Chapter 4
Spectral Analysis in the NIR Spectroscopy

Yukihiro Ozaki, Shigeaki Morita, and Yusuke Morisawa

Abstract This chapter is concerned with the introduction to spectral analysis in the NIR spectroscopy. It consists of two major parts, conventional spectral analysis and spectra pretreatments. In the former, various conventional spectral analysis methods such as group frequency analysis, derivative spectra, difference spectra, spectral analysis based on perturbation, comparison of a NIR spectrum with the corresponding IR spectrum, and isotope exchange experiments are explained. In the latter part smoothing, derivative methods, multiplicative scatter correction (MSC), standard normal variate (SNV), centering methods, and normalization are described.

Keywords Spectral analysis · Chemometrics · Group frequency · Derivative · Difference spectra · Spectral pretreatment · Baseline correction noise

4.1 Introduction to Spectral Analysis in the NIR Region

As described partly in Chaps. 1 and 2 there are various kinds and various types of NIR spectra [1–7]. First of all, NIR spectra can be divided into electronic spectra and vibrational spectra. However, in this chapter, we treat only vibrational spectra of solids and liquids. Compared with IR spectroscopy diversity of the types of NIR spectra is quite large because NIR spectroscopy is concerned with so many kinds of materials from pure samples such as pure liquids, solutions, and crystals to bulk materials including raw materials, industrial products, and natural products. To look at the diversity of NIR spectra let us compare the spectrum of methanol (0.005 M, in CCl_4; see Fig. 2.3) with that of flour (Fig. 4.1). The former is rather simple

Y. Ozaki (✉)
School of Science and Technology, Kwansei Gakuin University, Sanda, Japan
e-mail: yukiz89016@gmail.com

Toyota Physical and Chemical Research Institute, Nagakute, Japan

S. Morita
Faculty of Engineering, Osaka Electro-Communication University, Neyagawa, Japan

Y. Morisawa
School of Science and Engineering, Kindai University, Higashi-Osaka, Japan

Y. Ozaki et al. (eds.), *Near-Infrared Spectroscopy*,
https://doi.org/10.1007/978-981-15-8648-4_4

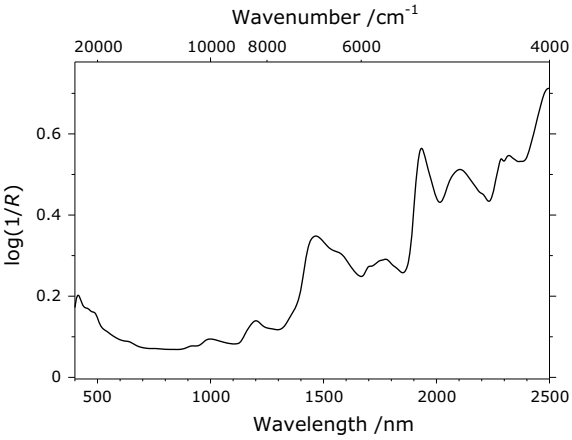

Fig. 4.1 A NIR spectrum of flour. Measured by A. Ikehata

although several bands are overlapped in the 4500–4000 cm^{-1} region. It has very little baseline change. On the other hand, the spectrum of flour consists mainly of the spectra of water, starch, and proteins. Bands are broad, and its baseline increases with the increase in the wavelength. Like this NIR spectra show significant diversity depending on samples and conditions. Therefore, spectral analysis methods are also diverse in NIR spectroscopy [1–3]. In other words, one must select the best spectral analysis method for a target. The spectral analysis methods must change with samples, sample conditions, measurement methods, and the purpose of analysis. The purpose of analysis is also wide spread varies from quantitative analysis, qualitative analysis, and sample identification to studies of molecular structure and chemical reaction. Therefore, when one selects a spectral analysis method, one must consider the purpose of analysis. For many purposes, chemometrics is very useful but for some purposes it is almost meaningful. In this chapter, general introduction to the spectral analysis methods and spectral pretreatments for NIR spectra are outlined. The detailed explanations of representative spectral analysis methods such as chemometrics (Chap. 7), two-dimensional correlation spectroscopy (2D-COS, Chap. 6), and quantum chemical calculations (Chap. 5) will be given in each chapter and session.

As in the cases of IR and Raman spectroscopy, band assignments are always the base for spectral analysis of NIR spectroscopy. However, the assignments are generally not straightforward since a number of bands originating from overtones and combinations overlap each other. In some cases, bands arising from combinations including combinations of overtones appear; for such cases, it is very difficult to make accurate band assignment. In NIR spectroscopy detailed band assignments are often not necessary, but even in such cases, one should know from which functional group a band arises.

Spectral analysis methods in NIR spectroscopy can be divided into conventional spectral analysis method, chemometrics [3], quantum chemical calculation [5, 8], and 2D-COS [1]. The conventional spectral analysis methods are, more or less, common among NIR, IR, Raman, and Terahertz/far-IR(FIR) spectroscopy. One must know

that they yield also a base for chemometrics, quantum chemical calculation, and 2D-COS. Chemometrics has most often been employed to extract rich quantitative and qualitative information from NIR spectra (Chap. 7). A major part of chemometrics is multivariate data analysis such as principal component analysis/regression (PCA/PCR) and partial least squares regression (PLSR), however, self-modeling curve resolution (SMCR), which is used to predict pure component spectra and pure component concentration profiles from a set of NIR spectra, is also becoming more and more significant (Chap. 7). Using quantum chemical calculations such as density function theory (DFT) calculations, one can calculate the intensities and frequencies of overtones and combination bands (Chap. 5). Quantum chemical calculation is still not always popular in NIR spectroscopy but it has already been applied not only to simple compounds but also to rather complicated molecules such as long-chain fatty acids, nucleic acid bases, and rosemaric acid [8]. 2D-COS is not a general method but it is often useful to unravel complicated NIR spectra (Chap. 6). In addition, neural network, AI, and machine learning have been started to be used to analyze NIR spectra. They are very promising methods for the spectral analysis in NIR spectroscopy (Chap. 7). However, one should know that in the early 1990 s neural network has already been tried to be applied to NIR spectra [9].

4.2 Conventional Spectral Analysis Method

Various kinds of conventional spectral analysis methods are used in NIR spectroscopy [1, 2, 6, 7]. They are summarized as follows:

(1) ***Spectral analysis based on group frequencies***

This is a traditional method established in IR and Raman spectroscopy. Spectral analysis based on group frequencies built for the fundamentals is modified for overtones and combinations. Each functional group such as OH and CH groups shows characteristic bands in particular regions. One can find tables for group frequencies in the NIR region in a few NIR textbooks [1, 6].

