

Chapter 2

Localization of Cereal Grain Components by Vibrational Microscopy and Chemometric Analysis

Anna-Stiina Jääskeläinen, Leonardo Galvis Rojas and Carlo G. Bertinetto

2.1 Introduction

The structure and composition of cereal grains are important factors when considering their usage in industrial food processes such as milling, bread-making, malting, and brewing. Grain composition also affects its digestion in human and animal gut, contributing to the nutritional value and possible health benefits. From these aspects follows an increasing importance in finding ways to characterize cereal structure quickly and accurately. This chapter shows how two spectroscopic imaging techniques, Fourier transform infrared (FTIR) microspectroscopy and confocal Raman microspectroscopy (CRM), can provide powerful tools for this purpose. The basic functioning and some practical issues of these experimental techniques will be illustrated in Sect. 2 and 3, respectively. An extensive overview of methods for chemometric analysis focusing on data preprocessing and spectral unmixing and resolution is provided in Sect. 4. Finally, Sect. 5 will present a case study demonstrating the application of CRM combined with chemometric processing to investigate samples of barley (*Hordeum vulgare* L.), one of the most commonly cultivated agricultural cereal crops. For information about the general structure of cereal kernels, the reader is invited to refer to Chap. 1 of this book.

A.-S. Jääskeläinen (✉)

VTT Technical Research Centre of Finland, Tietotie 2, Espoo FI-02044 VTT, Finland
e-mail: anna-stiina.jaaskelainen@vtt.fi

L. G. Rojas

Aalto University, Kirstinmäki 7, 02760 Espoo, Finland
e-mail: leonardo.galvisrojas@aalto.fi

C. G. Bertinetto

Aalto University, Joupinmäenrinne 4 B 33, 02760 Espoo, Finland
e-mail: carlo.bertinetto@aalto.fi

2.2 FTIR Microspectroscopy

Infrared spectroscopy is a widely applied, noninvasive technique to characterize the chemical structure of plant materials. In the vast majority of experiments, infrared spectroscopy has been utilized to characterize the bulk structure of the samples (Gholizadeh et al. 2014). However, a few research groups have applied the potential of location specificity of infrared microspectroscopy to determine the structural features of selected regions in cereal grains (Mills et al. 2005). Hence, the changes in, for example, aleurone and endosperm cell walls (Barron et al. 2005; Jamme et al. 2008; Saulnier et al. 2009) or secondary structure of protein (Bonwell et al. 2008) during grain maturation (Toole et al. 2007) or rumen degradation (Walker et al. 2009) have been investigated.

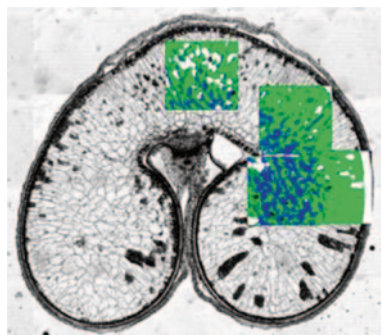
Infrared microspectroscopy is most commonly applied for thin-sectioned grain samples, which are deposited on infrared inactive discs, such as barium fluoride or zinc selenium. These samples are measured in transmission mode using either conventional global or synchrotron infrared light source. The main advantage of the latter is a much higher (1000x) light density but the utilization is limited due to few instruments available.

Alternatively, attenuated total reflectance (ATR) infrared spectroscopy can be applied in grain cell wall microspectroscopic analysis (Barron and Rouau 2008). In this technique, the sample is pressed against a high-refractive index crystal and the infrared spectrum is measured through the crystal in the total reflection angle. In this technique, the infrared irradiation does not actually penetrate the sample but the evanescent wave at the crystal-sample interface generates the infrared spectrum at the sample surface. Hence, the analysis depth of the sample is small, only ca. 0.5–4 μm , depending on the wavelength range and the refractive indices of the ATR crystal and sample materials. Barron and Rouau (2008) estimated that, in their setup, the sampling volume was ca. $<2\text{ mm} \times 2\text{ mm} \times 6\text{ }\mu\text{m}$.

