Review of the most common pre-processing techniques for near-infrared spectra

Åsmund Rinnan, Frans van den Berg, Søren Balling Engelsen

Pre-processing of near-infrared (NIR) spectral data has become an integral part of chemometrics modeling. The objective of the pre-processing is to remove physical phenomena in the spectra in order to improve the subsequent multivariate regression, classification model or exploratory analysis. The most widely used pre-processing techniques can be divided into two categories: scatter-correction methods and spectral derivatives. This review describes and compares the theoretical and algorithmic foundations of current pre-processing methods plus the qualitative and quantitative consequences of their application. The aim is to provide NIR users with better end-models through fundamental knowledge on spectral pre-processing.

Keywords: Multiplicative Scatter Correction; Near-infrared spectroscopy; Normalization; Norris-Williams derivation; Pre-processing; Savitzky-Golay derivation; Scatter correction; Spectral derivative; Standard Normal Variate; Review

Åsmund Rinnan*, Frans van den Berg, Søren Balling Engelsen
Quality and Technology, Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

*Corresponding author.
E-mail: aar@life.ku.dk

1. Introduction

There is no substitute for optimal data collection, but, after proper data collection, pre-processing of spectral data is the most important step before chemometric bi-linear modeling [e.g., Principal Component Analysis (PCA) and Partial Least Squares (PLS)]. There is substantial literature on multivariate spectroscopic applications of food, feed and pharmaceutical analysis, in which comparative pre-processing studies are an integral part. Near-infrared reflectance/transmittance (NIR/NIT) spectroscopy is the spectroscopic technique that has led to by far the largest amount of and greatest diversity in pre-processing techniques, primarily because the spectra can be significantly influenced by non-linearities introduced by light scatter. Due to the comparable size of the wavelengths in NIR electromagnetic radiation and particle sizes in biological samples, NIR spectroscopy is a battleground for undesired scatter effects (both baseline shift and non-linearities) that will influence the recorded sample spectra. However, by applying suitable pre-processing, these effects can largely be eliminated.

In application studies, comparisons are almost exclusively of the relative performances in the calibration models developed (quantitative descriptor-response relations). Almost no evaluation of the differences and the similarities between the alternative techniques has been reported, and the implications of corrections (e.g., spectral descriptor data) are seldom discussed. This article aims to discuss the relations between the established pre-processing methods for NIR/NIT, more specifically those techniques that are independent of response variables, so we discuss only methods that do not require a response value. We focus on both the theoretical aspects of the pre-processing technique and the practical effect that the operation has on the NIR/NIT spectra.

For solid samples, undesired systematic variations are primarily caused by light scattering and differences in the effective path length. These undesired variations often constitute the major part of the total variation in the sample set, and can be observed as shifts in baseline (multiplicative effects) and other phenomena called non-linearities. In general, NIR-reflectance measurement of a sample will measure diffusively reflected and specular reflected radiation (mirror-like reflections). Specular reflections are normally minimized by instrument design and sampling geometry, as they do not contain any chemical information. The diffusively reflected light, which is reflected in a broad range of directions, is the primary source of information in the NIR spectra. However, the

0165-9936/$ - see front matter © 2009 Elsevier Ltd. All rights reserved. doi:10.1016/j.trac.2009.07.007

diffusively reflected light will contain information on not only the chemical composition of the sample (absorption) but also the micro-structure (scattering). The primary forms of light scattering (that do not include energy transfer with the sample) are Rayleigh and Lorentz-Mie. Both are processes in which the electromagnetic radiation is scattered (e.g., by small particles, bubbles, surface roughness, droplets, crystalline defects, microorganelles, cells, fibers, and density fluctuations).

Rayleigh scattering, which is strongly wavelength dependent ($\sim 1/\lambda^4$), occurs when the particles are much smaller in diameter than the wavelength of the electromagnetic radiation ($<\lambda/10$).

When the particle sizes are larger than the wavelength, as is generally the case for NIR spectroscopy, Lorentz-Mie scattering is predominant. In contrast to Rayleigh scattering, Lorentz-Mie scattering is anisotropic, dependent on the shapes of the scattering particles and not strongly wavelength dependent.

For biological samples, the scattering properties are excessively complex, so soft, or model-free, spectral pre-processing techniques of NIR spectra, as we discuss in this article, are demanded to remove the scatter from the pure, desirable absorbance spectra.

Obviously, pre-processing cannot correct for specular reflectance (direct scattering), since the spectra do not contain any fine structure. Spectra dominated by specular reflectance should always be removed as outliers prior to multivariate data analysis, since they will remain outliers, even after pre-processing. Fig. 1 shows a set of 13 good sucrose samples with different particle sizes plus one bad sucrose example showing how (extreme) specular reflectance manifests itself compared to normal spectra.

Fig. 1 also illustrates the general layout of most figures in this article. In the upper part of the figure, a bar graph shows PCA-score values on the first principal component (PC) for the sample set after data mean centering [1]. The lower part shows the effect the pre-processing has on the dataset (or, in this case, no pre-processing). The squared correlation coefficient r between the bar values and a selected reference variable is included (in this case, known average particle sizes of 13 sucrose samples). For the sucrose dataset, this correlation should, e.g., be low when assuming particle-originating scatter is a hindrance; as little information as possible on the particle size should remain after the right pre-processing.

An example of the pre-processed sucrose data can be seen in Fig. 2, which also contains a standard-deviation

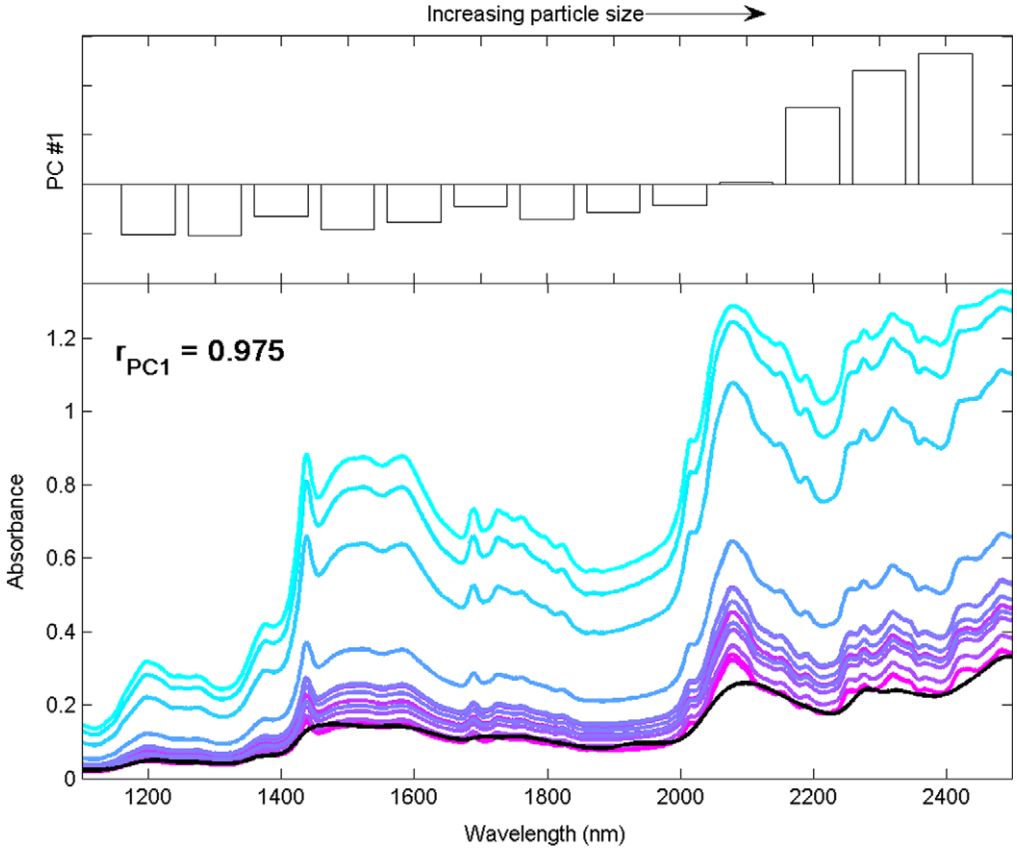

Figure 1. Near-infrared spectra of 13 sucrose samples with different particle sizes (smallest particles at the bottom, largest at the top; particle sizes are in the range 20–540 μm. The black spectrum shows a specular reflectance sucrose sample. Bars are the score values on the first Principal Component of the 13 sucrose samples in a Principal Component Analysis model on the full spectra.

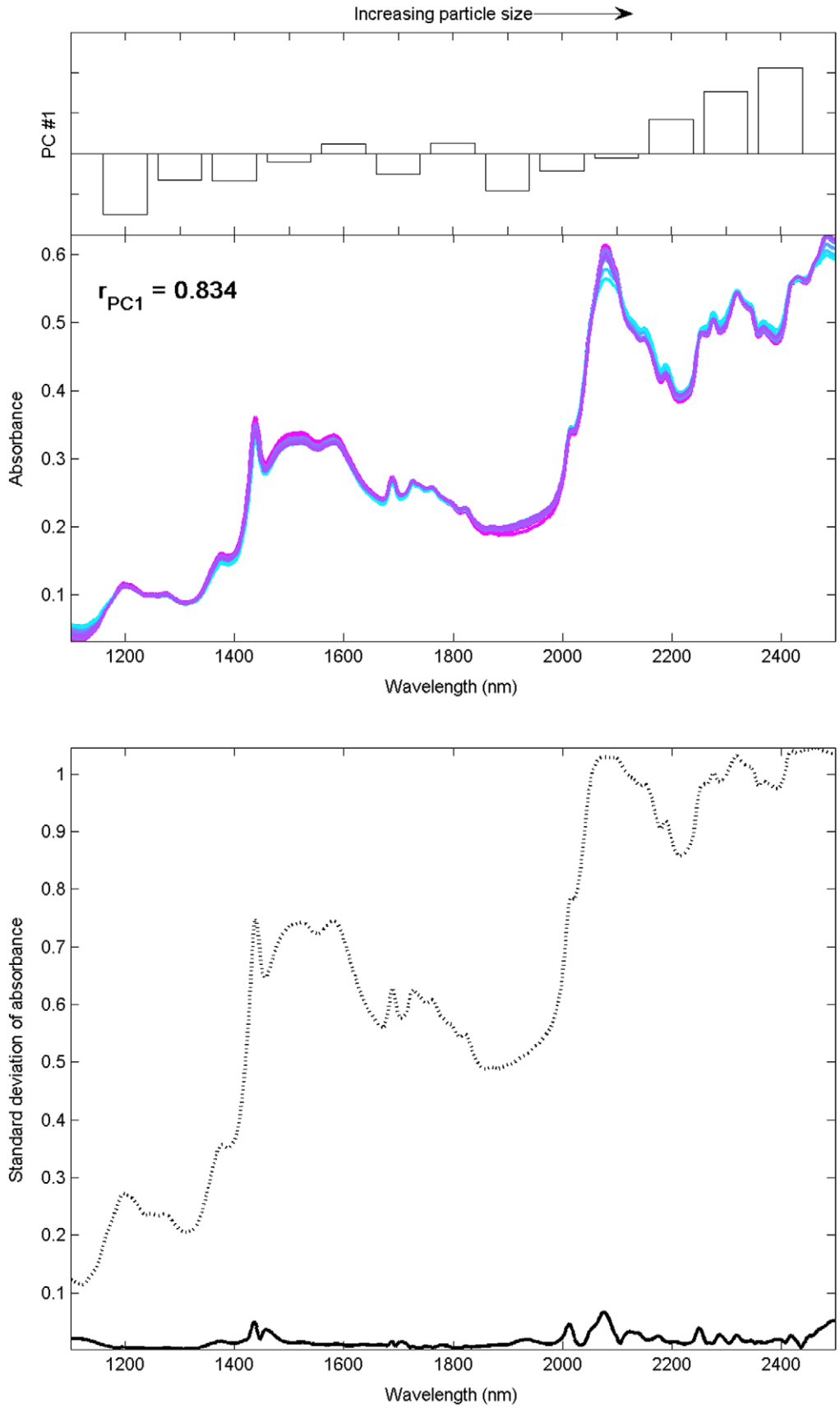

Figure 2. Top: Sucrose data treated by a second-order Multiplicative Scatter Correction; Bottom: the corresponding standard deviations per wavelength, dotted line is the raw/unprocessed data (see Fig. 1), solid is the pre-processed data.

plot, showing the effect that pre-processing has on the variation between samples for different wavelength regions. The selected pre-processing (detailed later on) removes some, but not all, of the undesired scatter or particle-size information in the spectra, as can be observed from, e.g., the first PC bars.

From now on in this article, we demonstrate the effect of different pre-processing techniques on a small pectin dataset containing only seven samples with very different degrees of esterification (%DE; in the range 0–93%) [2]. These samples were measured in NIR-reflectance mode in the spectral range 1100–2500 nm (collecting every 2-nm interval; Fig. 3). We present the corresponding first-factor PCA sample score after mean-centering as a bar graph, together with the centered absorbance value at wavelength 2244 nm. We selected this peak as it should, in theory, describe the %DE perfectly. For this article, we assume that the information in the spectra that is related to the pectin particle size and shape should be removed by the pre-processing technique, and that the bar graph should show a linear behavior correlated to %DE.