(2) ***Calculation of derivative spectra***

Derivative methods have long been popular in various spectroscopies [1–3, 6]. They are useful for resolution enhancement as well as baseline correction. Figure 4.2A, B shows a good example demonstrating the usefulness of the second derivative [10]. In Fig. 4.2A, NIR spectra in the 7500–5500 cm^{-1} region are shown for water-methanol mixtures with a methanol concentration of 0–100 wt% at increments of 5 wt% at 25 °C. Figure 4.2B, a gives an enlargement of the 6000–5700 cm^{-1} region of the NIR spectra as shown in Fig. 4.2a. In the 6000–5700 cm^{-1} region, many bands due to the overtones and combination of the CH stretching modes of CH_3 group of methanol are expected to appear. Figure 4.2B, b displays the second derivative of the spectra in Fig. 4.2B, a. Note that a broad feature in the 6000–5750 cm^{-1} region can be divided

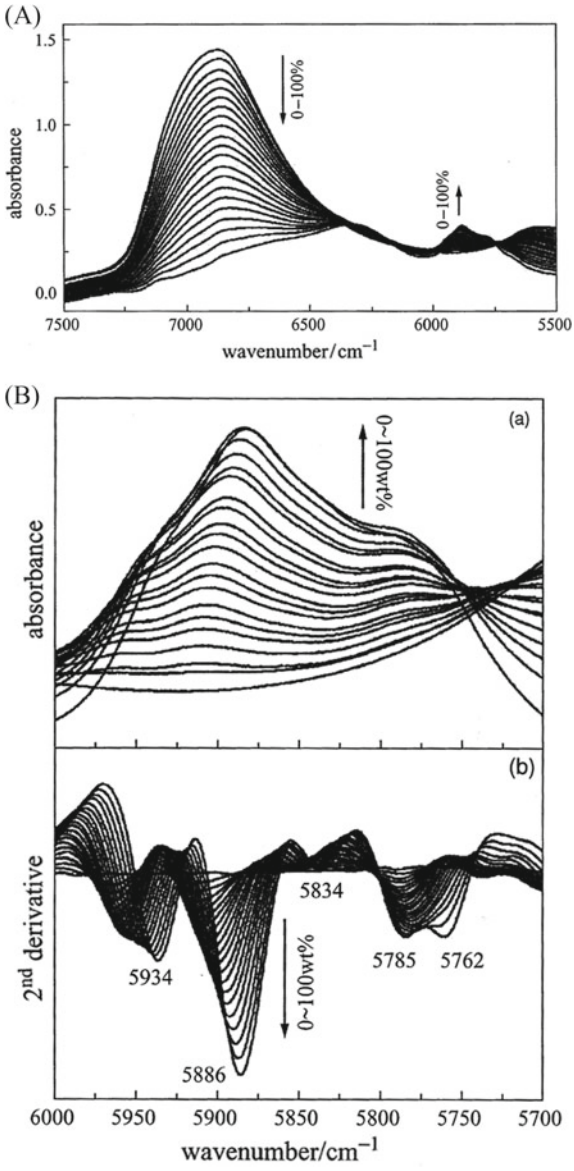

Fig. 4.2 **A** NIR spectra in the 7500–5500 cm^{-1} region of water-methanol mixtures with a methanol concentration of 0–100 wt% at increments of 5 wt% at 25 °C. **B a** An enlargement of the 6000–5700 cm^{-1} region of the NIR spectra shown in **A** and **B**, **b** the second derivative of the spectra in **a**. Reproduced from Ref. [10] with permission

into many bands which show clear concentration-dependent variations. Derivative methods will be explained in more detail in Chap. 4.3.2.

(3) ***Calculation of difference spectra***

Calculation of difference spectra is also useful to unravel overlapping bands and to find out a weak feature hidden by a strong band [1, 2, 5, 7]. The difference spectrum between a spectrum of sample *a* and that of sample *b* can be calculated by

subtracting the spectrum of sample *a* from that of sample *b*. The calculation of difference spectra is effective to analyze perturbation-dependent NIR spectra such as temperature-dependent, concentration-dependent, and pH-dependent spectra. To calculate accurate difference spectra one must obtain spectra with very high wavelength accuracy. Let us show very simple but important example of the calculation of difference spectra. Figure 4.3a displays NIR spectra of water collected over a temperature range of 5–85 °C [11]. From this figure, it is clear that the intensity at 7050 cm^{-1} increases while that at 6844 cm^{-1} decreases but it is not clear whether there is a band shift or not in the 7300–6200 cm^{-1} region. The calculation of the difference spectra clearly answers this question. Figure 4.3b displays the difference spectra of water obtained by subtracting the spectrum at 5 °C as a reference spectrum from other spectra in Fig. 4.3a [11]. It can be seen from Fig. 4.3b that the broad water feature consists of two bands at 7089 and 6718 cm^{-1} and that there is no significant band shift.

Generally speaking, difference spectra method is a reliable method but even so care must be taken. Using spectra of a model system, let us explain a problem of difference spectra. Figure 4.4a shows spectral changes of a system consisting of two components having Gaussian type bands with different peak wavenumbers (7300 and

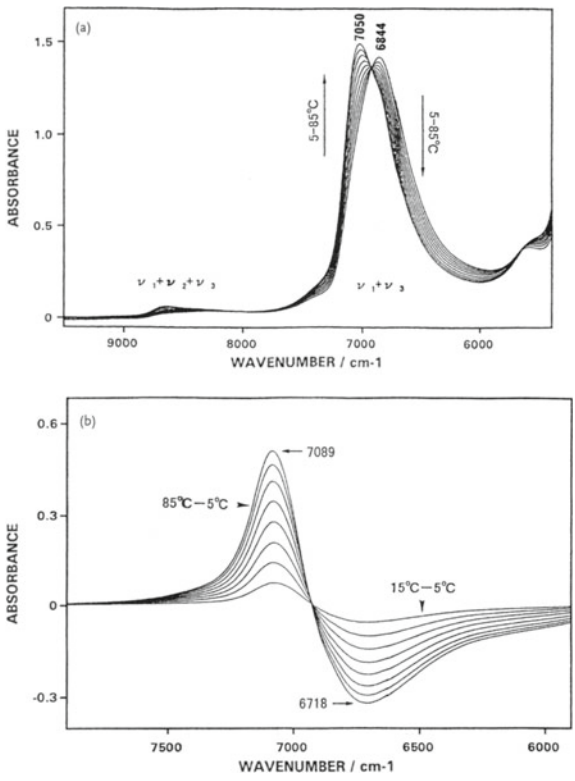

Fig. 4.3 **a** NIR spectra of water measured in a temperature range of 5–85 °C. **b** The difference spectra of water obtained by subtracting the spectrum at 5 °C as a reference spectrum from other spectra in **a**. Reproduced from Ref. [11] with permission