The spatial resolution of infrared spectroscopy is limited by diffraction. In theory, the highest achievable resolution depends on the microscope magnification (numerical aperture) and infrared wavelength. Theoretically, the calculated diffraction limit is ca. 5 μm , indicating that two points that are separated with a distance shorter than this are detected as a single object. In practice, it is commonly stated that the spatial resolution of FTIR microspectroscopy is approximately 5.5–6 μm (Barron et al. 2005; Bonwell et al. 2008). When this resolution is compared to the dimensions of cereal grain structures, it can be revealed that the cell walls of endosperm or aleurone cells are typically 2–3 μm thick (Bacic and Stone 1981) and hence much thinner than the lateral resolution of FTIR microspectroscopy. Therefore, even with the most careful focusing on the cell walls, some signals from the surrounding areas influence the spectra. Consequently, FTIR microspectroscopy is mainly employed to visualize grain features at a cellular scale, as shown in the example in Fig. 2.1.

Interpretation of FTIR spectra The spectral assignment is critical when interpreting the data. The FTIR spectra of several cereal grain components are illustrated in Fig. 2.2. The most important information on the chemical structure of components is in the fingerprint region of the spectra (800–1800 cm^{-1}). The high

Fig. 2.1 Overlaid FTIR images of cross-section of wheat (*Triticum aestivum*) kernel: *blue* = arabinoxylan (highly branched); *green* = arabinoxylan (low branched); *white* = starch; *black* = holes. (Toole et al. 2007). Reproduced with permission from Springer



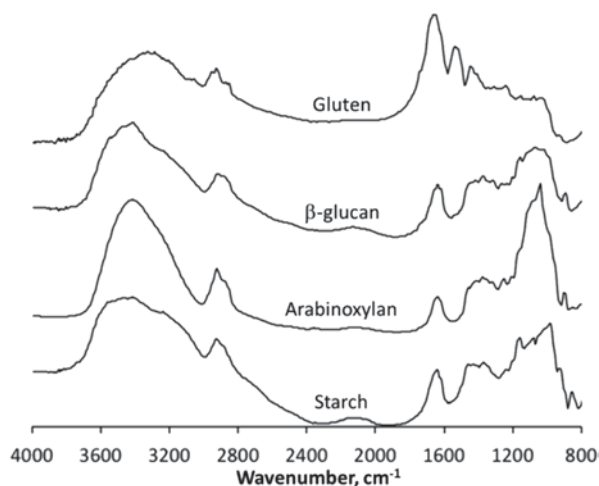
wavenumber range at $3800\text{--}2800\text{ cm}^{-1}$ contains mostly information on the OH, CH, and NH stretching vibrations. This spectral range is sensitive to sample moisture and, due to the broad and overlapping bands, this region is only occasionally used in quantitative analysis of the samples.

The characteristic infrared bands from protein and carbohydrates are clearly seen in the fingerprint region, as illustrated in Fig. 2.2. The wavenumber range of $1710\text{--}1500\text{ cm}^{-1}$ shows the most intense vibrational modes of cereal grain proteins.

Amide I and II have bands at $1710\text{--}1580\text{ cm}^{-1}$ and $1560\text{--}1500\text{ cm}^{-1}$, respectively (Walker et al. 2009). Typically, the highest intensities are observed at 1650 and 1543 cm^{-1} , as reported by Thygesen et al. (2003) and as illustrated in the FTIR spectrum of gluten in Fig. 2.2. When investigating the protein spectra in more detail, the absorption band at 1658 cm^{-1} has been assigned to originate from the alpha helix and beta-sheet of amide (Bonwell et al. 2008).

The strong absorption at spectral range $1300\text{--}800\text{ cm}^{-1}$ is characteristic for cereal carbohydrates (Kačuráková and Wilson 2001). These bands arise from different stretching and bending modes of CO, CC, COH, and COC structures in carbohy-

Fig. 2.2 FTIR spectra of main cereal grain components: protein (gluten) and three different carbohydrates (arabinoxylan, β -glucan, and starch)



hydrate polymers (Cael et al. 1973). It has not been possible to assign these bands unambiguously because most of these bands result from highly coupled vibrational modes. In addition, some of the bands are poorly resolved and severely overlapping, thus making the assignment challenging even when using advanced computational techniques, such as multivariate analysis or deconvolution (see below and Sect. 4.2). Nevertheless, all carbohydrate polymers have their own characteristic infrared absorption bands, which enable us to investigate the carbohydrate composition of the sample as well as identify selected features in the polymer structure at a molecular level (e.g., substitution, chain conformation, helicity, crystallinity, and retrogradation process) (van Soest et al. 1995).