To illustrate the impact of pre-processing on quantification, we use data taken from Christensen et al. [3]. They studied a set of 32 marzipan mixtures, based on nine different recipes, with data available on the Internet (www.models.life.ku.dk). All marzipan samples were measured with six different NIR instruments and chemical-reference analyses for moisture and sugar content were made. Before building a quantitative regression model, it is important to clean the predictor data from unsystematic scatter variations, since they can have a significant impact on the predictive model performance and the model complexity or parsimony. In this article, we use PLS to predict this quantitative response information [4].

2. Pre-processing techniques

The most widely used pre-processing techniques in NIR spectroscopy (in both reflectance and transmittance mode) can be divided into two categories: scatter-correction methods and spectral derivatives.

The first group of scatter-corrective pre-processing methods includes Multiplicative Scatter Correction (MSC), Inverse MSC, Extended MSC (EMSC), Extended Inverse MSC, de-trending, Standard Normal Variate (SNV) and normalization.

The spectral derivation group is represented by two techniques in this article: Norris-Williams (NW)

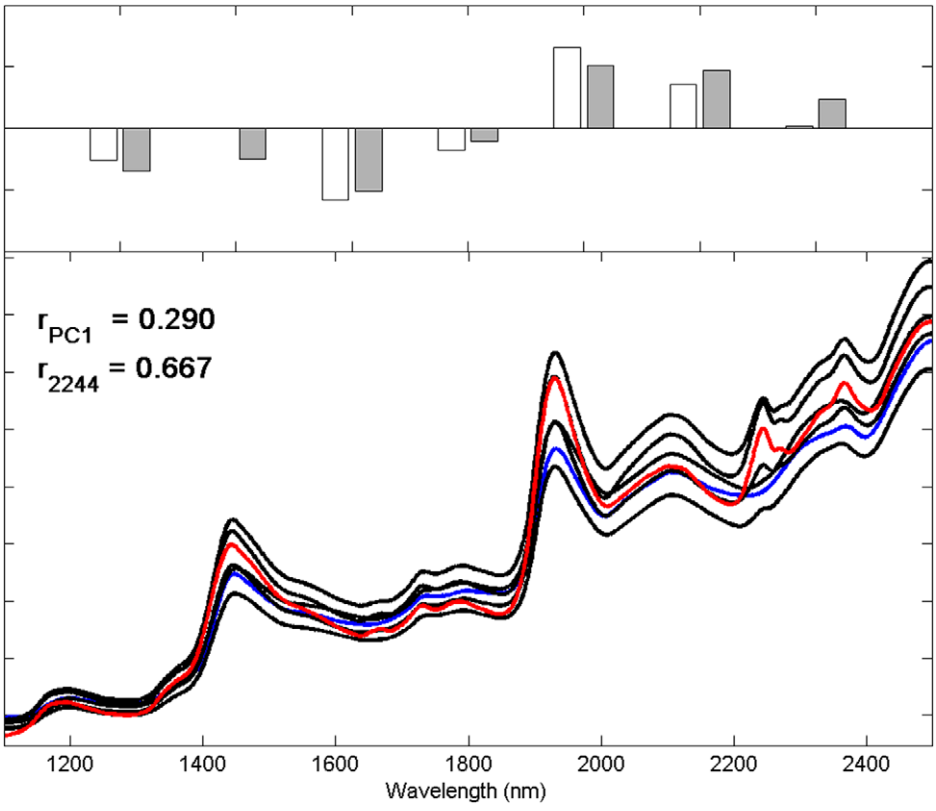

Figure 3. Raw/unprocessed spectra of 7 pectin samples. Blue line is a pectin sample with 0% degree of esterification (DE), red line is a sample with 93%DE. Open bars indicate the Principal Component Analysis (PCA) score values on the first PC for the full spectrum, after mean-centering, closed bars the spectral value at wavelength 2244 nm.

derivatives and Savitzky-Golay (SG) polynomial derivative filters. Both methods use a smoothing of the spectra prior to calculating the derivative in order to decrease the detrimental effect on the signal-to-noise ratio that conventional finite-difference derivatives would have.

The goal of the pre-processing step can be one of three:

- to improve a subsequent exploratory analysis,
- to improve a subsequent bi-linear calibration model (force the data to obey Lambert-Beer's law); or,
- to improve a subsequent classification model.

Lambert-Beer's law (Equation (1)) is empirical for NIR/NIT and suggests a linear relationship between the absorbance of the spectra and the concentration(s) of the constituent(s):

$$A_\lambda = -\log_{10}(T) = \varepsilon_\lambda \cdot l \cdot c \qquad (1)$$

where A_λ is the wavelength-dependent absorbance, T is the light transmittance, e_λ is the wavelength-dependent molar absorptivity, l is the effective path length of the light through the sample matrix and c is the concentration of the constituent(s) of interest. Lambert-Beer's law is strictly valid only for pure transmittance systems with no scatter. In reflectance measurements, Equation (1) is redefined in analogy to transmittance measurements as:

$$A_\lambda = -\log_{10}(R) \cong \varepsilon_\lambda \cdot l \cdot c$$

where R is the reflectance detected.

Selecting suitable pre-processing should always be considered in relation to the successive modeling stage. If, for example, the dataset of interest does not follow Lambert-Beer's law, additional factors or components in PLS regression can often compensate for this non-ideal behavior of the spectral predictor [5]. The disadvantage of including such additional factors is an increase in model complexity and, in turn, most probably a reduction of the model robustness for future predictions. All pre-processing techniques have the goal of reducing the un-modeled variability in the data in order to enhance the feature sought in the spectra, often a linear (simple) relationship with a phenomenon (e.g., a constituent) of interest. By using a suitable pre-processing technique, this can be achieved, but there is always the danger of applying the wrong type or applying a too severe pre-processing that will remove the valuable information. The proper choice of pre-processing is difficult to assess prior to model validation, but, in general, performing several pre-processing steps is not advisable, and, as a minimum requirement, pre-processing should maintain or decrease the effective model complexity.

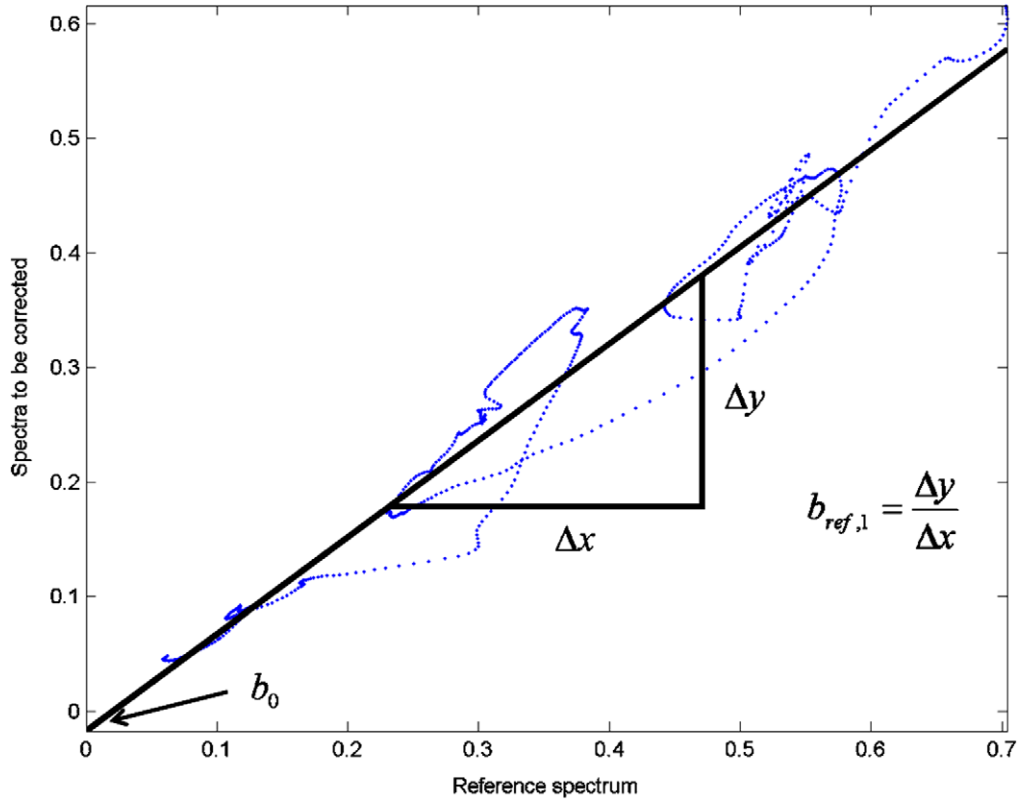

Figure 4. The sample spectrum (blue dots) plotted against a selected reference spectrum. The scalar correction terms are found as the intercept and the slope of the black line, which is found from the least-squares regression fit through all points.

3. Scatter corrections

Under scatter-correction methods, we consider three pre-processing concepts: MSC, SNV and normalization. These techniques are designed to reduce the (physical) variability between samples due to scatter. All three also adjust for baseline shifts between samples.

3.1. MSC

Multiplicative Scatter (or, in general, Signal) Correction (MSC) is probably the most widely used pre-processing technique for NIR (closely followed by SNV and derivation). MSC in its basic form was first introduced by Martens et al. in 1983 [6] and further elaborated on by Geladi et al. in 1985 [7]. The concept behind MSC is that artifacts or imperfections (e.g., undesirable scatter effect) will be removed from the data matrix prior to data modeling. MSC comprises two steps:

1. Estimation of the correction coefficients (additive and multiplicative contributions).

$$\mathbf{x}_{org} = b_0 + b_{ref,1} \cdot \mathbf{x}_{ref} + \mathbf{e} \quad (2)$$

2. Correcting the recorded spectrum.

$$\mathbf{x}_{corr} = \frac{\mathbf{x}_{org} - b_0}{b_{ref,1}} = \mathbf{x}_{ref} + \frac{\mathbf{e}}{b_{ref,1}} \quad (3)$$

where $\mathbf{x}_{org}$ is one original sample spectra measured by the NIR instrument, $\mathbf{x}_{ref}$ is a reference spectrum used for pre-processing of the entire dataset, $\mathbf{e}$ is the un-modeled part of $\mathbf{x}_{org}$, $\mathbf{x}_{corr}$ is the corrected spectra, and b_0 and $b_{ref,1}$ are scalar parameters, which differ for each sample. Fig. 4 illustrates the interpretation of the scalar parameters.

In most applications, the average spectrum of the calibration set is used as the reference spectrum. However, a generic reference spectrum can also be applied. In the original paper by Martens et al. [6], it was suggested to use only those parts of the spectral axis that do not include relevant information (baseline). While this makes good spectroscopic sense, it is difficult to determine such regions in practice, especially in NIR measurements, where the signals from different chemical components are strongly overlapping and correlated, and little or no true baseline is found. This is the reason why, in most cases, the entire spectrum is used to find the scalar correction parameters in MSC. Fig. 5 demonstrates the application of standard MSC to the pectin data. The spectral features of the pectin powder are conserved, while background offsets and slopes are largely removed (compare with Fig. 3). The linear relationship between the spectra and %DE is good, but not perfect.

