DOI: 10.58895/ksp/1000032143-1 erschienen in:
OCM 2013 - Optical Characterization of Materials - conference proceedings
DOI: 10.58895/ksp/1000032143 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000032143/

Spectral imaging in process analytics using chemometrics and first principles

S. Luckow-Markward, E. Ostertag, B. Boldrini,
W. Kessler and R. W. Kessler*

Reutlingen Research Institute, Process Analysis and Technology,
Reutlingen University, Alteburgstr. 150, 72762 Reutlingen
*corresponding author: rudolf.kessler@reutlingen-university.de

Abstract Optical spectroscopy is able to detect not only the chemical composition of the species by their wavelength specific absorption k but also the morphological feature through their wavelength dependent scattering s. In standard multivariate data analysis in hyperspectral imaging, the focus of the chemometric treatment of the data cube is given on the suppression of the unwanted perturbation of multiple scatter of photons. This paper describes an approach how to separate the morphological information s (scatter) from the chemical information k (absorption) using the radiative transfer equation or Kubelka Munk theory. When this "first principle spectroscopy" is integrated into most modern multivariate data analysis like multivariate curve resolution (MCR), causality is obtained between the spectral data and response variables like the concentration of an active pharmaceutical ingredient in a tablet. With this approach, the spatially resolved calculated k- and s-distribution of an aspirin particle in cellactose is shown. The optical set up for real life spectral imaging in industry is discussed and examples of spectral images to control the thickness of thin films on metals, the distribution of a resin on a wood chip and the differentiation of hard and soft maize kernels are shown.

1 Introduction

Spectral imaging or chemical imaging is the determination of the chemical identity of species and the visualization of their distribution. Optical spectroscopy is able to detect not only the chemical composition

of the species by their wavelength specific absorption but also the morphological feature through their wavelength dependent scattering [1]. Figure 1.1 visualizes the integration of the chemical and morphological information into an image. The most common tool to measure the distribution of components in a solid particulate system is spectroscopy, thus chemical imaging is also labeled as spectral or hyperspectral imaging [2].

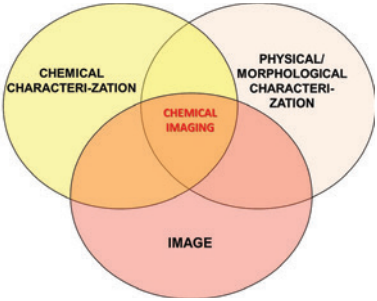

Figure 1.1: Visualization of spectral imaging.

The laterally resolved spectroscopy produces a three-dimensional data cube with two local axes, x and y, and a spectral z-axis with usually the intensity of the reflectance at different wavelengths λ. Figure 1.2 illustrates the essential differences of the techniques used to measure spectral images.

In the so-called whiskbroom imaging (= mapping), defined object areas or the entire object is measured point-by-point. This type of imaging is very flexible in relation to the object and the grid size and generally requires only a single detector; such as a monochromating element with a photomultiplier tube or a diode array. A staring imager (= imaging) takes two-dimensional images in a series at different wavelengths. A prerequisite for this technique is that the object must remain stationary during the measurement ("stop motion"), thus only atline applications can be realized [3].

In pushbroom imaging (= line scanning) the object is imaged along the y-axis using the line-scan method and is recorded in full through the movement of the object in the y-direction. Through an entrance gap in the spectrograph (x-spatial dimension), the light is routed usually into a

prism-grating-prism optical arrangement and then spectrally resolved onto the second dimension of the camera. The second spatial dimension (y) is achieved through the movement of the object. In contrast to whiskbroom and staring imaging, the pushbroom system is fully online/inline capable, where for each line, under time-defined conditions, images can be generated and evaluated.