The proportion of different carbohydrates in the sample can be performed using the FTIR spectra. Jamme et al. (2008) stated that arabinoxylan exhibits strong bands at 1045 and 970 cm^{-1} , whereas in β -glucan the band at 1155 and 1078 cm^{-1} is stronger. On the other hand, Barron et al. (2005) proposed that the band at 1075 cm^{-1} is influenced by the degree of branching and hydration of arabinoxylan, illustrating the difficulty of precise assignment of certain spectral bands.

Cereal grain arabinoxylan may have variable amount of ferulic acid substituents which can be observed by the presence of a band at 1595 cm^{-1} . This band is characteristic for the C=C stretching bond in the aromatic ring of ferulic acid. However, this spectral region overlaps with the bands from proteins and possible other aromatics (e.g., lignin) and therefore ferulic acids are not easy to detect or quantify.

The impact of, for example, starch can be removed by spectral subtraction, but this typically causes a distortion of the spectra. To overcome this problem, Barron et al. (2005) developed a method to prepare transverse thin sections of pure grain cell walls by removing the cell contents in the samples with 70% ethanol solution. This treatment leaves the cell walls intact, and then it has been possible to investigate the structure of endosperm cell walls at different positions of the grain without the interference of neighboring starch granules.

As illustrated above, the infrared bands are overlapping and hence univariate analysis, that is, considering a single peak at a time, is usually not possible. Hence, the utilization of multivariate methods is necessary to highlight the small changes in composition or architecture of different sampling positions, for example, in different positions of grain cell walls. The most common approach is to apply principle component analysis (PCA) (Barron et al. 2005; Saulnier et al. 2009). Alternatively, the second derivative of the spectrum can be utilized to identify the minor changes in the spectra, as has been performed by Bonwell et al. (2008), to determine the secondary structure of the endosperm proteins. The spectral analysis is discussed in more detail in Sect. 4.

2.3 Confocal Raman Microscopy

Raman spectroscopy is based on the inelastic scattering of light interacting with molecules in which the frequency shift between the incident and the scattered light is associated with a particular vibration mode of a chemical bond (Tsuboi

and Thomas JR 1997; Tanaka and Young 2006; Masic et al. 2011). When incident radiation encounters molecules, the electric field of the light induces the polarization of the electron cloud and the formation of a short-lived “virtual state” (Tanaka and Young 2006). Different energy transitions take place during this interaction. In elastic energy transition, molecules from the ground state are promoted to the “virtual state” level and return to the ground state with emission of scattered light in a process called Rayleigh scattering. However, the scattering process may also involve inelastic energy transitions to different energy levels from their initial one. These energy transitions are called *Raman scattering*. Raman scattering is inherently a very weak process, since only one in every 10^6 – 10^8 photos experiences a frequency shift (Gierlinger et al. 2013).

One tool that is suited to study cereal kernels is CRM, which can obtain maps of the chemical composition and orientation of materials present in a sample with high resolution (0.6–1 μm) in a nondestructive manner. The main advantage of confocal mapping is an improvement in lateral discrimination with simultaneous reduction of the fluorescence background and an important improvement in axial resolution (depth discrimination) that allows “optical sectioning” in transparent samples (Chalmers and Griffiths 2001). CRM has been used to image the distribution of different components with at least micron resolution in different plant-based samples, including cereal grains (Gierlinger and Schwanninger 2006). For instance, it allowed the study of the relationship between interstitial protein structure in wheat grain and its hardness, through the characterization of the arabinoxylans present in the cell wall during the development of wheat endosperm and the revelation of the structure of starch granules in the endosperm of wild and mutant maize kernels, thus obtaining the compound distribution on the aleurone layer in wheat and barley kernels (Piot et al. 2002; Wellner et al. 2011; Jääskeläinen et al. 2013).