The basic form of MSC has been expanded into more elaborate augmentations [8–12] commonly known as EMSC. This expansion includes both second-order

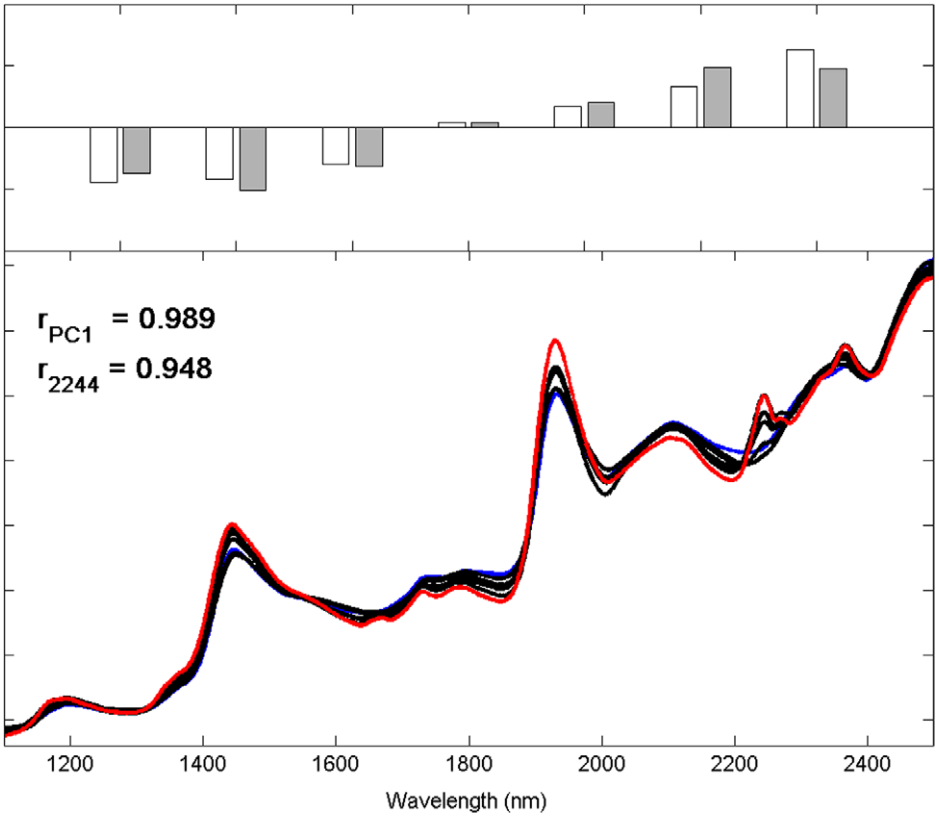

Figure 5. Data pre-processing by Multiplicative Scatter Correction using a first-order correction towards the average spectrum.

polynomial fitting to the reference spectrum, fitting of a baseline on the wavelength axis, and uses of *a priori* knowledge from the spectra of interest or spectral interferents. In this article, all these alternatives are called MSC for simplicity, as they can be summed up in one single equation:

$$\mathbf{x}_{org} = [1 \quad \mathbf{x}_{ref} \quad \mathbf{x}_{ref}^2 \quad \boldsymbol{\lambda} \quad \boldsymbol{\lambda}^2 \quad \mathbf{x}_{known,1} \quad \mathbf{x}_{known,2} \quad \ldots] \cdot \mathbf{b} + \mathbf{e} \tag{4}$$

where $\boldsymbol{\lambda}$ is the correction vector for the wavelength-axis dependency, and $\mathbf{x}_{\mathrm{known,i}}$ is the inclusion of *a priori* knowledge for wanted/unwanted spectral information (e.g., the spectrum of a known interfering species). Equation (4) can readily be expanded to include any other appropriate corrections. $\mathbf{b}$ is a set of scalars (correction coefficients) given by Equation (5).

$$\mathbf{b} = [b_0 \quad b_{ref,1} \quad b_{ref,2} \quad b_{\lambda,1} \quad b_{\lambda,2} \quad b_{known,1} \quad b_{known,2} \quad \ldots] \tag{5}$$

where b_0 is the offset correction, $b_{ref,i}$ is the correction according the ith order of the reference, $b_{\lambda,i}$ is the correction of the ith order wavelength-axis dependency, and $b_{known,i}$ is the correction of the ith known information. By comparison with Equation (2), it can be observed that Equation (4) is just a higher order expansion of the original idea. In this article, the $\mathbf{x}_{\mathrm{known}}$ will not be discussed further, since, in many practical cases, reference spectra for wanted and unwanted constituents are not readily available.

The reference correction is most commonly performed with only a first-order polynomial. Even though there are no mathematical limitations to expand to higher order additions, there are normally no spectroscopic arguments for doing so (except perhaps if significant Rayleigh scattering is present in the short wavelength region).

Fig. 6 shows the result of a second order polynomial correction to the pectin data. The correction terms used for the second-order polynomial reference correction are simply found by fitting a second-order (quadratic) polynomial to the points in Fig. 4. Only marginal improvements are achieved compared to the first-order correction in Fig. 5.

Wavelength-axis dependency is most often included as a second-order polynomial fitting on the wavelength axis to the spectra. When no reference correction is included, this simple wavelength fitting also goes under the name of spectral de-trending [13]; it can be viewed as a baseline correction. It is important to note that including the wavelength dependency in the full correction Equation (4) rather than having it as a separate step leads to a smaller corrective effect. This is due to a matrix-inversion operation performed simultaneously for all the

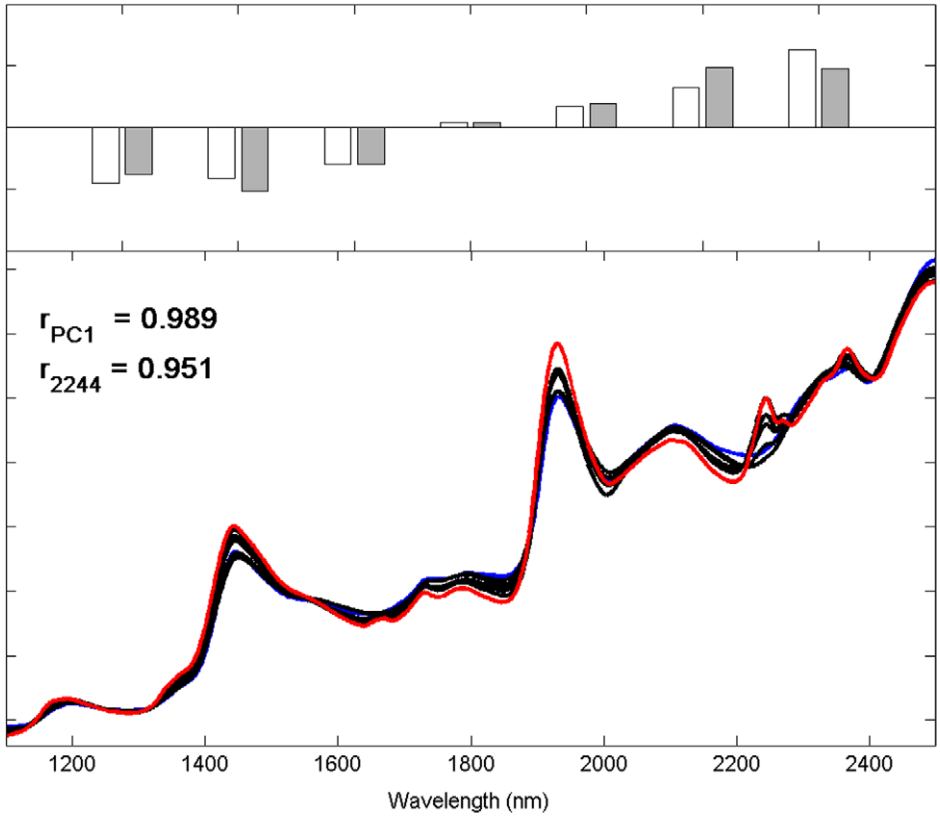

Figure 6. Multiplicative Scatter Correction of the spectra using a second-order reference correction towards the average spectrum.

correction parameters in MSC, where the different corrections will influence each other in the least squares fitting criterion. When a wavelength dependency is determined independently only the wavelength axis (and not the reference spectrum) influences the correction that will lead to a flattening of the processed spectrum. This effect can be seen by comparing Figs. 7 and 8.

As previously mentioned, more sophisticated corrections (e.g., higher order polynomials or other transformations of the wavelength dependency) can easily be incorporated in the MSC. Thennadil and Martin [12] suggested using the logarithmic values of the wavelengths, as this is judged more sound spectroscopically. However, the difference between using a logarithmic transformation of the wavelengths versus using a first-order polynomial correction is minimal, making these two approaches identical for all practical purposes.

As noted by Pedersen et al. [9], it is a quite simple procedure to apply the inverse version of MSC, called Inverse Signal Correction (ISC) [14]. The estimation of the correction parameters, the **b** coefficients, are found in a similar fashion to regular MSC:

$$\mathbf{x}_{ref} = [1 \quad \mathbf{x}_{org} \quad \mathbf{x}_{org}^2 \quad \boldsymbol{\lambda} \quad \boldsymbol{\lambda}^2 \quad \mathbf{x}_{known,1} \quad \mathbf{x}_{known,2} \quad \ldots] \cdot \mathbf{b} + \mathbf{e} \tag{6}$$

Note that $\mathbf{x}_{org}$ and $\mathbf{x}_{ref}$ have swapped places compared to Equation (4). An advantage of (Extended) ISC (EISC) is the simplicity of the correction equation:

$$\mathbf{x}_{corr} = [1 \quad \mathbf{x}_{org} \quad \mathbf{x}_{org}^2 \quad \boldsymbol{\lambda} \quad \boldsymbol{\lambda}^2 \quad \mathbf{x}_{known,1} \quad \mathbf{x}_{known,2} \quad \ldots] \cdot \mathbf{b} \tag{7}$$

In ISC and EISC, both the estimation of the correction coefficients and the correction itself are performed in what can be described as a forward manner, making it convenient to include additional terms and/or reference signals [9]. The previously mentioned matrix-inversion operation required for parameter estimation in MSC can easily become numerically ill-conditioned if it includes higher order polynomial reference corrections. This is an argument in favor of ISC. However, ISC assumes, in the least squares fitting, that the error in the recorded spectrum (to be corrected) is smaller than the error for the reference spectrum. In most practical applications, the reference is the average spectrum computed from n samples in the dataset (e.g., the calibration set). The expected noise level for this reference is of magnitude $\sqrt{n}$ smaller than the individual spectra (neglecting the bias due to scatter differences in the set). This is an argument against ISC, since a small error in the spectra will

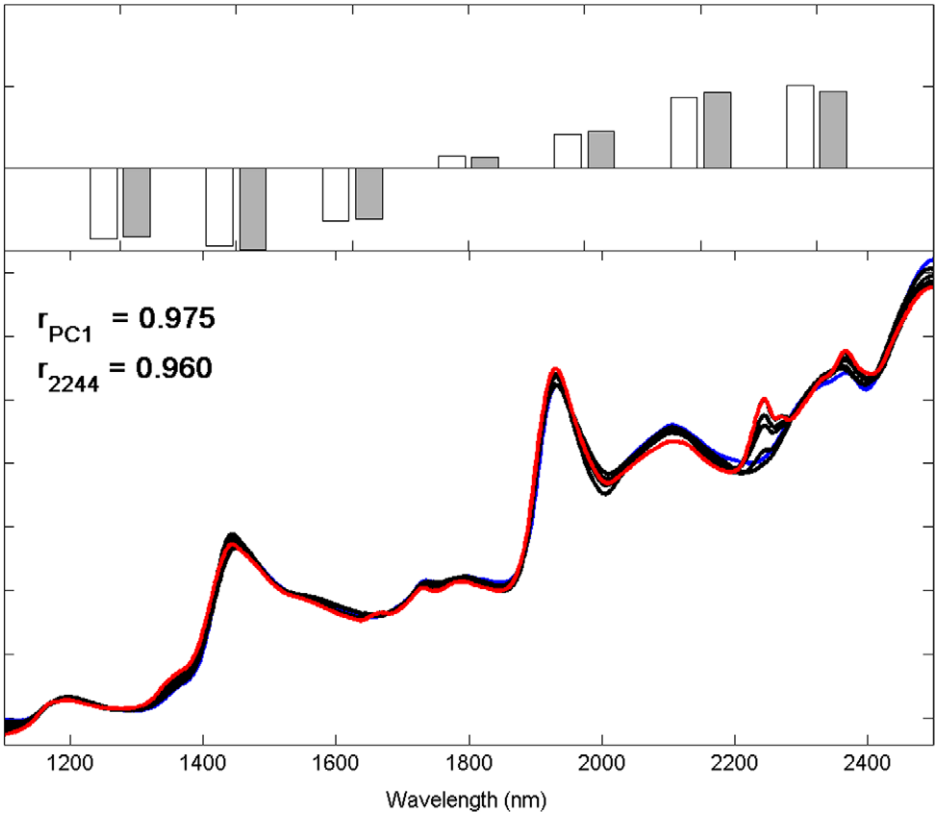

Figure 7. Multiplicative Scatter Correction with first-order polynomial reference correction towards the average spectrum and second-order polynomial wavelength correction.

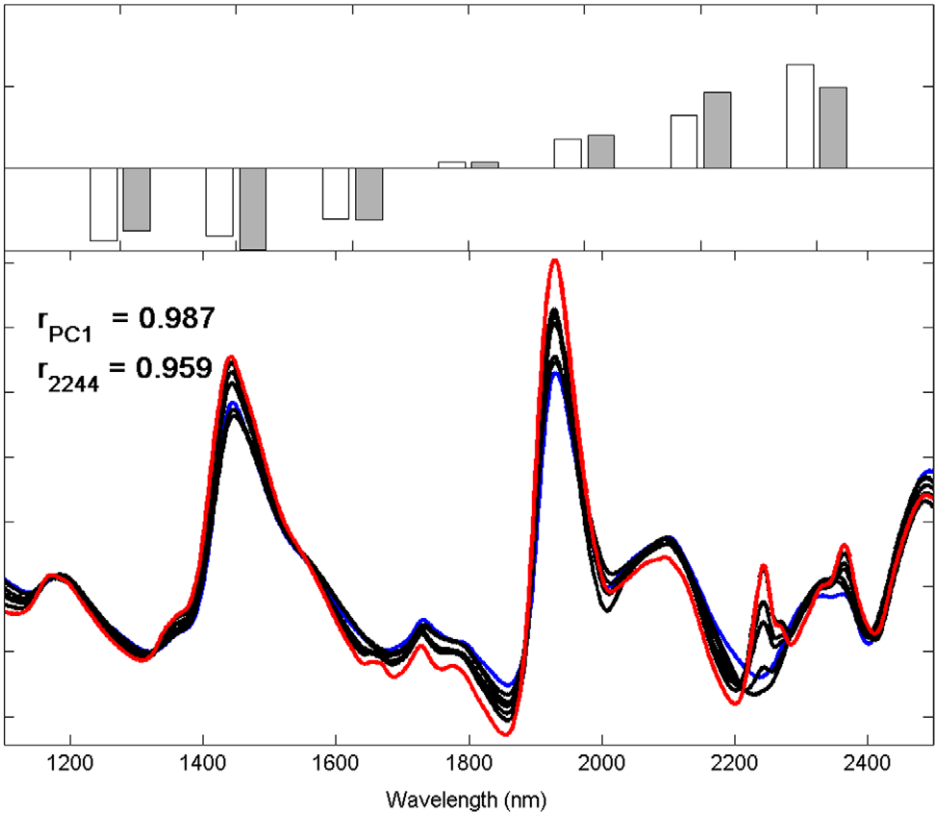

Figure 8. Multiplicative Scatter Correction (MSC) with a first-order polynomial reference correction towards the average spectrum, followed by a separate MSC with a second-order polynomial wavelength correction (detrend).

influence the pre-processing to a greater degree than the original MSC.