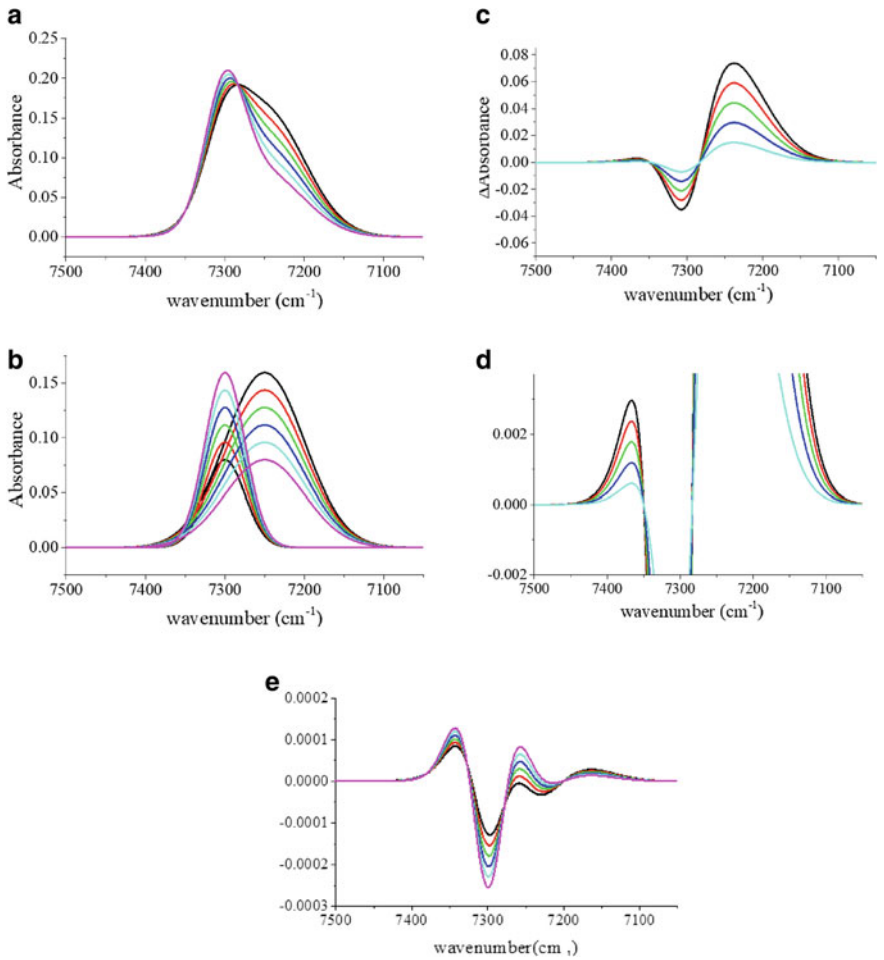

Fig. 4.4 **a** Spectral changes of a system consisting of two components having Gaussian type bands with different peak wavenumbers (7300 and 7250 cm^{-1}), different intensities (10→5 and 10→20), and different band widths (59 and 118 cm^{-1}). **b** Variations of each band. **c** The difference spectra calculated using the first spectrum (pink color) as a reference. **d** An enlargement of **c**. **e** The second derivative of the spectra shown in **a**. Prepared by Y. Morisawa

7250 cm^{-1}), different intensities (10→5 and 10→20), and different band widths(59 and 118 cm^{-1}). Figure 4.4b depicts variations of each band and Fig. 4.4c displays the difference spectra calculated using the first spectrum (pink color) as a reference. Note that peak positions (7309, 7239 cm^{-1}; Fig. 4.4c) are shifted in the difference spectra compared with positions in the original spectra. Figure 4.4d exhibits an enlargement of Fig. 4.4c. A ghost peak appears near 7380 cm^{-1} because the widths of these two bands are significantly different from each other. Therefore, a special care must be taken for feeble peaks.

(4) ***Spectra-structure correlations***

The NIR spectrum of a compound can be compared with those of similar compounds to make band assignments. For instance, NIR measurements of a series of alcohols or that of fatty acids allow one to make assignments of bands due to OH, CH_2, and CH_3 groups. Figure 4.5 compares NIR spectra in the 7500–4000 cm^{-1} region of saturated (stearic acid, arachidic acid, and palmitic acid) and unsaturated (oleic acid, linolenic acid, and linoleic acid) long-chain fatty acids (0.05 M in CCl_4) [12]. By comparison, one can easily discriminate saturated and unsaturated long-chain fatty acids; particularly see the 5800–5670 cm^{-1} region. A common band near 6908 cm^{-1} can be assigned to the first overtone of OH stretching mode. The unsaturated fatty acids yield characteristic bands near 4663 and 4590 cm^{-1}. Grabska et al. have assigned the 4663 cm^{-1} band to the combination of C=C stretching and CH stretching modes by quantum chemical calculation [12].

(5) ***Spectral analysis based on perturbation***

NIR measurements of perturbation-dependent spectral variations, such as temperature-dependent, concentration-dependent, and pH-dependent spectra variations, often give valuable information about the band assignments [1, 2, 7]. As a good example, temperature-dependent NIR spectra changes of octanoic acid in the pure liquid over a temperature range of 15–90 °C are shown in Fig. 4.6 [13]. Octanoic acid in the pure liquid forms a cyclic dimer with hydrogen bonds at room temperature but with the temperature increase the dimer dissociates gradually and a free OH

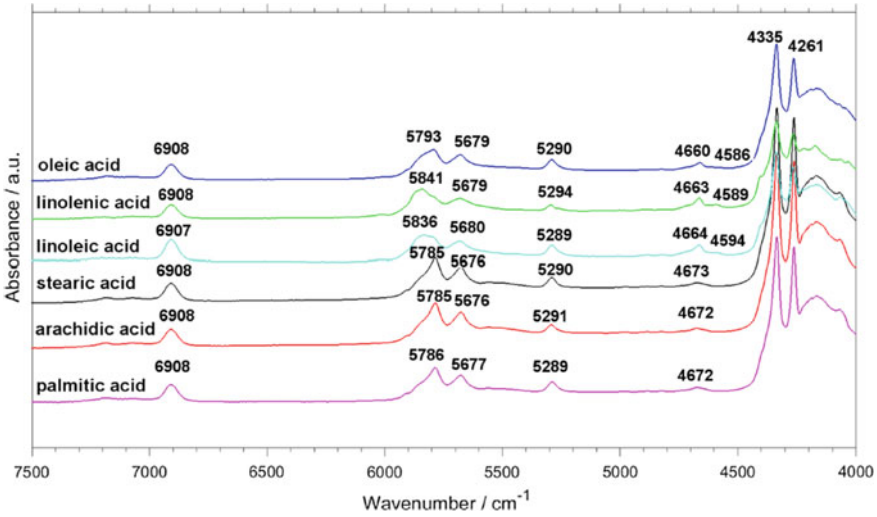

Fig. 4.5 NIR spectra in the 7500–4000 cm^{-1} region of saturated (stearic acid, arachidic acid, and palmitic acid) and unsaturated (oleic acid, linolenic acid, and linoleic acid) long-chain fatty acids (0.05 M in CCl_4). Reproduced from Ref. [12] with permission