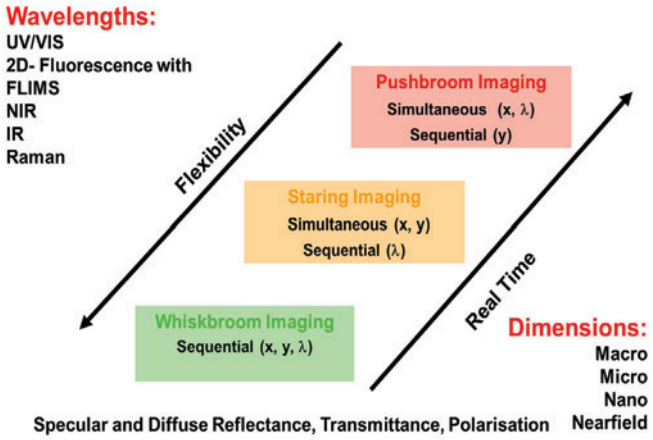

Figure 1.2: Taxonomy of spectral imaging techniques [1,3].

The objective of this paper is first to introduce fundamental principles into the evaluation of spectral imaging data with the objective to separate the chemical information from the morphology of the scattering system and then to show how this improves the robustness of the multivariate data analysis. Finally some examples of the optical setup of spectral imaging devices for inline control will be presented.

2 Integrating "first principles" into spectral imaging: separate absorption from scatter

Dispersions, emulsions or solids like powders show the wavelength dependent superposition of the scatter (s) and absorption (k) of light. In standard multivariate data analysis in process analytical technology, the

focus of the chemometric treatment of the spectroscopic data is given on the suppression of the unwanted perturbation of multiple scatter of photons. A better approach may be to extract not only the chemical information but also to use the morphological information from the spectra. This approach using first principles exploits the full potential of the spectral information rather than to eliminate the morphological features. One of the most appropriate theories to describe multiple scattering and absorption in opaque systems is the radiative transfer equation (RTE). The s- and the k-spectrum can be calculated using the (inverse) Monte Carlo simulation from the superposed spectra. The approach of Kubelka and Munk (K-M) is the simplified solution of the radiative transfer theory. In this case the diffuse reflectance and transmittance of a sample with defined thickness is described by the scattering effect s and and the absorption effect k. Thus at least two independent measurements are needed to separate s and k from the measured spectra [1, 3, 4]. S and k can be determined independently measuring just spectra in diffuse reflectance of two samples with known different layer thicknesses, or measuring one sample in diffuse reflectance and diffuse transmittance. After solving the equations, two spectra are obtained which more or less solely represent the spectrum of scatter and the unperturbed absorption spectrum of the component [3].

Figure 1.3 shows as an example the spectra of an Aspirin tablet measured in transmittance and reflectance and the calculated scatter and absorption spectra using the Kubelka Munk theory.

The scatter and absorption cross sections determine the penetration depth of the photons and therefore the information depth ("scale of scrutiny") [5]. Specular reflected light of the surface may produce also spectral artefacts [3]. However, these artifacts can easily be removed using parallel and crossed polarizer during measurement. For inline applications, in most cases diffuse illumination of the object is sufficient to minimize specular reflected light.

3 Integrating "chemometrics" into spectral imaging: reduce the data cube

Principle component analysis and data pretreatment For analysts used to interpret a single spectrum or a few averaged spectra for each

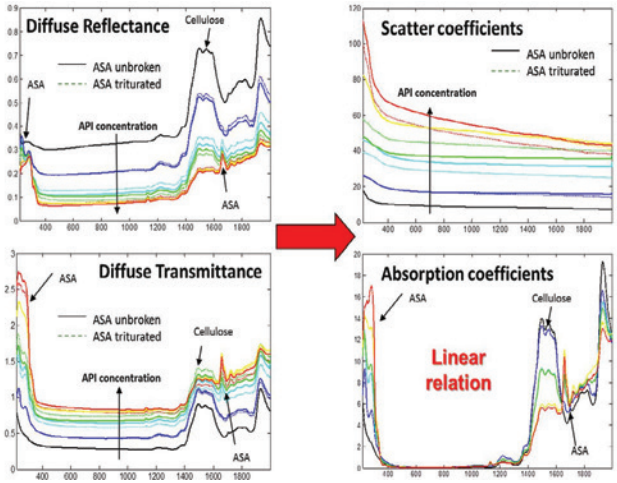

Figure 1.3: Spectra of Aspirin (ASA) tablets measured in diffuse reflectance and diffuse transmittance (left) as well as the resulting calculated scatter and absorption spectra using Kubelka Munk theory. The ASA particles show different particle size distribution (unbroken app. 80 μm, triturated app. 40 μm).