Experimental setups Different approaches can be used to produce Raman images such as scanning mapping (point-by-point scanning), line scanning, and wide-field illumination. Figure 2.3 depicts a point-by-point scanning where the laser is focused onto the sample surface using an objective, and a spectrum is taken by the detector at each spatial position over the region of interest (ROI) in the x and y directions. The integration values of a particular band (band intensity) are visualized as image where the bright pixels correspond to higher values of the integration. The Raman intensity value is linearly related not only to the number of chemical bonds but also to the orientation of molecules present in a confocal volume (Kneipp et al. 1999; Tanaka and Young 2006). Depth scan (x-z) direction is also possible as well as their reconstruction into 3D stack images (x, y, z). The time for the acquisition of every single spectrum at each pixel (integration time) depends usually on the Raman signal of the sample and also on the instrument setup. Integration time values between 0.1 and 1 s are commonly used in this mapping strategy; the duration of an experiment will be proportional to the integration time and also to the number of image pixels.

For the line imaging approach, cylindrical optics is used to distribute the laser beam in one direction over the sample. The laser line is oriented in a parallel fashion to the direction of the entrance slit of the dispersive spectrometer and a 2D charge-coupled device (CCD) detector located at the exit port of the spectrograph collects simultaneously spatial and spectral data of the illuminated sample (Chalmers and

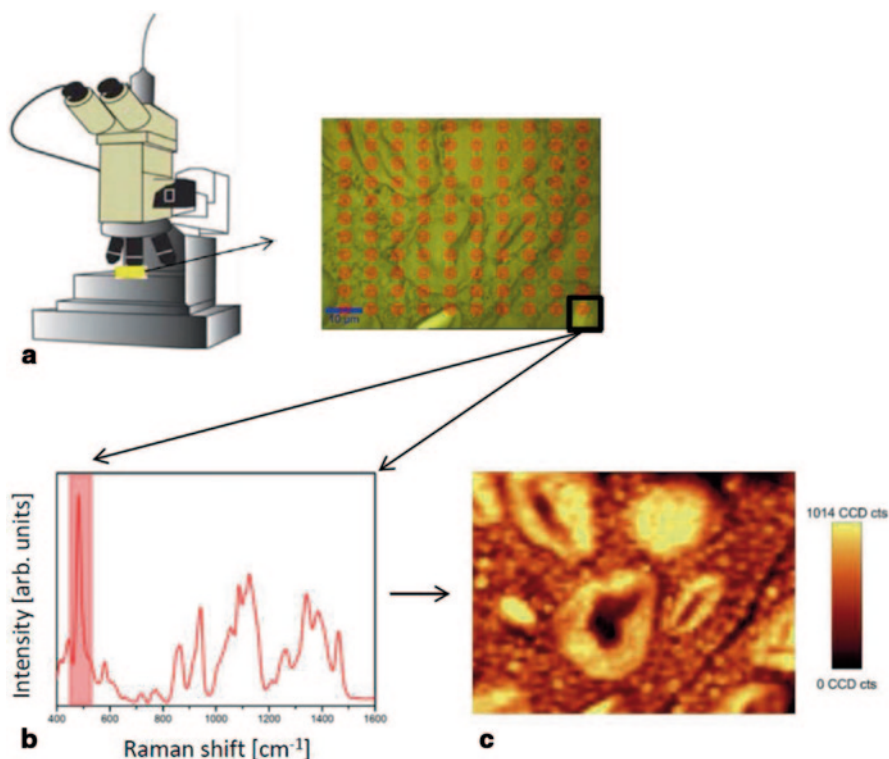
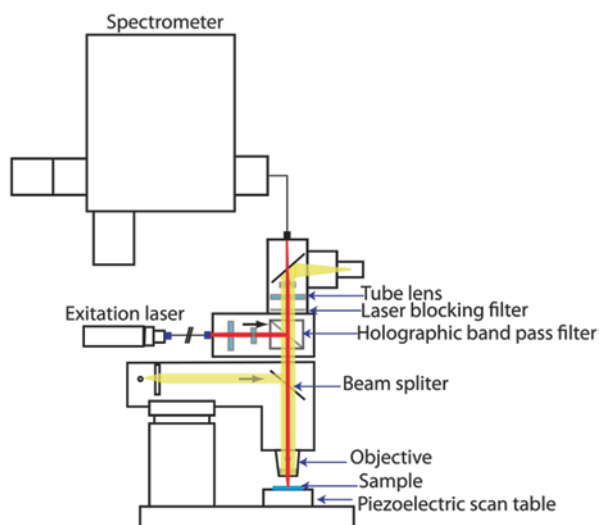


