Journal of Pharmaceutical and Biomedical Analysis 48 (2008) 554–561

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms—Comparing common calibration approaches

Carsten Ravn[a,b,*], Erik Skibsted[b], Rasmus Bro[c]

[a] Department of Pharmaceutics and Analytical Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark
[b] CMC Formulation and Analysis, Novo Nordisk A/S, Novo Nordisk Park, 2760 Maaloev, Denmark
[c] Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark

ARTICLE INFO

Article history:
Received 27 May 2008
Received in revised form 21 July 2008
Accepted 21 July 2008
Available online 31 July 2008

Keywords:
NIR chemical imaging (NIR-CI)
Chemical imaging
Hyperspectral image
Hyperspectral data cube
Image analysis
Solid dosage forms
Classical/partial least squares

ABSTRACT

Near-infrared chemical imaging (NIR-CI) is the fusion of near-infrared spectroscopy and image analysis. It can be used to visualize the spatial distribution of the chemical compounds in a sample (providing a chemical image). Each sample measurement generates a hyperspectral data cube containing thousands of spectra. An important part of a NIR-CI analysis is the data processing of the hyperspectral data cube. The aim of this study was to compare the ability of different commonly used calibration methods to generate accurate chemical images. Three common calibration approaches were compared: (1) using single wavenumber, (2) using classical least squares regression (CLS) and (3) using partial least squares regression (PLS1). Each method was evaluated using two different preprocessing methods.

A calibration data set of tablets with five constituents was used for analysis. Chemical images of the active pharmaceutical ingredient (API) and the two major excipients cellulose and lactose in the formulation were made. The accuracy of the generated chemical images was evaluated by the concentration prediction ability. The most accurate predictions for all three compounds were generated by PLS1. The drawback of PLS1 is that it requires a calibration data set and CLS, which does not require a calibration data set, therefore proved to be an excellent alternative. CLS also generated accurate predictions and only requires the pure compound spectrum of each constituent in the sample. All three calibration approaches were found applicable for hyperspectral image analysis but their relevance of use depends on the purpose of analysis and type of data set. As expected, the single wavenumber method was primarily found useful for compounds with a distinct spectral band that was not overlapped by bands of other constituents.

This paper also provides guidance for hyperspectral image (or NIR-CI) analysis describing each of the typical steps involved.

1. Introduction

Near-infrared chemical imaging (NIR-CI) is an emerging technology within the pharmaceutical industry compared to the now well-established traditional NIR spectroscopy. Pharmaceutical NIR spectroscopy applications range from raw material testing through process monitoring to final product analysis [1–5]. The conventional single point NIR spectroscopy measures a bulk average NIR spectrum and reflects an average composition of the sample. NIR-CI adds spatial distribution information to the spectral information by combining traditional NIR spectroscopy with digital imaging. In NIR-CI, a NIR spectrum is recorded in each pixel of the sample image resulting in a hyperspectral data cube. Translating the spectral signature from each pixel into, for example, chemical concentrations will generate a set of chemical images showing the distribution of each ingredient within the sample matrix. This visualization of the internal structure and elucidation of the distribution and cluster size of each constituent in the sample is valuable in formulation development and manufacturing of solid dosage forms as well as for troubleshooting quality defects. NIR-CI has the potential to provide increased process and product understanding which goes well in hand with the process analytical technology (PAT) initiative of the FDA [6]. Briefly, the concept of PAT is to build in quality by design instead of merely passively testing the quality of the products and manufacturing processes. PAT promotes technologies that can identify and monitor critical process parameters and the goal is to enhance understanding and control the manufacturing process. NIR-CI is such a technology and has received attention by the FDA, which has evaluated NIR-CI for different pharmaceutical applications [7–10].

* Corresponding author at: Novo Nordisk A/S, CMC Formulation and Analysis, Novo Nordisk Park, B6.1.070, 2760 Maaloev, Denmark. Tel.: +45 44431481.
E-mail addresses: cra@novonordisk.com, carstenravn@privat.dk (C. Ravn).

0731-7085/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2008.07.019

The majority of the early NIR-CI literature in pharmaceutical analysis describes the general principle of this new technology and its potential use. The applications include root-cause analysis of manufacturing problems, product development, quality assurance and quality control but are mostly feasibility studies on relatively simple model systems or examples on a single pharmaceutical sample [7,11–17]. The pharmaceutical NIR-CI research later moved into developing methods to analyse hyperspectral NIR images and investigating the factors affecting NIR-CI of solid dosage forms [18–23]. In the past few years the number of pharmaceutical applications using NIR-CI has increased significantly [3,9,24–29] and recently NIR-CI is seen integrated in formulation development [30,31], used for mechanistic powder blending studies [32,33] and a review has also been published [34].

For NIR-CI to develop into a useful and well accepted technology in pharmaceutical analysis it is important to have a thorough understanding of how to properly measure and analyse such data. The analytical work of a NIR-CI experiment can be divided into three overall steps:

- **Data acquisition** Includes sample preparation, instrumental settings and basic spectral transformation. The raw data output from a NIR-CI measurement is organised in a 3D data structure with two spatial axes and one wavelength axis, also called a hyperspectral data cube.
- **Data processing** The processing of the hyperspectral data cube into a, typically chemical, image by univariate or multivariate image analysis approaches. This part includes wavelength selection, spectral preprocessing and the subsequent data analysis to generate the chemical images showing the distribution of each of the ingredients within the imaged sample.
- **Image processing** The processing of the generated chemical images into relevant and 'useful' information that will qualitatively or quantitatively describe the properties of a sample in relation to the problem investigated. This could, for example, be a total concentration or a measure of the distribution of the concentration of the active ingredient.

Each of the three steps in a full hyperspectral image analysis is important for a successful NIR-CI experiment. If the spectral quality from the data acquisition is poor, no multivariate image analysis method is able to compensate for this and still generate accurate results. If the data processing method is suboptimal, inaccurate chemical images will be generated that will lead to erroneous conclusions in the subsequent image processing analysis. And finally, even when an accurate chemical image is generated, poor image processing methods may extract the wrong product or process-related information from the images. It is therefore imperative that each of the three overall steps is thoroughly investigated and their strengths and limitations are known.