sample, the idea of getting hundreds or thousands of spectra which are spatially resolved, is confusing and may be even hindering to use spectral imaging for quality or process control. Therefore the implementation of chemometric tools is very advisable when analyzing such large amounts of data. Chemometrics offers the possibility to extract the relevant information from the full chemical imaging data set instead of using single-wavelength channels only. And additionally, chemometrics reduces this relevant information into one or a few quality defining parameters by applying either multivariate classification or regression models to the hyperspectral data.

A very effective data reduction is achieved with the principal component analysis. The PCA gives a compressed representation of the image that retains all of the relevant information in the spectral dimension [6,7]. Often three to five principal components capture most of the relevant information of several hundred spectral pixels. The principal components are linear combinations of the original spectral variables.

PCA is a chemometric method, which decomposes a two- or multi-dimensional data table $\mathbf{X}$ into a bilinear model of latent variables, the so-called principal components, according to the following expression:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}$$

where $\mathbf{T}$ is the scores matrix and $\mathbf{P}^T$ the transposed loadings matrix. The matrix $\mathbf{E}$ is the residual matrix and accounts for the experimental error (noise), which is not part of the model. The principal components are calculated so that they explain as much variance of the data as possible. The first principal component captures most of the variance in the data set. This information is then removed from the data and the next principal component is calculated, which again captures most of the remaining variance, this continues until a predefined stopping criteria of too little variance explained by a new component is fulfilled. All principal components are linearly independent, that means there is no correlation among them and they can therefore serve as a new coordinate system with reduced dimensions. An image spectrum can have hundreds or even thousands of pixels, but the relevant information can be contained in a very small number of principal components and each spectrum can be described by the first few scores of the principal component model. A picture of 256 lateral pixels in x- and y- direction and 1000 pixels in the spectral dimension is then reduced from to e.g. 3 latent variables: from $256 \times 256 \times 1000$ down to $256 \times 256 \times 3$, more than 300 times less.

The state of the art approach in chemometrics of spectroscopic data from particulate systems is to exclude the scattering information from the spectral features by data pre-treatment procedures like standard normal variate (SNV), multiplicative scattering correction ((extended) MSC) or orthogonal signal correction (OSC) to obtain unperturbed quantitative information [3, 6, 7]. In this case, the information scatter is often regarded as unwanted and therefore eliminated instead of being used as supplementary information on the morphology of the substrate. A better approach is to integrate the information morphology into the model as described in the previous chapter.

Multivariate curve resolution A more advanced technique in comparison to PCA is multivariate curve resolution (MCR). The major reason of an increasing interest in multivariate curve resolution (MCR) solved

by alternating least squares (MCR-ALS) is its ability to extract from a complex spectral feature a) the number of involved components b) to attribute the resulting spectra to chemical compounds and c) to quantify the individuel spectral contributions. Thus interpretable loadings which represent spectra are obtained. In addition, MCR provides a perfect means to integrate knowledge into the chemometric approach. E.g., known spectra of the components can be integrated into the model or e.g. the s- and k- "pure" spectra of the system under investigation [8].

Example Figure 1.4 shows the results using the unperturbed k - absorption spectra from figure 1.3 for a quantitative calculation based on Lambert-Beer's law with a single wavelength or using multivariate partial least square analysis (PLS).

The separated k-spectrum shown in the centre of the chart is comparable to the spectrum in solution. In the visible range no absorption is measured as it should be the case for transparent materials. It is important to emphasize that the absorbance of aspirin is more pronounced in the UV than in the NIR region and increases with increasing concentration. The spectral features from 1400nm – 1600nm in the NIR spectra can be attributed to the excipient cellulose and decreases with increasing API content. It is remarkable that only one latent variable is necessary in PLS calculations to quantify the API content due to the "first principle" separation of the scatter from the spectrum. This increases the robustness of the chemometric model. Standard procedures in NIR spectroscopy often need many more principal components to adjust for the nonlinearity of the scatter in the spectral information. Alternatively, single wavelengths can be used to calculate the concentration of aspirin in a tablet just by Lambert-Beer's law.

