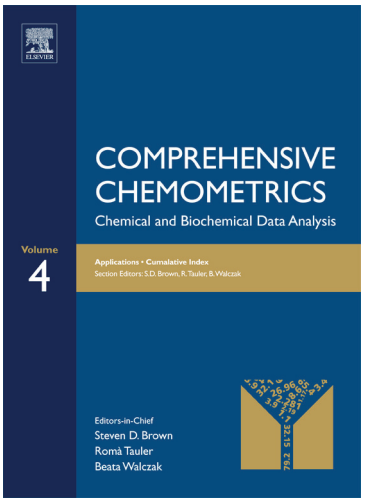
Provided for non-commercial research and educational use.
Not for reproduction, distribution or commercial use.

This article was originally published by Elsevier in *Comprehensive Chemometrics*, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

http://www.elsevier.com/locate/permissionusematerial

J. A. Fernandez Pierna, V. Baeten and P. Dardenne, J. Dubois, E. N. Lewis, and J. Burger 2009 Spectroscopic Imaging. In: Brown S, Tauler R, Walczak R (eds.) Comprehensive Chemometrics, volume 4, pp. 173-196 Oxford: Elsevier.

Author's personal copy

4.06 Spectroscopic Imaging

J. A. Fernandez Pierna, V. Baeten and P. Dardenne, Walloon Agricultural Research Centre (CRA-W), Gembloux, Belgium

J. Dubois and E. N. Lewis, Malvern Instruments Inc., Columbia, MD, USA

J. Burger, BurgerMetrics SIA, Jelgava, Latvia

© 2009 Elsevier B.V. All rights reserved.

Symbols

b_j	offset	$<w, x>$	inner product of w and x
C	penalty included in SVM that has to be added in order to take into account those samples that cannot properly separated	x, y	first and second dimensions in the hypercubes representing the spatial distances
$f(x)$	function representing the separating surface	x_i	training points
K	kernel function	y_i	reference or group value (± 1)
n	number of training data points	α_i	Lagrangian multipliers
t	score value	λ	third dimension in the hypercubes representing spectral variation (wavelengths)
w	weight vector	ξ_i	variables that measure the error made at point (X_i, Y_i)

4.06.1 Introduction

In the past decade, multispectral and hyperspectral chemical imaging have become powerful analytical approaches in several areas including quality control and troubleshooting of pharmaceutical products, remote sensing to tackle environmental, agricultural or mineralogy challenges, assessment of painting features, and safety evaluation and quality control of agrofood products.[1] In fact, these optical imaging techniques are increasingly considered the preferred tools in the design of nondestructive food/feed inspection instruments with applications in sample characterization, the measurement of chemical species distribution and the detection of contamination and defects in agrofood products,[2] and for a variety of applications for troubleshooting and quality assurance of pharmaceutical products. Of particular interest from a public health standpoint are instruments designed for multispectral or hyperspectral near-infrared (NIR) imaging analysis, which already play and will increasingly hold a key role for automatic food and feed inspection. For analysts and chemometricians, this is something of a revolution with hundreds or thousands of spectra (including tens or hundreds of variables) being collected for each sample, instead of the unique average spectrum typically

Author's personal copy

collected with classical spectroscopic instrumentation.[2] The challenge is to handle, extract, and exploit the relevant information contained in this large amount of data now available.

The aim of this chapter is to discuss the integration of chemometric processing tools for studying NIR imaging data. The first part of this chapter briefly introduces the instrumentation and some preprocessing requirements specific to imaging data. Some of the most popular chemometric tools used in NIR image analysis are then discussed, with a short introduction to a tool only recently added to the imaging spectroscopists' box. Finally, a number of applications of chemometrics in the development of NIR imaging analytical methods are explored, with a focus on agrofood and pharmaceutical products.

4.06.2 Introduction to NIR Imaging

NIR imaging instruments probe the spatially resolved chemical composition of samples. The spectral range probed by the instrument depends on its illumination source, wavelength selection mechanism and detector. Long-wavelength infrared imaging systems probe the combination, first, and second overtone bands of NH, CH, and OH bonds; intermediate-wavelength instruments measure the first, second, and third overtones; short-wavelength systems access only the third overtone. The chemical specificity of the spectral information, greater in the combination bands and gradually decreasing through the first, second, and third overtone bands, is the key to the value of this family of techniques aptly named chemical imaging. The recent success of NIR chemical imaging can be attributed to a combination of different factors: (1) nondestructive method, (2) high performance and availability of uncooled or Sterling-cooled NIR-sensitive two-dimensional array detectors, (3) digitally tunable infrared optical filters, and (4) drastic increase in both speed and capacity of laboratory computing platforms. The integration of these elements has already shown promising results in the determination of quality parameters for complex matrices such as pharmaceutical blends (e.g., Lyon *et al.*,[3] Reich[4]), food products such as the detection of apple surface defects (e.g., Lu[5]) and contaminations (e.g., Mehl *et al.*[6]), or for the feed industry (e.g., Fernández Pierna *et al.*[7] Fernández Pierna *et al.*[8]). Indeed, it differs from classical NIR spectroscopy (NIRS) in that the quality attributes can be assessed from both spatial and spectral (and therefore chemical) information. This makes the hyperspectral imaging technique better than NIRS in extracting details on a smaller scale with the objective of obtaining a better quality evaluation.

Hyperspectral images or hypercubes are three-dimensional data sets containing light intensity measurements where two dimensions (x and y) represent spatial distances, while a third dimension (λ) represents spectral variation such as wavelength (**Figure 1**). They can be interpreted as stacks of typically hundreds of two-dimensional spatial images at different wavelengths, or tens of thousands of spectra, aligned in rows and columns. Three instrumentation approaches are used to acquire hyperspectral images (**Figure 2**). These different approaches can be termed 'point', 'line', or 'plane' scan, based on the orientation of the scanning dimension relative to the two-dimensional spatial sample axes. These three acquisition modalities are described in detail later.

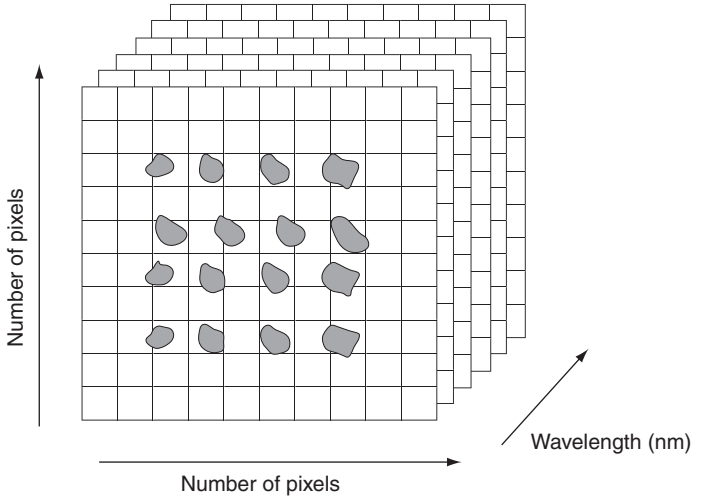


Figure 1 The three-dimensional hypercube.

Author's personal copy

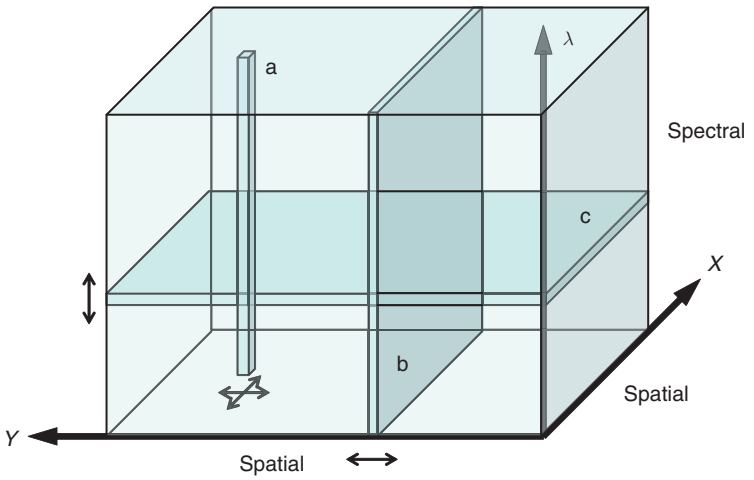


Figure 2 Modes of acquisition used for hyperspectral image data collection: a, single-point mapping; b, line scanning, and; c, global imaging.

Regardless of the scanning acquisition method, instruments may be multispectral and hyperspectral in nature. With multispectral instruments, only a limited number of wavelengths (below 10) are collected. Typically, these make use of a set of interference filters to measure the radiation of a fixed or predefined number of spectral bands. The band pass of the filters used may vary from a few tenths up to hundreds of nanometers. On the contrary, hyperspectral instruments allow collection of continuous spectral information from many contiguous or discrete wavelengths (typically >100 wavelength channels) within a specified spectral range.

4.06.3 Chemometrics and NIR Imaging

Traditionally, chemometrics are used with spectroscopic data to yield a measure of composition. The advent of imaging spectrometers adds a new dimension to the data set and it is imperative that the chemometric tools be designed to use this information as well. Structurally intricate samples are well suited for chemical imaging because of spatial and chemical complexity. Pharmaceutical products are a good example of chemically complex architectures and they have become one of the most prolific fields of use of NIR chemical imaging (e.g., Veronin and Youan,[9] Dubois *et al.*[10] Lee *et al.*[11] Lewis *et al.*[12]). Combining the spatial and chemical information brings about a need for different data-processing modalities in order to exploit the information that is present in the chemical imaging data set. In the following sections, the multistep chemometric approaches that attempt to include the spatial information present in the image as well as the chemistry to better describe and segment complex mixtures are presented and discussed.

The fundamental requirement for NIR chemical imaging experiments is that sample building blocks should possess different chemistries. A chemometric model is used to segment the image based on chemical parameters that are specific to that system, just like it would be used on individual spectra obtained from regular NIR spectroscopy. The contrast contained in the results of the chemometrics allows a segmentation of the sample into smaller parts that are characterized by their chemistry; these segments may further be measured and described with size and shape parameters such as their elongation, circularity, and convexity. The value of this combination of information, chemistry and morphology, is discussed later in the chapter with an example relating it with dissolution profiles of pharmaceutical granules. We also explore how the same type of processing applied to individual samples can provide valuable phenotypic information or insight into insect and rot occurrence in corn kernels.

The strength of chemical imaging resides in the availability of a massive amount of information in the data set; it can never be overemphasized that carefully designed data processing is required for accessing this information. Image-processing techniques can be applied to individual images measured at single-wavelength

Author's personal copy

channels (or combinations of channels) to obtain spatial information content such as feature recognition. Chemometric techniques can also be applied to the spectra from complete hyperspectral data sets as well as subsets of spectra measured within spatial regions of interest (ROIs). The first section of this chapter explores data conditioning or pretreatment and the application of some chemometric techniques (principal component analysis (PCA), partial least squares (PLS), support vector machine (SVM)) to aid in the extraction of information contained within hyperspectral images.

4.06.3.1 Image Acquisition and Data Conditioning

Spectroscopic data typically comprise variations in sample illumination and systematic throughput dependencies on wavelength that must be accounted and corrected for; this is especially true when working with spectroscopic imaging systems. Equally important are any variations due to spatial differences not linked to the chemical composition of the sample, for example, variations of physical origin such as density or glare, and they must be corrected for to focus the chemometric analysis on the chemical information. As introduced earlier, there are different approaches to collect hyperspectral image data. The process of converting instrument measurement signals to units of reflectance or absorbance depends on the type of instrumentation used for image acquisition.