Fig. 2.3 Raman imaging procedure: **a** The selected region of interest (ROI) is scanned and Raman spectra of the ROI are taken at certain spatial steps. **b** Subsequently, the intensity of a single Raman band is obtained by integration (univariate analysis) for each spatial location on the ROI. **c** The result of this integration procedure is shown as chemical distribution image where the brightness of each pixel is associated with the intensity of the selected Raman band

Griffiths 2001; Schlücker et al. 2003). As a result of this setup, only the dimension perpendicular to the laser line is required to be scanned, thus reducing the experimental time. Finally, in wide-field Raman imaging, the entire ROI of the sample is illuminated and analyzed simultaneously by recording an image at discrete wave-number increments through the spectrometer (Chalmers and Griffiths 2001).

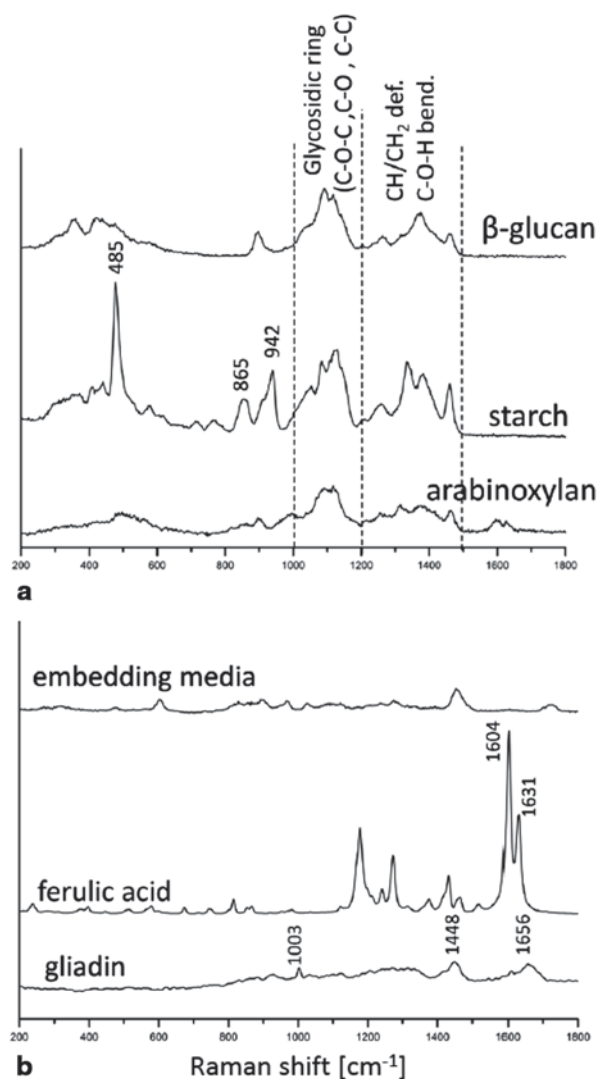
Typical point-by-point Raman microspectroscopy maps are performed in a confocal Raman microscope equipped with linearly polarized lasers with a wavelength comprised in the visible range (e.g., Ar^+ , He-Ne, Kr^+ , and doubled Nd:YAG lasers) coupled to a dispersive spectrometer and a CCD detector. Figure 2.4 shows a standard configuration of the confocal microscope in Raman mode where the excitation laser is focused onto the sample with an objective, and the reflected Raman radiation resulting from the interaction of the laser and the sample is collected with the same objective and focused into a pinhole in front of the detector. The spectra are acquired using an air-cooled CCD camera behind a grating spectrograph. The Raman images are obtained by moving the sample with a piezoelectric scan table in x and y directions. The lateral resolution of the technique is calculated according