A variety of factors affect the quality of the output for each of the three steps. The present study focuses on the data processing part generating the chemical images. Unfortunately, there exists no universal data processing method that is superior for all hyperspectral data cubes. The choice of proper analysis will depend on the data set and the purpose of analysis. Table 1 presents an overview of calibration methods demonstrated in the NIR-CI literature analysing hyperspectral data cubes of pharmaceutical samples. Common for most of the studies in Table 1 is that the pharmaceutical application is the main purpose. The data processing methods is of course an important part of the studies but often not critically evaluated for its appropriateness. Many of the early NIR-CI studies used univariate approaches (single wavenumber, peak–height ratio etc.). More attention has since been drawn to develop multivariate approaches to extract more information from the hyperspectral images. For example, Jovanovic et al. [26] evaluated four different data processing approaches to analyse mixtures of lysozyme and trehalose. The contrast in the chemical images were compared by methods using intensity of a single wavelength, peak–height ratio of two wavelengths, correlation coefficient with a reference spectrum and principal component analysis (PCA). The correlation coefficient method was also compared with partial least squares (PLS) regression for further homogeneity investigations. Gendrin et al. [27] compared classical least squares (CLS) and PLS regression for best content prediction of the active pharmaceutical ingredient (API) and two excipients in pharmaceutical solid dosage forms.

Table 1
Calibration methods used for analysing hyperspectral data cubes in pharmaceutical applications

Calibration approach	Reference
Single wavelength	[7,12,24,26,30,36,37]
Peak–height ratio	[26,38]
Correlation coefficient	[26]
PCA	[19,21,24,31,33,36,38]
CLS	[27,30]
PLS2 (pure spectra)	[7,9,14,19,24,29,32,33]
PLS2 (calibration set)	[26,27]

Although different univariate and multivariate data processing approaches are applied to pharmaceutical applications it is often not easy to compare the results. Chemical images generated from different data processing methods may visually look similar but actually provide different chemical information. Thus, there is a lack of objective criteria or, for example, a 'NIR-CI calibrated tablet' to assess what the 'best' or most accurate image is [20]. Studies are often seen comparing data processing methods by differences in contrast of the generated images. However, the goal of the data processing step in a NIR-CI experiment is not to generate *high* contrast but to generate the *right*, i.e. accurate, contrast.

In this study, three common calibration approaches were evaluated for their ability to generate accurate chemical images of the API and two major excipients in a five-compound pharmaceutical solid dosage form. The aim of the study is to investigate the ability of different commonly used hyperspectral image data processing methods to generate accurate chemical images. The three calibration approaches compared were (1) using a single wavenumber for calibration, (2) using classical least squares (CLS) where estimates of pure spectra are used to obtain concentration estimates [35] and (3) partial least squares regression (PLS) where a regression model is built between measured spectra and known concentrations [35]. Two different spectral preprocessing methods were investigated for each of the three data processing approaches. They were selected as the two best performing preprocessing methods selected from a comparative study of a range of different preprocessing approaches applied to each of the data processing methods. Further, this paper delineates the general steps involved in data processing of hyperspectral data cubes and can thus also be used as practical guidance for this analytical approach.

2. Materials and methods

2.1. Materials

Due to intellectual property rights, the name and structure of the active pharmaceutical ingredient (API) cannot be shown. It is simply denoted API. The excipients for the tablet formulation were silicified microcrystalline cellulose (ProSolv SMCC® HD90, JRS Pharma, Germany), α-lactose monohydrate (Tablettose®70, Meggle, Germany), magnesium stearate (Liga MF-2-V, Peter Greven Fett-Chemie, Germany) and talc (Unikem, Denmark).

2.2. Samples

A five-compound conventional pharmaceutical tablet formulation was used to produce the calibration data set analysed throughout this study. The nominal composition was active pharmaceutical ingredient (API: 6.3%, w/w), microcrystalline cellulose (MCC: 20.0%, w/w), lactose (lact: 71.5%, w/w), and the lubricants magnesium stearate (0.75%, w/w) and talc (1.5%, w/w). From this nominal composition a calibration data set of 9 batches was designed by a D-optimal formulation design using Modde software [39]. The design was constructed to vary the API and cellulose ±30% from their nominal values. The content of the lubricants magnesium stearate and talc were fixed and lactose was adjusted to make a total of 100%. Table 2 shows the concentrations of the five compounds for each of the 9 calibration batches.

The dry-blend formulations were all mixed in a drum-mixer and compressed into tablets of 175 mg by direct compression on a 6-punch station rotary tablet press. A flat punch-set was used to obtain a flat sample surface. The diameter of the tablets was 8 mm and the thickness 2.6 mm. Batch sizes were 500 g and tablets were collected from start, mid and end of the tabletting process. Pure compound reference samples of the API and the excipients were also produced. Approximately 250 mg of each raw material was compressed into 8 mm diameter wafers on a hydraulic tablet press using 10 kN pressure for 10 s. The wafers were analysed similar to the pharmaceutical tablets and used to generate pure compound reference spectra.

2.3. Data acquisition

To get a representative sampling from each batch two tablets from start, mid and end of the tabletting process were analysed from each of the 9 calibration batches, i.e. a total of 54 samples (6 tablets from each of 9 batches).

Each tablet was fixed onto a microscope slide using cyanoacrylate glue and measured directly on the flat tablet surface. Samples were analysed on a NIR line mapping system (Spectrum Spotlight 350 FT-NIR Microscope, PerkinElmer, UK) from which 16 spectra were collected in each acquisition from a linear MCT detector array. An area of 2 mm × 2 mm were analysed using pixel size 25 μm × 25 μm thus obtaining a total of 6400 spectra (= pixels) for each image. Each spectrum was the average of 8 scans from wavelength region 7800–4000 cm^{-1} using a 16 cm^{-1} spectral resolution.