A point scan instrument acquires a spectrum at a single spatial location using a Fourier transform or grating-type spectrometer. Hyperspectral images are obtained by successively measuring spectra while the sample is repositioned in the X and Y spatial dimensions. Assuming the lighting source remains fixed relative to the spectrometer, instrument calibration can be performed utilizing a one-dimensional (spectral) calibration model. This instrument configuration is often used in microscopy utilizing a high-precision X–Y motion stage.

For push-broom systems, which project a line of light onto a two-dimensional focal plane array (FPA), a two-dimensional calibration model (spatial–spectral) is needed to account for variation in sample illumination and instrument throughput. This instrumentation is best suited for remote sensing by aircraft or online process measurement since the Y spatial axis may be arbitrarily long.

Finally, plane scan imaging systems position the measurement camera parallel to the sample surface, obtaining X–Y spatial images with fixed sizes limited by the dimensions (pixels) of the camera detector. Hyperspectral images are obtained by modulating the radiation reaching the camera using band pass or tunable filters positioned in front of the camera. In this configuration, a complete three-dimensional (spatial–spatial–spectral) calibration model must be determined. This calibration typically uses only two points (dark and bright images), but calibrations with intermediate values have also been designed to account for system nonlinearity.[13]

Figure 3 shows a single-wavelength image obtained at 1430 nm of a mixture of grains containing soybean, barley, maize, and wheat. The NIR chemical imaging instrument used for this acquisition is a MatrixNIR system (Malvern Instruments Ltd.) based on an InGaAs FPA detector with 240×320 pixels (76 800 spectra per

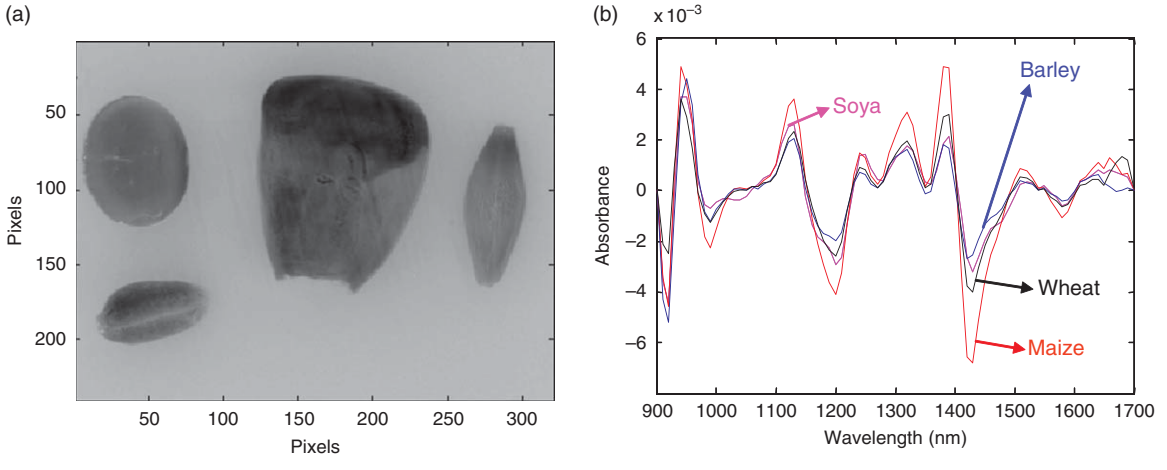


Figure 3 Image obtained by the NIR camera at 1430 nm of a mixture of grains containing soybean, barley, maize, and wheat.

Author's personal copy

scan) and a liquid crystal tunable filter (LCTF) for wavelength selection. The effective field of view (FOV) is variable and selection is made using microscope objectives.

The spectral imaging system was calibrated prior to further analysis with a dark image and a white (background) image collected from a standard white reference board. The reflectance was calculated for each pixel at each wavelength using the following equation:

$$R = \frac{\text{sample} - \text{dark}}{\text{back} - \text{dark}}$$

where sample = image of the sample, dark = dark current image, and back = reference image (background).

This is performed to remove noise and to compensate for offset (b_j) due to dark current, light source temperature drift, and spatial lighting nonuniformity across the scene.

The image was preprocessed by applying a second derivative using the Savitzky–Golay algorithm.[14] The mean spectrum of each grain was calculated following the application of a morphological mask obtained through a process of erosion in order to determine the contour of each grain. The resulting mean spectra from each kernel are shown in **Figure 3(b)**.

An additional issue common to all spectroscopic instrumentation is calibration transfer: How can spectral data sets obtained from one instrument be corrected to allow direct comparison with data sets collected on different days or with different instruments? Instrument standardization procedures employing known sample standards with unique spectral signatures carefully selected to span the spectral space of expected sample materials have been successfully applied to hyperspectral imaging data sets.[15–17] Such chemometric-based techniques determine correction models which 'stretch' the measured spectra of a slave instrument to match the corresponding spectra of a master instrument. The selection of NIR imaging standards is still a matter of personal preference and creativity; there are currently no official NIR chemical imaging standards available. Common NIR spectroscopic standards, such as NIST 2036, can be employed to verify wavelength accuracy, but these do not possess the spatial patterns required for spatial accuracy and resolution. Reflectance standard materials such as Spectralon can be used; however, these materials may appear inhomogeneous at high spatial magnification.

Spectral transferability can be achieved quite simply, thanks to the large spatial dimension of the image; small spatial areas within the sample image can be allocated to the reference sample materials necessary for instrument standardization. What was previously a time-consuming calibration step can be automatically incorporated into every hyperspectral image by including these standards *in situ* within the FOV of the image.

One final consideration in data conditioning is that an FPA detector may contain a small number of bad or 'dead' pixels originating from sensor elements that either fail to respond or respond erroneously. The type of imaging instrumentation setup implies a different impact of such bad sensor pixels on the data set and dictates different approaches to their detection and removal.

A bad sensor element in a push-broom line scan configuration affects only the part of the spectrum imaged onto that element. Removing this data value would require eliminating either a complete wavelength channel or the affected spatial channel from the entire hypercube. Such 'bad' data elements are routinely simply replaced instead, using median filters that examine the surrounding neighborhood of data values.

Bad pixels in a plane scan image system affect all values within a single spectrum. Correction steps depend on the type of analysis. In the case of chemometric processing of sets of spectra, bad spectra can be detected as outliers and removed. In the case of spatial image processing, bad pixels can be replaced with the median value of neighboring pixels. The impact of the number of bad pixels must be considered both as a function of their spatial distribution and a proportion to the total number of detectors. Clusters of tens of bad pixels greatly impact the image by removing all spatial resolution for the entire area of the cluster, while tens of isolated bad pixels in an array containing tens of thousands of pixels have relatively no impact.

How are bad data values detected? Simple thresholding will identify sensors stuck 'on' or 'off'. For example, with 12-bit digital data values ranging between 0 and 4095, data values < 100 or > 4000 may be suspect. Principal component analysis (PCA) score plots of image spectra will often reveal significant outliers where an individual sensor may give an erroneous result that otherwise lies within the acceptable data value range. For example, **Figure 4** shows $t_1 - t_2$ score plots (1) before and (2) after replacement of bad line scan data with neighborhood values. The data are taken from 160 000 reflectance spectra of a single cheese sample. Each line

Author's personal copy

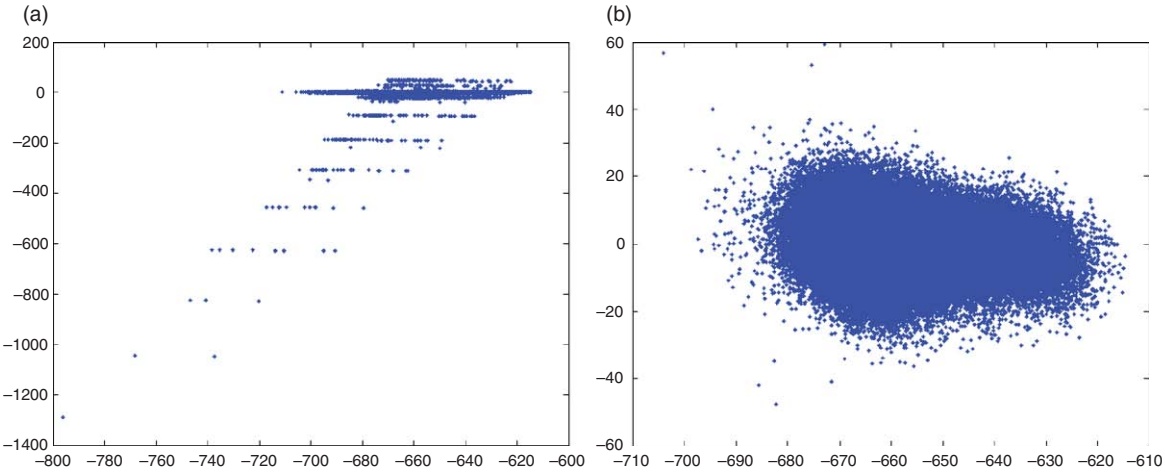


Figure 4 $t_1 - t_2$ score plots (a) before and (b) after replacement of bad line scan data with neighborhood values. The x-axis is the score on PC1 and the y-axis is the score on PC2.

scan image (320×256 pixels) contained 298 bad pixels. The strange clustering of points observed is due primarily to sets of spectra containing individual values near 0 or 4095. These extreme outliers have very significant leverage and must be detected and removed to obtain accurate PCA or PLS calibration models. Because hyperspectral images often contain tens or even hundreds of thousands of spectra, outlier spectra can be liberally removed while retaining quality spectra representative of the true sample nature.

4.06.3.2 Spectral Preprocessing

As with conventional spectroscopy, hyperspectral image reflectance spectra can be further transformed or filtered to remove unwanted nonlinear, additive, and multiplicative effects. Transformation to absorbance units estimated as the base 10 logarithm of reflectance, or application of the Kubelka–Munk transform[18] may help to 'linearize' diffuse reflectance spectra making them proportional to chemical constituent concentrations, which is frequently the objective of spectroscopic analysis.

Changes in sample orientation, particle size distributions, packing, instrumentation hardware, and/or analytical environment such as lamp intensity, temperature, or detector response may result in background signal that is added throughout the spectrum. The application of first and second derivative transforms[14,19–21] can compensate for constant additive effects.

Multiplicative scatter correction (MSC)[22,23] can be applied to correct for particle light scatter effects when sample preparation or data acquisition tools cannot be adapted to limit this effect. Since the scatter effect may not be exactly the same for all wavelength ranges, Isaksson and Kowalski[24] proposed correcting the spectral value at each wavelength with independent offset and slope correction terms. This technique is called piecewise multiplicative scatter correction (PMSC). Alternatively, the standard normal variate (SNV) transform was proposed by Barnes *et al.*[25] and has been shown to be equivalent to the MSC transform, differing only in scaling factors.[26]

Hyperspectral imaging offers unique opportunities to closely examine the interaction between preprocessing treatments, instrumentation, and sample problems. Fundamental particle scattering effects can be explored, for instance, by examining the large populations of spectra contained within individual hyperspectral images. Large population statistics and data visualization tools can be employed to compare the effectiveness of different preprocessing treatments.