The main challenge in MSC is to define an appropriate reference spectrum. As mentioned before, this is most often set to the average of the calibration spectra. Gallagher et al. [15] presented a natural variation to MSC by including a weighting scheme in the pre-processing step. Two alternatives were presented:

- the use of a pre-defined weight vector for the wavelength axis; and,
- an iterative search for the optimal weighting vector.

The iterative solution is found by giving lower weight to variables or wavelengths with high residual differences between the raw data and the corrected solution. The calculation of the weights continues until the difference between corrected spectra for two subsequent iterations is less than the assumed noise level in the data. Unfortunately, this fairly straightforward method does not always work well with NIR data, since the spread in the higher wavelength range typically indicates more scatter, and should be corrected for rather than given less weight. Fig. 9 shows that the weights used in the final correction give strong emphasis to the short-wavelength region, while the long-wavelength region does not contribute to the correction at all.

Another suggestion for finding the reference correction in MSC was suggested by Windig et al. – so-called Loopy MSC [16]. This method finds the average spectrum from the MSC-corrected dataset. Next, MSC is repeated multiple times updating the reference spectrum as the average of the corrected dataset after each iteration step.

Fig. 10 shows the result of Loopy MSC applied to the pectin dataset – in this case the performance of Loopy MSC is very similar to that of the simple MSC. In Loopy MSC, it is possible to follow the increase in model statistics and then stop upon convergence (two iteration steps are typically sufficient). Superimposed in Fig. 10 is the change of reference spectrum from the average of the raw spectra.

3.2. Standard Normal Variate (SNV)

SNV pre-processing is probably the second most applied method for scatter correction of NIR/NIT data [13]. In this article, normalization (also called object-wise standardization) of the spectra will be examined in the same sub-section because of the obvious similarity between the two principles. The basic format for SNV and Normalization correction is the same as that for the traditional MSC:

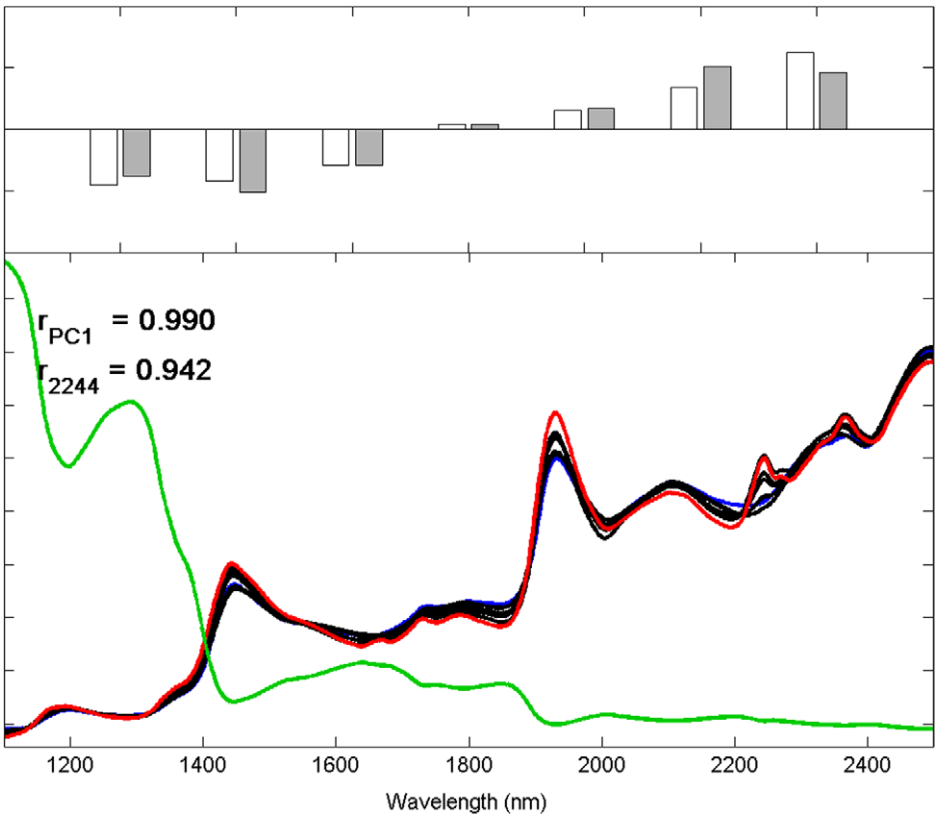

Figure 9. Weighted Multiplicative Scatter Correction, based on iterative weight determination. Green line shows the final weight vector (arbitrary scale, relative contributions).

$$\mathbf{x_{corr}} = \frac{\mathbf{x_{org}} - a_0}{a_1} \quad (8)$$

For SNV, a_0 is the average value of the sample spectrum to be corrected, while, for normalization, a_0 is set equal to zero. For SNV, a_1 is the standard deviation of the sample-spectrum.

Fig. 11 demonstrates the SNV correction for the pectin dataset. For normalization, different vector-norms can be used for scaling factor a_1, the most common being the total sum of the absolute values of the elements in the vector (city-block or Taxicab norm) or the square root of sum of the squared elements (Euclidian norm). Other options that are sometimes used are normalizing to the maximum absorbance variable and normalizing towards a single selected wavelength. Both these last options should be used with caution, since they can have undesirable effects on the subsequent analysis in cases of noisy data.

Fig. 12 shows the effect of Euclidian normalization, by far the most used normalization, of the pectin dataset.

The signal-correction concepts behind SNV and Normalization are the same as for MSC except that common reference signal is not required. Instead, each observation is processed on its own, isolated from the remainder of the set. The lack of need for a common reference might be a practical advantage.

Since SNV and normalization do not involve a least squares fitting in their parameter estimation, they can be sensitive to noisy entries in the spectrum. Instead of using the average and the standard deviation as the correction parameters, one might consider using the more robust equivalents of these statistical moments. Guo et al. [17] suggested using the median or the mean of the inner quartile range and the standard deviation of the inner quartile as estimates for a_0 and a_1, respectively, and named the method ''robust normal variate''. This would be especially appropriate in cases where the spectra are noisy [e.g., in ultra-fast on-line NIR applications, where the robust measures would be less affected by shot noise (e.g., wavelength selective reflections of particles in liquid streams)]. The effects of robust estimates were shown by Guo et al. [17] for both simulated and real data.

As already discussed by Dhanoa et al. [18], there is an obvious similarity between SNV and MSC. This relation can be presented by the following simple approximation:

$$\mathbf{x}_{MSC} \approx \mathbf{x}_{SNV} \cdot \overline{s_{\mathbf{X}}} + \overline{\overline{\mathbf{X}}} \quad (9)$$

where $\overline{s_{\mathbf{X}}}$ is the average standard deviation of all spectra, and $\overline{\overline{\mathbf{X}}}$ is the grand mean over all spectra, both found from the raw/uncorrected spectra (see Fig. 13).

As Equation (9) suggests, MSC and SNV are similar up to a simple rotation and offset correction. For the pectin data

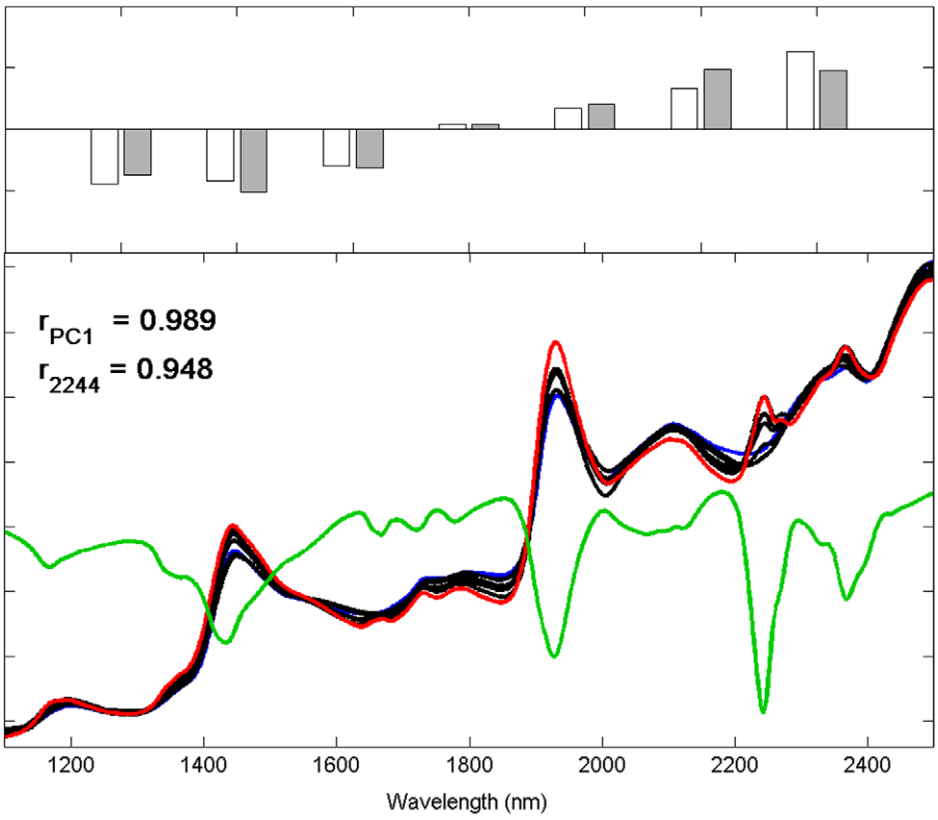

Figure 10. Loopy Multiplicative Scatter Correction with a first-order reference correction. Green line shows the difference between the final and the starting reference spectrum (arbitrary scale, relative contributions).

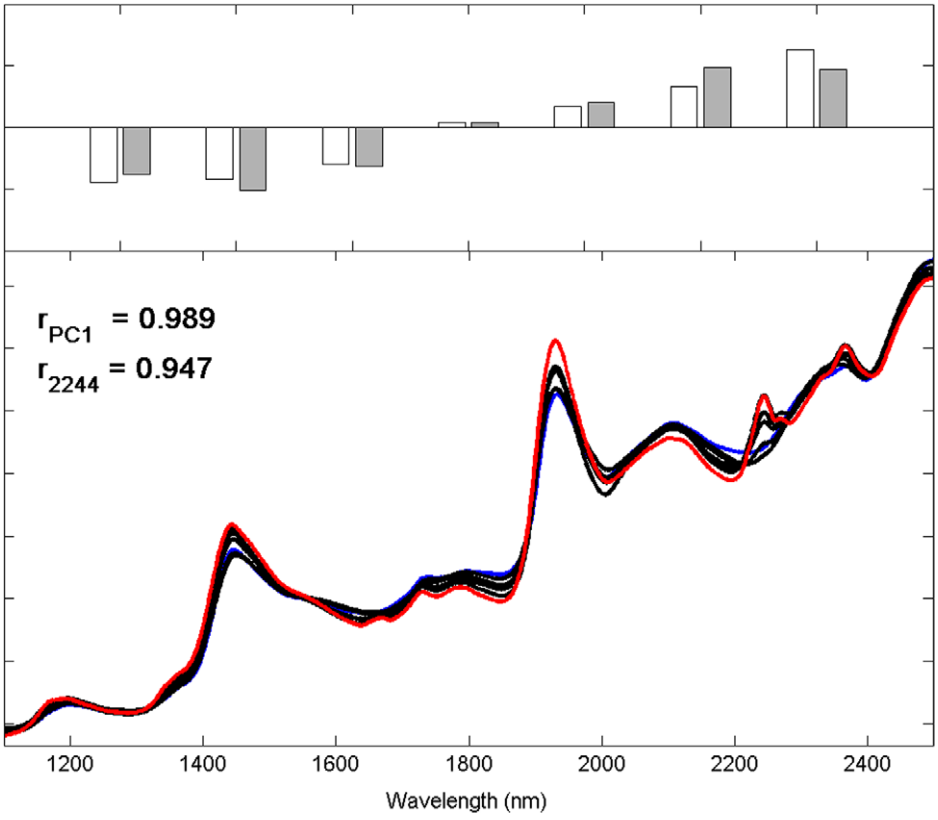

Figure 11. Standard Normal Variate corrected spectra.