2.3.1. Spectral correction

As the spectral responses obtained from a NIR-CI measurement contain information from both the sample and the instrument it is necessary to correct for the instrument response by using a background reference. The raw data from the data acquisition is thus relative NIR diffuse reflectance data ($R = R_{sample}/R_{background}$) organised in a 3D structure (hyperspectral data cube). The high-reflectance standard Spectralon™ (Labsphere, Inc., North Sutton, New Hampshire) was used as background reference in this study.

The background corrected 3D image data files were imported into Matlab software [40]. All image data processing was performed using in-house scripts together with PLS_Toolbox [41].

2.3.2. Conversion to absorbance

Prior to data analysis all raw reflectance data (R) were transformed into absorbance (A) by the relation $A = -\log_{10}(1/R)$. Assuming the path length on average is constant for the NIR diffuse reflectance mapping measurements of the sample, a linear relationship exists between absorbance and chemical compound concentration (Beer–Lambert law).

2.3.3. Unfold 3D hyperspectral data cube

Hyperspectral image data can be analysed by both ordinary two-way and three-way methods but the two-way methods have been found most suitable for this type of data [42]. In our study, ordinary two-way multivariate methods are compared and to make hyperspectral image data amenable for two-way methods it is necessary to unfold the 3D hyperspectral data cube to a 2D matrix, in which each row is a spectrum related to one of the pixels. Once all data acquisition and data processing has been performed the resulting 2D matrix is refolded to retain the pixel location of each spectrum and generate the chemical image.

2.4. Data processing

Prior to applying the actual data analysis method that generates the chemical image the wavenumber range and spectral preprocessing methods must be selected. These two steps are described below together with the specific settings used in this study.

2.4.1. Variable selection

Multivariate methods often excel above univariate methods because of their ability to use the entire measured wavenumber range. Nevertheless, the precision of a multivariate method can, in some cases, be improved by a proper variable selection. In this study, variable selection by variable importance in the projection (VIP) [43,44] was used to select the optimal wavenumber range(s) for the PLS1 model. For CLS the noisy wavenumber ends were removed and the spectral range was reduced to 7500–4200 cm^{-1}. For the single wavenumber method the wavenumber was selected at positions with a distinct spectral absorption band having little spectral overlap from the other compounds. This was assessed visually.

2.4.2. Spectral preprocessing

The raw NIR diffuse reflectance spectra obtained from a NIR-CI measurement contain both chemical and non-chemical information about the solid sample [45]. The source of the non-chemical information may be from the sample (e.g. uneven sample surface or differences in sample density) and/or the instrumentation (e.g. changes in lamp intensity or detector response). The effects are typically observed as spectral baseline offsets or a sloping baseline.

As it is the chemical information that is of interest the non-chemical biases are sought and removed by different preprocessing techniques. These preprocessing techniques are routinely used in conventional NIR spectroscopy and their effects on hyperspectral NIR images have also been investigated [21,27]. The most common preprocessing approaches used in NIR-CI experiments on pharmaceutical solid dosage forms are first and second Savitzky–Golay derivative transformation [46], standard normale variate (SNV) [47], multiplicative scatter correction (MSC) [48] or a combination hereof.

In this study a calibration data set was available. It was thus possible to perform regression analysis by all three calibration approaches trying different preprocessing treatments and choosing the preprocessing giving best results. Using this approach the two best preprocessing methods found for each data processing method were selected and used for comparison throughout this study. Savitzky–Golay derivative transform implies choosing derivative order, filter width and polynomial order. For example a first derivative transform with a nine-point filter width and polynomial order three is denoted here as "first (or 1st) derivative (9/3)".

2.4.3. Calibration methods

At this stage of the overall NIR-CI analysis the spectral data are translated into concentrations producing the NIR chemical images.

Table 2
Composition (% w/w) of the five-compound pharmaceutical tablet formulations constituting the 9 calibration batches

Ingredients (particle size*)	1	2	3	4	5	6	7	8	9
API (2.4/11/129)	4.38	8.14	4.38	8.14	4.38	8.14	6.26	6.26	6.26
Cellulose (43/121/272)	14.00	14.00	26.00	26.00	20.00	20.00	14.00	26.00	20.00
Lactose (13/62/152)	79.37	75.61	67.37	63.61	73.37	69.61	77.49	65.49	71.49
Magnesium stearate (1.7/4.7/19)	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
Talc (3.5/13/44)	1.50	1.50	1.50	1.50	1.50	1.50	1.50	1.50	1.50

* Particle size measures (μm) of the cumulative volume distribution given by the 10%, 50% and 90% percentiles (D[v,0.1]/D[v,0.5]/D[v,0.9]) obtained using a Malvern Mastersizer 2000 laser diffraction system.

Similar to preprocessing treatments there do not exist any standard calibration method that is superior for analysing all hyperspectral data cubes. The choice of calibration method will depend on the type of data set and the purpose of analysis.

The goal of this study was to compare the ability of three common calibration methods to generate accurate chemical images of the major ingredients in a solid dosage form. Hyperspectral NIR image data from 54 tablets (6 tablets from 9 calibration batches) were analysed by a single wavenumber method, CLS and PLS1.

The nine predicted concentrations (mean of 6 replicates) were calculated for API, cellulose and lactose for each of the three methods. The method that generated the most accurate chemical images was evaluated by the accuracy of prediction of the overall concentration in the image, which was assessed by the model prediction error (root mean square error of cross-validation; RMSECV). Characteristics of the three common calibration methods are summarized in Table 3.

All commonly used calibration approaches share the feature that they do not actively exploit the spatial information in the images. Alternatives that do exploit this are available [49] but they are not commonly used and are hence not included in this work. A general description of the principles for each of the three calibration methods used to analyse hyperspectral NIR images follows together with the specific settings used for each data processing method in this study.

2.4.3.1. Single wavenumber method. The single wavenumber method is a univariate approach and the chemical images generated from this method are based on the absorbance intensity values in each pixel at one specific wavenumber. To obtain an image with chemical information for a specific compound it is important to select a wavenumber with a strong and distinct absorption band for that particular compound, i.e. with as little absorption interference as possible from the other compound's spectra in the sample. The most distinct absorption band is best selected from the pure compounds NIR spectra if available and preferably from a derivative form of the spectra as this transform will enhance the spectral resolution of overlapping bands and highlight subtle spectral peaks. Second derivatives spectra are often preferred as the absorption peaks from this preprocessing technique appear at the same position as for the original peaks.