A series of hyperspectral NIR images (960–1662 nm) were acquired of ordinary table salt and sugar, separated into eight different particle size fractions ranging between >400 and $<63\,\mu m$. From each image 250 spectra were randomly selected from the two sets of images (salt and sugar), thereby providing a combined spectral data set containing 4000 spectra. PCA was applied to mean-centered data, following different spectral preprocessing treatments. The results are shown in **Figure 5(a)**, where the plus symbols represent sugar and the

Author's personal copy

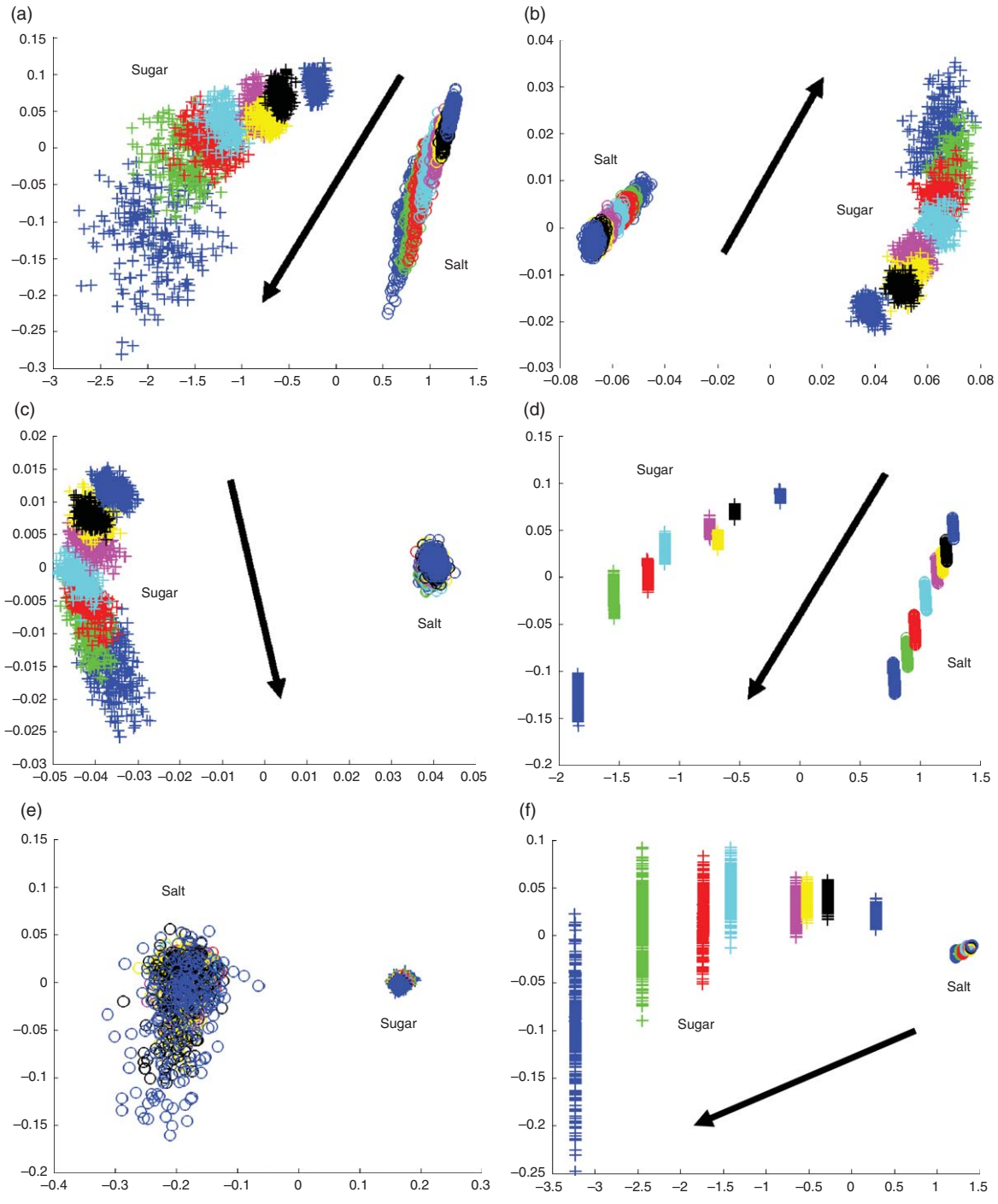


Figure 5 PCA applied to salt and sugar mixture. The plus symbol represents sugar and the circle represents salt. An arrow has been added to indicate the direction from small to large particle size fractions. (a) Absorbance, (b) first derivative, (c) second derivative, (d) absorbance + image MSC, (e) absorbance + global PMSC, and (f) Kubelka–Monk + image MSC. In all plots above, the x-axis is the score on PC1 and the y-axis is the score on PC2.

circles represent salt. The arrow indicates the direction from small to large particle size fractions. Each size fraction has been color coded as well.

Sugar has a much stronger NIR absorbance than salt; consequently, the scattering effects are greater. This is indicated in the first plot for absorbance spectra (**Figure 5(a)**). The variation in score values for each particle

Author's personal copy

size fraction is much greater for the sugar samples, and increases with particle size. The first derivative transform (Figure 5(b)) reduces the variance in sugar relative to that in salt; however, the trend in particle size group clusters is still clearly evident. The second derivative transformation (Figure 5(c)) results in a similar reduction in relative variance in the sugar samples.

One approach to MSC corrections was based on first computing individual mean spectra for each of the eight size fractions of both sugar and salt samples. These 16 spectra then became the target spectra for the respective size fraction MSC corrections. As indicated in Figure 5(d), this technique greatly reduces the within-fraction variance – each cluster is significantly tighter when compared to the untreated absorbance (Figure 5(a)). The general trend between different size fractions is the same. Figure 5(e) indicates the results when a single target spectrum is used for each salt and sugar (the mean of the eight respective size fraction spectra) and piecewise MSC is applied. Here, the variation in sugar spectra has been dramatically reduced. Compared with sugar, the salt has very little absorbance – consequently, the piecewise MSC models a much higher noise contribution indicated by the larger relative cluster sizes. The overall trend in particle size dependency has been reduced for both salt and sugar.

To indicate the exploratory nature of this process, Figure 5(f) shows the results of applying MSC with targets based on particle size fractions as in Figure 5(d), but to spectra first transformed using the Kubelka–Munk transformation on reflectance spectra. As with the second derivative transform (Figure 5(c)), the size dependency of the salt fractions has been nearly completely removed. These figures indicate some of the possible ways of combining exploratory analysis with preprocessing and PCA to examine the group populations of spectra obtained with hyperspectral imaging. Such exploratory analysis of class populations is realistically possible only with the immense number of spectra available from hyperspectral images.

4.06.3.3 Unsupervised Data Analysis of Imaging Data

In pattern classification problems (e.g., discrimination of incoming batches, authentication of food/feed products, or detection of contaminants), all the variables used to describe the data may not be equally distinctive and informative. In terms of hyperspectral imaging data, it means that it is not difficult to get overwhelmed by the high dimensionality of the data set, which leads to the so-called curse of dimensionality. For this reason, optimal variable selection and variable combination methods are important topics in these fields.

In chemometrics, the most widely used method for exploratory analysis is the PCA.[27,28] PCA is simple, easy to use, and amply discussed, especially for its use as a pattern recognition and data compression method for signal processing. In the case of image processing, volume reduction is a common objective of the PCA. PCA is a way of identifying patterns in data by reducing the number of dimensions without much loss of information. In other words, PCA is a linear transformation that tries to reduce the dimensionality. This is performed in such a way that the reduced number of dimensions captures most of the informative variance of the data. The new variables are called principal components (PCs) and correspond to the largest eigenvalues of the covariance matrix that account for the largest possible variance in the data. The theory concerning PCA has been explained elsewhere in this book. This section will focus only on its application to classical imaging and NIR imaging.

Applied on digital photographs, PCA performs a coordinate transformation of the color image represented in the space of fundamental colors (red, green, and blue, known as RGB). After this transformation, the new axes are the largest eigenvalues of the covariance matrix of the three input images. Then three new images are obtained by projecting the RGB axes onto the three resulting axes. The first axis corresponds to the largest eigenvalue. The two other axes, which must be orthogonal to the previous axis, are linear combinations of the input images that lead to the remaining variability or information that is not correlated to the first axis or to any other PC of higher order.

Figure 6(a) represents a digital photograph of size $M \times N \times 3$ of a worldwide known Belgian scene. The picture includes the information about intensity of color components stored in three planes. These R, G, and B components are represented in Figure 6(b). Figure 6(c) represents the pseudocolor maps of the three reconstructed components for each eigenvalue and of the PCA residuals. The first PC contains most of the

Author's personal copy

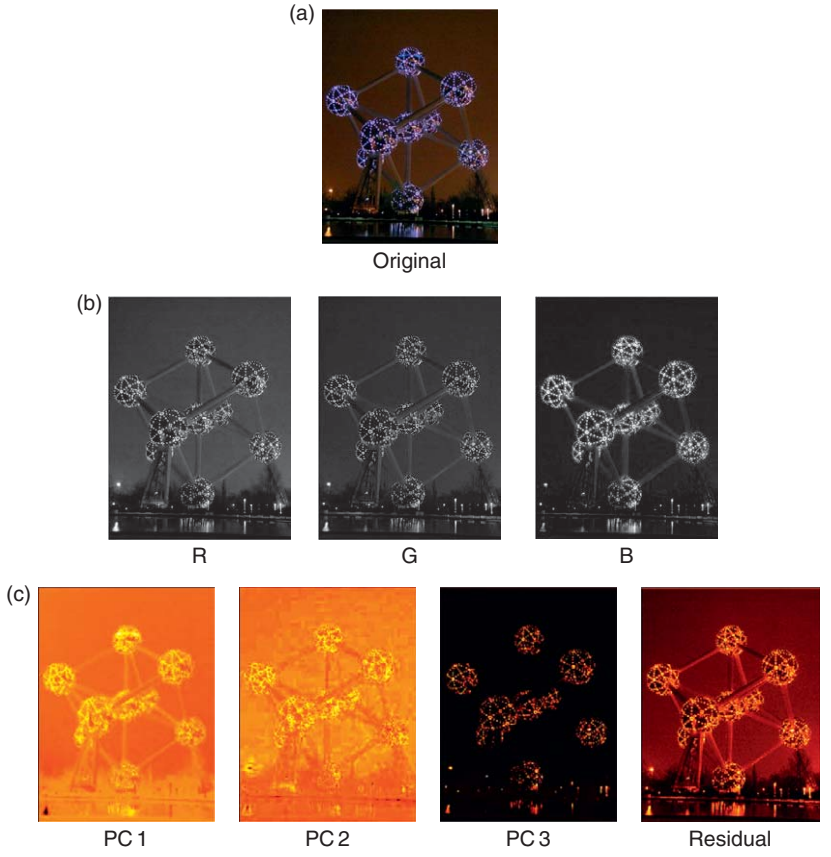


Figure 6 A worldwide known Belgian scene: (a) digital image, (b) R, G, and B components, (c) pseudocolor maps of the three reconstructed components for each eigenvalue and of the PCA residuals.

intensity information. The other two PCs have a lower signal-to-noise ratio and they would require a filter in order to improve the visualization; this is clear when looking at PC 2.

Owing to the recent availability of imaging spectrometers with high spectral resolution, hyperspectral image compression has become increasingly important.[29] In these cases, the hyperspectral image cube can be considered as a high-dimensional feature space where each feature is represented as a spectral image. The (reduced) dimensions obtained after PCA are defined as the ones that preserve the most information among the hyperspectral data cube, that is, the ones that present the best representation of the original data.[30] Examples of the use of PCA in chemical image analysis of agrofood and pharmaceutical products are presented later in the chapter.