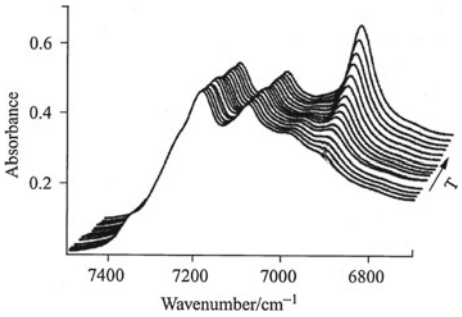

Fig. 4.6 Temperature-dependent NIR spectra changes of liquid octanoic acid collected in a temperature range of 15–90 °C. Reproduced from Ref. [13] with permission

group emerges. It is noted that the intensity of a band at 6920 cm^{-1} increases as a function of temperature, while those of other bands in the 7300–7000 cm^{-1} region are almost temperature independent. Thus, the band at 6920 cm^{-1} is assigned to the first overtone of the OH stretching mode of the monomeric species of octanoic acid, and the rest is due to combinations of CH vibrations [13].

The spectra of Figs. 4.2 and 4.3 are also good examples of perturbation-dependent spectra changes.

(6) ***Comparison of an NIR spectrum with the corresponding IR spectrum***

The significance of comparison of an NIR spectrum with the corresponding IR spectrum was pointed out in Chap. 2.1.6. Here, let us compare NIR spectra of poly(3-hydroxybutyrate) (PHB) with the corresponding IR spectra again but this time in a different region. Figure 4.7a shows time-dependent changes in the NIR spectra in the 5200–5060 cm^{-1} region of a PHB film during the melt-crystallization process at 125 °C [14]. The corresponding IR spectra in the region of 1780–1670 cm^{-1} are depicted in Fig. 4.7b [14]. Of note is that a band at 5127 cm^{-1} in the NIR spectra gradually increases during the crystallization process, while a broad feature centered at 5160 cm^{-1} decreases with time, suggesting that the former band is assigned to the crystalline band and the latter band to the amorphous one. In the corresponding IR spectra bands at 1722 and 1743 cm^{-1} are ascribed to the crystalline and amorphous C=O bands, respectively. Of note is that the NIR spectra and the IR spectra show very clear correspondence. The comparison of the NIR spectra with the IR spectra led Hu et al. [14] to ascribe the band at 5127 cm^{-1} to the second overtone of the C=O stretching mode of the C–H…O=C hydrogen bonding in the crystalline state and the broad feature near 5160 cm^{-1} to the corresponding band due to the amorphous state.

Figure 4.8a, b gives rise to another example of comparison between a NIR spectrum and an IR spectrum. They are the IR spectrum in the 3800–3000 cm^{-1} region and the NIR spectrum in the 7600–6000 cm^{-1} region of diluted methanol in CCl_4. A peak at 3630 cm^{-1} and that at 7090 cm^{-1} are due to a fundamental and a first overtone of a stretching mode of free OH group of methanol while broad features in the regions of 3500–3200 and 6900–6100 cm^{-1} arise from a fundamental and a first overtone of a stretching mode of hydrogen-bonded OH groups of methanol dimer, trimer, and oligomers. Thus, there is the correspondence between the NIR spectrum

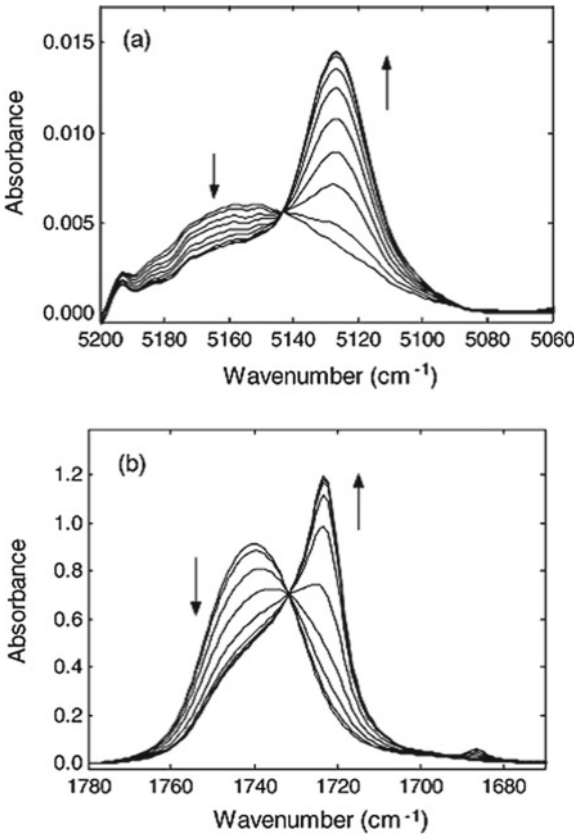

Fig. 4.7 **a** Time-dependent NIR spectra changes in the 5200–5060 cm^{-1} region of a PHB film during the melt-crystallization process at 125 °C. **b** The corresponding IR spectra variations in the 1780–1670 cm^{-1} region. Reproduced from Ref. [14] with permission

and the IR spectrum, but the relative intensity of bands between the free OH bands and the hydrogen-bonded OH bands is largely changed between them. Accordingly, special care must be taken in comparison with the relative intensity between a NIR spectrum and an IR spectrum.

(7) ***Spectral interpretation by polarization measurement***

Polarization measurement, which is popular in IR and Raman spectroscopy, is not often used in NIR spectroscopy, but it is useful for the determination of the molecular-orientation of solid-oriented compounds such as uniaxially stretched polymers. For more details, see Ref. [14].