In this study, pure compound spectra normalised using SNV followed by second derivative (9/3) were used to select the most distinct absorption band for each of the three analysed compounds in the tablets. This preprocessing method was selected together with a second derivative (15/3) preprocessing for the single wavenumber regression analyses.

2.4.3.2. Classical least squares (CLS). The multivariate classical least squares (CLS) algorithm is often used in simpler spectroscopic applications and also appears to be an obvious choice for analysing hyperspectral images such as those in this paper [35]. The CLS model is based on the assumption that the measured spectra are the sum of pure compound spectra weighted by the concentration of the compounds. The relative concentrations of the compounds in the sample can thus be estimated using only the pure compound spectra according to Beer–Lambert's law.

In our study, the pure compound spectrum for each of the five constituents was calculated as the mean spectrum of the hyperspectral data cube for each pure compound reference sample. Using all five pure compound spectra and the image cube mean spectrum of a sample the relative concentrations of the compounds in the sample matrix were estimated by the CLS model. To generate a chemical image for each compound in a sample, all spectra from the hyperspectral data cube were used to estimate the concentration values for each compound in each pixel.

The wavenumber range used for CLS in our study was selected by removing the noisy end channels thus reducing the wavenumber range to 7500–4200 cm^{-1}. The lowest prediction error was obtained using either the first derivative (9/3) or the second derivative (9/3) preprocessing of spectra and these two preprocessing methods were therefore selected for comparison.

2.4.3.3. Partial least squares (PLS1). PLS is a multivariate regression method used to build quantitative calibration models [35,50]. It is a regression method that relates two data matrices, X (spectra) and Y (reference values), with each other. PLS requires a calibration data set composed of several samples spanning an appropriate concentration range to build a model for new predictions. This may limit the use of the PLS method as calibration data sets can often be difficult to obtain, for example, in the early development of a pharmaceutical formulation.

Table 3
Comparison of the three calibration methods used to analyse the hyperspectral NIR data cubes in this study

Calibration method	Samples required for method	Characteristics of method and preferred use
Single wavenumber (univariate)	NIR spectra of pure compounds (to identify distinct absorption peaks)	Qualitative use. Easy, seemingly intuitive and fast. Exploratory analysis, valid for simple sample matrices. Distinct spectral band for compounds required.
CLS (multivariate—supervised)	NIR spectra of pure compounds	Quantitative use. Relatively accurate, fast and easy. Requires only pure compound spectra. Assumes Beer's law valid.
PLS (multivariate—supervised)	Calibration data set of known compositions	Quantitative use. Accurate, robust predictive precision. Requires full calibration data set.

Table 4
Results from concentration predictions of API, cellulose and lactose for the single wavenumber method, CLS and PLS1

	Single wavenumber			CLS			PLS1		
Compounds	API	Cellulose	Lactose	API	Cellulose	Lactose	API	Cellulose	Lactose
Preprocessing	2nd derivative (15/3)			1st derivative (9/3)			1st derivative (9/3) + mean center		
Wavenumber region (cm^{-1})	5984	4280	5168	4200-7500			4256-4744; 5840-6176 (105 var.)	4088-4488 (52 var.)	4248-4520; 5008-5400 (85 var.)
Corr. coef. R^2	0.98	0.95	0.92	0.98	0.98	0.96	0.99	0.98	0.99
Slope	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.01	0.99
RMSEC	0.29	1.56	2.03	0.30	0.93	1.50	0.05	0.24	0.32
RMSECV	0.38	1.94	2.46	0.40	1.16	1.82	0.18	0.70	0.62
Preprocessing	SNV + 2nd derivative (9/3)			2nd derivative (9/3)			2nd derivative (9/3) + mean center		
Wavenumber region (cm^{-1})	5984	4280	5168	4200-7500			4304-4792; 5816-6200 (112 var.)	4104-4496 (50 var.)	4264-4536; 5016-5400 (86 var.)
Corr.coef. R^2	0.96	0.97	0.86	0.99	0.97	0.95	0.98	0.98	0.97
Slope	1.00	1.00	1.00	1.00	1.00	1.00	1.04	1.04	0.92
RMSEC	0.44	1.53	2.59	0.17	1.11	1.60	0.07	0.33	0.35
RMSECV	0.54	1.97	3.38	0.23	1.34	1.94	0.23	0.87	0.88

For each calibration method results are presented for the two preprocessing treatments showing best predictions. The grey-shaded area points out the best predictions obtained.

When a new hyperspectral data cube of a sample with identical ingredients but unknown concentrations is applied to a PLS model it will convert the spectral information of each pixel into predicted concentrations. The generated PLS prediction image of each compound therefore shows the predicted concentration in each pixel (= chemical image). The predicted compound concentration of the imaged sample is calculated by the mean value of all predicted pixel concentrations.

In our study, PLS1 models were developed for each of the three major compounds API, cellulose and lactose. The image cube mean spectrum for each of the 54 calibration samples was first calculated followed by computing the mean spectrum of the 6 replicates for each of the 9 calibration batches. The resulting 9 mean spectra formed the data matrix X. The reference values in the Y matrix were the 9 theoretical concentrations (% w/w) for each of the three major constituents given in Table 2. Each PLS1 model was developed and optimised regarding number of PLS components, wavenumber range(s) and preprocessing methods evaluated from regression analysis. Cross-validation was performed by leave-one-out cross-validation. The chosen preprocessing methods were first derivative (9/3) and second derivative (9/3) both followed by mean centering and the wavenumber ranges optimised for each model can be read from Table 4.