The same approach can be applied in the multidimensional space of spectral imaging. Figure 1.4 also shows the spatially resolved calculated k- and s-distribution of an aspirin particle in cellactose in the UV and NIR using the Kubelka Munk approach. As can be seen, the main scatter is observed directly at the phase boundary of the particle and is much higher at shorter wavelengths. The combined effect of scatter and absorption may even hinder the penetration of the photons into the particle. In this case quantitative analysis of the composition is a challenge.

When this "first principle spectroscopy" is integrated into most mod-

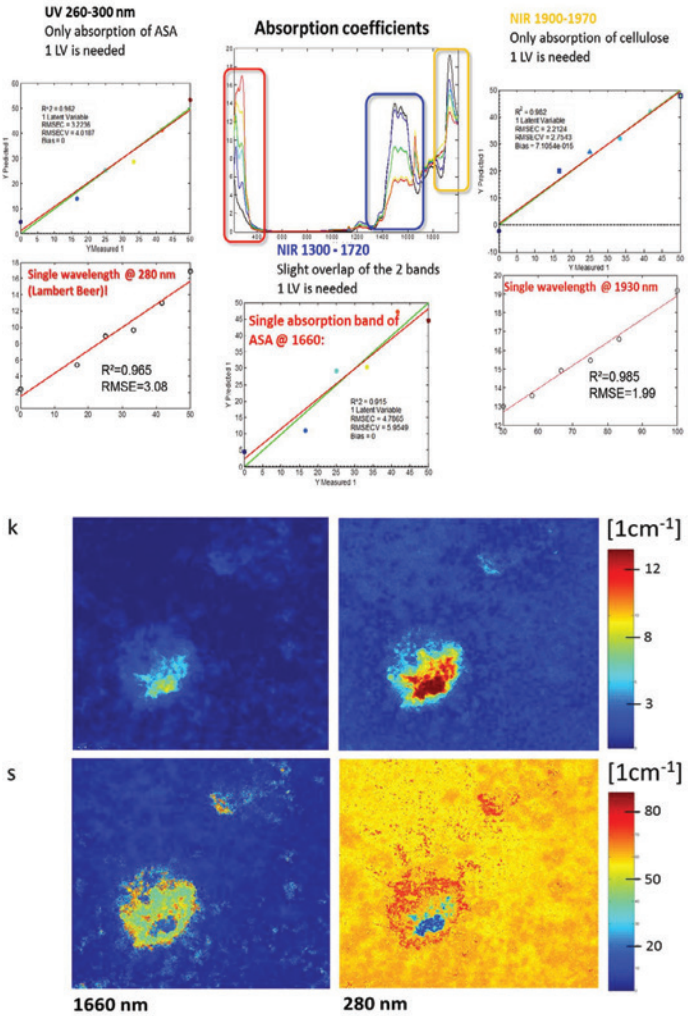

Figure 1.4: Top: quantitative calculation of the API concentration of an aspirin (ASA) tablet using the unperturbed absorption spectra (details see text), lower part: calculated s- and k- spatial distribution of an aspirin particle in cellactose measured at 1600nm and 280nm. The data are extracted from transmittance and reflectance measurements using the Kubelka Munk approach.

ern MVA methods like Multivariate Curve Resolution (MCR), causality is obtained between the data and response variables. This closes the gap between empirical correlative and first principle process information [6,8,9].

4 Inline control optical set up: some selected examples

Since 2002, the food and drug administration (FDA) has strongly encouraged the process analytical technique (PAT) for a better understanding of the process and to achieve a higher control of the pharmaceutical manufacturing process [9–11]. An ideal situation would be to control inline 100% of the tablet and the particle size as well as the homogeneous distribution of the active ingredient in the excipient. In the literature, there are numerous methods which use NIR, IR, Terahertz and Raman spectroscopic imaging [1–3,11]. Figure 1.5a shows the sketch of a pushbroom imaging device as described in the previous chapters and a set of optical arrangements for inline process control (figure 1.5).