(8) ***Isotope exchange experiments***

The use of an isotope shift, which is associated with isotopic substitution, is a traditional method for band assignment in IR and Raman spectroscopy. It is also useful for NIR spectra analysis, particularly, a deuteration shift. The isotope shift provides convinced assignment of bands in a number of cases. Force constants can be assumed

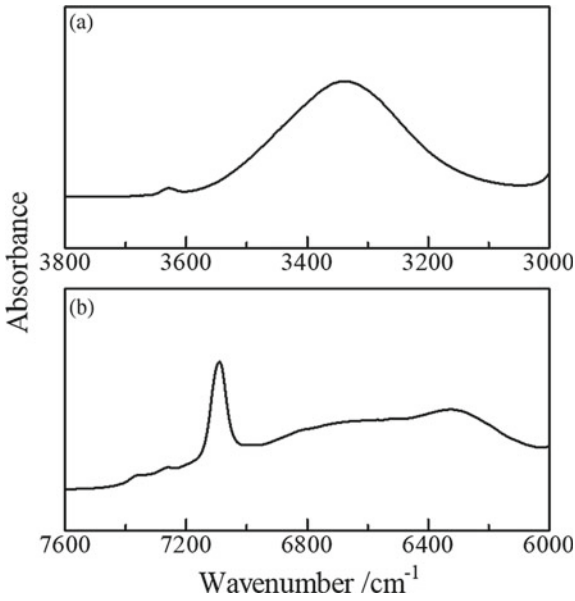

Fig. 4.8 **a** An IR spectrum in the 3800–3000 cm^{-1} region and **b** a NIR spectrum of the 7600–6000 cm^{-1} region of diluted methanol in CCl_4

not to vary due to isotopic substitution, and hence, isotope shifts involve only with mass effects. Taking a diatomic molecule as an example, one can calculate the magnitude of an isotope shift. Frequency of a stretching vibration of the diatomic molecule is given by Eq. (4.1)

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{4.1}$$

if ν' is the frequency for replacing an atom having a mass m_1 with an isotope having a mass m_1', the following relation holds:

$$\frac{\nu}{\nu'} = \sqrt{\frac{\mu'}{\mu}} \tag{4.2}$$

where $\mu' = m_1' m_2/(m_1' + m_2)$. As Eq. (4.2) reveals, the larger the difference between m_1 and m_1', the larger the isotope shift is. Since $\nu/\nu' = 1.36$ if H is replaced with D, a C–H stretching vibration of saturated hydrocarbon, which is located in the vicinity of 2900 cm^{-1}, shifts close to 2100 cm^{-1}. Figure 4.9 shows calculated three vibrational modes (Amide I', II', and III') of deuterated *N*-methylacetamide and the corresponding modes of the nondeuterium-substituted one (Amide I, II, and III modes) are displayed in Fig. 2.11 [16]. Band shifts induced by the deuterium substitution are rather large for the Amide II and III modes since NH bending vibrations contribute to these two modes, but the Amide I mode, being principally a C=O

Fig. 4.9 Amide I', Amide II', and Amide III' modes of deuterated *N*-methyacetamide. Reproduced from Ref. [16] with permission

stretching vibration, yields a very small isotope shift [15]. It is noted with respect to isotope shifts of polyatomic molecules that vibrational modes vary more or less with isotopic substitution, which can be evidently understood from the comparison of Fig. 4.9 and Fig. 2.11. In studies of IR, Raman, and NIR spectra, [15] *N*-substitution, [13] *C*-substitution, and the like are often used in addition to deuterium substitution. Although an isotope shift is small when such a heavy atom is replaced, a variation in a vibrational mode associated with the isotopic substitution is also small.

4.3 Pretreatment Methods in NIR Spectroscopy

NIR spectra often encounter the problems of unwanted spectral variations and baseline shifts [1–3]. They come from the following sources.

1. Various kinds of noises such as those from a detector, an amplifier, and an AD converter.
2. Light scattering from cloudy liquids or solid samples.
3. Changes in temperature, density, and particle size of samples.
4. Poor reproducibility of NIR spectra caused by, for example, path length variations.
5. The use of optical fiber cable may cause baseline shifts.

The above interferences may become obstacles for conventional spectral analysis and 2D-COS. More importantly, they may easily violate the assumptions on which chemometrics equations are based [3]. For instance, the simple linear relationship stated by Beer's law does not hold any more, and the additivity of individual spectral responses is not guaranteed. Accordingly, data pretreatment is often necessary [1–3]. (Chap. 7) Whenever one attempts to improve SN ratio or to correct baseline fluctuations, one should explore the cause of poor SN ratio and that of baseline changes. Otherwise, one cannot find proper pretreatment methods. One interesting example of studies of baseline changes was reported by Geladi et al. [17] They modeled the reflectance spectra of milk by optical effects and chemical light absorption effects. The former induces variations in the direction of the light, and the latter is concerned

with light absorption. In some cases, the former brings about more prominent variations to spectra than the latter. The response of the spectral data to the physical effects is significant baseline variations. On the basis of this study, Geladi et al. [17] proposed multiplicative scattering collection (MSC) as a preprocessing tool to correct the light scattering problems in the NIR spectra.

This section explains four kinds of data pretreatment methods, noise reduction methods, baseline correction methods, resolution enhancement methods, and centering and normalization methods [1–3].

4.3.1 Noise Reduction Methods

In NIR spectroscopy, several kinds of noise are caused by a variety of interfering physical and/or chemical process [1–3]. The most general noise is high-frequency noise associated with the instrument's detector and electronic circuits. There are other forms of noise as well; for example, low-frequency noise and localized noise. Low-frequency noise is induced, for instance, by instrument drift during the scanning measurements. The reduction of the low-frequency noise may be more difficult because it often resembles the real information in the data.

Most standard method to improve SN ratio in spectra is accumulation-average processing that requires to increase the accumulation number and calculate an average. This reduces the effects of high-frequency noise significantly, but technically it is not a "pretreatment" but a normal, integrated part of collecting spectra. If the noise reduction by the accumulation average is still insufficient, one can employ smoothing to remove high-frequency noise. The most commonly used smoothing methods are moving-average method and Savitzky–Golay method [1, 2].

The moving-average method is the simplest type of smoothing [1–3]. In this method, the reading A_i' (A is, for instance, absorbance) at each variable $i = 1, 2, ---k$ is replaced by a weighted average of itself and its nearest neighbors. From $i-n$ to $i+n$:

$$A_i = \sum_{k=-n}^{n} w_k A_{i+k} \tag{4.3}$$

w_k, defining the smoothing, is called the convolution weights.

The Savitzky–Golay method originated from the idea that in the vicinity of a measurement point a spectrum can be fitted by low-degree polynomials [18]. Practically, w_k is determined by fitting the spectrum with low-degree polynomials using least squares regression. Savitzky and Golay calculated w_k for the different orders of polynomials and N ($N = 2n+1$) [18]. One can find these calculated convolution weights in a numeral table. For example, when N is equal to 5, smoothed values can

be obtained by substituting $w_k = -3/35$, 12/35, 17/35, 12/35, $-3/35$ ($k = -2, -1$, 0, 1, 2) into Eq. (4.3). It is noted that if one tries to increase the effect of smoothing by increasing the number of the point of w_k, a band shape would be distorted. This distortion may lead to the decrease in spectral resolution and band intensity.

There are other methods for the noise reduction such as wavelets, eigenvector reconstruction, and artificial neural networks (ANN) [2, 3, 19, 20].