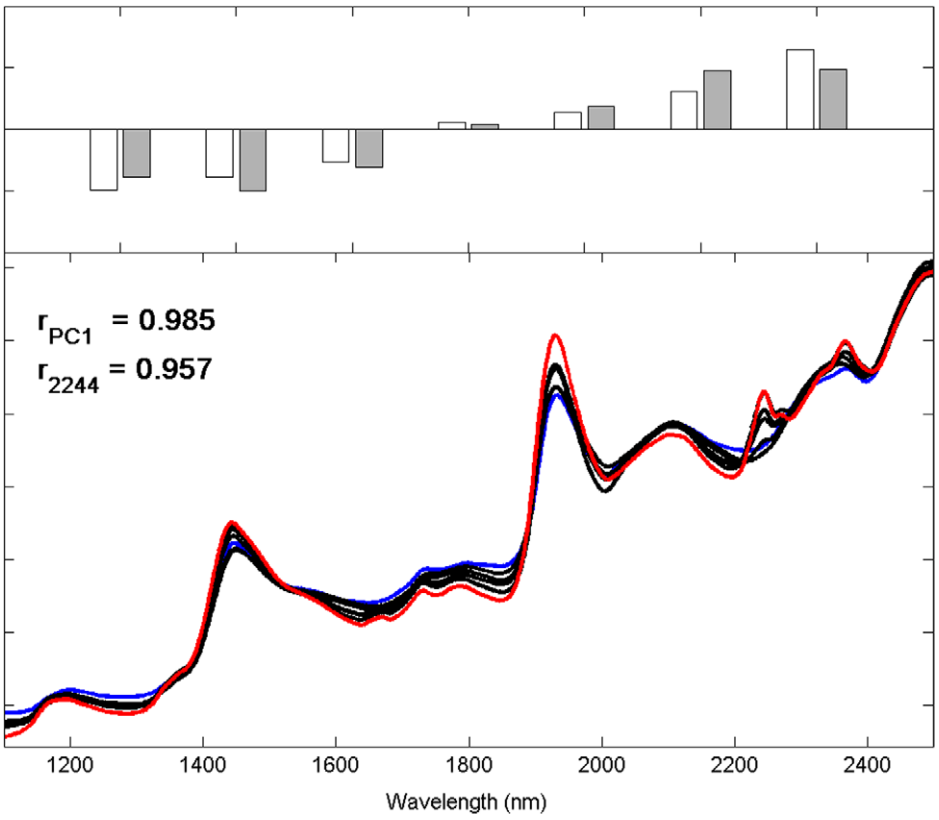

Figure 12. Euclidean norm normalized spectra.

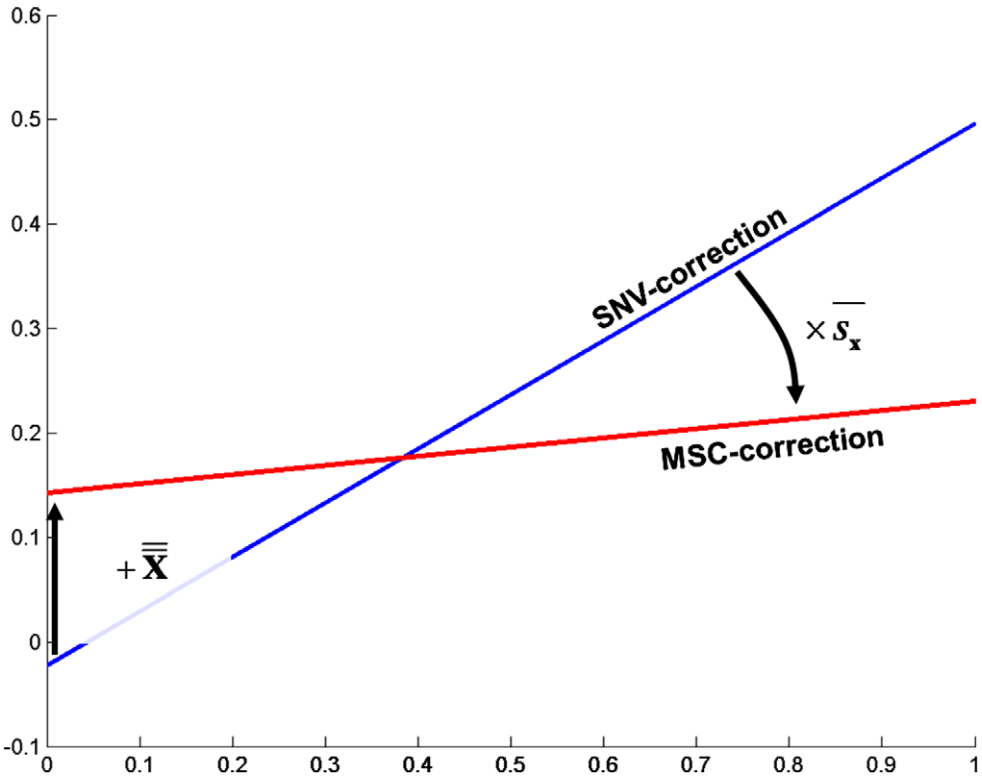

Figure 13. The relationship between Standard Normal Variate and Multiplicative Scatter Correction. The blue and red lines are representative of the trend line estimated in Fig. 4.

used in this article, the correlation of the pre-processed SNV data (Fig. 11) and basic MSC corrected data (Fig. 5) after mean-centering is 0.9995. In other words, MSC and SNV are the same for most practical applications.

4. Spectral derivatives

Derivatives have the capability to remove both additive and multiplicative effects in the spectra and have been used in analytical spectroscopy for decades. This concept is demonstrated in Fig. 14 for a simple Gaussian peak with added baseline and baseline plus multiplicative effect. The first derivative removes only the baseline; the second derivative removes both baseline and linear trend. In this article, we discuss two different methods: SG and NW. Both derivation techniques use smoothing in order not to reduce the signal-to-noise ratio in the corrected spectra too much.

The most basic method for derivation is finite differences: the first derivative is estimated as the difference between two subsequent spectral measurement points; the second order derivative is then estimated by calculating the difference between two successive points of the first-order derivative spectra:

$$x'_i = x_i - x_{i-1} \tag{10}$$

$$x''_i = x'_i - x'_{i-1} = x_{i-1} - 2 \cdot x_i + x_{i+1} \tag{11}$$

where x'_i denotes the first derivative and x''_i the second derivative at point (wavelength) i. This method is extremely simple, but, unfortunately, it is not feasible for most real measurements due to noise inflation; it should almost always be avoided in practice.

4.1. Norris-Williams derivation

The NW derivation is a basic method developed to avoid the noise inflation in finite differences. This technique was suggested by Norris in 1983 [19] and elaborated on by Norris and Williams in 1984 [20] as a way to calculate the derivative of NIR/NIT spectra. The NW derivation includes two steps (see Fig. 15):

1. Smoothing of the spectra, where averaging over a given number of points is performed:

$$x_{smooth,i} = \frac{\sum_{j=-m}^{m} x_{org,i+j}}{2m+1} \tag{12}$$

where m is the number of points in the smoothing window centered around the current measurement point i.

2. For first-order derivation, take the difference between two smoothed values with a given gap size between them (larger then zero); for second-order derivation, take twice the smoothed value at point i and the smoothed value at a gap distance on either side:

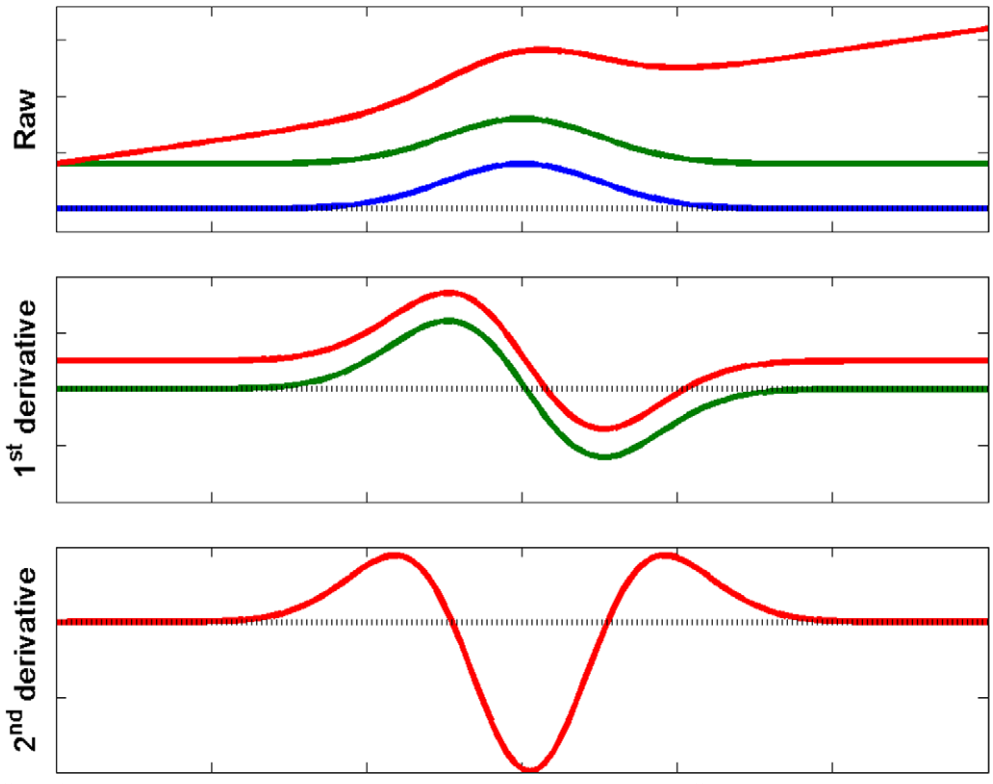

Figure 14. The effect of derivation on additive (green) and additive plus multiplicative (red) effects. The blue spectrum is the spectra without any offsets, and the black dotted line is the zero line.

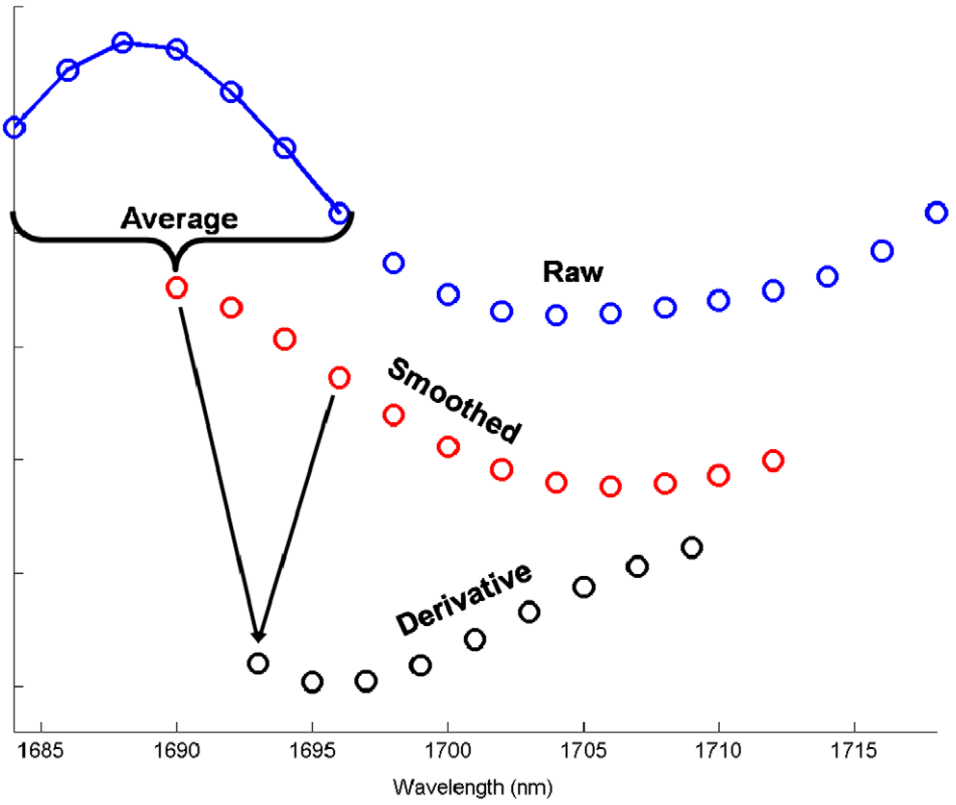

Figure 15. Estimation of the first derivative by Norris-Williams. A 7-point window is used for smoothing, and a gap size of 3 is applied in derivation.

$$x_i' = x_{smooth,i+gap} - x_{smooth,i-gap}$$
$$x_i'' = x_{smooth,i-gap} - 2 \cdot x_{smooth,i} + x_{smooth,i+gap} \quad (13)$$

As can be seen from Equation (13), the actual derivation mimics a finite difference (Equations (10) and (11)). By applying a smoothing prior to the calculation and by introducing a gap size, the problem of decreasing the signal-to-noise ratio is reduced.

