tissue in fish meal or in compound feeds containing fish meal in order to allow only fish meal to be used in ruminant feed. The use of pure fish material in the animal production chain poses no risk, and it is accepted that fish do not carry Transmissible Spongiform Encephalopathy (TSE). Thus, they have worked on methods permitting the detection and identification, at species level, of animal by-products included in compound feed. NIRM spectra allowed the construction of discrimination equations using support vector machines (SVM) as a chemometric tool. As a result, clear discrimination between fish meal and meal of other animal species is possible, with a high average success rate of 95%. In contrast to optical microscopy, NIRM offers one clear advantage: it is

not dependent on the subjectivity of the analyst, because particles are identified from their NIR spectral fingerprint and not by visual inspection.

NIRM has also been applied for authentication in different areas. Wilson and Moffat (2004) used a Perkin-Elmer FT-NIR IdentiCheck spectrophotometer coupled with a microscope for the authentication of Viagra tablets. A single spectrum for each tablet was taken, and some chemometric procedures (such as PCA or PLS) were used to identify the active ingredient by comparing it with the real pure compound. This study is important in the fight against counterfeit pharmaceuticals by using the spatial distribution of the active compounds. Another area of application of NIRM is in forensic laboratories to determine, among other things, the type of explosive used in terrorist attacks, or to detect the presence of illegal drugs.

An alternative to NIRM is the use of a more recent technology called NIR imaging. This technology is a powerful approach to remote sensing in precision agriculture and mineralogy, among other areas. The success of NIR imaging can be considered as due to a combination of different factors: high-performance and uncooled NIR sensitive focal plane array detectors, digitally-tunable infrared optical filters, the drastic increase in computer speed, and the increased capacity of laboratory computing platforms. The integration of these elements has already shown promising results in the determination of quality parameters for complex matrices such as pharmaceutical blends, detection of apple surface defects and contamination or the detection of animal compounds in feeding stuff (Fernández Ocana *et al*., 2004). NIR imaging allows the contemporaneous collection of spatial and spectral (and therefore chemical) information characterizing samples under test.

Conclusions

This chapter has shown applications of FT-NIR spectroscopic and microscopy techniques for the authentication of agro-food products. The potential of this technique to assess different authenticity issues is obvious, and will be used as new approaches for at-line, on-line and in-line control. The forthcoming challenge for this technique is to develop adequate strategy for the construction of large datasets in order to calibrate any given instrument. The strategy should include the extraction of the information in such a way that it can be used in qualitative and quantitative analysis as well as the implementation in routine control.

References

Baeten, V. and Aparicio, R. (2000). Edible oils and fats authentication by Fourier transform Raman spectrometry. *Biotechnology, Agronomy, Society and Environment*, **4**(4), 196–203.

Baeten, V. and Dardenne, P. (2002). Spectroscopy: developments in instrumentation and analysis. *Grasas y Aceites*, **53**(1), 45–63.

Baeten, V., Hourant, P., Morales, M.T. and Aparicio, R. (1998a). Oil and fat classification by FT-Raman spectroscopy. *Journal of Agricultural and Food Chemistry*, **46**, 2638–2646.

Baeten, V., Morales, M. T., Aparicio, R. (1998b). In: J. Sanchez, E. Cerda-Olmedo and E. Martinez-Force, (eds), *Oil and Fat Analysis by FT-Raman Spectroscopy*, 13th International Symposium on Plant Lipids, 5–10 July, Seville, Spain. Seville: Université de Séville, pp 18–21.

Baeten, V., Morales, M.T. and Aparicio, R. (1998c). Oil and fat analysis by FT-Raman spectroscopy. In: J. Sanchez, E. Cerda-Olmedo and E. Marinez-Force (eds), *Advances in Lipids Research*. Seville: University of Seville, pp. 18–21.

Baeten, V., Dardenne, P. and Aparicio, R. (2001a). Interpretation of Fourier transform Raman spectra of the unsaponifiable matter in a selection of edible oils. *Journal of Agricultural and Food Chemistry*, **49**, 5098–5107.

Baeten, V., Michotte-Renier, A., Sinnaeve, G., and Dardenne, P. (2001b). Analysis of feeding stuffs by near infrared microscopy (NIRM): detection and quantification of meat and bone meal (MBM). In: *8th International Symposium on Food Authenticity and Safety (FASIS)*, October, Eurofins Secretariat, Nantes, France. Eurofins.