4.3.2 Baseline Correction Methods

As described above in NIR spectra baselines vary for various reasons [1–3]. An observed NIR spectrum, $A(\lambda)$, can be represented as follows;

$$A(\lambda) = \alpha A_0(\lambda) + \beta + e(\lambda) \tag{4.4}$$

Here, A_0 (λ), α, β, and $e(\lambda)$ are a real spectrum, a multiplicative scatter factor (amplification factor), an additive scatter factor (offset deviation), and noise, respectively. There are several methods to eliminate or reduce the effects of α and β. We explain three of them.

Derivative methods

Derivative methods are utilized in NIR spectra for both resolution enhancement and baseline correction [1–3]. (Chap. 7) A derivative spectrum is an expression of derivative values, $d^nA/d\lambda^n$ ($n = 1, 2, \ldots$), of a spectrum A (λ) as a function of λ. The second derivative, $d^2A/d\lambda^2$, is most often encountered. The superimposed peaks in an original spectrum turn out as clearly separated downward peaks in a second-derivative spectrum. Another important property of second-derivative method is the removal of the additive and multiplicative baseline changes in an original spectrum. Figure 4.10a displays NIR spectra of 16 kinds of linear low-density polyethylene (LLDPE) and one kind of high-density polyethylene (HDPE), and Fig. 4.10b shows the second derivative obtained with the Savitzky–Golay method of the spectra as shown in Fig. 4.10a [21]. It can be seen from Fig. 4.10b that the second derivative is powerful in removing additive and multiplicative baseline variations of the spectra, and at the same time, it enables to detect a number of bands clearly. A drawback in the derivative methods is that the SN ratio deteriorates every time a spectrum is differentiated.

Let us explain derivative methods using equations. If the band shape is a Gaussian shape as below;

$$A(x) = \alpha \exp\left\{-\left(\frac{x - x_0}{w}\right)^2\right\}, \tag{4.5}$$

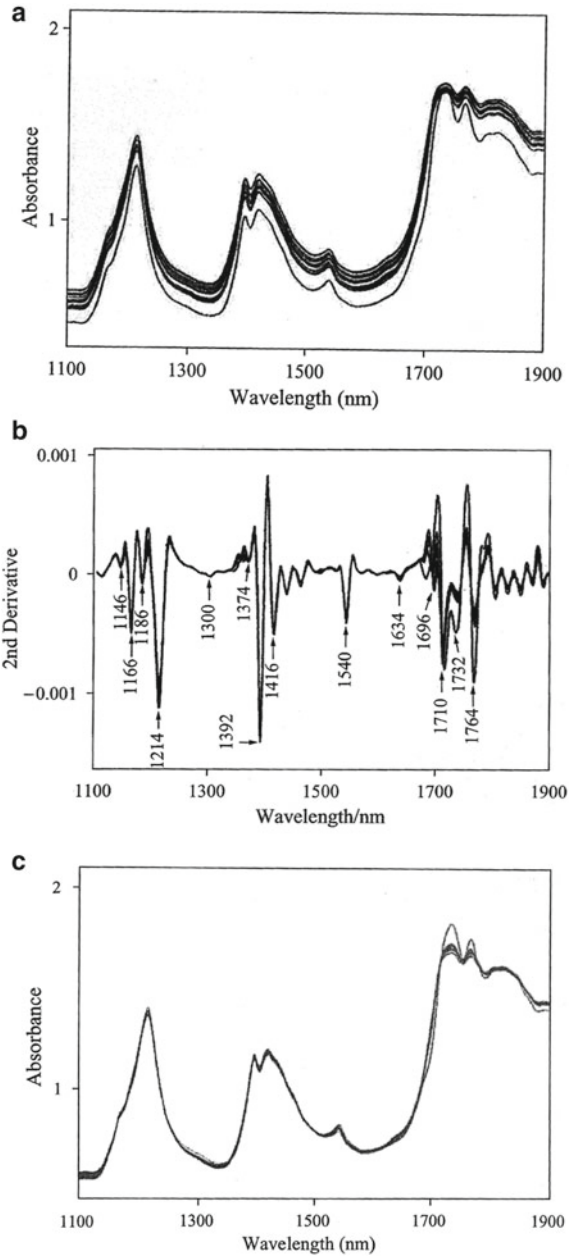

Fig. 4.10 **a** NIR spectra of 16 kinds of linear low-density polyethylene (LLDPE) and one kind of high-density polyethylene (HDPE). **b** The second derivative was calculated with the Savitzky–Golay method of the spectra in **a**. **c** NIR spectra shown in **a** after the MSC treatment. Reproduced from Ref. [21] with permission

Analytical solutions of the first and second derivatives of the band are

$$\frac{\mathrm{d}A}{\mathrm{d}x} = -2\frac{x - x_0}{w^2}A = -2\frac{\alpha(x - x_0)}{w^2}\exp\left\{-\left(\frac{x - x_0}{w}\right)^2\right\} \tag{4.6}$$

$$\frac{\mathrm{d}^2A}{\mathrm{d}x^2} = -2\frac{A}{w^2}\left(1 - 2\frac{(x - x_0)^2}{w^2}\right) = -2\frac{\alpha}{w^2}\left(1 - 2\frac{(x - x_0)^2}{w^2}\right)\exp\left\{-\left(\frac{x - x_0}{w}\right)^2\right\} \tag{4.7}$$

respectively, where α is a peak height of the band, and w is proportional to a full width at the half maximum (FWHM) of the band (ω = FWHM/{Ln (4)}$^{0.5}$). As shown in Fig. 4.11, at the peak position ($x = x_0$), the first derivative coefficient is 0, and the second-derivative coefficient is $-2\alpha/\omega^2$. The positions and intensities of the maximum and minimum of the first derivative spectra are $x_0 \pm \omega\sqrt{2}$ and $\mp\sqrt{2}\alpha\omega e^{0.5}$, respectively. As can be seen in these formulas, the maximum and minimum of the first- and second-derivative coefficients are proportional to an area of the band, if the width of the band is not changed.

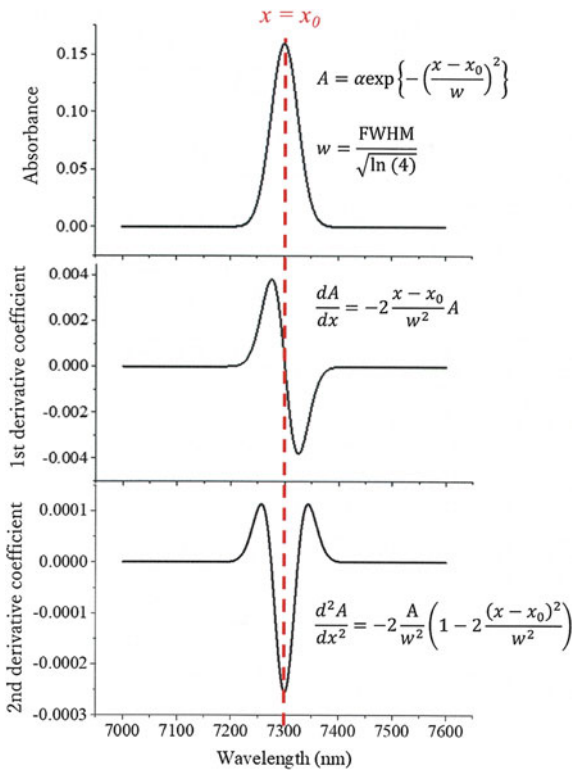

Fig. 4.11 Original spectrum, its first and second derivatives. Prepared by Y. Morisawa