In literature, NW derivation is often followed by normalization of the corrected spectra. Norris and Williams [20] proposed to normalize the spectra to equal intensities at a single selected wavelength, but more sophisticated normalization methods can be used. Using a gap is difficult to defend in NIR spectroscopy. The concept of a gap is often used if there is a (fixed) frequency component in the data, where the size of the gap would correspond to the distance between two peak values in the signal. However, in spectroscopy, there is normally no such background-frequency contribution. The NW derivative works due to the high degree of co-variation and smoothing of the NIR spectra and not necessarily due to spectroscopic reasoning (see Fig. 16).

An interesting note on NW derivation is that there exist several settings (combinations of gap and smoothing window) that give identical estimates for the derivative. For the first derivative, a three-point smoothing with a gap of five equals a four-point smoothing with a gap size of three; similarly, a three-point smoothing with a gap of seven is the same as a six-point smoothing with a gap size of three. This can be generalized to m point smoothing with a gap size of k equals a k-1 point smoothing and a gap size of m.

4.2. Savitzky-Golay derivation

Savtizky and Golay (SG) [21] popularized a method for numerical derivation of a vector that includes a smoothing step. In order to find the derivative at centre point i, a polynomial is fitted in a symmetric window on the raw data (see Fig. 17). When the parameters for this polynomial are calculated, the derivative of any order of this function can easily be found analytically, and this value is subsequently used as the derivative estimate for this centre point (see Fig. 18). This operation is applied to all points in the spectra sequentially. The number of points used to calculate the polynomial (window size) and the degree of the fitted polynomial are both decisions that need to be made. The highest derivative that can be determined depends on the degree of the polynomial used during the fitting (i.e. a third-order polynomial can be used to estimate up to the third-order derivative).

We note that there is an intrinsic redundancy in the hierarchy of SG derivation. For each derivation, two

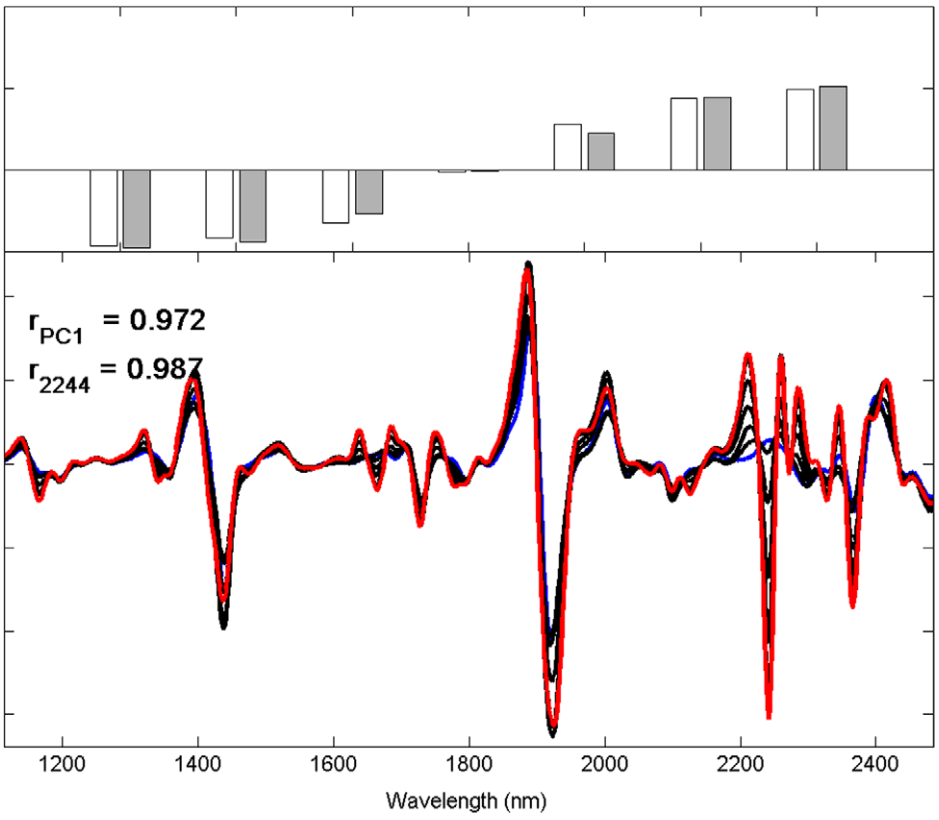

Figure 16. Norris-Williams second derivative using a 9-point smoothing and a gap size of 3.

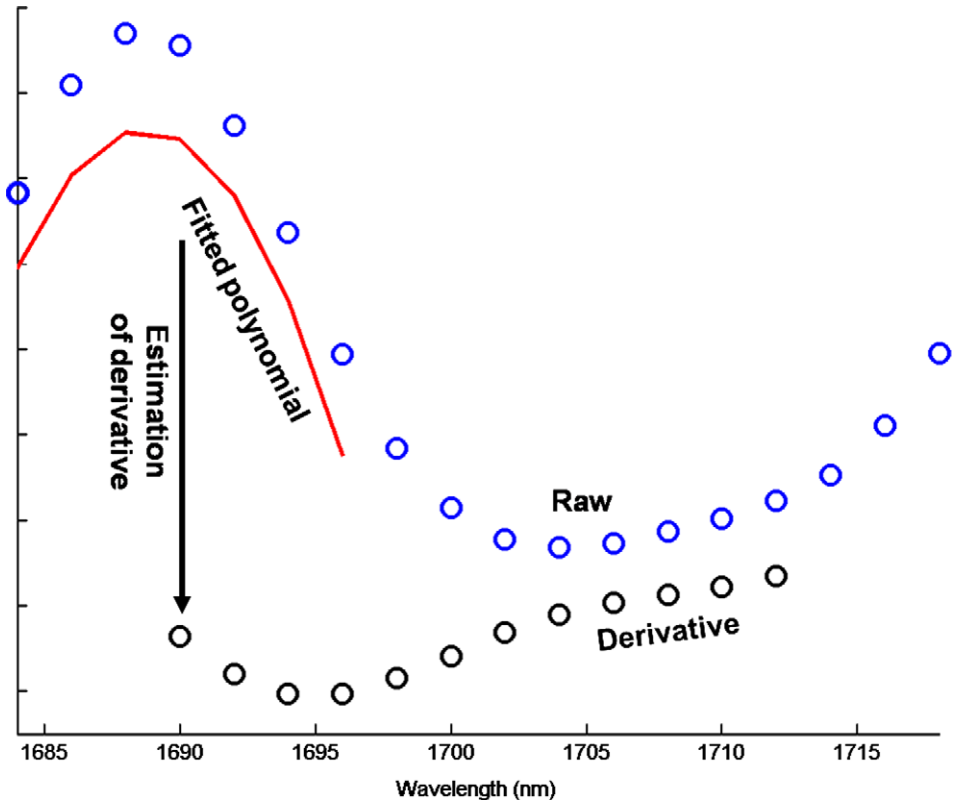

Figure 17. Estimation of the first derivative by Savitzky-Golay. A 7-point window and a second-order polynomial is used for smoothing.

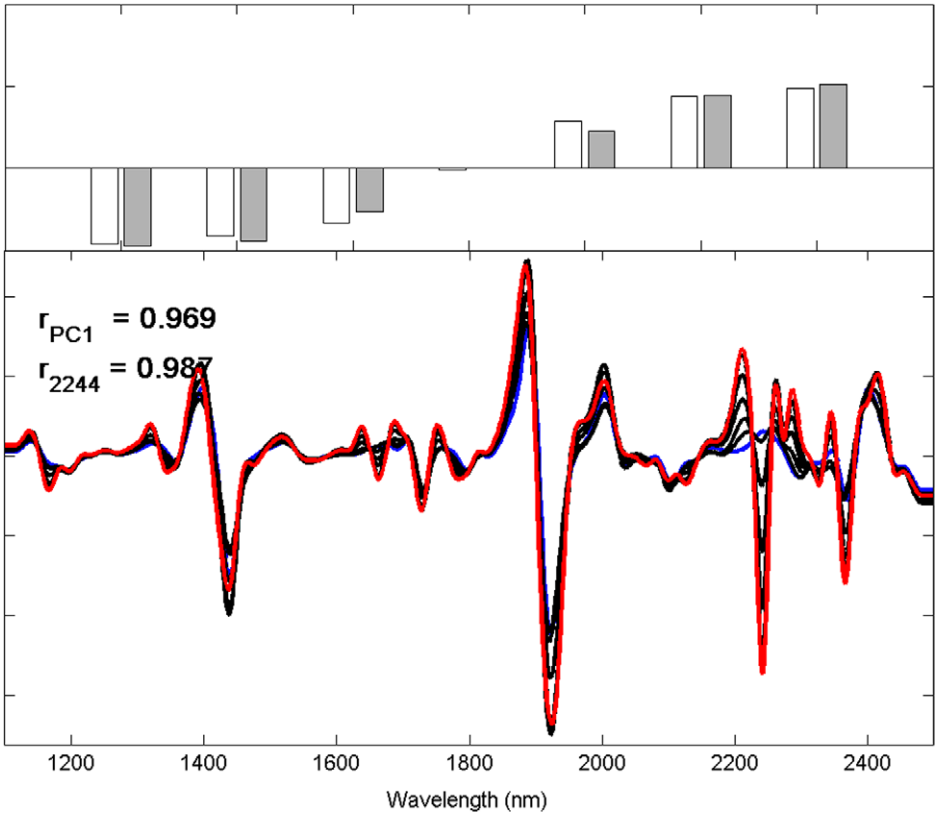

Figure 18. Savitzky-Golay estimate of the second derivative using 9 points and a second-order polynomial for the smoothing.

subsequent polynomial fits will give the same estimate of the coefficients. For the first derivative, a first-degree polynomial and second-degree polynomial will give the same answer (as will the third and fourth degrees). For the second derivative, a second and third-degree polynomial will give the same answer (as will the fourth and fifth degrees), etc. When this method was first introduced by Savitzky and Golay [21], it was still computationally cumbersome to calculate the parameters in estimating the derivative. For that reason, the authors reported a set of tabulated values for several different types of derivatives and polynomial combination. However, errors were introduced in their first article, and Steinier et al. [22] published a corrected and expanded version of the original tables. The tables were later even further expanded by Madden [23]. However, with modern computers, there is no longer any real need for these tables.

The original forms of NW and SG derivation use a symmetric window smoothing, requiring the number of data points on each side of the center point to be the same. As a consequence, the techniques neglect a number of points at each end of the spectrum during the pre-processing. For NW derivation, the number of points lost equals the number of points used for smoothing plus the size of the gap minus one. For SG derivation, the number of points lost equals the number of points used for smoothing minus one. NW derivation thus absorbs more points than SG derivation. If the spectral vector is long (i.e. more than 500 points), this issue is not important, but, for shorter spectra (e.g., diode-array instruments), this loss of wavelengths can be important.

Proctor and Sherwood in 1980 [24] and Gorry in 1990 [25] suggested a solution that involves using a fitted polynomial based on an asymmetric window for the end-points. In practice, this means that the m first points of the spectra are estimated from the $2m+1$ first points in the spectra, and a similar estimate for the last m points. However, such a solution will evidently introduce artifacts, as the accuracy of the derivative decreases with the distance from the centre point ($m+1$). Furthermore, the estimation of the end-points does not possess the inherent redundancy mentioned for SG: no two subsequent polynomial order fittings will give the same estimates. In addition to this, the estimate of the d^{th} derivative will be equal for all the end-points if the spectrum is smoothed by a d^{th}-order polynomial.

NW derivation is similar to finite differences, but introduces smoothing and gap-size as counteractions in the estimated derivate spectra to preserve the signal-to-noise ratio. These two steps in NW derivation are more or less independent. However, SG derivation uses more common filtering techniques to estimate the derivative spectra, and, instead of using the finite-difference approach, fits a polynomial through a number of points to

maintain an acceptable signal-to-noise ratio. In general, the NW and SG derivations do not give the same estimates. The only pair of settings that gives identical results is three smoothing points for both, SG using a first-order polynomial fit, and the gap size in NW equal to 1. However, more complex (and realistic) settings for SG and/or NW automatically lead to (slightly) different derivation results.

5. Interval and combined versions

Of the pre-processing techniques mentioned thus far, only the estimation of the derivatives operates by a moving-window operation, where only a local part (window) of the spectra is used at any time to estimate the correction. However, all the other methods can equally well be performed in a window-wise fashion.

Isaksson and Kowalski [26] suggested this for MSC, and named it piecewise MSC (PMSC). Andersson [27] compared alternative pre-processing methods with two versions of PMSC: moving-window or local pre-processing (dividing the wavelength axis into a few sections and performing pre-processing on each section separately).

The moving-window version of the pre-processing techniques has received little interest from the NIR community, probably because the right choice of window size is crucial and it is far from trivial to do this correctly. Too small a window will lead to the introduction of large artifacts in the corrected spectra and to a reduced signal-to-noise ratio. However, the larger the size of the window, the smaller the distinction between full and moving-window pre-processing (see Fig. 19). Local window pre-processing can be useful, especially in cases where the recorded spectra are measured all the way from the visual range or shortwave NIR up to the mid-IR range. In this wide spectral region, several different scatter issues coexist, and the spectra should be divided accordingly, performing separate scatter corrections on the different parts. However, since this is not essentially different from dividing the spectra in regions and applying the pre-processing methods independently, we do not discuss it further.