3. Results and discussion

3.1. Pure compound spectra and mean spectra of calibration samples

The normalised (SNV) NIR absorbance spectra of the pure components API, cellulose and lactose were first examined to select specific wavenumbers for the single wavenumber analysis (Fig. 1, left). A distinct, sharp absorption band for API at 5984 cm^{-1} was easily identified. It was more difficult to identify distinct absorption bands for cellulose and lactose. They both showed broad absorption bands and were difficult to distinguish from each other due to their similar spectral pattern caused by their chemical resemblance. The second derivative (9/3) of the NIR normalised (SNV) spectra were therefore generated (Fig. 1, right). A distinct absorption band for lactose at 5168 cm^{-1} was now resolved but it was still

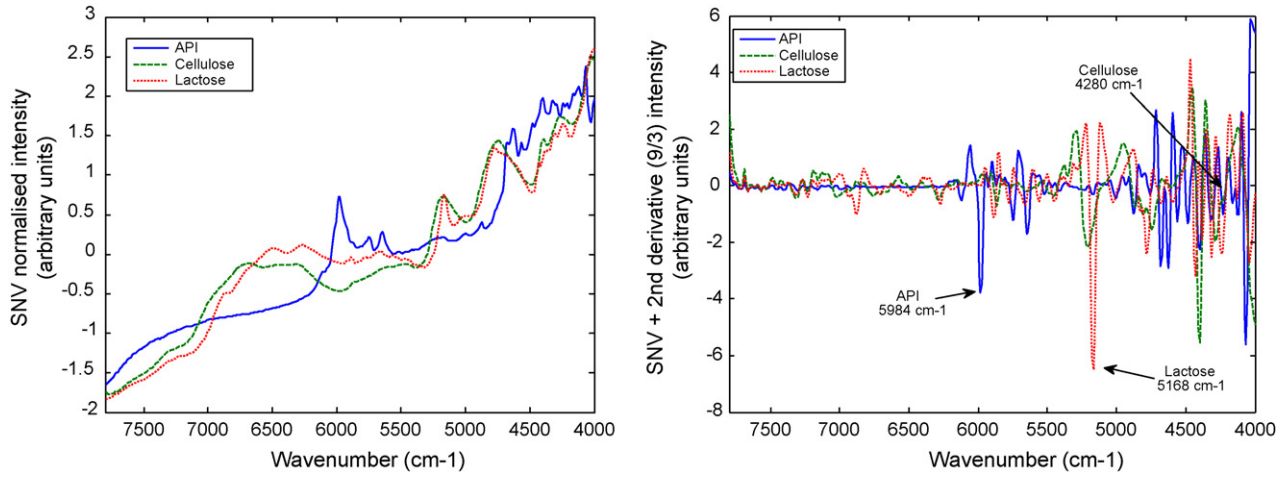

Fig. 1. Pure compound reference spectra of the three major compounds constituting the calibration batches. Spectra are preprocessed by SNV (left) and SNV followed by second derivative (9/3) (right). The arrows point out the wavenumbers used for the single wavenumber analysis.

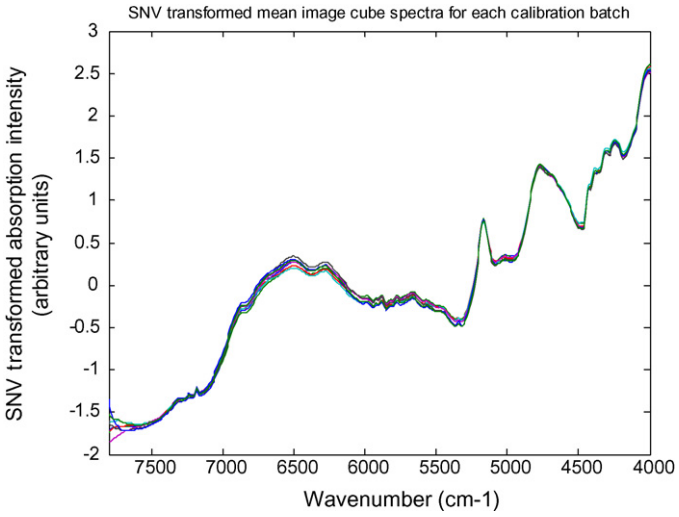

Fig. 2. Image cube mean spectra of the 9 calibration batches normalised by standard normal variate (SNV).

difficult to find a characteristic band for cellulose. Enlargement of spectral regions finally identified wavenumber at 4280 cm^{-1} for cellulose single wavenumber analysis although it was not as clear and well resolved as for the two other compounds and selection of other wavenumbers could be argued.

The CLS method used the pure compound spectra shown in Fig. 1 together with those of the two minor ingredients (not shown here). The PLS1 method used the mean spectra from each of the 9 calibration samples as matrix X (Fig. 2).

3.2. Prediction of concentrations

The concentration predictions of API, cellulose and lactose evaluated by correlation coefficient, slope and prediction error (RMSECV) are shown in Table 4. In general, reasonable prediction results (low RMSECV values) were obtained by all three methods and for each of the two preprocessing methods examined.

For all three calibration methods the predictions of API (RMSECV: 0.18–0.54%) were more accurate compared to predictions of the two excipients (RMSECV: 0.62–3.38%). This is not surprising as the API component showed a very distinct absorption peak with no large spectral interference from the other components in the pharmaceutical tablet (Fig. 1). The predictions of API for all three methods with the preprocessing giving the best predictions are shown in Fig. 3.

Excipient predictions were less accurate for API and with a generally poorer prediction of lactose compared to cellulose. This may be surprising as lactose had a more resolved and characteristic absorption peak than could be found for cellulose. The difference could probably be explained by the more narrow lactose concentration range between the different calibration batches (9 levels in steps of ~ 2%) compared to cellulose (3 levels in steps of 30%, cf. Table 2).

PLS1 was the superior regression method to predict concentrations for all three compounds. The best predictions were obtained for PLS1 using first derivative (9/3) followed by mean centering (Table 4, shaded area). For this PLS1 model API prediction was highly accurate with low prediction error (RMSECV = 0.18%) and correlation 0.99. The API prediction for the two other calibration methods showed less accurate but still reliable results with CLS predictions being slightly better than the single wavenumber method.

The better concentration predictions of cellulose compared to lactose were observed for CLS and single wavenumber but not for PLS1. The PLS1 predictions of the two major excipients were similar and quite accurate with RMSECV < 0.88% and correlation > 0.97. For CLS and single wavenumber the prediction error values and correlations followed API < cellulose < lactose.