4.06.3.4 Supervised Data Analysis of Imaging Data

In order to locate features and to extract and analyze the information from spectral image data, a combination of image-processing and chemometric techniques may be applied. Classical chemometric methods, such as PLS[31] or artificial neural networks (ANNs),[32] are well-known, proven techniques for both classification and regression analysis of multivariate data, such as NIR spectra. New chemometric techniques such as SVMs are being introduced in the hyperspectral imaging treatment. The use of SVM as a chemometric tool is recent, and its application to spectroscopic data has been proposed by Cogdill and Dardenne[33] among others.[34,35] The use of SVM in the analysis of hyperspectral imaging data is still sparse and the focus is on discrimination problems.[7–8,36]

Author's personal copy

Because the main theory about SVM has been explained elsewhere in this book, only the main ideas are presented here, with examples in Section 4.06.4.

In brief, the main feature of SVM is that this method is able to construct a mathematical hyperplane ($f(x) = wx + b$) based on a high-dimensional space to discriminate between groups. The problem of calculating an SVM model can be solved by finding an optimal solution for the quadratic programming problem:

$$\min\left(\frac{1}{2}||w||^2 + C\sum_{i=1}^{n}\xi_i\right) \rightarrow \text{subject to} \begin{array}{l} y_i(w^{\mathrm{T}}x_i) + b \geq 1 - \xi_i \\ \xi_i \geq 0, \quad \text{for} i = 1, \ldots, n \end{array} \tag{1}$$

where n is the number of training data points, y_i is the reference or group value (± 1 for training points x_i), C is a penalty that has to be added in order to take into account those samples that cannot be properly separated, and ξ_i represents the variables that measure the error made at point (x_i, y_i).

This equation can be solved using the method of the Lagrangian multipliers (α_i) and the support vectors are the points for which α_i is positive and correspond to the nearest points to the boundary. In order to solve the problem of the dimensionality, the kernel K function is introduced. The K function maps the data into some dot product space (feature space) through a nonlinear map and in such a way that it constructs an optimal hyperplane in a high-dimensional space and is returned to the original space as a nonlinear decision frontier.

Finally, the separating surface $f(x)$ that maximizes the distance between the two groups of cases is found of the form:

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x, x_i) + b \tag{2}$$

The kernel function, as well as the associated parameters, has to be defined prior to applying SVM.

4.06.4 Example Applications

4.06.4.1 Hyperspectral Imaging of Agrofood Products

The following examples of NIR images aim to demonstrate the potential of this technique to assess damaged kernels and identify kernel types. **Figure 7** shows a NIR chemical image at 1430 nm following processing with a second derivative using the Savitzky–Golay algorithm (third polynomial degree and a gap of 15 points). The image shows three maize grains with different characteristics: The grain at the left part of the image corresponds to a damaged grain containing an insect (weevil larvae), the grain in the middle is a damaged empty grain, and the grain at the right is a healthy grain. An enlargement of the image at the left that contains the grain with the insect allows to study the spectral difference between the 'healthy' part and the damaged part (containing the insect) of the grain.

Figure 8 shows the PCA results for another example of two maize grains analyzed by NIR imaging. The left grain is a healthy one and the right grain a damaged one. In the middle is placed an insect. From the PCA analysis on the multispectral images, it has been found that mainly the first PC score of images, which explains $> 71\%$ of the variance, is suitable for identifying the presence of the insect. This is also the case when looking at the residual matrix obtained after reconstruction using the three first PCs, which explain $> 89.95\%$.

NIR imaging and PCA can also be used for the detection of the presence of rot in cereals. **Figure 9** shows an example of this problematic with two grains of maize, one of them being rotted (the left part of the image) and the other being healthy. This figure shows the images taken at two different wavelengths: At 1160 nm, the difference between the two grains is enhanced as it is also shown in the spectra, and at 1560 nm, the rotten part of the grain becomes clearer. PCA, and mainly the seventh PC, is also useful in order to put in evidence the rotten part of a grain mainly because of the large loading at 1560 nm as indicated in **Figure 10**.

The examples of the use of PCA to handle visible and NIR imaging data are also reported in several papers. For instance, in a work aiming to study the ripeness of tomatoes, Polder *et al.*[37–39] used PCA and Fisher's linear discriminant analysis (LDA; a classical classification method) in order to visualize the data and calibrate the instrument, respectively. The technique, based on hyperspectral NIR imaging to determine the maturity stage

Author's personal copy

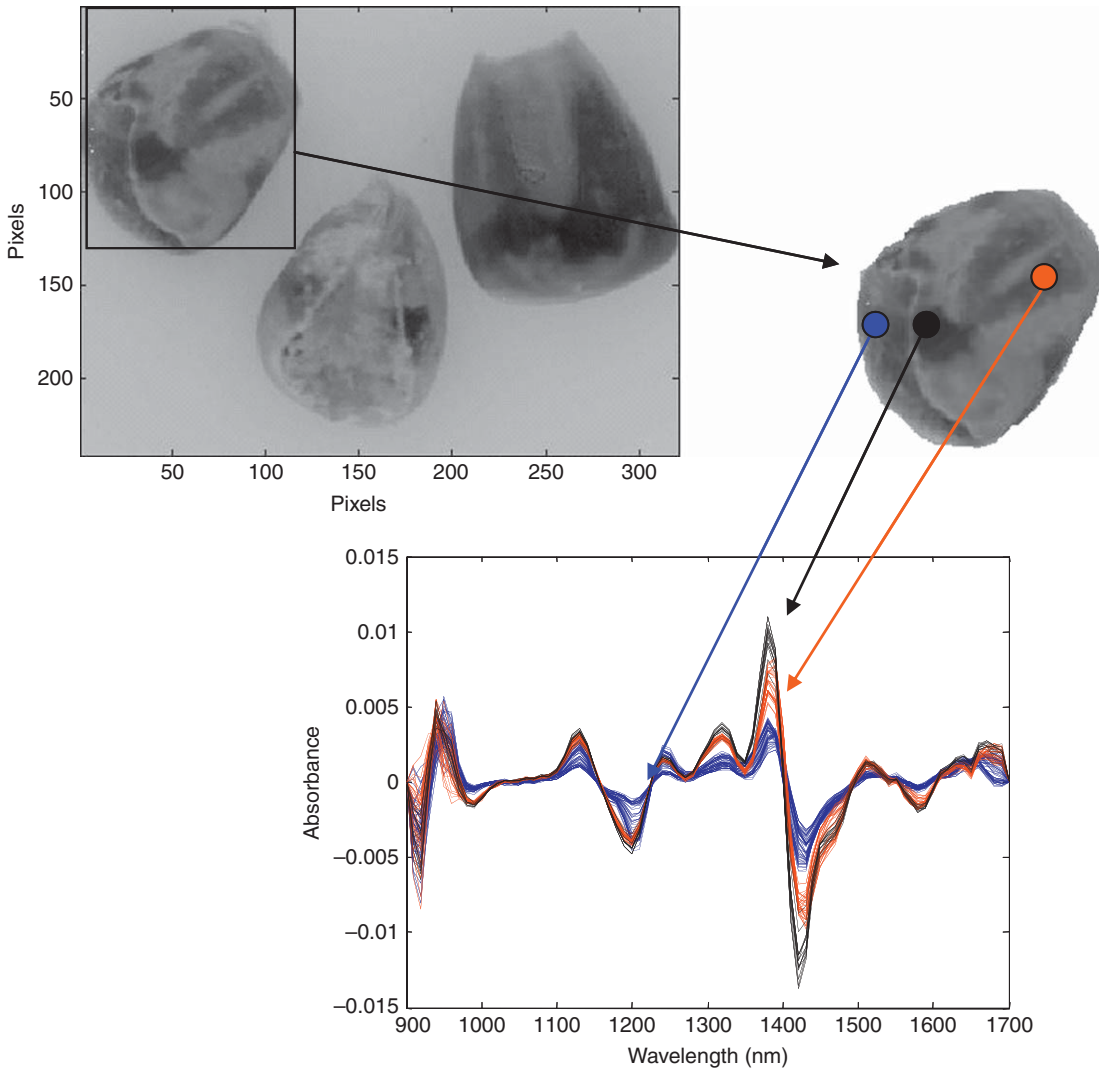


Figure 7 NIR chemical image at 1430 nm of three maize grains with different characteristics.

of preclimacteric apples,[40] applied PCA in order to distinguish the starch concentrations within one apple and among several apples during maturation. Lawrence *et al.*[41] demonstrated the usefulness of PCA for distinguishing the contaminants inside poultry carcasses. In their study to detect poultry skin tumors, Chao *et al.*[42] used PCA to select useful bands for detecting tumorous regions. Mehl *et al.*[43] demonstrate that hyperspectral imaging system allowed them to determine scabs, fungal, and soil contamination, and bruising using either PCA or the absorption intensities at a specific frequency.

NIR chemical imaging has shown the potential to provide both speed for high-throughput analysis of large numbers of kernels for quality assessment and a high content of chemical information – two important criteria for plant breeding studies (e.g., Smail *et al.*[44] Weinstock *et al.*[45]). The selection of a larger FOV for imaging multiple kernels to extract average measurements per unit described previously highlights the value of the spectroscopic image for high-throughput analysis. If kernels are spectroscopically imaged using a smaller FOV, that is., at greater magnification, the data acquired may contain a breath of additional valuable information, which may help establish phenotypic differences. The previous examples made this point clear. In this next example, a corn kernel was sequentially sectioned and imaged plane by plane in the NIR spectral range 1150–1750 nm for 43 times. Each image is 250 μm deeper into the kernel than the previous one. It is noteworthy

Author's personal copy

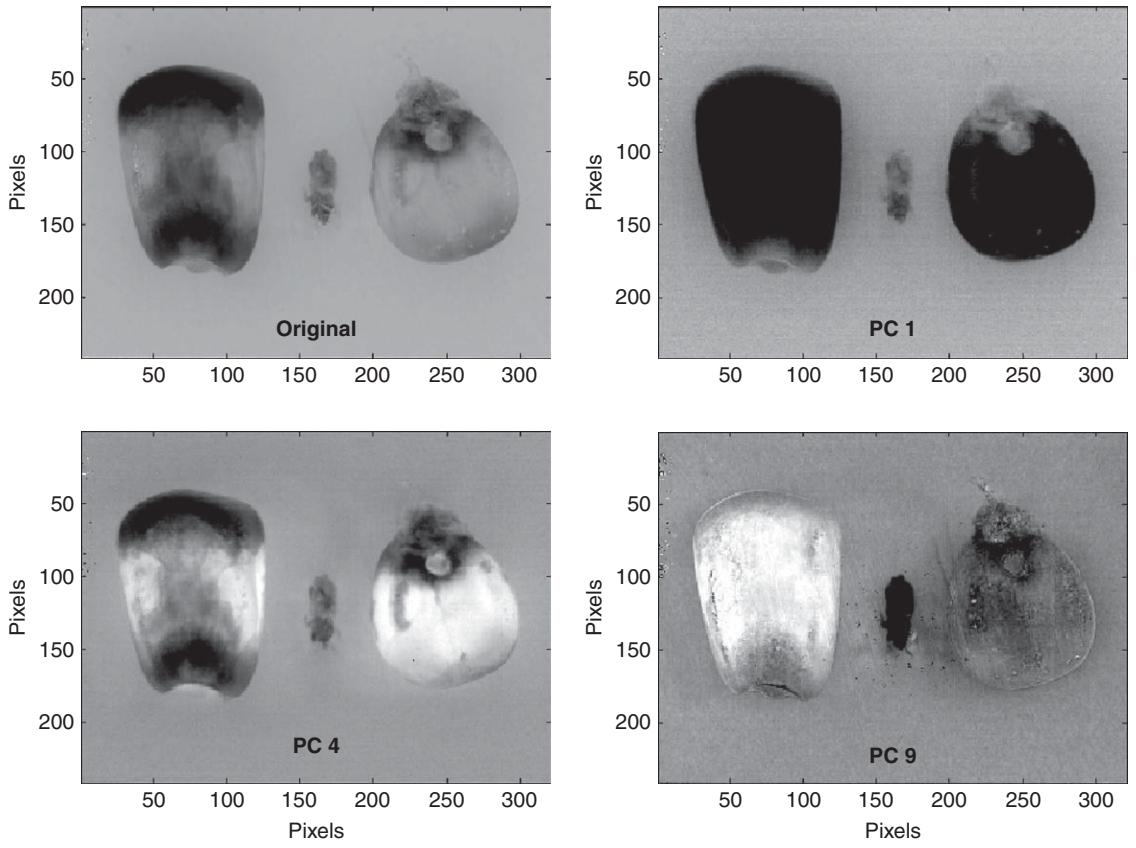


Figure 8 PCA results for two maize grains analyzed by NIR imaging.

that a single image acquired from a cross-section of the core of the kernel could suffice to access a large amount of phenotypic information. The interest of the multiple slices is the possibility to render the final chemically relevant results in three spatial dimensions, which allows volume calculations in addition to surface measurements.