Illumination e.g. at 45° (with respect to the macroscopic surface) and detection of the reflected light at 45° (45R45) measures mainly the specular reflected light. With this arrangement the spectral interference pattern is measured and from these measurements the thickness of e.g. an oxide film on a glass or metal substrate can be calculated using the Fresnel equations [1,12]. When pushbroom imagers with a high pixel number are used, the film thickness along the imaging line can be detected. An example how different the distribution of the oxide film thickness on aluminium can be is shown in figure 1.5c (left).

Particulate systems are commonly measured in diffuse reflectance. Here an optical arrangement with an illumination at 45° (may be from both sides) and detection at 0° (45R0) is favorable. The example in figure 1.5c (middle) shows the PCA analysis of the distribution of a resin on a wood chip [3]. However, when high specular reflectance of the object is observed together with a curvature, strong specular reflectance often superposes the diffuse reflectance with an optical setup 45R0. These spectral artifacts can hardly be mathematically eliminated. A solution is to illuminate the object with diffuse light (e.g. dR0) or a more complex arrangement by illumination with diffuse light and detection with an integrating sphere. Some possible set ups are explained in [3]. An

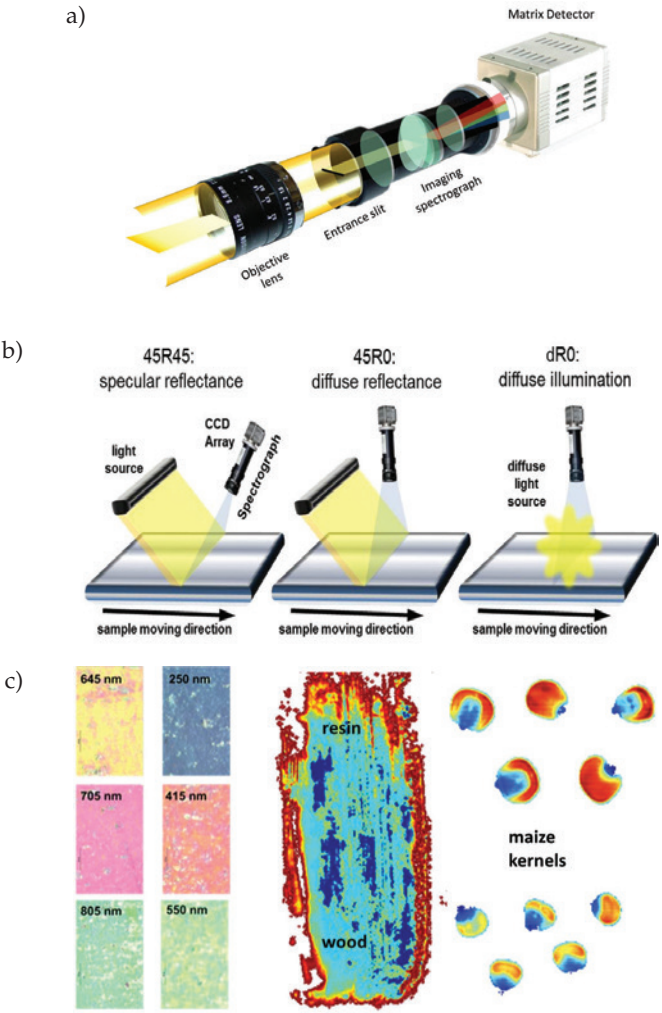

Figure 1.5: Sketch of a pushbroom imaging device (a) and a set of optical arrangements for inline process control (b). Bottom (c): examples of spectral images to control the thickness of thin films (left), the distribution of a resin on a wood chip (middle) and the differentiation of hard (bottom) and soft (top) maize kernels.

example for a typical application for dR0 is shown in figure 1.5c (right) where maize kernels of different origin are measured with a pushbroom imager. The figure shows the result of a PCA and false color representation.