Fig. 2.4 Confocal Raman setup and optical pathway in Raman mode. The laser is delivered by an optical fiber and focused onto the sample with an objective lens. The reflected Raman radiation is collected with the same objective and focused into a multimode fiber, which sends the radiation to the spectrometer



to the Rayleigh criterion: $\Delta x = 0.61\lambda/NA$, where Δx is the smallest distance between two point objects that will appear separated in the image plane, λ is the wavelength of the laser, and NA is the numerical aperture of objective that is described by the equation: $NA = n \sin\theta$, where n is the refractive index in which the objective is immersed (e.g., 1.0 for air and up to 1.56 for oils) and θ is the half-angle of the maximum cone of light that can enter or exit the objective. In order to obtain reliable spectral data from Raman mapping experiments, different aspects should be taken into consideration such as sample preparation, damage threshold of the sample under excitation radiation (laser), and presence of fluorescent compounds. The first requirement is to have a very flat sample surface because correct spectral analysis is only possible by mapping even surfaces; otherwise, it is recommended to normalize all spectra with a reference band (see Sect. 4.1). Histological sections of kernels are usually selected for point-by-point Raman mapping. There are several protocols for preparation of kernel sections, but in general they involve several steps like dehydration of kernels in a series of ethanol solutions followed by embedding (in resin, paraffin, polyethylene glycol) and cutting of the kernels into semi-thin sections using a rotary microtome (Holopainen et al. 2005; Gierlinger et al. 2012). The use of polyethylene glycol as embedding media has proven to be useful in plant materials because the dehydration step is unnecessary and the embedding polymer can be easily removed from sections by rinsing with water. It is also important that the section thickness is never lower than the confocal depth of laser beam (depending on the particular setup). For the most common CRM setups, kernel sections with a thickness of 4–8 μm are suitable for Raman mapping. Another important aspect to be considered is the selection of the appropriate excitation laser. The Raman scattering intensity is proportional to ν^4 , where ν is the wavelength of the excitation laser (Gierlinger et al. 2013). Therefore, excitation lasers at 400 nm produce a Raman scattering intensity about 16 times higher than others at 800 nm. However, in confocal microscopes equipped with lasers at 532 nm, the improvement in lateral

Fig. 2.5 **a** Raman spectra of model compounds of main carbohydrates present in sectioned wheat kernels; **b** spectra of other components as embedding media, ferulic acid and protein (gliadin)



resolution and Raman intensity is accompanied by an increase in fluorescence that can hamper the Raman signal. In this case, it is advisable to use setups with higher wavelength lasers, such as 633 or 785 nm. Finally, kernels sections also need to be tested for laser power density to determine the level at which samples do not suffer from degradation or burning. A laser power between 5 and 20 mW in green laser setups is commonly used in plant materials (Gierlinger et al. 2012).

Raman signature of main components in cereal starchy endosperm kernels As the first step in Raman spectral analysis of cereal kernels, a previous knowledge of the Raman fingerprint of pure components present in the cereal kernel material is necessary. For example, Fig. 2.5 shows the Raman spectra of the main components

in embedded sections of cereal starchy endosperm. This information will serve to elaborate chemical images based on the Raman intensity of a particular band (univariate analysis) in case there is no band overlapping from other components in the sample. Polysaccharides spectra in Fig. 2.5a are represented by β -glucan, starch, and arabinoxylan that show bands in the three major regions: 1000–1200 cm^{-1} that are the typical fingerprint of carbohydrates assigned to C-O-C, C-O, and C-C stretching; 1200–1500 cm^{-1} associated to CH/CH₂ deformation and C-O-H bending modes, and below 800 cm^{-1} that is attributed to the skeletal and torsional vibration modes (Barron et al. 2006; Philippe et al. 2006). It is in the last region that starch displays a strong band at 485 cm^{-1} that is assigned to the skeletal vibration of the glucopyranose unit (Kizil et al. 2002; Liu et al. 2004).

The starch spectrum also exhibits a band located at 865 cm^{-1} assigned to stretching C-O-C/ ring-breathing that could indicate sensitivity to the molecular orientation of ordered structures in starch grains (Wellner et al. 2011). The protein component is represented by a spectrum of gliadin (wheat protein) that shows the characteristic signature given by the amide I band that is assigned to the vibration of the trans-peptide group (CONH) in the range 1600–1700 cm^{-1} , the amide III band (1230–1340 cm^{-1}) that arises from the combination of N-H bending and C-N stretching of the peptide group, and the sharp band at 1003 cm^{-1} assigned to the phenylalanine ring vibration (Tsuboi and Thomas JR 1997; Movasaghi et al. 2007). Phenolic compounds are represented by ferulic acid, which exhibits sharp bands at 1604 and 1631 cm^{-1} that are assigned to the aromatic ring vibration and to C=C side-chain stretching, respectively (Sebastian et al. 2009). The same doublets on the region 1590–1630 cm^{-1} are observed in the arabinoxylan spectrum due to the presence of covalently bonded ferulic acid to its structure (Kačuráková et al. 1999; Calheiros et al. 2008). The spectrum of the embedding media needs to be taken into account for further Raman spectral analysis of histological sections.