For the PLS1 models the results illustrate that first derivative (9/3) preprocessing gave better prediction than second derivative (9/3) preprocessing both followed by mean centering. This may only be the case for this data set and it should also be noted that the results were only slightly better for the first derivative preprocessing.

3.3. Chemical images

The aim of this study was to evaluate the selected methods ability to generate accurate chemical images. As discussed in the previous section, PLS1 was the method that generated the most accurate overall predictions and therefore presumably the most accurate chemical images can be obtained using PLS. Fig. 4 shows the chemical images of API, cellulose and lactose for a tablet of batch 9 composition (cf. Table 2) analysed by single wavenumber, CLS and PLS1 methods using their best performing preprocessing treatment. The images show the distribution of the predicted concentrations of the three major compounds.

The chemical images of API appeared visually similar for the single wavenumber, CLS and PLS1 methods with only minor differences in the distribution information. Again, this can be explained by the distinct, well resolved absorption band for API which makes all three methods suitable for mapping API within the sample.

The single wavenumber method did not extract the distribution information of cellulose and lactose very well which was also

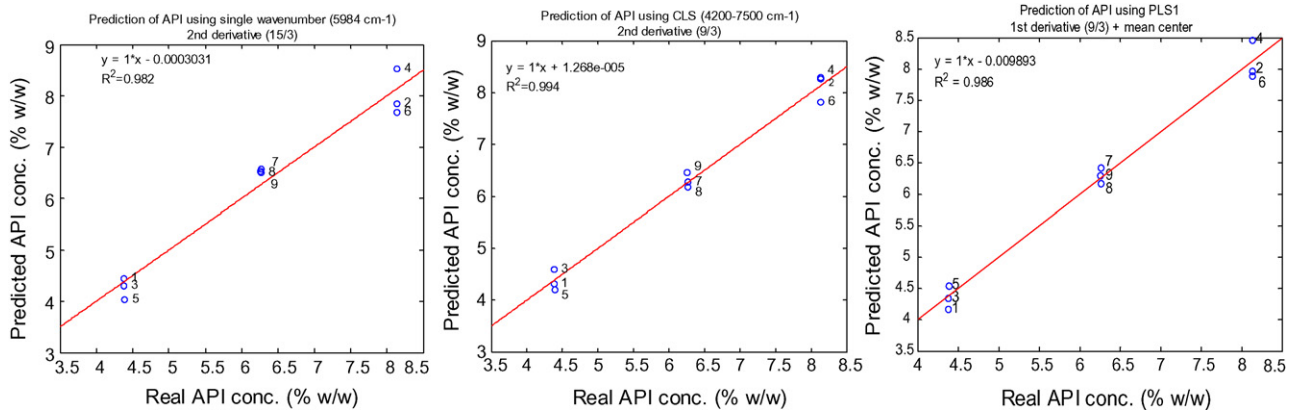

Fig. 3. Linear regression results of API concentration for single wavenumber method (left), CLS (middle) and PLS1 (right). Results are shown for the preprocessing methods giving the best regression and concentration predictions.