Since, in general, spectrum data take discrete values, and the calculation of derivatives with various orders is performed by algebraic differences between data taken at closely spaced wavelengths. Transformation to first- and second-derivatives is then

$$\mathrm{d}A_i = A_{i+k} - A_{i-k}$$

$$\begin{aligned} \mathrm{d}^2 A_i &= \mathrm{d}(A_{i+k} - A_{i-k}) \\ &= A_{i+2k} - 2A_i + A_{i-2k} \end{aligned} \tag{4.8}$$

Multiplicative scatter correction (MSC)

MSC is an effective method for correcting vertical variations of the baseline (additive baseline variation) and inclination of the baseline (multiplicative baseline variation) [17]. The idea of MSC originates from the fact that light scattering has the wavelength dependence different from that of chemically based light absorbance. Thus, we can use data from a number of wavelengths to distinguish between light absorption and light scattering.

MSC corrects spectra according to a simple linear univariate fit to a standard spectrum. α and β are estimated by least squares regression using the standard spectrum. As the standard spectrum, a spectrum of a particular sample or an average spectrum is used. Figure 4.10c shows the NIR spectra as shown in Fig. 4.10a after the MSC treatment [21]. The spectra demonstrate the potential of MSC in correcting offset and amplification in the NIR spectra. Generally, MSC improves essentially the linearity in NIR spectra. While it is generally an very useful technique, care must be exercised since the use of MSC may generate unwanted artifacts.

Standard normal variate (SNV)

Standard normal variate (SNV) is also a powerful method for correcting vertical baseline drift of a set of spectra [22]. For each spectrum $A(\lambda)$, SNV is calculated as $A_{\mathrm{SNV}}(\lambda) = \left(A(\lambda) - \dot{A}\right)/\sigma$, where $\dot{A}$ and σ are mean and standard deviation of the intensities in the spectrum, respectively. Therefore, mean and standard deviation of the intensities for each spectrum after SNV are standardized as 0 and 1, respectively.

4.3.3 Resolution Enhancement Methods

Resolution enhancement methods are very important to unravel overlapping bands and elucidating the existence of obscured bands [2, 3]. In NIR spectroscopy derivative methods, difference spectra, mean centering, and Fourier self-deconvolution are employed as resolution enhancement methods. Note that PCA loadings plots are often effective for resolution enhancement [21]. 2D correlation spectroscopy can also make resolution enhancement, but since it is not a pretreatment method, it will be

outlined in Chap. 6. Mean centering will be discussed in centering and normalization section.

Derivative methods

We already explained the usefulness of derivative method in resolution enhancement as shown in Fig. 4.10a, b. Here, a problem in the derivative methods which we often encounter is pointed out. Figure 4.4e shows the second derivative of the spectra as shown in Fig. 4.4a. It can be seen from Fig. 4.4e that the 7250 cm^{-1} peak is much weaker than the 7300 cm^{-1} peak. In the second-derivative spectra, a broad band is often underestimated, and thus, care must be taken for the second derivative of a broad band.

4.3.4 Centering and Normalization Methods

Centering and normalization are often effective in chemometrics analysis of NIR data [1–3]. (Chap. 7) Mean centering is simply an adjustment to a data set to reposition the centroid of the data to the origin of the coordinate system [2, 3]. Normalization is an adjustment to a data set that equalizes the magnitude of each spectrum. [2, 3]

Centering methods

Mean centering is a method where from every element of the *j*th spectrum (row) the column mean is subtracted:

$$\boldsymbol{X}_{jcent} = \boldsymbol{X}_j - \left(\frac{1}{n} \sum_{j=1}^{n} \boldsymbol{X}_{ij} \right) \tag{4.9}$$

$\boldsymbol{X}_j$ and $\boldsymbol{X}_{ij}$ are an element of the *j*th spectrum and that of a data matrix $\boldsymbol{X}$, respectively. After this step, all means are zero and variances are spread around zero. Each mean centering spectrum can be regarded as a difference spectrum between the individual spectrum and an averaged spectrum. Mean centering is often powerful in resolution enhancement. Figure 4.12a displays NIR spectra in the region of 6000–5500 cm^{-1} of nylon 12 collected in a temperature range from 30 to 150 °C, and Fig. 4.12b exhibits their mean-centered spectra [23]. The mean-centered spectra show that the intensity of a band at 5770 cm^{-1} arising from the first overtone of CH_2 stretching mode varies markedly with temperature. Mean centering is used also as a pretreatment for constructing 2D correlation spectra (Chap. 6).

Normalization

Two popular normalization procedures have been known in common practice [2, 3]. Most normalization methods employ vectors normalized to constant Euclidean norm. That is,

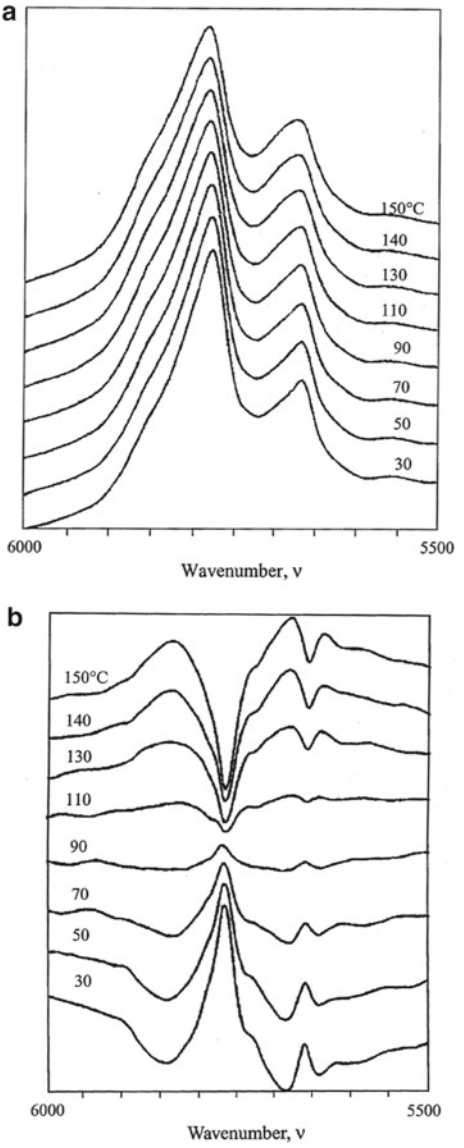

Fig. 4.12 **a** NIR spectra in the 6000–5500 cm^{-1} region of nylon 12 measured over a temperature range from 30 to 150 °C. **b** Mean-centered spectra of the spectra in **a**. Reproduced from Ref. [23] with permission