2.4 Data Analysis of Spectral Images

Further chemical and structural knowledge can be extracted from spectral images by means of computational processing. Several methods are available, depending on the analyzed sample and desired information, which can be classified into two main categories. The first is commonly referred to as preprocessing and comprises algorithms for correcting noises, distortions, or any other unwanted effect in the signal. The second category consists of those methods that perform qualitative and quantitative evaluation of chemical composition and structure. In this chapter, we focus on multivariate methods for unmixing and resolution of spectral components. There are several other uses of spectral data by chemometric techniques, such as classification of substances or prediction of their properties, for which the reader is invited to consult other texts (Tauler et al. 2009). These operations are normally executed either using programming languages (most codes are in MATLAB, R, and C++) or by means of dedicated software often incorporated with the instrument.

2.4.1 Preprocessing

The various operations collectively called preprocessing range from simple format conversions to sophisticated algorithms to filter the signal from certain types of noise, instrumental artifacts, nonrelevant sources of variation and/or nonlinear behavior (Rinnan et al. 2009). These issues may not hamper peak detection or basic univariate analysis but are usually detrimental to precise quantification by multivariate analysis. Choosing the most suitable combination of preprocessing methods is not always easy: it depends on the particular case under study, it often involves additional human knowledge on the data, and there are normally no objective criteria to help the decision (Engel et al. 2013). Some studies have been carried out to address this issue (Bocklitz et al. 2011; Laxalde et al. 2011; Polshin et al. 2011; Bertinetto and Vuorinen 2014a), but they only apply to certain problems and contexts. As a general rule, it is better not to perform a computational transformation on the data when it is not strictly necessary.

For the spectral techniques treated in this chapter, the most common preprocessing operations include restriction of the wavenumber range, removal of cosmic ray peaks, correction of random noise, of baseline distortion, and of intensity variations not related to the phenomenon under study. These operations can be performed in different order and their *effect is illustrated in Fig. 2.6*.

Restriction of the wavenumber range The purpose of this operation is to remove the region that does not contain meaningful information (e.g., singling out the fingerprint region as discussed in Sect. 2 and 3) to simplify the calculations and reduce the influence of noise on the result. The selection is usually based on experience and chemical knowledge of the problem, although there are ways to automate this step using partial least squares, genetic algorithms, or other machine learning tools (Devos and Duponchel 2011; Polshin et al. 2011).

Removal of Cosmic Ray (CR) peaks CR peaks are large, positive, unidirectional, and narrow features that emerge when a CCD detector, such as those typically present in Raman spectrometers, is accidentally stricken by a cosmic ray. They vary randomly in magnitude, bandwidth, and wavenumber, often appearing on top of other spectral features. They are computationally detected by observing a sudden jump in signal intensity that is anomalous with respect to the overall standard deviation and signal-to-noise ratio. The detection can be done by comparing adjacent wavenumbers (Zhang and Henson 2007), adjacent pixels for the same wave number (Cappel et al. 2010; Schulze and Turner 2013), or using wavelets (Ehrentreich and Sümmchen 2001). The CR peaks thus found are then removed by substituting the affected spectral points with an interpolation curve or with an average of neighboring spectra at the same wavenumbers.

Very small CR peaks are likely to be missed by most methods but they can be handled in a subsequent smoothing step (see next paragraph).

Smoothing of random noise Signals are also contaminated by random high-frequency oscillations, which can be reduced by smoothing, that is, some sort of weighted averaging that decreases the curvature of the sharpest signal features. The

Imaging Technologies and Data Processing for Food
Engineers

Sozer, N. (Ed.)

2016, X, 353 p. 143 illus., 82 illus. in color., Hardcover

ISBN: 978-3-319-24733-5