Data preprocessing consists of conversion to absorbance following subtraction of the dark image and normalization by mean centering and scaling to unit variance by spectrum to reduce the effects of physical parameters such as density. Data analysis is limited to a PCA, where the fourth PC was found to display good discrimination between the various known structures of the kernel. **Figure 11(a)** shows a gray-scale image of the scores of PC 4 at all pixels in the image and the histogram plot of the distribution of the scores. Histogram-based segmentation was then performed using a curve-fitting algorithm. The graph in **Figure 11(b)** clearly indicates three main fractions in the distribution, and histogram fitting further reveals that the blue area in the image represents 23% of the area (i.e., of the pixels) of the kernel. Fine features of the root and coleorhiza are also distinguished and highlighted in red, quantitatively representing about 0.4% of the area of the kernel. The two colored regions in the image are delimited by segmentation of the image using the chemical response, obtained by the calculation of a PCA. In this particular example, the segmented images of all slices are recombined and rendered in three dimensions to create the image shown in **Figure 11(c)**.

A method for the detection of animal proteins in compound feeds was developed using NIR chemical imaging with SVM data processing.[7,8] A large spectral library that groups spectra representative of the wide diversity of ingredients usually present in the preparation of compound feeds was gathered.[35] The objective was to investigate the limit of detection, specificity, and reproducibility of the NIR imaging method using SVM for the detection and quantification of animal ingredient in feed. In brief, an SVM discriminant model was applied to a single spectral image cube collected from animal sediment (bones) particles and nonanimal sediment positioned separately on the same surface. **Figure 12** shows the

Author's personal copy

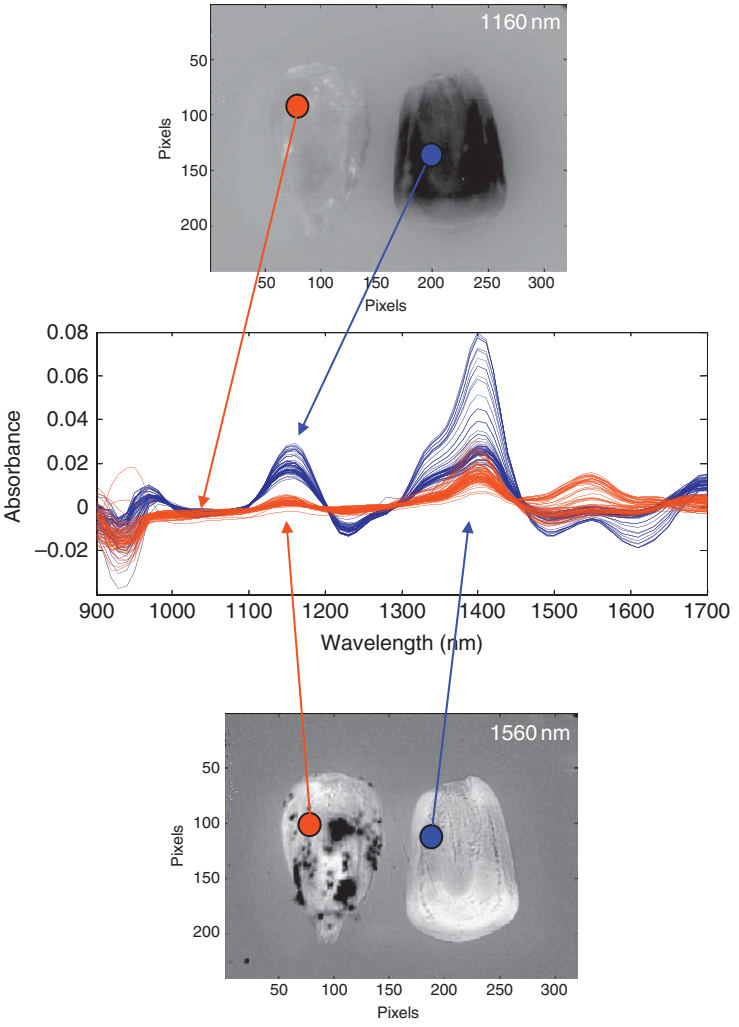


Figure 9 Example of two grains of maize: one of them being rotted (at the left part of the image) and the other being healthy.

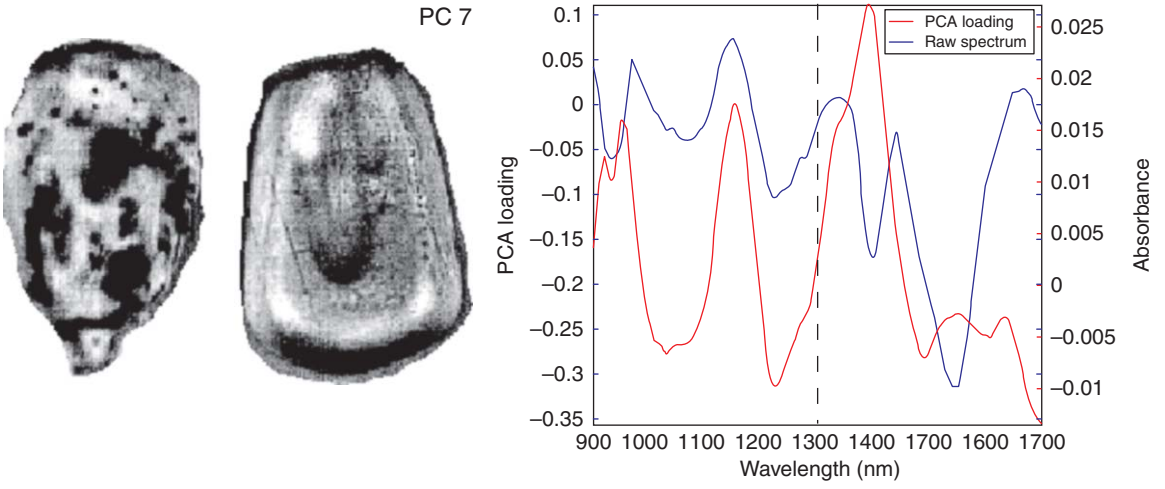


Figure 10 The seventh principal component of **Figure 9**.

Author's personal copy

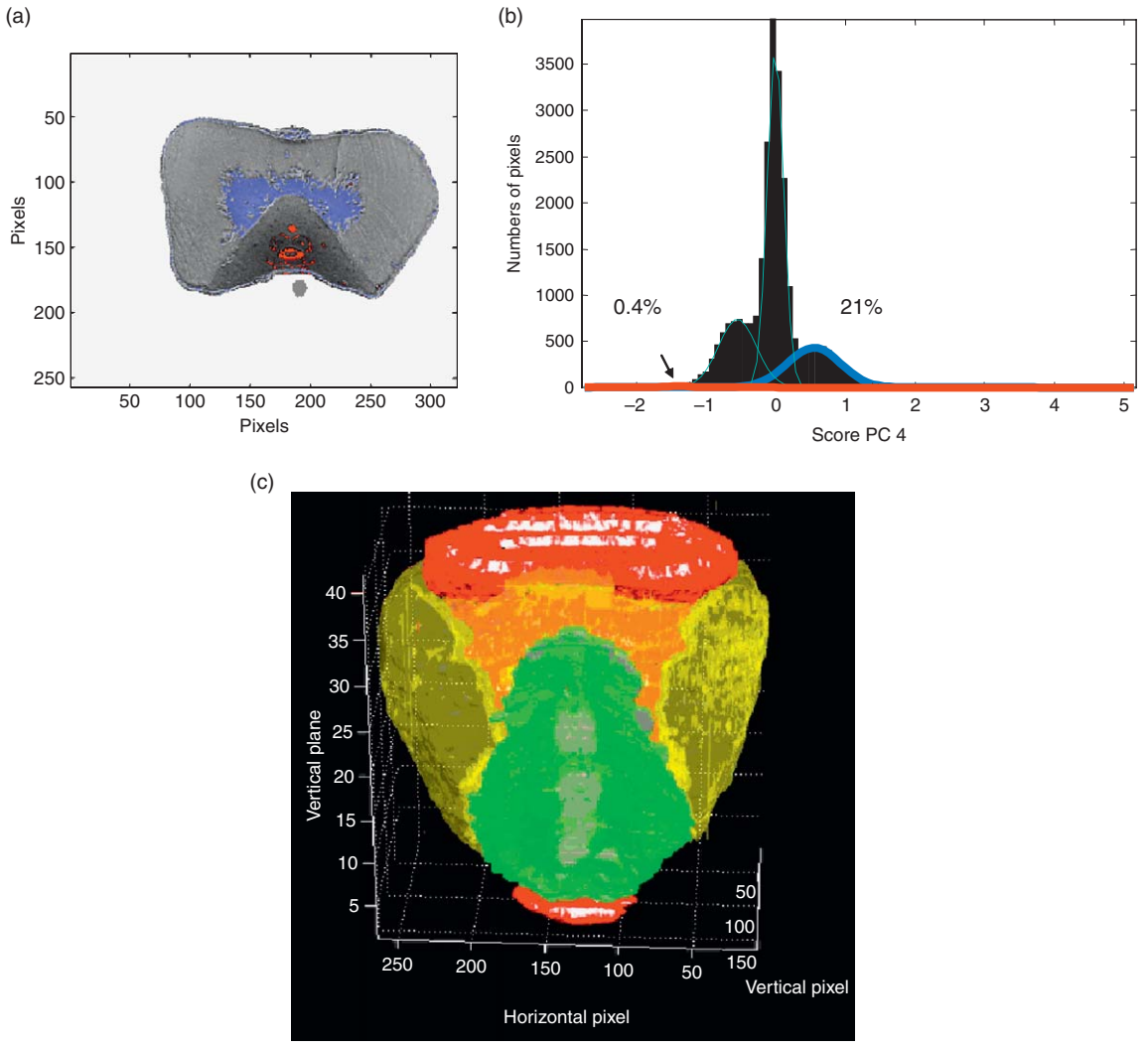


Figure 11 (a) False-color NIR chemical image of the central slice in the corn kernel. (b) Histogram plot of the distribution of score for PC 4. (c) Three-dimensional false-color rendition of the structure of the corn kernel based on the PCA of NIR chemical images of consecutive slices through the kernel.

discrimination results for this model. Animal sediment material (121 particles) is shown on the left side of the image and nonanimal sediment material (62 particles) on the right side. The results indicate that 101 of the 121 animal particles were correctly detected as animal material, while 4 of the 62 nonanimal particles were detected as being of animal origin. A visual inspection of the spectra of these four particles confirmed that they were false-positive results. This wrong classification by the model was mainly due to the fact that the discriminant models were constructed with nonsediment materials and applied here on sediment fraction. However, the results indicate that the discrimination of animal particles in the sediment fraction is possible using SVM.