Baeten, V., Von Holst, C., Garrido Varo, A. *et al*. (2005). Detection of banned meat and bone meal in feedstuffs by near-infrared microscopic analysis of the dense sediment fraction. *Analytical Bioanalytical Chemistry*, **382**, 149–157.

Bertrand, D. and Baeten, V. (2006). Instrumention. In: D. Bertrand and E. Dufour (eds), *La spectroscopie infraouge*. Paris: Lavoisier, pp. 247–305.

Bewig, K., Clarke, A.D. and Roberts, C. (1994a). Discriminant analysis of vegetable oils by near infrared reflectance spectroscopy. *Journal of American Oil Chemical Society*, **71**(2), 195–200.

Bewig, K., Clarke, A.D., Roberts, C. and Unklesbay, N. (1994b). Discriminant analysis of vegetable oils using near-infrared spectroscopy. In: G.D. Batten, P.C. Flinn, L.A. Welsh and A.B. Blakeney (eds), *Leaping Ahead with Near Infrared Spectroscopy*. Melbourne: NIR Spectroscopy Group, Royal Australian Chemical Institute.

Brettel, T.A., Butler, J.M. and Saferstein, R. (2005). Forensic science. *Analytical Chemistry*, **77**(12), 3839–3860.

Capron, X., Walczak, B., de Noord, O.E. and Massart, D.L. (2005). A modification of the ICOMP criterion for estimation of optimum complexity of PCR models. *Journal of Chemometrics*, **19**, 308–316.

Capron, X., Verbeke, J. and Massart, D.L. (2007). Multivariate determination of the geographical origin of wines from four different countries. *Food Chemistry*, **101**, 1585–3597.

Cattaneo, T.M.P., Giardina, C., Sinelli, N. *et al*. (2005). Application of FT-NIR and FT-IR spectroscopy to study the shelf-life of Crescenza cheese. *International Dairy Journal*, **15**, 693–700.

Chen, Q., Zhao, J., Fang, C.H. and Wang, D. (2007). Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochimica Acta Part A*, **66**, 568–574.

Ciurczak, E.W. (1991). What's new in spectroscopy instrumentation?. *Spectroscopy International*, **3**(3), 18–30.

Cogdill, R.P. and Dardenne, P. (2004). Least-squares support vector machines for chemometrics; an introduction and evaluation. *Journal of Near Infrared Spectroscopy*, **12**, 93–100.

Davies, T. (1998). The history of near infrared spectroscopic analysis: past, present and future – from sleeping technique to the morning star of spectroscopy. *Analysis Magazine*, **26**, M17–M19.

De la Haba, M.J., Fernández Pierna, J.A., Fumiere, O. *et al*. (2007). Discrimination of fish bones from other animal bones in the sedimentation fraction of compound feeds by near infrared microscopy. *Journal of Near Infrared Spectroscopy*, **15**(2), 81–88.

Dennis, J. (1998). Recent developments in food authentication. *Analyst*, **123**, 151R–156R.

Despagne, F., Walczak, B. and Massart, D.L. (1998). Transfer of calibrations of near infrared spectra using neural networks. *Applied Spectroscopy*, **52**(5), 732–745.

Dong, J., Van de Voort, F.R. and Ismail, A.A. (1997). Stochiometric determination of hydroperoxides in oils by Fourier transform near-infrared spectroscopy. *Journal of American Official Analytical Chemists*, **80**(2), 345–352.

Downey, G., Daniel Kelly, J. and Petisco Rodriguez, C. (2006). Food authentication – has near infrared spectroscopy a role? *Spectroscopy Europe*, **18**(3), 10–14.

Fernández Ocana, M., Neubert, H., Przyborowska, A. *et al*. (2004). BSE control: detection of gelatine-derived peptides in animal feed by mass spectrometry. *The Analyst*, **129**, 111–115.

Fernandez Pierna, J.A., Baeten, V., Michotte Renier, A. *et al*. (2005a). Combination of SVM and NIR imaging spectroscopy for the detection of MBM in compound feeds. *Journal of Chemometrics*, **18**(7–8), 341–349.