The use of combinations of pre-processing methods is abundant in literature, and, in principle, any sequence of pre-processing is possible. However, the following simple rules can serve as initial guidelines.

- Scatter correction (with the exception of normalization) should always be performed prior to differentiation. These techniques are all designed for correction of raw spectrum, and have never been thought of as corrections to a differentiated or baseline-corrected spectrum.

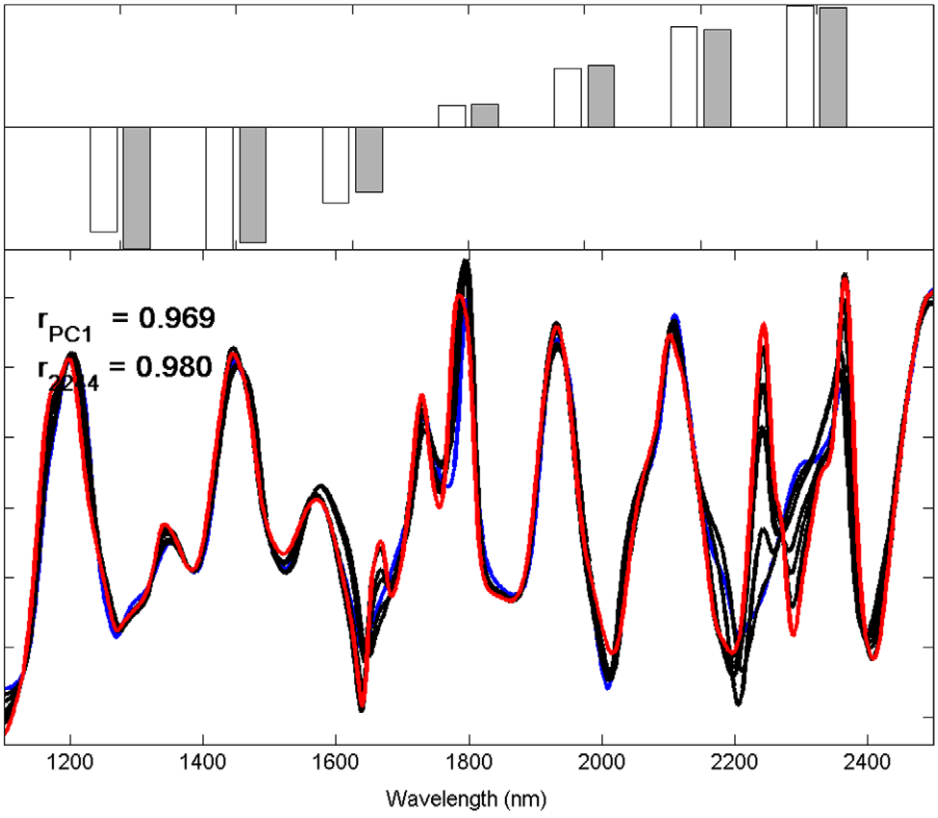

Figure 19. Moving window Standard Normal Variate using a 129-nm moving window (65 measurement points).

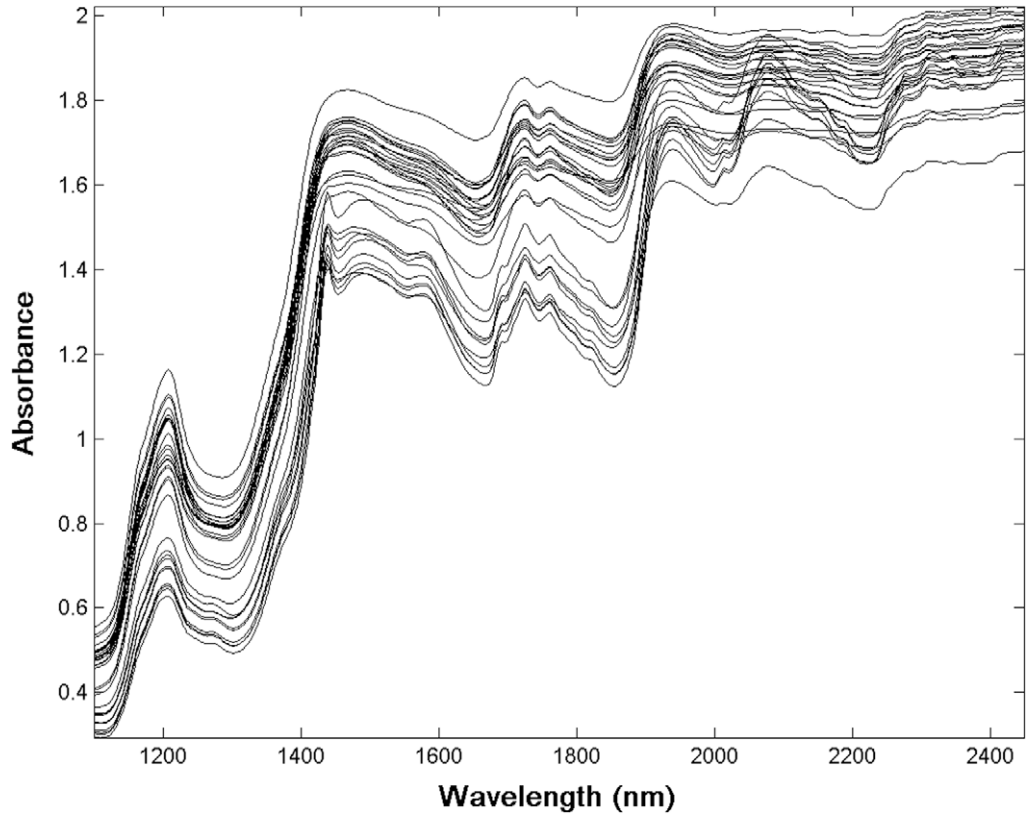

Figure 20. 32 marzipan samples measured by Instrument 2, in the interval 1100–2500 nm.

- Normalization can be used at both ends of the correction, although it is easier to assess the effect of Normalization if it is done prior to any other operation.
- The basic difference between SNV with subsequent de-trending and MSC with reference and baseline correction is that, in MSC, both corrections are applied simultaneously, not consecutively. Thus, MSC will generally give a smaller baseline correction than SNV plus de-trending.
- Performing de-trending followed by SNV was not recommended by Barnes et al. [13], and, for the reasons given above, it is not recommended to perform de-trending first.

6. A quantitative example

We will now apply all the pre-processing methods discussed to a quantitative spectroscopic task involving 32 marzipan samples measured on six very distinct spectrometers as predictor variables for two different response variables: moisture and sugar content. The data are taken from a study by Christensen et al. [3]. Fig. 20 shows one of the spectral sets. For a summary of the data, see Table 1. Here, we show the PLS-regression models, which were built for all the six NIR instruments, and responses separately (so-called PLS1 models [5]).

The marzipan NIR dataset was treated with the different pre-processing techniques described in this article. In addition to the settings used in the theoretical sections, some more extreme parameter settings were investigated to estimate piecewise MSC, to show the importance of using reasonable choices. No samples were removed as outliers, as all samples behaved well in the initial exploratory analysis. Bootstrapping-error estimates [28] were used as the validation method. A total of 1000 bootstrap drawings were performed for each combination of dataset, reference and pre-processing. The same set of drawings was used for all datasets, except for Instrument 1, where only 15 of 32 samples were measured. The 0.632-bootstrap estimate of the prediction error was calculated as shown in Equation (14), in accordance with Wehrens et al. [28].

$$RMSE_f = 0.368 \cdot \overline{RMSE_f} + 0.632 \cdot \overline{RMSEP_f} \qquad (14)$$

where $RMSE_f$ is the estimated prediction error, $\overline{RMSE}_f$ and $\overline{RMSEP}_f$ are the average calibration (samples selected per one bootstrap draw) and prediction (samples not selected per one bootstrap draw) errors across all bootstrap drawings. The optimal number of factors, f, is determined based on the 0.632-bootstrap estimates, selecting the first minimum or the point where the $RMSE_f$ curve as a function of factors flattens out (where the slope of the $RMSE_f$ curve is constant).

By applying all the pre-processing techniques to the same sample set recorded with six different instruments and/or optical measurement geometries (dispersive, interferometer, reflectance, transmission and fiber-probe) using two different responses (moisture and sugar), some general performance differences are revealed (see Tables 2 and 3).

As a very first observation, it is consoling that nearly all pre-processed models are simpler or more parsimonious (i.e. use fewer PLS factors) than the global model, independent of the spectrometer set-up and independent of the response variable.

A second general observation is that filter-based Instrument 1 in reflectance mode is not competitive in measuring the marzipan samples ($RMSE_{moisture}$ = 0.75, 4 latent variables (LVs) and $RMSE_{sugar}$ = 2.30, 3 LVs; but, we need to keep in mind that only 15 of 32 samples are measured) and pre-processing does not help to get it

Table 1. Overview of the instruments used for the measurements of 32 marzipan samples [3]

Instrument	Optical principle	Spectral range	Spectral interval	Sample unit
1[a]	Filter instrument	19 wavelengths[b]	-	Rotating sample cup, reflection
2[c]	Dispersive scanning	1100–2500 nm	2 nm	Sample cup, reflection
3	Dispersive scanning	1100–2100 nm	2 nm	Fiber probe, reflection
4	Interferometer (FT)	833–2500 nm	11.6 cm^{-1}	Sample cup, reflection
5	Interferometer (FT)	1000–2222 nm	12 cm^{-1}	Fiber probe, reflection
6	Dispersive scanning	850–1050 nm	2 nm	Sample cup, transmission

[a] Only 15 marzipan samples were measured for this instrument.
[b] The filters are centered at: 1110, 1128, 1138, 1154, 1188, 1200, 1212, 1238, 1254, 1262, 1276, 1320, 1670, 1680, 1940, 2100, 2180, 2230 and 2310 nm.
[c] Same instrument and set-up as used for measuring the pectin samples in this article.

Table 2. Prediction results for the moisture content (w/w%) of the 32 marzipan samples on six different NIR instruments with different pre-processing, given as *RMSE* with number of latent variables in brackets. Moisture content range Instrument 1 = 7.4–18.6 (w/w%), Instruments 2–6 = 6.8–18.6 (w/w%)

Method	Settings	Instruments					
		1	2	3	4	5	6
None	-	0.75 (4)	0.42 (7)	0.42 (11)	0.46 (6)	0.48 (7)	0.48 (4)
MSC	First-order reference correction	0.64 (3)	0.36 (6)	0.37 (10)	0.45 (4)	0.35 (8)	0.46 (4)
MSC	Second-order reference correction	0.62 (4)	0.41 (5)	0.41 (7)	0.50 (3)	0.37 (5)	0.45 (4)
MSC	First-order reference correction + second-order wavelength correction	0.61 (4)	0.39 (5)	0.48 (9)	0.43 (5)	0.40 (5)	0.50 (5)
SNV+Detrend	-	0.61 (4)	0.35 (5)	0.49 (9)	0.42 (4)	0.40 (5)	0.45 (3)
WMSC	First-order reference correction	0.75 (3)	0.39 (5)	0.40 (10)	0.43 (5)	0.41 (5)	0.44 (4)
LMSC	First-order reference correction	0.64 (3)	0.36 (5)	0.37 (10)	0.43 (5)	0.41 (5)	0.46 (4)
SNV	-	0.65 (3)	0.35 (6)	0.37 (10)	0.42 (5)	0.41 (5)	0.46 (4)
Norm	Euclidean	0.70 (3)	0.35 (6)	0.39 (10)	0.42 (6)	0.52 (6)	0.38 (3)
NW	Seven-point smoothing, gap-size 3, first derivative	0.66 (4)	0.38 (5)	0.40 (9)	0.41 (5)	0.43 (7)	0.45 (3)
SG	Seven-point smoothing, second-order polynomial, first derivative	0.58 (3)	0.38 (5)	0.41 (8)	0.40 (5)	0.47 (6)	0.46 (3)
SG	15-point smoothing, second-order polynomial, first derivative	3.73 (4)	0.41 (7)	0.40 (10)	0.46 (7)	0.49 (7)	0.49 (3)
Finite difference	First derivative	0.71 (4)	0.34 (5)	0.39 (9)	0.58 (4)	0.73 (4)	0.46 (3)
NW	Nine-point smoothing, gap size 3, second derivative	0.88 (4)	0.38 (5)	0.38 (9)	0.39 (7)	0.51 (6)	0.45 (3)
NW	Three-point smoothing, gap size 3, second derivative	0.58 (4)	0.35 (6)	0.41 (9)	0.53 (4)	0.71 (6)	0.46 (3)
SG	Nine-point smoothing, second-order polynomial, second derivative	0.58 (4)	0.34 (6)	0.39 (9)	0.51 (4)	0.67 (5)	0.46 (3)
Finite difference	Second derivative	0.76 (4)	0.72 (4)	0.65 (5)	0.88 (5)	0.91 (5)	0.42 (6)
PSNV	Window width 129	- (-)	0.30 (4)	0.57 (6)	0.38 (4)	0.45 (5)	- (-)
PMSC	Window width 129 First-order reference correction	- (-)	0.40 (6)	0.62 (7)	0.37 (5)	0.42 (6)	- (-)
PMSC	Window width 65 First-order reference correction	- (-)	0.33 (6)	4.17 (1)	0.35 (6)	0.45 (6)	0.47 (4)
PMSC	Window width 17 First-order reference correction	0.64 (3)	5.06 (1)	7.54 (1)	3.79 (1)	3.83 (1)	3.26 (11)
PMSC	Window width 129 Second-order reference correction	- (-)	0.38 (4)	0.71 (4)	0.40 (6)	0.52 (4)	- (-)