SVM was also used in the development of an automatic system for pollen identification based on its texture classification.[36] In this study, a texture feature extractor computes image properties on selected ROIs; these texture features are then used for pollen load classification. In a comparison of the classification ability of *k*-NN (*k*-nearest neighbors), MLP (multilayer perceptron), and SVM using optical microscopy, SVM showed a 76% classification rate for the discrimination of the different geographical origin compared to 69% for MLP and 67% for *k*-NN.

Author's personal copy

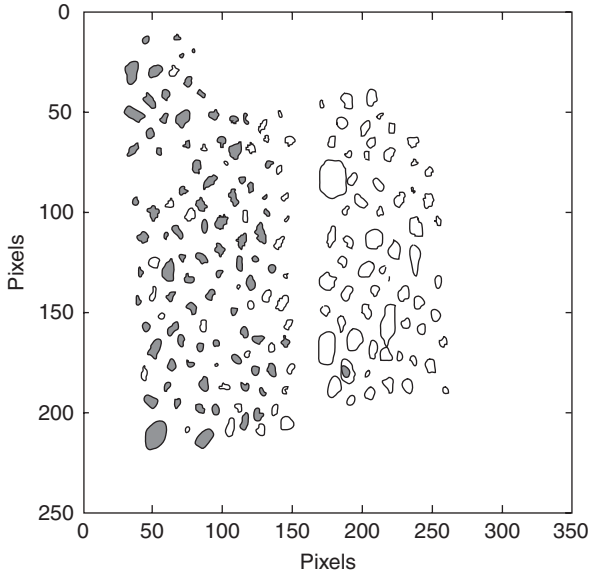


Figure 12 Result of the SVM discriminant model on a series of particles. Gray particles are the ones classified as animal particles.

Several studies have shown that hyperspectral imaging in combination with PLS and ANN present an important opportunity for rapid chemical imaging of biological and agricultural products (for more information, see dedicated papers and books, e.g., Martens and Naes[31]). Several studies performed on that topic concern mainly quality parameter determination in the case of single kernel analysis. Cogdill *et al.*[46] have used hyperspectral imaging spectroscopy for the study of single maize kernel analysis. After comparison of different spectral preprocessing methods, they have developed predictive calibrations for moisture and oil content using PLS and principal component regression (PCR). The most accurate results were obtained for moisture calibration using PLS on the raw data. The results for the oil content calibration were not as accurate as the moisture calibration mainly due to the reference method rather than the spectrometer. In their paper, Stevermer *et al.*,[47] have designed an automated single kernel analysis and sorting system that allows them to construct models for protein determination. The performance of this system has been tested by predicting constituents using PLS and comparing the predicted values with the reference values. Gorretta *et al.*[48] proposed a combination of a hyperspectral system with partial least squares discriminant analysis (PLS-DA) for the classification of durum wheat kernels according to their vitreousness in order to create an automatic method to replace the visual method stipulated by the European Union regulations. They obtained a classification rate of up to 94% when separating two nonvitreous classes, and a 100% separation when separating vitreous and nonvitreous kernels.

4.06.4.2 Hyperspectral Imaging of Pharmaceutical Products

Pharmaceutical solid dosage forms are ideal candidates for NIR chemical imaging, because they are chemically complex systems where the distribution of the chemistry affects product quality and performance. The size scales of the architectural blocks match the spatial resolution achievable with NIR diffuse reflectance measurements, thereby permitting nondestructive analysis of samples for which other tests must be performed. The increasing popularity and broadening of application of the technique in this area is due to the access to information that was previously hidden about the sample. Indeed, much detail is known about the chemistry and morphology of pure ingredients and this information is valuable in the implementation of quality by design strategies, but only bulk analysis information, such as content and dissolution profile, was measurable in the finished product. The following paragraphs describe an example of the new information accessed with NIR chemical imaging measurements and their effect on product quality and performance.

Author's personal copy

In this example, three lots of pharmaceutical granules are used to illustrate the handling of hyperspectral images for the assessment of compliance to bioavailability specifications as typically measured by dissolution testing. These granules have been produced according to a single formulation, but yield different dissolution profiles. The focus of this analysis is to find discriminant parameters that explain the differences in dissolution, as well as to ensure that an unrelated product (which may, e.g., be a counterfeit product that does not contain the active pharmaceutical ingredient of the true product) is flagged. For such tasks, individual measurements of chemical composition,[10] size, and shape can be employed, but we describe how a combination of these parameters can be used to better answer the question by gathering all this information from a single modality, namely, near-infrared chemical imaging (NIR-CI), and the chemometrics used to extract the information from the data cube.

Chemical images were collected on a Sapphire™ NIR-CI instrument (Malvern Instruments Ltd., Malvern, UK) using a magnification of 88 μm per pixel. In brief, the Sapphire consists in a two-dimensional infrared FPA detector (320×256 pixels), broad band illumination, a wavelength filtering unit, and diffractive optics. Diffuse reflectance images were acquired in the spectral range 1200–2400 nm with a 10-nm step. Each resulting image data cube is of dimension 320×256×121, thereby containing 9 912 320 data points. The granules were deposited on a stainless-steel mesh resting on a mirror made of the same material, and images of discreet sections were acquired, for a total of three images per lot. In total, nine images of granules and one background and one dark image were acquired, in a total collection time of 33 min.

All data processing was performed using ISys 4.1 (Malvern Instruments). PCA was applied to mean-centered and scaled spectral images. PCA, as explained previously in this chapter, is well suited for exploratory work in spectroscopic imaging data because it requires no *a priori* knowledge about the samples. In the particular case of pharmaceutical products, PLS analysis often offers better performance and is accessible because the ingredients are generally available to create a library. In this example, an unsupervised analysis is sufficiently discriminating to achieve the chemical separation needed in the exploration of the causes of dissolution failures. The result of the PCA prediction is a series of scores for each pixel in the image, regardless of which granule the pixels belong to. It is therefore an unbiased measurement of the chemistry, that is, one that is not influenced by the shape or size of the individual granules. A data set combining granules from a passing lot and from a failing lot of the common formulation was created by concatenation and used to calculate the loading vectors. The loading vectors are saved and used to calculate score images for the other sample images.

Shape and size measurements are performed on binary images obtained through intensity threshold at 1990 nm, where pixels with a value greater than the threshold are set to 0 and the others set to 1. There is little specific NIR absorption at this wavelength in these samples; the intensity measured in pixels that do not contain granules is very high because the metallic mesh and support do not produce diffuse reflectance and, therefore, create an absorbance value that is > 1. An erosion morphological filter[49] is applied to clean the boundaries of objects. In brief, pixels straddling a boundary between objects and the background have an average intensity value that falls between the high values of the background and the low values of the sample, producing what is often referred to as an edge effect. The erosion filter is used to replace the 1 with a 0 in pixels that have three neighbors with a value of 0, the result being the removal of a line at the edge of the object and of small protrusions (less than two pixels in length) at each iteration. Many size and shapes of structural elements can be employed for the erosion as long as the element selected preserves the centers of mass of the object. The watershed filter is applied once or twice depending on the sample image and is used to separate touching granules by performing a single cut for particles within the binary image with convexity (shape parameter) smaller than 0.94. Convexity is defined as the ratio of the area of the object and the convex area of the object. This is where preserving the center of mass for the objects affects the results; if the erosion biases the mask toward one side of all objects, the spectra of that edge, which may be noisy and distorted at the boundary, will affect the perceived chemistry of the objects. Finally, all granules touching the border of the image are excluded because only a fraction of the whole granule is measured, which would generate erroneous shape characteristics. Size-independent shape information, namely, elongation, circularity, and convexity, are calculated based on this image and the binary image is used as a mask for its corresponding chemical image (the PCA scores image of the sample).

Once all scores images and masks are created, it is possible to manipulate the data one more time to compare the chemical signatures, either average chemical characteristics per granules or all individual pixels of each

Author's personal copy

granule, with the shape parameters measured for each granule. The first data set of information about the sample, the result of PCA, is a chemical description of the samples. In many cases, only chemical composition is needed to discriminate between samples and to relate to quality parameters. When a sample is made up of many units (like granules making up a dosage form), the chemistry per unit and the proportions of units presenting various chemistries, that is, the distribution of the chemical composition, may be related to the common quality parameter (like active ingredient dosage or dissolution profile). This is the information generated in the average score vector per granule. We look into the shape parameters for the units to add information that may be correlated to the quality parameter and may help go one step further and point to the cause of the different results for the quality test. For the example of pharmaceutical granules, one could measure size and relate this to the dissolution in a relationship as follows: If the granules are of the expected shape and size, they should dissolve as expected. Once could go in the opposite direction and use chemistry only in a relationship that would read: If the distribution of the chemical composition is correct, then dissolution should be as expected. The difficulty often arises in establishing the thresholds that characterize a good distribution. **Figure 13(a)** shows the distribution of the chemical composition (score for PC 4) for three sets of granules that arise from lots with known differences for a particular quality parameter. The set represented in black appears somewhat different and this is quantitatively measured by the difference in the mean and skew of the distributions. In many situations, thresholds established on the basis of these distributions could be used to predict the quality parameter response of a new data set.