Fernández Pierna, J.A., Volery, P., Besson, R. *et al*. (2005b). Classification of modified starches by FTIR spectroscopy using support vector machines. *Journal of Agricultural and Food and Chemistry*, **53**(17), 6581–6585.

Fu, X., Zhou, Y., Ying, Y. *et al*. (2007). Discrimination of pear varieties using three classification methods based on Near-Infrared spectroscopy. *American Society of Agricultural and Biological Engineers*, **50**(4), 1–7.

Gizzi, G., Van Raamsdonk, L.W.D., Baeten, V. *et al*. (2003). An overview of tests for animal tissues in feeds applied in response to public health concerns regarding bovine spongiform encephalopathy. *Scientific and Technical Review – Office International des Epizooties*, **22**(1), 311–331.

Hourant, P., Baeten, V., Morales, M.T. *et al*. (2000). Oil and fat classification by selected bands of near infrared spectroscopy. *Applied Spectroscopy*, **54**(8), 1168–1174.

Karoui, R. and De Baerdemaeker, J. (2007). A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. *Food Chemistry*, **102**, 621–640.

Karoui, R., Dufour, E., Pillonel, L. *et al*. (2005). The potential of combined infrared and fluorescence spectroscopies as a method of determination of the geographic origin of Emmental cheeses. *International Dairy Journal*, **15**, 287–298.

Kasemsumran, S., Kang, N., Christy, A. and Ozaki, Y. (2005). Partial least squares processing of near-infrared spectra for discrimination and quantification of adulterated olive oils. *Spectroscopy Letters*, **38**, 839–851.

King, P.L., Ramsey, M.S., McMillan, P.F. and Swayze, G. (2004). Laboratory Fourier transform infrared spectroscopy methods for geologic samples. *Mineral Association of Canada*, **33**, 57–91.

Lachenal, G. (1998a). Structural investigations and monitoring of polymerisation by IR spectroscopy. *Journal of Near Infrared Spectroscopy*, **6**, 299–306.

Lachenal, G. (1998b). Analyze par spectroscopie proche infrarouge (PIR) et applications aux polymères. *Analysis Magazine*, **26**, M20–M29.

Li, H., Van De Voort, F.R., Ismail, A.A. *et al*. (2000a). Discrimination of edible oil products and quantitative determination of their iodine value by FT NIR spectroscopy. *Journal of the American Oil Chemical Society*, **77**(1), 29–36.

Li, H., Van De Voort, F.R., Ismail, A.A. and Cox, R. (2000b). Determination of peroxide value by fourier transform near-infrared spectroscopy. *Journal of the American Oil Chemical Society*, **77**(2), 137–142.

Li, H., Van De Voort, F.R., Ismail, I.A. *et al*. (2000c). Trans Determination of Edible Oils by Fourier Transform Near-Infrared Spectroscopy. *Journal of the American Oil Chemical Society*, **77**(10), 1061–1067.

Li, Y., Brown, C.W., Sun, F.M. *et al*. (1999). Non-invasive fermentation analysis using an artificial neural network algorithm for processing near infrared spectra. *Journal of Near Infrared Spectroscopy*, **7**, 101–108.

Li, Z., Chu, X., Mouille, G. *et al*. (1999). The localization and expression of the class II starch synthases of wheat. *Plant Physiology*, **120**, 1147–1155.

López, M.G. (2002). Authentification and classification of strawberry varieties by near infrared spectral analysis of their leaves. In: A.M.C. Davies and R.K. Cho (eds), Proceeding of the 10th International Conference on Near-Infrared Spectroscopy (NIR2001), Korea, June. Wellington: NIR Publications, pp. 335–338.

Manley, M., Van Zyl, A. and Wolf, E.E.H. (2001). The evaluation of the applicability of Fourier transform near-infrared (FT-NIR) spectroscopy in the measurement of analytical parameters in must and wine. *South African Journal for Enology and Viticulture, Stellenbosch*, **22**(2), 93–100.

Manley, M., De Bruyn, N. and Downey, G. (2003a). Classification of South African brandy with near infrared spectroscopy. *NIR News*, **14**(5), 3–19.

Manley, M., De Bruyn, N. and Downey, G. (2003b). Classification of three year old, unblended South African brandy with near infrared spectroscopy. *NIR News*, **14**(5), 8–11.