Table 3. Sugar prediction results for the sugar content of the 32 marzipan samples on six different NIR instruments with different pre-processing, given as *RMSE* with number of latent variables in brackets. Sugar content range 32.2–68.7 (w/w%)

Method	Settings	Instruments					
		1	2	3	4	5	6
None	-	2.30 (3)	1.48 (5)	2.47 (7)	1.09 (9)	1.89 (8)	1.50 (5)
MSC	First-order reference correction	2.11 (2)	1.31 (4)	1.94 (8)	0.95 (9)	1.64 (6)	1.69 (4)
MSC	Second-order reference correction	1.90 (5)	1.40 (3)	2.11 (8)	0.97 (8)	1.57 (6)	1.41 (5)
MSC	First-order reference correction + second-order wavelength correction	1.95 (2)	1.43 (6)	1.87 (8)	0.93 (9)	1.71 (5)	1.84 (3)
SNV+Detrend	-	2.09 (2)	1.32 (4)	1.88 (8)	1.00 (8)	1.68 (6)	1.71 (2)
WMSC	First-order reference correction	2.23 (2)	1.41 (4)	1.96 (8)	0.95 (9)	1.65 (4)	1.68 (4)
LMSC	First-order reference correction	2.11 (2)	1.31 (4)	1.94 (8)	0.92 (10)	1.64 (6)	1.69 (4)
SNV	-	2.11 (2)	1.31 (4)	1.99 (8)	0.90 (10)	1.64 (6)	1.37 (5)
Norm	Euclidean	2.23 (3)	1.46 (4)	2.33 (6)	1.09 (10)	1.99 (8)	1.39 (5)
NW	Seven-point smoothing, gap size 3, first derivative	2.07 (2)	1.55 (6)	1.99 (7)	1.18 (9)	1.75 (8)	1.43 (4)
SG	Seven-point smoothing, second-order polynomial, first derivative	2.07 (3)	1.54 (6)	1.93 (8)	1.23 (8)	1.76 (7)	1.42 (4)
SG	15-point smoothing, second-order polynomial, first derivative	12.50 (2)	1.57 (6)	2.11 (7)	1.47 (7)	2.10 (6)	1.44 (5)
Finite difference	First derivative	2.10 (3)	1.25 (6)	1.94 (6)	1.50 (5)	2.35 (5)	1.42 (4)
NW	Nine-point smoothing, gap size 3, second derivative	2.15 (2)	1.68 (3)	1.81 (8)	1.25 (7)	2.04 (7)	1.43 (2)
NW	Three-point smoothing, gap size 3, second derivative	2.10 (2)	1.22 (6)	1.97 (8)	1.39 (8)	2.47 (6)	1.42 (2)
SG	Nine-point smoothing, second-order polynomial, second derivative	2.05 (3)	1.26 (6)	2.25 (4)	1.31 (7)	2.28 (6)	1.42 (2)
Finite difference	Second derivative	2.05 (3)	1.90 (5)	2.45 (5)	2.48 (5)	3.06 (4)	1.39 (2)
PSNV	Window width 129	- (-)	1.30 (6)	2.25 (5)	1.08 (8)	1.62 (7)	- (-)
PMSC	Window width 129 First-order reference correction	- (-)	1.54 (5)	2.50 (6)	1.23 (6)	1.59 (5)	- (-)
PMSC	Window width 65 First-order reference correction	- (-)	1.54 (4)	14.63 (1)	3.72 (1)	1.73 (5)	1.52 (6)
PMSC	Window width 17 First-order reference correction	2.11 (1)	17.84 (1)	24.11 (1)	13.97 (1)	13.13 (1)	12.01 (6)
PMSC	Window width 129 Second-order reference correction	- (-)	1.33 (5)	2.28 (6)	1.45 (6)	1.71 (5)	- (-)

down to the level of the other instruments. Filter-based instruments are not really compatible with spectral derivation techniques, but the other pre-processing techniques also fail to reach the desired performance.

A third general comment can be made on the holographic information content of NIR, in which information (the overtones) is repeated multiple times. The small spectral region 850–1050 nm (covered by Instrument 6) that contains the second overtone of the O-H and N-H stretches and the third overtone of the C-H stretches is fully competitive with the more sophisticated instruments covering the complete or traditional NIR region. Moreover, it is interesting to note that the models created from spectra from transmission-based Instrument 6 in general are the simplest, even before pre-processing. Apparently, the scatters from the density fluctuations in the samples measured in transmission mode are less demanding than the reflective scatters measured in reflection mode. When it comes to pre-processing, it is surprising that, in contrast to all other instruments, the Euclidean norm works very well and gives the best results for Instrument 6 ($RMSE_{moisture}$ = 0.38, 3 LVs, and $RMSE_{sugar}$ = 1.39, 5 LVs). The reason might be that this small NIR region contains all hydrogen-bond stretches from the sample and a normalized approach thus corresponds to integrating all the proton signals and setting the proton-density equal between samples. Besides the normalization approach, it would appear that derivatives are a good pre-processing strategy for this type of data, as they can consistently simplify the models, as is particularly obvious in the case of sugar prediction.

For the remaining full NIR-region Instruments 2–5, we find some interesting and strong differences dependent on the response variable, presumably because the moisture content is a low-resolution spectral task while sugar content is a high-resolution problem.

In the case of the moisture models, dispersive Instruments 2 and 3 are almost consistently better than the models based on Fourier-transform Instruments 4 and 5.

The best overall model is found for Instrument 2 with $PSNV_{window\text{-}width\ 129}$ pre-processing ($RMSE$ = 0.30, 4 LVs) and the best Fourier-transform model is found for Instrument 5 with fiber optics using MSC with second-order reference correction ($RMSE$ = 0.37, 5 LVs). When adding a fiber probe to Instrument 2 (= Instrument 3), the complexity of the models increases (on average by 3 LVs).

This large difference can be assigned to the more complex optical geometry of the later system. Furthermore, the performance without pre-processing remains the same (*RMSE* = 0.42), but the pre-processed performance of Instrument 3 is inferior (*RMSE* = 0.37, 10 LVs using SNV and MSC) to that of the best model of Instrument 2.

In the case of sugar models, the situation is almost reversed. Here, interferometer-based Instrument 4 displays consistently the best models, albeit more complex, presumably due to the much better spectral resolution of this instrument. The best overall model is found for Instrument 4 with $MSC_{1st\ order\ ref.\ 2nd\ order\ wave}$ pre-processing (*RMSE* = 0.92, 9 LVs), which is far better than the best dispersive results (*RMSE* = 1.30, 4 LVs for Instrument 2) but also much more complex. Again, in the case of the sugar models, adding a fiber probe to Instrument 2 (= Instrument 3) makes the models inferior and much more complex (on average two more LVs and an increase in *RMSE* between the two best models from 1.22 for Instrument 2 to 1.81 for Instrument 3).

The moving-window versions of SNV and MSC show varying results. In general, moving-window versions give results similar to or better than the best remaining pre-processing method. However, the *RMSE* is at best 10% better than the best normal pre-processing technique, but window selection could easily become a critical parameter. For comparison, some sub-optimal moving-window approaches are included as the last three rows in Tables 2 and 3.

Discrepancy between the finite-difference approach to derivation and the more sophisticated methods is not apparent in the estimate of the first derivative for some measurements (Instruments 2, 3 and 6). This fits well with the smooth behavior of these systems, indicating that additional smoothing is not necessary. The interferometers (Instrument 4 and 5) have a better spectral resolution, giving rise to a higher degree of fine structure, which leads to a lower signal-to-noise ratio in the estimate of the first derivative by the finite-difference method. This, in the end, leads to inferior models. The results for using the finite difference for the second derivatives are generally that they are all inferior to the more sophisticated methods. This indicates that the signal-to-noise ratio in the spectra has been lowered significantly (as expected), letting noise influence the models to a greater degree. However, even though finite difference may lead to good models for the estimate of the first derivative, it is safer to always stick with smoothing derivation methods.

7. Concluding remarks

Obviously, our quantitative example does not give the authoritative answer about which pre-processing to use in any given case. However, it does appear sensible to use normalization for short-wave NIR-transmission spectra and to use MSC (with first-order reference correction) or standard SNV for most other cases.

While it is difficult to find the best pre-processing, it is indeed possible to use wrong pre-processing. This will primarily happen due to incorrect parameter settings of window size and/or smoothing functions in the estimate of the derivative and the moving-window techniques.

Finally, we emphasize that the maximum improvement of any pre-processed model when compared to the global model is approximately 25% in *RMSE* in our study. While a 25% reduction might be important for industrial applications [29], this is hardly what makes the difference in the many multivariate feasibility studies that flourish in the scientific literature, for which we could recommend selecting pre-processing so as to achieve parsimonious, interpretable models.

References

[1] S. Wold, K. Esbensen, P. Geladi, Chemom. Intell. Lab. Syst. 2 (1987) 37.
[2] S.B. Engelsen, E. Mikkelsen, L. Munck, Progr. Colloid Polym. Sci. 108 (1998) 166.
[3] J. Christensen, L. Nørgaard, H. Heimdal, J.G. Pedersen, S.B. Engelsen, J. Near Infrared Spectrosc. 12 (2004) 63.
[4] S. Wold, H. Martens, H. Wold, Lect. Notes Math. 973 (1983) 286.
[5] H. Martens, T. Næs, Multivariate Calibration, Wiley, New York, USA, 1989.
[6] H. Martens, S.A. Jensen, P. Geladi, Multivariate linearity transformations for near infrared reflectance spectroscopy, in: O.H.J. Christie (Editor), Proc. Nordic Symp. Applied Statistics, Stokkland Forlag, Stavanger, Norway, 1983, pp. 205–234.
[7] P. Geladi, D. MacDougal, H. Martens, Appl. Spectrosc. 39 (1985) 491.
[8] H. Martens, E. Stark, J. Pharm. Biomed. Anal. 9 (1991) 625.
[9] D.K. Pedersen, H. Martens, J.P. Nielsen, S.B. Engelsen, Appl. Spectrosc. 56 (2002) 1206–1214.
[10] H. Martens, J.P. Nielsen, S.B. Engelsen, Anal. Chem. 75 (2003) 394.
[11] M. Decker, P.V. Nielsen, H. Martens, Appl. Spectrosc. 59 (2005) 56.
[12] S.N. Thennadil, E.B. Martin, J. Chemom. 19 (2005) 77.
[13] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Appl. Spectrosc. 43 (1989) 772.
[14] I.S. Helland, T. Næs, T. Isaksson, Chemom. Intell. Lab. Syst. 29 (1995) 233.
[15] N.B. Gallagher, T.A. Blake, P.L. Gassman, J. Chemom. 19 (2006) 271.
[16] W. Windig, J. Shaver, R. Bro, Appl. Spectrosc. 62 (2008) 1153.
[17] Q. Guo, W. Wu, D.L. Massart, Anal. Chim. Acta 382 (1999) 87.
[18] M.S. Dhanoa, S.J. Lister, R. Sanderson, R.J. Barnes, J. Near Infrared Spectrosc. 2 (1994) 43.
[19] K.H. Norris, Extracting information from spectrophotometric curves - Predicting chemical composition from visible and near infrared spectra, in: H. Martens, H. RusswurmJr. (Editors), Food Research and Data Analysis–Proc. IUFOST Symposium, Applied Science Publishers, London, UK, 1983, pp. 95–113.
[20] K.H. Norris, P.C. Williams, Cereal Chem. 61 (1984) 158.
[21] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627.
[22] J. Steinier, Y. Termonia, J. Deltour, Anal. Chem. 44 (1972) 1906.
[23] H.H. Madden, Anal. Chem. 50 (1978) 1383.

[24] A. Proctor, P.M.A. Sherwood, Anal. Chem. 52 (1980) 2315.
[25] P.A. Gorry, Anal. Chem. 62 (1990) 570.
[26] T. Isaksson, B.R. Kowalski, Appl. Spectrosc. 47 (1993) 702.
[27] C.A. Andersson, Chemom. Intell. Lab. Syst. 47 (1999) 51.
[28] R. Wehrens, H. Putter, L.M.C. Buydens, Chemom. Intell. Lab. Syst. 54 (2000) 35.
[29] C.B. Zachariassen, J. Larsen, F. van den Berg, S.B. Engelsen, Chemom. Intell. Lab. Syst. 76 (2005) 149.