Since the data were acquired using an imaging modality, it appears wasteful to discard the image information. **Figure 13(b)** shows the loading vector of an LDA performed using a library comprising ~40 granules from each lot. LDA could be used as a data reduction method, but we use it here to investigate the discriminating power of the variables available.[50] The information used is a vector including both chemical, in the form of the

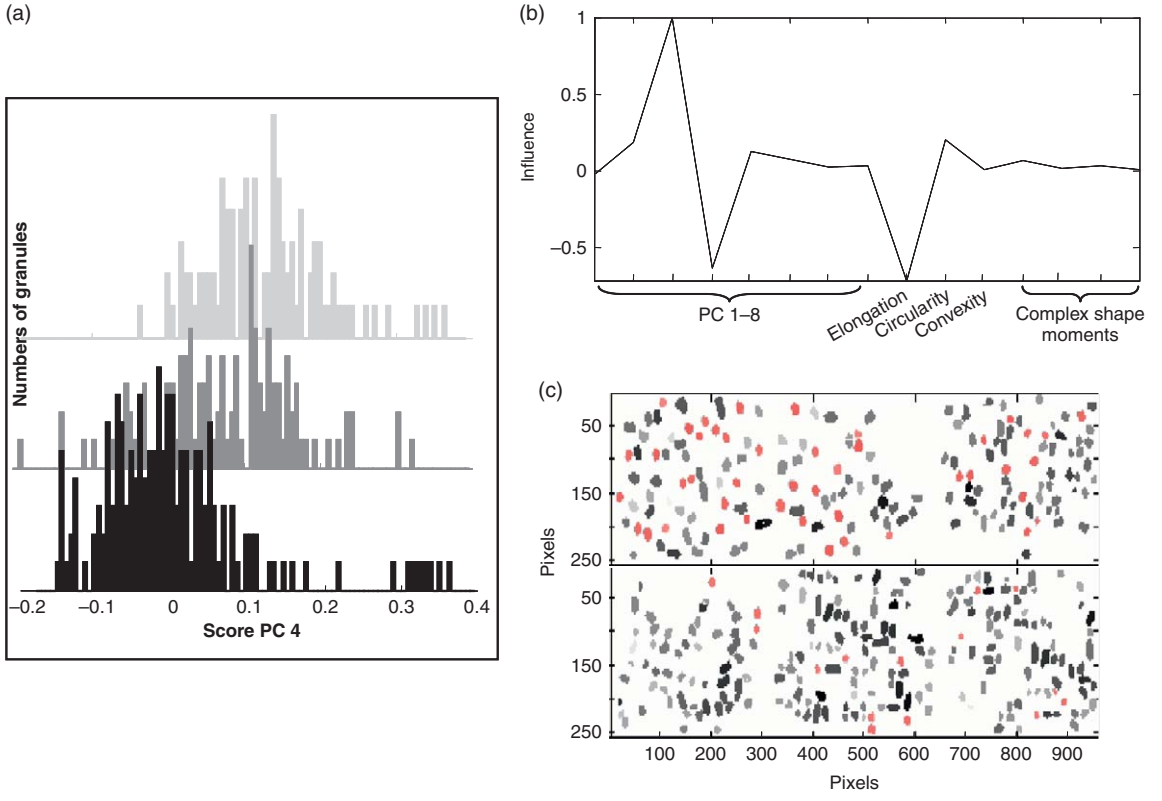


Figure 13 (a) Histogram plot of scores for PC 4. (b) Linear discriminant analysis loading vector 2. (c) False-color image of the granules passing the AND logical test superimposed on the gray-scale image of the score for PC 4 for granules from lots showing different quality attributes.

Author's personal copy

first eight scores of the PCA, and shape information, namely, elongation, circularity, convexity, and four complex shape moments. The LDA loading vector indicates that the score for PCs 3 and 4 and the elongation have better discriminating power. Both score for PC 4 and elongation also showed greater influence in the first loading vector, so they were combined to characterize the three lots. **Figure 13(c)** shows images from subsets of two of these lots. The granules identified in red satisfied both conditions of presenting a high score for PC 4 and having a small elongation (i.e., being rounder granules). Quantitatively, 26% of granules in the upper sample satisfy both criteria compared to 7% in the lower sample. This may be indicative of a relationship between the chemistry expressed in the fourth PC and elongation and the quality attribute that separated the lots, that is, the fact that one passed dissolution and the other failed. Accessing size, shape, and chemical information by both chemometric analysis of the spectroscopic information measured at each pixel and image analysis of the chemically contrasted image makes more efficient use of chemical imaging data. The relationships between chemistry, shape and size parameters, and specific quality parameters are often not direct ones and calibrations must be developed to establish the appropriate links.

4.06.5 Conclusion

Spectroscopic imaging, in particular chemical imaging, has rapidly developed over the past decade into a technology that is deployed in numerous fields of endeavor. In this particular section, we have described a few applications in the areas of pharmaceutical biological sciences; both have proven to be fruitful areas for the exploitation of this particular technology. One of the reasons for this is the significant spatial/chemical heterogeneity that exists in both these types of systems, where structure–function relationships drive performance. On the contrary, because of the complexity of the data, this presents challenges for the traditional data mining methodologies. We have shown how one may add a completely new multivariate descriptor of the data, which relies on the shape and size of intrinsic 'objects' within a complex chemical system as an additional discriminator. We consider this to be one of the new important steps forward in the chemometric analysis of these complex systems.

References

1. Baeten, V.; Dardenne, P. Applications of Near-Infrared Imaging for Monitoring Agricultural Food and Feed Products. In *Spectrochemical Analysis Using Infrared Multichannel Detectors*; Bhargava, R., Levin, I. W., Eds.; Blackwell Publishing Ltd.: Oxford, UK, 2005.
2. Grahn, H. F.; Geladi, P., Ed. *Techniques and Applications of Hyperspectral Image Analysis*; John Wiley & Sons Ltd: Chichester, UK, 2005.
3. Lyon, R. C.; Lester, D. S.; Lewis, E. N.; Lee, E.; Yu, L. X.; Jefferson, E. H.; Hussain, A. S. Near-Infrared Spectral Imaging for Quality Assurance of Pharmaceutical Products: Analysis of Tablets to Assess Powder Blend Homogeneity. *AAPS PharmSciTech* **2002**, *3* (3) article 17.
4. Reich, G. Near-Infrared Spectroscopy and Imaging: Basic Principles and Pharmaceutical Applications. *Adv. Drug Deliv. Rev.* **2005**, *57*, 1109–1143.
5. Lu, R. Detection of Bruises on Apples Using Near-Infrared Hyperspectral Imaging. *Trans. ASAE* **2003**, *46* (2), 523–530.
6. Mehl, P. M.; Chao, K.; Kim, M.; Chen, Y. R. Detection of Defects on Selected Apple Cultivars Using Hyperspectral and Multispectral Image Analysis. *Appl. Eng. Agric.* **2002**, *18* (2), 219–226.
7. Fernández Pierna, J. A.; Baeten, V.; Michotte Renier, A.; Cogdill, R. P.; Dardenne, P. Combination of Support Vector Machines (SVM) and Near Infrared (NIR) Imaging Spectroscopy for the Detection of Meat and Bone Meat (MBM) in Compound Feeds. *J. Chemom.* **2004**, *18*, 341–349.
8. Fernández Pierna, J. A.; Baeten, V.; Dardenne, P. Screening of Compound Feeds Using NIR Hyperspectral Data. *Chemom. Intell. Lab. Syst.* **2006**, *84*, 114–118.
9. Veronin, M. A.; Youan, B.-B. C. Magic Bullet Gone Astray: Medications and the Internet. *Science* **2004**, *305*, 481.
10. Dubois, J.; Wolff, J.-C.; Warrack, J. K.; Schoppelrei, J.; Lewis, E. N. NIR Chemical Imaging for Counterfeit Pharmaceutical Products Analysis. *Spectroscopy* **2007**, *22* (2), 40–50.
11. Lee, E.; Huang, W. X.; Chen, P.; Lewis, E. N.; Vivilecchia, R. V. High-Throughput Analysis of Pharmaceutical Tablet Content Uniformity by Near-Infrared Chemical Imaging. *Spectroscopy* **2005**, *21* (11), 25–32.
12. Lewis, E. N.; Schoppelrei, J.; Lee, E. Near-Infrared Chemical Imaging and the PAT Initiative. *Spectroscopy* **2004**, *19* (4), 22–31.
13. Tsai, F.; Lin, E. K.; Yoshino, K. Spectrally Segmented Principal Component Analysis of Hyperspectral Imagery for Mapping Invasive Plant Species. *Int. J. Remote Sens.* **2007**, *28* (5), 1023–1039.

Author's personal copy

14. Savitzky, A.; Golay, M. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639.
15. Dardenne, P.; Cowe, I.; Berzaghi, P.; Flinn, P. C.; Lagerholm, M.; Shenk, J. S.; Westerhaus, M. Standardisation of Near Infrared Instruments, Influence of the Calibration Methods and the Size of the Cloning Set. In *Near Infrared Spectroscopy: Proceedings of the 10th International Conference*; Davies, A. M. C., Cho, R. K., Eds.; NIR Publications: Norwich, UK, 2001, pp 23–28.
16. Wang, Y.; Veltkamp, D.; Kowalski, B. Multivariate Instrument Standardization. *Anal. Chem.* **1991**, *63*, 2750–2756.
17. Wang, Y.; Kowalski, B. Calibration Transfer and Measurement Stability of Near-Infrared Spectrometers. *Appl. Spectrosc.* **1992**, *46*, 764–771.
18. Kubelka, P.; Munk, F. Ein Beitrag zur Optik der Far-banstriche. *Z. Tech. Physik* **1931**, *12*, 593–604.
19. Hopkins, D. Derivatives in Spectroscopy. *Near Infrared Anal.* **2001**, *2*, 1–13.
20. Giesbrecht, F.; McClure, W.; Hamid, A. The Use of Trigonometric Polynomials to Approximate Visible and Near Infrared Spectra of Agricultural Products. *Appl. Spectrosc.* **1981**, *35*, 210–214.
21. Norris, K.; Williams, P. Optimization of Mathematical Treatments of Raw Near-Infrared Signal in the Measurement of Protein in Hard Red Spring Wheat, I. Influence of Particle Size. *Cereal Chem.* **1984**, *61*, 158–165.
22. Martens, H.; Jensen, S.; Geladi, P. Multivariate Linearity Transformations for Near-Infrared Spectrometry. In *Proceedings of the Nordic Symposium on Applied Statistics*; Christie, O., Ed.; Stokkand Forlag: Stavanger, 1983; pp 205–233.
23. Geladi, P.; McDougall, D.; Martens, H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Appl. Spectrosc.* **1985**, *39*, 491–500.
24. Isaksson, T.; Kowalski, B. Piece-Wise Multiplicative Scatter Correction Applied to Near-Infrared Diffuse Transmittance Data from Meat Products. *Appl. Spectrosc.* **1993**, *47*, 702–709.
25. Barnes, R.; Dhanoa, M.; Lister, S. Standard Normal Variate Transformation and Detrending of Near Infrared Diffuse Reflectance. *Appl. Spectrosc.* **1989**, *43*, 772–777.
26. Dhanoa, M.; Lister, S.; Sanderson, R.; Barnes, R. The Link between Multiplicative Scatter Correction (MSC) and Standard Normal Variate (SNV) Transformations of NIR Spectra. *J. Near Infrared Spectrosc.* **1994**, *2*, 43–47.
27. Massart, D. L. M.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, J. P.; Smeyers-Verbeke, J. *Chemometrics: A Textbook*; Elsevier: Amsterdam, 1988; Vol. 2.
28. Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Analysis of Measurement Tables In Handbook of Chemometrics and Qualimetrics: Part B*; Vandeginste, B. G. M., Rutan, S. C., Eds.; Elsevier: Amsterdam, The Netherlands, 1998; Chapter 31, pp 87–160.
29. Burger, J.; Geladi, P. Hyperspectral NIR Image Regression Part I: Calibration and Correction. *J. Chemom.* **2005**, *19*, 355–363.
30. Cheng, X.; Chen, Y. R.; Tao, Y.; Wang, C. Y.; Kim, M. S.; Lefcourt, A. M. A Novel Integrated PCA and FLD Method on Hyperspectral Image Feature Extraction for Cucumber Chilling Damage Inspection. *Trans. ASAE* **2004**, *47* (4), 1313–1320.
31. Martens, H.; Naes, T. *Multivariate Calibration*, 2nd ed.; Wiley: Chichester, UK, 1989; Vol. 1.
32. Despagne, F.; Massart, D. L. *Tutorial Review: Neural Networks in Multivariate Calibration Analyst* **1998**, *123*, 157R–178R.
33. Codgill, R. P.; Dardenne, P. Least-Squares Support Vector Machines for Chemometrics: An Introduction and Evaluation. *J. Near Infrared Spectrosc.* **2004**, *12* (1), 93–100.
34. Dardenne, P.; Fernández Pierna, J. A. Soil Parameter Quantification by NIRS as a Chemometric Challenge at 'Chimiométrie 2006'. *Chemom. Intell. Lab. Syst.* **2008**, *91*, 94–98.
35. De la Haba, M. J.; Fernández Pierna, J. A.; Fumière, O.; Garrido-Varo, A.; Guerrero, J. E.; Pérez-Marín, D. C.; Dardenne, P.; Baeten, V. Discrimination of the Class Origin of Bones Present in the Sediment Fraction of Animal By-Products Using Near Infrared Microscopy (NIRM). In *Proceedings of the NIR2005;* NIR Publications: New Zealand, 2005. *Near Infrared Spectroscopy: Proceedings of the 12th International Conference*; Burling-Claridge, G. R., Holroyd, S. E., Sumner, R. M. W., Eds.; New Zealand Near Infrared Spectroscopy Society Incorporated, 2006.
36. Fernández Delgado, M.; Carrion Pardo, P.; Cernadas García, E.; Gálvez Gálvez, J. F.; Sá Otero, P. Improved Classification of Pollen Texture Images Using SVM and MLP.; In *3rd IASTED International Conference on Visualization, Imaging and Image Processing (VIIP2003)*; Hamza, M. H., Ed.; Benidorm, Spain, 2003; Vol. 2, pp 686–691.
37. Polder, G.; van der Heijden, G. W. A. M.; Young, I. T. Hyperspectral Image Analysis for Measuring Ripeness of Tomatoes. *Presented at the 2000 ASAE International Meeting*, *Paper Number 003089*, Milwaukee, WI, July 9–12, 2000.
38. Polder, G.; van der Heijden, G. W. A. M.; Young, I. T. Spectral Image Analysis for Measuring Ripeness of Tomatoes. *Trans. ASAE* **2002**, *45* (4), 1155–1161.
39. Polder, G.; van der Heijden, G. W. A. M.; Young, I. T. Tomato Sorting Using Independent Component Analysis on Spectral Images. *Real-Time Imaging* **2003**, *9*, 253–259.
40. Peirs, A.; Scheerlinck, N.; De Baerdemaeker, J.; Nicolaï, B. M. Starch Index Determination of Apple Fruit by Means of a Hyperspectral Near Infrared Reflectance Imaging System. *J. Near Infrared Spectrosc.* **2003**, *11*, 379–389.
41. Lawrence, K. C.; Windham, W. R.; Park, B.; Buhr, R. J. A Hyperspectral Imaging System for Identification of Faecal and Ingesta Contamination on Poultry Carcasses. *J. Near Infrared Spectrosc.* **2003**, *11*, 269–281.
42. Chao, K.; Mehl, P. M.; Chen, Y. R. Use of Hyper- and Multi-Spectral Imaging for Detection of Chicken Skin Tumors. *Appl. Eng. Agric.* **2002**, *18* (1), 113–119.
43. Mehl, P. M.; Chen, Y. R.; Kim, M. S.; Chan, D. E. Development of Hyperspectral Imaging Technique for the Detection of Apple Surface Defects and Contaminations. *J. Food Eng.* **2004**, *61* (1), 67–81.
44. Smail, V. W.; Fritz, A. K.; Wetzel, D. L. Chemical Imaging of Intact Seeds with NIR Focal Plane Array Assists Plant Breeding. *Vib. Spectrosc.* **2006**, *42*, 215–221.
45. Weinstock, B. A.; Janni, J.; Hagen, L.; Wright, S. Prediction of Oil and Oleic Acid Concentrations in Individual Corn (*Zea mays* L.) Kernels Using Near-Infrared Reflectance Hyperspectral Imaging and Multivariate Analysis. *Appl. Spectrosc.* **2006**, *60* (1), 9–16.
46. Codgill, R. P.; Hurburgh, C. R.; Rippke, G. R. Single Kernel Maize Analysis by Near-Infrared Hyperspectral Imaging. *Trans. ASABE* **2004**, *47* (1), 311–320.
47. Stevermer, S. W.; Steward, B. L.; Codgill, R. P.; Hurburgh, C. R. Automated Sorting and Single Kernel Analysis by Near-Infrared Hyperspectral Imaging. *Presented at the 2003 ASAE International Meeting, Paper Number 036159*, Las Vegas, NV, USA, July 27–30; American Society of Agricultural and Biological Engineers: St. Joseph, MI, USA, 2003.