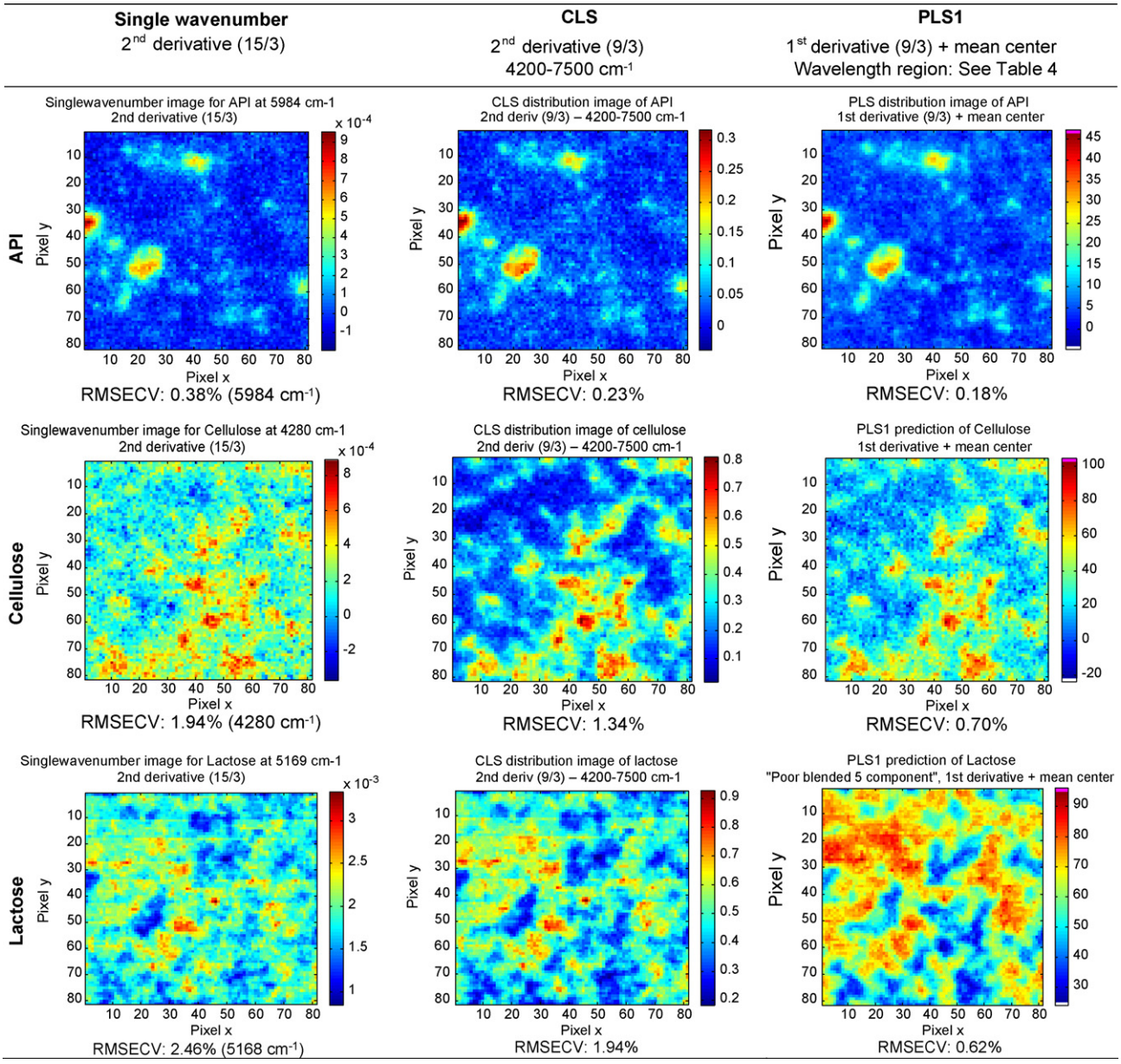

Fig. 4. Chemical images of a tablet of batch 9 composition (cf. Table 2) generated from data processing using single wavenumber, CLS or PLS1. The chemical images show the distribution of the three major compounds API, cellulose and lactose. The prediction errors (RMSECV) from Table 4 are shown beneath each image. High/low (red/blue) color intensities relates to high/low concentration of the component of interest. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

reflected in the higher RMSECV values. CLS and PLS1 provided comparable visual distribution information of cellulose and lactose but with differences in the image contrast. The highest contrast for the images of cellulose was obtained by CLS whereas for lactose it was by PLS1. In both instances the best concentration prediction, and thus presumably the most accurate chemical image, was obtained by PLS1. The observation that the best concentration prediction not necessarily produces images with the highest contrast is not unusual (and could, for example, also be shown for images by the other preprocessing methods used in this study but not presented here).

This also delineates one of the limitations of NIR chemical imaging. No tablet calibration standard exists for NIR chemical imaging and it is therefore not possible to set any objective criteria or confirm what the 'correct' chemical image is. Only with a calibration data set available is it possible to statistically evaluate the most accurate generated images as performed in this study. A calibration data set is a prerequisite and always available for PLS but not for the general use of the single wavenumber method and CLS (cf. Table 3). This is both the advantage and disadvantage of PLS. Data processing using PLS may generate the most accurate chemical images but it requires a calibration data set which is time consuming to produce and often not available (e.g. for a single trouble-shooting case or in early formulation development).

4. Conclusion

This study emphasizes the importance of data processing as part of a successful near-infrared chemical imaging analysis. Comparing a single wavenumber method, CLS and PLS1 by their ability to

predict API and excipient concentrations from hyperspectral data cubes of pharmaceutical solid dosage forms, PLS1 proved most accurate. This means that PLS1 can be assumed to provide the most accurate chemical images when using the model on single-pixel spectra. PLS1 is therefore also the preferred method when further image processing to extract process-related information from the chemical images is needed.

All three calibration approaches were found applicable for analysing hyperspectral data cubes and generate chemical images. But their use depends on the purpose of analysis, type of data set and the accuracy of the generated chemical images required. The single wavenumber method should primarily be used for initial exploration of compound distribution in a sample and it requires a distinct NIR absorption band for the compound of interest. CLS proved to be an excellent alternative to PLS1 generating only slightly less accurate concentration predictions. An advantage of CLS is that it is relatively fast as it only requires pure compound spectra of the sample constituents to perform the data processing. PLS1 is the method of choice when accurate concentration predictions are required but the disadvantage is that a calibration data set is needed which in many cases may not be available. Other calibration approaches to analyse hyperspectral image data cubes than investigated in this study may also be used with similar success. However, the three common calibration approaches presented here will cover a wide range of possible pharmaceutical samples and applications.

The wavenumber and preprocessing selections were also found to be an important part of data processing hyperspectral images. This study indicated the importance of a careful selection of both wavenumber range(s) and preprocessing treatment in order to obtain the most accurate results but a more thorough investigation of the issue is needed to fully understand the impact of these two factors.

The general principles of each calibration approach and the typical steps involved in a NIR-CI analysis is described in this work. This paper may therefore be used as practical guidance for analysing hyperspectral image data of pharmaceutical solid dosage forms.

This study demonstrates the usefulness of NIR chemical imaging when spatial distribution information of compounds in a solid dosage form is needed. But the message is also that care should be taken not to over-interpret the chemical images. Chemical images can be obtained by several different data processing methods but the obtained accuracy might be quite different as shown in this study. One should not be misled by images with high contrast as it is not high contrast but the right, i.e. accurate, contrast that is the goal. And the right contrast or accuracy of the chemical images can so far only be evaluated using a calibration data set as demonstrated here. Generating accurate chemical images is of high importance for the subsequent image processing analyses used to extract useful information and, for example, numerically describe the quality of the images. Different pharmaceutical conclusions may be drawn from chemical images of the same sample analysed by different processing methods having different accuracies as shown in this study. Developing image processing tools is not simple but is highly needed to further develop the technology of NIR chemical imaging.