5 Outlook

Focus in the pharmaceutical industry is given mainly on three different uses: blend uniformity of powders and tablets, composition and morphological features of coated tablets and granules, spatial changes during hydration, degradation and active release. Counterfeit pharmaceutical products are a real threat to the health of the patients. NIR chemical imaging provides a rapid method for detecting and comparing suspected counterfeit products without sample preparation. The advantage of imaging is that the discrimination of the tablets is not only caused by changes in the chemical composition, but also from its spatial distribution and texture of the tablet.

Online chemical imaging in agriculture is mainly remote sensing. Satellite or aerial remote sensing (RS) technology uses nowadays Pushbroom Imaging Technology in the Vis, s-NIR and NIR-range. Vegetation images show crop growth from planting through to harvest, changes as the season progresses and abnormalities such as weed patches, soil compaction, watering problems etc. This information can help the farmer make informed decisions about the most feasible solution. In food industry, numerous online controls are still made by human vision, especially for sorting bad looking products. Chemical imaging in food and agriculture can also be used to identify diseases, rot and contaminations by insects e.g. larvae.

Instead of using at each individual production step a single spectrometer, a pushbroom imager with attached fiber bundles on its slit allows individual control of the quality at every intermediate and final step. In this case, the pushbroom imager is used as a multipoint information source and can substitute a moving multiplexer.

Diffuse optical imaging (DOI) is a new emerging technique for functional imaging of biological tissues. It involves generating images using measurements in the visible or s-NIR-light scattered across large and thick tissues for e.g. detecting cancer.

A detailed description of the future trends in chemical imaging is given in [3, 11, 12].

References

1. R. W. Kessler, *Prozessanalytik: Strategien und Fallbeispiele aus der industriellen Praxis*. Wiley-VCH Weinheim, 2006.
2. R. Salzer and H. W. Siesler, *Infrared and Raman Spectroscopic Imaging*. Wiley-VCH Weinheim, 2009.
3. B. Boldrini, W. Kessler, K. Rebner, and R. W. Kessler, "Hyperspectral imaging: a review of best practice, performance and pitfalls for inline and online applications," *Journal of Near Infrared Spectroscopy*, vol. 20, pp. 438–508, 2012.
4. Z. Shi and C. A. Anderson, "Pharmaceutical applications of separation of absorption and scattering in near-infrared spectroscopy (NIRS)," *Journal of Pharmaceutical Sciences*, vol. 99, no. 12, pp. 4766–4783, 2010.
5. D. Oelkrug, M. Brun, K. Rebner, B. Boldrini, and R. W. Kessler, "Penetration of light into multiple scattering media: Model calculations and reflectance experiments. Part I: The axial transfer," *Appl Spectrosc*, vol. 66, pp. 934–943, 2012.
6. S. D. Brown, R. Tauler, and B. Walczak, *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*. Elsevier, Amsterdam, 2009.
7. W. Kessler, *Multivariate Datenanalyse: Für die Pharma-, Bio- und Prozessanalytik*. Wiley-VCH Weinheim, 2006.
8. W. Kessler, D. Oelkrug, and R. W. Kessler, "Using scattering and absorption spectra as MCR-hard model constraints for diffuse reflectance measurements of tablets," *Analytica Chimica Acta*, vol. 642, pp. 127–134, 2009.
9. U.S. Department of Health and Human Services, Food and Drug Administration, Guidance for Industry, "Process validation: General principles and practices," (Revison January 2011).
10. A. Bogomolov, "Multivariate process trajectories: Capture, resolution and analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 108, no. 1, pp. 49–63, 2011.
11. E. N. Lewis, J. W. Schoppelrei, L. Makein, L. H. Kidder, and E. Lee, "Near-infrared chemical imaging for product and process understanding," in *Process Analytical Technology*, K. Bakeev, Ed. Wiley, 2010, pp. 245–279.

12. A. Kandelbauer, M. Rahe, and R. W. Kessler, "Process control and quality assurance – industrial perspectives," in *Handbook of Biophotonics, Vol. 3: Photonics in Pharmaceutics, Bioanalysis and Environmental Research*. Wiley, 2012, pp. 1–69.