$$x_{j,\,\mathrm{norm}} = x_j / ||\mathbf{x}|| \tag{4.10}$$

where $||\mathbf{x}||$ is the Euclidean norm of the spectral vector $\mathbf{x}$. This normalization transforms the spectral points on a unit hypersphere, and all data are approximately in the same scaling. This normalization has a good property that the similarity between two spectral vectors may be estimated by the scalar product of these two vectors. However, the normalization induces the geometric configuration of the data points,

either the clustering structure or the spreading directions, substantially different from the original one, which may result in a misleading in the understanding of the data in exploratory data analysis. In addition, the variation in spreading directions has a significant effect on principal component analysis (PCA)-related analysis. Therefore, one must be careful enough in using normalization in situations where exploratory data analysis and PCA-related procedures, such as PCA, partial least squares (PLS), and so on, are concerned.

Another normalization procedure is so-called mean normalization, where all points of the *j*th spectrum are divided by its mean value

$$\boldsymbol{X}_{j\text{norm}} = \boldsymbol{X}_j / \left(\frac{1}{m} \sum_{i=1}^{m} \boldsymbol{X}_{ij} \right) \tag{4.11}$$

where m is a total number of spectral points. After mean normalization, all the spectra have the same area. Essentially, mean normalization is equivalent to normalize the spectral vectors to constant 1-norm, that is, the sum of spectral values (always positive) equals to a constant. This means that the geometry of mean normalization is to transform the spectral points to be contained in a convex set, and the dimensionality of the spectral space is thus decreased by 1. This transformation is very useful in self-modeling curve resolution (SMCR).

References

1. H. W. Siesler, Y. Ozaki, S. Kawata, H. M. Heise, Eds., *Near-Infrared Spectroscopy, Principles, Instruments, Applications* (Wiley-VCH, 2002)
2. Y. Ozaki, W. F. McClure, A. A. Christy, eds., *Near-Infrared Spectroscopy in Food Science and Technology* (Wiley-Interscience, 2007)
3. H. Martens, M. Martens, *Multivariate Analysis of Quality; An Introduction* (John Wiley and Sons, 2001)
4. D. A. Burns, E. W. Ciurczak eds., *Handbook of Near-Infrared Analysis*, 3rd edn. (Practical Spectroscopy) (CRC Press, 2007)
5. Y. Ozaki, C. W. Huck, K. B. Beć, Near-IR spectroscopy and its applications, in *Molecular and Laser Spectroscopy: Advances and Applications,* edited by V. P. Gupta (Elsevier, 2017), p. 11
6. J. Workman, Jr., L. Weyer, *Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy*, 2nd edn. (CRC Press, 2012)
7. M.A. Czarnecki, Y. Morisawa, Y. Futami, Y. Ozaki, Advances in molecular structure and interaction studies using near-infrared spectroscopy. Chem. Rev. **115**, 9707–9744 (2015)
8. K.B. Beć, J. Grabska, C. W. Huck, Y. Ozaki, Quantum mechanical simulation of near-infrared spectra: Applications in physical and analytical chemistry, in *Molecular Spectroscopy; A Quantum Chemistry Approach*, edited by Y. Ozaki, M. J. Wojcik, J. Popp, Wiley-VCH, vol. 1, pp. 353–388 (2019)
9. P.J. Gemperline, J.R. Long, V.G. Gregoriou, Nonlinear multivariate calibration using principal components regression and artificial neural networks. Anal. Chem. **63**, 2313–2323 (1991)
10. Y. Katsumoto, D. Adachi, H. Sato, Y. Ozaki, Useless of a curve fitting method in the analysis of overlapping overtones and combinations of CH stretching modes. J. NIR Spectrosc. **10**, 85–91 (2002)

11. H. Maeda, Y. Ozaki, M. Tanaka, N. Hayashi, T. Kojima, Near infrared spectroscopy and chemometrics studies of temperature-dependent spectral variations of water: relationship between spectral changes and hydrogen bonds. J. NIR Spectrosc. **3**, 191–201 (1995)
12. J. Grabska, K.B. Beć, M. Ishigaki, C.W. Huck, Y. Ozaki, NIR spectra simulations by anharmonic DFT-saturated and unsaturated long-chain fatty acids. J. Phys. Chem. B **122**, 6931–6944 (2018)
13. M.A. Czarnecki, Y. Liu, Y. Ozaki, M. Suzuki, M. Iwahashi, Potential of Fourier transform near-infrared spectroscopy in studies of the dissociation of Fatty acids in the liquid phase. Appl. Spectrosc. **47**, 2162–2168 (1993)
14. Y. Hu, J. Zhang, H. Sato, Y. Futami, I. Noda, Y. Ozaki, C-H…O=C hydrogen bonding and isothermal crystallization kinetics of poly(3-hydroxybutyrate) investigated by near-infrared spectroscopy. Macromol. **39**, 3841–3847 (2006)
15. L. Bokobza: Origin of near-infrared absorption bands, in Ref. [1], p. 11–42
16. Y. Sugawara, A.Y. Hirakawa, M. Tsuboi, In-plane force constants of the peptide group: Least-squares adjustment starting from ab initio values of *N*-methylacetamide. J. Mol. Spectrosc. **108**, 206–214 (1984)
17. P. Geladi, D. MacDougall, H. Martens, Linearization and scatter-correction for near-infrared reflectance spectra of meat. Appl. Spectrosc. **39**, 491–500 (1985)
18. A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. **36**, 1627–1639 (1964)
19. F.T. Chau, T.M. Shih, J.B. Gao, C.K. Chan, Application of the fast wavelet transform method to compress ultraviolet-visible spectra. Appl. Spectrosc. **50**, 339–348 (1996)
20. C.L. Stok, D.L. Veltkamp, B.R. Kowalski, Detecting and identifying spectral anomalies using wavelet processing. Appl. Spectrosc. **52**, 1348–1352 (1998)
21. M. Shimoyama, T. Ninomiya, K. Sano, Y. Ozaki, H. Higashiyama, M. Watari, M. Tomo, Near infrared spectroscopy and chemometrics analysis of linear low-density polyethylene. J. NIR Spectrosc. **6**, 317–324 (1998)
22. R. Barnes, M. Dhanoa, S.J. Lester, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. **43**, 772–777 (1989)
23. Y. Ozaki, Y. Liu, I. Noda, Two-dimensional near-infrared correlation spectroscopy study of premelting behavior of nylon 12. Macromol. **30**, 2391–2399 (1997)