References

[1] M. Blanco, J. Coello, H. Iturriage, S. Maspoch, C. de la Pezuela, Analyst 123 (1998) 135R–150R.
[2] E.W. Ciurczak, J.K. Drennen III (Eds.), Pharmaceutical and Medical Applications of Near-Infrared Spectroscopy, Marcel Dekker Inc., New York, 2002.
[3] G. Reich, Adv. Drug Deliv. Rev. 57 (2005) 1109–1143.
[4] E. Räsanen, N. Sandler, J. Pharm. Pharmacol. 59 (2007) 147–155.
[5] Y. Roggo, P. Chalus, L. Maurer, C.L. Martinez, A. Edmond, N. Jent, J. Pharm. Biomed. Anal. 44 (2007) 683–7000.
[6] U.S. Food and Drug Administration (FDA), Guidance for Industry, PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, FDA (CDER), 2004 (http://www.fda.gov/cder/OPS/PAT.htm).
[7] R.C. Lyon, D.S. Lester, E.N. Lewis, E. Lee, L.X. Yu, E.H. Jefferson, A.S. Hussain, AAPS Pharm.Sci.Tech. 3 (2002) 1–15.
[8] R.C. Lyon, E.H. Jefferson, C.D. Ellison, L.F. Buhse, J.A. Spencer, M.M. Nasr, A.S. Hussain, Am. Pharm. Rev. 6 (2003) 62–70.
[9] B.J. Westenberger, C.D. Ellison, A.S. Fussner, S. Jenney, R.E. Kolinski, T.G. Lipe, R.C. Lyon, T.W. Moore, L.K. Revelle, A.P. Smith, J.A. Spencer, K.D. Story, D.Y. Toler, A.M. Wokovich, L.F. Buhse, Int. J. Pharm. 306 (2005) 56–70.
[10] M.L. Hamad, C.D. Ellison, M.A. Khan, R.C. Lyon, J. Pharm. Sci. 96 (2007) 3390–3401.
[11] S.V. Hammond, Eur. Pharm. Rev. 3 (1998) 47–51.
[12] A.S. El-Hagrasy, H.R. Morris, F. D'Amico, R.A. Lodder, J.K. Drennen III, J. Pharm. Sci. 90 (2001) 1298–1307.
[13] E.N. Lewis, J.E. Carroll, F. Clarke, NIR News 12 (2001) 16–18.
[14] F.W. Koehler, E. Lee, L.H. Kidder, E.N. Lewis, Spectrosc. Eur. 14 (2002) 12–19.
[15] F. Clarke, S. Hammond, Eur. Pharm. Rev. 8 (2003) 41–50.
[16] E.N. Lewis, J. Schoppelrei, E. Lee, Spectroscopy 19 (2004) 26–36.
[17] E.N. Lewis, E. Lee, L.H. Kidder, Microsc. Today (2004) 8–12.
[18] F.C. Clarke, S.V. Hammond, R.D. Jee, A.C. Moffat, Appl. Spectrosc. 56 (2002) 1475–1483.
[19] F. Clarke, Vib. Spectrosc. 34 (2004) 25–35.
[20] F. LaPlant, Am. Pharm. Rev. 7 (2004) 16–24.
[21] J. Burger, P. Geladi, J. Near Infrared Spectrosc. 15 (2007) 29–37.
[22] S.J. Hudak, K. Haber, G. Sando, L.H. Kidder, E.N. Lewis, NIR News 18 (2007) 6–8.
[23] H. Ma, C.A. Anderson, J. Near Infrared Spectrosc. 15 (2007) 137–151.
[24] E.N. Lewis, J. Schoppelrei, E. Lee, L.H. Kidder, in: K.A. Bakeev (Ed.), Process Analytical Technology, Blackwell Publishing Ltd, Oxford, 2005, pp. 187–225.
[25] E. Lee, W.X. Huang, P. Chen, E.N. Lewis, V. Vivilecchia, Spectroscopy 21 (2006) 24–32.
[26] N. Jovanovic, A. Gerich, A. Bouchard, W. Jiskoot, Pharm. Res. 23 (2006) 2002–2013.
[27] C. Gendrin, Y. Roggo, C. Collet, Talanta 73 (2007) 733–741.
[28] E.N. Lewis, J. Dubois, L.H. Kidder, K.S. Haber, in: H.F. Grahn, P. Geladi (Eds.), Techniques and Applications of Hyperspectral Image Analysis, John Wiley & Sons Ltd, 2007, pp. 335–361.
[29] T. Furukawa, H. Sato, H. Shinzawa, I. Noda, S. Ochiai, Anal. Sci. 23 (2007) 871–876.
[30] C. Gendrin, Y. Roggo, C. Spiegel, C. Collet, Eur. J. Pharm. Biopharm. 68 (2008) 828–837.
[31] L.R. Hilden, C.J. Pommier, S.I.F. Badawy, E.M. Friedman, Int. J. Pharm. 353 (2008) 283–290.
[32] L. Weiyong, A. Woldu, R. Kelly, J. McCool, R. Bruce, H. Rasmussen, J. Cunningham, D. Winstead, Int. J. Pharm. 350 (2008) 369–373.
[33] H. Ma, C.A. Anderson, J. Pharm. Sci. 97 (2008) 3305–3320.
[34] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, S.E.J. Bell, Eur. J. Pharm. Biopharm. 69 (2008) 10–22.
[35] H. Martens, T. Naes, Multivariate Calibration, John Wiley & Sons, Chichester, 1989.
[36] S. Sasic, Appl. Spectrosc. 61 (2007) 239–250.
[37] Y. Roggo, N. Jent, A. Edmond, P. Chalus, M. Ulmschneider, Eur. J. Pharm. Biopharm. 61 (2005) 100–110.
[38] Y. Roggo, A. Edmond, P. Chalus, M. Ulmschneider, Anal. Chim. Acta 535 (2005) 79–87.
[39] Modde software version 8.0, Umetrics (www.umetrics.com).
[40] Mathworks Inc., Matlab version 7.2.0 R2006a (www.mathworks.com).
[41] Eigenvector Research Inc., PLS_Toolbox version 4.1.1 (www.eigenvector.com).
[42] J. Huang, H. Wium, K.B. Qvist, K.H. Esbensen, Chemom. Intell. Lab. Syst. 66 (2003) 141–158.
[43] S. Wold, E. Johansson, M. Cocchi, in: H. Kubinyi (Ed.), 3D-QSAR in Drug Design, Theory, Methods and Applications, Escom Science Publishers, Leiden, 1993, pp. 523–550.
[44] I.-G. Chong, C.-H. Jun, Chemom. Intell. Lab. Syst. 78 (2005) 103–112.
[45] M.W. Borer, X. Zhou, D.M. Hays, J.D. Hofer, K.C. White, J. Pharm. Biomed. Anal. 17 (1998) 641–650.
[46] A. Savitzky, M.J.E. Golay, Anal. Chem. 36 (1964) 1627–1639.
[47] R.J. Barnes, M.S. Dhanoa, S. Lister, Appl. Spectrosc. 43 (1989) 772–777.
[48] P. Geladi, D. MacDougall, H. Martens, Appl. Spectrosc. 39 (1985) 491–500.
[49] R. Larsen, J. Chemom. 16 (2002) 427–435.
[50] S. Wold, M. Sjöström, L. Eriksson, Chem. Intell. Lab. Syst. 58 (2001) 109–130.