Müller, A. and Steinhart, H. (2007). Recent developments in instrumental analysis for food quality. *Food Chemistry*, **102**(2), 436–444.

Oliveira, F.C., Brandao, C.R., Ramalho, H.F. *et al*. (2007). Adulteration of diesel/biodiesel blends by vegetable oil as determined by Fourier transform (FT) near infrared spectrometry and FT-Raman spectroscopy. *Anaytical Chimica Acta*, **587**, 194–199.

Osborne, B.G. and Fearn, T. (1986). Near infrared data handling and calibration by multiple linear regression. In: G.D. Batten, P.C. Flinn, L.A. Welsh and A.B.

Blakeney (eds), *Leaping Ahead with Near Infrared Spectroscopy*. Melbourne: NIR Spectroscopy Group, Royal Australian Chemical Institute.

Pasquini, C. (2003). Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of Brazilian Chemistry Society*, **14**(2), 198–219.

Pillonel, L., Luginbühl, W., Picque, D. *et al*. (2003). Determining the geographic origin of emmental cheese by mid- and near-infrared spectroscopy. *NIR News*, **15**, 14–16.

Piraux, F. and Dardenne, P. (1999). Feed authentification by near infrared microscopy. In: A.M.C. Davies and R. Giangiacomo (eds), *Proceedings of the 9th International Conference, Verona, Italy*. Wellington: NIR Publications, pp. 535–541.

Planck, M. (1925). *A Survey of Physical Theory* (transl. R. Jones and D.H. Williams). London: Methuen & Co. Ltd. (Dover editions 1960 and 1993).

Pontes, M.J.C., Santos, S.R.B., Araujo, M.C.U. *et al*. (2006). Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry. *Food Research International*, **39**, 182–189.

Rodriguez-Saona, L.E., Fry, F.S. and Calvey, E.M. (2000). Use of Fourier transform near-infrared reflectance spectroscopy for rapid quantification of castor bean meal in a selection of flour-based products. *Journal of Agricultural and Food Chemistry*, **48**, 5169–5177.

Ruoff, K., Luginbühl, W., Bogdanov, S. *et al*. (2006). Authentication of the botanical origin of honey by near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, **54**(18), 6867–6872.

Van der Reyden, D. (1996). *Identifying the Real Thing* Document prepared for School for Scanning, sponsored by the National Park Service and managed by the Northeast. New York, NY: Document Conservation Center.

VerDuin, W.H. (1991). Neural nets for custom formulation. In: R. Maus and J. Keyes (eds), *The Handbook of Expert Systems in Manufacturing*. New York, NY: McGraw-Hill, pp. 515–524.

Williams, P.C. and Norris, K. (eds) (2001). *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd edn. St Paul, MN: American Association of Cereals Chemists.

Wilson, N.D. and Moffat, A.C. (2004). The use of near-infrared microscopy for the identification of counterfeit viagra tablets. *Journal of Pharmacy and Pharmacology*, **September**(Suppl.), S3.

Yang, C., Everit, J.H. and Davis, M.R. (2003). A CCD camera-based hyperspectral imaging system for stationary and airborne applications. *Geocarto International*, **18**(2), 71–80.

Yu, H., Ying, Y.B., Fu, X. and Lu, H. (2006). Classification of Chinese rice wine with different marked ages based on near infrared spectroscopy. *Journal of Food Quality*, **29**, 339–352.

Yu, H., Zhou, Y., Fu, X. *et al*. (2007). Discrimination between Chinese rice wines of different geographical origins by NIRS and AAS. *European Food Research and Technology*, **225**, 313–320.

Zalacain, A., Ordoudi, S.A., Diaz-Plaza, E.M. *et al*. (2005). Near-infrared spectroscopy in saffron quality control: determination of chemical composition and geographical origin. *Journal of Agricultural and Food Chemistry*, **53**, 9337–9341.

Zhao, J., Chen, Q., Huang, X. and Fang, C.H. (2006). Qualitative identification of tea categories by near infrared spectroscopy ADN support vector machine. *Journal of Pharmaceutical and Biomedical Analysis*, **41**, 1198–1204.

Zhou, J., Gu, J. and Zhang, D. (2007). Singular points analysis in fingerprints based on topological structure and orientation field. In: *Advances in Biometrics*. Heidelberg: Springer, pp. 261–270.