Author's personal copy

48. Gorretta, N.; Roger, J. M.; Aubert, M.; Bellon-Maurel, V.; Campan, F.; Roumet, P. Determining Vitreousness of Durum Wheat Kernels Using Near Infrared Hyperspectral Imaging. *J. Near Infrared Spectrosc.* **2006**, *14*, 231–239.
49. Russ, J. C. *The Image Processing Handbook*, 3rd ed.; CRC Press LLC: Boca Raton: Florida 1999; 771 p.
50. Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. *A User Friendly Guide to Multivariate Calibration and Classification*; NIR Publications: Chichester, UK, 2002.

Author's personal copy

Biographical Sketches

Juan Antonio Fernandez Pierna received his Ph.D. in Pharmaceutical Sciences (Chemometrics) from the Analytical Chemistry Department of the Vrije Universiteit Brussel (Professor D. L. Massart) with a thesis entitled 'Improvements in the multivariate calibration processes'. Since 2003, he has been working as a research assistant at the CRA-W where he has been focusing on the statistical treatment of data, the application of chemometrics, and the validation of chemometric methods. Between October 2004 and September 2005, he collaborated with the CRA-W while working as scientific collaborator for F.N.R.S. During the past few years, his main work area is related to the application of imaging to agricultural products and the extraction of all the possible chemical and statistical information from the data as well as the optimization and validation of the different procedures. He is author or coauthor of four book chapters and more than 25 articles mainly related to the statistical treatment of spectroscopic data (including homogeneity detection and uncertainty estimation) and imaging techniques. He is a member of the Belgian Chemometric Society and he was and is still involved in different EU projects: STRATFEED, TYPIC, TRACE, FEED SAFETY, SAFEED-PAP.

Vincent Baeten is part of the statutory staff and is project leader at the Quality Department of Agricultural Products of CRA-W since 1999. He got his Engineering degree (1993) and Ph.D. (1998) in Agricultural Sciences from the Catholic University of Louvain. Since 2006, he leads the group of spectroscopy and chemometry at CRA-W, which develops alternative physical methods for the quality, the authentication, and the traceability of agrofood products, and for feed safety. He has more than 13 years of experience on European projects. During 1996–98, he was awarded with a Marie Curie Fellowship at the Instituto de la Grasa of Seville (Spain)

Author's personal copy

where he was involved in demonstrating the potential of alternative techniques for the detection and quantification of food product adulteration. Since 1999, he has been participating in the improvement of reference methods and the development of alternative methods for the detection of constituents of animal origin in feed. He has a wide expertise in the management and administration of European projects. In the last 5 years he has participated in several projects dealing with food and feed safety and has been closely involved in the coordination of the STRATFEED project. He has more than 25 publications related to the MBM detection problems. He is a member of the scientific committee of the Belgian Federal Agency for the Safety of the Food Chain and is member of the scientific committee of the TRACE IP project. Since July 2006, he is the director of the Community Reference Laboratory for animal proteins in feedingstuffs (http://www.crl.cra.wallonie.be). He is also the coordinator of the European SAFEED-PAP project, which aims at the development of analytical methods for the species detection of processed animal proteins (http://safeedpap.feedsafety.org).

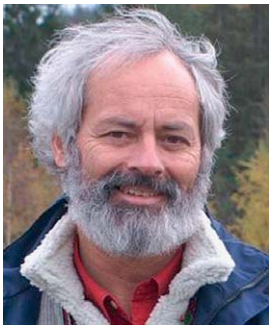
Pierre Dardenne is Agronomy Engineer from Gembloux Agricultural University. In 1980, he was employed by the Walloon Agricultural Research Centre (CRA-W) to lead researches in the near-infrared spectroscopy (NIRS) domain, which he has not left since. In 1991, he got his Ph.D. in Agronomical Sciences at Gembloux Agricultural University in the field of spectroscopy and chemometrics. During the past 20 years, the agronomical and agroindustrial applications developed in NIRS are extremely wide. Pierre Dardenne was and is still involved in many European programs. He is currently Head of the Quality Department of the Walloon Agricultural Research Centre and continues to manage the section entitled 'Application of the physical methods of non destructive analyses for the global management of the quality of the primary and transformed products: development of softwares and management of networks'. He is leading other groups of scientists working on feed and food chemical composition, contaminants (heavy metals, antibiotics), milk microbiology, and GMO detection. Authenticity, antifraud, and food safety are keywords in many research programs of his department. He is referee of *Journal of Near Infrared Spectroscopy* and chairman of NIR working group for the BIPEA (Bureau Interprofessionnel d'Etudes Analytiques, France). Pierre Dardenne received the Tomas Hirschfeld award in 2002 for his contribution to NIRS.

Author's personal copy

Janie Dubois received her Ph.D. from McGill University in Montreal, where she worked on biological applications of FTIR spectroscopy and imaging. In 2000, she joined the National Research Council Canada, Institute for Biodiagnostics, where she became a Research Officer with a specialization in imaging of infectious diseases. In 2003, she moved to the Joint Institute for Food Safety and Nutrition at the US Food and Drug Administration, where her work focused on NIR imaging as a rapid identification system for food-borne outbreaks. She has been with Spectral Dimensions, now Malvern Instruments, since 2005, where she was Senior Application Scientist and is now Product Manager for the Americas for the Analytical Imaging product line. Janie is also the secretary of ASTM subcommittee E13.10 on Molecular Imaging.

E. Neil Lewis received his Ph.D. in chemistry from the Polytechnic of Wales in the United Kingdom and did his postdoctoral fellowship at the National Institutes of Health (NIH) in the United States. He was tenured by the NIH in 1992 and held the position of Senior Biophysical Researcher. In 1999 he founded Spectral Dimensions, Inc., a company that develops hyper-spectral imaging systems. He has been at the forefront of the development of these technologies and has published more than 60 papers in the field. He has received numerous awards for his contributions including the Meggers Award in 1992, and again in 1994, presented by the Society for Applied Spectroscopy; the Heinrich award in 1995 presented by the Microbeam Analysis Society; and the Washington Academy of Sciences Outstanding Contribution to the Physical Sciences award in 1997. He is also the recipient of the 2004 Williams–Wright Award, awarded by the Coblentz Society to a person who has made significant contributions to the field of vibrational spectroscopy while working in the industry.

Author's personal copy



James Burger began his chemometrics career at the University of Washington (Bruce Kowalski), receiving a Master's degree in 1985. After many years of commercial, scientific, and business-related software development, he returned to Ph.D. studies in 2002 at Aalborg University, Esberg, Denmark (Kim Esbensen) and the Swedish University of Agricultural Sciences in Umea, Sweden (Paul Geladi), focusing on the extension and application of chemometric principles to hyperspectral imaging. He received his Ph.D. in 2006 with a thesis entitled 'Hyperspectral NIR Image Analysis: Data Exploration, Correction, and Regression'. James is the founder and director of two related businesses: BurgerMetrics, Inc., Shoreline, WA, USA, and BurgerMetrics SIA, Jelgava, Latvia, offering consulting and training in hyperspectral image analysis and NIR imaging systems. He continues to explore novel extensions of chemometrics and the development of interactive software to aid in the interpretation of hyperspectral image data. His interests include both exploratory laboratory systems and online process automation.